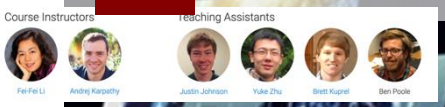
CS231n: Deep Learning for Computer Vision

Lecture 1: Introduction

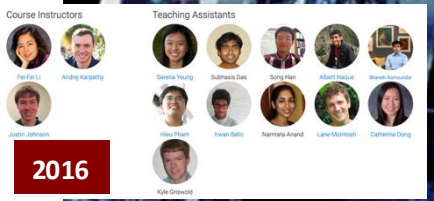
Welcome to CS231n

Welcome to CS231n

2015

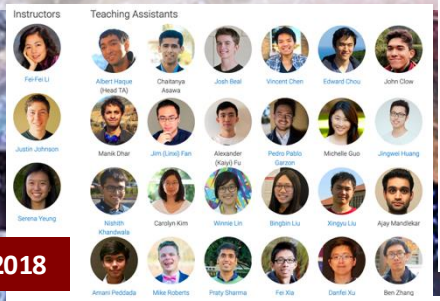


2016

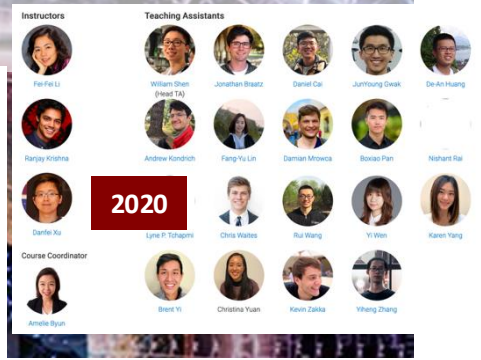


2017

