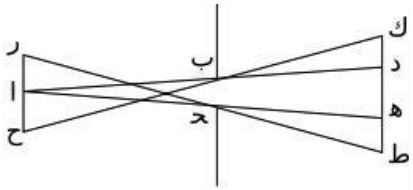
CSE 455 - Computer Vision

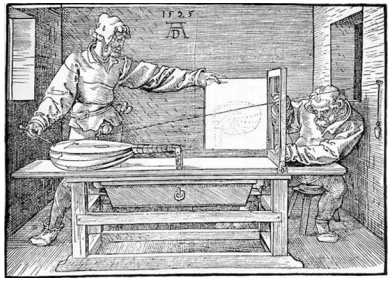
Lecture 1: Brief history of computer vision

Science stands on the shoulder of giants

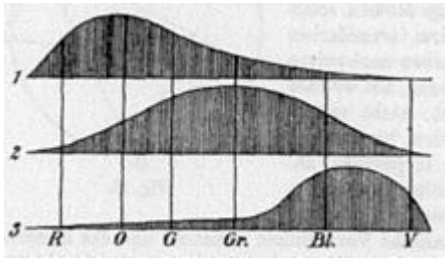
Computer vision draws origins from math & physics



Pinhole projection, optics



Projective geometry



Models of color vision (trichromacy)

Early theories of visual perception: Helmholtz, others

Two big technologies changed how computer vision was studied and how we understand them today.

Q. Can anyone here guess what those two events were?

First technology

Aside from physics and math, computer vision also has connections to art

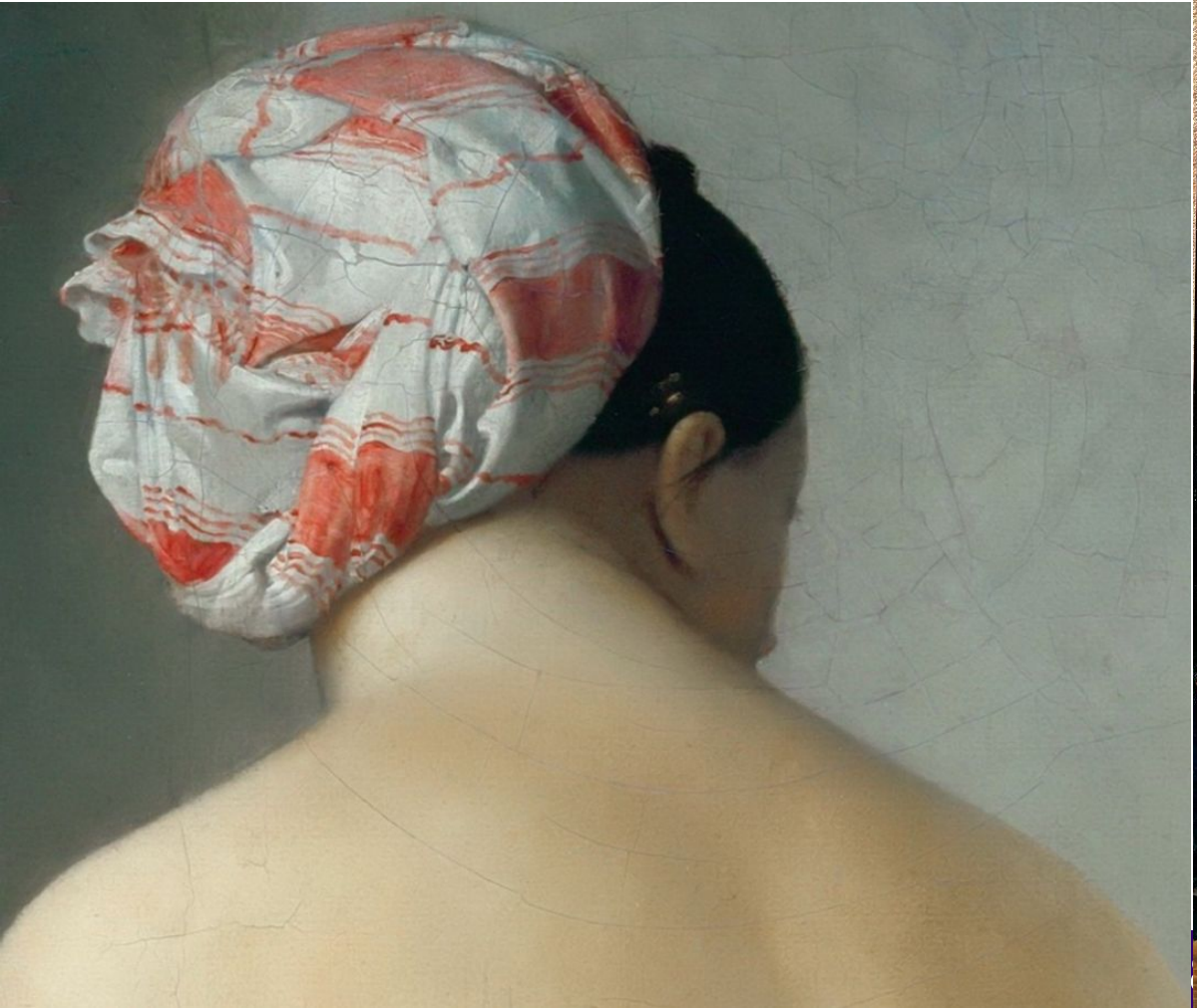
Pictures before 1838

Portraiture - artists would spend hours/days drawing their subjects who stood still in front of them

1812: Jacques-Louis-David

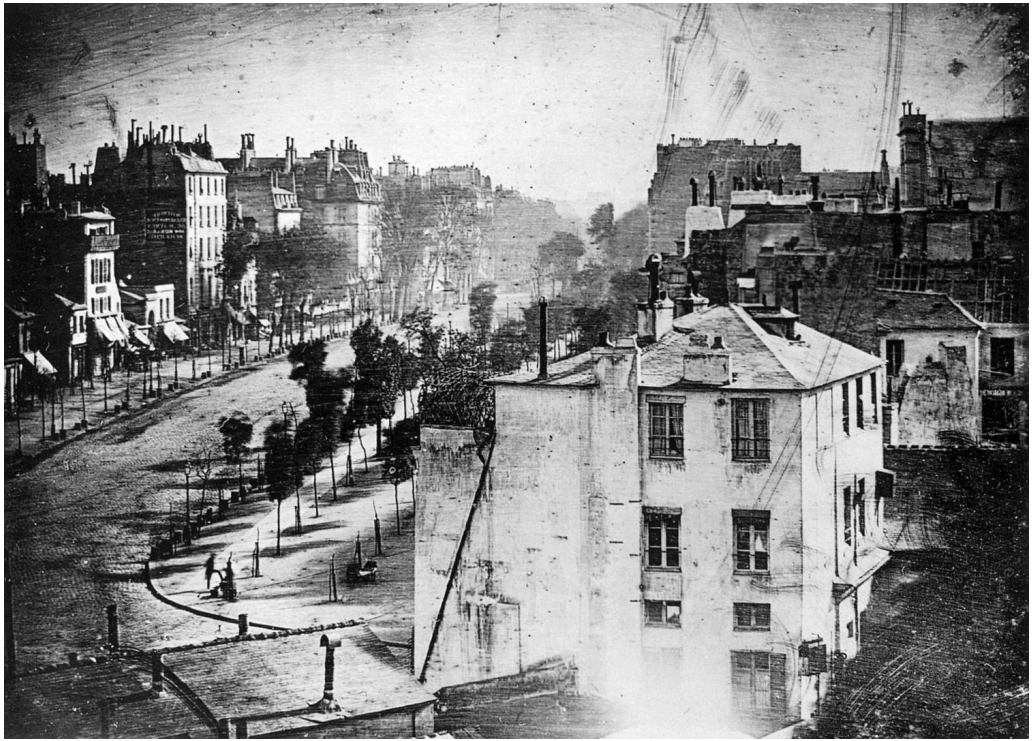
The Emperor Napoleon at his Study at the Tuileries

1808: Ingres, La grande baigneuse



1837: Niépce, First photo of one’s meal

1838: Boulevard du Temple, Daguerre



1838: First selfie, Robert Cornelius

Technology often begets fear

“From today, painting is dead”
— painter Paul Delaroche
at a demonstration of the Daguerreotype, 1839

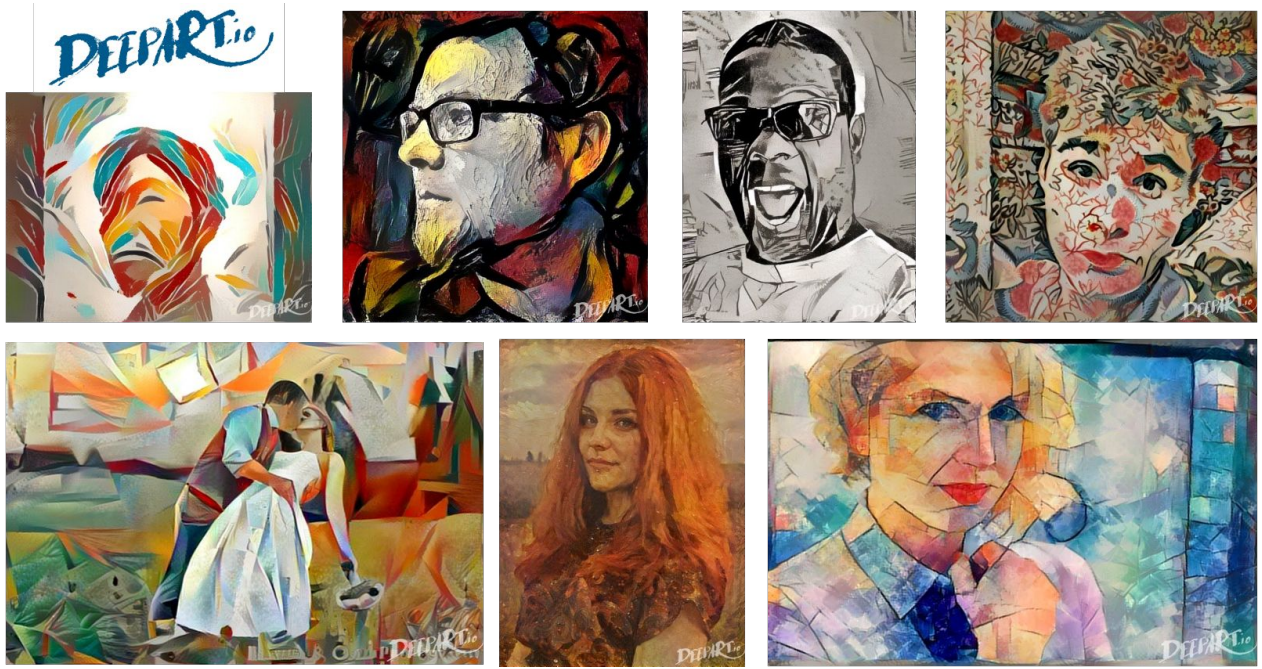
Second technology

- 1957: Digital scanner invented at NIST

With smaller cameras and larger storage,

We began curating large scale databases of images online

With those images, we now have Generate vision models

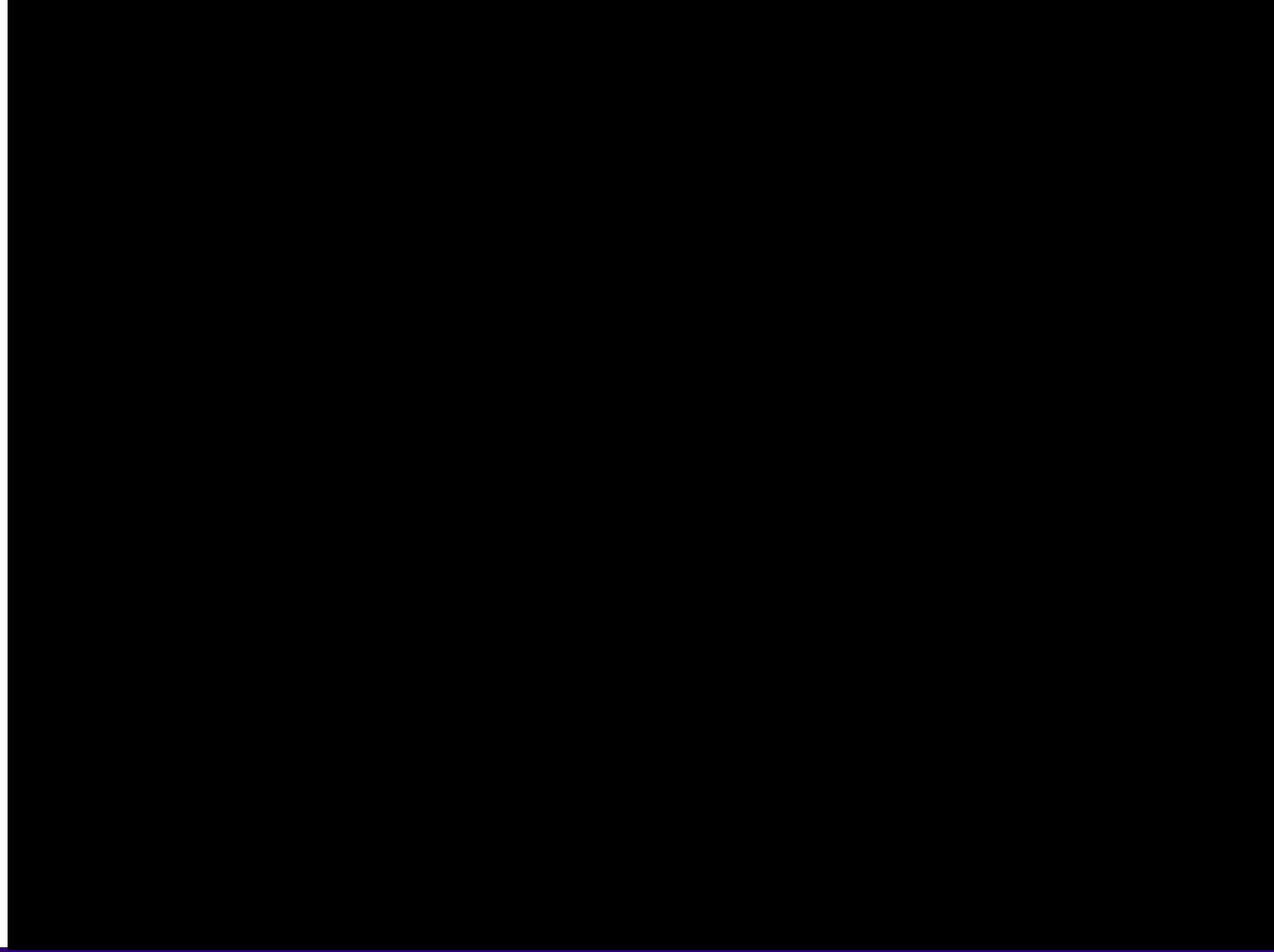


Neural Style Transfer [Gatys et al. 2015]

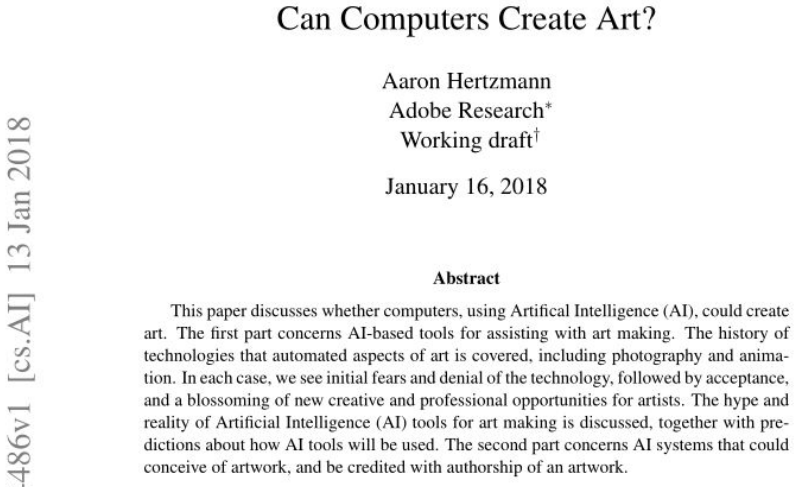
AI Art


DeepDream:
The art
of neural
networks
February 26 – 27th, 2016
DeepDream

New interactive art



New technology begets fear



.486v1 [cs.AI] 13 Jan 2018

Can Computers Create Art?

