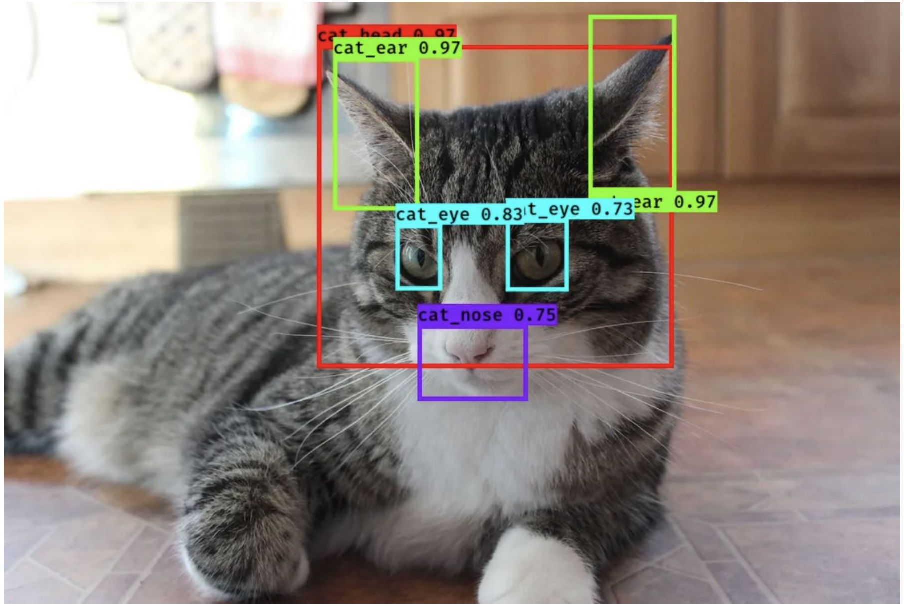
CS-GY 6643
Computer Vision

Lecture 1: Course introduction and survey of topics

Prof. Erdem Varol



Welcome to Computer Vision!

Instructor: Prof. Erdem Varol

Email: ev2240@nyu.edu

Office Hours: On Zoom,15 min appointments.

TA: Rishabh Raj

Email: rr4574@nyu.edu

Office Hours: Tuesdays 1-2pm on Zoom.

TA: Malhar Patel

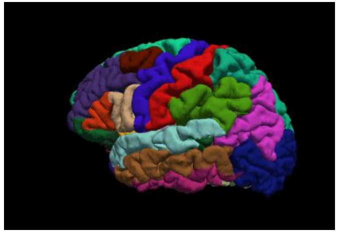
Email: mkp6112@nyu.edu

Office Hours: Mondays 1-2pm on Zoom

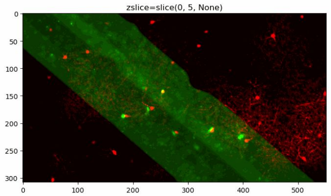
Course website: https://www.neuroinformaticslab.com/computer-vision-cs-gy-6643

Briefly about myself…

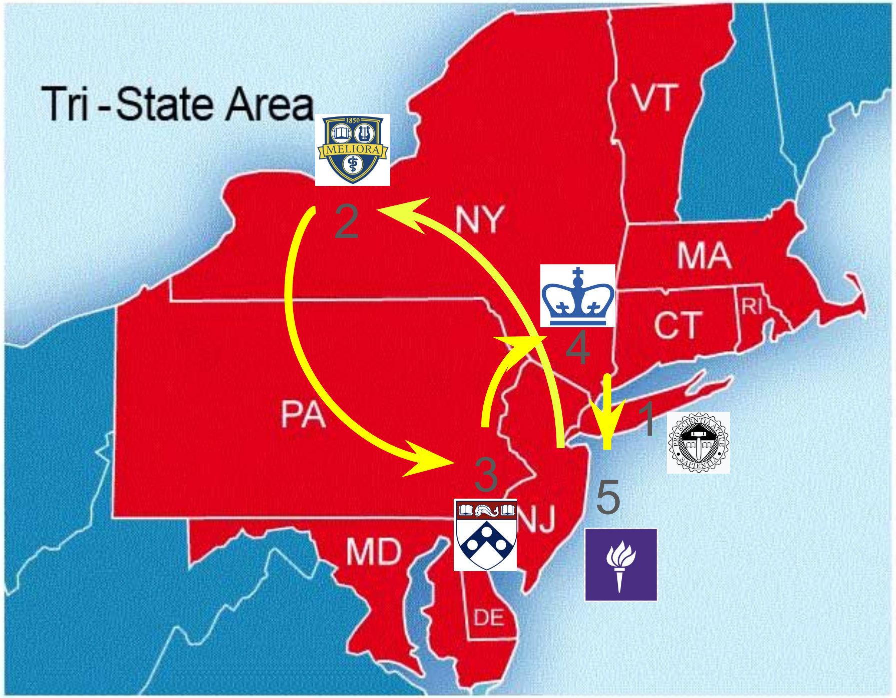
Ph.D. in EE at



Postdoc in Comp. Neuroscience at



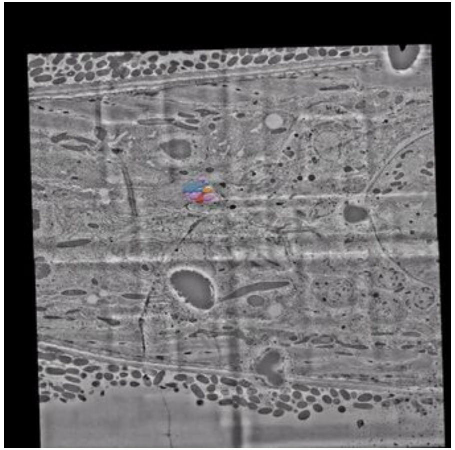
At NYU as an asst. prof. since 2023.



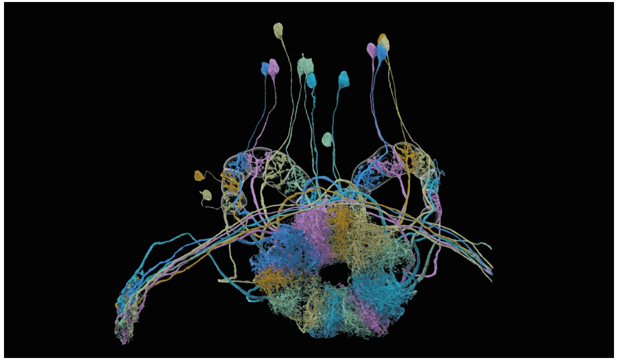
Our research lab focuses on computer vision with neuroscience applications

There might be potential to do a final project that is also publishable as research.

We encourage you to come talk to us if you are interested in any bio/neuro focused computer vision projects.



Witvliet 2022



FlyEM project

NEUROINFORMATICS LAB

Lab website: https://www.neuroinformaticslab.com/

Today’s menu

Why is computer vision cool?

Survey of topics

What is this class going to cover and not cover?

Course information, logistics and grading

Why is computer vision cool?

To appreciate its coolness, (I think) we must appreciate its roots.

It all started with a cat and a electrical wire (1957)

SCIENCE, VOL. 125

Tungsten Microelectrode for Recording from Single Units

DAVID H. HUBEL
Department of Neurophysiology, Walter Reed Army Institute of Research, Walter Reed Army Medical Center, Washington, D.C.

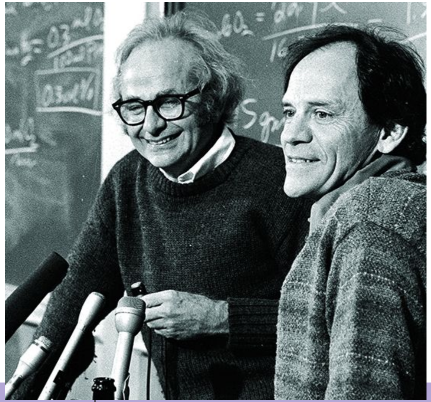
22 MARCH 1957

J. Physiol. (1959) 148, 574–591

RECEPTIVE FIELDS OF SINGLE NEURONES IN THE CAT'S STRIATE CORTEX

BY D. H. HUBEL* AND T. N. WIESEL*
From the Wilmer Institute, The Johns Hopkins Hospital and University, Baltimore, Maryland, U.S.A.

(*Received* 22 *April* 1959)



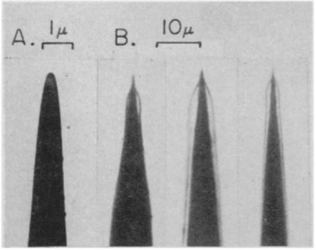


Fig. 1. (*A*) Electromicrograph of an uncoated, sharpened tungsten wire; (*B*) optical photomicrographs of coated electrodes immersed in water to show the coating.

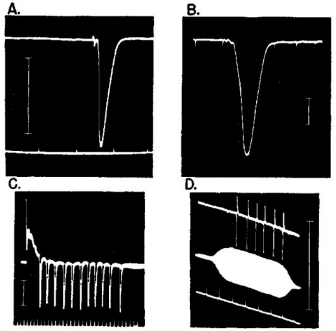


Fig. 2. Single-unit action potentials re-

Cat visual system is hierarchically composed of cells that perform "simple" tasks

106

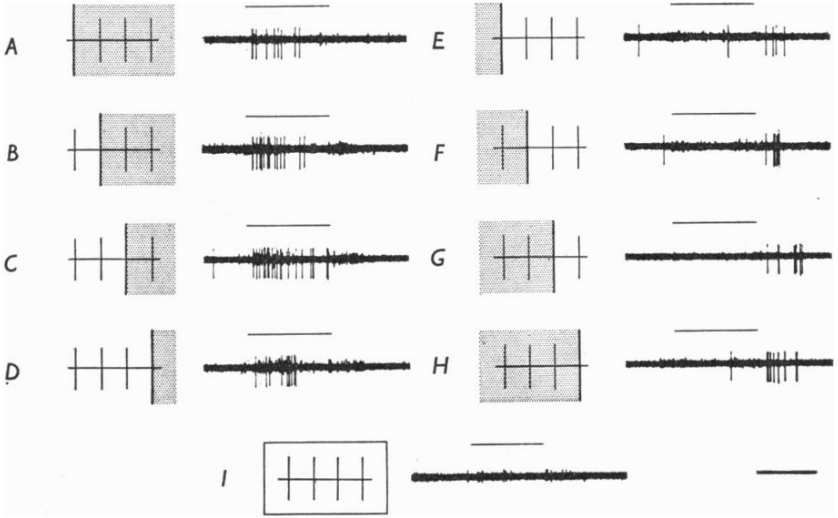
J. Physiol. (1962), **160**, *pp.* 106–154
With 2 plates and 20 text-figures
Printed in Great Britain

RECEPTIVE FIELDS, BINOCULAR INTERACTION AND FUNCTIONAL ARCHITECTURE IN THE CAT'S VISUAL CORTEX

BY D. H. HUBEL AND T. N. WIESEL

From the Neurophysiology Laboratory, Department of Pharmacology Harvard Medical School, Boston, Massachusetts, U.S.A.