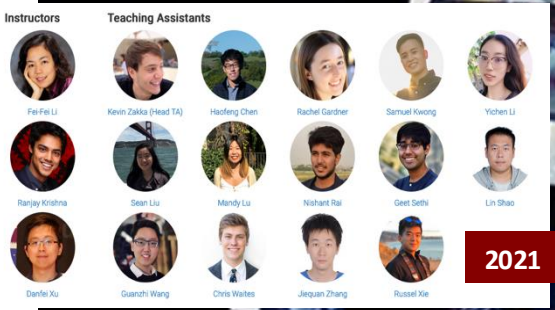
2018



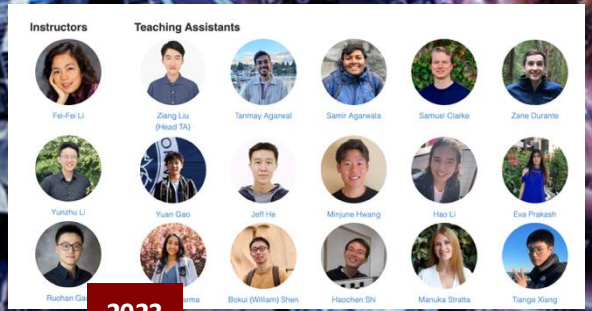
2020



2021



2023



2025

CS231n 10th Anniversary