Aaron Hertzmann
Adobe Research*
Working draft†

January 16, 2018

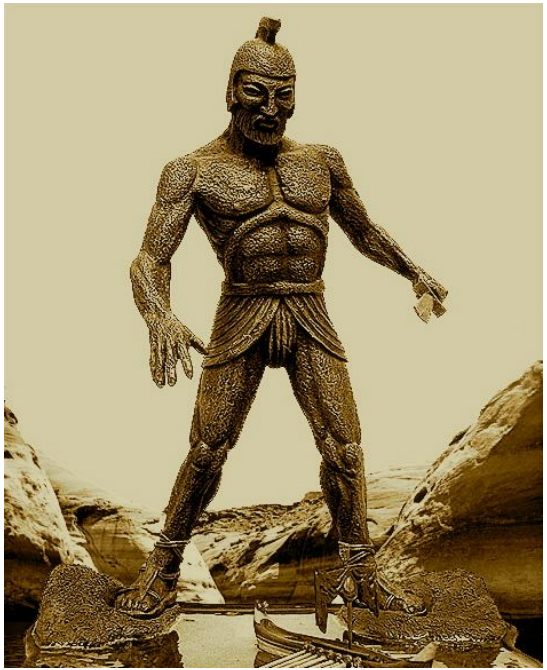
Abstract

This paper discusses whether computers, using Artifical Intelligence (AI), could create art. The first part concerns AI-based tools for assisting with art making. The history of technologies that automated aspects of art is covered, including photography and animation. In each case, we see initial fears and denial of the technology, followed by acceptance, and a blossoming of new creative and professional opportunities for artists. The hype and reality of Artificial Intelligence (AI) tools for art making is discussed, together with predictions about how AI tools will be used. The second part concerns AI systems that could conceive of artwork, and be credited with authorship of an artwork.

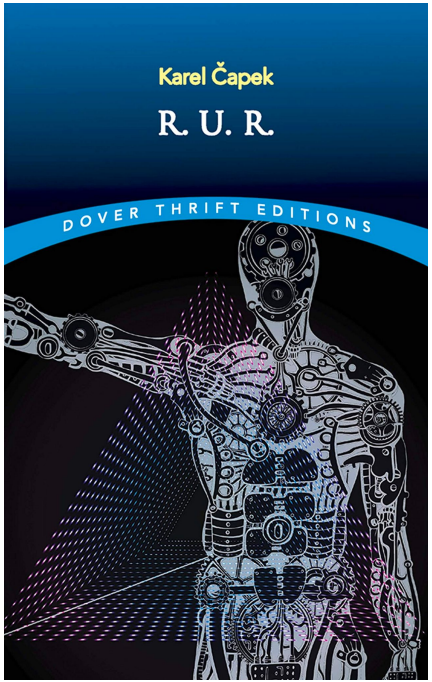
Aside from art

Computer Vision has often been depicted by popular media

Depictions of AI: Myths and Stories



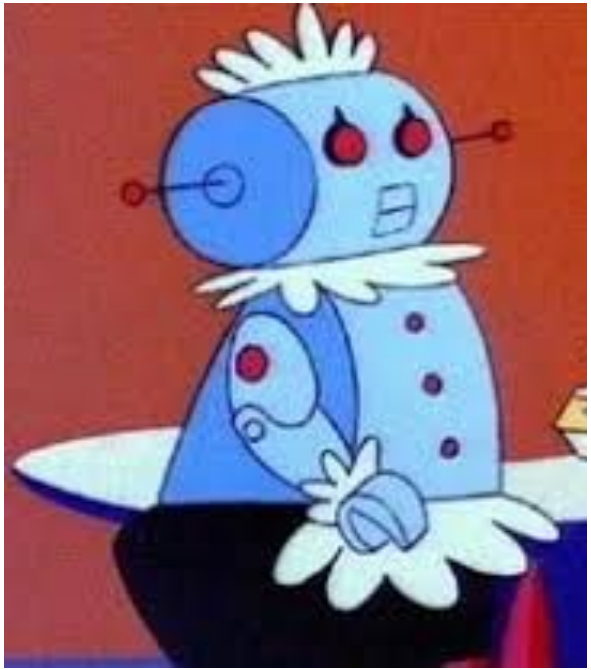
Legend of Talos
Adrienne Mayor, *Gods and Robots*



R. U. R. (1920)

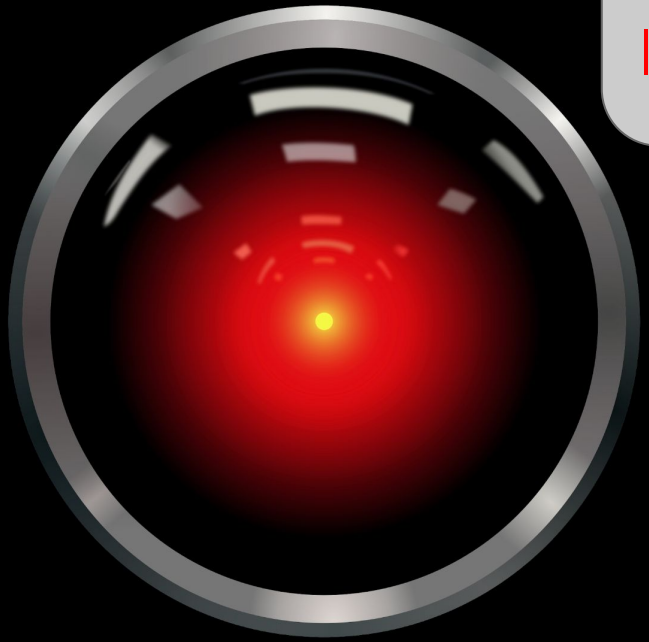
Talking Computer in *Star Trek* (1966)

1956

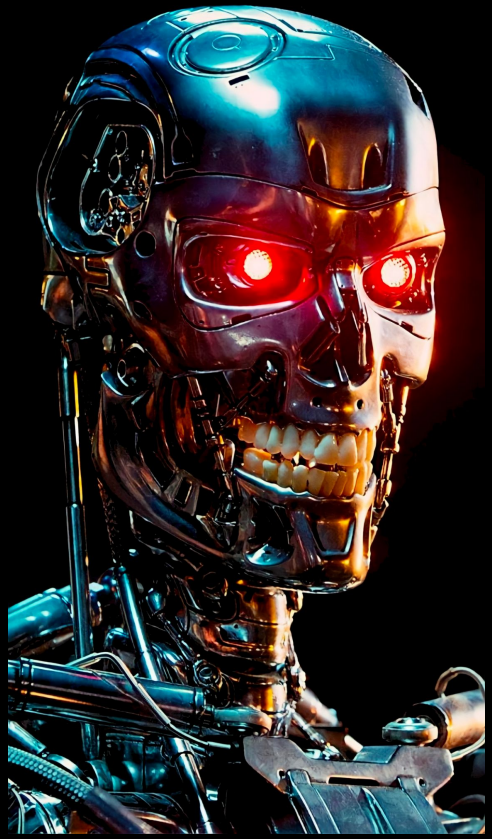


1962

1965



I'm sorry, Dave.
I'm afraid
I can't do that.

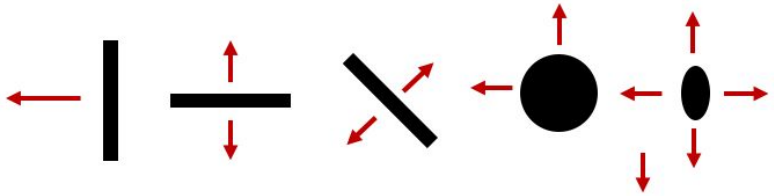


1968

1984

Aside from physics, math, art, popular media,

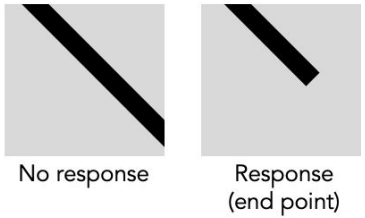
Computer Vision also draws on fundamental findings in neuroscience

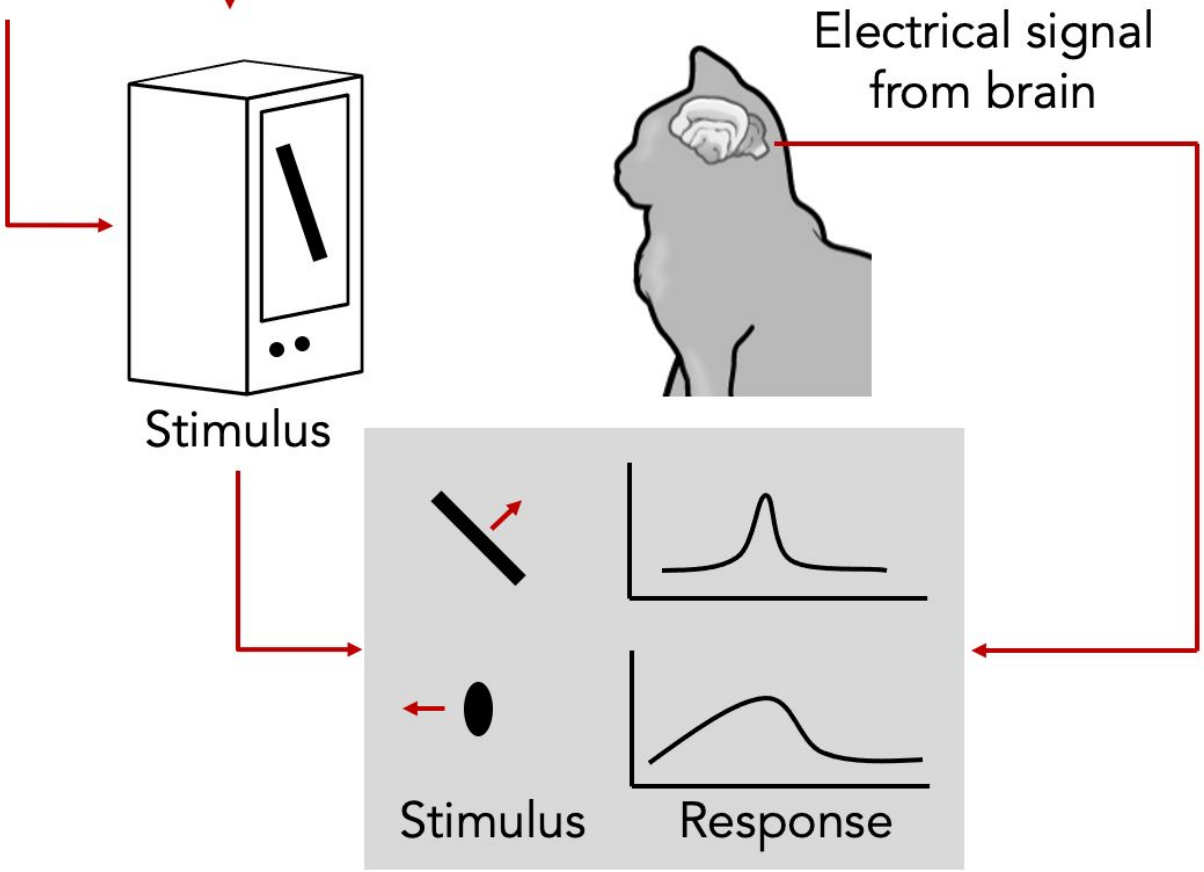


Hubel & Wiesel, 1959

How does animal vision work?

Won Nobel Prize in 1981
Visual processing is hierarchical, involving recognizing simpler structures, edges, etc.

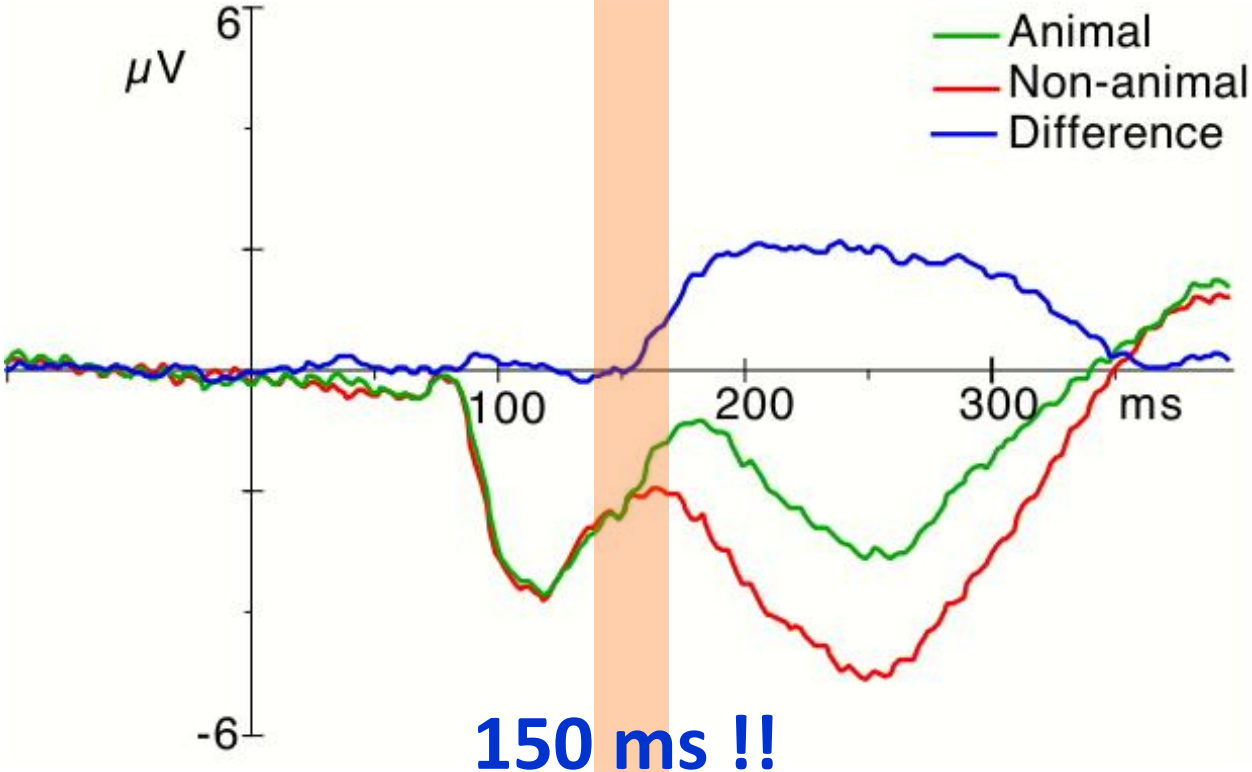




Human vision is superbly efficient

Potter, Biederman, etc. 1970s

Thorpe, et al. *Nature,* 1996



150 ms !!

Thorpe, et al. *Nature,* 1996

Aside from physics, math, art, popular media, neuroscience

Computer Vision is also influenced by cognitive science explorations

Change Blindness

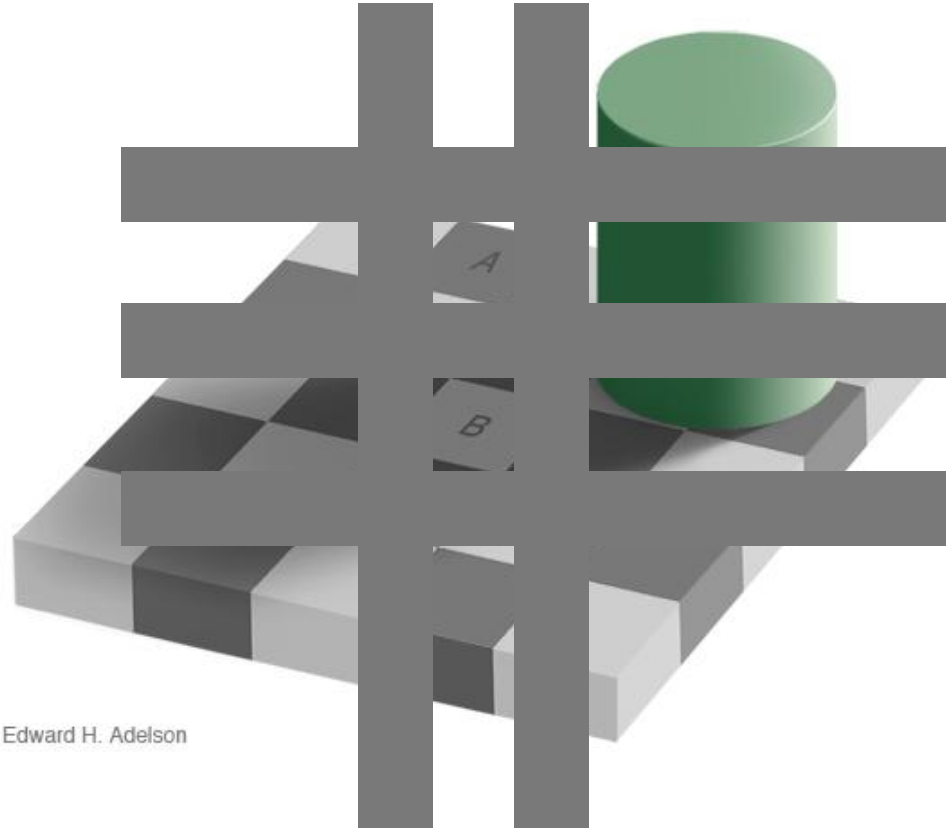
Rensink, O’regan, Simon, etc.