(*Received* 31 *July* 1961)



Text-fig. 6. Same cell as in Text-fig. 5. *A–H*, responses to a vertical edge in various parts of the receptive field: *A–D*, brighter light to the left; *E–H*, brighter light to the right; *I*, large rectangle, 10 × 20°, covering entire receptive field. Time, 1 sec.

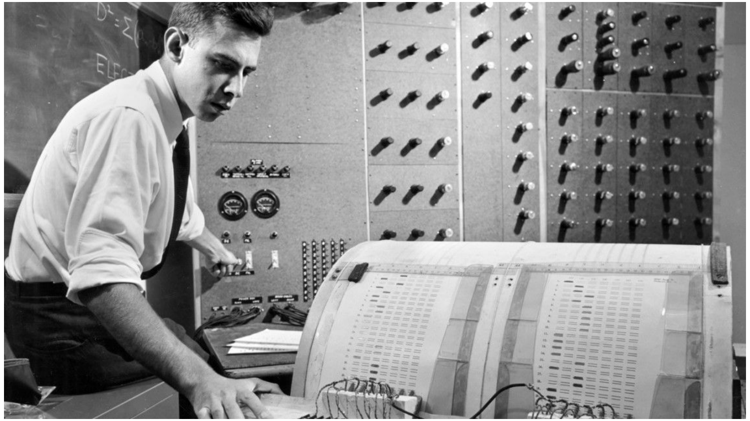
The beginnings of computer vision = beginnings of AI

Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN[1]

F. ROSENBLATT

Cornell Aeronautical Laboratory



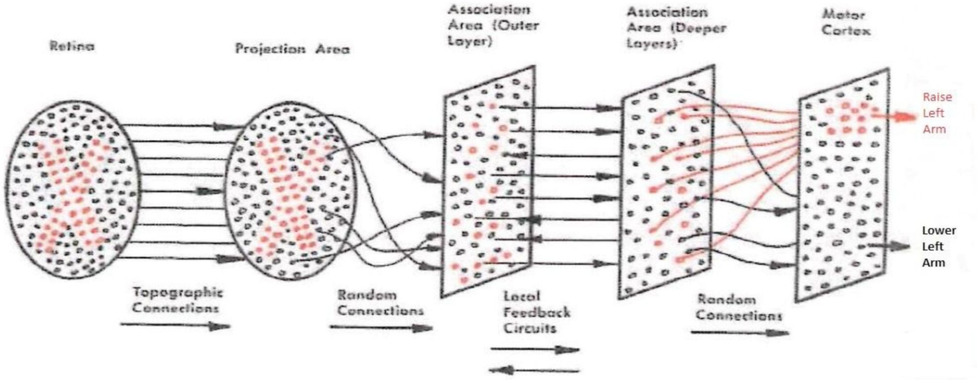


FIG. 1 — Organization of a biological brain. (Red areas indicate active cells, responding to the letter X.)

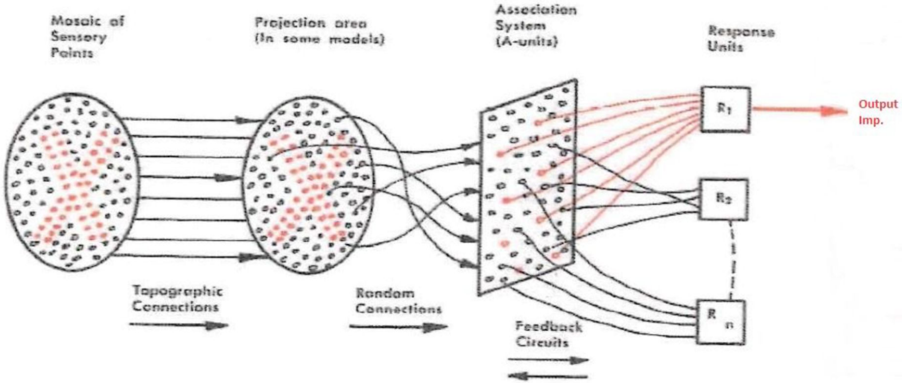
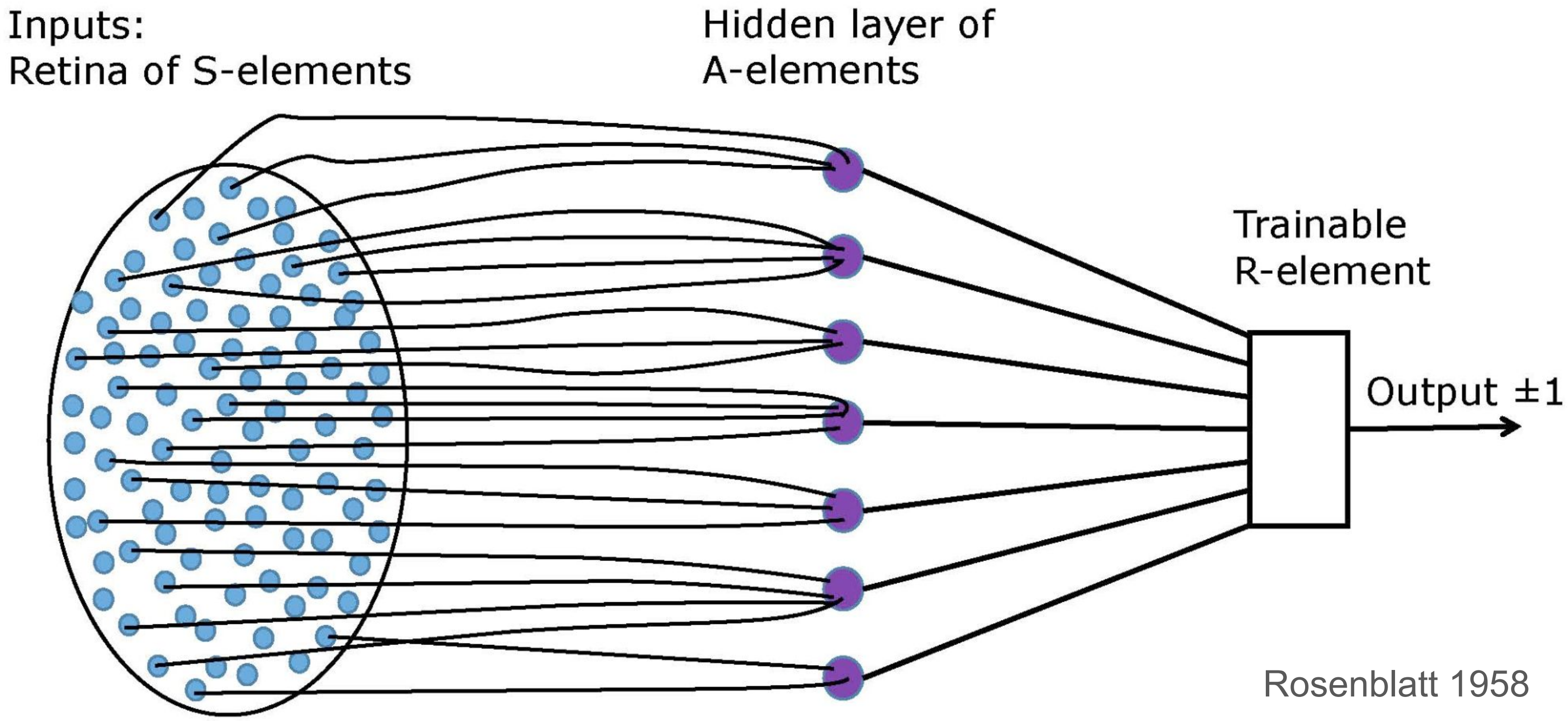


FIG. 2 — Organization of a perceptron.

Rosenblatt’s perceptron



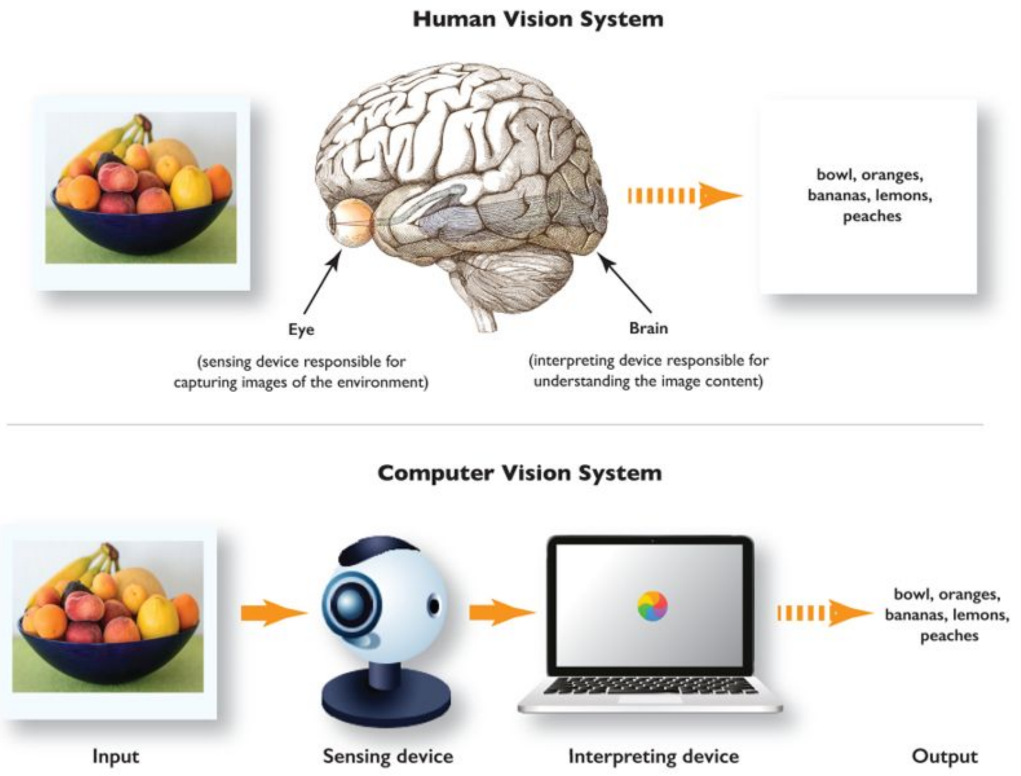
Rosenblatt 1958

The birth of computer vision

Computer vision is the discipline of automating tasks that the human visual system does.

