Vision and Perception
2022-2023

SAPIENZA
UNIVERSITÀ DI ROMA

General information

- Lectures and tutoring sessions

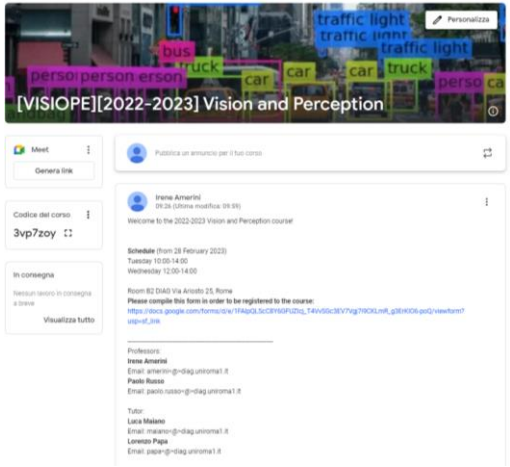
  Tuesday 10:00-14:00
  Wednesday 12:00-14:00

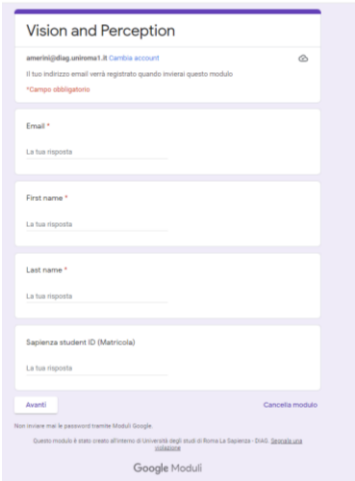
  In presence: room B2 DIAG, Via Ariosto 25, Rome

Registration to the course

- **Google Classroom** for slides, course material, announcements
  Subscribe now: code 3vp7zoy

- **Compile the form:**
  https://forms.gle/1PrQJXhVS2vrQDYw9





Instructors

- Prof. Irene Amerini amerini@diag.uniroma1.it

- Prof. Paolo Russo paolo.russo@diag.uniroma1.it

Teaching Assistants

- Luca Maiano maiano@diag.uniroma1.it
- Lorenzo Papa papa@diag.uniroma1.it

- In email messages, please use the prefix **[VISIOPE]** to the subject
- **Office hours** by appointment

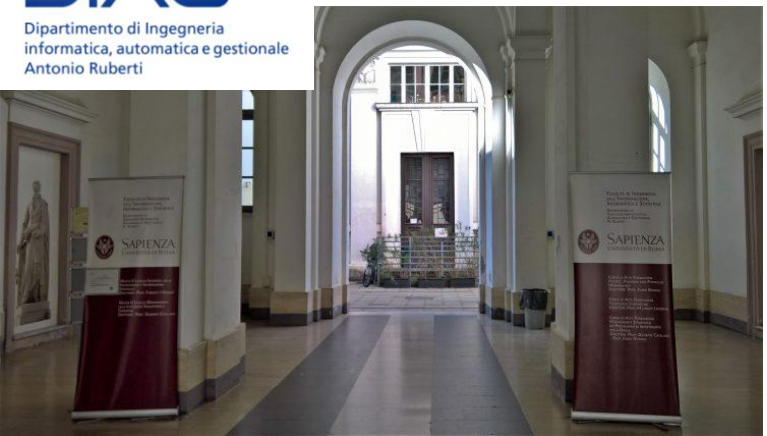
Research group ALCORLab

ALCORLab research laboratory at Department of Computer, Control, and Management Engineering A. Ruberti

Research at ALCORLab covers topics of Computer Vision, Pattern Recognition & Machine Learning, Artificial Intelligence, and Multimedia forensics applied to images and videos.

https://sites.google.com/diag.uniroma1.it/alcorlab-diag

DIAG
Dipartimento di Ingegneria informatica, automatica e gestionale Antonio Ruberti



Research Topics

Image and Video Understanding

- Deep learning algorithms for classification and detection (also on embedded devices)
- Multimodal analysis
- Image similarity and object re-identification

Multimedia Forensics and Security

- Tampering detection and deepfake detection
- Source identification
- Adversarial machine learning, security and robustness of deep learning algorithms

Perception

- Depth estimation
- Multichannel learning RGB +
- Semantic-segmentation / Panoptic-segmentation

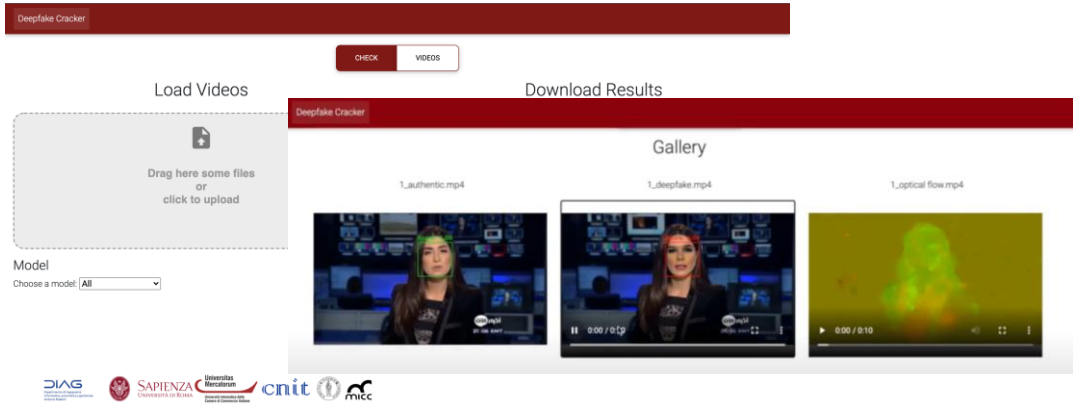
Resource Constraint

- Deep learning on resource constraint edge devices
- Monocular depth estimation on embedded devices

Master Thesis Computer Vision/Deep Learning

- Availability of **research thesis** at ALCORlab: some example but not limited to monocular depth estimation, deepfake detection, semantic forensics, multimodal disinformation detection, computer vision, adversarial attacks.

- **Thesis projects** in collaboration with startup, companies and research centers: some example but not limited to Automatic Licence Plate Recognition, Abnormal behaviour detection in cctv footage, People re identification, Earth Observation.



Overview of the course 1/2

1. Image Processing
Image Formation and Filtering
Feature Detection and Matching
Frequency analysis
Optical flow and video recognition

2. Multiview Geometry
Cameras, Multiple Views
Projective geometry and camera geometry

Overview of the course 2/2

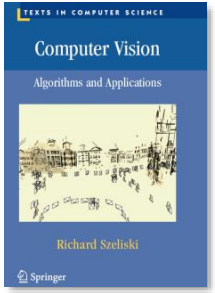
3. Deep Learning for Computer Vision
Deep Learning for computer vision and basic architectures
Object classification and detection, semantic segmentation and instance segmentation
Efficient evaluation / Tips and Tricks
RNN and image captioning
Generative Adversarial Networks and diffusion models
Domain adaptation and transfer learning
Transformers in vision and attention model

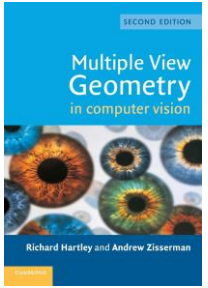
4. Seminars on different topics
Multimedia forensics and deepfake detection
Monocular depth estimation
Adversarial attacks

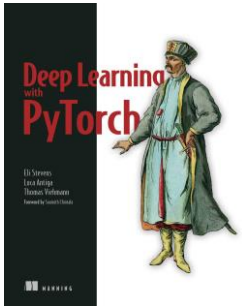
Materials

Suggested (not mandatory)

- Computer Vision: Algorithms and Applications, Richard Szeliski , 2022
  - available for free at: https://szeliski.org/Book/
- MultiView Geometry in Computer Vision, Richard Hartley & Andrew Zisserman, Cambridge ed. 2019
- Understanding Deep Learning. J.D. Prince, https://udlbook.github.io/udlbook/
- Deep Learning with PyTorch, Eli Stevens, Luca Antiga, Thomas Viehmann
- Slides, scientific papers and articles







Exam

- The exam covers the different sections of the course:
    1. Image Processing
    2. Multiview Geometry
    3. Deep Learning for Computer Vision + computer vision research topics

- The assessment of the exam consist of a **project + project presentation** (worth 1/3) and a **final written exam** (worth 2/3). Cum laude 32 points= 11+11+10 (project)

- Final project:
    - Algorithms, objectives and topics for the final project may be freely chosen (a list of topics will be given at the half of the course).
    - It requires a project abstract to be presented at the end of the course and approved by the instructors
    - Groups from 1 to 3 people

Questions?

Alcor Lab Via Ariosto 25, Rome
Department of Computer, Control, and Management Engineering A. Ruberti,
Sapienza University of Rome

Multimedia Forensics
Green AI (Vertical Farming and Beekeeping)
Edge-Vision
Deep Learning Theory
Visual Knowledge acquisition: Activity Recognition & Object Detection

Computer Vision and Nearby Fields

Simplistic summary of computer vision:

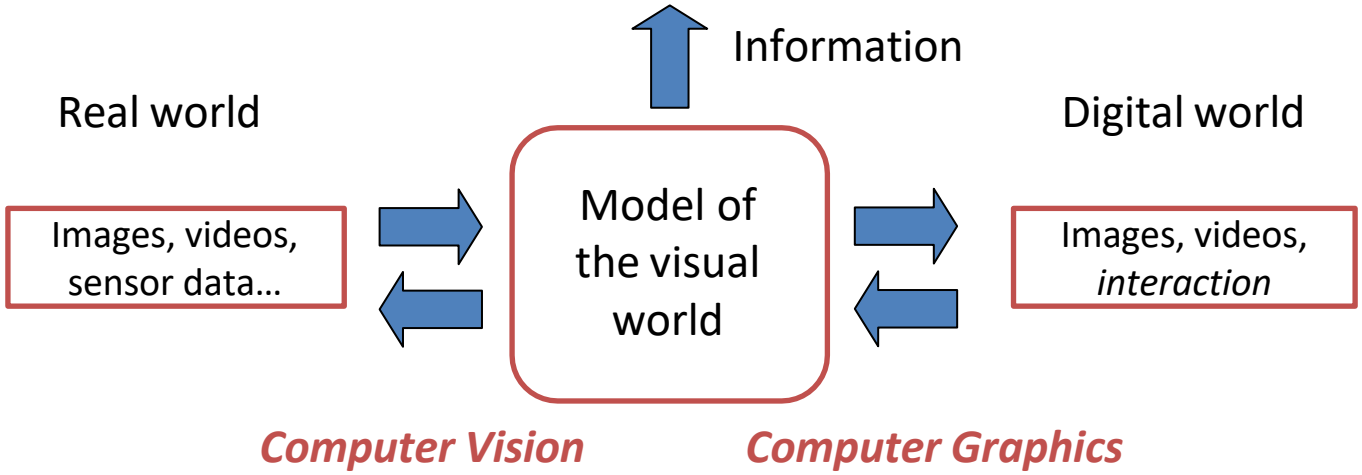
“Machine learning applied to visual data.”

