Deep Learning

Lecture 1 - Introduction

Who am I?

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 insights

Past courses:
- UW CSE 599H: Artificial Intelligence vs Intelligence Augmentation
- Stanford CS 231N: Convolutional neural networks for computer vision
- Stanford CS 131: Computer vision fundamentals and applications

Are you in the right place?

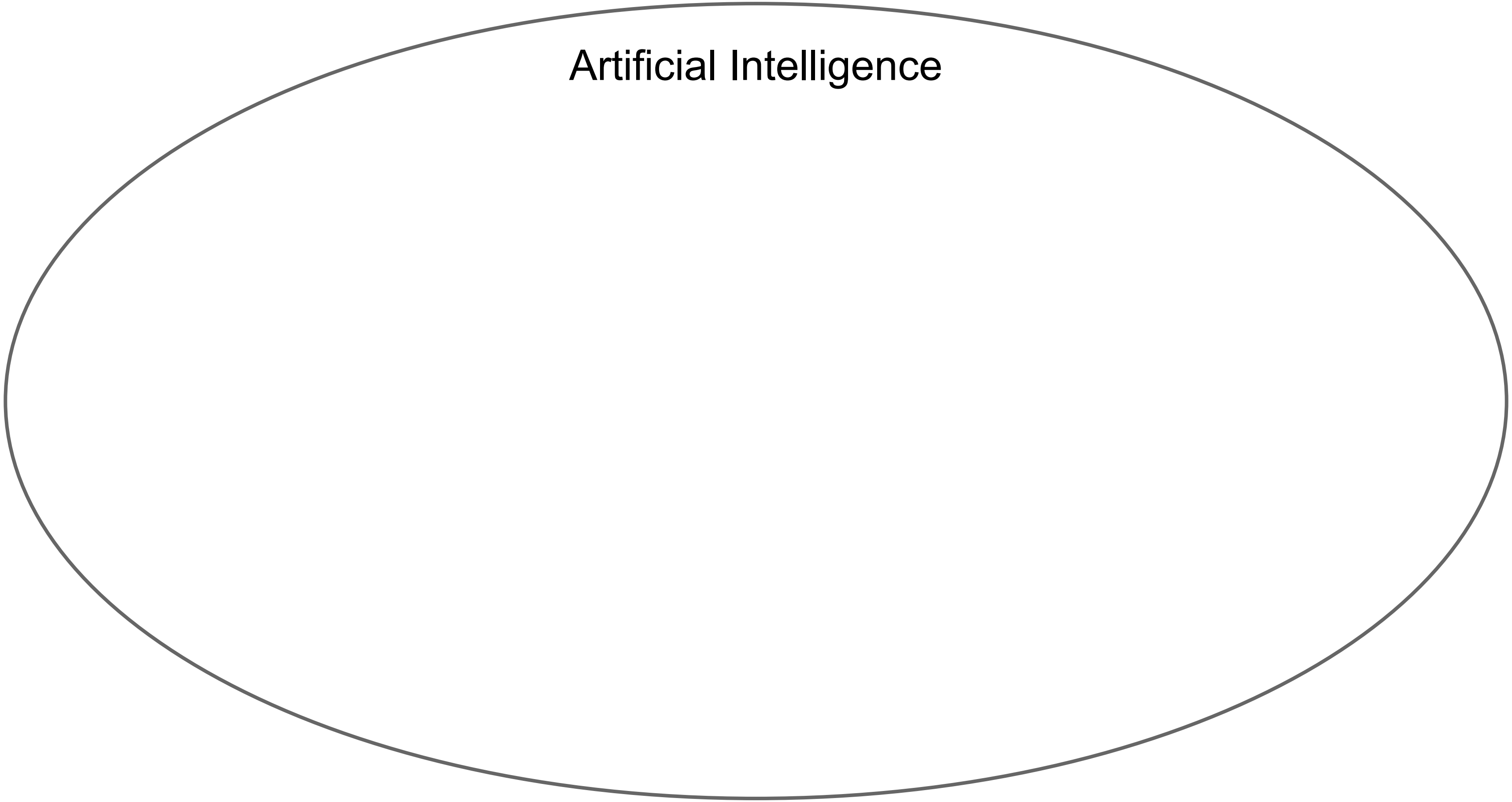
Location: CSE2 G01
Lectures: Tuesdays and Thursdays @ 10-11:20am
Recitations: Fridays @ TBD
Canvas: https://canvas.uw.edu/courses/1653282
Gradescope: https://www.gradescope.com/courses/522621
Website: https://courses.cs.washington.edu/courses/cse493g1/23sp/
EdStem: https://edstem.org/us/courses/38318/

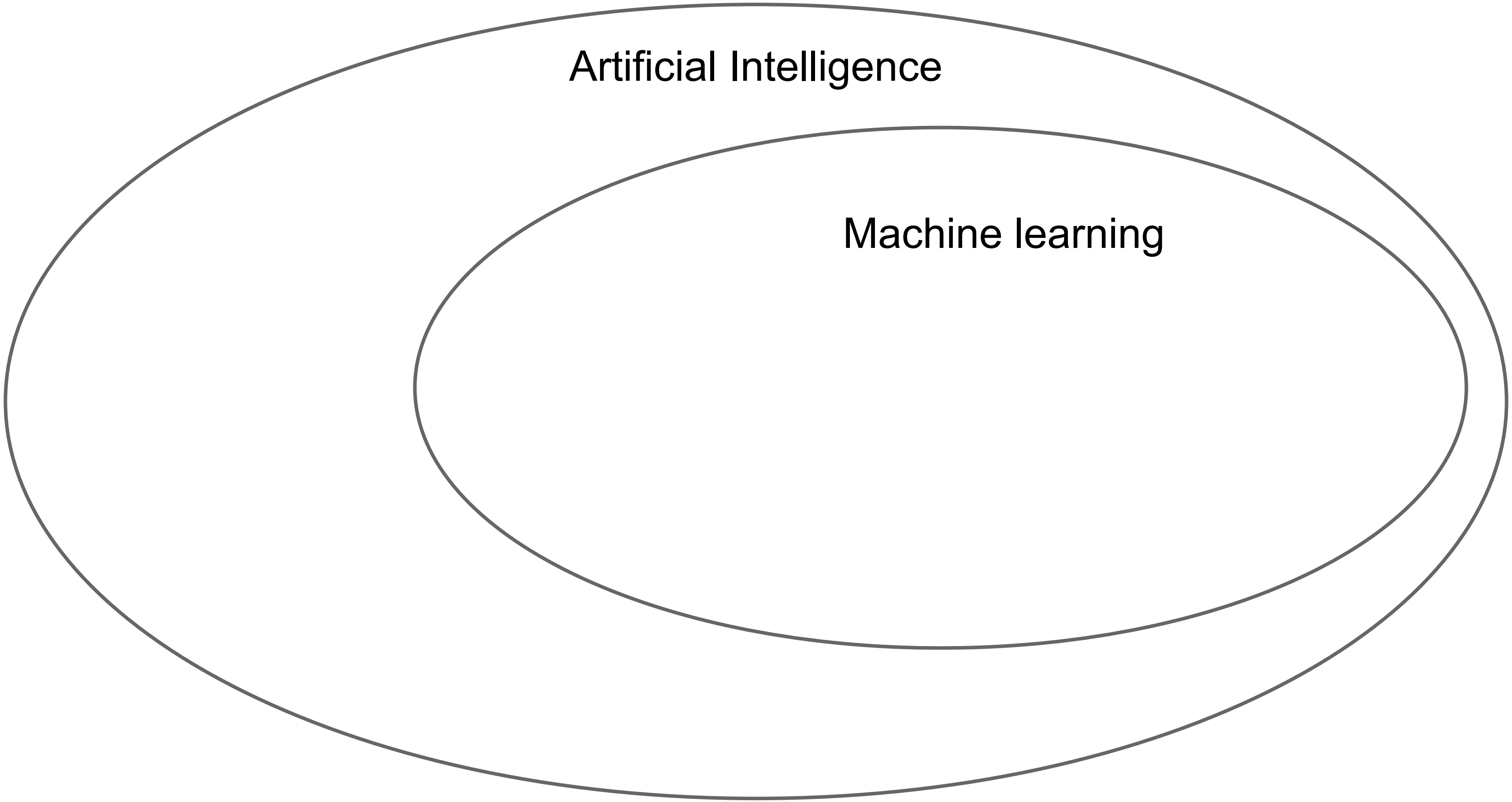
What is ~~Deep~~ Learning?

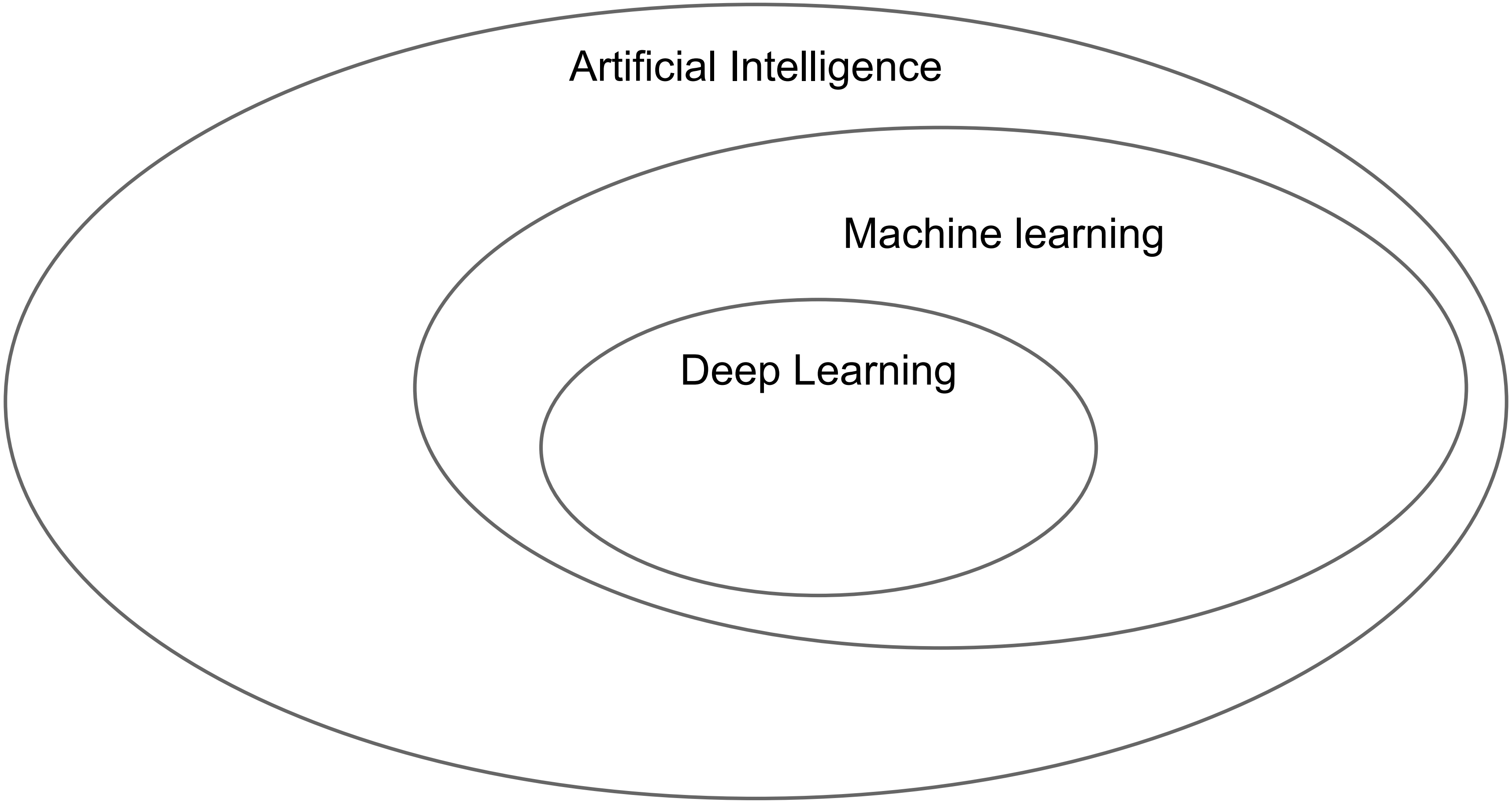
Building artificial systems that learn from data and experience

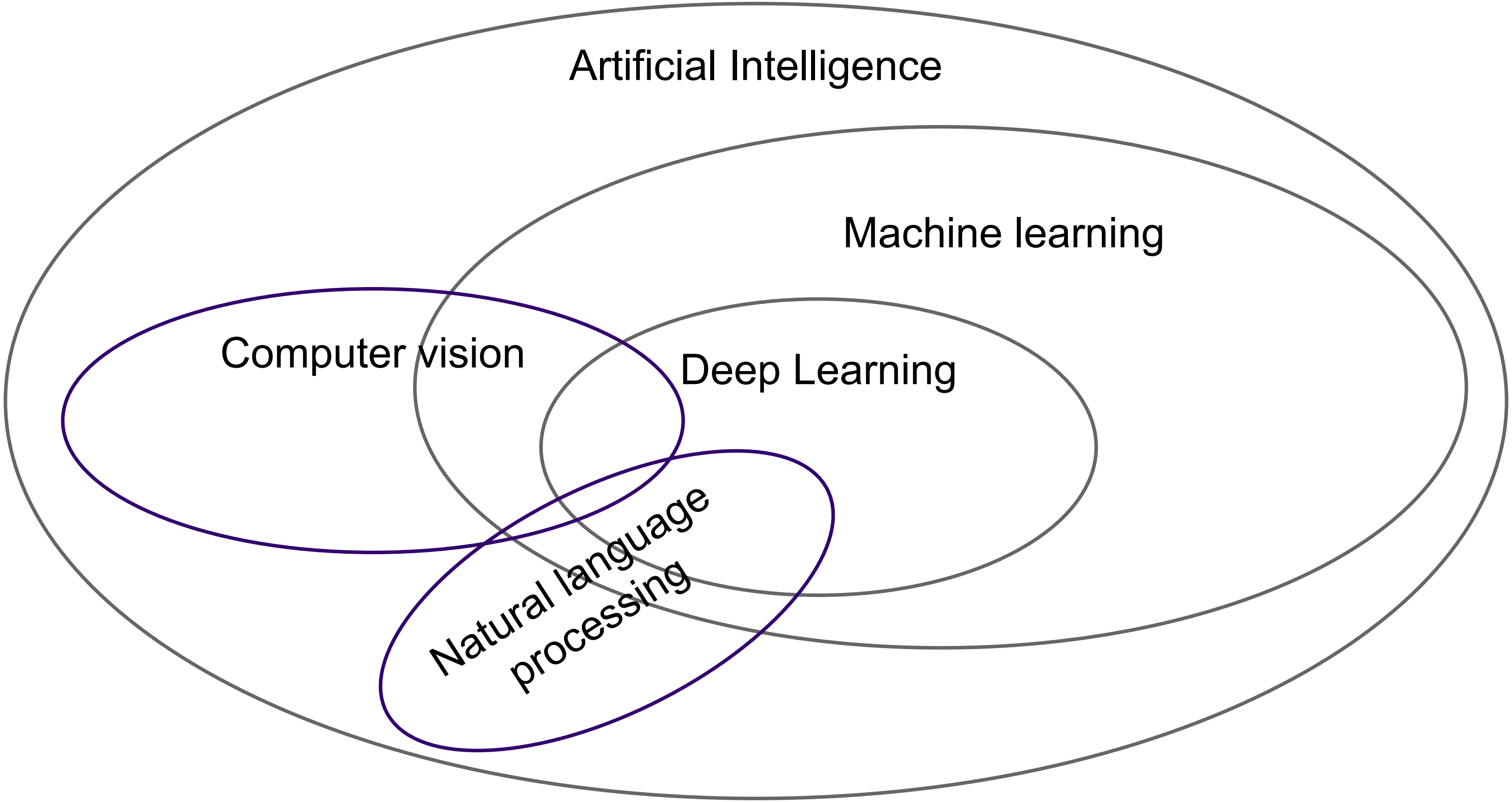
What is Deep Learning?