Change Blindness

Rensink, O’regan, Simon, etc.

Segmentation

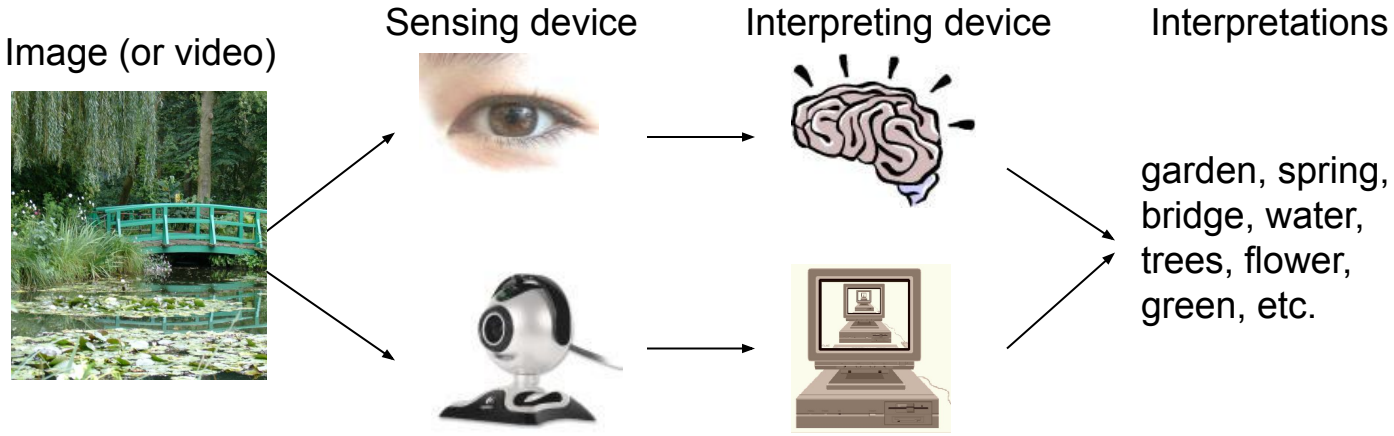
Perception


A
B
Edward H. Adelson

Motion without movement

Common theme in computer vision: which parts of human vision are necessary for intelligent systems?

So, what is computer vision?



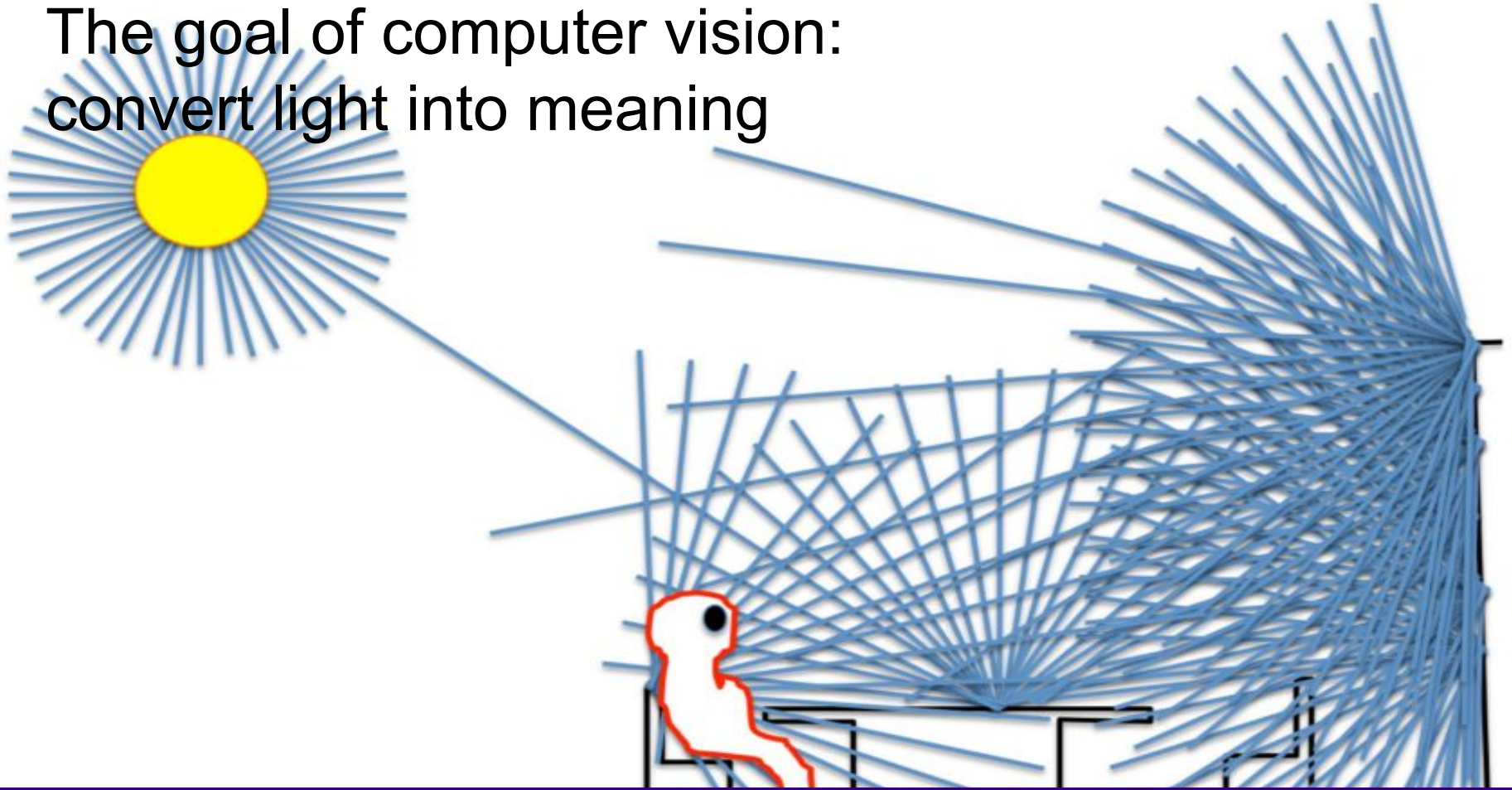
Today's agenda

- History of understanding perception
- Introduction to computer vision
- Course overview

Today's agenda

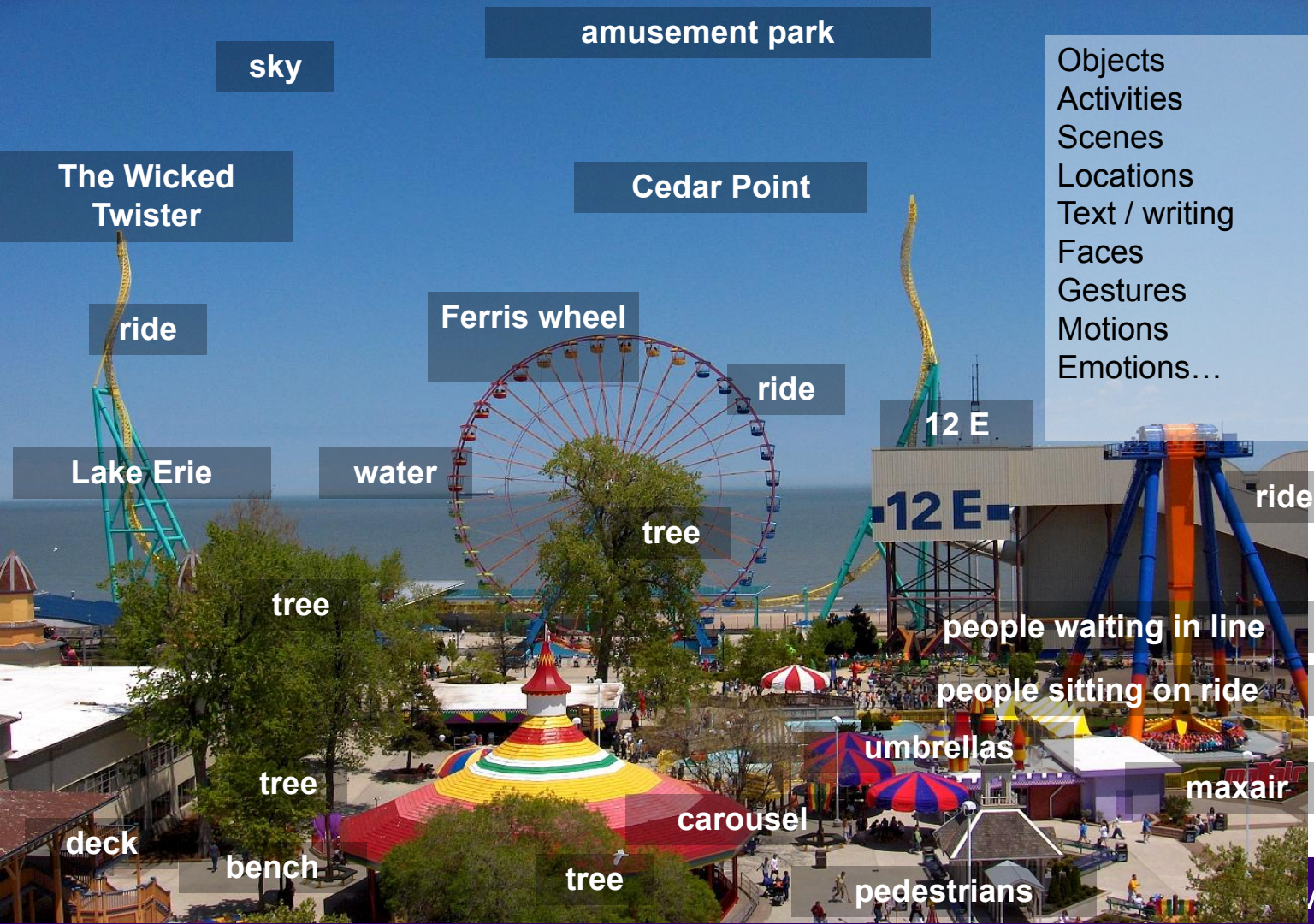
- History of understanding perception
- Introduction to computer vision
- Course overview

The goal of computer vision:
convert light into meaning



What kind of information can we extract from an image?

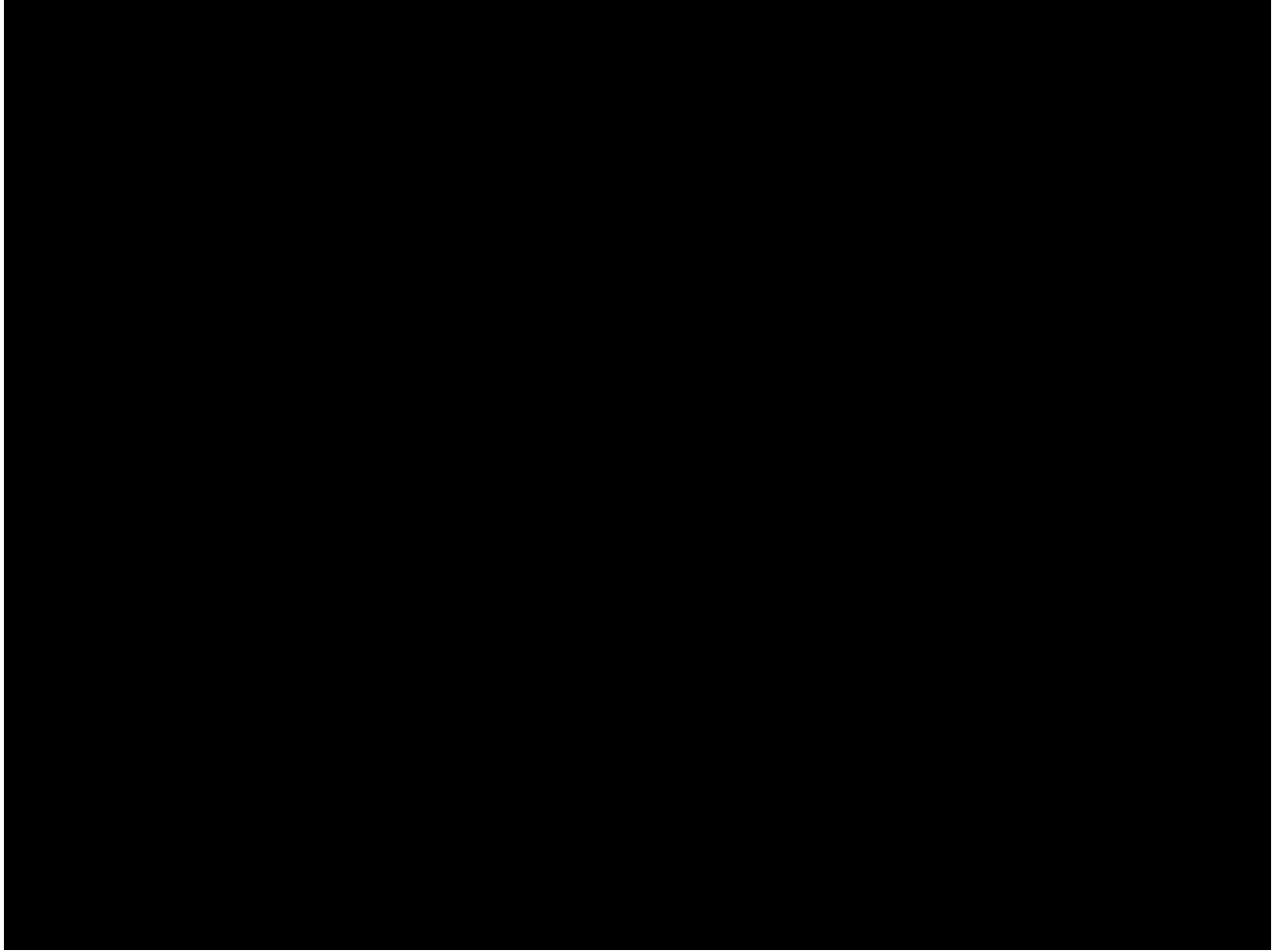
1. Semantic information
2. Geometric 3D information



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

Vision as a source of semantic information

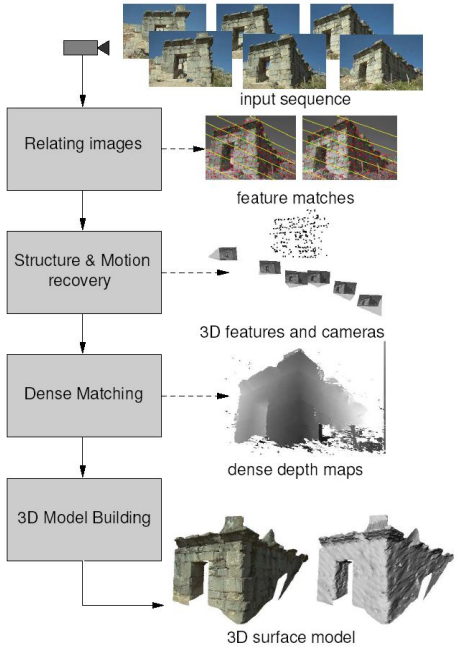
Extracting Semantic



Segment Anything
2023

Extracting geometric information

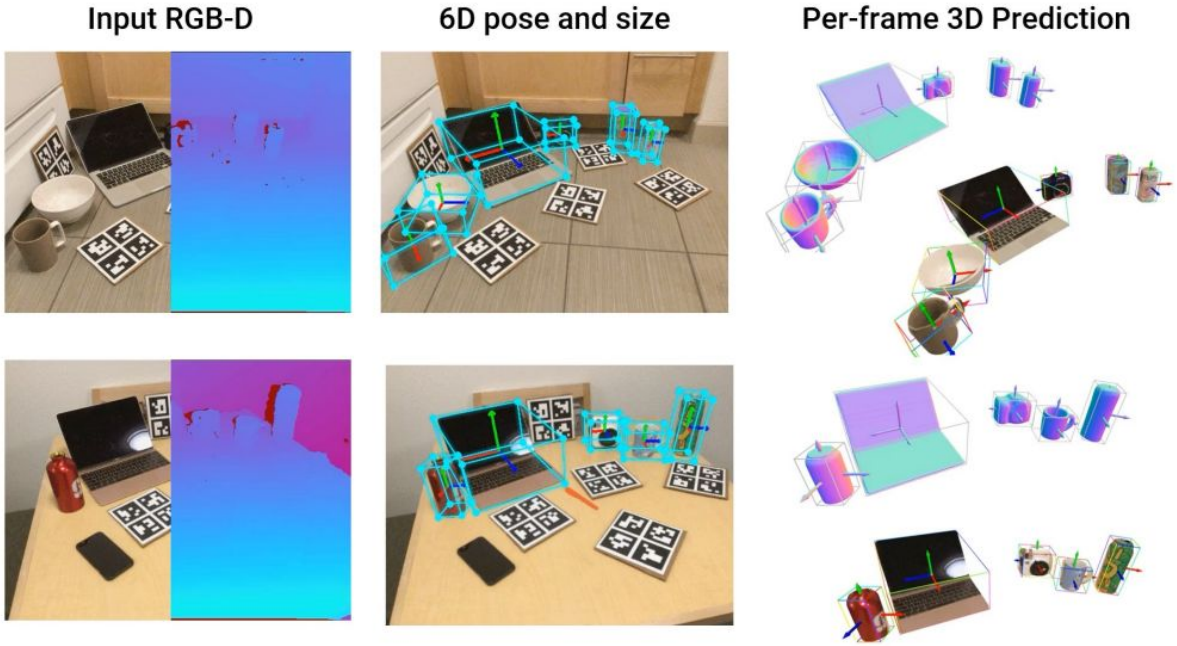
Real-time stereo



Pollefeys et al.