Computer vision originates from attempts to understand/model the visual system that lead to building autonomous systems that can mimic / outperform biological visual systems.

(What do we mean by that?)

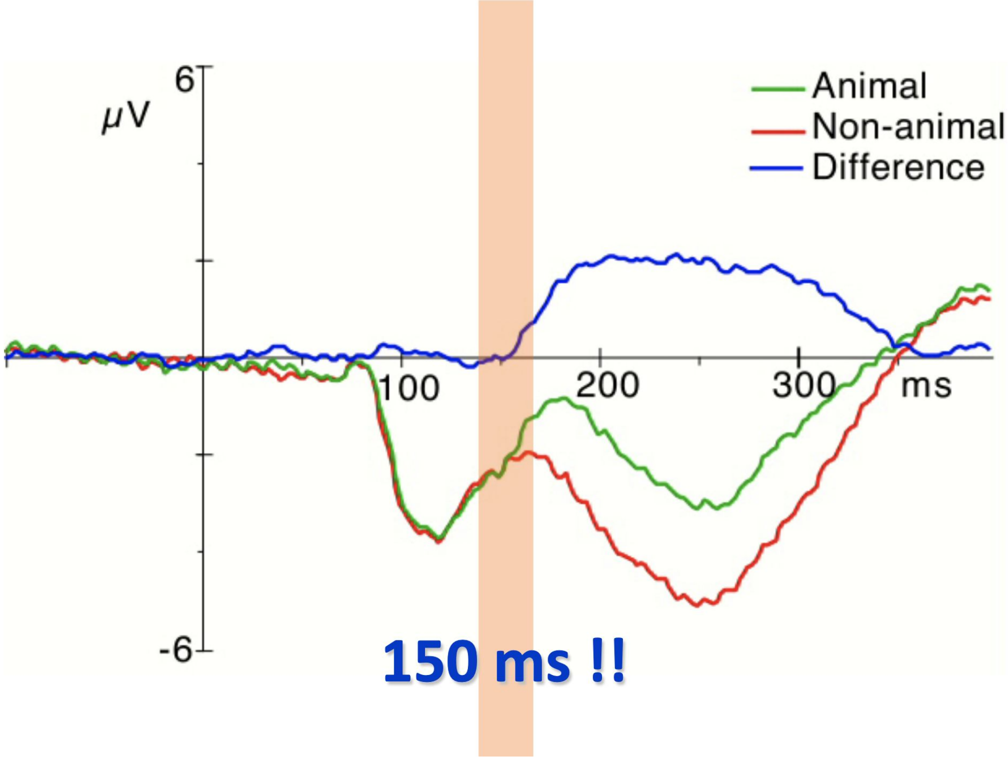


Humans are insanely good at object recognition…

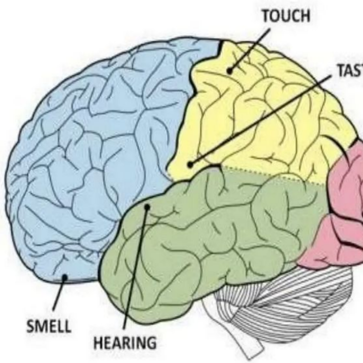
Let’s do a test - raise your hand as soon as you can “see” a living object in the next set of images.

(experiment taken from Carl Vondrick’s lecture notes)

150 ms to recognize animals - our brains are evolved to see!



- Percentage of neurons in brain devoted to each sense
  - Sight – 30%
  - Touch – 8%
  - Hearing – 2%
  - Smell - < 1%
- Over 60% of brain involved with vision in some way

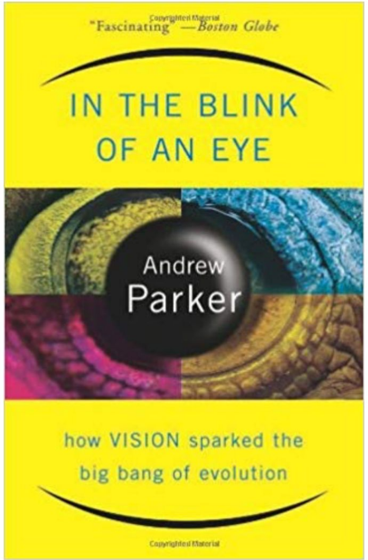


Speed of processing in the human visual system

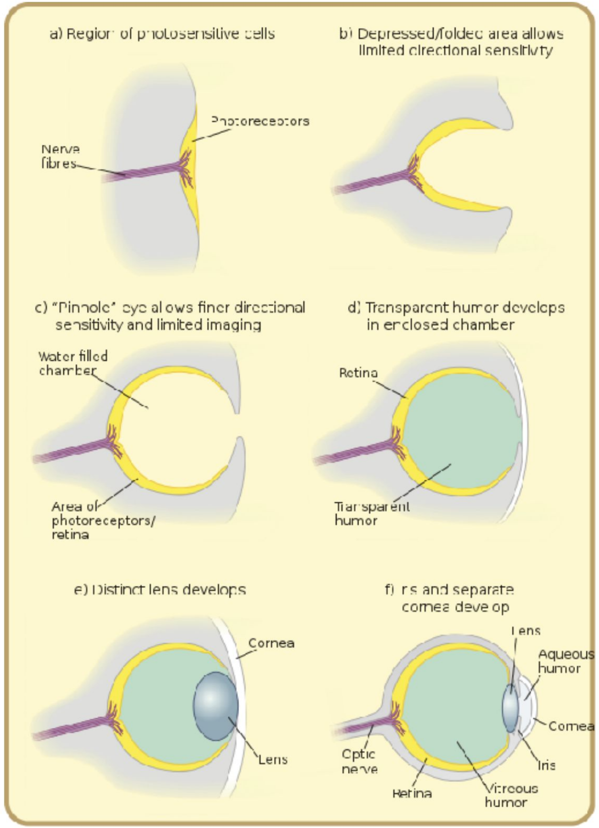
Simon Thorpe, Denis Fize & Catherine Marlot

Nature **381**, 520–522 (1996) | Cite this article

Evolution of the eye



"The Cambrian Explosion is triggered by the sudden evolution of vision," which set off an evolutionary arms race where animals either evolved or died.
— Andrew Parker



If the human brain is so good at vision, why don’t we copy it?

1. Too many neurons (10^11). Even more connections (10^14).
    a. We don’t even know how cells are wired to one another.

2. It’s not the best at **everything…**

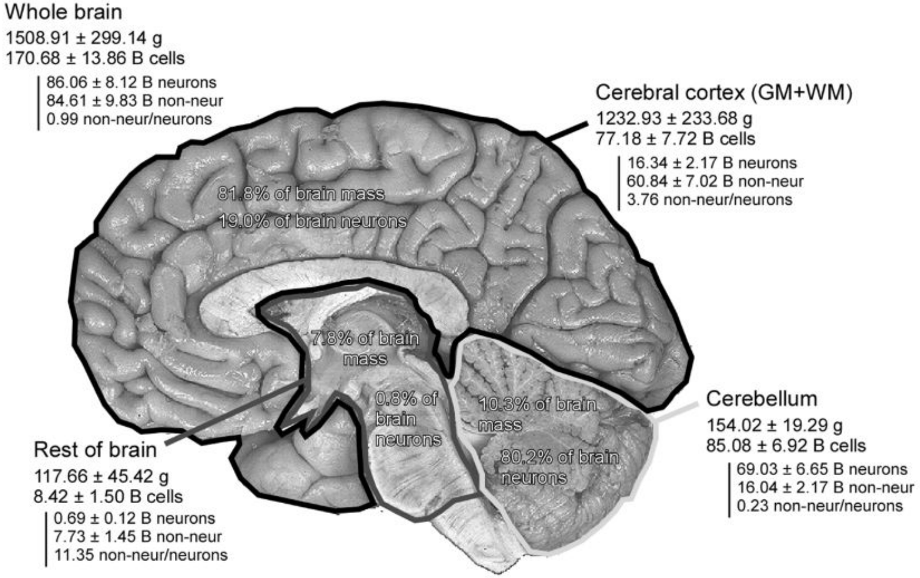
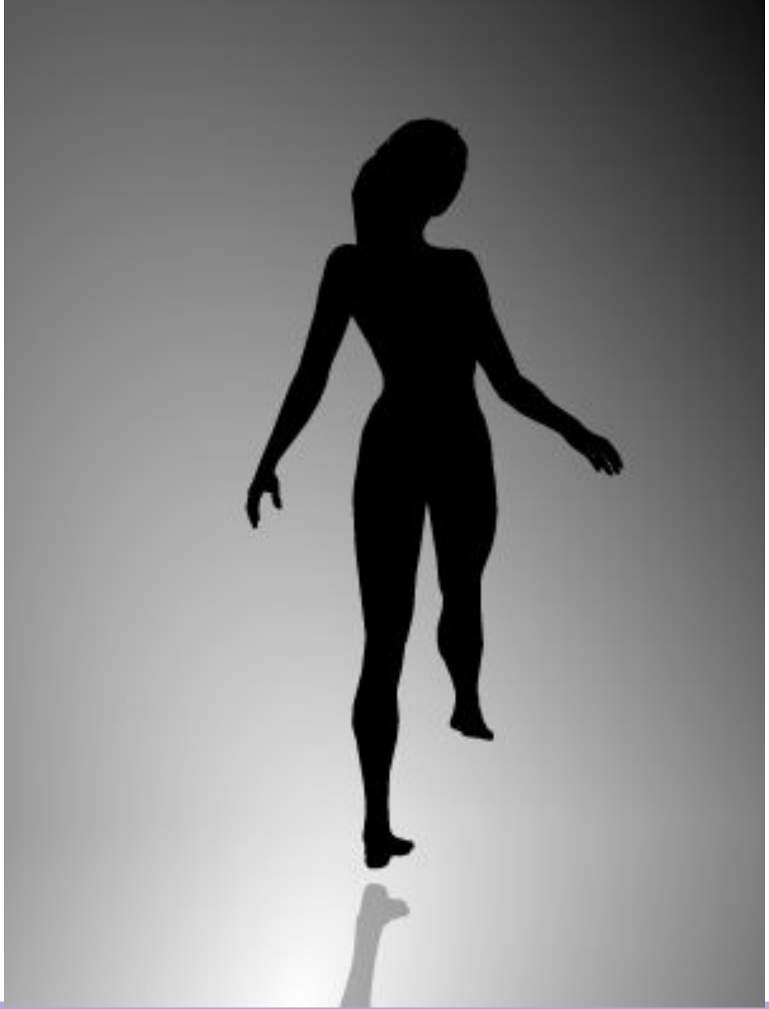


Figure 2.
Absolute mass, numbers of neurons, and numbers of nonneuronal cells in the entire adult human brain. Values are mean ± SD and refer to the two hemispheres together. B, billion.