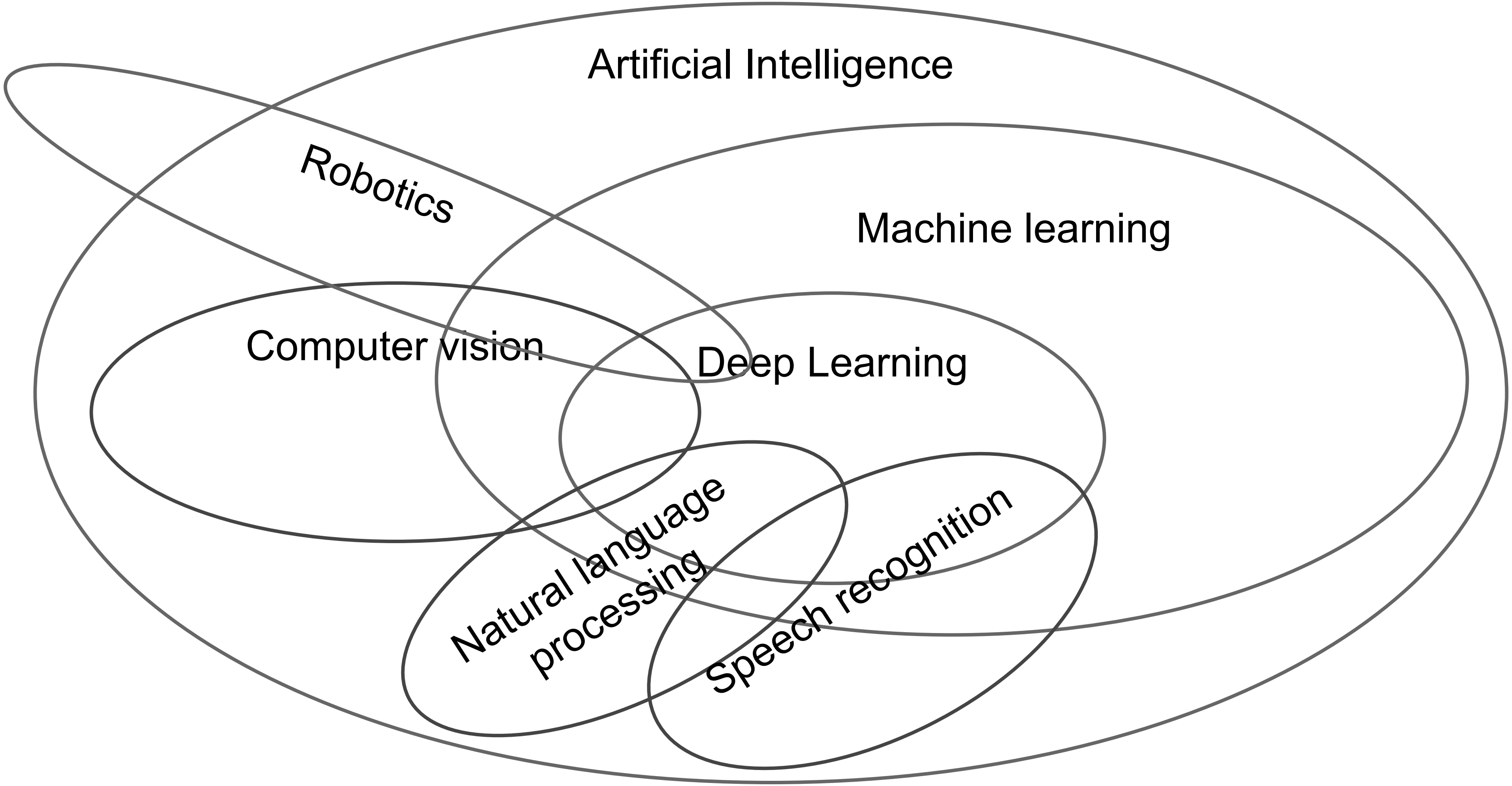
Hierarchical learning algorithms with many “layers”, (very) loosely inspired by the brain

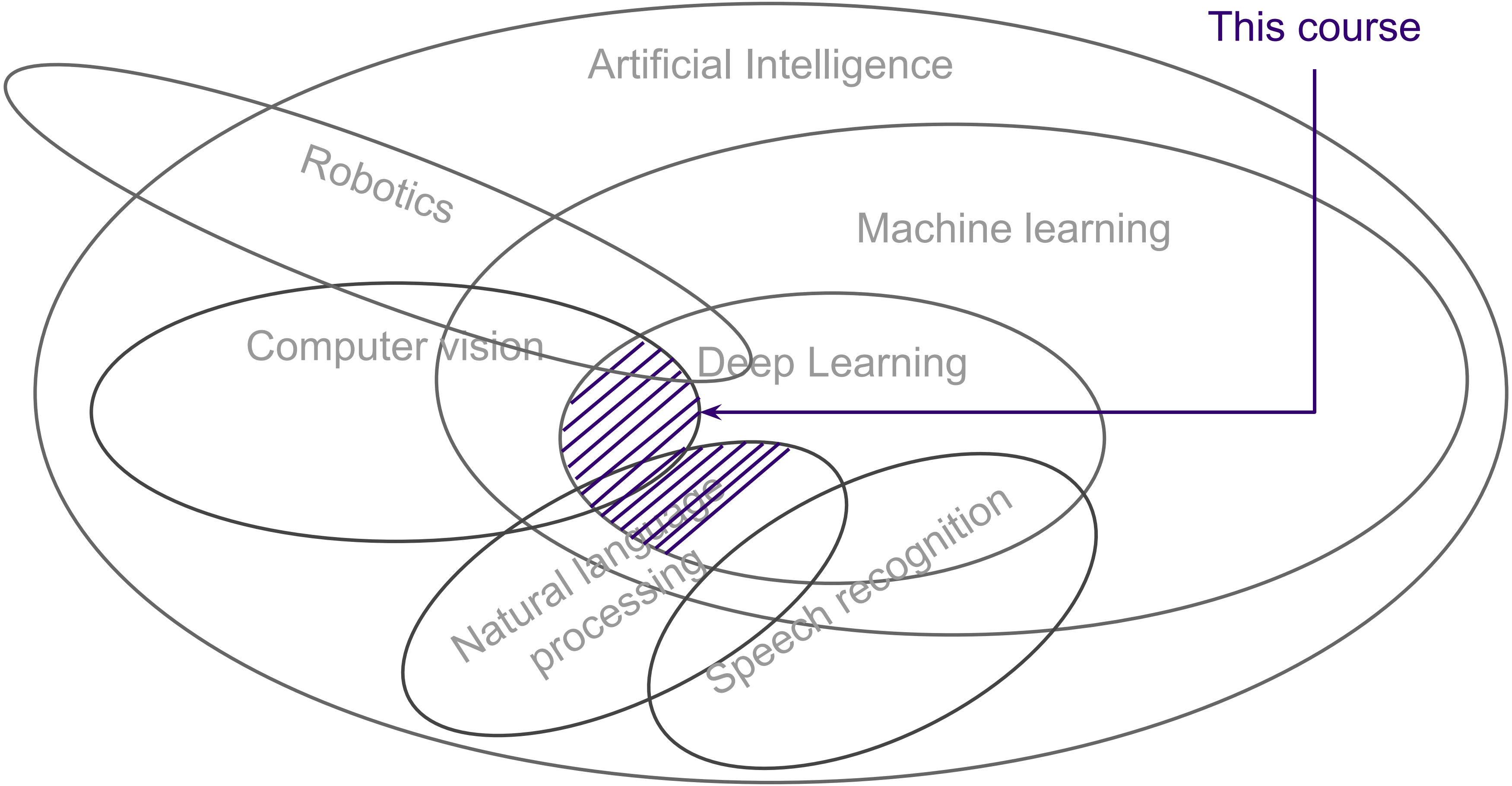

Artificial Intelligence

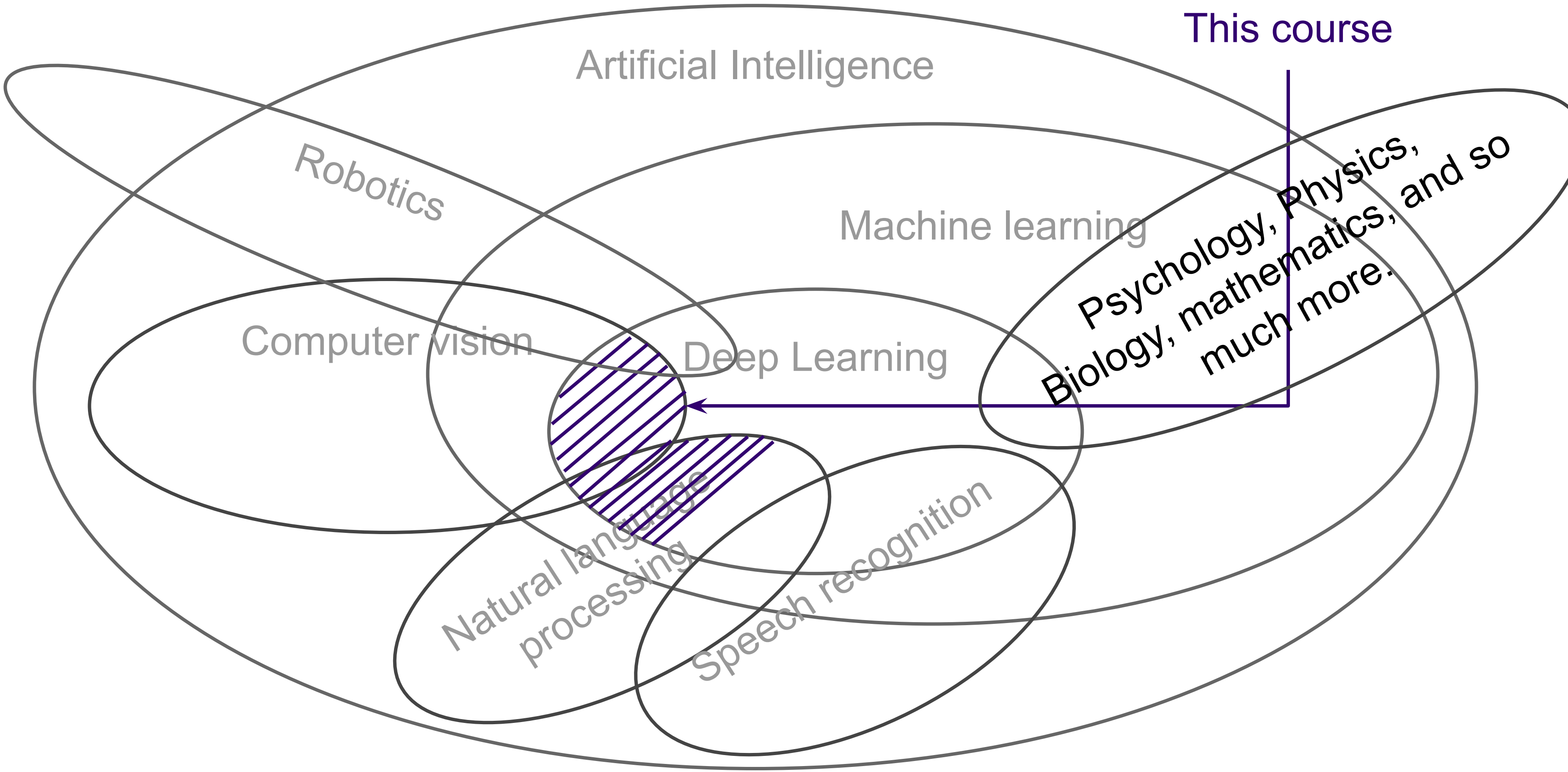

Artificial Intelligence
Machine learning


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


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


This course
Artificial Intelligence
Robotics
Machine learning
Psychology, Physics,
Biology, mathematics, and so
much more.
Computer vision
Deep Learning
Natural language
processing
Speech recognition

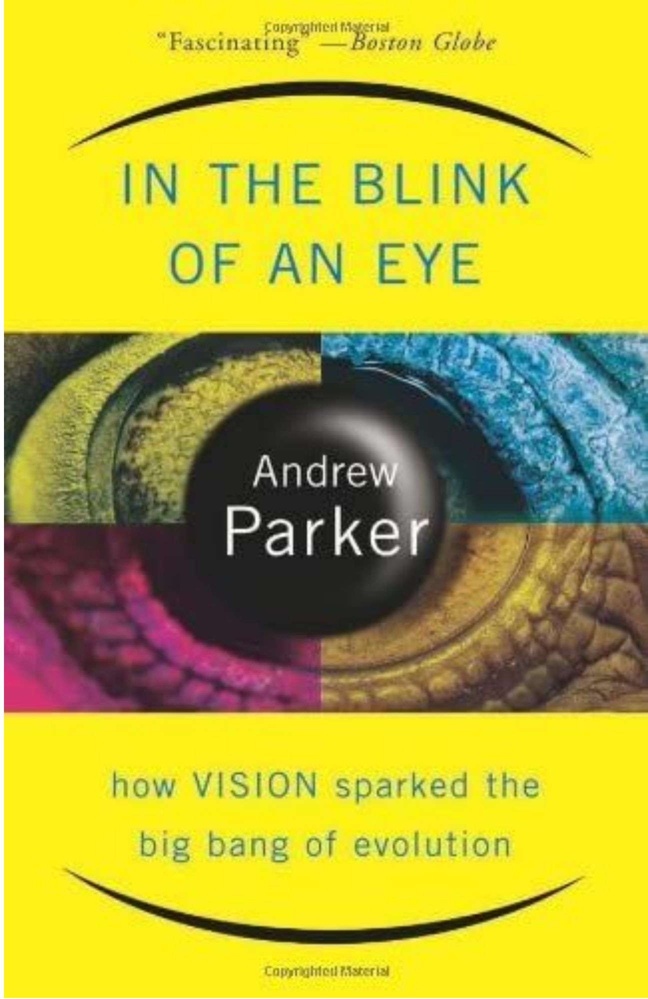
Today's agenda

- A brief history of deep learning
- CSE 493G1 overview

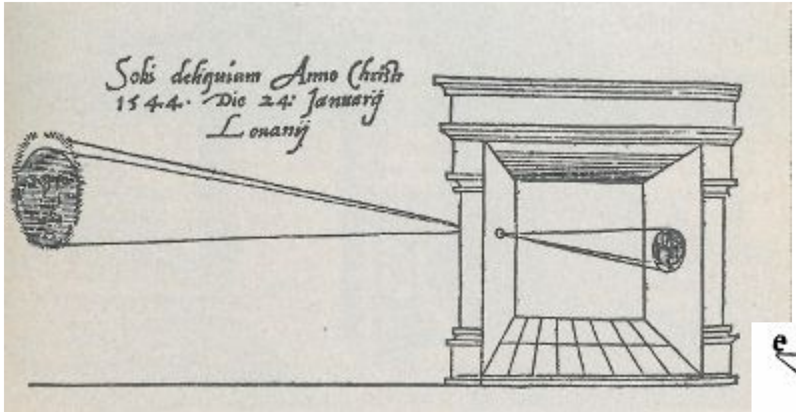

The European Camera of the Year 1983.
Fortunately, fine photography is a universal language.
PENTAX
PENTAX

Vision is core to the evolution of intelligence

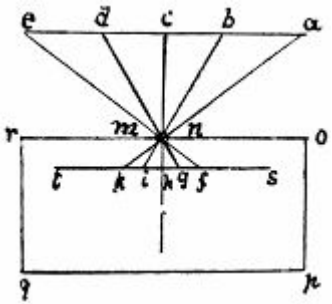
543 million years ago.



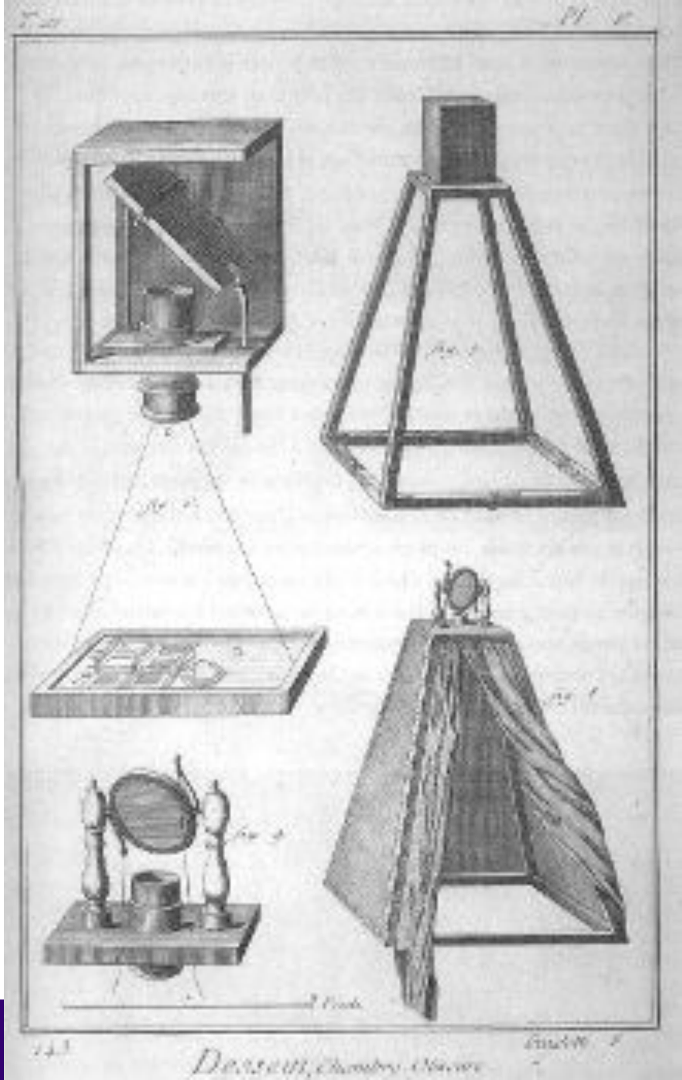
The first attempts at capturing the visual world