Goesele et al.

Geometric 3D information from 2D images

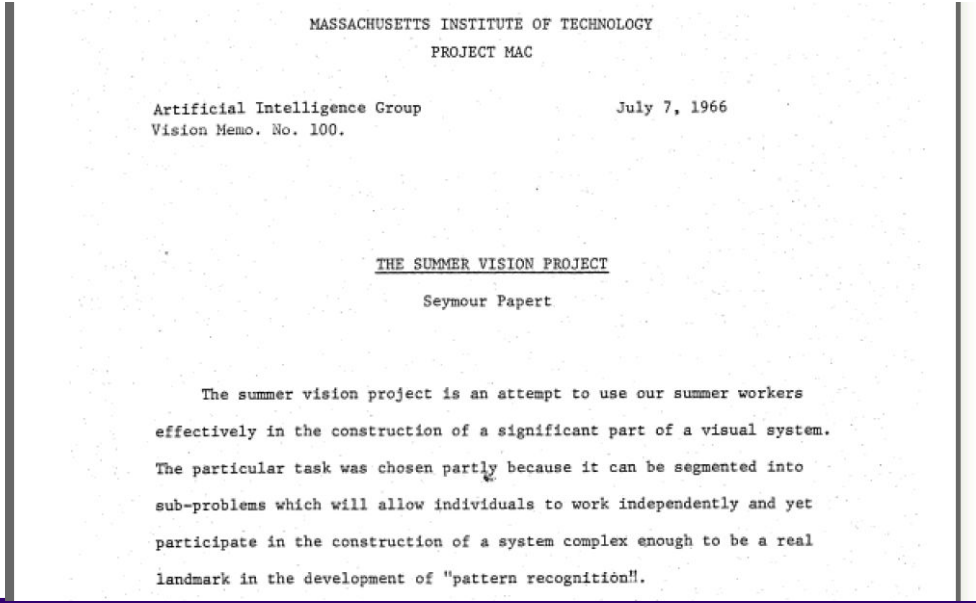


TRI & GATech's ShaPO (ECCV'22): https://zubair-irshad.github.io/projects/ShAPO.html

MIT thought that computer vision would be solved as an undergraduate summer project

"The primary goal of the project is to construct a system of programs which will divide a [...] picture into regions such as likely objects, likely background areas and chaos."

"The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects."



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

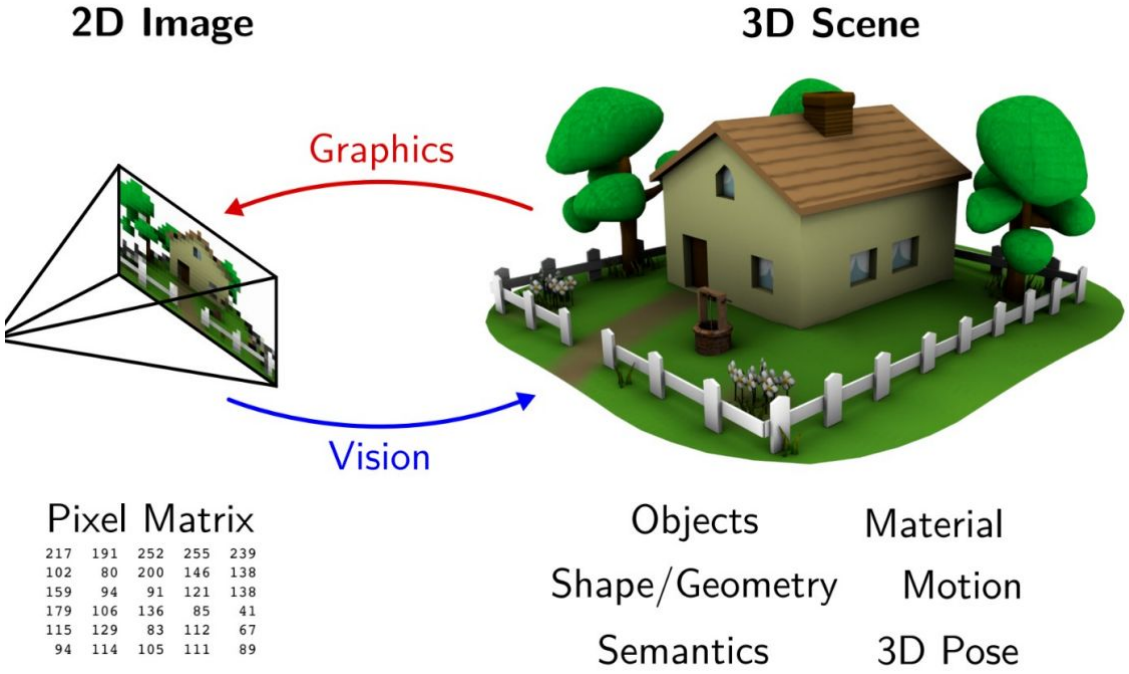
July 7, 1966

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".

But why is computer vision so hard?



It is an ill posed problem

Computers need to convert pixel intensities into meaning

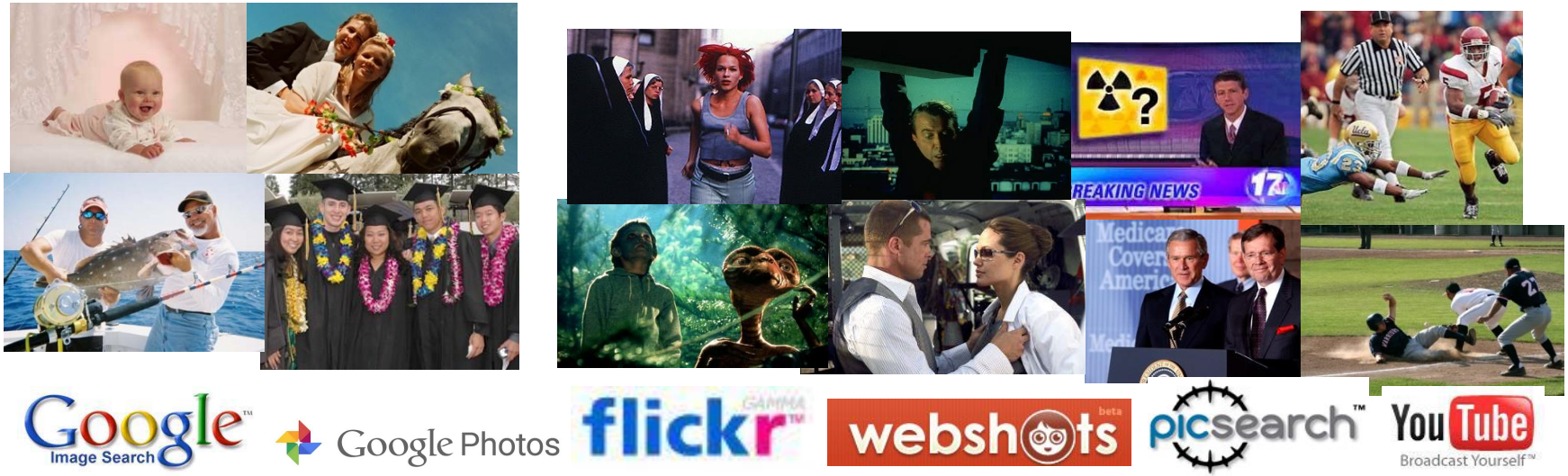
What we see

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

What a computer sees

Why study computer vision?

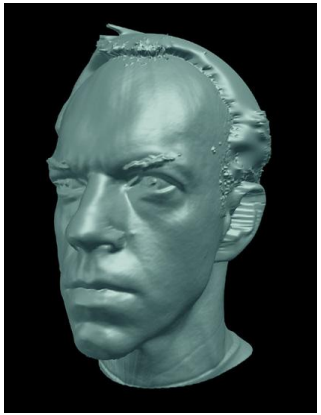
Vision is useful: Images and video are everywhere!



80% of all web traffic is images and videos

Majority of the internet is dark matter without computer vision

Special effects: shape and motion capture



3D urban modeling

Google Streetview - custom campus tours

3D urban modeling: Microsoft Photosynth

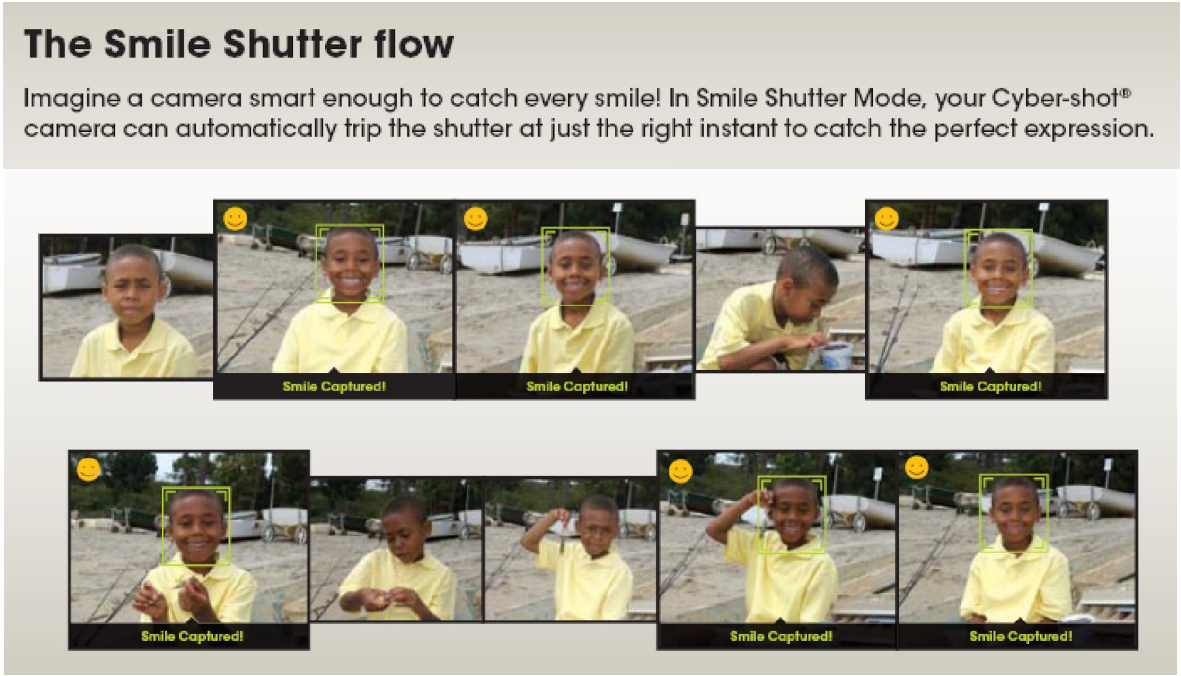
http://photosynth.net

Face detection

- Many digital cameras now detect faces
  - Canon, Sony, Fuji, …

Source: S. Seitz

Smile detection

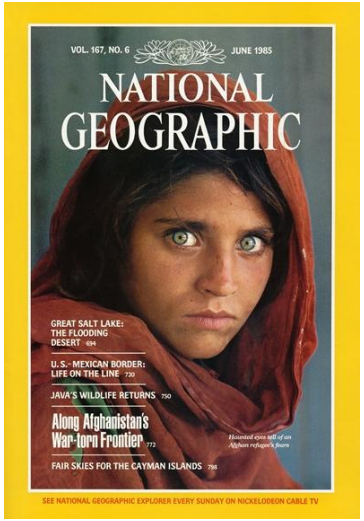


Sony Cyber-shot® T70 Digital Still Camera

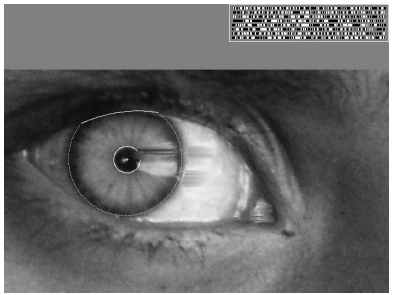
Source: S. Seitz

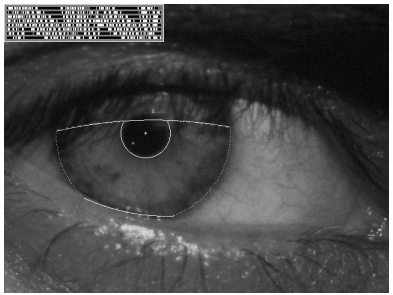
Face recognition: Apple iPhoto software

Biometrics



[How the Afghan Girl was Identified by Her Iris Patterns]

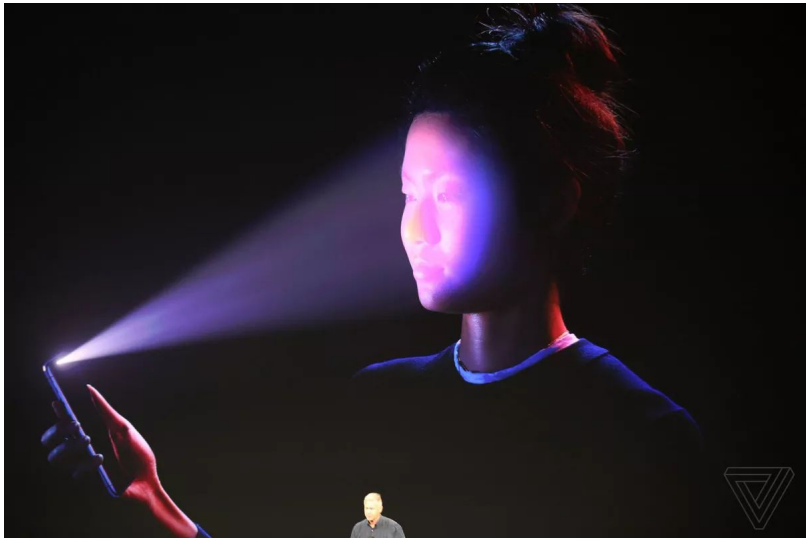




Biometrics

Fingerprint scanners on many new laptops, other devices

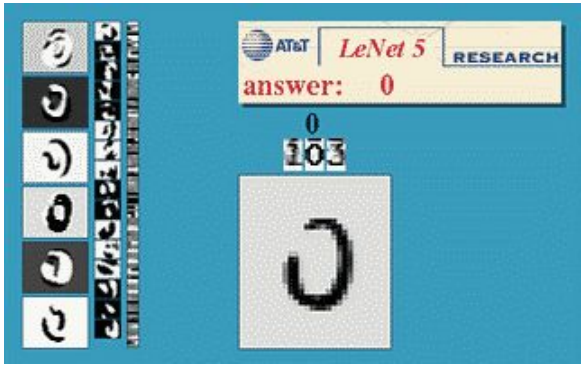
Face recognition systems now on iphones and samsungs



Optical character recognition (OCR)

Technology to convert scanned docs to text

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



Digit recognition, AT&T labs

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

Google maps: Annotate all houses and streets

Avenue des Sapins

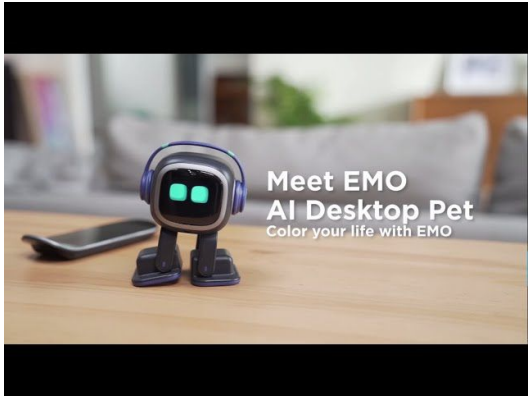
Goodfellow et al. 2014

Vision-powered toys and robots in the 2000s

Vision-powered toys and robots in the 2020s

Scout home security robot - monitors your house

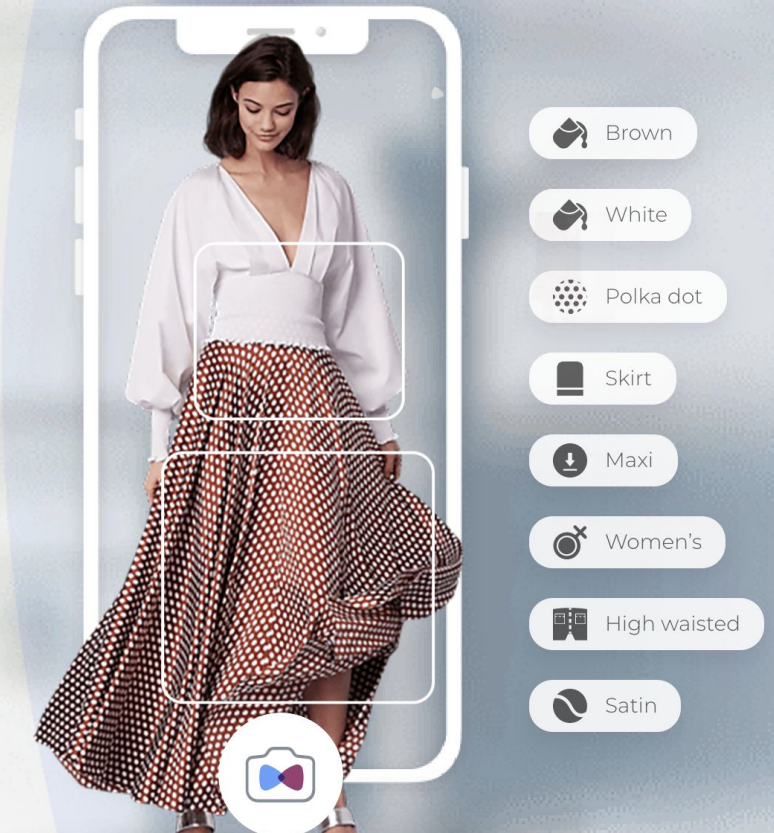
Unitree Go1’s companion robot - like a dog



The Leader in Visual AI for Retail

Syte changes the way retailers connect shoppers with the products that inspire them by delivering the best Visual AI technology for retail. Discover our solutions that empower retailers to increase customer engagement, and boost conversion and sales.