Computer Vision and Nearby Fields

Simplistic summary of computer vision:

~~"Machine learning applied to visual data."~~

More in general: the ability of computers to see



Computer Vision and Nearby Fields

The ability of computers to see:

- Image and video analysis
- Image and video understanding
- Machine and robot vision

Vision and Society

Lots of data = lots of potential bias in the data.

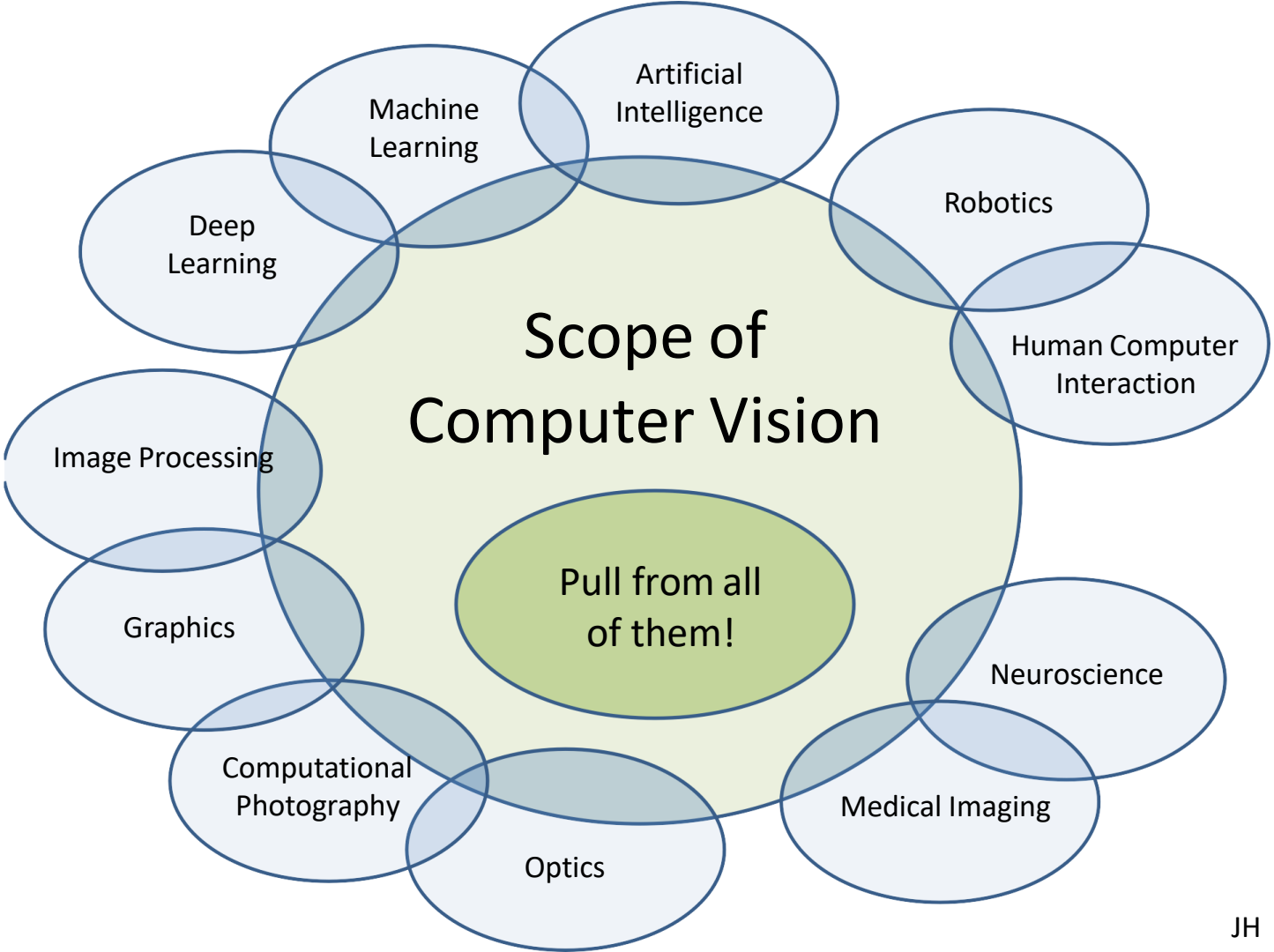
Needs understanding of possible failures.

+

Responsible approach.

+

Techniques to overcome bias.


Scope of
Computer Vision
Pull from all
of them!
Artificial
Intelligence
Machine
Learning
Deep
Learning
Robotics
Human Computer
Interaction
Image Processing
Graphics
Computational
Photography
Optics
Neuroscience
Medical Imaging

JH

Every image tells a story

- Goal of computer vision: perceive the “story” behind the picture
- Compute properties of the world
  - Image formation
  - 3D shape from 2D image
  - Names of people or objects
  - What happened?

The goal of computer vision

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

Every image tells a story

Alex Honnold

person/rock climber-→ free soloist

Rock wall/ Mountain
El Capitan

Trees / Yosemite
Park

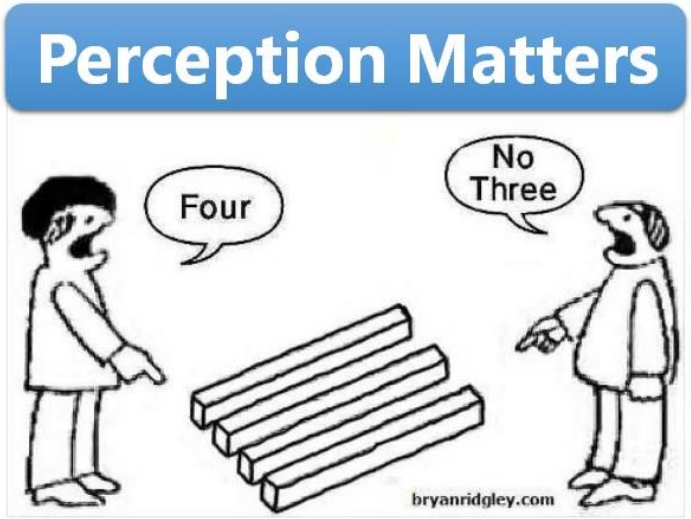
A frame of the
documentary
«Free Solo»

Can computers match human perception?

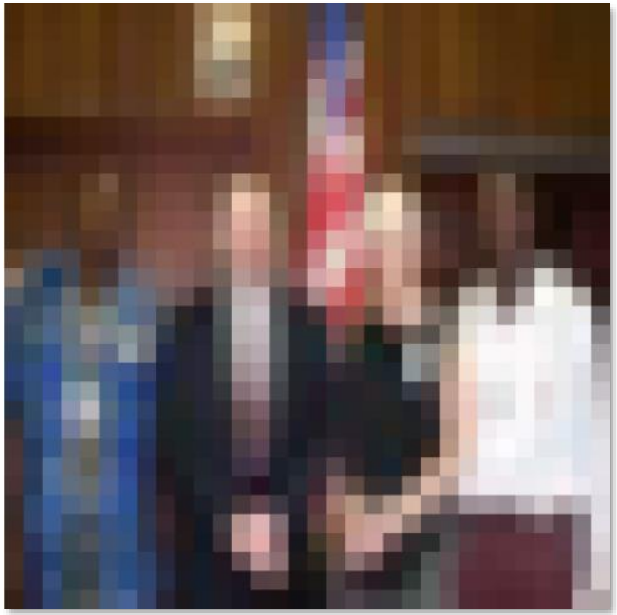
- Yes and no (mainly no)
  - computers can be better at “easy” things
  - humans are better at “hard” things
  - the brain is billion times faster than PCs
- But huge progress
  - accelerating in the last five years due to deep learning
  - what is considered “hard” keeps changing

Human perception has its shortcomings

Sinha and Poggio, *Nature*, 1996
“The Presidential Illusion”



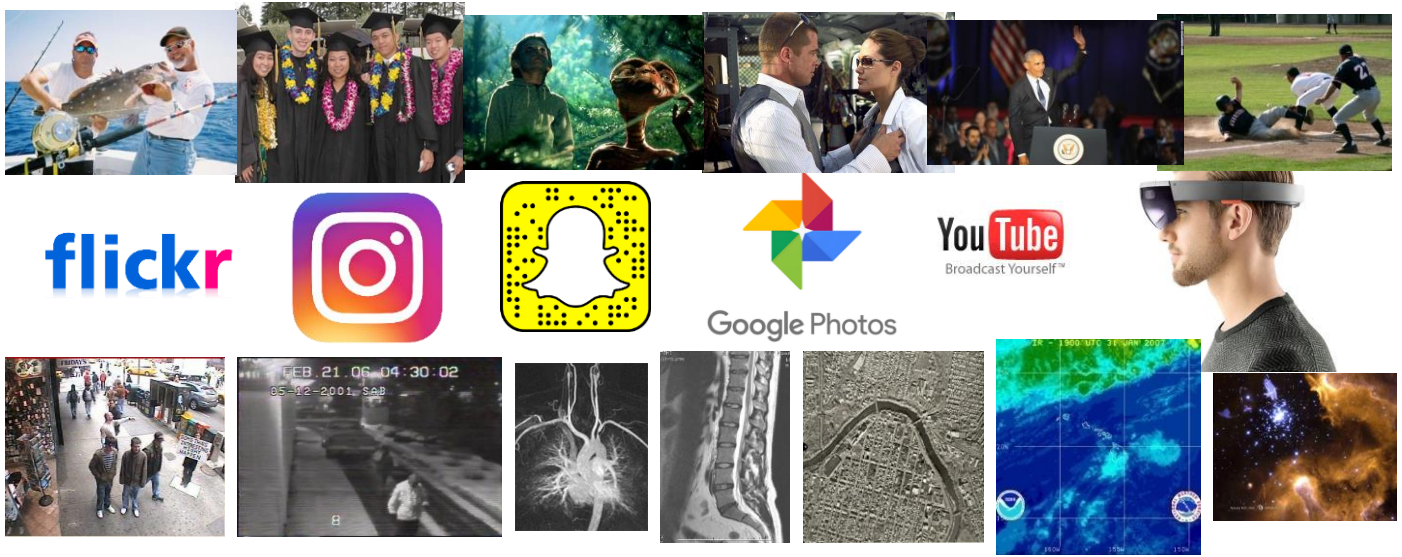
But humans can tell a lot about a scene from a little information…



Source: “80 million tiny images” by Torralba, et al.