Camera obscura by Gemma Frisius, 1545

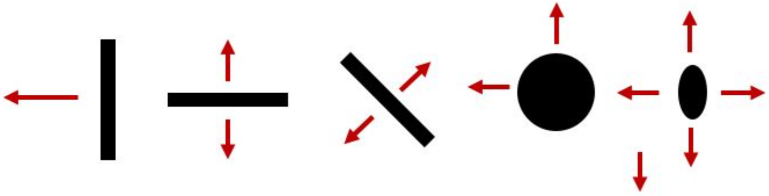


Inspired Leonardo da Vinci,
16th Century AD

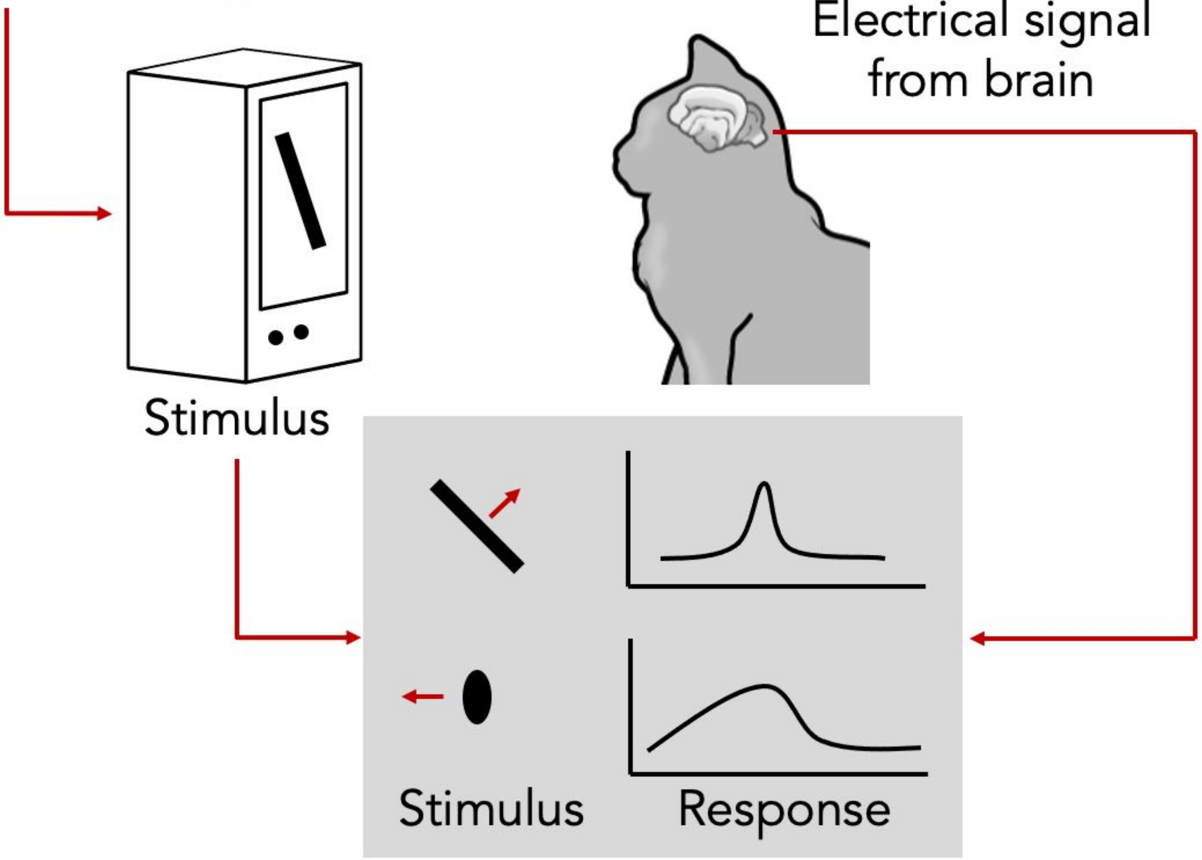


Examples from 18th century Encyclopedia

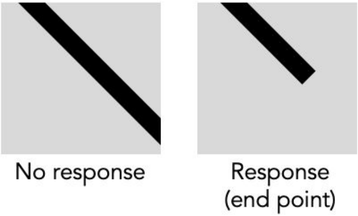
Hubel & Wiesel, 1959



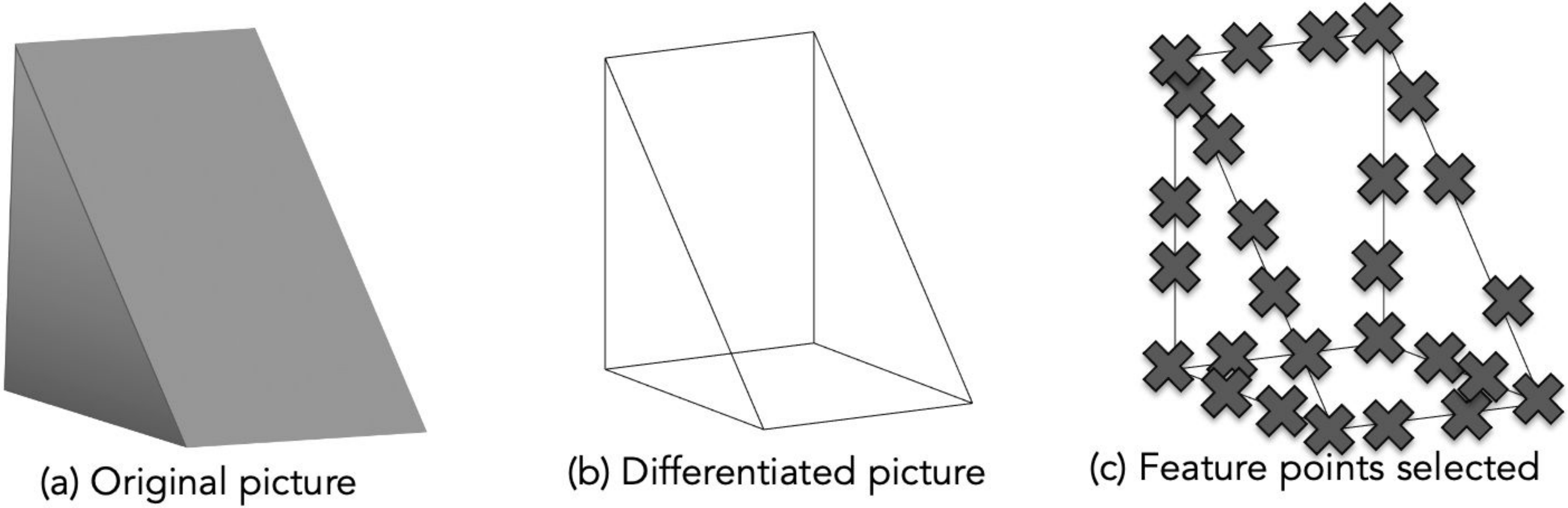
How does animal vision work?



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



Larry Roberts - Father of computer vision



Synthetic images, building up the visual world from simpler structures

The summer vision project

Organized by Seymour Papert

Computer vision was meant to be just a simple summer intern project

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group   July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

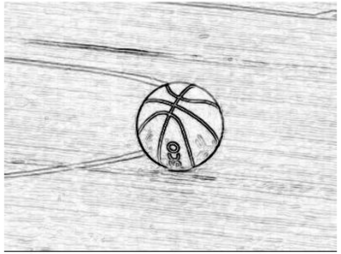
Seymour Papert

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".

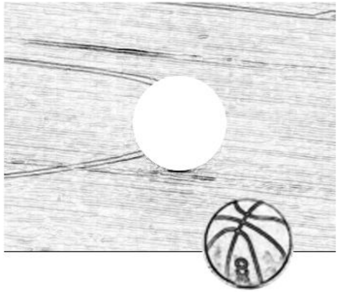
Input image

This image is CC0 1.0 public domain

Edge image



2 ½-D sketch



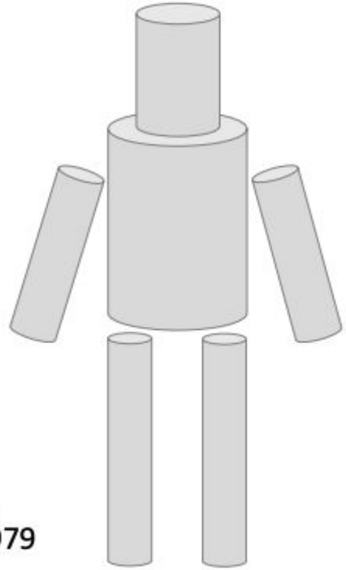
3-D model

This image is CC0 1.0 public domain

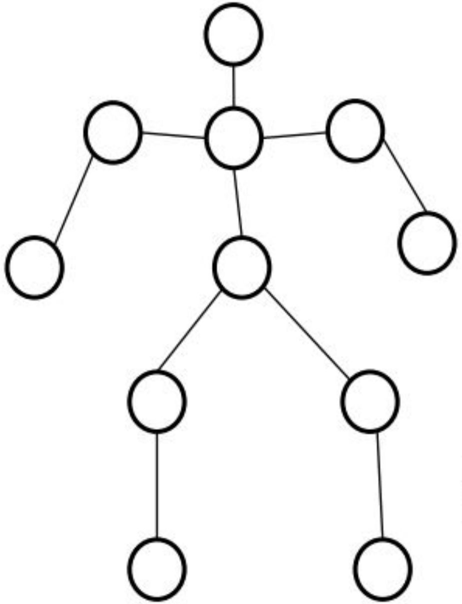
Input Image	Primal Sketch	2 ½-D Sketch	3-D Model Representation
Perceived intensities	Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries	Local surface orientation and discontinuities in depth and in surface orientation	3-D models hierarchically organized in terms of surface and volumetric primitives

David Marr, Stages of Visual Representation, 1970

Recognition via parts (1970s)



Generalized Cylinders,
Brooks and Binford, 1979

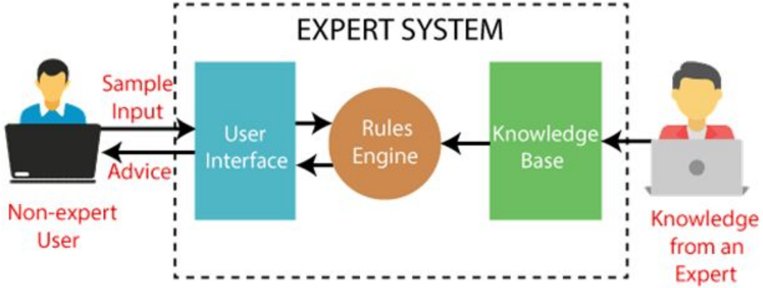


Pictorial Structures,
Fischler and Elshlager, 1973

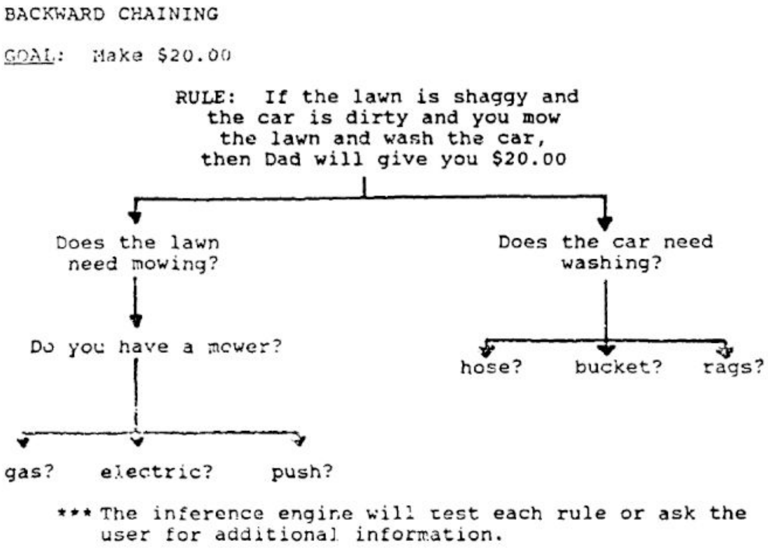
Recognition via edge detection (1980s)

John Canny, 1986 David Lowe, 1987

1980s caused one of the larger AI winters (the second AI winter)



Originally called heuristic programming project.

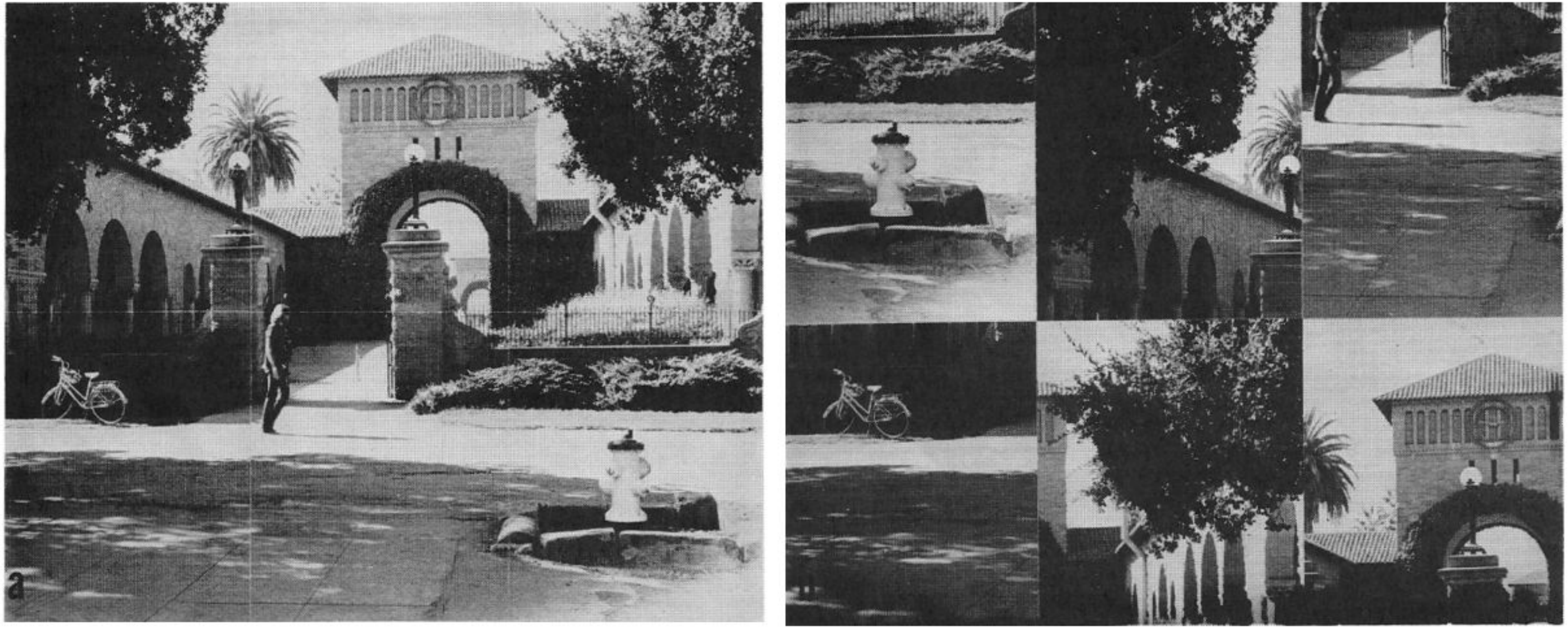


- Enthusiasm (and funding!) for AI research dwindled
- “Expert Systems” failed to deliver on their promises
- But subfields of AI continued to grow
    - Computer vision, NLP, robotics, compbio, etc.

In the meantime…seminal work in cognitive and neuroscience

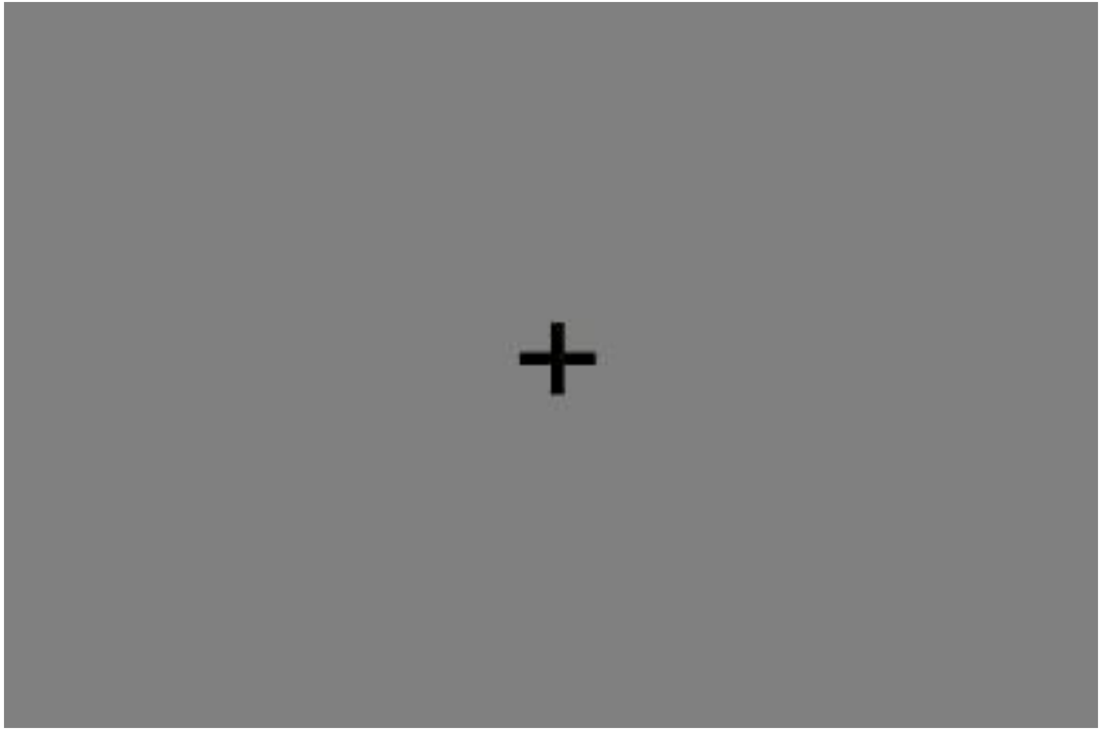
Perceiving real-world scenes

Irving Biederman



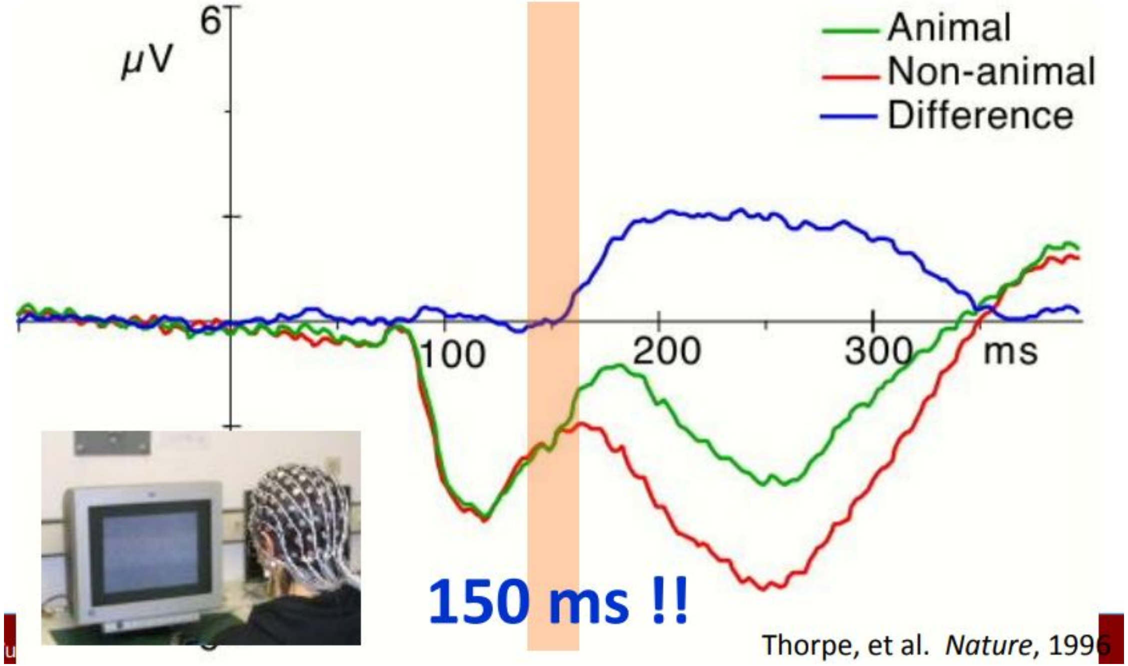
I. Biederman, *Science*, 1972

Rapid Serial Visual Perception (RSVP)