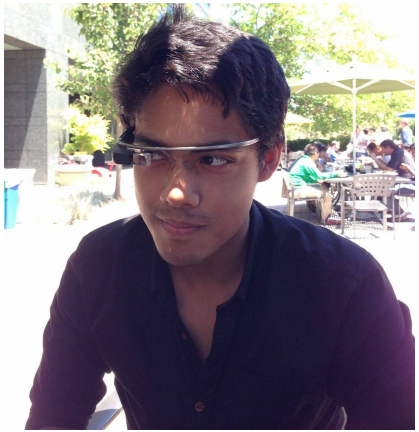
Enter Business Email | See Syte in Action! | Watch Video



TOMMY HILFIGER FARFETCH bon prix SHOPSTYLE

Apple Vision Pro, Snapstacles and Google glasses

me in undergrad ->



Automotive safety

Mobileye: Vision systems in high-end BMW, GM, Volvo models

Claimed that they would release self-driving cars by 2015. Still not there.

Sources: A. Shashua, S. Seitz

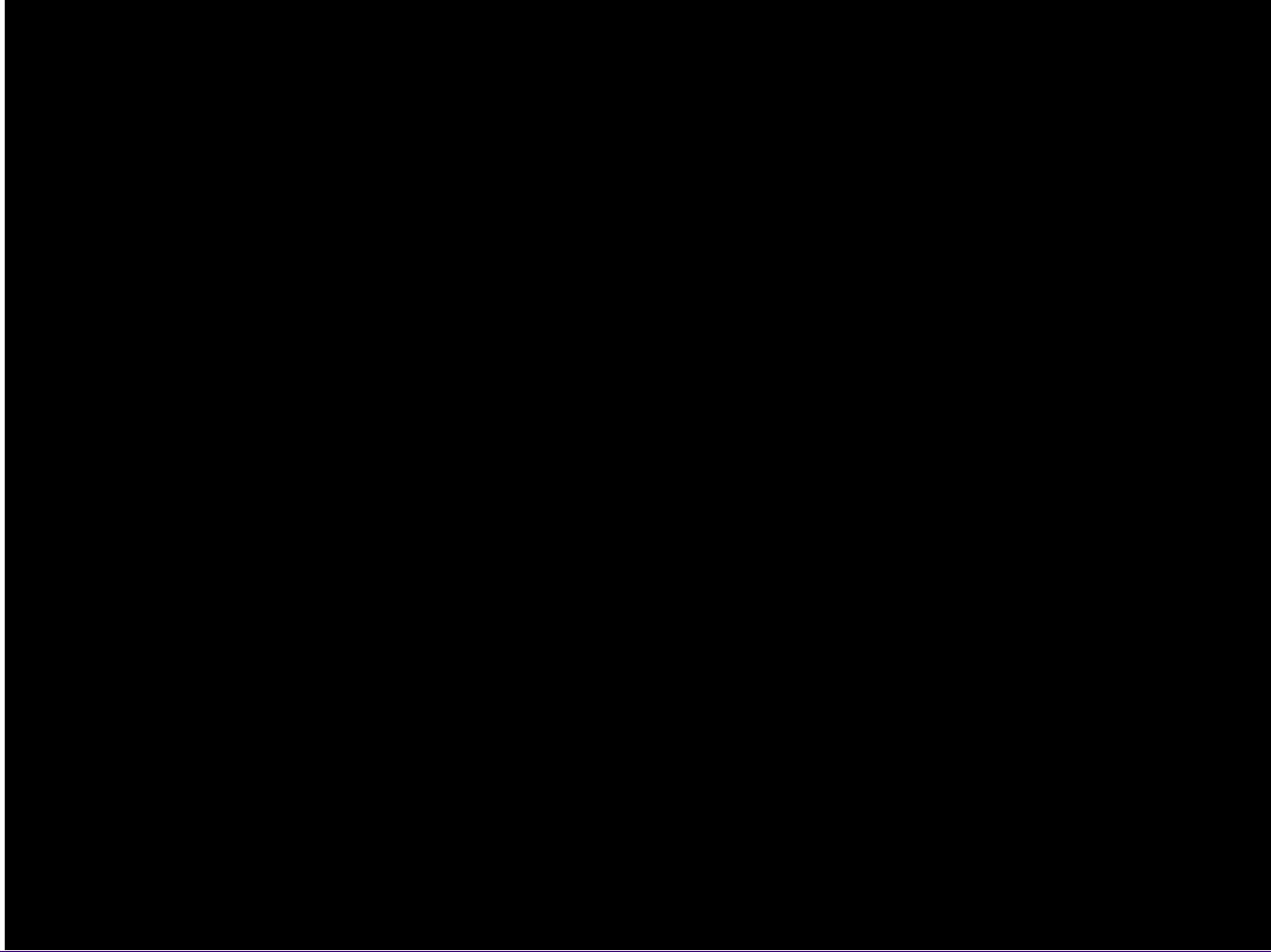
Vision in supermarkets

LaneHawk by EvolutionRobotics
(acquired by iRobot for $74M in 2012)

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk, you are assured to get paid for it… “

Source: S. Seitz

Amazon Go



Vision-based interaction (and games)

Microsoft's Kinect

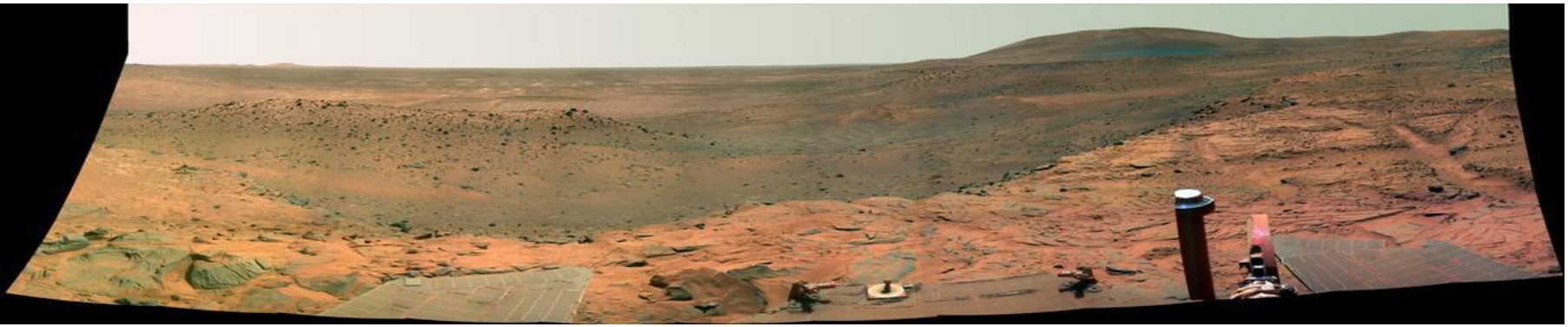
Assistive technologies

Sony EyeToy

Augmented Reality

Virtual Reality

Vision for robotics, space exploration



NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Source: S. Seitz

Machine Bias

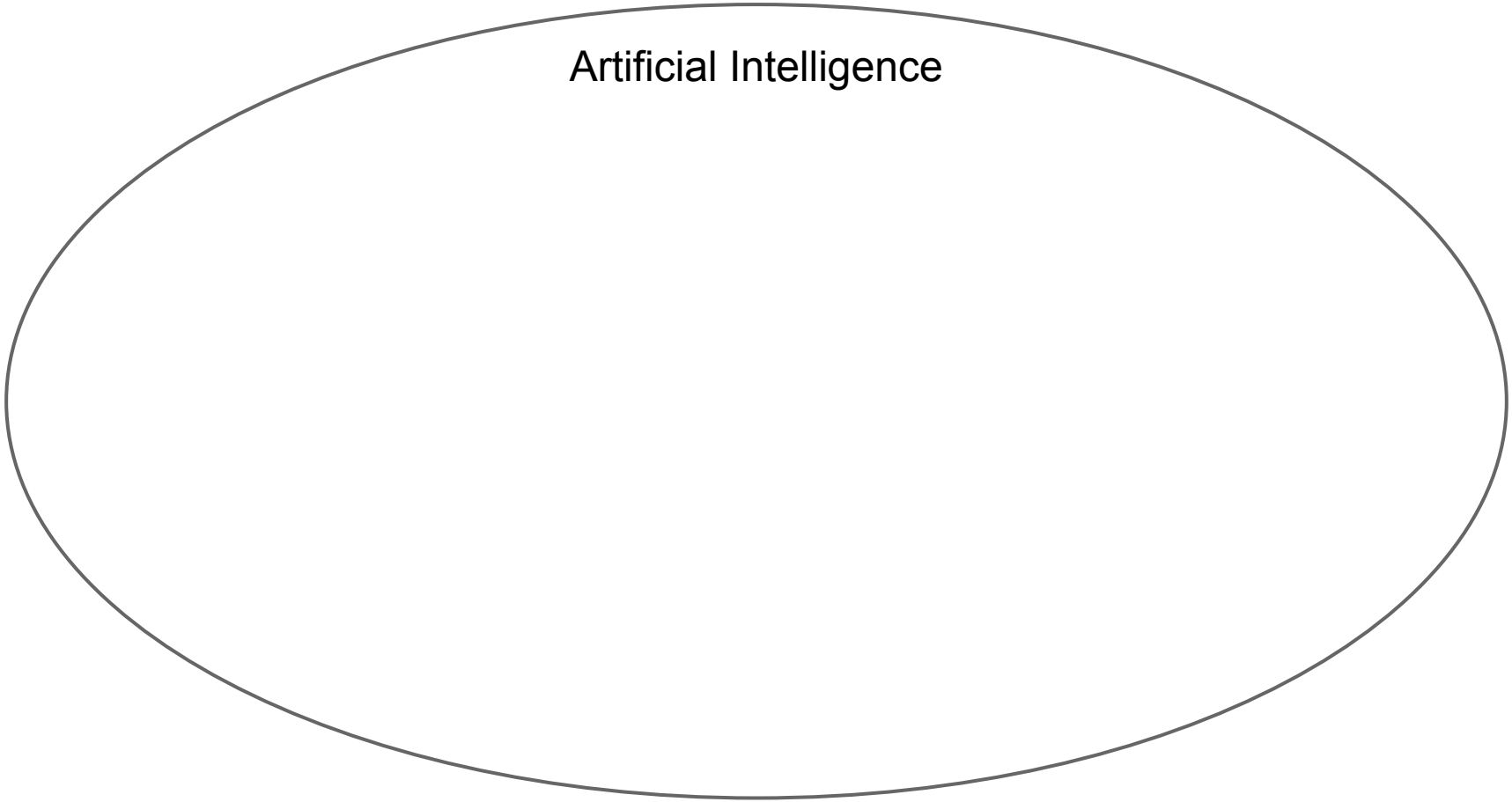
There's software used across the country to predict future criminals. And it's biased against blacks.

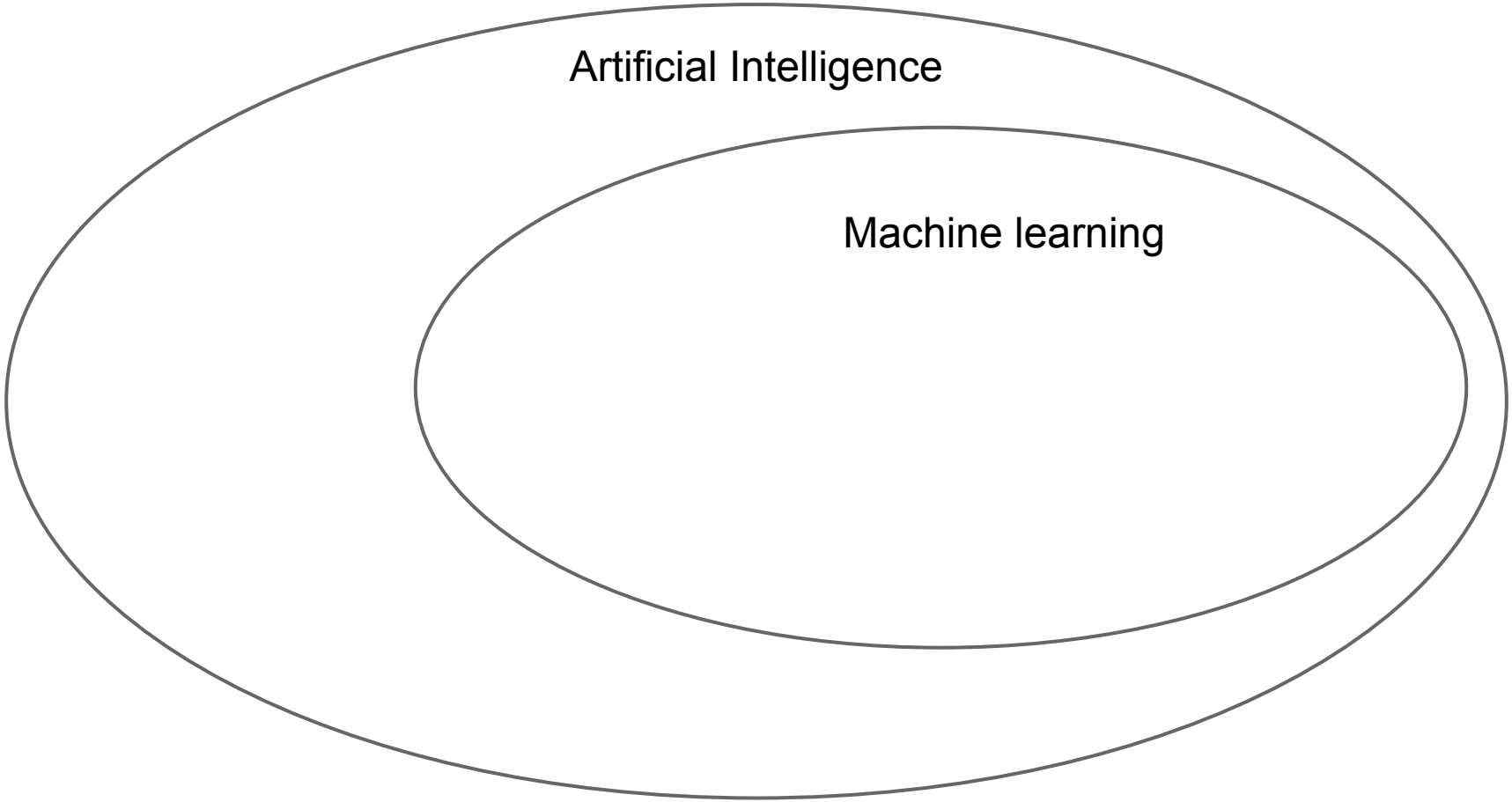
MGMT “When You Die”