Let’s do another test -

Raise your right hand if you think the next animation is rotating to the right.



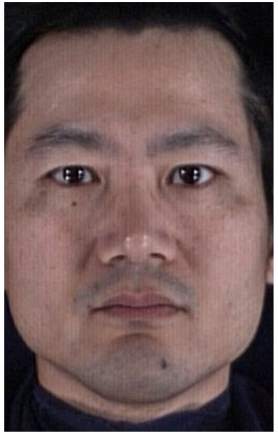
https://en.wikipedia.org/wiki/Spinning_dancer

Our brains still have difficulty in many vision tasks

Because vision is a generally “ill-posed” problem - ambiguities exist.

Other ill-posedness and confounds

illumination



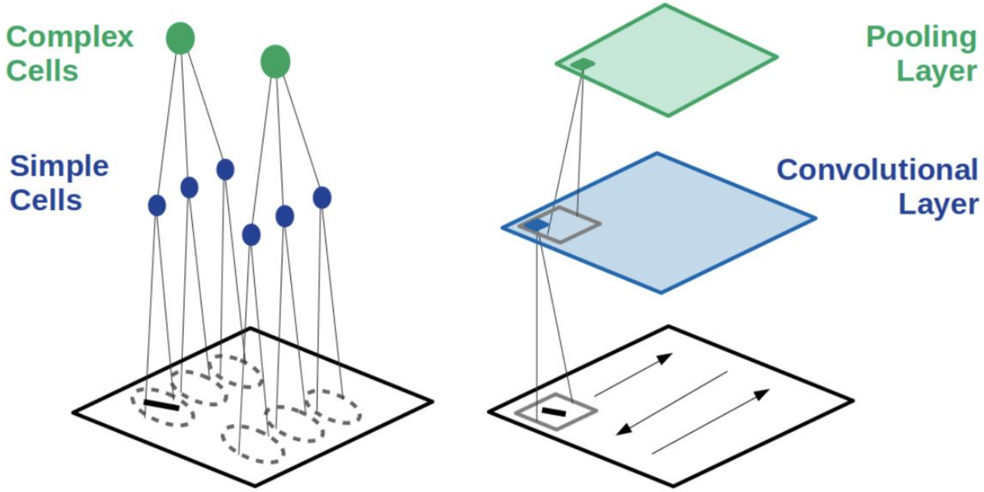
occlusion

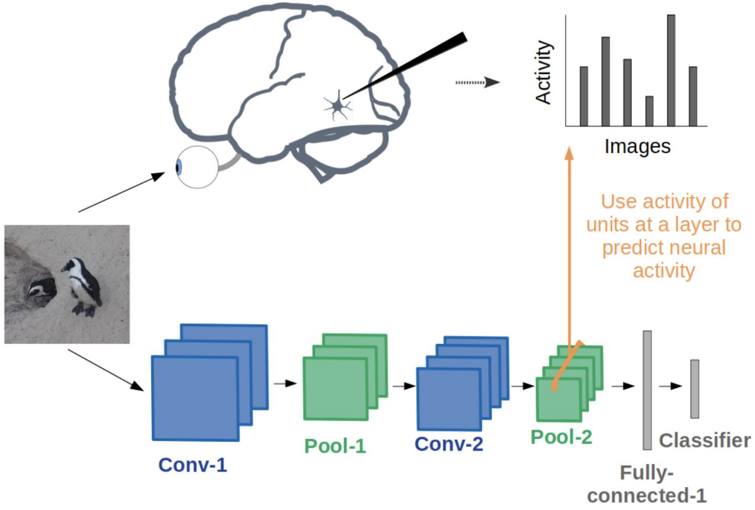
motion

Bio-mimicry* + lots of data is the reason behind the initial success in computer vision