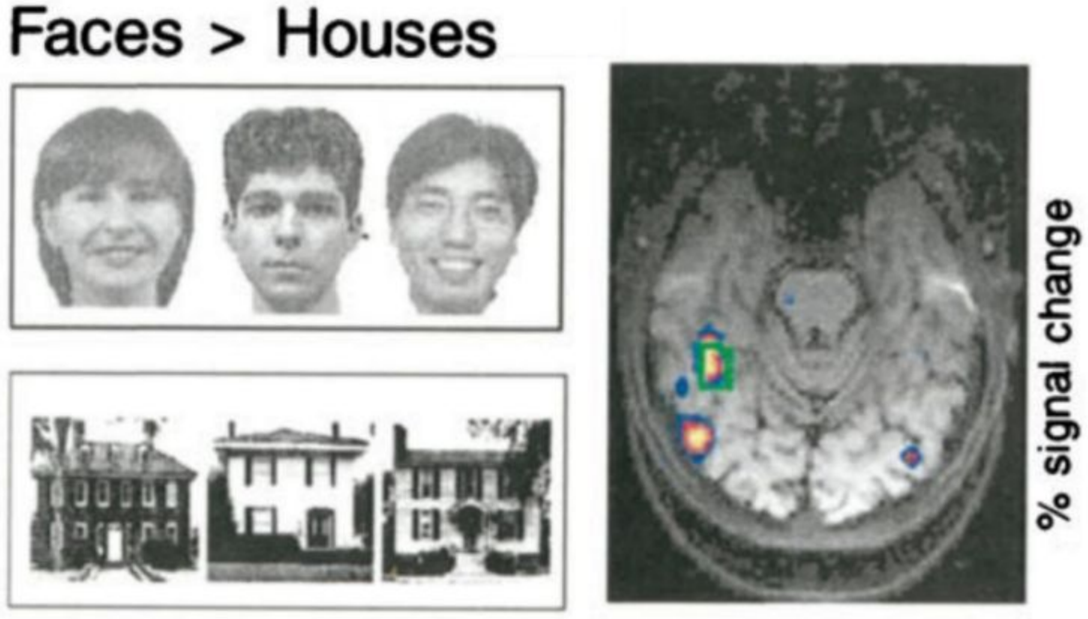


Potter, etc. 1970s

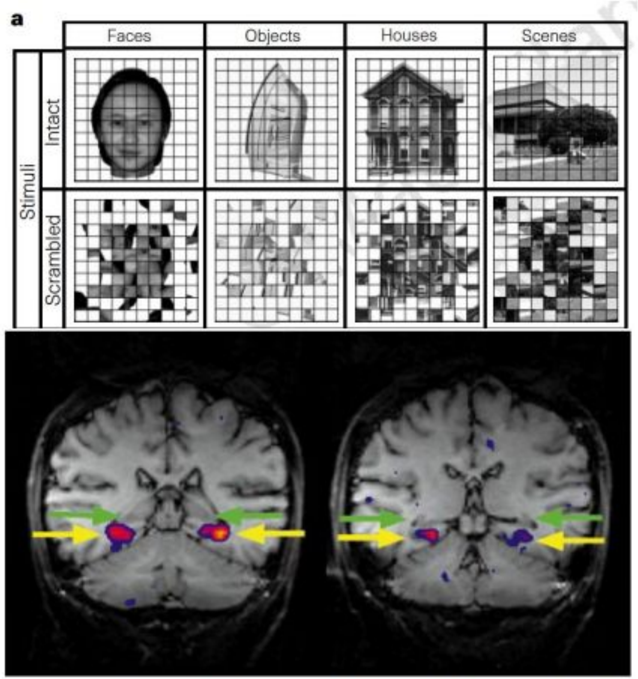
Speed of processing in the human visual system (Thorpe et al. Nature 1996)



Neural correlates of object & scene recognition



Kanwisher et al. J. Neuro. 1997



Epstein & Kanwisher, Nature, 1998

Visual recognition is a fundamental to intelligence

Searching for Computer Vision North Stars

AUTHORS: Fei-Fei Li and Ranjay Krishna

Until the 90s,
computer vision was not broadly applied to real world images

The focus was on algorithms! Recognition via Grouping (1990s)

Shi & Malik, *Normalized Cut*, 1997

Recognition via Matching (2000s)

Image is public domain

Image is public domain

SIFT, David Lowe, 1999

First commercial success of computer vision

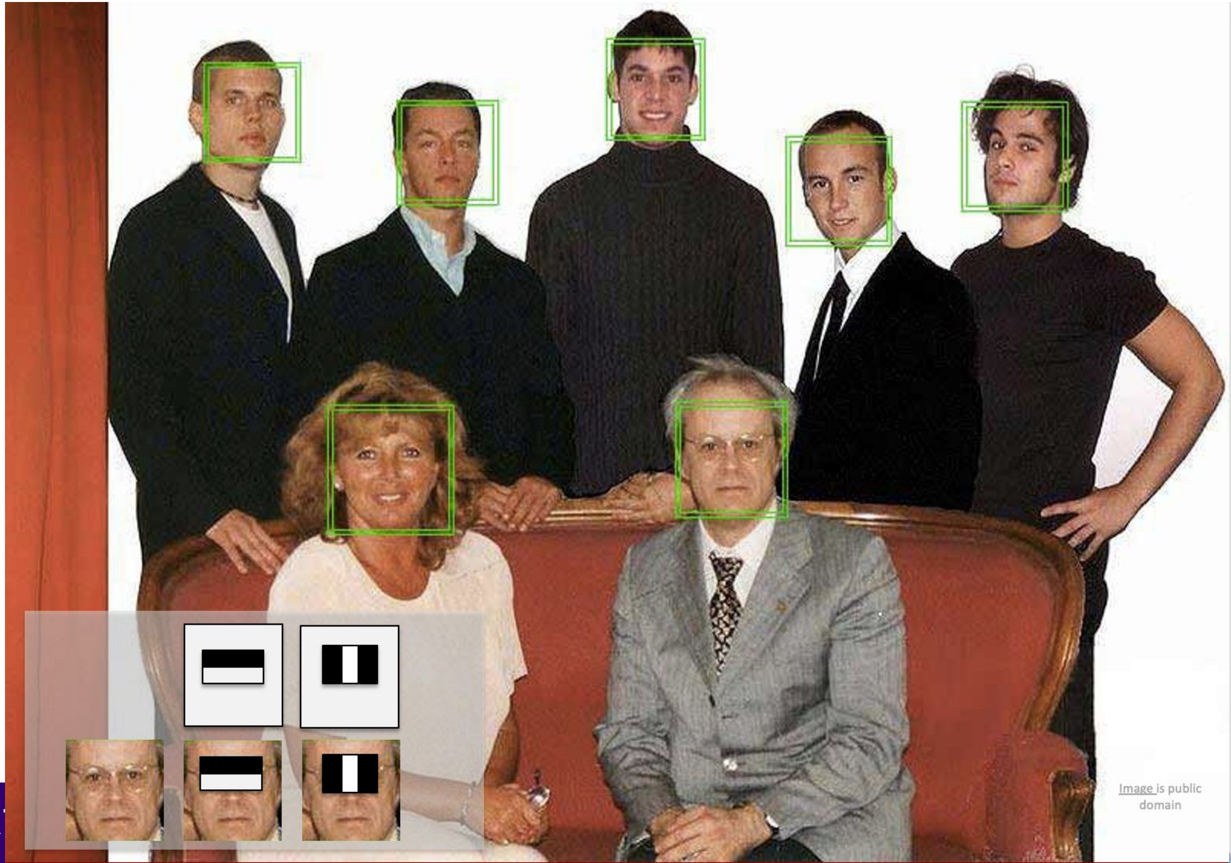
It came from embracing machine learning in 2001.

Does anyone know what it was?

First commercial success of computer vision

Real time face detection using using an algorithm by Viola and Jones, 2001

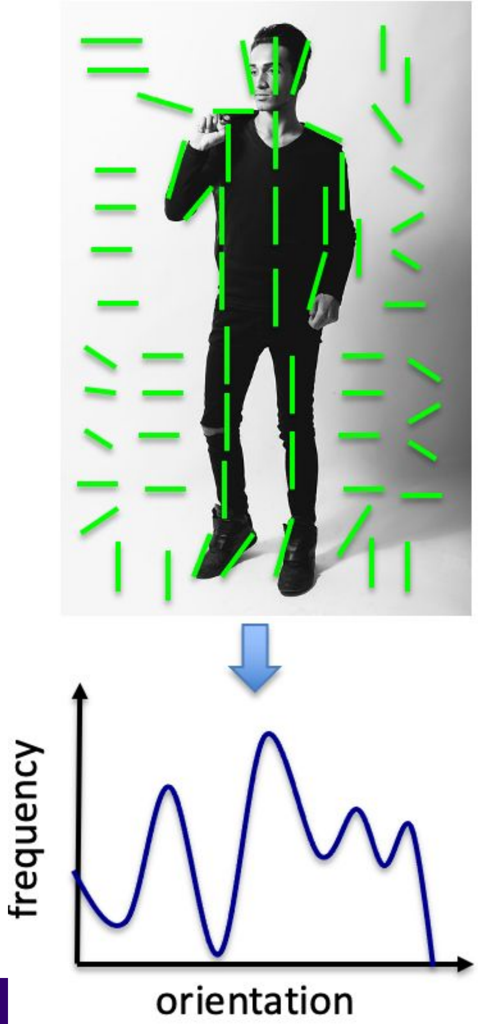
- Fujifilm face detection in cameras
- HP patent immediately



Designing better feature extraction became the focus

HoG features
- Histogram of oriented gradients
- Handcrafted

[Dalal & Triggs, HoG. 2005]



Caltech 101 images

PASCAL Visual Object Challenge

Image is CC0 1.0 public domain

Image is CC0 1.0 public domain

IMAGENET

www.image-net.org

22K categories and **14M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

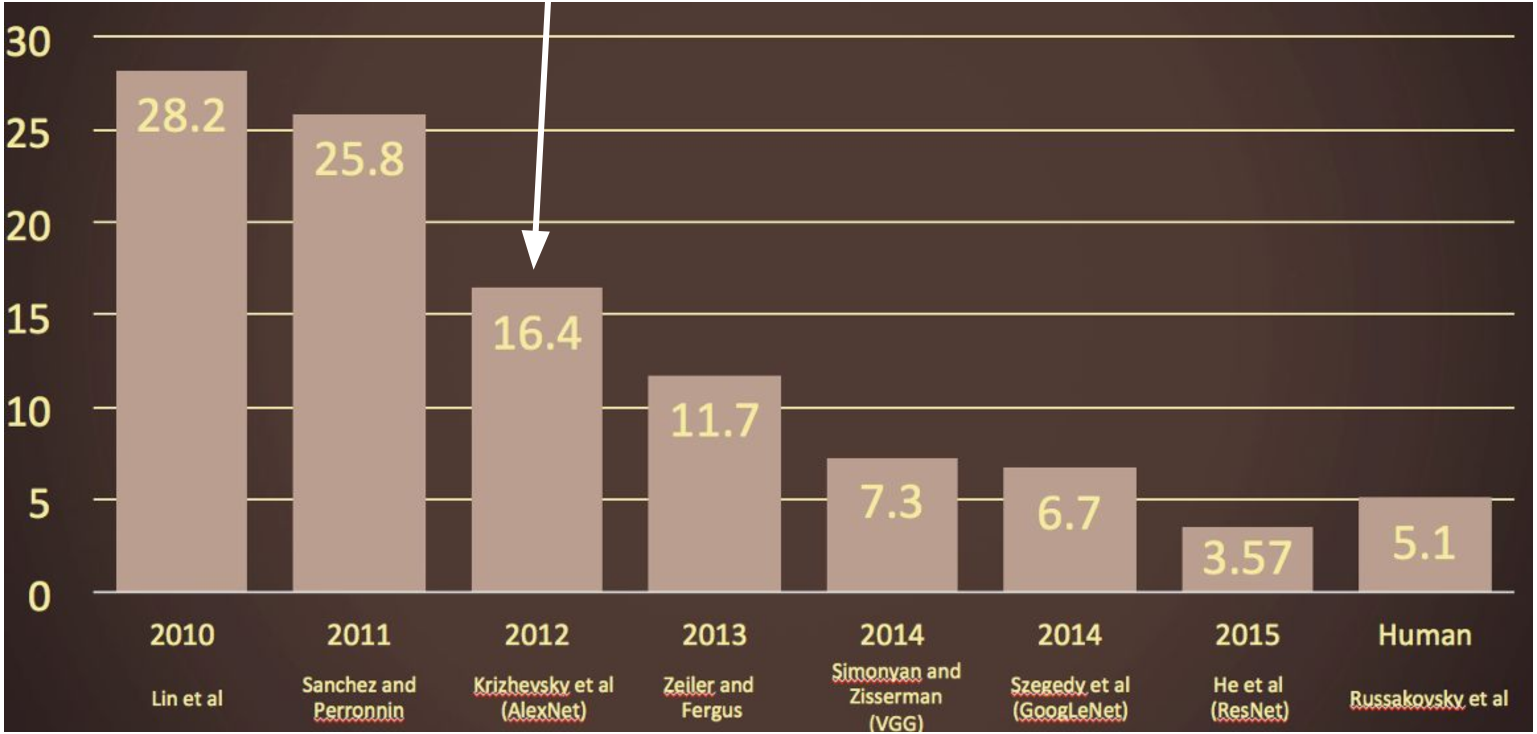
Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

Hypothesis behind ImageNet

- A child sees nearly 3K unique objects by the age of 6
- Calculated by Irving Biederman
    - [Biederman. Recognition-by-components: a theory of human image understanding. 1983]

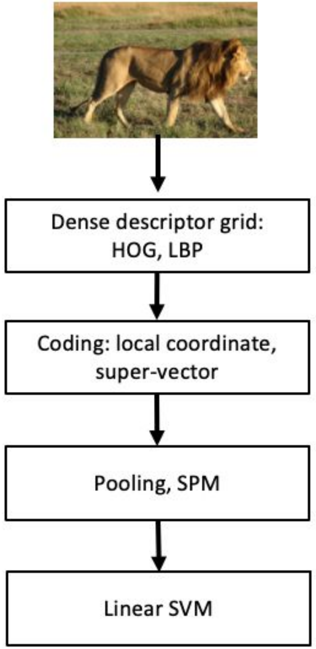
- But computer vision algorithms are trained on a handful of objects.