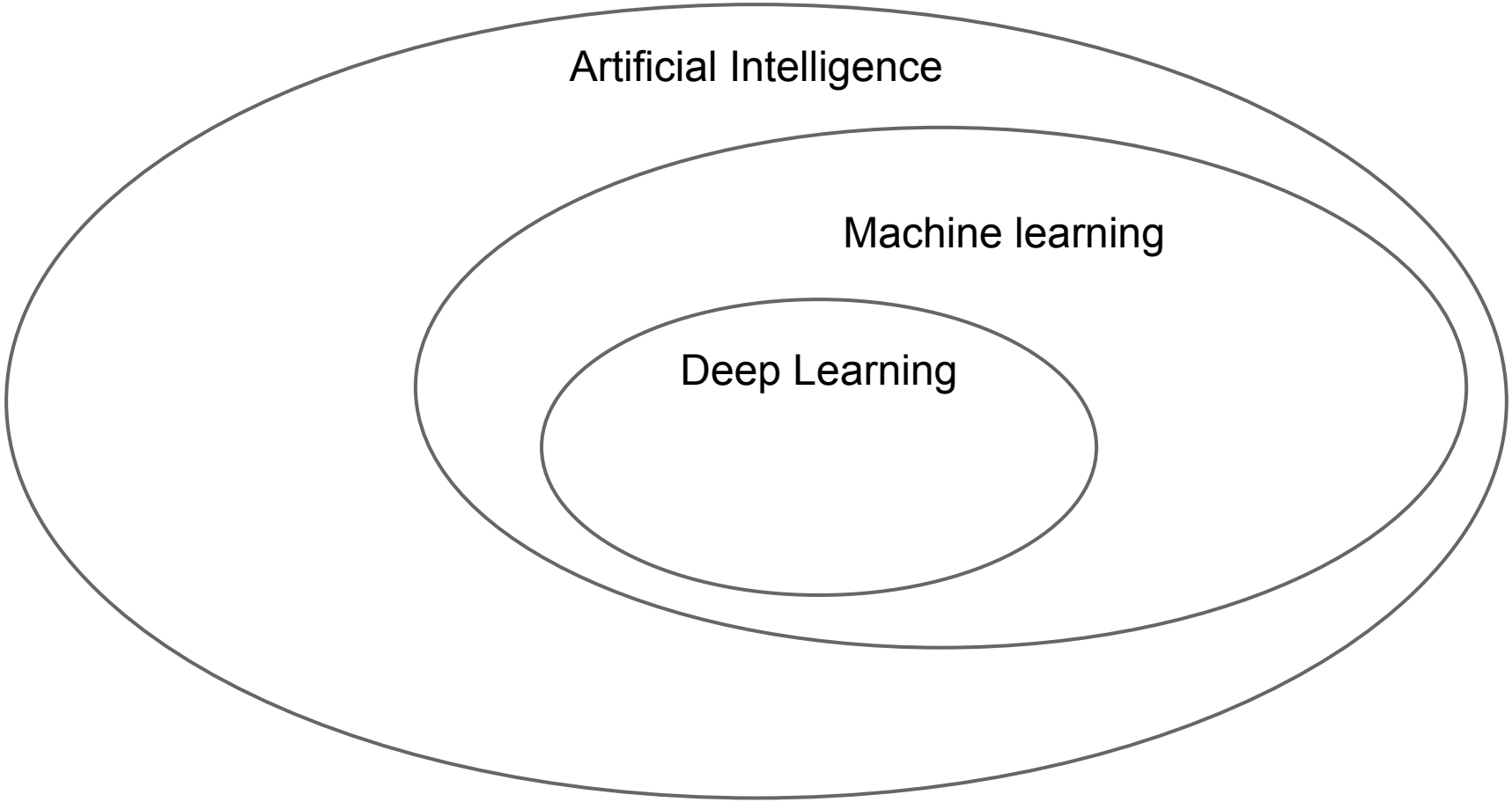
How should you make sense of computer vision as a field?

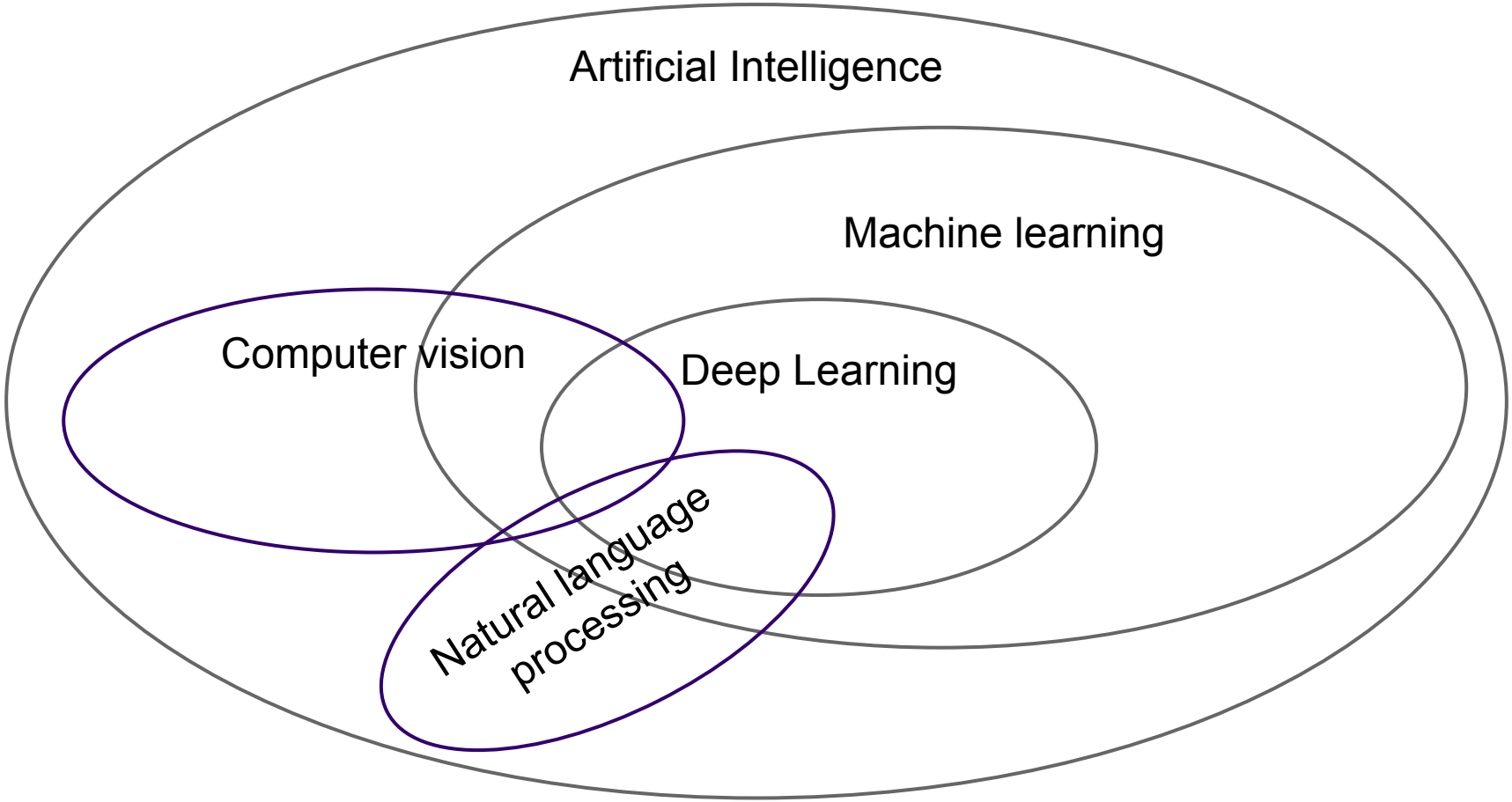
Let’s situate computer vision in the broader context of AI

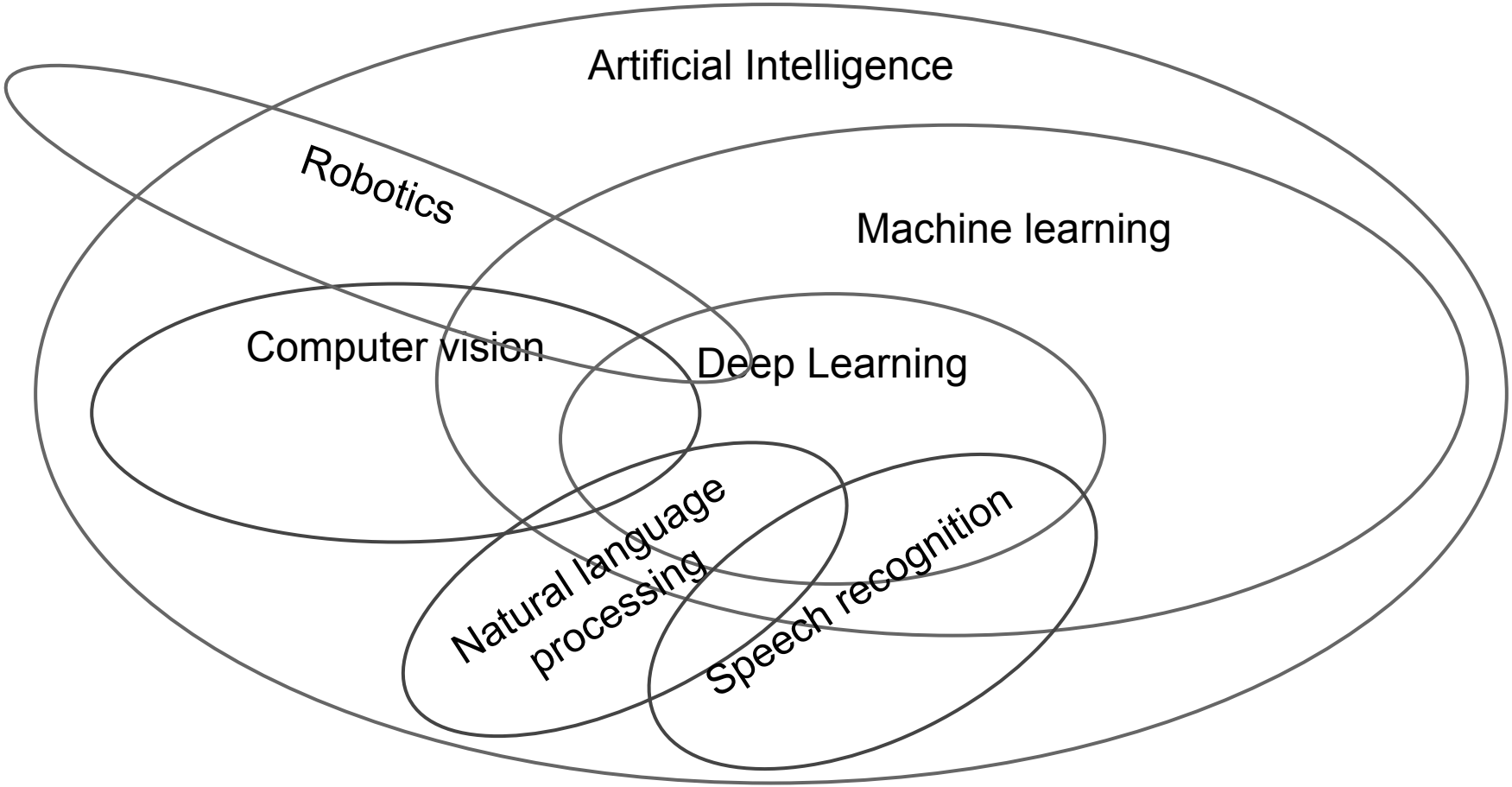
I want to change your conception of computer vision is