*initially

Convolutional neural nets are inspired by layers of the visual system

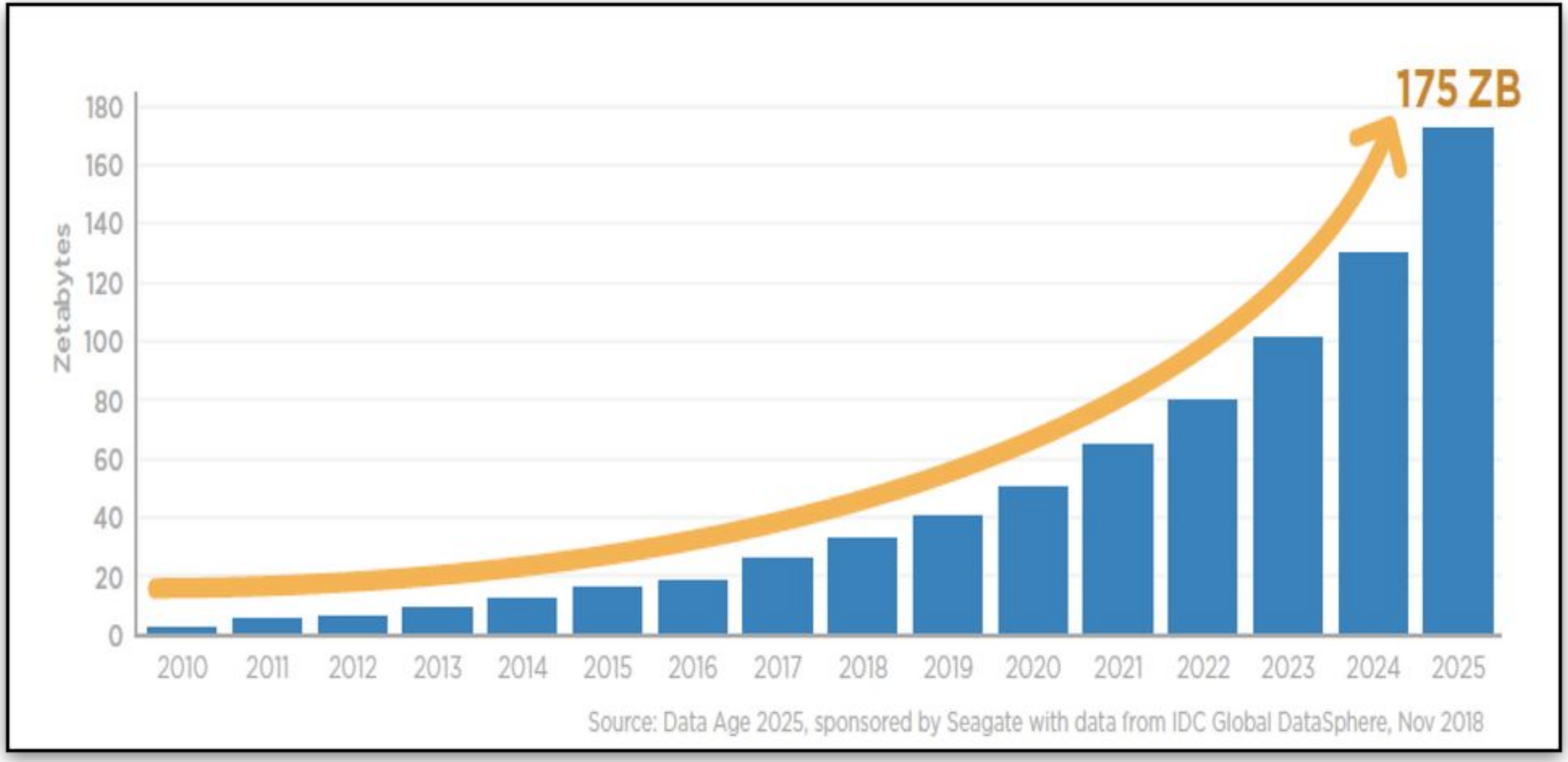




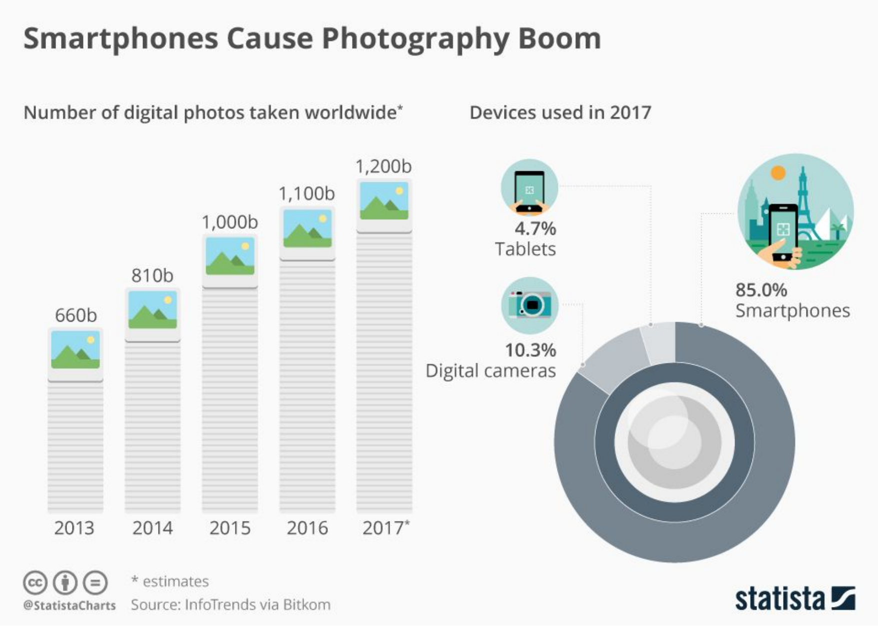
Lindsay 2020

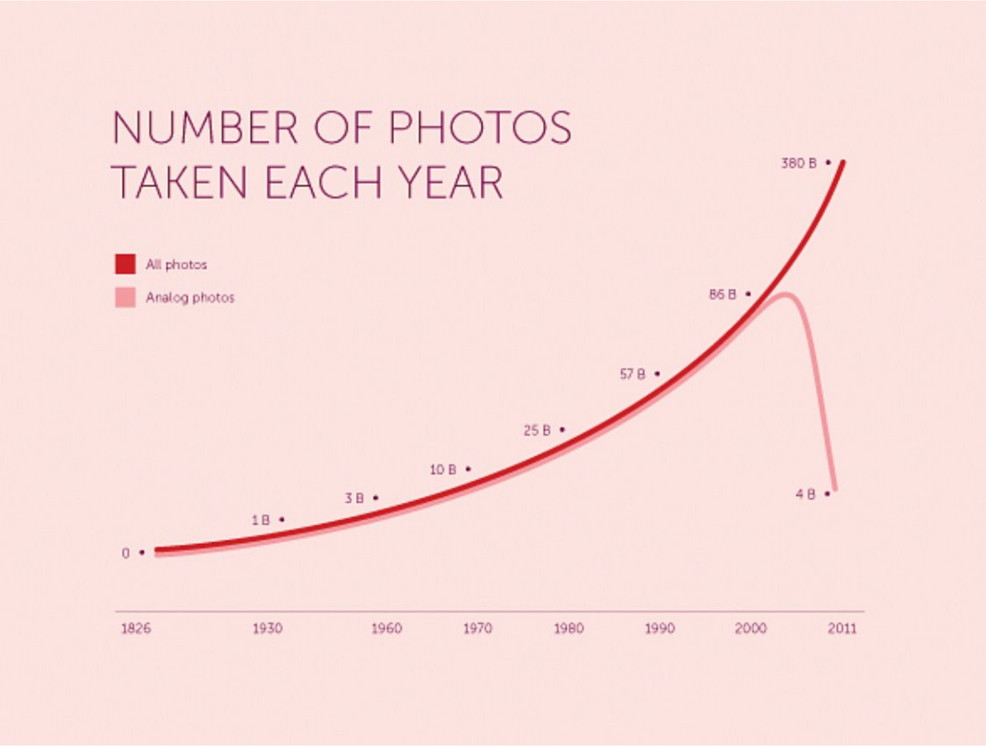
Data explosion has fueled the sustained breakthroughs

A golden age of data science

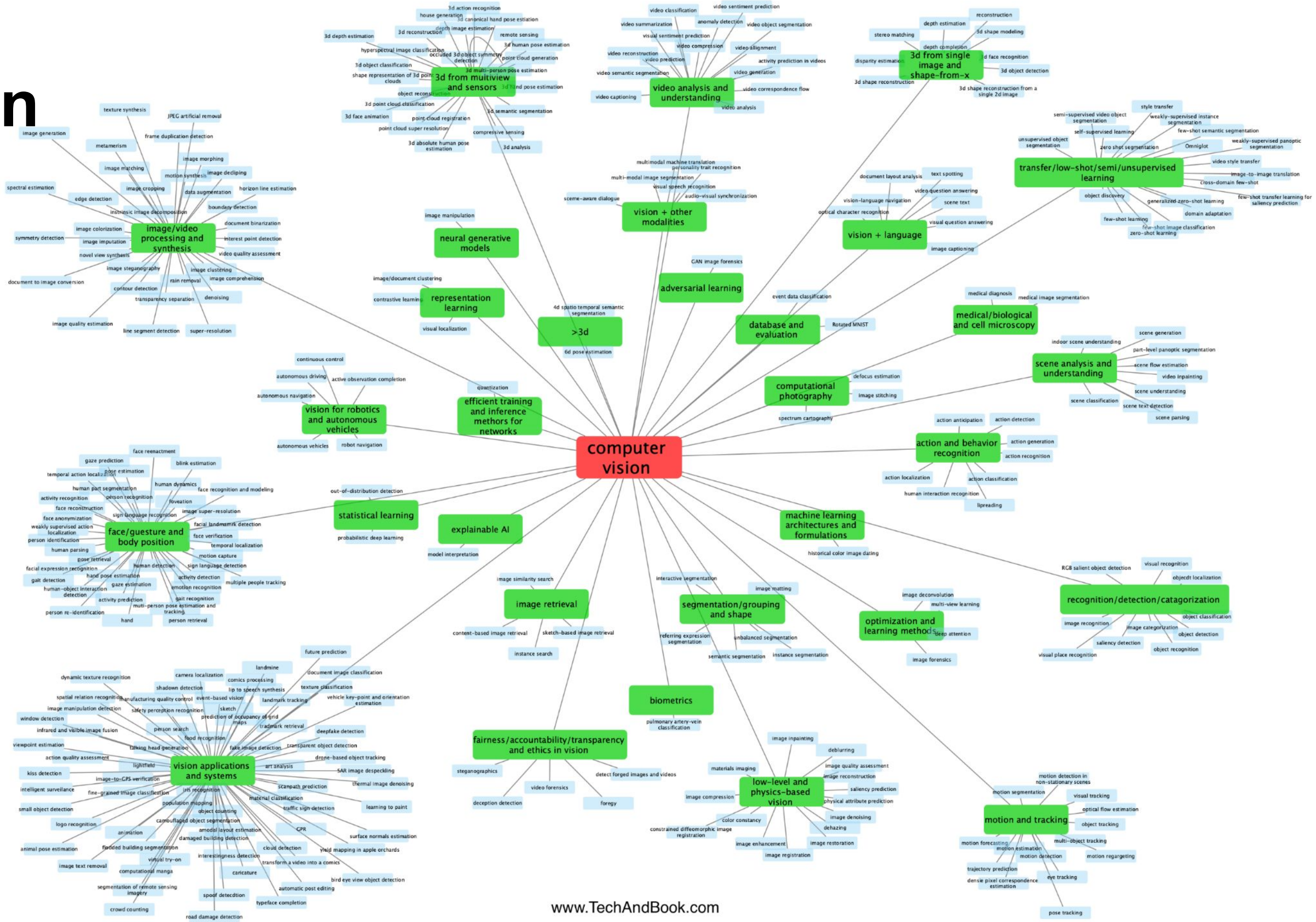


Images are by far the largest contributors of data explosion

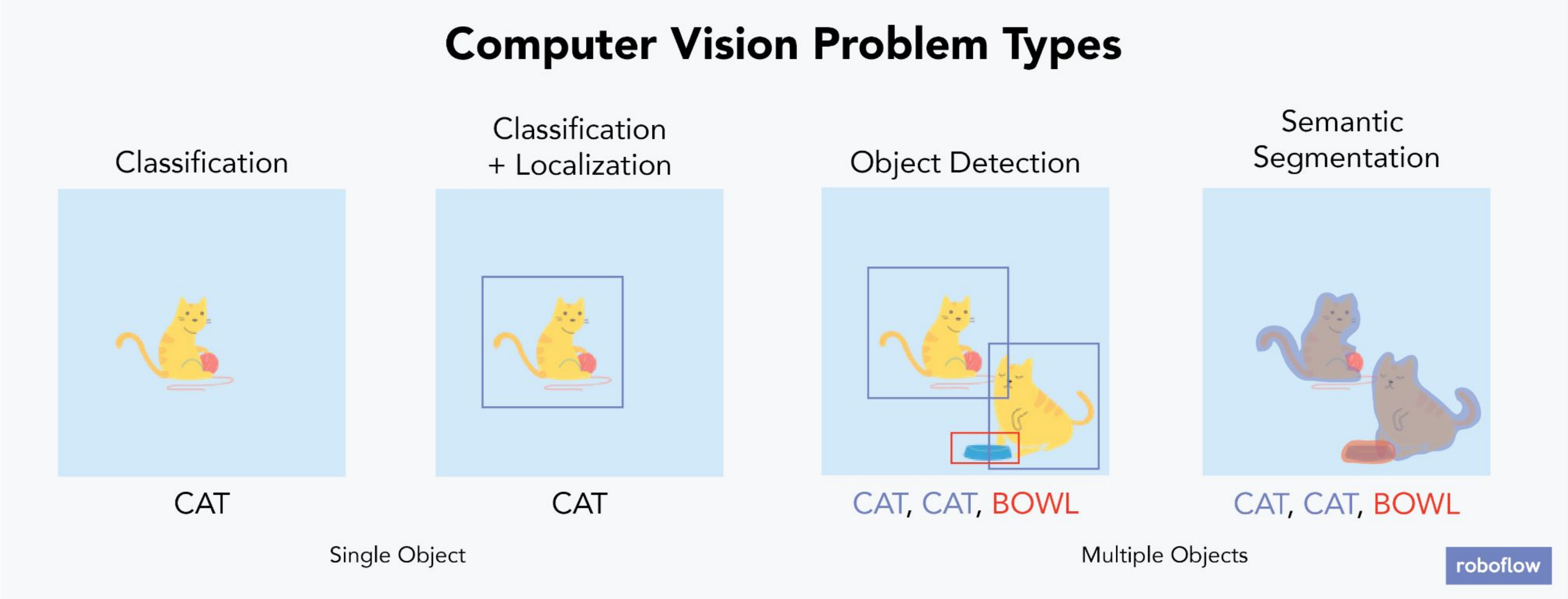




Taxonomy of computer vision circa 2021



Broadly speaking, we can categorize CV areas based on complexity of the task



These next slides should serve as seeds for what you want to do your final project on.