Object recognition accuracy drops by half in 2012 (Enter **deep learning**)



Year 2010

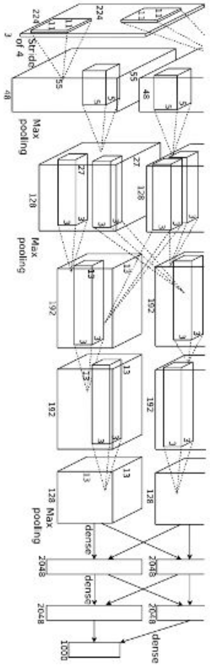
NEC-UIUC



[Lin CVPR 2011]

Year 2012

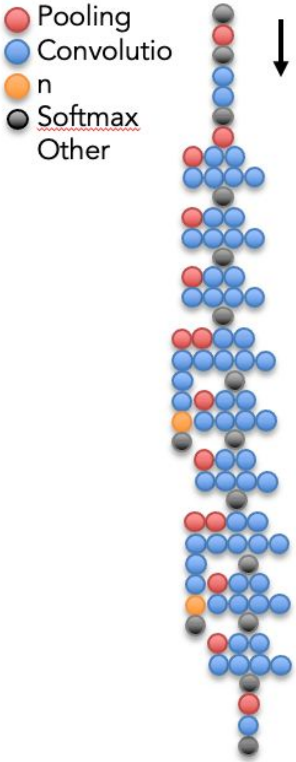
SuperVision



[Krizhevsky NIPS 2012]

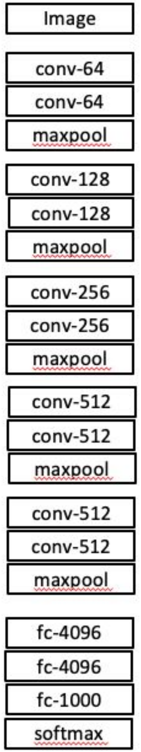
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

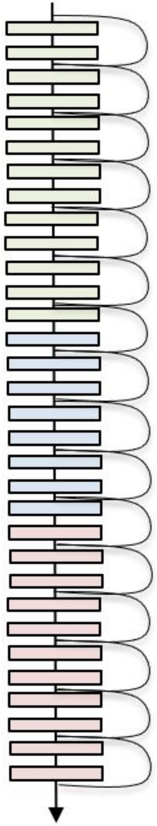
VGG



[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

AlexNet goes mainstream across computer vision

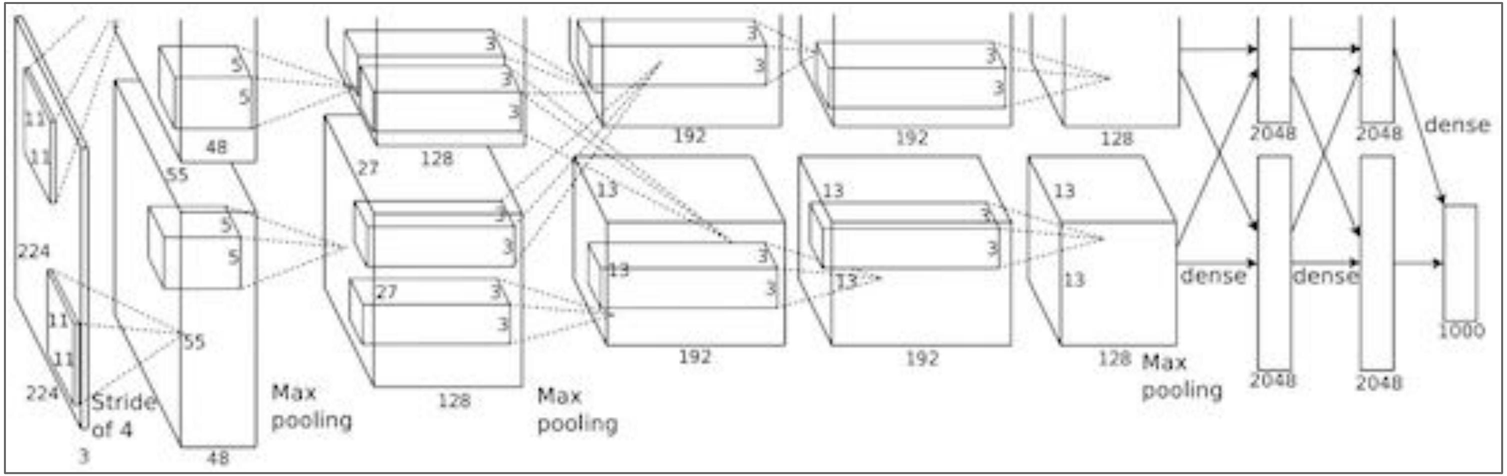


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

"AlexNet"

Core ideas go back many decades!

The **Mark I Perceptron** machine was the first implementation of the perceptron algorithm.

The machine was connected to a camera that used 20×20 cadmium sulfide photocells to produce a 400-pixel image.

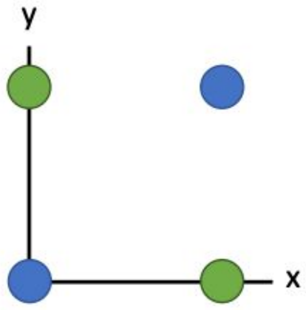
recognized
letters of the alphabet

Frank Rosenblatt, ~1957: Perceptron

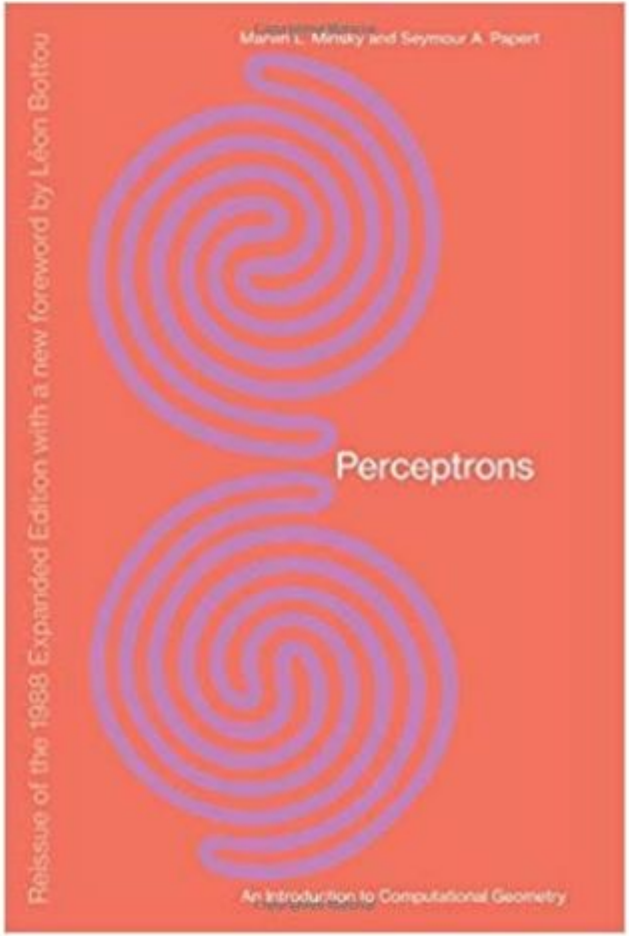
This image by Rocky Acosta is licensed under CC-BY 3.0

Minsky and Papert, 1969

X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0



Showed that Perceptrons could not learn the XOR function
Caused a lot of disillusionment in the field

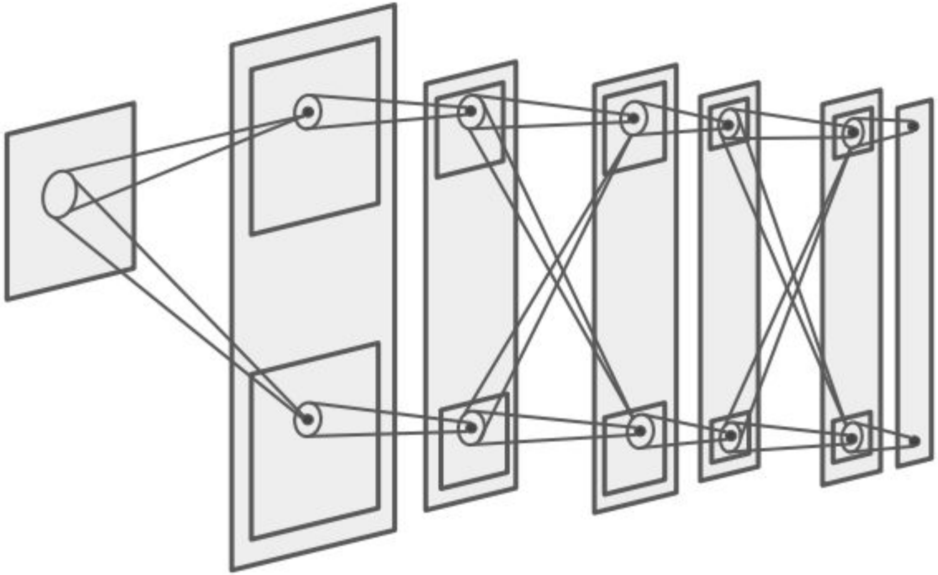


Neocognitron: Fukushima, 1980

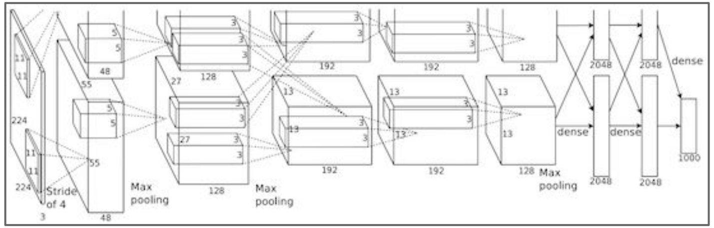
Computational model the visual system, directly inspired by Hubel and Wiesel's hierarchy of complex and simple cells

Interleaved simple cells (convolution)
and complex cells (pooling)

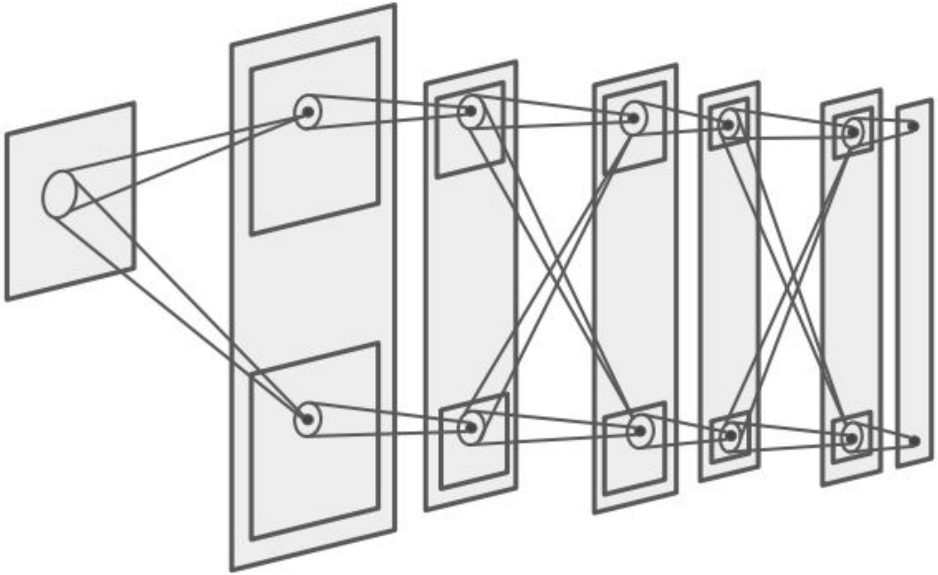
No practical training algorithm



A lot like AlexNet today



"AlexNet"



Backprop: Rumelhart, Hinton, and Williams, 1986

Introduced backpropagation for computing gradients in neural networks

Successfully trained perceptrons with multiple layers

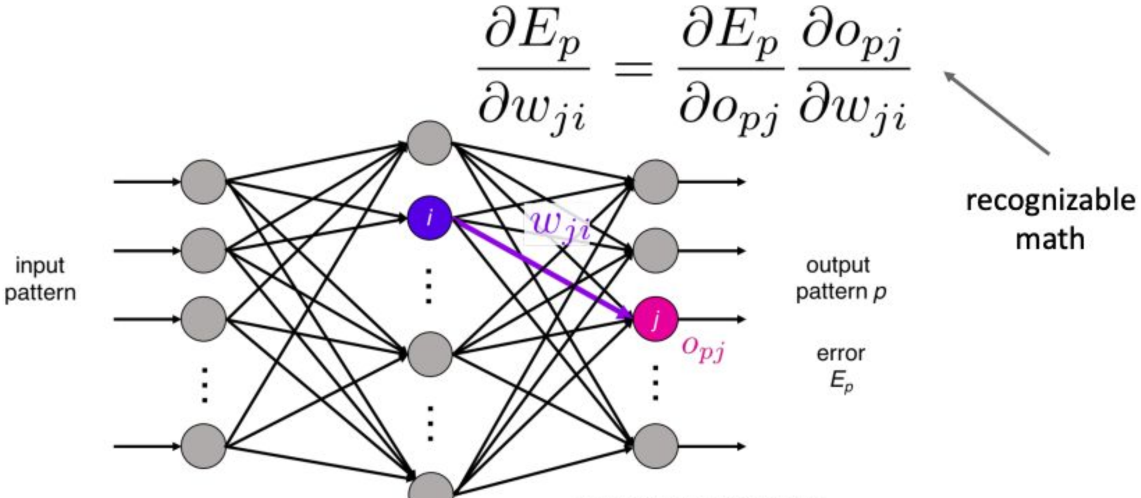


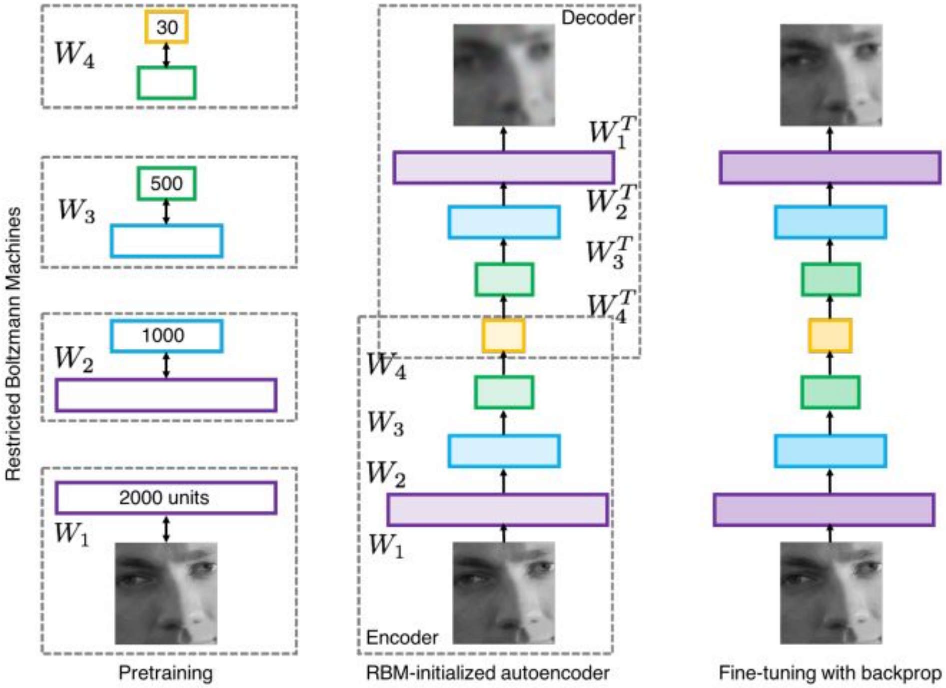
Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017

2000s: “Deep Learning”

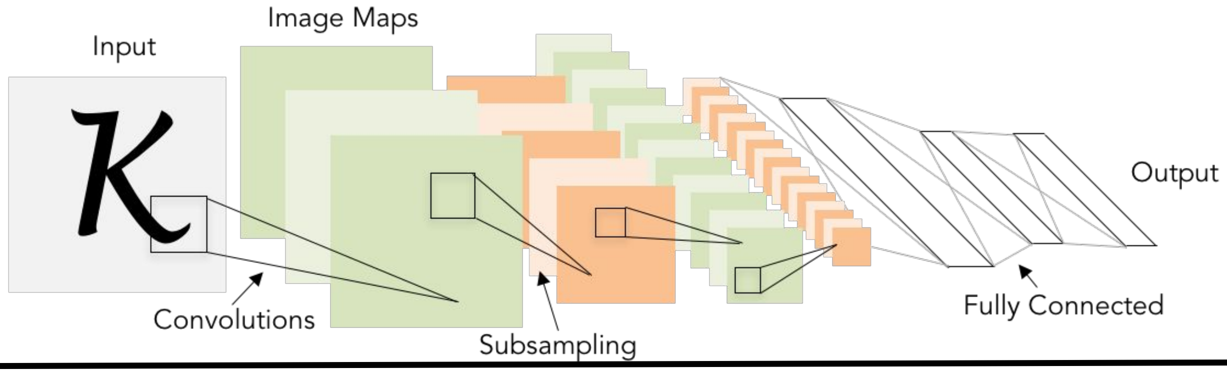
People tried to train neural networks that were deeper and deeper

Not a mainstream research topic at this time

Hinton and Salakhutdinov, 2006
Bengio et al, 2007 Lee et al, 2009
Glorot and Bengio, 2010