CS231n – 2026 Edition

Fei-Fei Li

Ehsan Adeli

Zane Durante

Justin Johnson

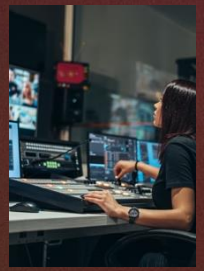
Tiange Xiang

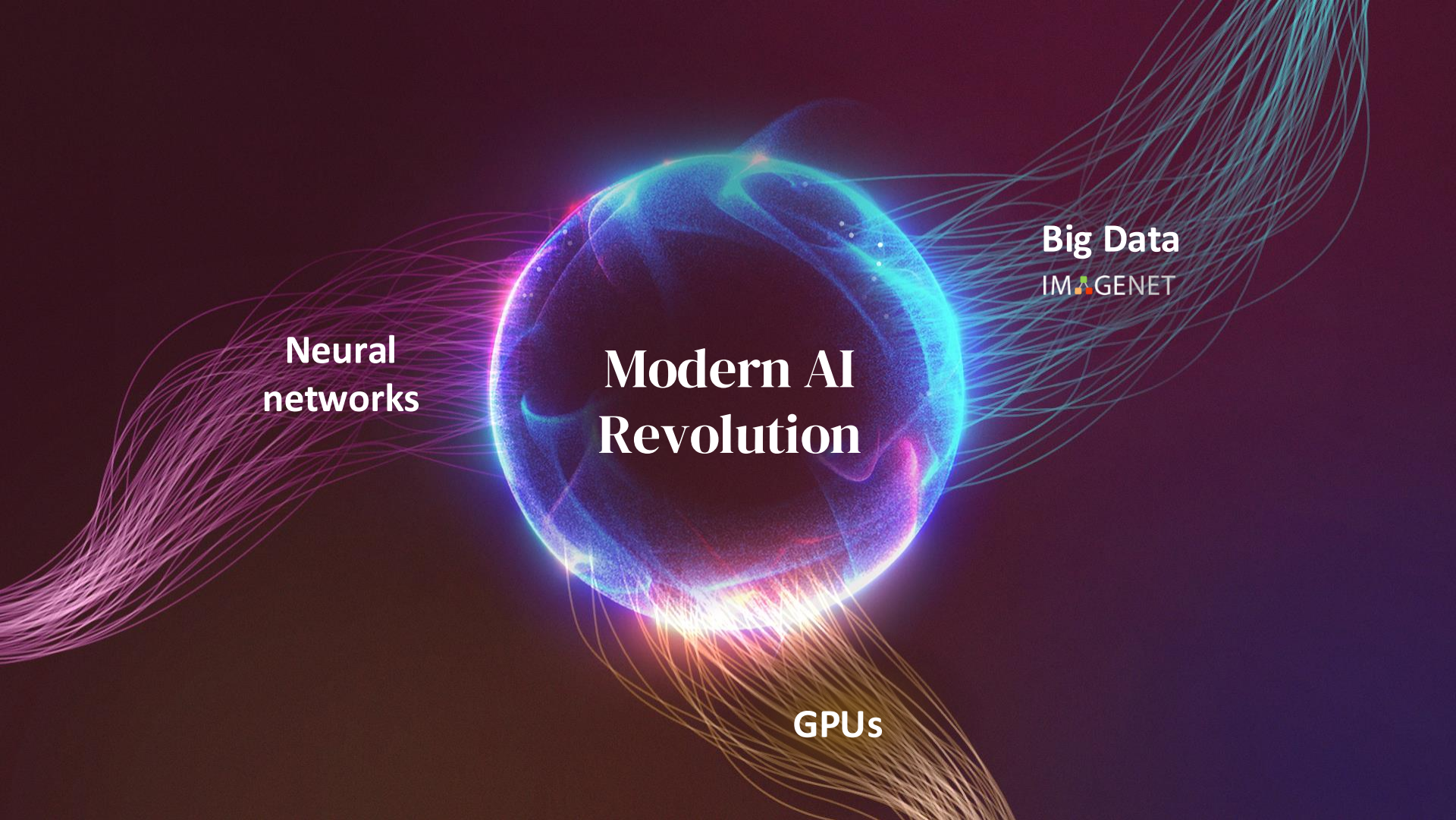


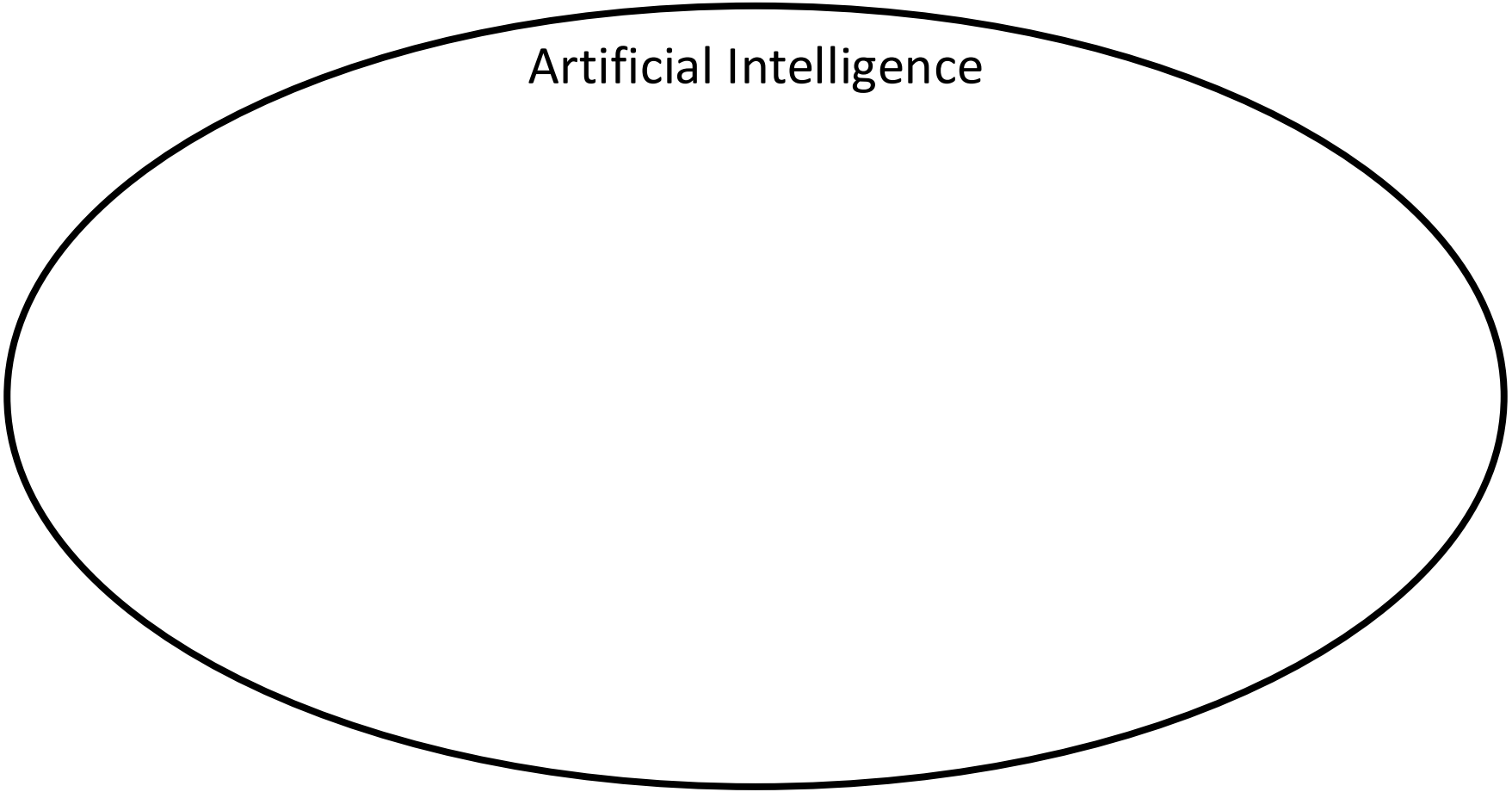




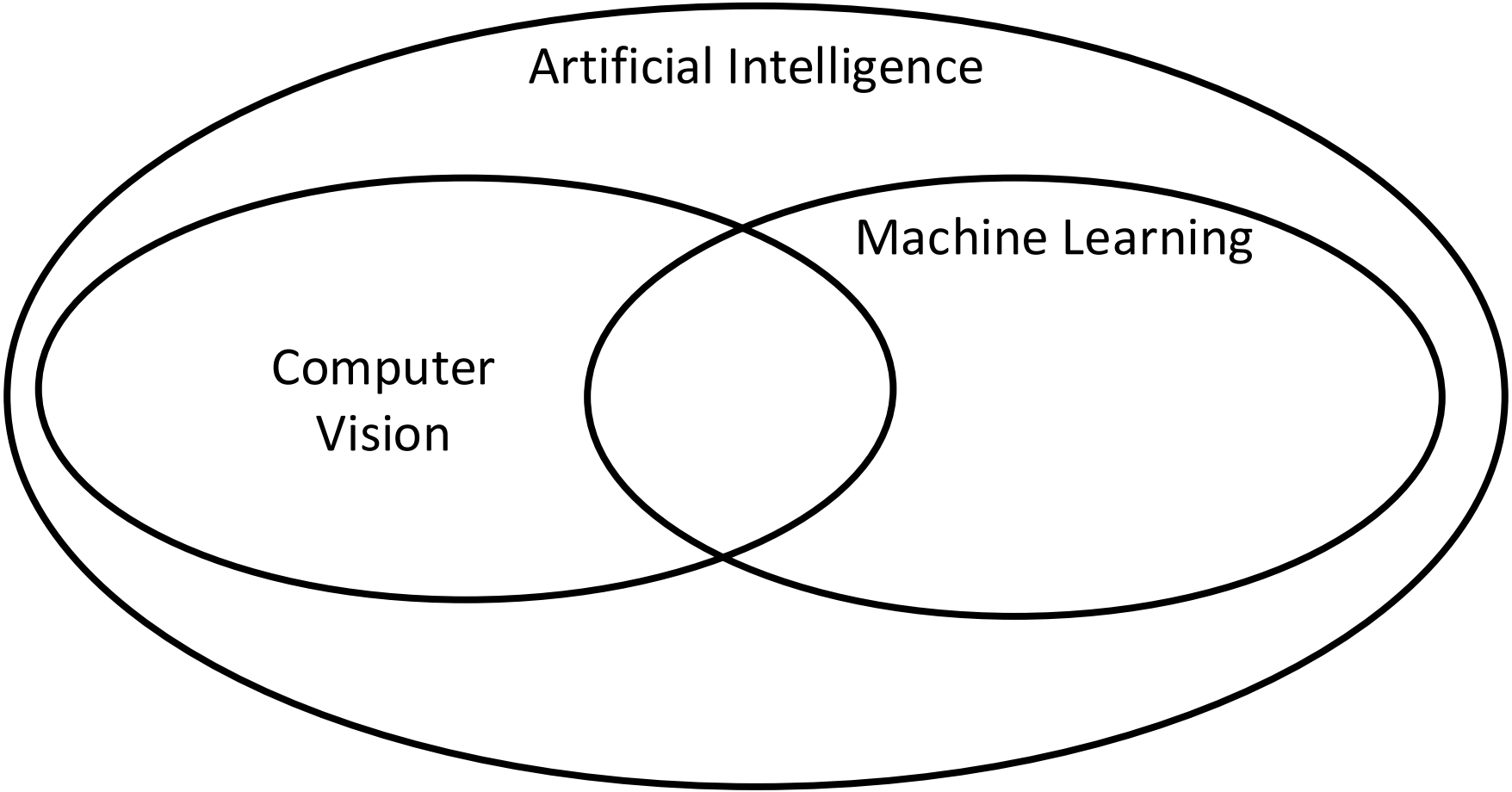


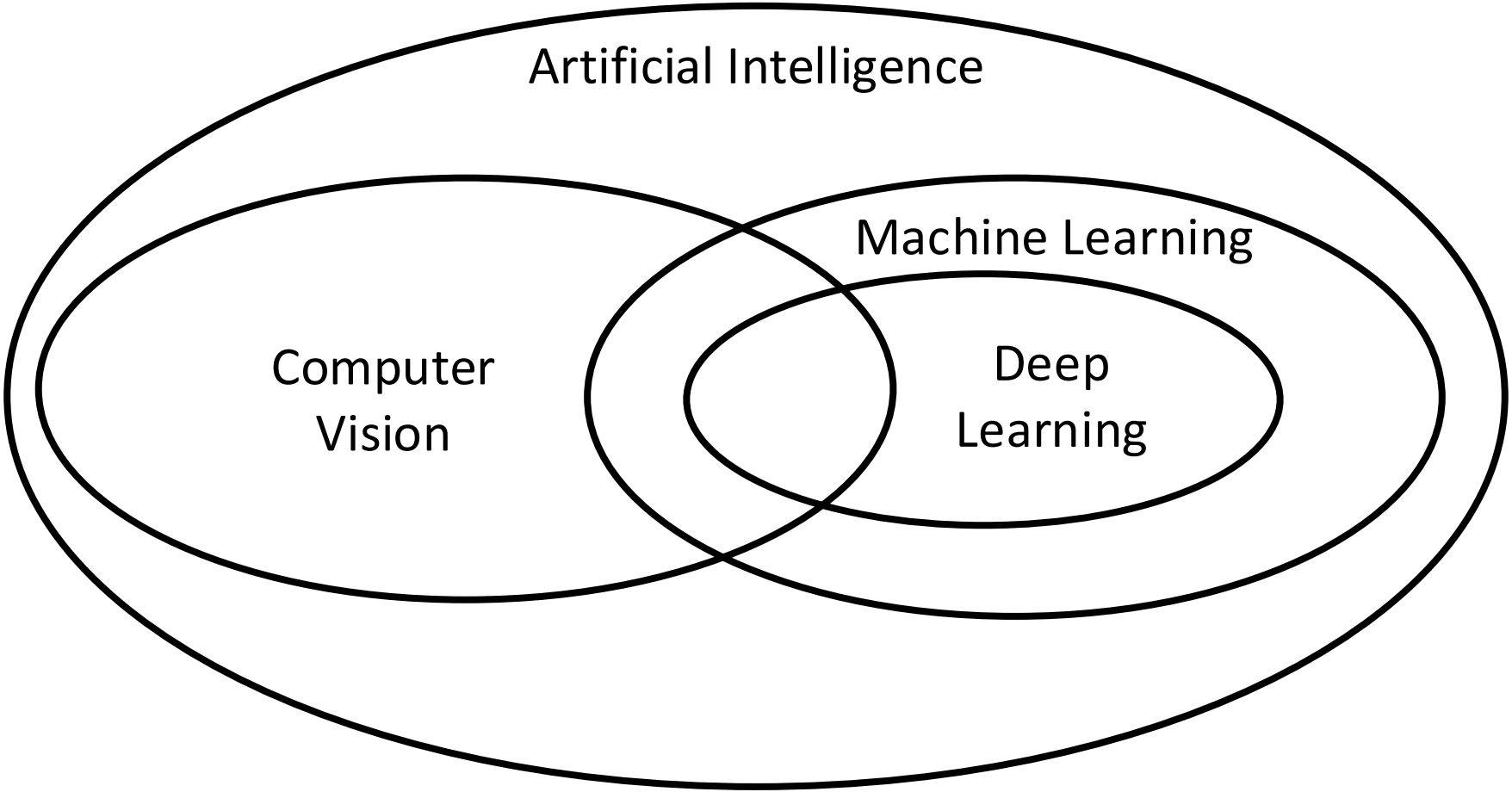