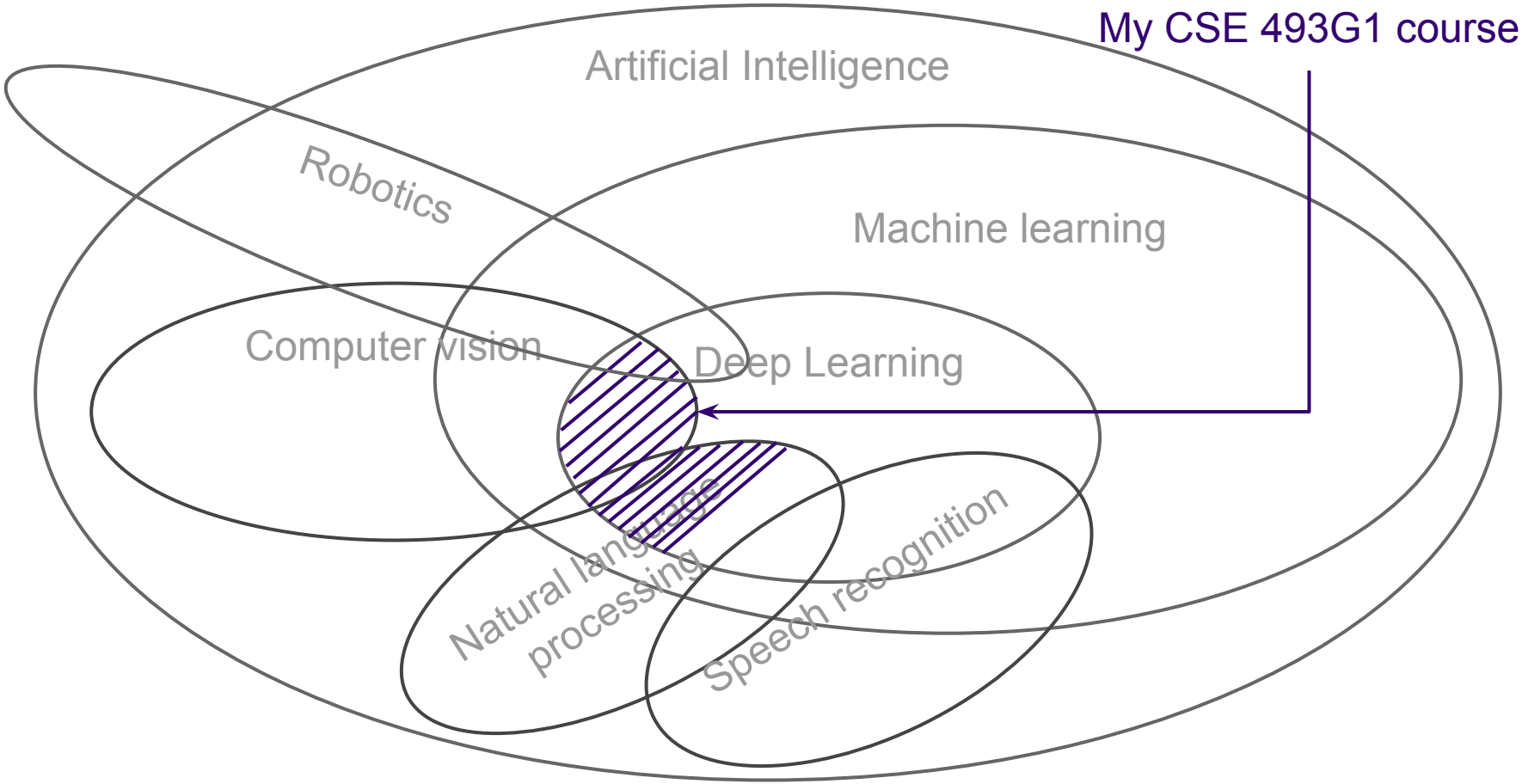

Artificial Intelligence

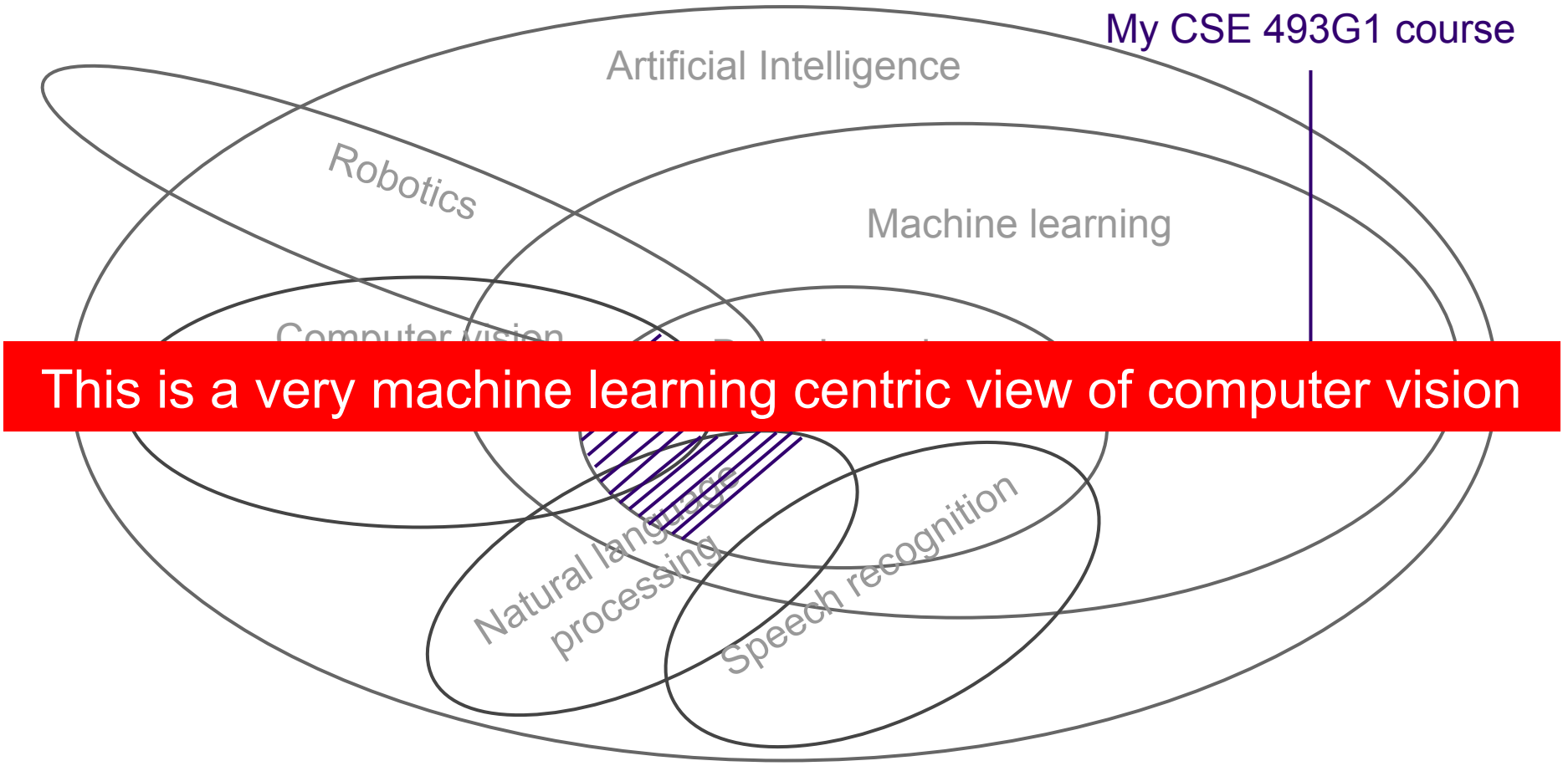

Artificial Intelligence
Machine learning


Artificial Intelligence
Machine learning
Deep Learning


Artificial Intelligence
Machine learning
Computer vision
Deep Learning
Natural language processing

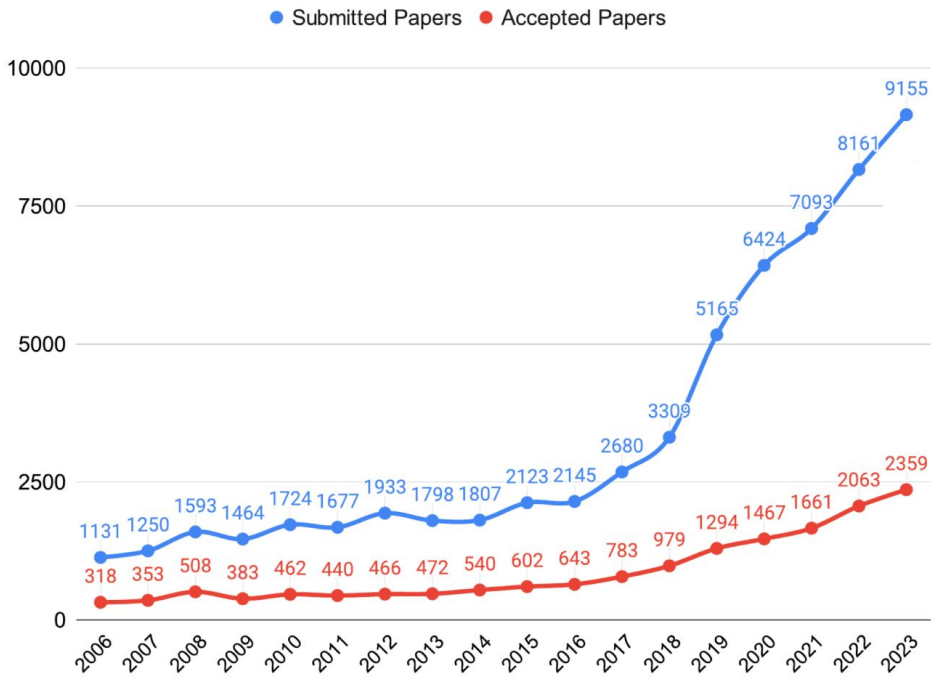

Artificial Intelligence
Robotics
Machine learning
Computer vision
Deep Learning
Natural language processing
Speech recognition


My CSE 493G1 course
Artificial Intelligence
Robotics
Machine learning
Computer vision
Deep Learning
Natural language processing
Speech recognition

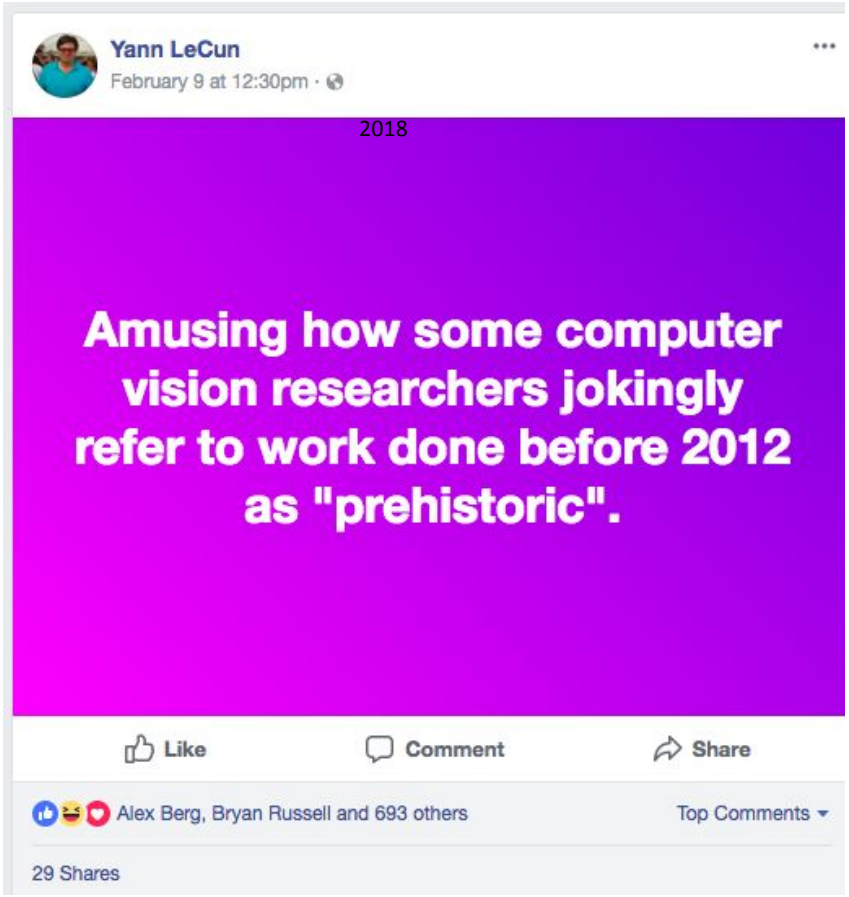


This is a very machine learning centric view of computer vision

CVPR has seen a large number of deep learning people enter



CVPR


Yann LeCun
February 9 at 12:30pm ·
2018
Amusing how some computer vision researchers jokingly refer to work done before 2012 as "prehistoric".
Like
Comment
Share
Alex Berg, Bryan Russell and 693 others
Top Comments
29 Shares


Serge Belongie I call it "antedeepluvian" (before the flood of deep learning papers)
Like · Reply · 5d
46
View 1 more reply
Yann LeCun Yeah, I like this porte manteau.
3
Like · Reply · 4d
Alex Berg "some"...
Like · Reply · 4d
Yann LeCun Yes, some. And jokingly.
1
Like · Reply · 4d
Aaron Hertzmann I'm just thrilled when I get a paper to review that even acknowledges the existence of related work published prior to 2012.
9
Like · Reply · 4d
Alyosha Efros Around 2005, I was at a vision workshop in MSRI where one of the Gemans said: "physics before Newton is now called 'miscellaneous early efforts'. Vision is at this same stage." Jury still out if we've seen our Newton yet. But looking hopeful.
9
Like · Reply · 5d · Edited

The Affective Growth of Computer Vision

Norman Makoto Su David J. Crandall
Luddy School of Informatics, Computing, and Engineering
Indiana University Bloomington

Abstract

The success of deep learning has led to intense growth and interest in computer vision, along with concerns about its potential impact on society. Yet we know little about how these changes have affected the people that research and practice computer vision: we as a community spend so much effort trying to replicate the abilities of humans, but so little time considering the impact of this work on ourselves. In this paper, we report on a study in which we asked computer vision researchers and practitioners to write stories about emotionally-salient events that happened to them. Our analysis of over 50 responses found tremendous affective (emotional) strain in the computer vision community. While many describe excitement and success, we found strikingly frequent feelings of isolation, cynicism, apathy, and exasperation over the state of the field. This is especially true among people who do not share the unbridled enthusiasm for normative standards for computer vision research and who do not see themselves as part of the "in-crowd." Our findings suggest that these feelings are closely tied to the kinds of research and professional practices now expected in computer vision. We argue that as a community with significant stature, we need to work towards an inclusive culture that makes transparent and addresses the real emotional toil of its members.

Academics are starting to get worried…

Choose Your Weapon:
Survival Strategies for Depressed AI Academics

Julian Togelius and Georgios N. Yannakakis*

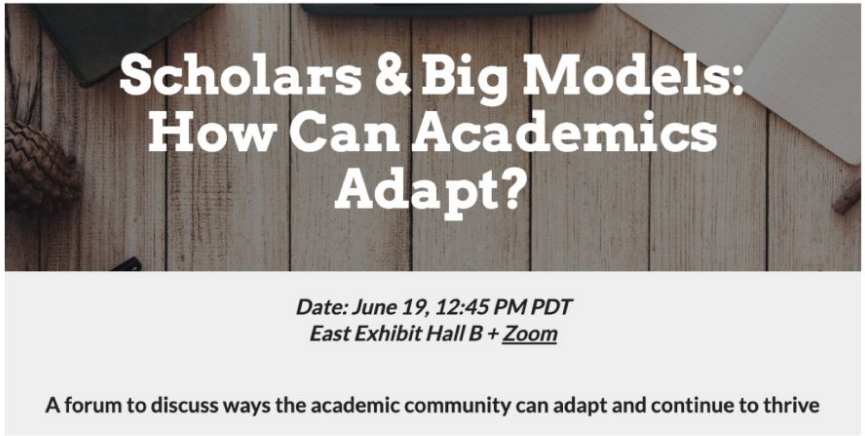
April 14, 2023

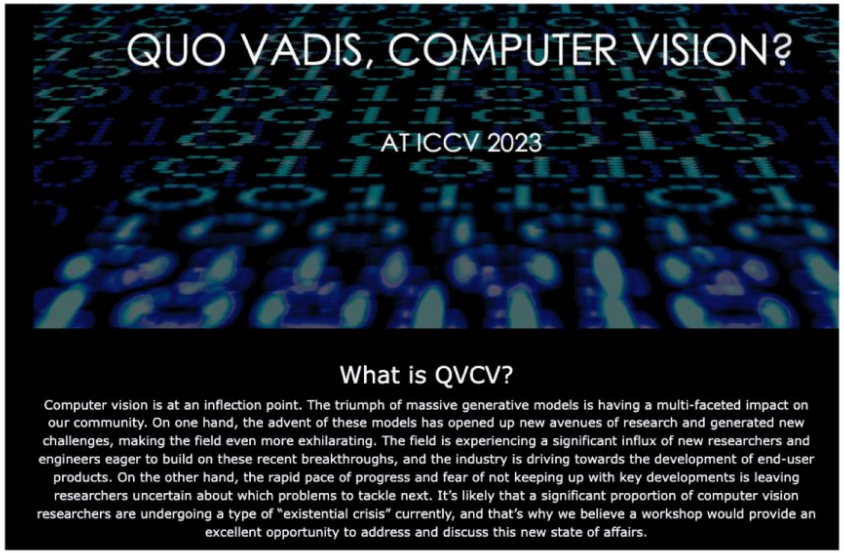
Abstract

Are you an AI researcher at an academic institution? Are you anxious you are not coping with the current pace of AI advancements? Do you feel you have no (or very limited) access to the computational and human resources required for an AI research breakthrough? You are not alone; we feel the same way. A growing number of AI academics can no longer find the means and resources to compete at a global scale. This is a somewhat recent phenomenon, but an accelerating one, with private actors investing enormous compute resources into cutting edge AI research. Here, we discuss what you can do to stay competitive while remaining an academic. We also briefly discuss what universities and the private sector could do improve the situation, if they are so inclined. This is not an exhaustive list of strategies, and you may not agree with all of them, but it serves to start a discussion.

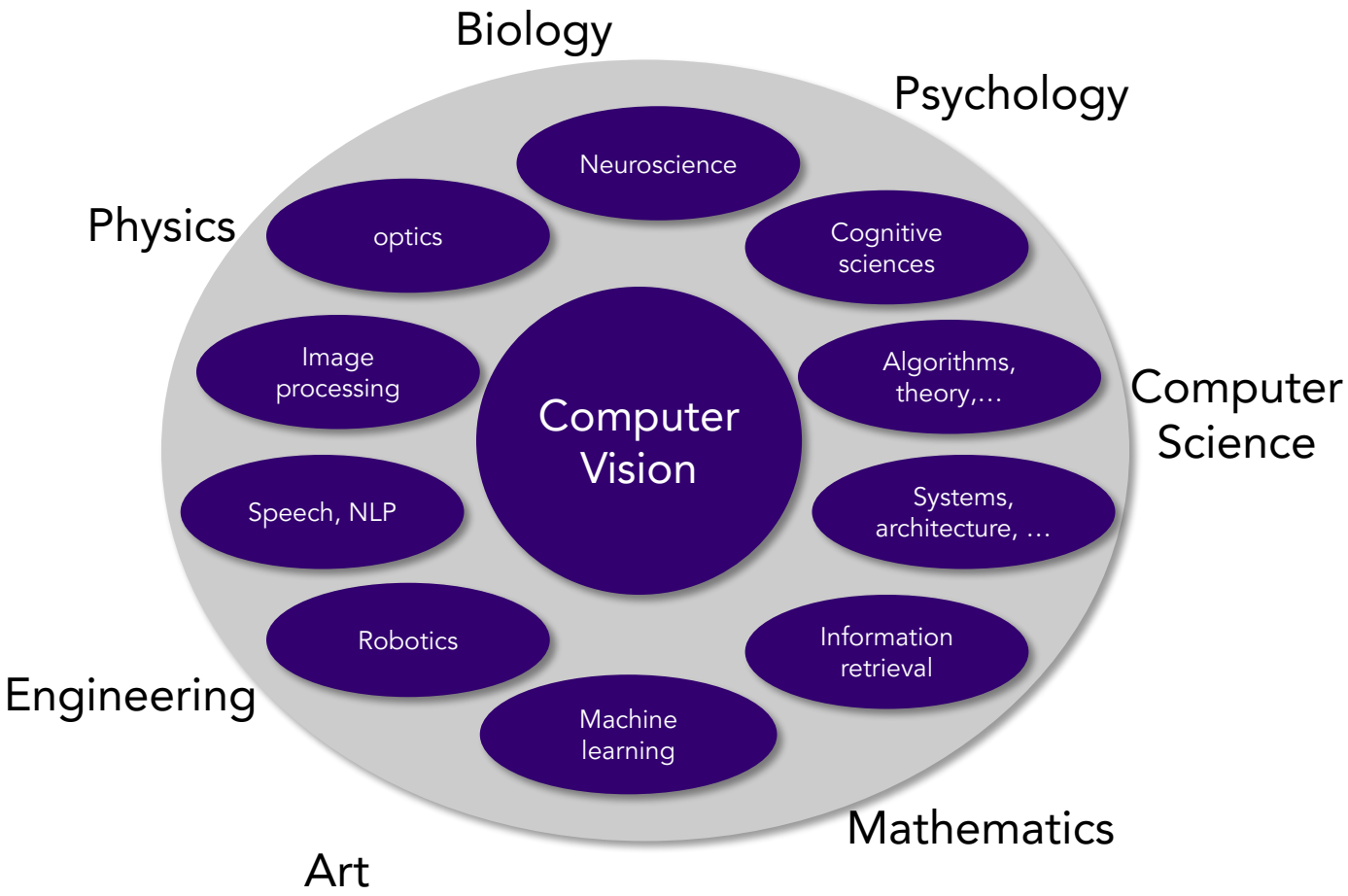
https://arxiv.org/pdf/2304.06035.pdf
https://vision.soic.indiana.edu/papers/affective2021cvpr.pdf