Why study computer vision?

- Billions of images/videos captured per day



- Huge number of useful applications

The goal of computer vision

*"**Computer Vision** is a cutting edge field of Computer Science that aims to enable computers to understand **what is being seen in an image**"*

The study of recovering useful properties of the world from one or more images
with an algorithmic level of specification.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966

Vision Memo. No. 100.

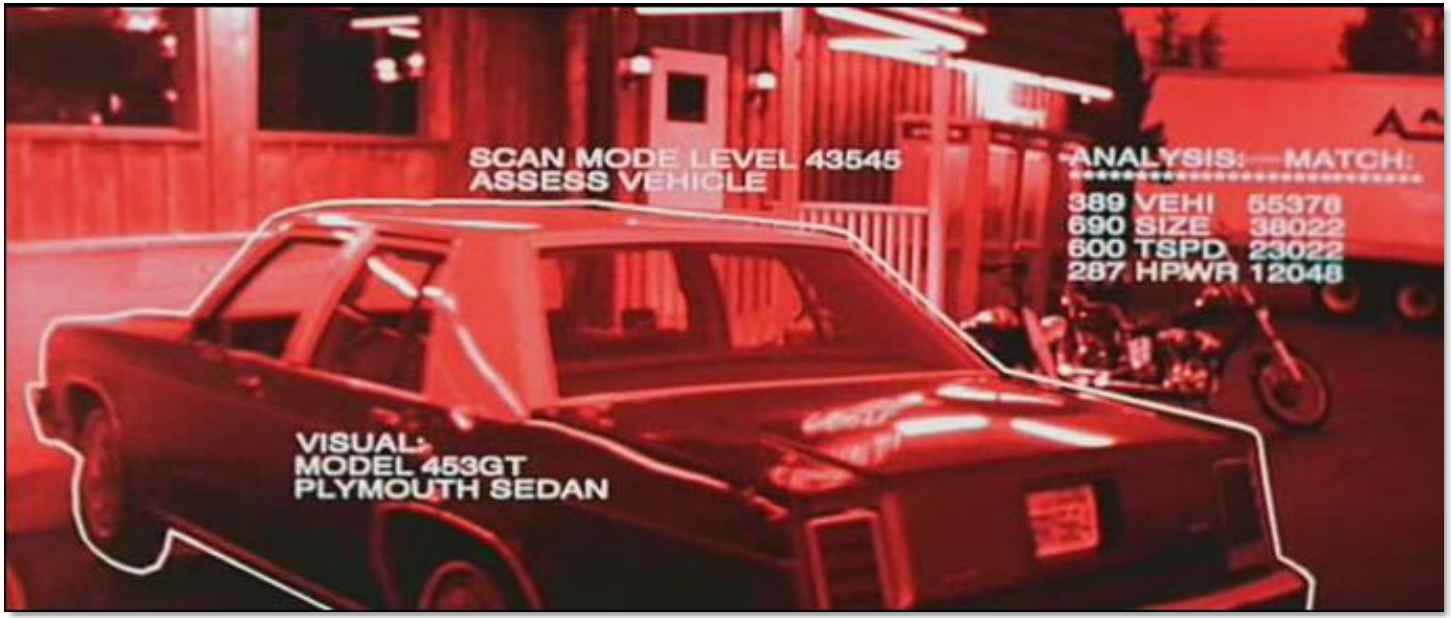
THE SUMMER VISION PROJECT

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".

The goal of computer vision

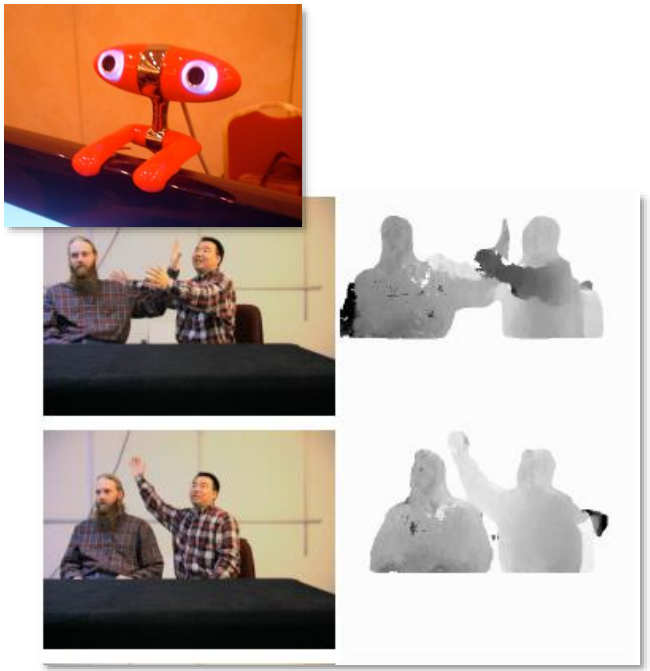
- Recognize objects and people

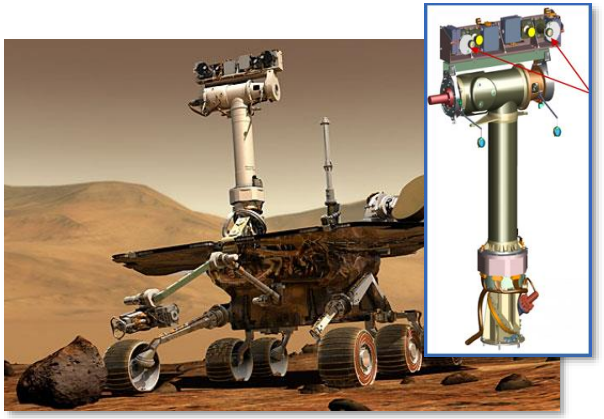


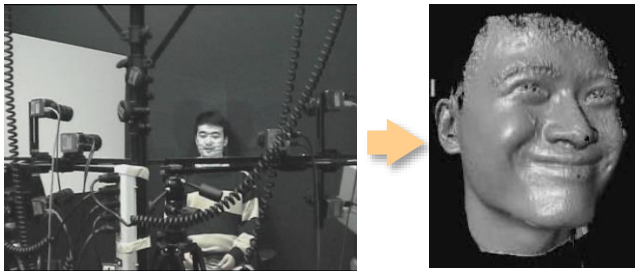
Terminator 2, 1991

The goal of computer vision

- Compute the 3D shape of the world







The goal of computer vision

- Improve photos (“Computational Photography”)

Super-resolution (source: 2d3)

Low-light photography
(credit: Hasinoff et al., SIGGRAPH ASIA 2016)

Depth of field on cell phone camera
(source: Google Research Blog)

Inpainting / image completion
(image credit: Hays and Efros)

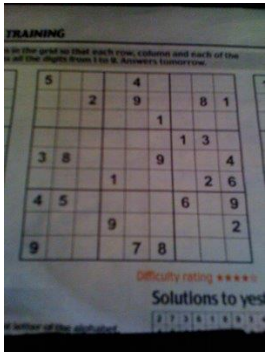
Optical character recognition (OCR)

- If you have a scanner, it probably came with OCR software

Digit recognition, AT&T labs (1990's)
http://yann.lecun.com/exdb/lenet/

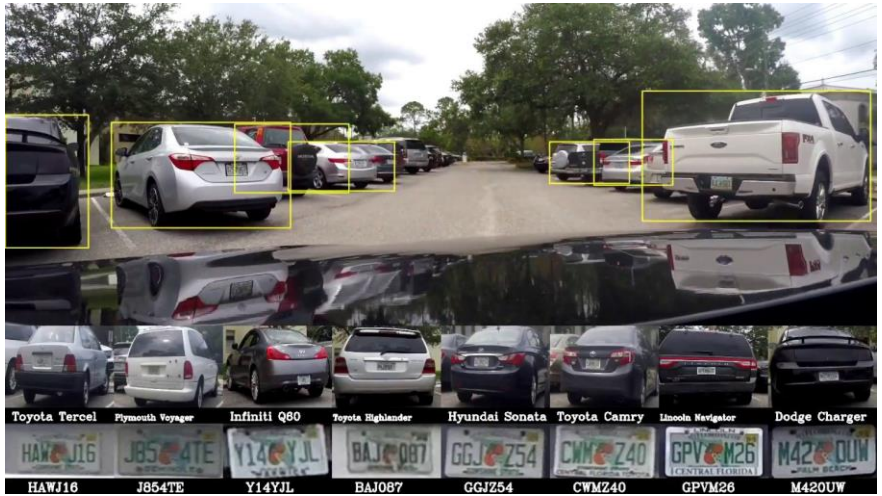
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