Biometrics

1984

- "the most recognized photograph” in the history of the *National Geographic* magazine
- No one knew her identity…

Biometrics

1984 2002

Augmented reality

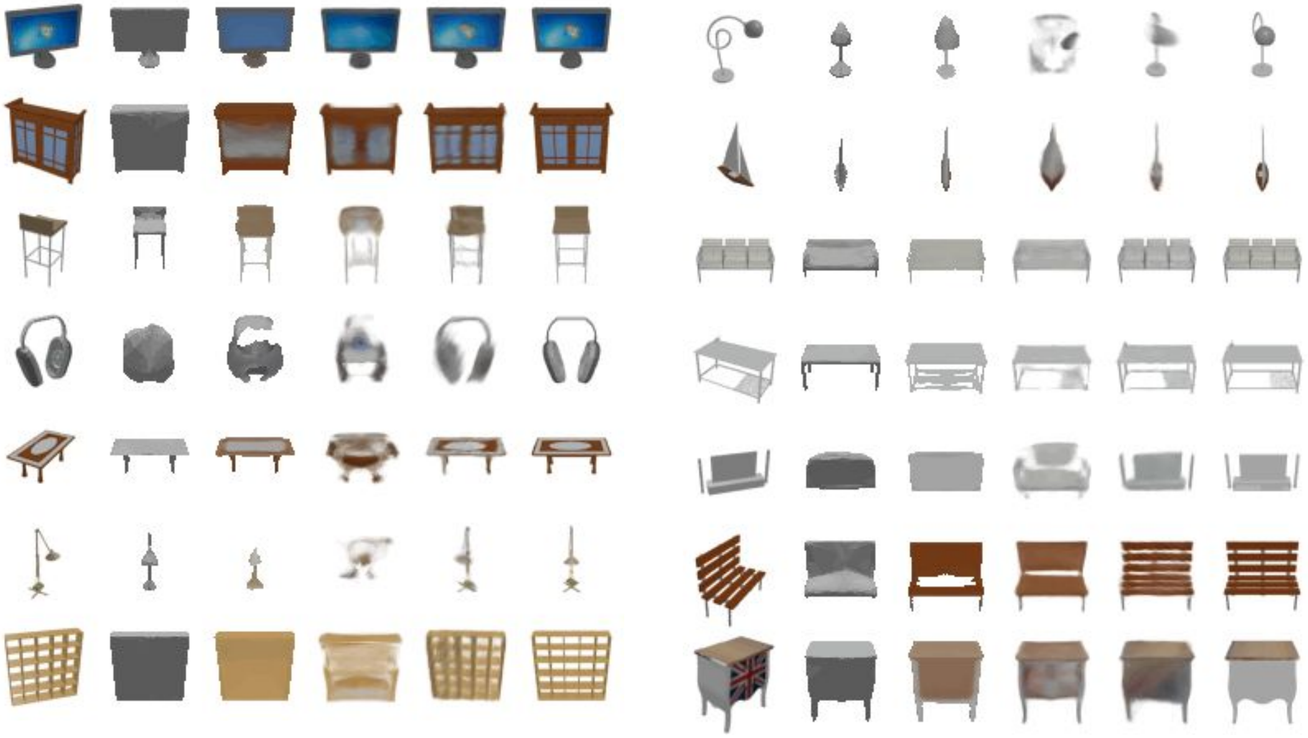
Motion capture and special effects

Avatar

Computational photography

Adobe

3D object reconstruction



Scene reconstruction

Yi et al. CVPR 2022

Autonomous driving

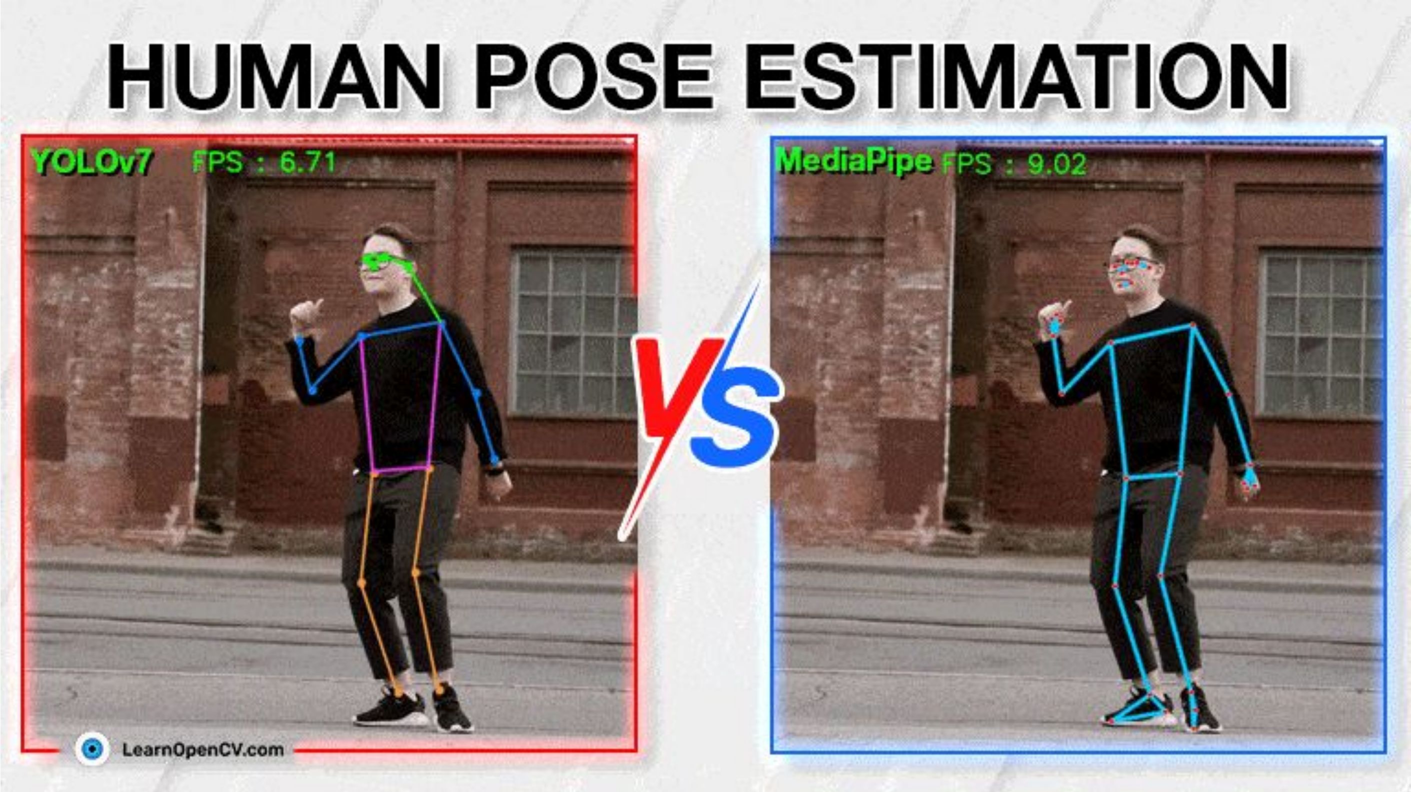
Tesla

Other types of automation - robotics

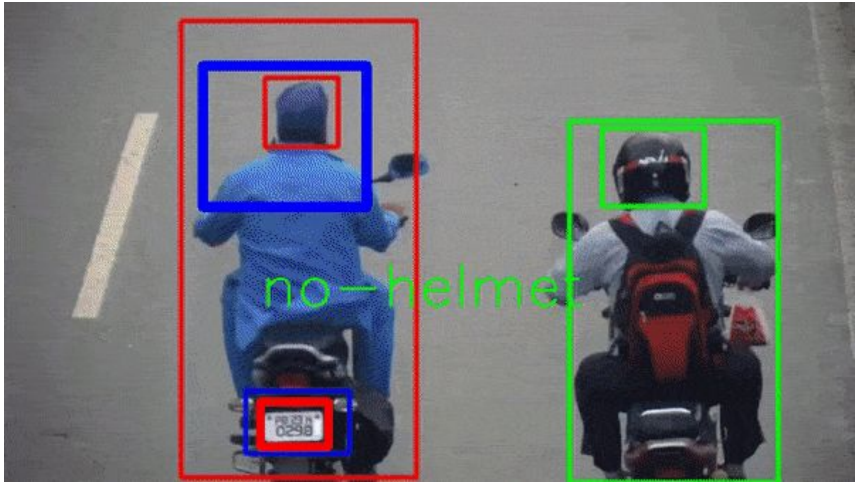
Boston dynamics

Image colorization

Pose estimation and tracking



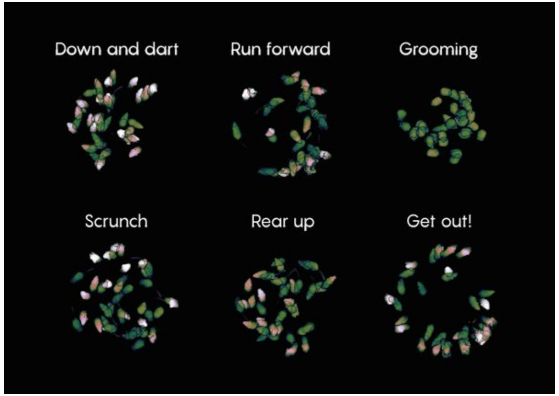
Optical character recognition



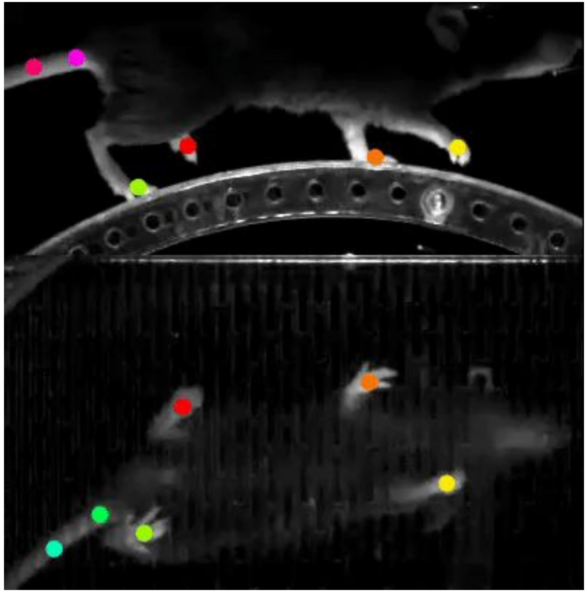
Visual search

TechCrunch

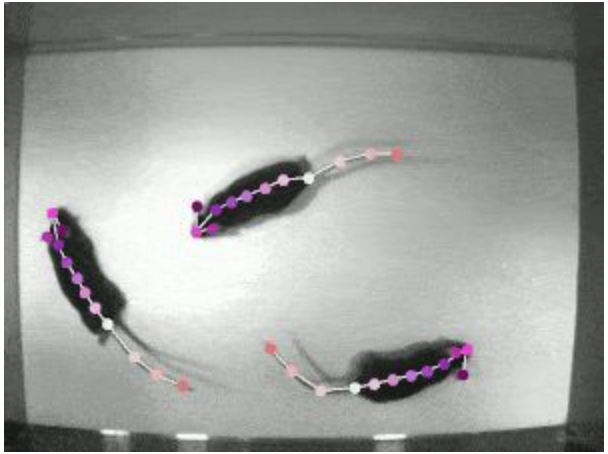
Biospecimen tracking

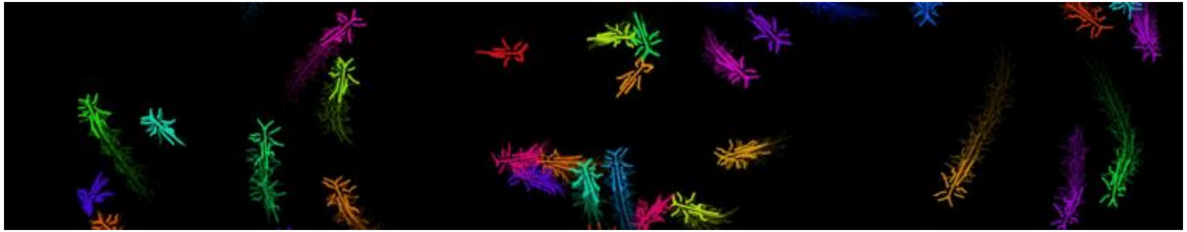


Datta Lab



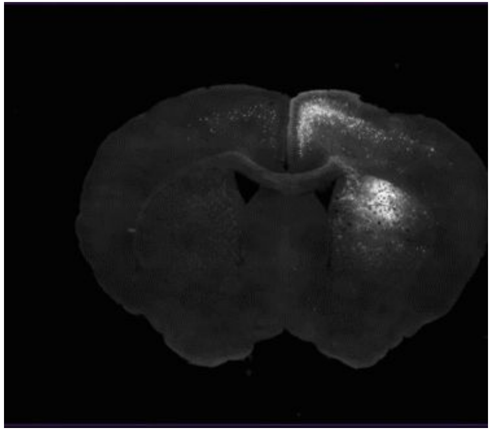
DeepLabCut (Mathis et al 2018)