Workshops in 2023 in response

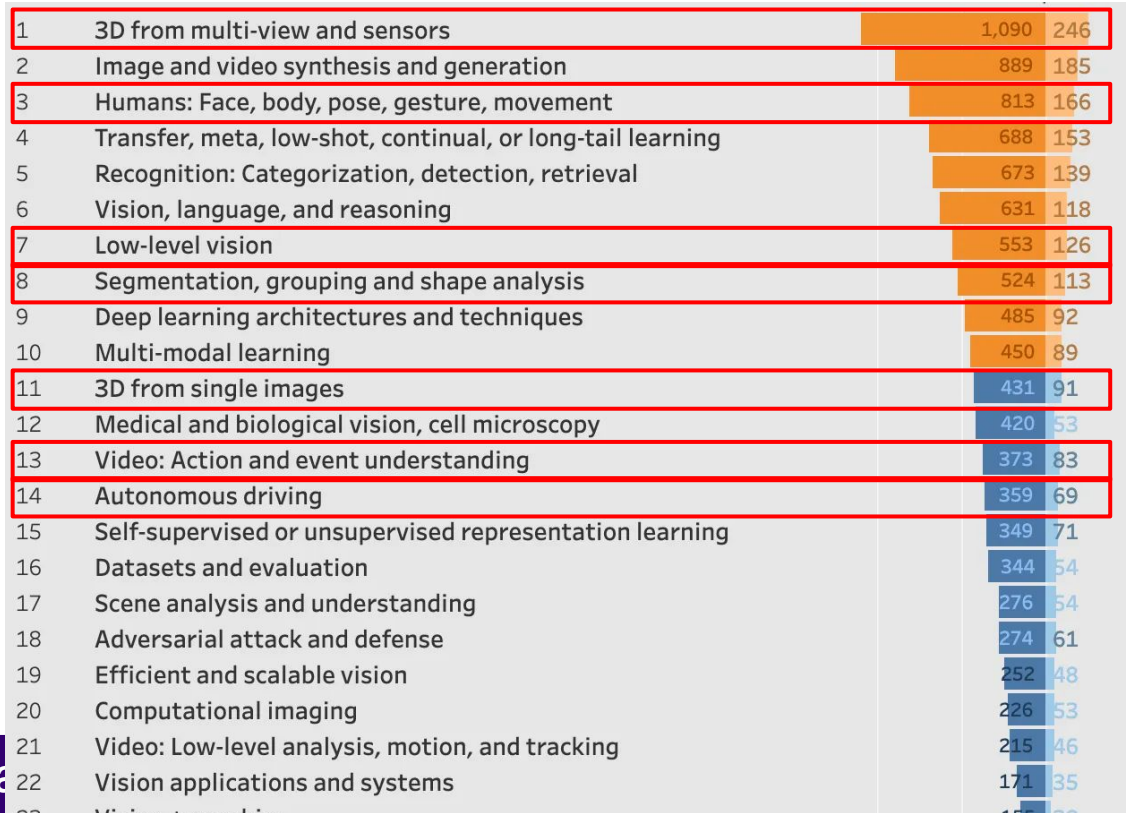




https://sites.google.com/view/academic-cv/
https://gkioxari.github.io/Tutorials/iccv2023/

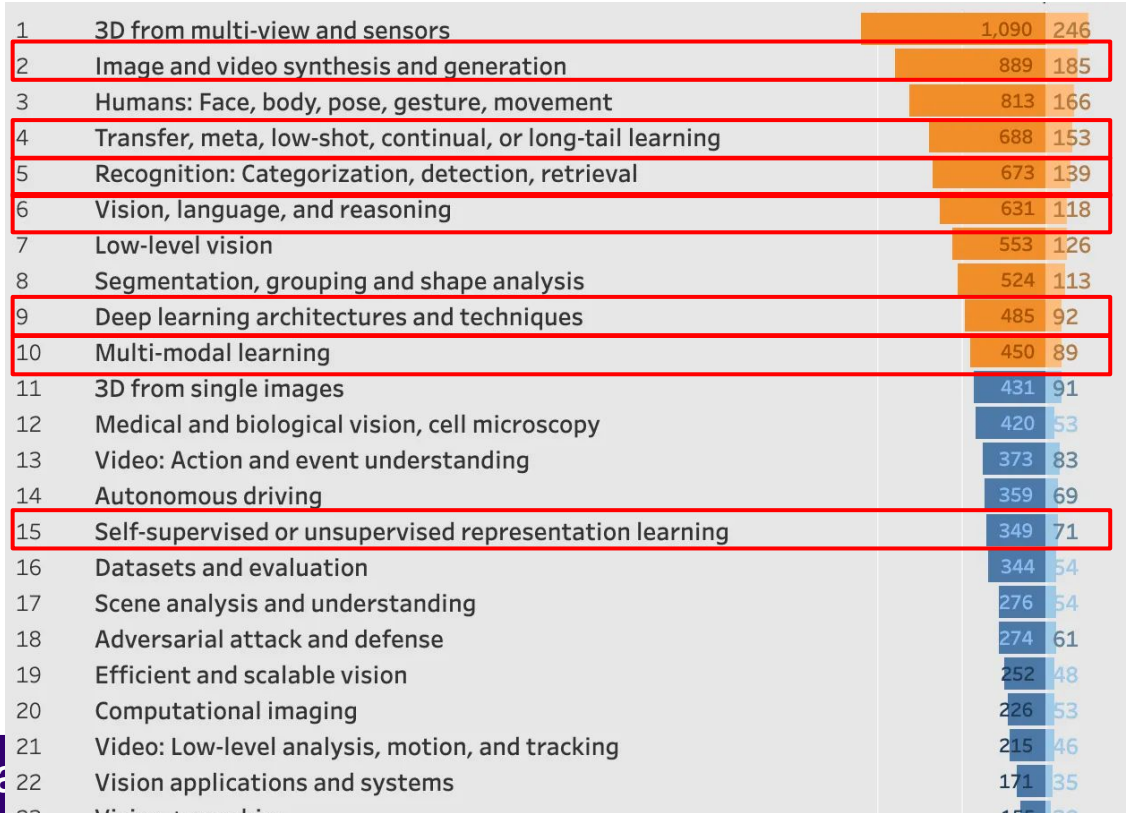

Biology
Psychology
Physics
Computer Science
Engineering
Mathematics
Art
Neuroscience
optics
Cognitive sciences
Image processing
Computer Vision
Algorithms, theory,…
Speech, NLP
Systems, architecture, …
Robotics
Information retrieval
Machine learning

CVPR is a lot more than just deep learning and semantics - what CSE455 will cover



1	3D from multi-view and sensors	1,090	246
2	Image and video synthesis and generation	889	185
3	Humans: Face, body, pose, gesture, movement	813	166
4	Transfer, meta, low-shot, continual, or long-tail learning	688	153
5	Recognition: Categorization, detection, retrieval	673	139
6	Vision, language, and reasoning	631	118
7	Low-level vision	553	126
8	Segmentation, grouping and shape analysis	524	113
9	Deep learning architectures and techniques	485	92
10	Multi-modal learning	450	89
11	3D from single images	431	91
12	Medical and biological vision, cell microscopy	420	53
13	Video: Action and event understanding	373	83
14	Autonomous driving	359	69
15	Self-supervised or unsupervised representation learning	349	71
16	Datasets and evaluation	344	54
17	Scene analysis and understanding	276	54
18	Adversarial attack and defense	274	61
19	Efficient and scalable vision	252	48
20	Computational imaging	226	53
21	Video: Low-level analysis, motion, and tracking	215	46
22	Vision applications and systems	171	35

What I cover in my CSE493G1 deep learning course



1	3D from multi-view and sensors	1,090	246
2	Image and video synthesis and generation	889	185
3	Humans: Face, body, pose, gesture, movement	813	166
4	Transfer, meta, low-shot, continual, or long-tail learning	688	153
5	Recognition: Categorization, detection, retrieval	673	139
6	Vision, language, and reasoning	631	118
7	Low-level vision	553	126
8	Segmentation, grouping and shape analysis	524	113
9	Deep learning architectures and techniques	485	92
10	Multi-modal learning	450	89
11	3D from single images	431	91
12	Medical and biological vision, cell microscopy	420	53
13	Video: Action and event understanding	373	83
14	Autonomous driving	359	69
15	Self-supervised or unsupervised representation learning	349	71
16	Datasets and evaluation	344	54
17	Scene analysis and understanding	276	54
18	Adversarial attack and defense	274	61
19	Efficient and scalable vision	252	48
20	Computational imaging	226	53
21	Video: Low-level analysis, motion, and tracking	215	46
22	Vision applications and systems	171	35

Decade by decade

- **1960s**: Image processing and pattern recognition, blocks world
- **1970s**: Key recovery problems defined: structure from motion, stereo, shape from shading, color constancy. Attempts at knowledge-based recognition
- **1980s**: Fundamental and essential matrix, multi-scale analysis, corner and edge detection, optical flow, geometric recognition as alignment
- **1990s**: Multi-view geometry, statistical and appearance-based models for recognition, first approaches for (class-specific) object detection
- **2000s**: Local features, generic object recognition and detection
- **2010s**: Deep learning, big data

Adapted from J. Malik

CVPR is happening here in June 2024

https://cvpr.thecvf.com/

Why should you go to CVPR? It is ranked #4 amongst all scientific publications across all disciplines

	Publication	h5-index	h5-median
1.	Nature	467	707
2.	The New England Journal of Medicine	439	876
3.	Science	424	665
4.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	422	681
5.	The Lancet	368	688
6.	Nature Communications	349	456
7.	Advanced Materials	326	415
8.	Cell	316	503

Source: Google scholar

Today's agenda

- History of computer vision
- Introduction to computer vision
- Course overview

Who is Ranjay?

Ranjay Krishna (Assistant Professor at UW CSE)
- PhD from Stanford
- I worked with Fei-Fei Li (**AI**)
- And with Michael Bernstein (**HCI**)

I conduct two types of **research inquiries**:
- I study emergent **human behaviors** when they interact with AI systems
- I develop better **AI** (specifically **computer vision**) systems with these insi

Past courses:
- UW CSE 493G1 [2023, 2024]: Deep learning for computer vision
- UW CSE 599H [2023]: Artificial intelligence vs intelligence augmentation
- Stanford CS 231N [2020, 2021]: Convolutional neural networks for computer vision
- Stanford CS 131 [2017, 2018, 2019]: Computer vision fundamentals and applications

Who is Jieyu?

Jieyu Zhang (4th year PhD student at UW CSE)
- I work with Ranjay Krishna

My research focus on data-centric AI/ML:
- I study how to build comprehensive and faithful AI evaluation system
- I study how to create high-quality training dataset for training AI models
- I develop AI-based autonomous agents capable of solving real-world tasks

Course staff (Office hours coming soon)

Instructors

Ranjay Krishna
Hours: Tue, 9:00-10:00am
CSE2 304
ranjay@cs.washington.edu

Jieyu Zhang
Hours: ,
CSE2 TBD
jieyuz2@cs.washington.edu

Teaching Assistants

Mahtab Bigverdi
Hours: ,
TBD
mahtab@cs.washington.edu

Xiaojuan Wang
Hours: ,
TBD
xiaojwan@cs.washington.edu

Vivek Jayaram
Hours: ,
CSE2 TBD
vjayaram@cs.washington.edu

Zihan Wang
Hours: ,
CSE2 TBD
avinwang@cs.washington.edu

Fatemeh Ghezloo
Hours: ,
CSE2 TBD
fghezloo@cs.washington.edu

Minh Hoang
Hours: ,
CSE2 TBD
minh257@cs.washington.edu

Class times

Lectures

- Tuesdays and Thursdays
  10:00am to 11:20am
  @CSE2 G20

Recitations

- Fridays
  12:30-1:20pm
  @ JHN 102

Lecture recordings

Will be made available on canvas:
https://canvas.uw.edu/courses/1718581/

Come to class!

Contacting instructor and TAs

- All announcements, Q&A in EdStem
  - https://edstem.org/us/courses/57280
  - All course related posts should be public.

- All private correspondences to course staff should post private (instructors only) post on EdStem.
  - Use this for personal problems, and debugging help to avoid showing other people your solutions.
  - If you have questions that others can benefit from, do a public post.

How to think about computer vision?

Breadth

- Computer vision is a huge field
- It can impact every aspect of life and society
- It is driving the current generative AI revolution
- Pixels are everywhere in our lives and cyber space
- CSE455 is meant as an broad overview course,
  - we will not cover all topics of CV
- Lectures are mixture of detailed techniques and high level ideas
- I want to teach you to speak our “language”

Depth

- …

How to think about computer vision?

Breadth

- …

Depth

- Computer vision is a highly technical field, i.e. know your math!
- Master bread-and-butter techniques: face recognition, corners, lines, features, optical flows, clustering and segmentation
- Programming assignments: be a good coder AND a good writer
- Math problem questions: know your concepts!
- Final Exam: your chance to shine!

Syllabus

Official website

https://courses.cs.washington.edu/courses/cse455/24sp/

You can compare this course against the winter quarter and last year's.

Everything has changed!

What is new?

All assignments are new!

- No more coding in C. Everything is in Python
- We are moving everything to Google Colab

All slides are all new.

- There is some overlap with topics from previous years but taught differently
- This is how I understand things in vision.

What can go wrong with all the changes?

We might invariably introduce errors in the assignments.

We have have mistakes in the slides.

Help us detect and fix them! I will give you extra credit if you post errors on EdStem.

Grading policy

75%: 5 Assignments.

- **15%** per Assignment.

24%: 1 Final Exam.

1-3%: Course Participation in lectures.

10%: Extra Credit - in assignments and in final exam.

Grading policy - Assignments

- **Assignment 0** (Using Colabs, Python basics)
  - Recommended Due by Apr 04 (Ungraded)
- **Assignment 1** (Filters, Convolutions, Edges)
  - Due Apr 16, 11:59 PST
- **Assignment 2** (Keypoints, Panaromas, Seam Carving)
  - Due Apr 25, 11:59 PST
- **Assignment 3** (Cameras, Clustering, Segmentation)
  - Due May 7, 11:59 PST
- **Assignment 4** (kNN, PCA, LDA, Detection)
  - Due May 16, 11:59 PST
- **Assignment 5** (Optical Flow, Tracking, Machine Learning)
  - Due May 28, 11:59 PST

Grading policy - assignments

- Most assignments will have an extra credit worth 1% of your total grade.
- Late policy
  - 5 free late days – use them in your ways
  - Maximum of 2 late days per assignment
  - Afterwards, 25% off per day late
- Collaboration policy
  - Read the student code book, understand what is ‘collaboration’ and what is ‘academic infraction’
  - We have links to this on the course webpage

Submitting homeworks

- **Homeworks** will consist of python files with code and jupyter notebooks.

- **Jupyter notebooks**:
  - Will guide you through the assignments.
  - Might contain written questions
  - Once you are done, convert the ipython notebook into a pdf and submit on Gradescope (https://www.gradescope.com/courses/755852).
    - Access code: **6G2NBR**

- **Python files**:
  - All code must be submitted to Gradescope as well.
  - Check our course website for details on submissions.

- **A0** will be live soon, you can start working on it immediately. We will try and get all the assignments out to you as soon as they are ready.

Final exam

- Monday June 3rd 10:30am - 12:20pm @ CSE2 G20
  - Optional make up exam: 9:30am-11:20am CSE2 371
  - We will send out form for students to apply to take the make up
- Will contain written questions from the concept covered in class or any questions in the homeworks.

- Can require you to solve technical math problems.

- Will contain a lot of multiple choice and true-false questions. We will release a practice final towards the end of the quarter.

Why should you take the class?

- Become a vision researcher
  - CVPR 2019 conference
  - ICCV 2019 conference
- Become a vision engineer in industry
  - Perception team at Google AI
  - Vision at Google Cloud
  - Vision at Facebook AI
- General interest

CSE 455 Roadmap

Pixels	Segments	Images	Videos	Web
Convolutions Edges Descriptors	Resizing Segmentation Clustering	Recognition Detection Machine learning	Motion Tracking	Neural networks Convolutional neural networks

From Convolutions to Convolutions

Welcome to CSE455

Let's have a fun quarter!