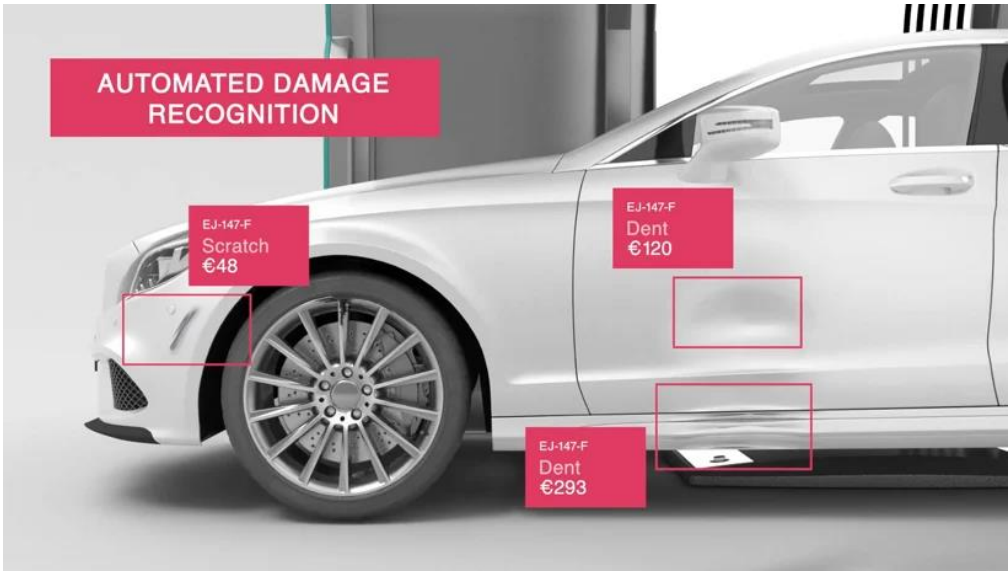
Automatic check processing



Sudoku grabber
http://sudokugrab.blogspot.com/

License plate and damage recognition

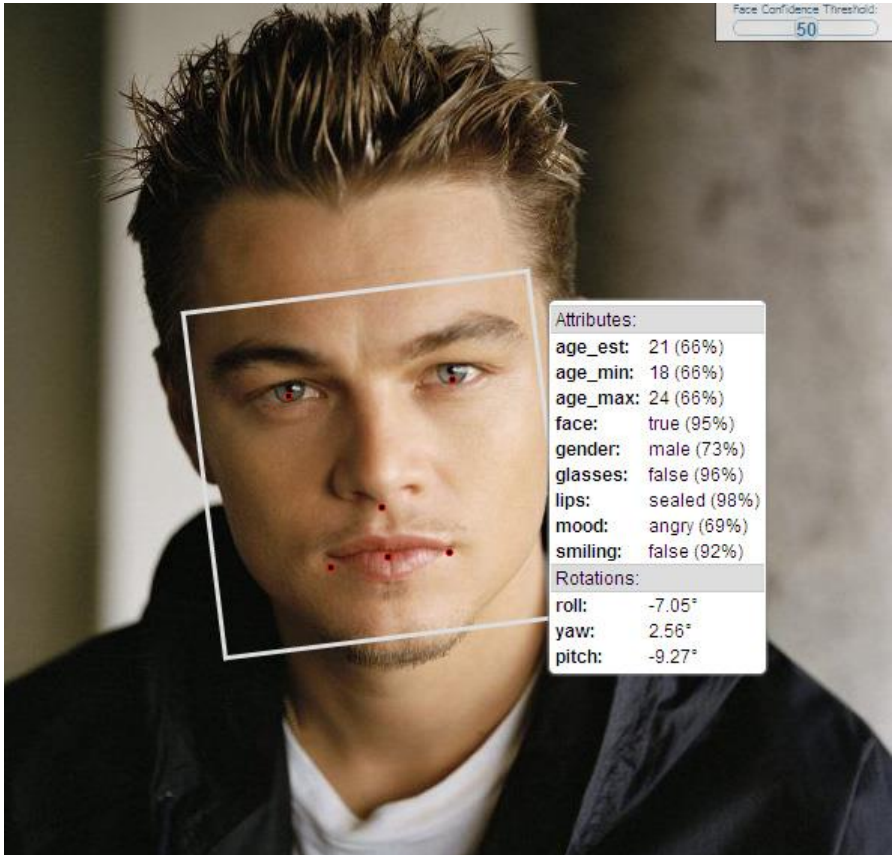




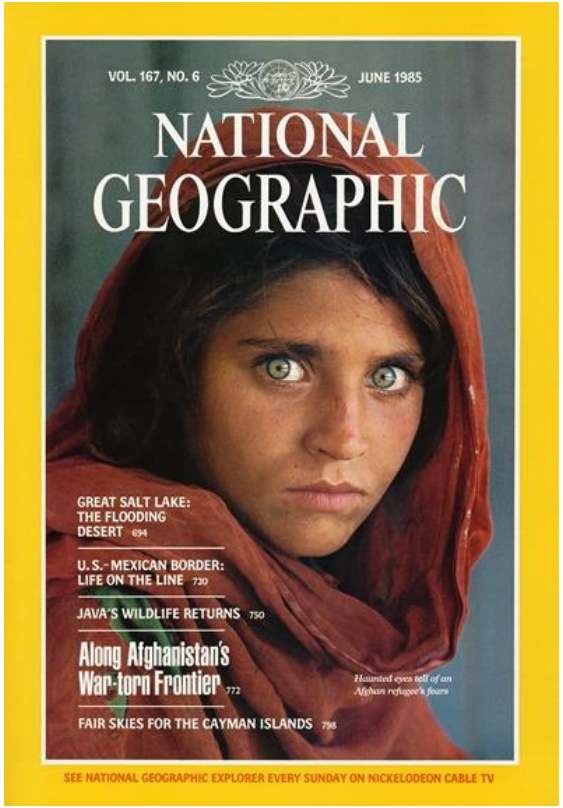
Face detection

- Nearly all cameras detect faces in real time
- Why would this be useful? Main reason is focus. Also enables “smart” cropping.

Face analysis and recognition



Face recognition

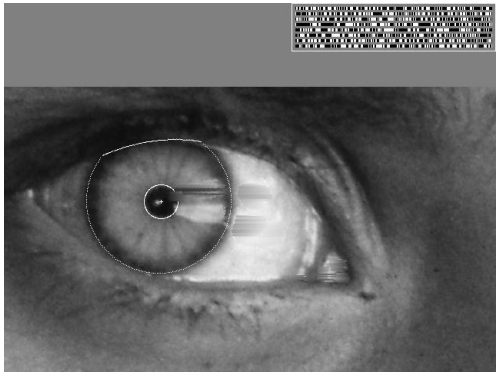


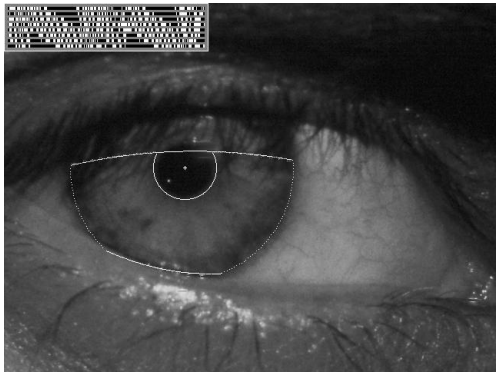
Who is she?

Source: S. Seitz

Vision-based biometrics

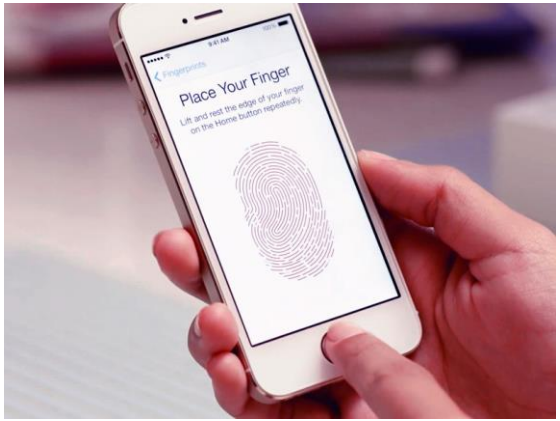
"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story





Source: S. Seitz

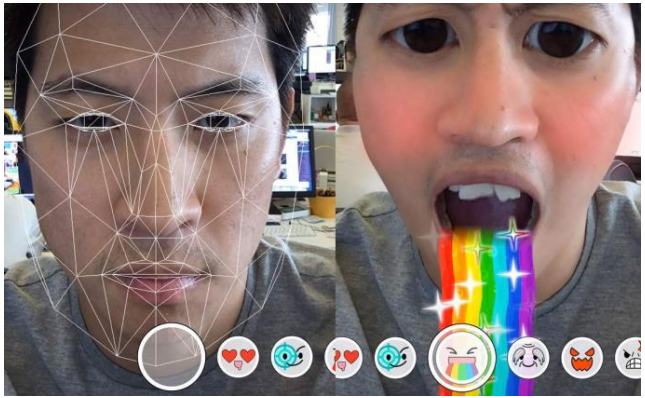
Login without a password



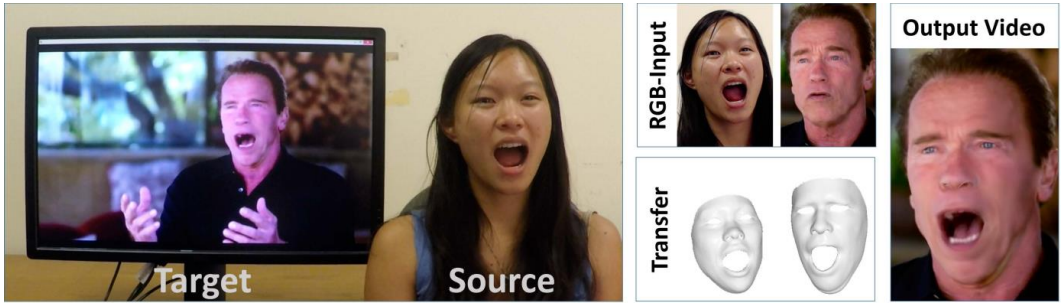
Fingerprint scanners on many new smartphones and other devices