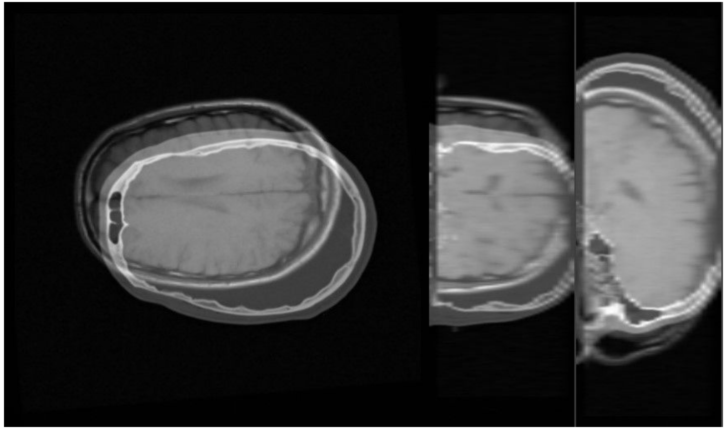


DeepPoseKit (Graving et al 2019)

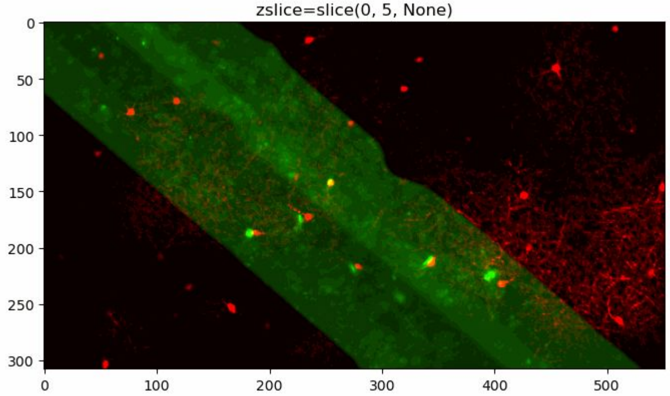
Image alignment



Allen Mouse Brain Atlas

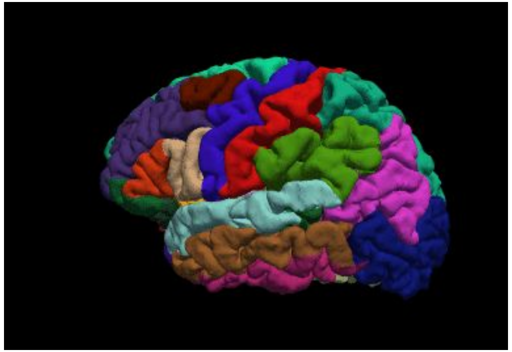


SimpleITK

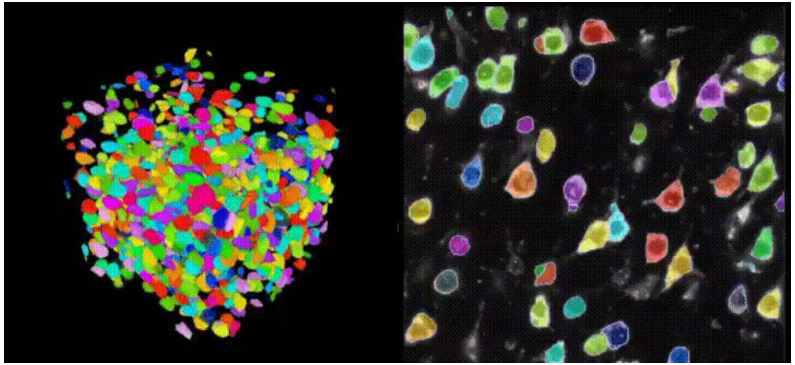


Chen et al. 2023

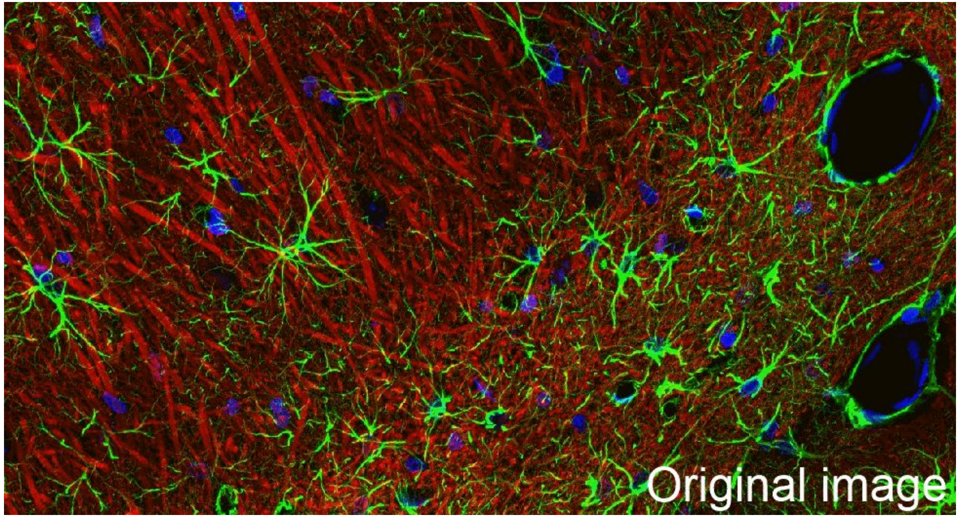
Segmentation



Freesurfer (Fischl et al 2012)



Cellpose (Stringer et al. 2021)



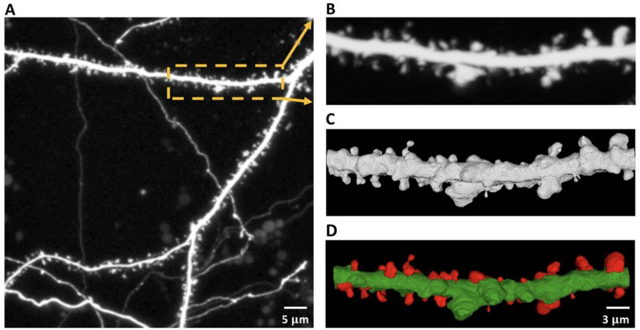
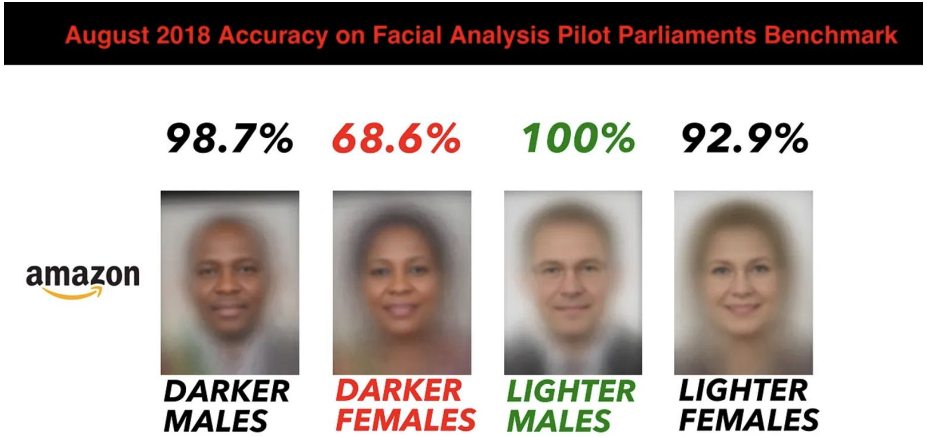


Figure 1. Confocal light microscopy image of hippocampal dendrite covered with dendritic spines.

(Basu et al. 2018)

Ethical issues - mass surveilance, deepfakes, racial bias and many others



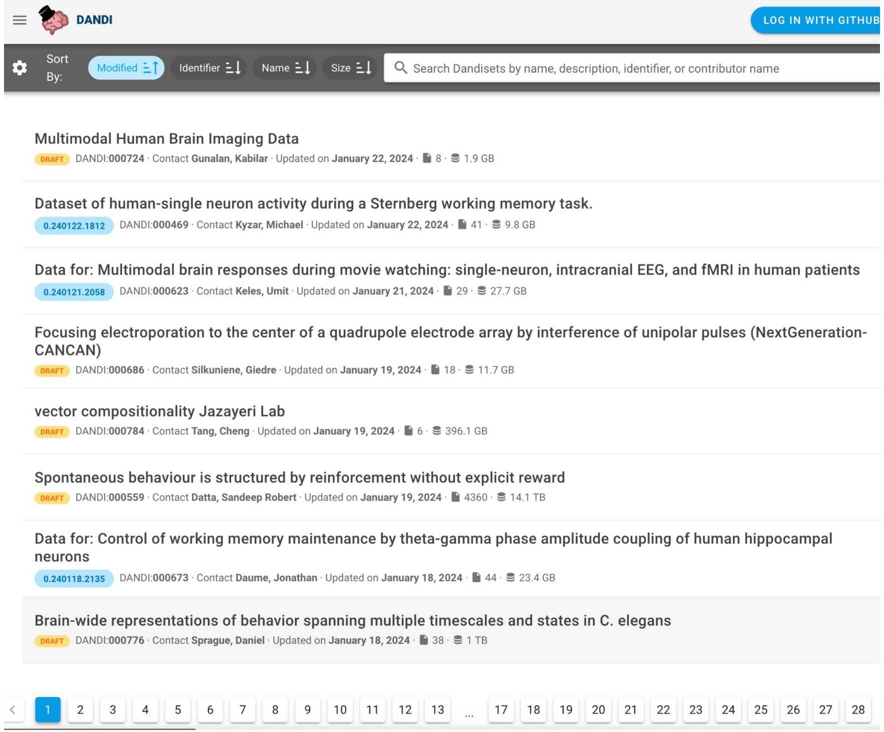
All of these approaches need lots of data to test

It's good to start thinking of project areas early.