1998 LeCun et al.



# of transistors

10^6

# of pixels used to train:

10^7 NIST

2012 Krizhevsky et al.

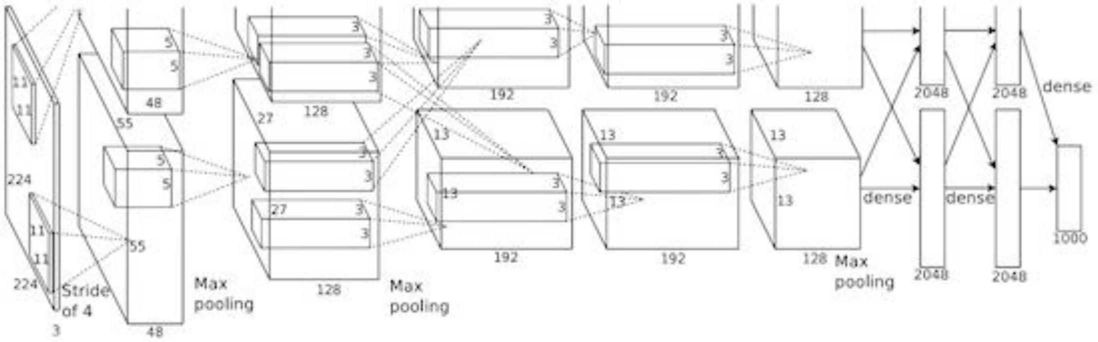


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

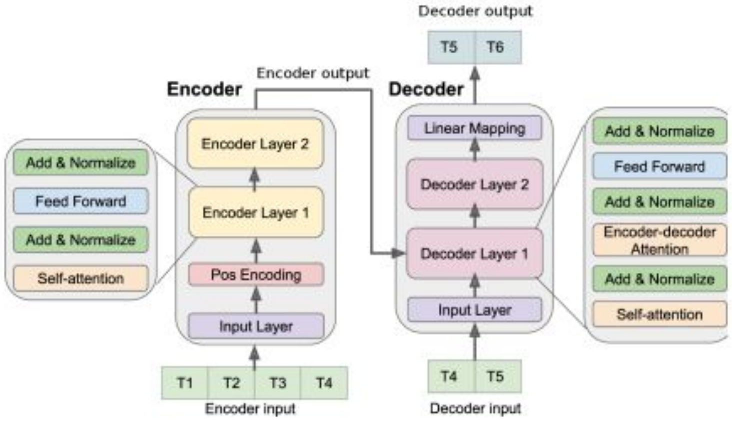
# of transistors

10^9

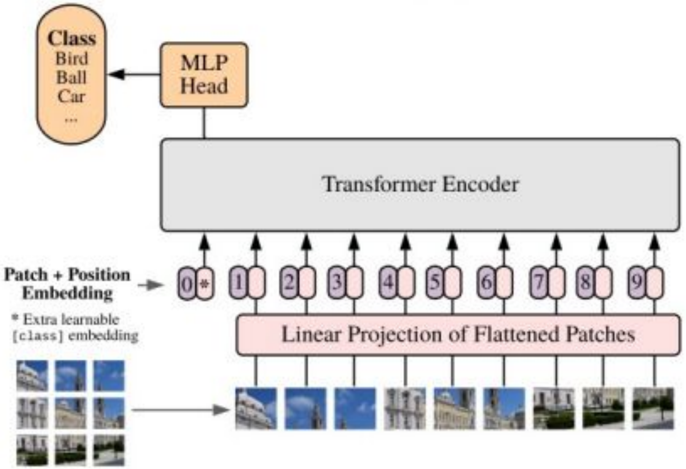
# of pixels used to train:

10^{14} IMAGENET

Today: Homogenization of Deep Learning
Same models for GPT-4 and image recognition



Transformer Models originally designed for NLP



Almost identical model (Visual Transformers) can be applied to Computer Vision tasks

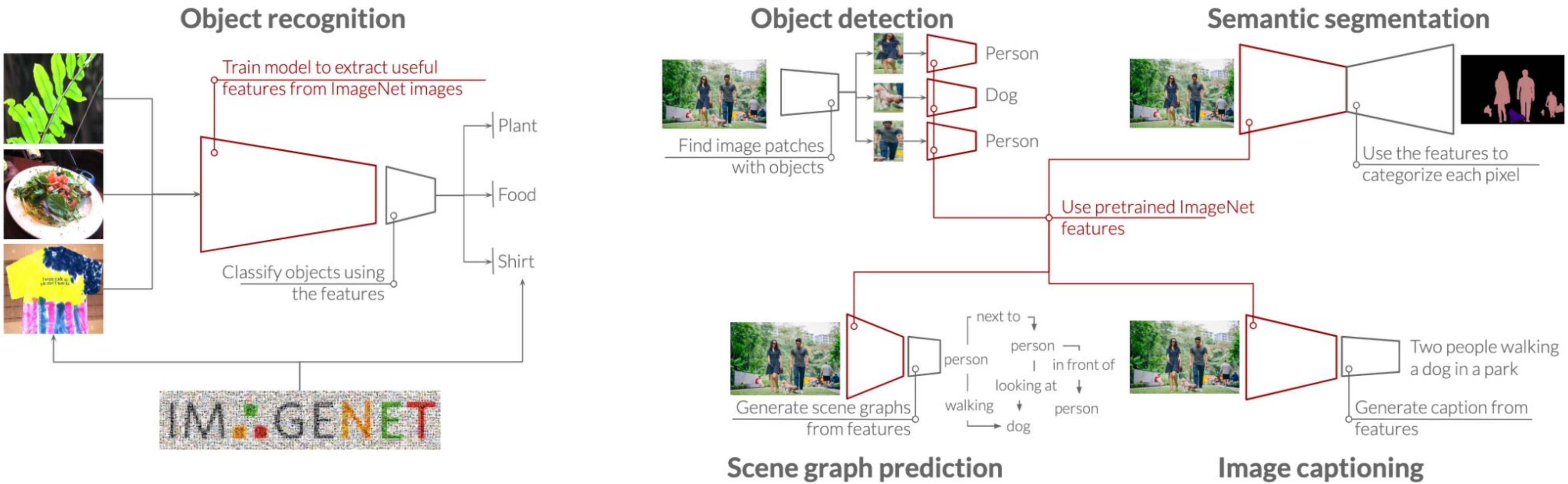
2012 to present: deep learning is everywhere

Image Classification

Image Retrieval

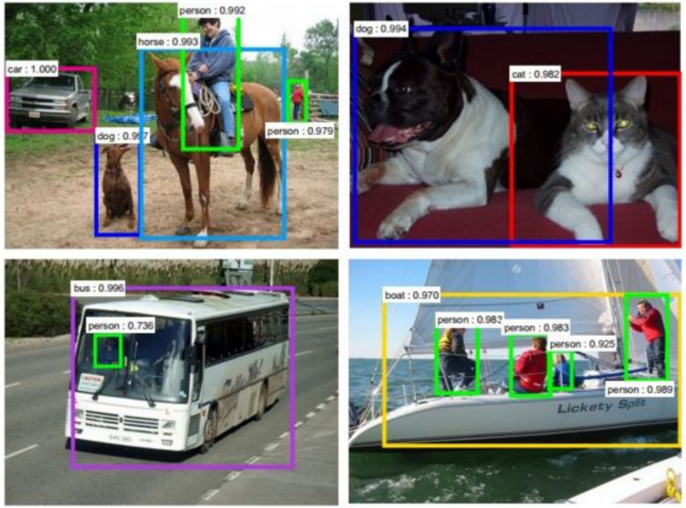
Data hungry machine learning models are now everywhere

Pretraining on ImageNet for object classification → Transfer ImageNet features for many other tasks:



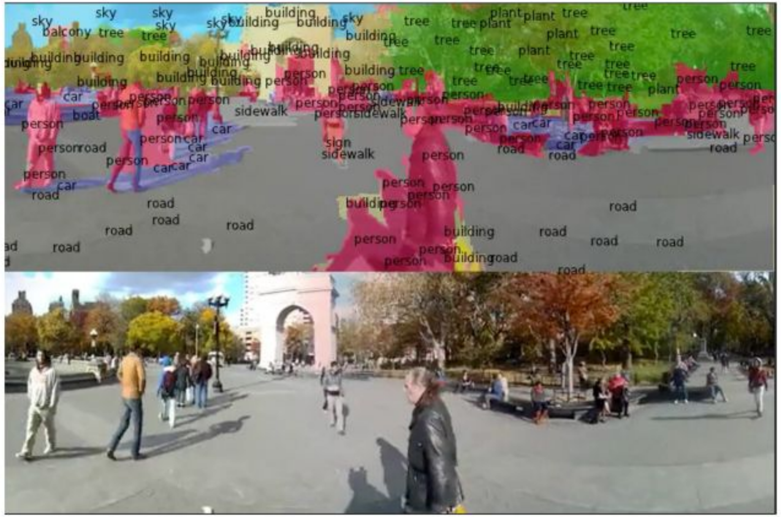
2012 to Present: Deep Learning is Everywhere

Object Detection



Ren, He, Girshick, and Sun, 2015

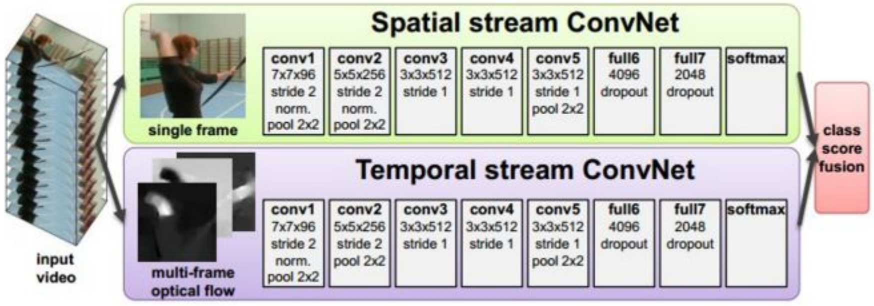
Image Segmentation



Fabaret et al, 2012

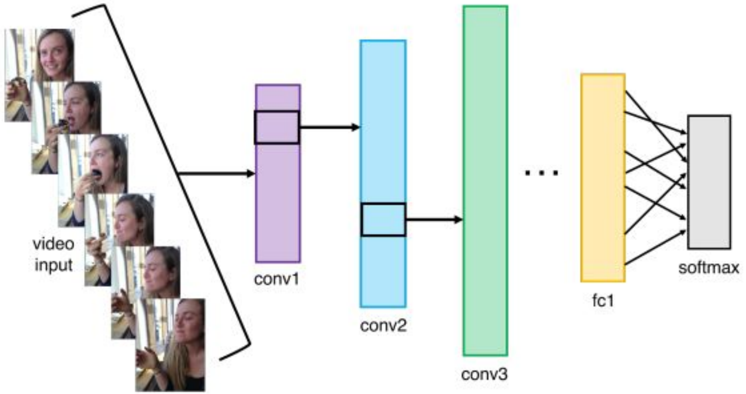
2012 to Present: Deep Learning is Everywhere

Video Classification



Simonyan et al, 2014

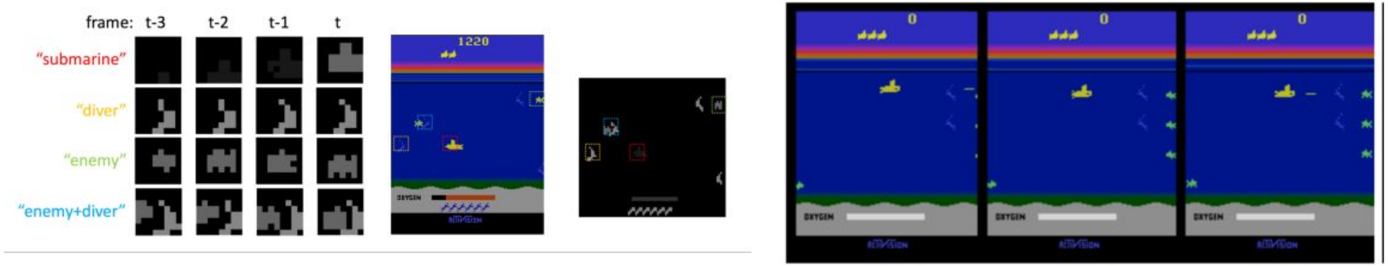
Activity Recognition



2012 to Present: Deep Learning is Everywhere

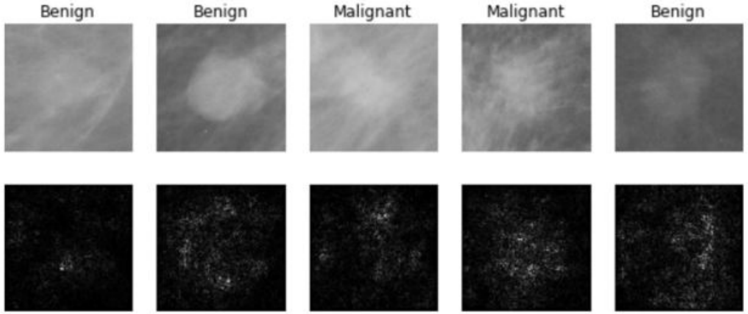
Pose Recognition (Toshev and Szegedy, 2014)

Playing Atari games (Guo et al, 2014)



2012 to Present: Deep Learning is Everywhere

Medical Imaging



Levy et al, 2016 Figure reproduced with permission

Galaxy Classification

Dieleman et al, 2014

From left to right: public domain by NASA, usage permitted by ESA/Hubble, public domain by NASA, and public domain.

Whale recognition

Kaggle Challenge

This image by Christin Khan is in the public domain and originally came from the U.S. NOAA.

2012 to Present: Deep Learning is Everywhere

Image Captioning

Vinyals et al, 2015
Karpathy and Fei-Fei, 2015

A white teddy bear sitting in the grass

A man in a baseball uniform throwing a ball

A woman is holding a cat in her hand

A man riding a wave on top of a surfboard

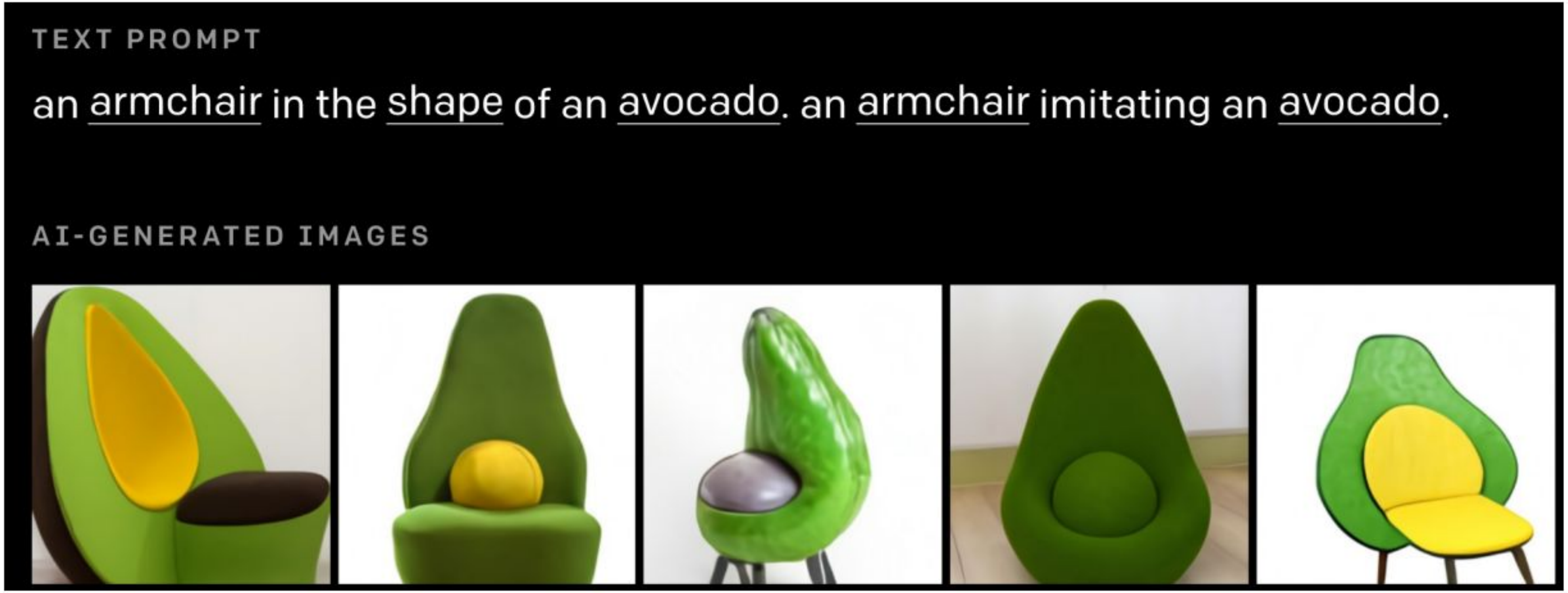
A cat sitting on a suitcase on the floor

A woman standing on a beach holding a surfboard

All images are CC0 Public domain:
https://pixabay.com/en/luggage-antique-cat-1643010/
https://pixabay.com/en/teddy-plush-bears-cute-teddy-bear-1623436/
https://pixabay.com/en/surf-wave-summer-sport-litoral-1668716/
https://pixabay.com/en/woman-female-model-portrait-adult-983967/
https://pixabay.com/en/handstand-lake-meditation-496008/
https://pixabay.com/en/baseball-player-shortstop-infield-1045263/