Face unlock on Apple iPhone X
See also http://www.sensiblevision.com/

3D face tracking w/ consumer cameras



Snapchat Lenses



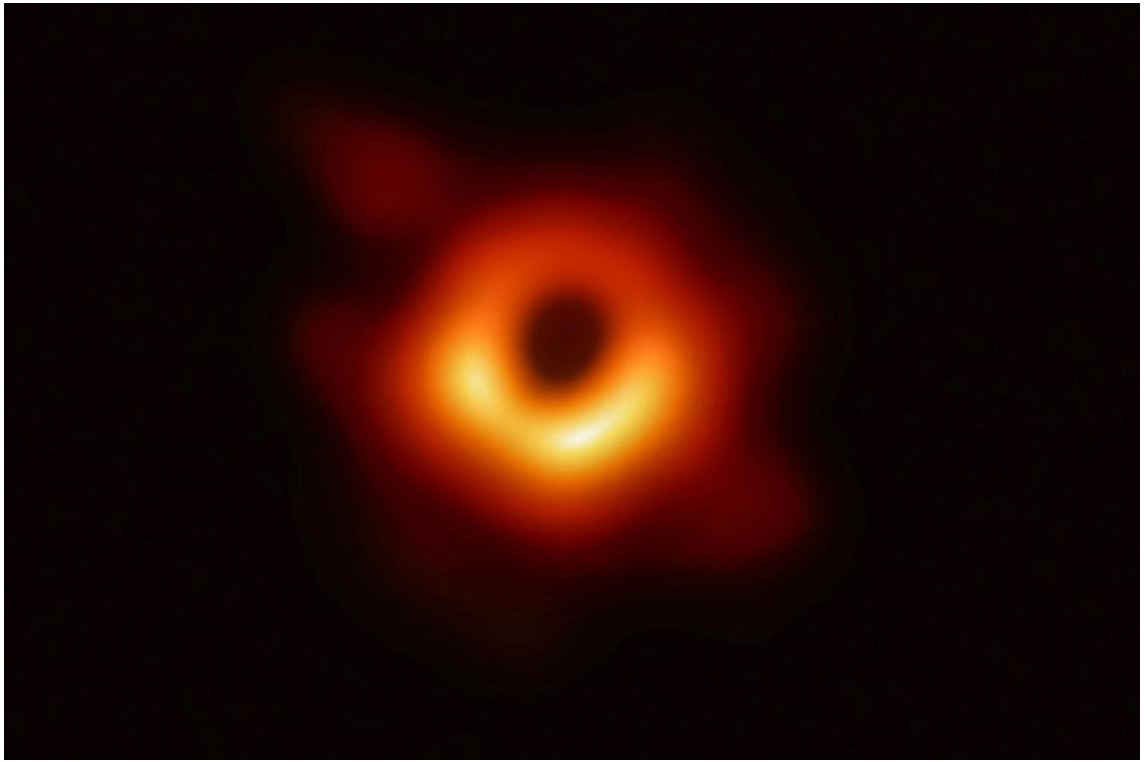
Face2Face system (Thies et al.)

Age simulation and «filters» on consumer camera

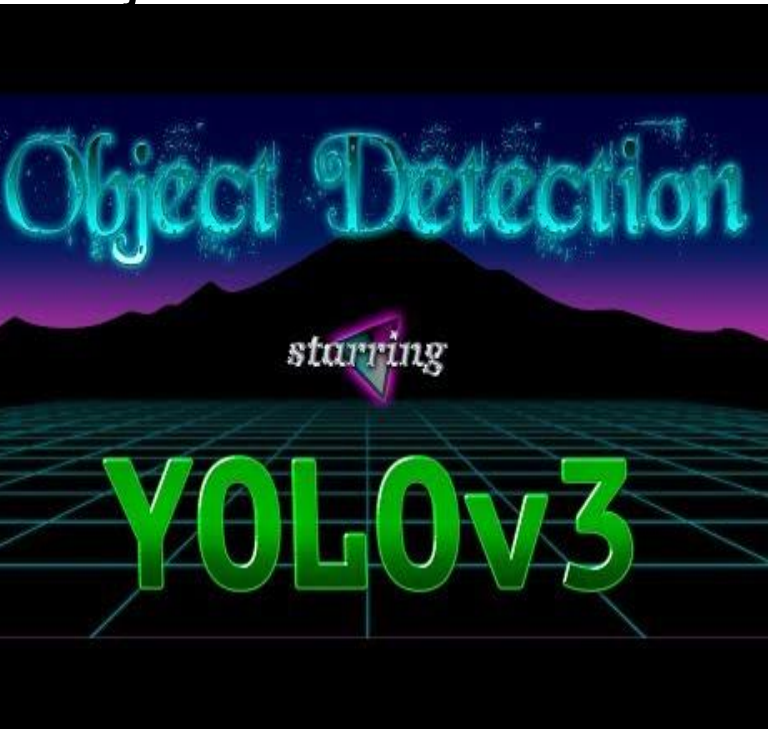
- Snapchat Time Machine, Instagram and TikTok filters