There is lots of free data!

See spreadsheet of data repos on slack channel.

Visit office hours to discuss.



Public benchmark datasets and state of the art code

Browse State-of-the-Art

11,438 benchmarks **5,079** tasks **139,771** papers with code

Computer Vision

Semantic Segmentation
328 benchmarks
5732 papers with code

Image Classification
469 benchmarks
4093 papers with code

Object Detection
365 benchmarks
4089 papers with code

Browse SoTA > Computer Vision

Computer Vision

4935 benchmarks • 1501 tasks • 3209 datasets • 51950 papers with code

Semantic Segmentation

Semantic Segmentation
328 benchmarks
5732 papers with code

Tumor Segmentation
4 benchmarks
248 papers with code

Panoptic Segmentation
28 benchmarks
230 papers with code

3D Semantic Segmentation
21 benchmarks
185 papers with code

Weakly-Supervised Semantic Segmentation
9 benchmarks
155 papers with code

See all 28 tasks

Representation Learning

Representation Learning
16 benchmarks
4032 papers with code

Disentanglement
3 benchmarks
634 papers with code

Graph Representation Learning
2 benchmarks
405 papers with code

Sentence Embeddings
3 benchmarks
231 papers with code

Network Embedding
156 papers with code

See all 17 tasks

Course information, logistics and grading

Course website:

https://www.neuroinformaticslab.com/computer-vision-cs-gy-6643

Goals of this course

- Understand the computer vision landscape:

  What are the fundamental building blocks of a visual sensing system? (Lectures)

- Understand the theory

  How are computer vision tools applications of fundamental mathematical concepts? (Homeworks)

- Develop tool building skills:

  Can I build a central computer vision tools from first principles? (Projects)

- Gain insights on how to innovate:

  Can I put these tools together in a team setting to solve a real life problem? (Final project)

Course approach

Central philosophy: computer vision is learned by doing

This course will be interdisciplinary and interactive:

- Lectures will cover
  - **Central computer vision areas**
  - **Principles of tool building**
- Later lectures will cover
  - **Case studies on cutting edge tools by guest lecturers**
- Course projects and homeworks will encourage
  - **Building new tools**
  - **Asking new questions**

Prerequisites: Linear algebra, probability and data structures knowledge

What won't be covered

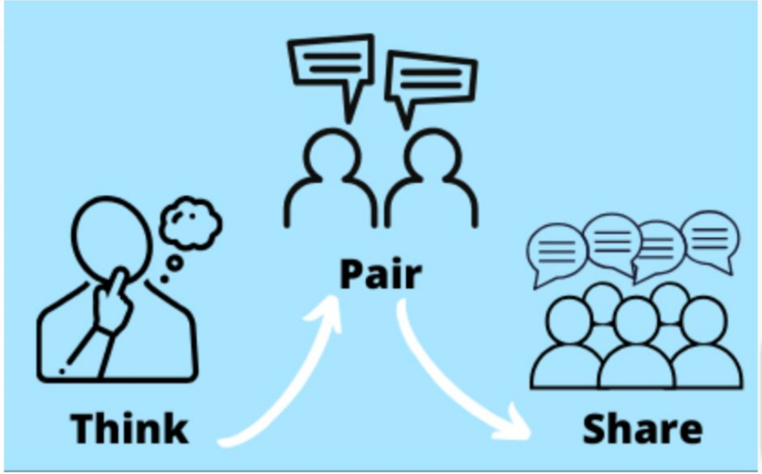
Software engineering

I assume you're familiar with python. If not, please see provided tutorials on course website.

Latest state of the art cutting edge computer vision architectures

This course serves as an entry point for computer vision, building from first principles. For more advanced CV, see applied lecture (with a neuroscience flair) CS-GY 9663

Course format*



1-1:10 lecture

10-20 min break

1 hr interactive coding demo by TAs
(bring your laptop with Colab set up)

*except for today

Primary textbook

Computer Vision: Algorithms and Applications, 2nd ed.

© 2022 [Richard Szeliski](), The University of Washington

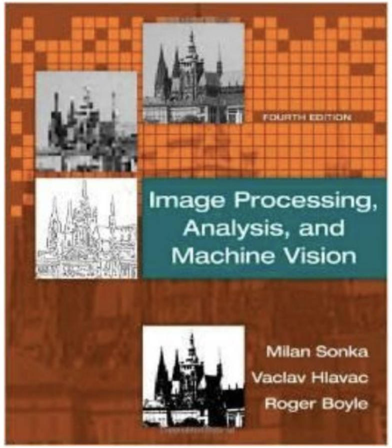
Primary Book: Richard Szeliski, Computer Vision: Algorithms and Applications.

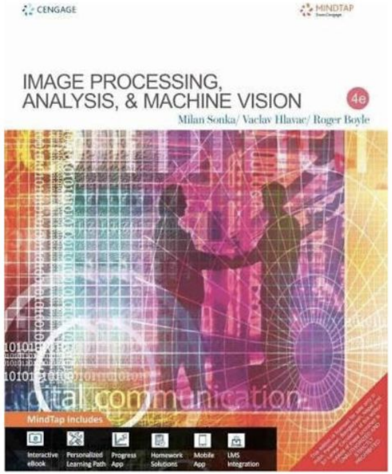
Freely available online: Link http://szeliski.org/Book/

Secondary Textbooks

Some course material from:

Milan Sonka, Vaclav Hlavac and Roger Boyle, *Image Processing, Analysis, and Machine Vision, 4th Edition*, Cengage Learning, Stamford, CT, 2015. (ebook, paperback, hardcover)





Course load

There will be…

3 individual homework assignments, focusing on theory and algorithms.

3 individual projects that require programming. We will primarily support Python.

A midterm exam (in-class October 24), covering the first part of the semester.

A final group project (3-4 students per project). Projects will be presented during our final class on December 12, write-up and code due December 19.

Grading

Homeworks (3) - 18% of grade, to be done individually

Programming projects (3) - 45% of grade, to be done individually.

Midterm exam - 15% of grade, in class.

Final project - 20% of grade, to be done in groups of 3-4.

Final project team and idea selection - 2% of grade, to encourage forming groups early.

Total = 100%

Homeworks (3 x 6% of the grade)

- Homeworks will quiz the ability to execute theoretical concepts in CV in mathematical form.
- No coding will be necessary for these.
- Write ups in Latex will be highly encouraged.

Programming projects (3 x 15% of the grade)

- Programming projects will quiz the ability to implement foundational computer vision tools from first principles, execute on real data and present results in a a short report format.
- A python notebook with executed outputs + a pdf write up of each step explained will be required.
- Write ups in Latex highly encouraged.

Midterm exam (15% of grade)

- Will test the ability to derive key mathematical concepts on the spot.
- It will also test the ability to “think like AI” in several image analysis scenarios.
- The exam will be designed to be completed in 1:15 mins although you are welcome to stay the full 2.5 hour time slot if needed.

Final project (20% of grade + 2% for forming a team on time)

- The project can involve either the implementation and justification of a new computer technique, or a existing technique implemented and applied by you to a novel data set. Students must work in **groups of 3 or 4 (ideally 4)**.
- Project ideas and datasets will be provided in the slack channel.
    - Come to office hours for early advice
- Project team selection is due Sep 19 (2 out of 22 points of grade).
- Project idea proposal is due Oct 3. (3 out of 22 points of grade).
- 10 minute presentation (**December 12 in class**)
- 6 page write-up + code notebook (Due **December 19, 11:59PM** on Brightspace)
    - Overleaf LaTeX template will be provided
- Evaluation will be on the scientific rigor, justification of methods, clarity of written and oral presentations and ability to field questions.

Important: We don't want you to just download a CV package off github and run it. The goal of this class is to test your ability to connect foundational knowledge towards building a running system.

Late policy

Every hour that a project is late (rounded down) will cause 1% penalization of the total allotted grade.

For example, a project or homework that is 11 hours 45 minutes late will have a maximum possible score of 89%.

Tentative schedule and topics

Date	Topic	Material	Homework/Projects
September 5, 2024	Intro and survey of topics	Szeliski 2.1, 2.2, 2.3, Python tutorial, Google Colab Setup	Homework 0 out: Join slack nyucomputervision.slack.com (Due September 11, 11:59pm)
September 12, 2024	Image formation and filtering	Szeliski 2.1, 2.2, 2.3, 3.1, 3.2, 3.3, 3.6	Homework 1 out (due September 18, 11:59pm), Project 1 out (due October 2 11:59pm)
September 19, 2024	Image recognition	Szeliski 6.1,6.2,6.3,6.4	Final project team selection due
September 26, 2024	Feature detection and matching	Szeliski 7.1,7.2	
October 3, 2024	Contour tracking and Hough transforms	Szeliski 7.3,7.4	Final project proposal due
October 10, 2024	Segmentation	Szeliski 7.5	
October 17, 2024	Image alignment	Szeliski 8.1,8.2	
October 24, 2024	**Midterm exam (in class)**		
October 31, 2024	TBD		
November 7, 2024	TBD		
November 14, 2024	TBD		
November 21, 2024	**THANKSGIVING BREAK (No class)**		
November 28, 2024	TBD		
December 5, 2024	TBD		
December 12, 2024	**Final project presentations**		Write-up and code due December 19, 11:59pm

FAQ

- Can I be enrolled in the class?
  - We are at full or near full capacity.
  - If we are below capacity of the classroom and you’re a grad student - yes.
  - If you are an undergrad, share me your transcript and we can make a case by case decision.
- Where is the syllabus?
  - On the course website, slack and brightspace.
- Are we going to cover deep learning methods?
  - Not in this class. For advanced methods and applications - see CS-GY 9223 (that I teach) or other CV classes at NYU.
- Can I use C, Java or any other programming language?
  - No, the projects and problem sets will use Python.
- Best way to contact instructors or TAs
  - Slack >> Brightspace >> Email

Last but not least

This is a new course (for me, I’m new here).

Any feedback is appreciated.

Let us know what works and what doesn’t!

Next class: Image formation and filtering