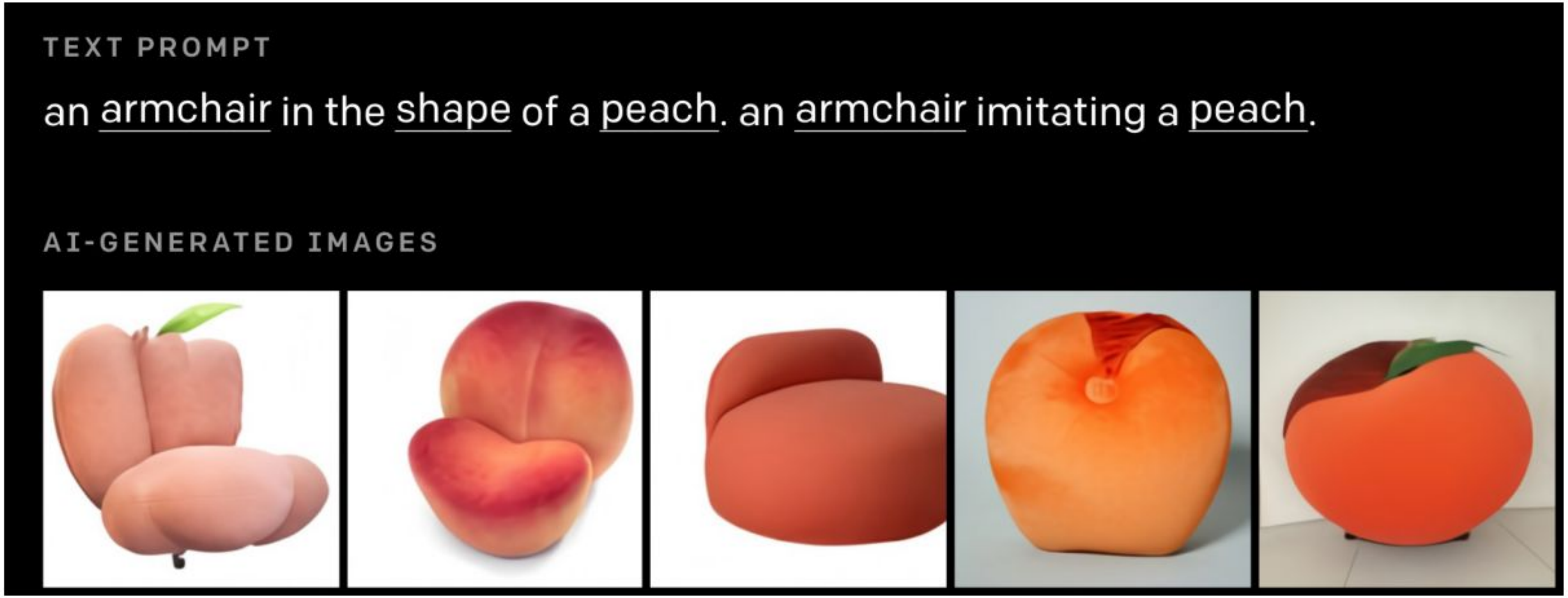
Captions generated by Justin Johnson using Neuraltalk2

2012 to Present: Deep Learning is Everywhere



Ramesh et al, “DALL·E: Creating Images from Text”, 2021. https://openai.com/blog/dall-e/

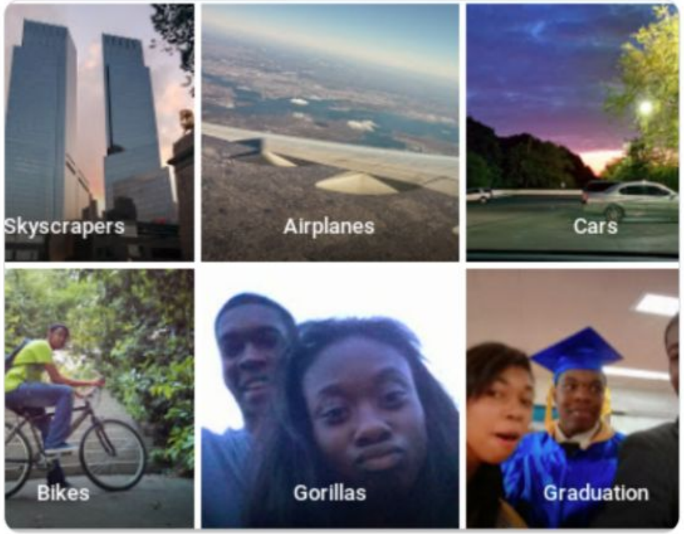
2012 to Present: Deep Learning is Everywhere



Ramesh et al, “DALL·E: Creating Images from Text”, 2021. https://openai.com/blog/dall-e/

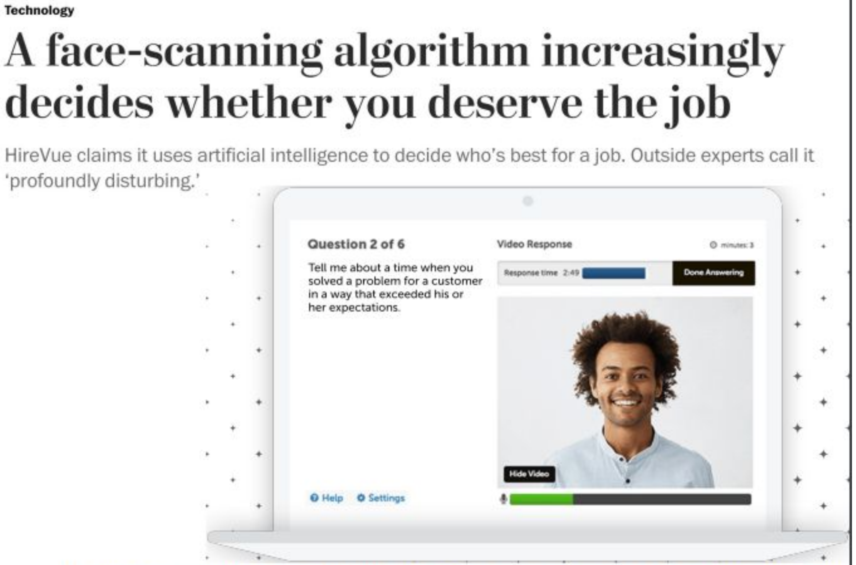
Despite progress, deep learning can be harmful

Harmful Stereotypes



Barocas et al, "The Problem With Bias: Allocative Versus Representational Harms in Machine Learning", SIGCIS 2017
Kate Crawford, "The Trouble with Bias", NeurIPS 2017 Keynote
Source: https://twitter.com/jackyalcine/status/615329515909156865 (2015)

Affect people's lives

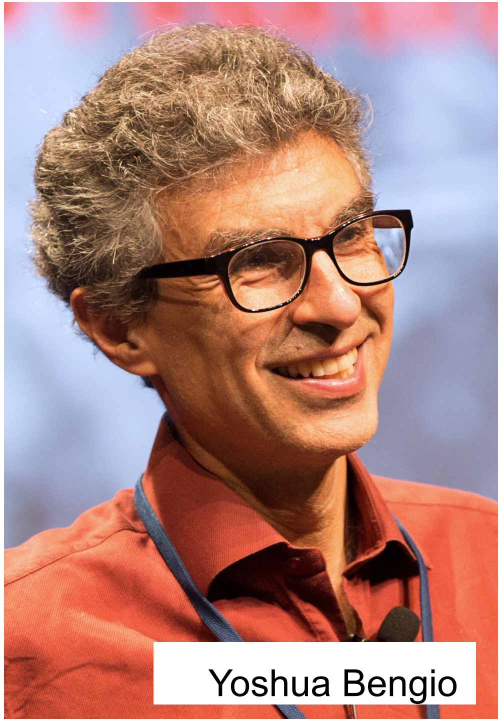


Source: https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/
https://www.hirevue.com/platform/online-video-interviewing-software
Example Credit: Timnit Gebru

2018 Turing Award for deep learning

most prestigious technical award, is given for major contributions of lasting importance to computing.

This image is CC0 public domain



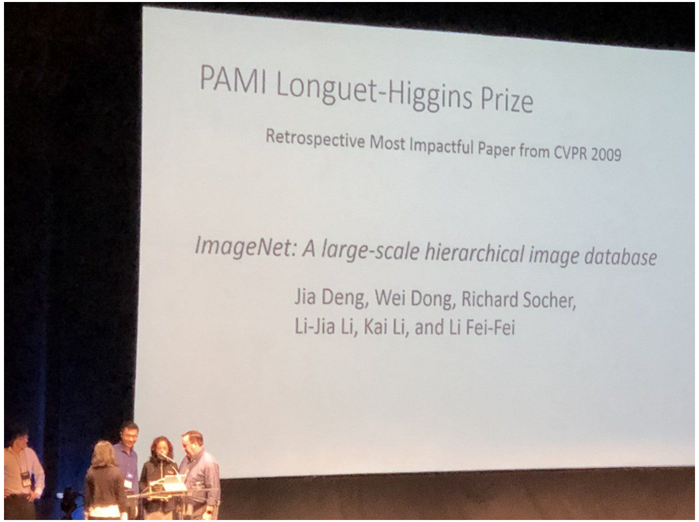
This image is CC0 public domain

This image is CC0 public domain

IEEE PAMI Longuet-Higgins Prize

Award recognizes ONE Computer Vision paper from **ten years ago** with **significant impact on computer vision** research.

In 2019, it was awarded to the 2009 original ImageNet paper




Algorithms
Data
Computation
13
13
13
192
192
128 Max
dense
dense
1000
2048
2048
GTX
GTX

In this course, we will study these algorithms and architectures starting from a grounding in Visual Recognition

A fundamental and general problem in Computer Vision, that has roots in Cognitive Science

Image Classification: A core task in Computer Vision

This image by Nikita is
licensed under CC-BY 2.0

→ cat

Image by US Army is licensed under CC BY 2.0

Image is CC0 1.0 public domain

Image by Kippelboy is licensed under CC BY-SA 3.0

Image by Christina C. is licensed under CC BY-SA 4.0

Object detection
car

This image is licensed under CC BY-NC-SA 2.0; changes made

Action recognition
bicycling

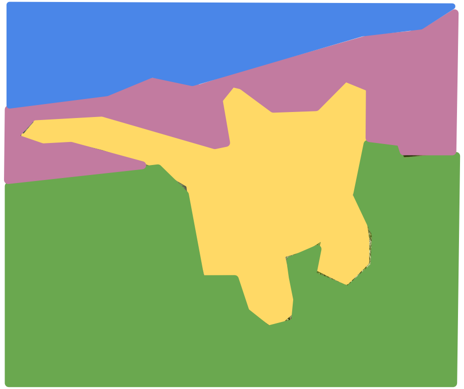
This image is licensed under CC BY-SA 3.0; changes made

Scene graph prediction
<person - holding - hammer>

Captioning:
a person holding a hammer

This image is licensed under CC BY-SA 3.0; changes made

Beyond recognition: Segmentation, 2D/3D Generation



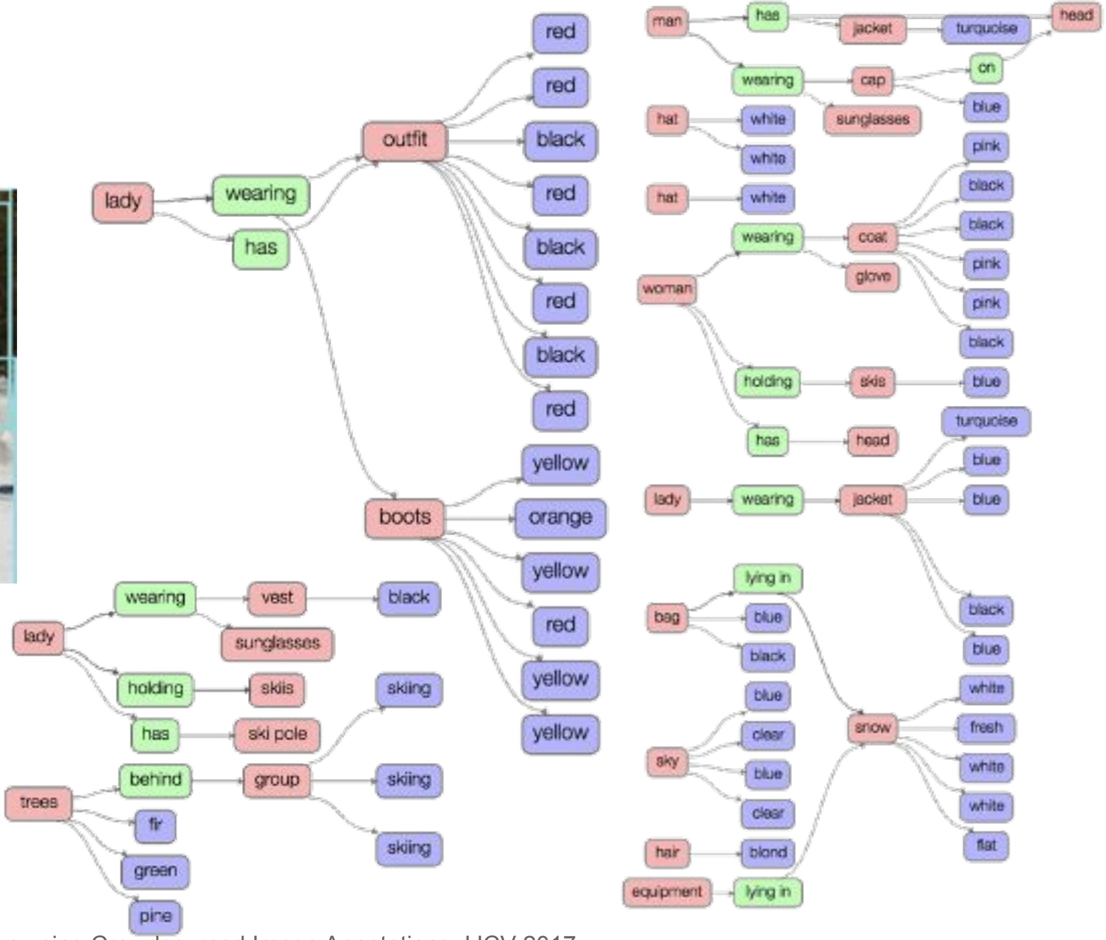
This image is CC0 public domain

Progressive GAN, Karras 2018.

Wang et al, “Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images”, ECCV 2018

Scene Graphs

This image is CC0 public domain



[Three Ways Computer Vision Is Transforming Marketing]

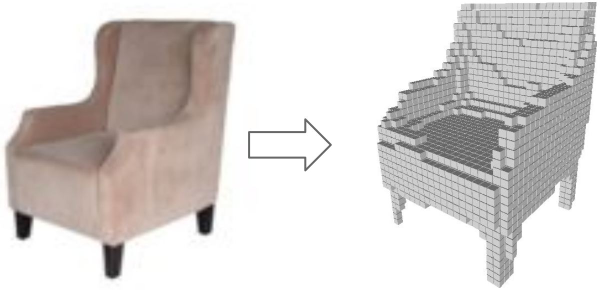
- Forbes Technology Council

Krishna et al., Visual Genome: Connecting Vision and Language using Crowdsourced Image Annotations, IJCV 2017

Spatio-temporal scene graphs

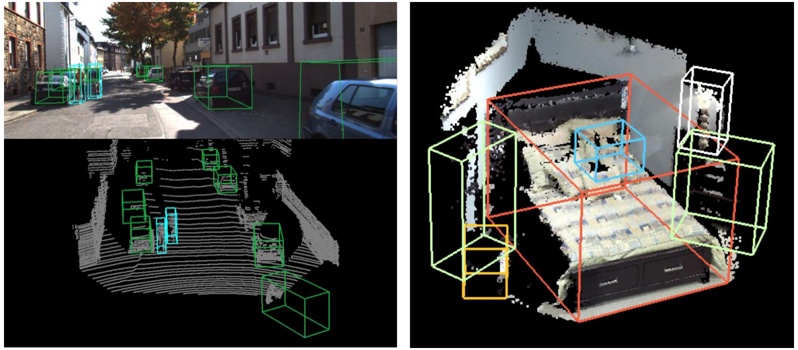
Ji, Krishna et al., Action Genome: Actions as Composition of Spatio-temporal Scene Graphs, CVPR 2020

3D Vision & Robotic Vision

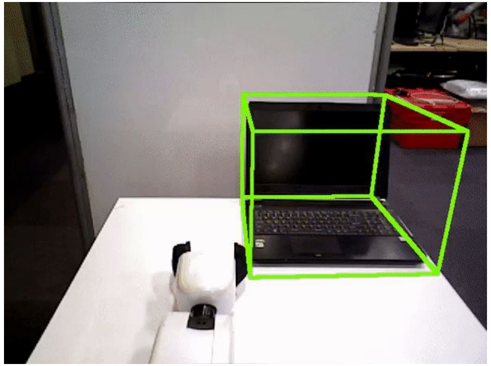


Choy et al., 3D-R2N2: Recurrent Reconstruction Neural Network (2016)

Mandlekar and Xu et al., Learning to Generalize Across Long-Horizon Tasks from Human Demonstrations (2020)



Xu et al., PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation (2018)



Wang et al., 6-PACK: Category-level 6D Pose Tracker with Anchor-Based Keypoints (2020)

PT = 500ms

Image is licensed under CC BY-SA 3.0; changes made

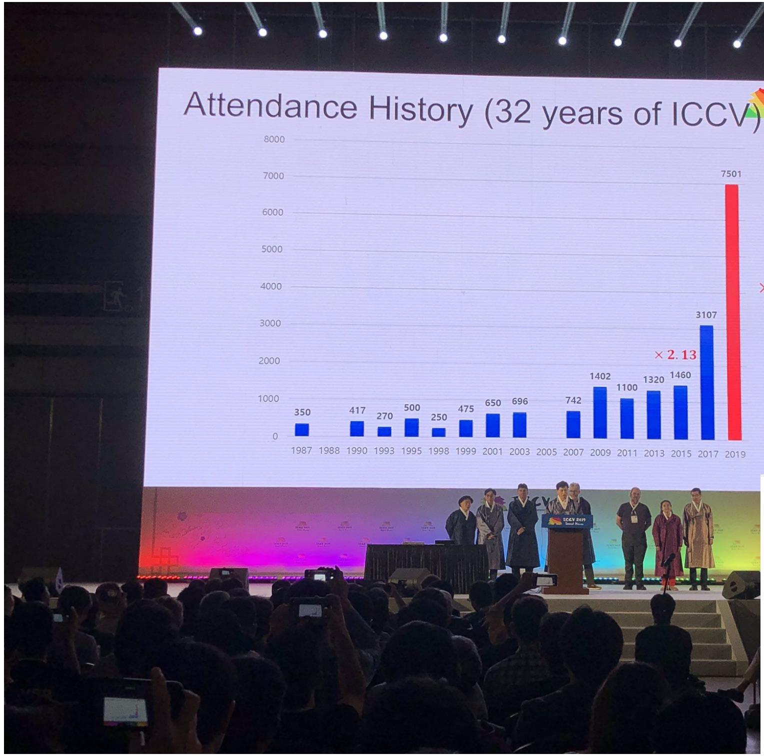
Some kind of game or fight. Two groups of two men? The man on the left is throwing something. Outdoors seemed like because i have an impression of grass and maybe lines on the grass? That would be why I think perhaps a game, rough game though, more like rugby than football because they pairs weren't in pads and helmets, though I did get the impression of similar clothing. maybe some trees? in the background.