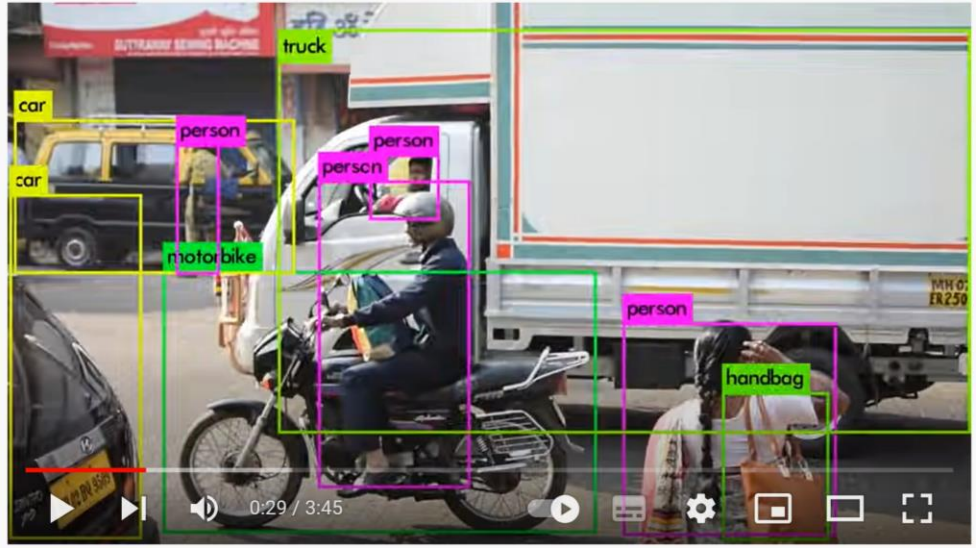
The black hole

How scientists captured the first image of a black hole, 2019



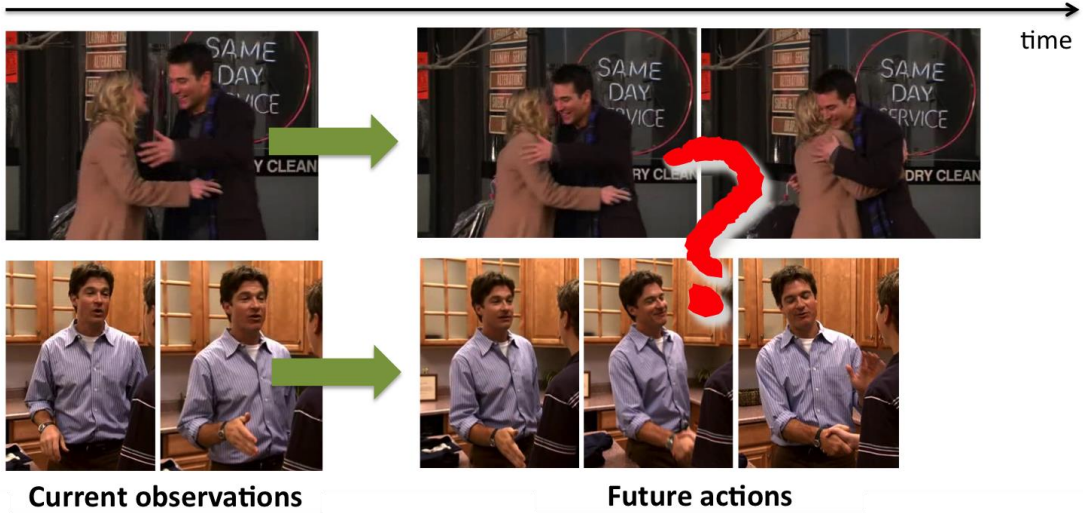
Object Detection: Yolo

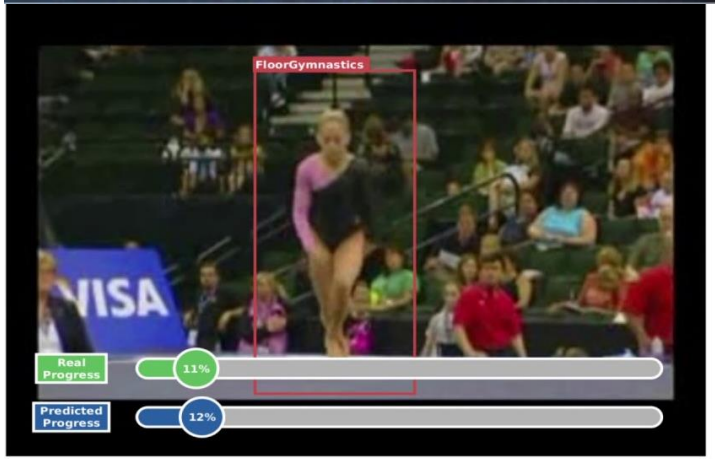




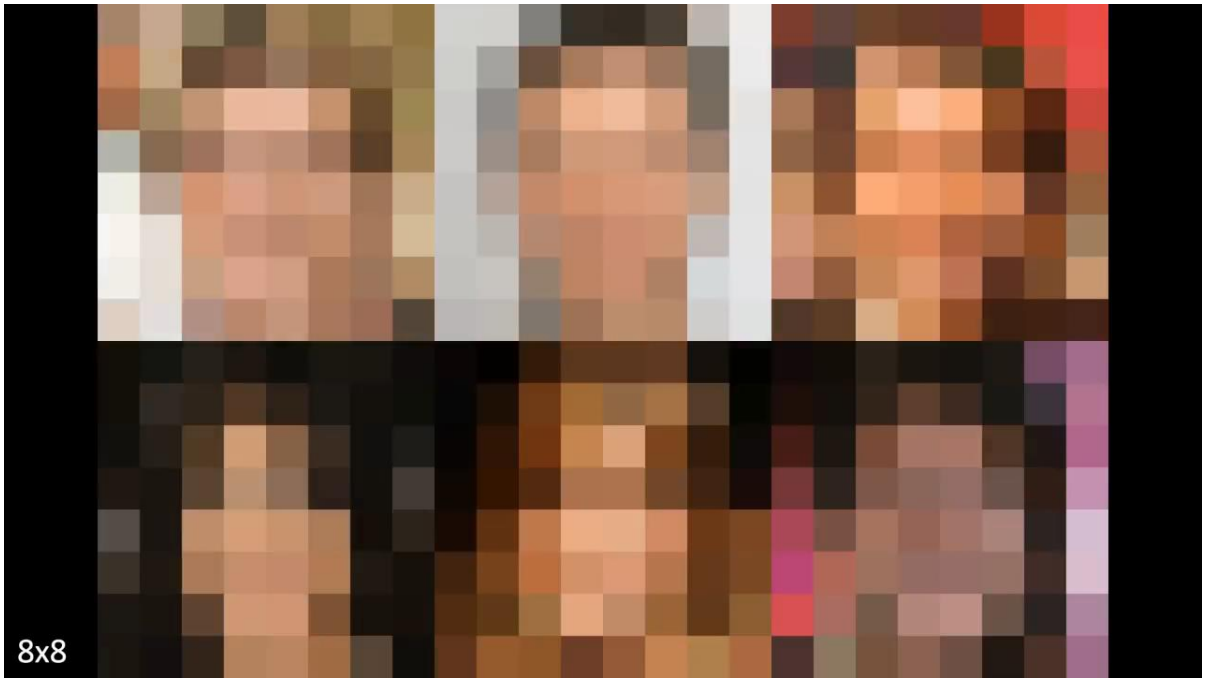
You only look once (YOLO) is a state-of-the-art, real-time object detection system.

Action recognition and prediction



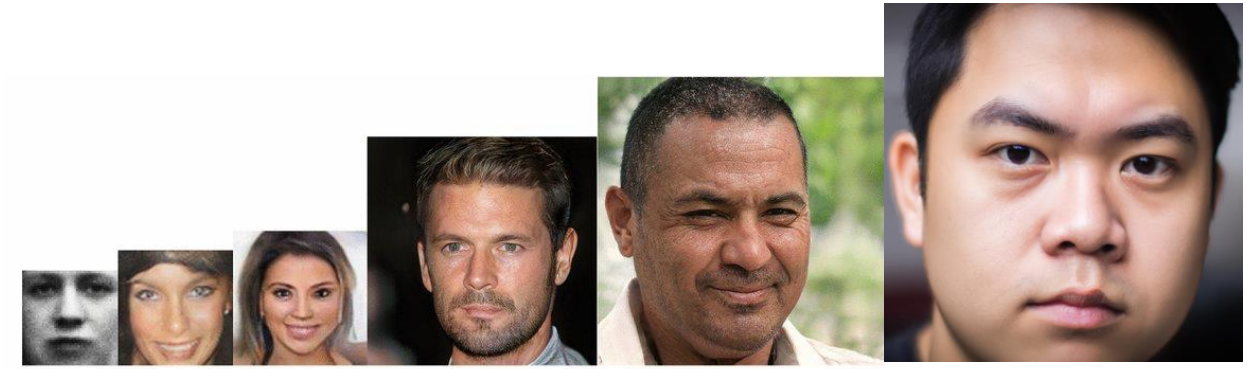


Photorealistic Human Faces with GAN



Karras, et al., *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, ICLR 2018

…and nowadays Diffusion models

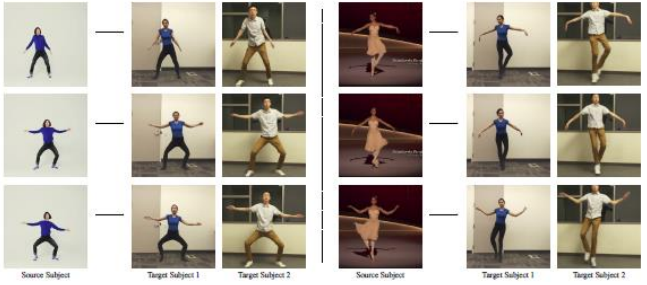


2014 GAN **2015** DCGAN **2016** CoGAN **2017** Progressive growing of GANs **2018** StyleGAN **2021** DALL-E

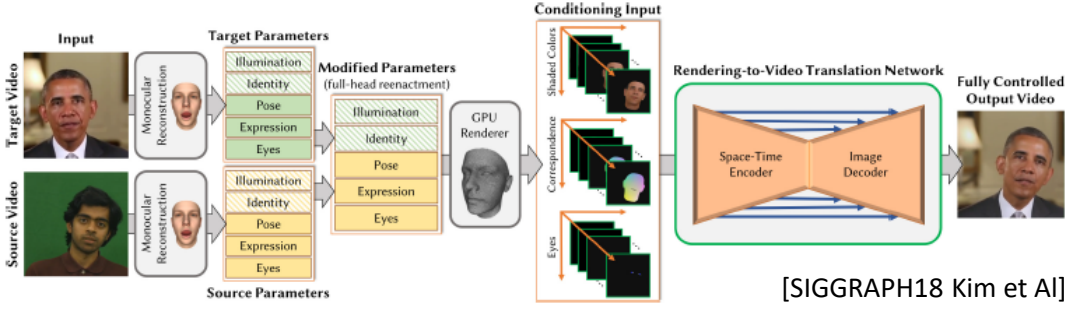
Deepfake phenomena with AI

- Many techniques: FaceTransfer, Face2Face, DeepFake, Deep Video Portaits, FaceSwap etc..

reddit



Everybody can dance now
[Chan, Efros 2018]

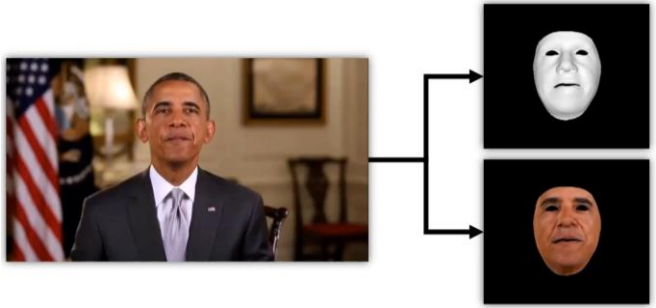


[SIGGRAPH18 Kim et Al]

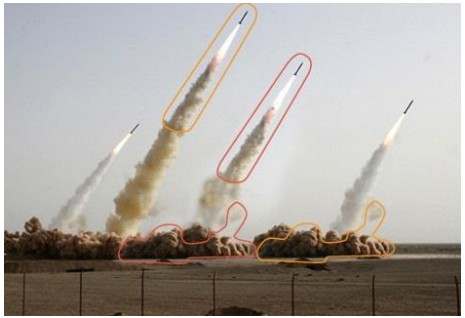
Deepfake detection

- Can you trust what you see?

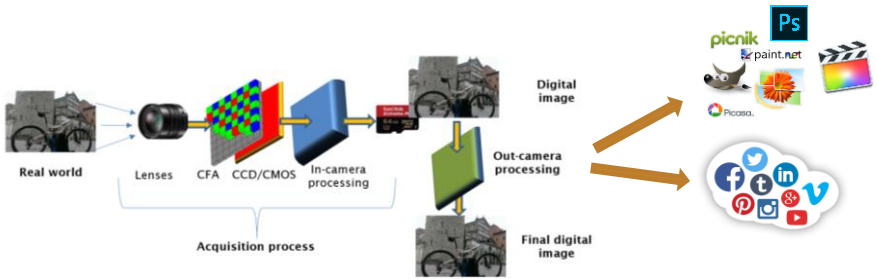
Media forensics



Authenticity: *deepfake detection*



Integrity: *forgery detection*

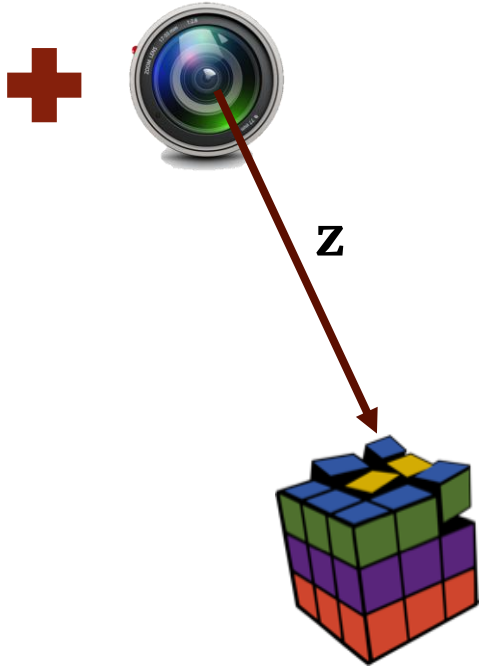


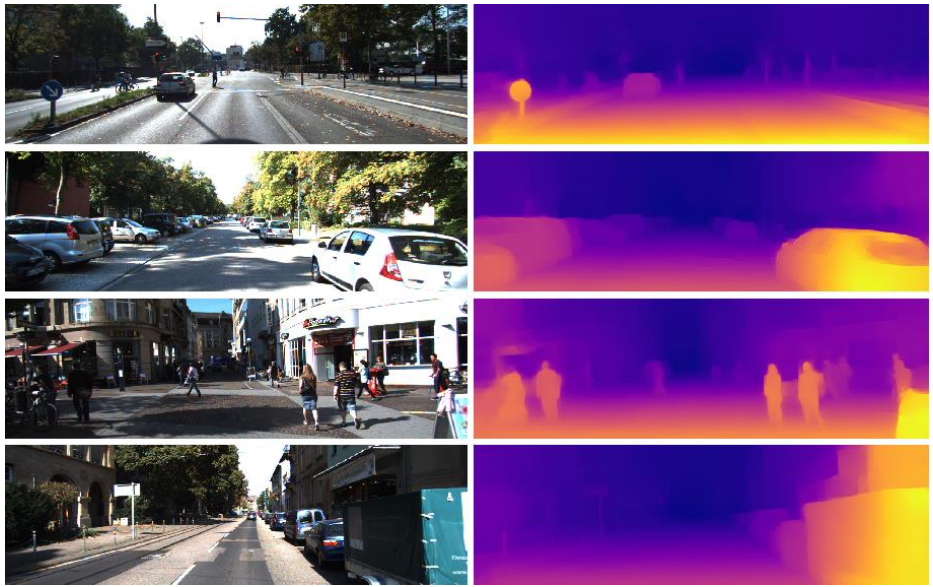
Origin: *source identification*

Monocular Depth Estimation

Task: estimate depth from a single frame/image

Embedded devices





Research and innovation examples

You just saw examples of current systems.

- Many of these are less than 5 years old

This is a very active research area, and rapidly changing

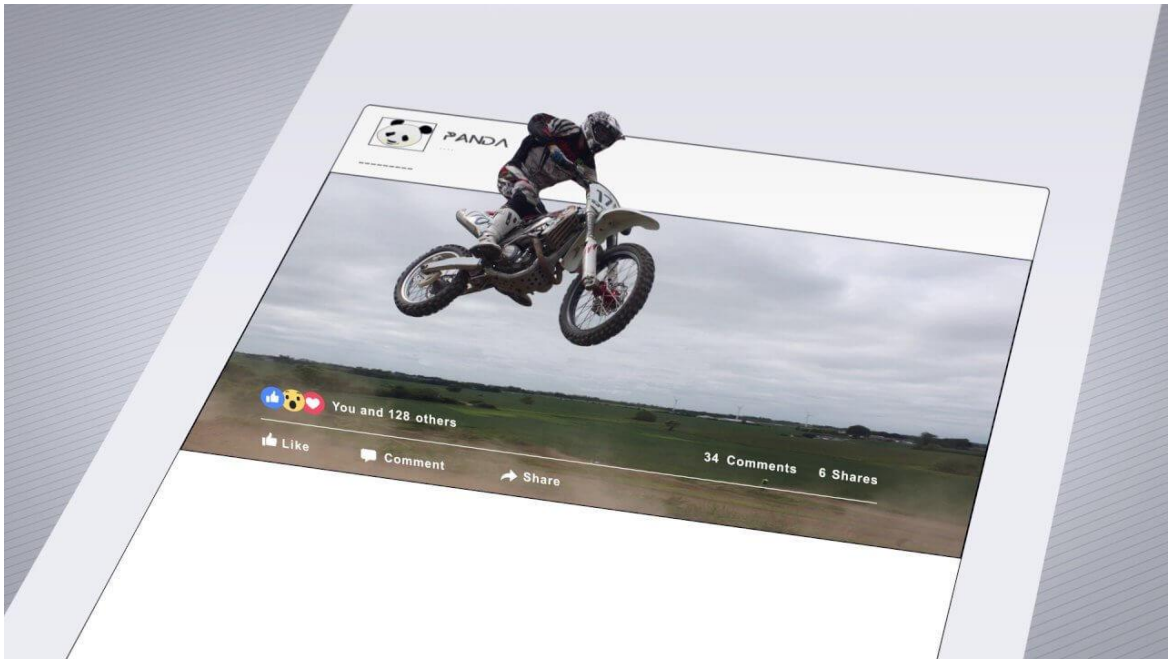
- Many new apps in the next 5 years
- Deep learning is powering many modern applications

Many startups across different of areas

- Deep learning, robotics, autonomous vehicles, medical imaging, construction, inspection, VR/AR, …

Applications

3D photos

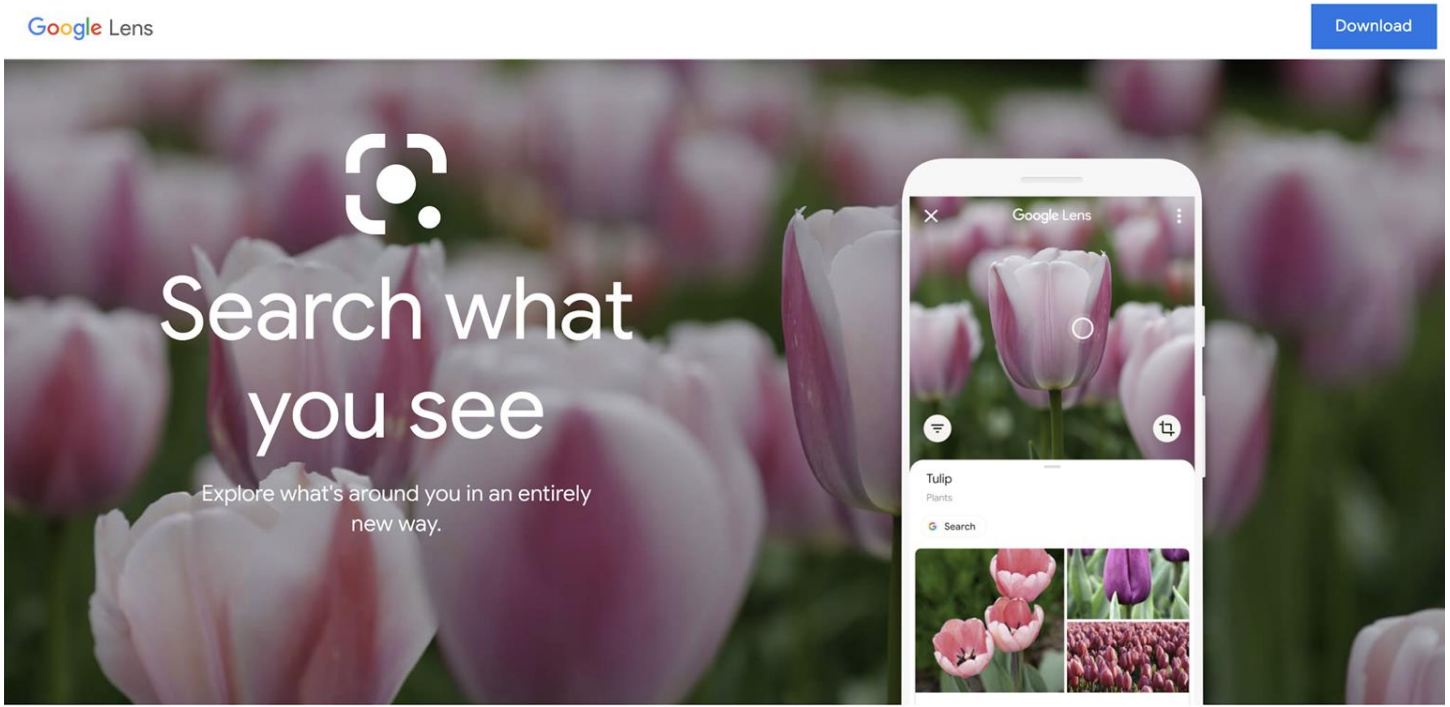


3D Photos on Facebook

Estimate depth from photo to create animation

https://ai.facebook.com/blog/-powered-by-ai-turning-any-2d-photo-into-3d-using-convolutional-neural-nets/

Object recognition



Word Lens


PLAYA CERRADA
RECIENTE ATAQUE DE TIBURÓN
BEACH CLOSED


ВЫХОД В ГОРОД
ACCESS TO CITY


現金のみ
cash only

Games

Microsoft’s XBox Kinect

Virtual reality

Oculus 2 Quest, Beat Saber

Augmented reality

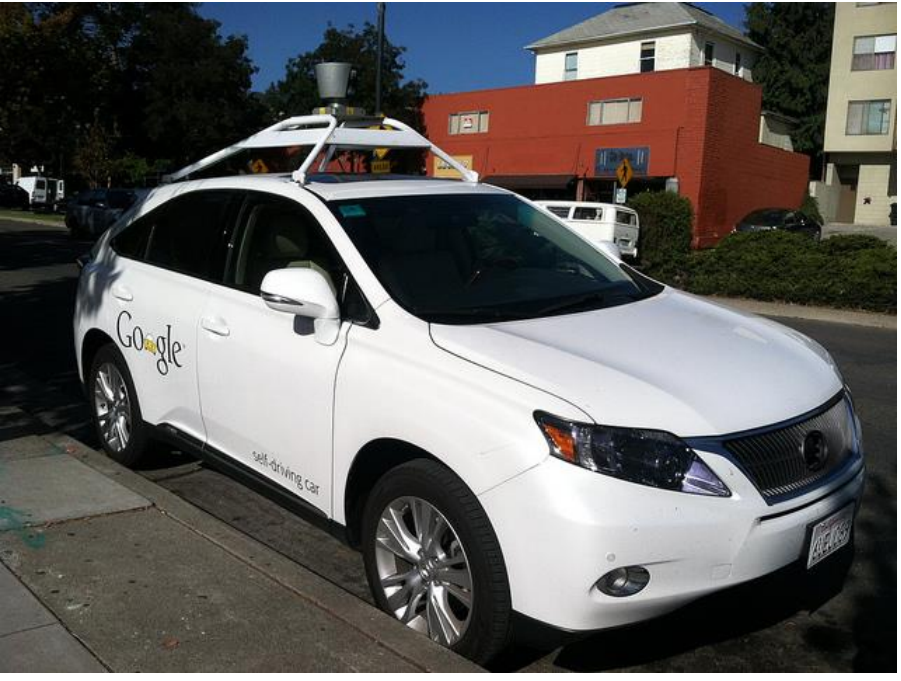
Microsoft Hololens 2

Virtual Fitting

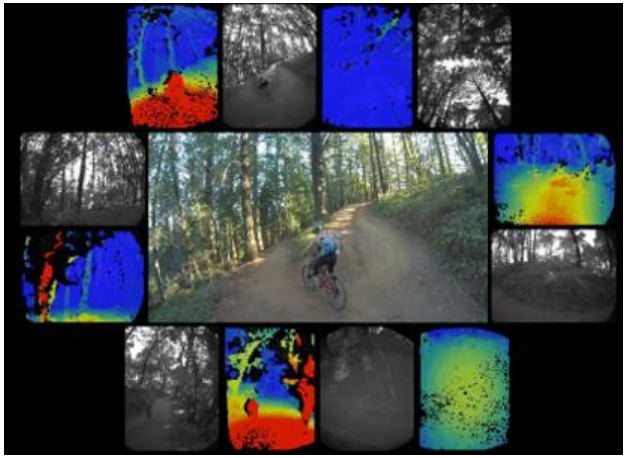
Food advisor

Smart cars / Self driving cars

/



Drones



Amazon Prime Air

Despite our success, computer vision still have long way to go…

Why is computer vision difficult?