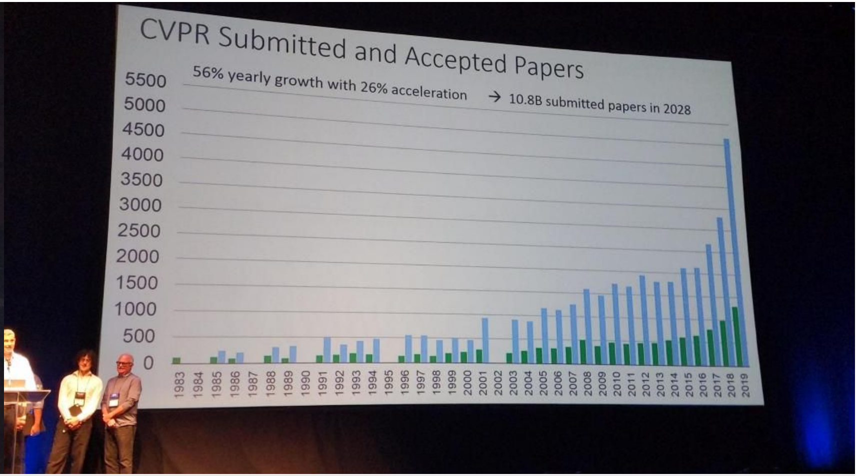
Fei-Fei, Iyer, Koch, Perona, *JoV,* 2007

And there is a lot we don't know how to do

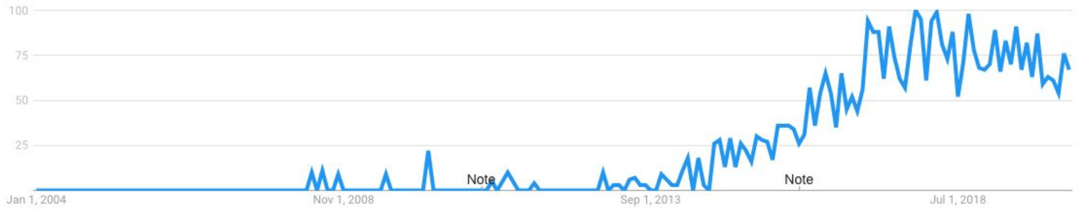
https://fedandfit.com/wp-content/uploads/2020/06/summer-activities-for-kids_optimized-scaled.jpeg

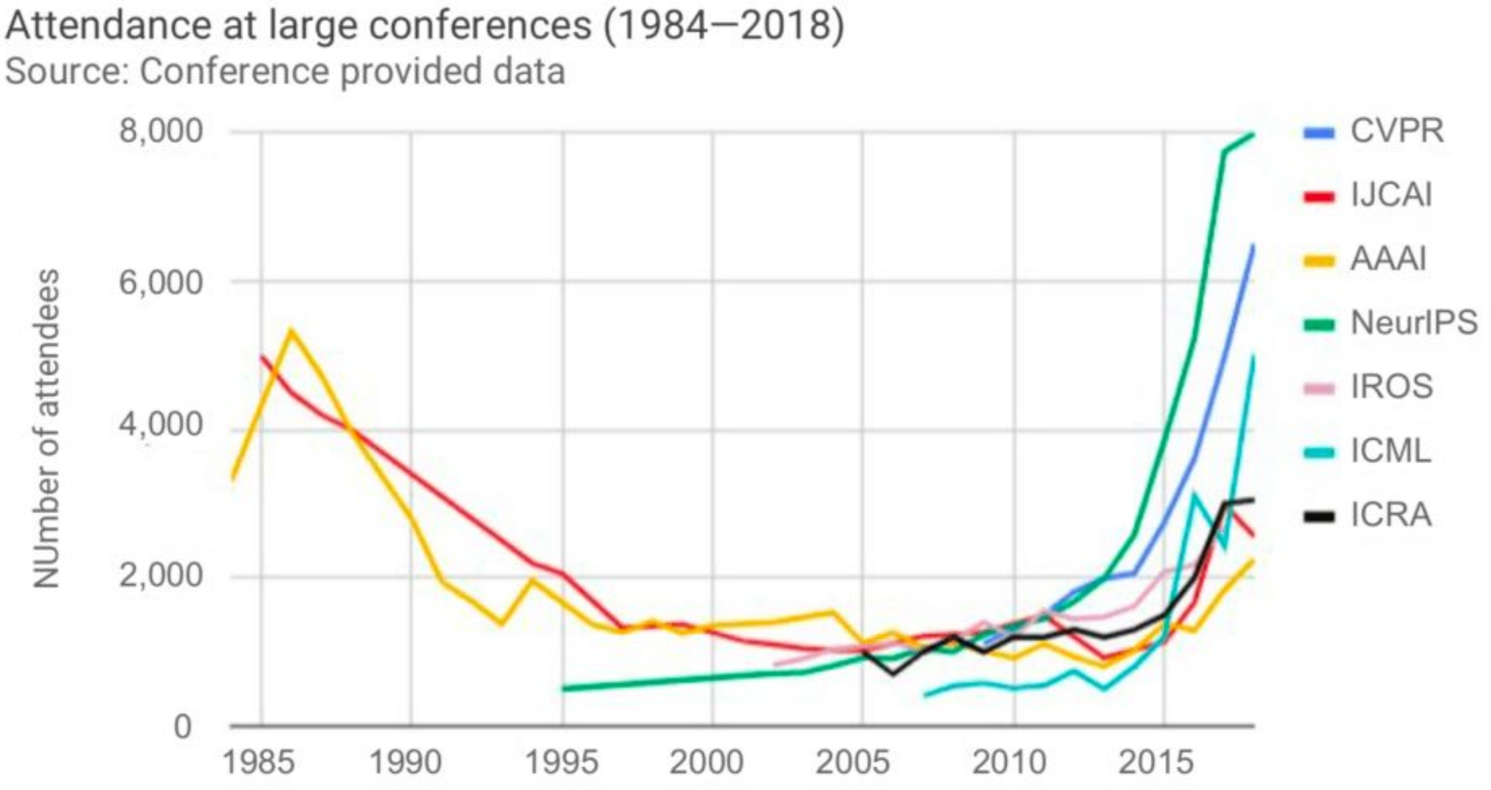
Why is deep learning its own course?





Google search trends for convolutional neural networks




Attendance at large conferences (1984—2018)
Source: Conference provided data
NUmber of attendees
8,000
6,000
4,000
2,000
0
1985
1990
1995
2000
2005
2010
2015
CVPR
IJCAI
AAAI
NeurIPS
IROS
ICML
ICRA

Today's agenda

- A brief history of computer vision
- CSE 493G1 overview

Survey - A show of hands

Undergrad?
M.S.?
Ph.D.?

CSE / EE?
Other Engineering?
Math / Natural Science?
Others?

Instructors		Teaching Assistants			Friday Lecturer
					
Ranjay Krishna	Aditya Kusupati	Sarah Pratt	Aniket Rege	Benlin Liu	Shubhang Desai
Hours: Thursday,	Hours: Tuesday,	Hours: Friday,	Hours: Wednesday,	Hours: Monday	---
11:30am - 12:30pm	11:30am - 1:30pm	1:00pm - 3:00pm	5:00pm - 7:00pm	2:00pm - 4:00pm	---
CSE2 304	CSE2 151	CSE2 274	CSE2 276	CSE2 274	---

Syllabus

Deep learning Fundamentals	Practical training skills	Applications
Data-driven approaches	Pytorch 1.4 / Tensorflow 2.0	Image captioning
Linear classification & kNN	Activation functions	Interpreting machine learning
Loss functions	Batch normalization	Generative AI
Optimization	Transfer learning	Fairness & ethics
Backpropagation	Data augmentation	Data-centric AI
Multi-layer perceptrons	Momentum / RMSProp / Adam	Deep reinforcement learning
Neural Networks	Architecture design	Self-supervised learning
Convolutions		Diffusion
RNNs / LSTMs		LLMs
Transformers		

Lectures

In person in Gates building: CSE2 G01

- Zoom links and recordings will be shared via canvas:
  - Due to security reasons, please do not share zoom links publicly
- **Tuesdays** and **Thursdays** between **10am to 11:20am**
  - To watch the lectures later, you must login to canvas. We highly recommend coming in person
- Slides posted to our website:
  - https://courses.cs.washington.edu/courses/cse493g1/23sp/

Friday recitation sections

Fridays TBD

Hands-on concepts, some tutorials, more practical details than tuesday/thursday lectures

Check the syllabus page for more information on what is going to be covered when.

This Friday: Python / numpy / Google Cloud (Presenter: Sarah Pratt)

EdStem discussions

For questions about assignments, midterm, projects, logistics, etc, use EdStem!

SCPD students: Use your @uw.edu address to register for EdStem;

Office Hours

See course webpage for schedule.

- Add your name to a queue when you arrive for a particular office hours
- TAs will usually conduct 1-1 conversations in front of the whole group unless otherwise requested for a private conversation.

Optional textbook resources

- *Deep Learning*
  - by Goodfellow, Bengio, and Courville
  - Here is a free version
- Mathematics of deep learning
  - Chapters 5, 6 7 are useful to understand vector calculus and continuous optimization
  - Free online version
- Dive into deep learning
  - An interactive deep learning book with code, math, and discussions, based on the NumPy interface.
  - Free online version

Grading

All assignments, coding and written portions, will be submitted via Gradescope.

We use an **auto-grading system**

- A consistent grading scheme,
- Public tests:
  - Students see results of public tests immediately
- Private tests
  - Generalizations of the public tests to thoroughly test your implementation

Grading

3 Problem Sets: 10% + 20% + 20% = 45%

Take home 24hr Midterm Exam: 15%

Course Project: 35%

- Project Proposal: 5%
- Milestone: 5%
- Final report: 15%
- Poster presentation: 10%

Participation **Extra Credit** in lectures: up to 5%

Grading

Late policy

- 2 free late days
- Afterwards, 25% off per day late
- No late days for project report
- Weekends count as 1 day. So using 1 late day for a Friday 11:50pm deadline means you can submit by Sunday 11:59pm

Overview on communication

Course Website: https://courses.cs.washington.edu/courses/cse493g1/23sp/

- Syllabus, lecture slides, links to assignment downloads, etc

EdStem:

- Use this for most communication with course staff
- Ask questions about assignments, grading, logistics, etc
- Use private questions if you want to post code

Gradescope:

- For turning in homework and receiving grades

Canvas:

- For watching lecture videos

Assignments

All assignments will be completed using **Google Colab**

- We have a tutorial for how to use Google Colab on the website

Assignment 1: IS OUT!!!, due 4/14 by 11:59pm

- K-Nearest Neighbor
- Linear classifiers: SVM, Softmax
- Two-layer neural network
- Image features

Pre-requisite

Proficiency in Python

- All class assignments will be in Python (and use numpy)
- Later in the class, you will be using Pytorch and TensorFlow
- We will go over a Python tutorial on this Friday’s recitation.

College Calculus, Linear Algebra

No longer need Machine Learning as a prerequisite

Collaboration policy

Please follow UW student code of conduct – read it!

Here are our course specific rules:

- **Rule 1**: Don't look at solutions or code that are not your own; everything you submit should be your own work. We have automatic tools that detect plagiarism.
- **Rule 2**: Don't share your solution code with others; however discussing ideas or general strategies is fine and encouraged.
- **Rule 3**: Indicate in your submissions anyone you worked with.

Turning in something late / incomplete is better than violating the code

Learning objectives

Formalize deep learning applications into tasks

- Formalize inputs and outputs for vision-related problems
- Understand what data and computational requirements you need to train a model

Develop and train deep learning models

- Learn to code, debug, and train convolutional neural networks.
- Learn how to use software frameworks like TensorFlow and PyTorch

Gain an understanding of where the field is and where it is headed

- What new research has come out in the last 0-9 years
- What are open research challenges?
- What ethical and societal considerations should we consider before deployment?

What you should expect from us

Fun: We will discuss fun applications like image captioning, GPT, generative AI

What we expect from you

Patience.

- This is new for us as much as it is new for you
- Things will break; we will experience technical difficulties
- Bear with us and trust us to listen to you

Contribute

- Build a community with your peers
- Help one another - discuss topics you enjoy
- Give us (annonymous) feedback

Why should you take this class?

Become a deep learning researcher (an incomplete list of conferences)

- Get involved with research at UW: apply using this form.

Conferences:

- CVPR 2023, ACL 2023, NeurIPS 2023, ICML 2023

Become a deep learning engineer in industry (an incomplete list of industry teams)

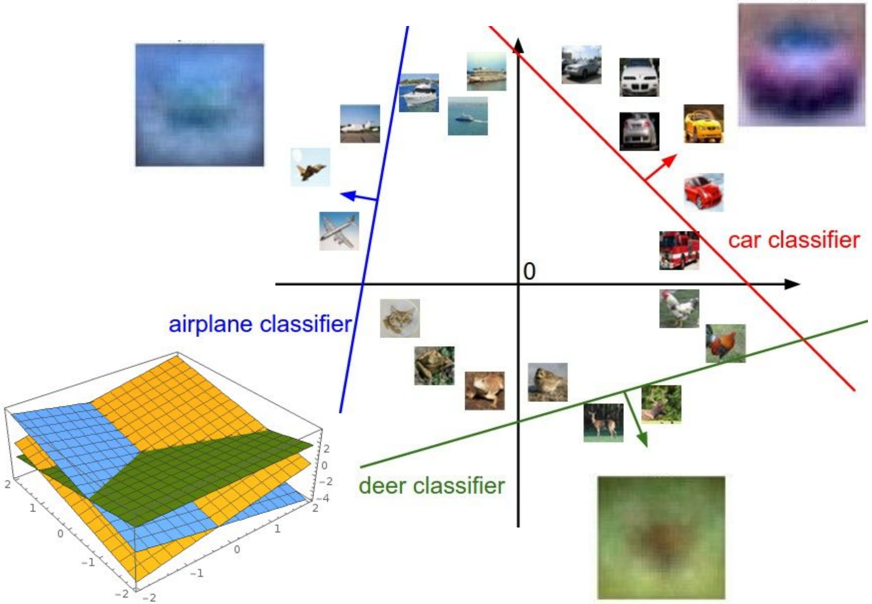
- Brain team at Google AI
- OpenAI
- Meta's Fundamental AI research team
- Microsoft's AI research team

General interest

Next time: Image classification

k- nearest neighbor

Linear classification



Plot created using [Wolfram Cloud]

References

- Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005. [PDF]
- Felzenszwalb, Pedro, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008 [PDF]
- Everingham, Mark, et al. "The pascal visual object classes (VOC) challenge." International Journal of Computer Vision 88.2 (2010): 303-338. [PDF]
- Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009. [PDF]
- Russakovsky, Olga, et al. "Imagenet Large Scale Visual Recognition Challenge." arXiv:1409.0575. [PDF]
- Lin, Yuanqing, et al. "Large-scale image classification: fast feature extraction and SVM training." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011. [PDF]
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012. [PDF]
- Szegedy, Christian, et al. "Going deeper with convolutions." arXiv preprint arXiv:1409.4842 (2014). [PDF]
- Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014). [PDF]
- He, Kaiming, et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." arXiv preprint arXiv:1406.4729 (2014). [PDF]
- LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324. [PDF]
- Fei-Fei, Li, et al. "What do we perceive in a glance of a real-world scene?." Journal of vision 7.1 (2007): 10. [PDF]

References

- Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005. [PDF]
- Felzenszwalb, Pedro, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008 [PDF]
- Everingham, Mark, et al. "The pascal visual object classes (VOC) challenge." International Journal of Computer Vision 88.2 (2010): 303-338. [PDF]
- Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009. [PDF]
- Russakovsky, Olga, et al. "Imagenet Large Scale Visual Recognition Challenge." arXiv:1409.0575. [PDF]
- Lin, Yuanqing, et al. "Large-scale image classification: fast feature extraction and SVM training." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011. [PDF]
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012. [PDF]
- Szegedy, Christian, et al. "Going deeper with convolutions." arXiv preprint arXiv:1409.4842 (2014). [PDF]
- Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014). [PDF]
- He, Kaiming, et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." arXiv preprint arXiv:1406.4729 (2014). [PDF]
- LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324. [PDF]
- Fei-Fei, Li, et al. "What do we perceive in a glance of a real-world scene?." Journal of vision 7.1 (2007): 10. [PDF]