Viewpoint variation

Illumination

Credit: Flickr user michael

Scale

Why is computer vision difficult?

Intra-class variation

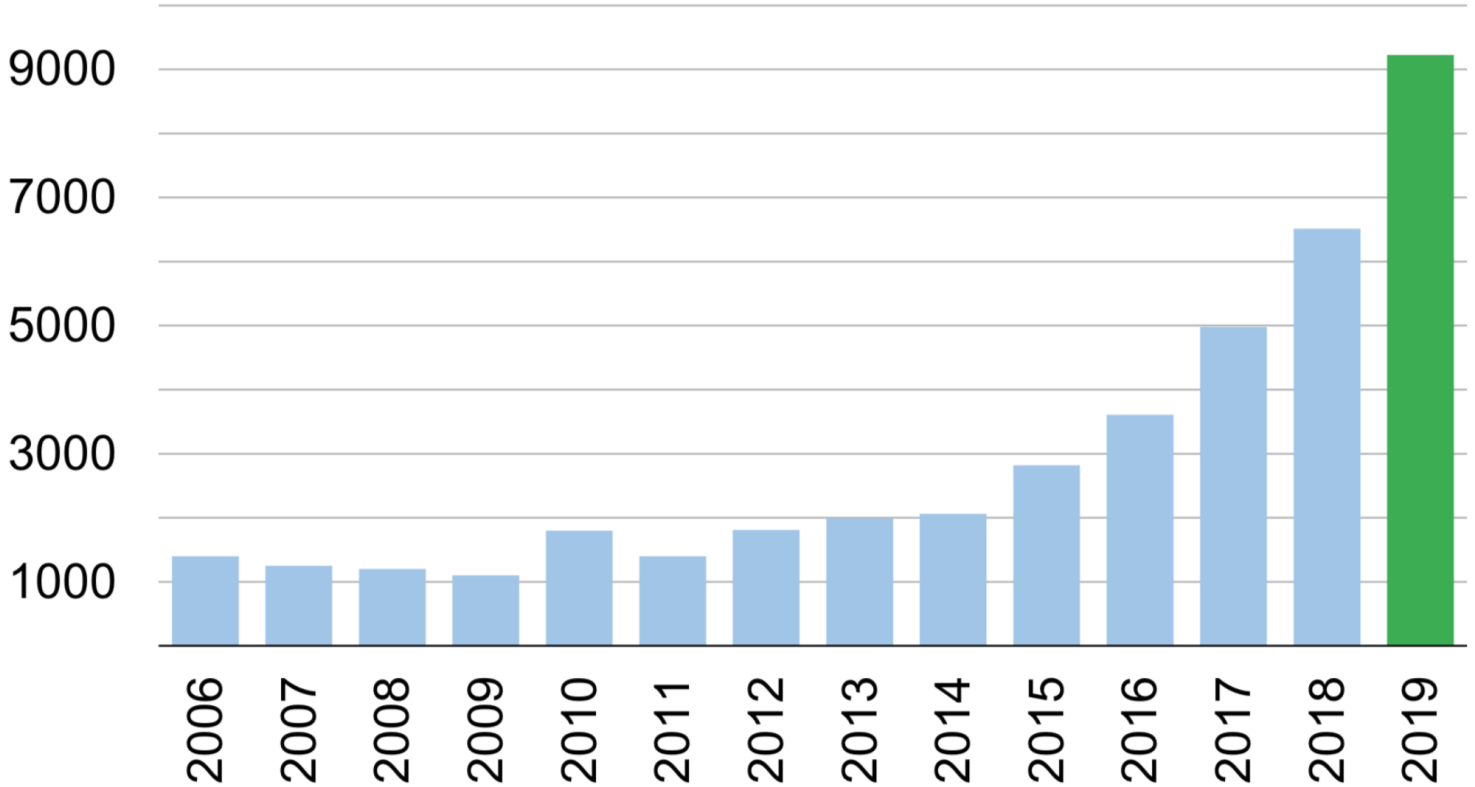
Motion (Source: S. Lazebnik)

Background clutter

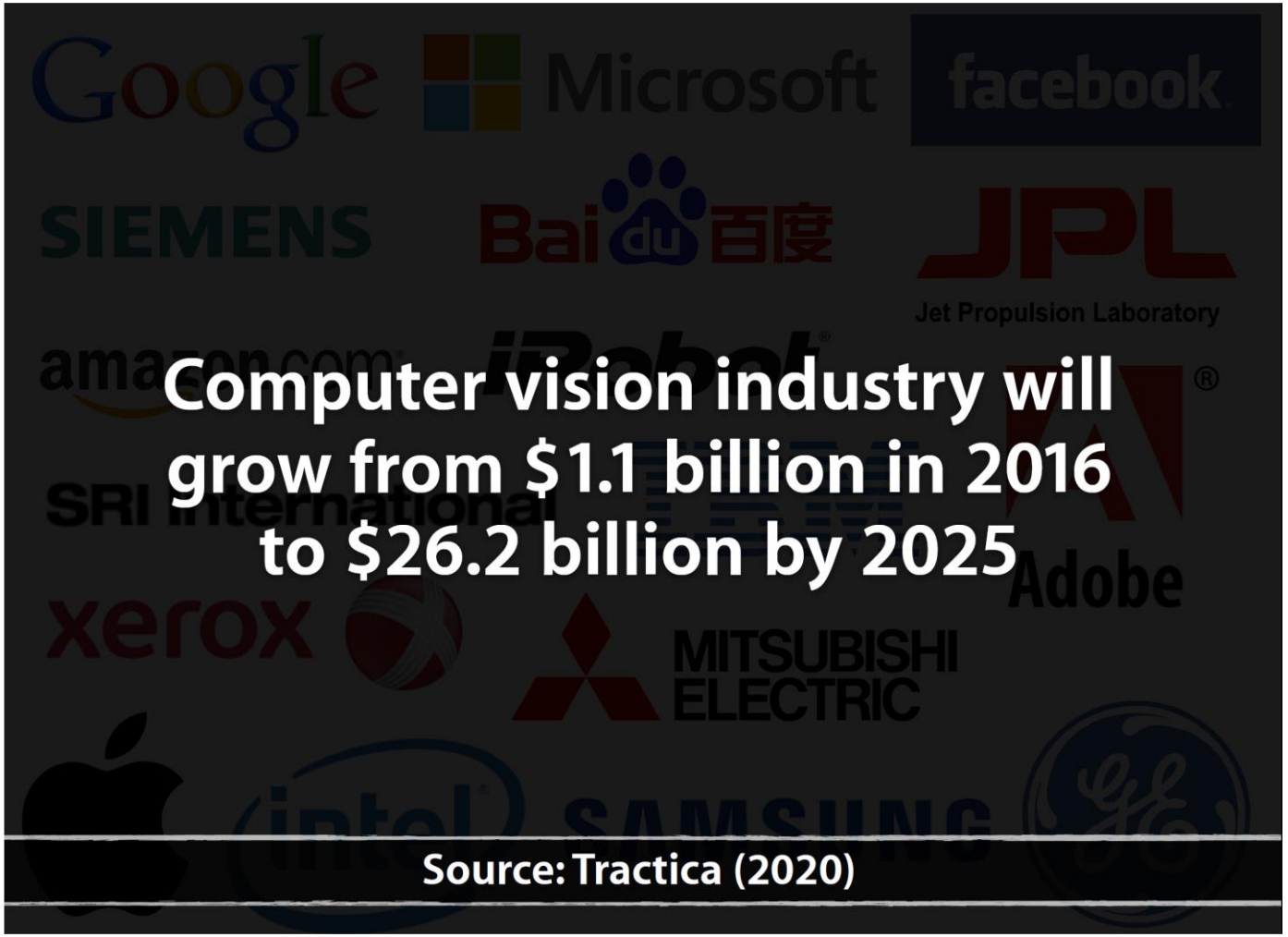
Occlusion

CVPR attendance

IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) is the premier annual computer vision event



But then COVID


Google
Microsoft
facebook
SIEMENS
Baidu 百度
JPL
Jet Propulsion Laboratory
amazon.com
iRobot
SRI International
Adobe
xerox
MITSUBISHI ELECTRIC
intel
SAMSUNG
Computer vision industry will grow from $1.1 billion in 2016 to $26.2 billion by 2025
Source: Tractica (2020)


Road ahead

Foundational analyses of visual information processing

Algorithms exploiting these analyses

Acknowledgements: some slides and material from Bernt Schiele, Mario Fritz, Michael Black, Bill Freeman, Fei-Fei, Justin Johnson, Serena Yeung, R. Szelisky, Fabio Galasso, Ioannis Gkioulekas. Kosta Derpanis