Neural networks
Modern AI Revolution
Big Data
IMAGENET
GPUs



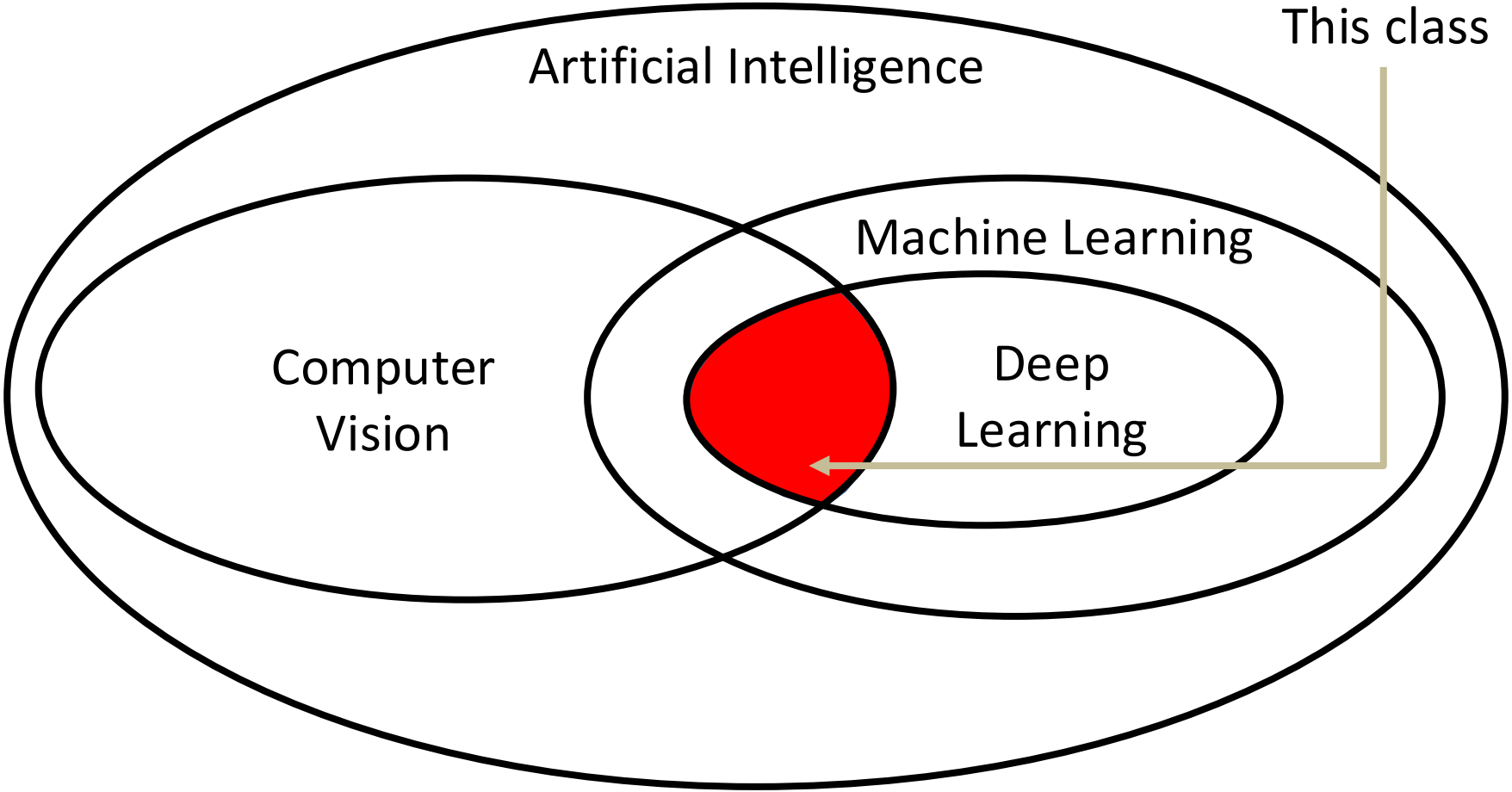
Slide inspiration: Justin Johnson



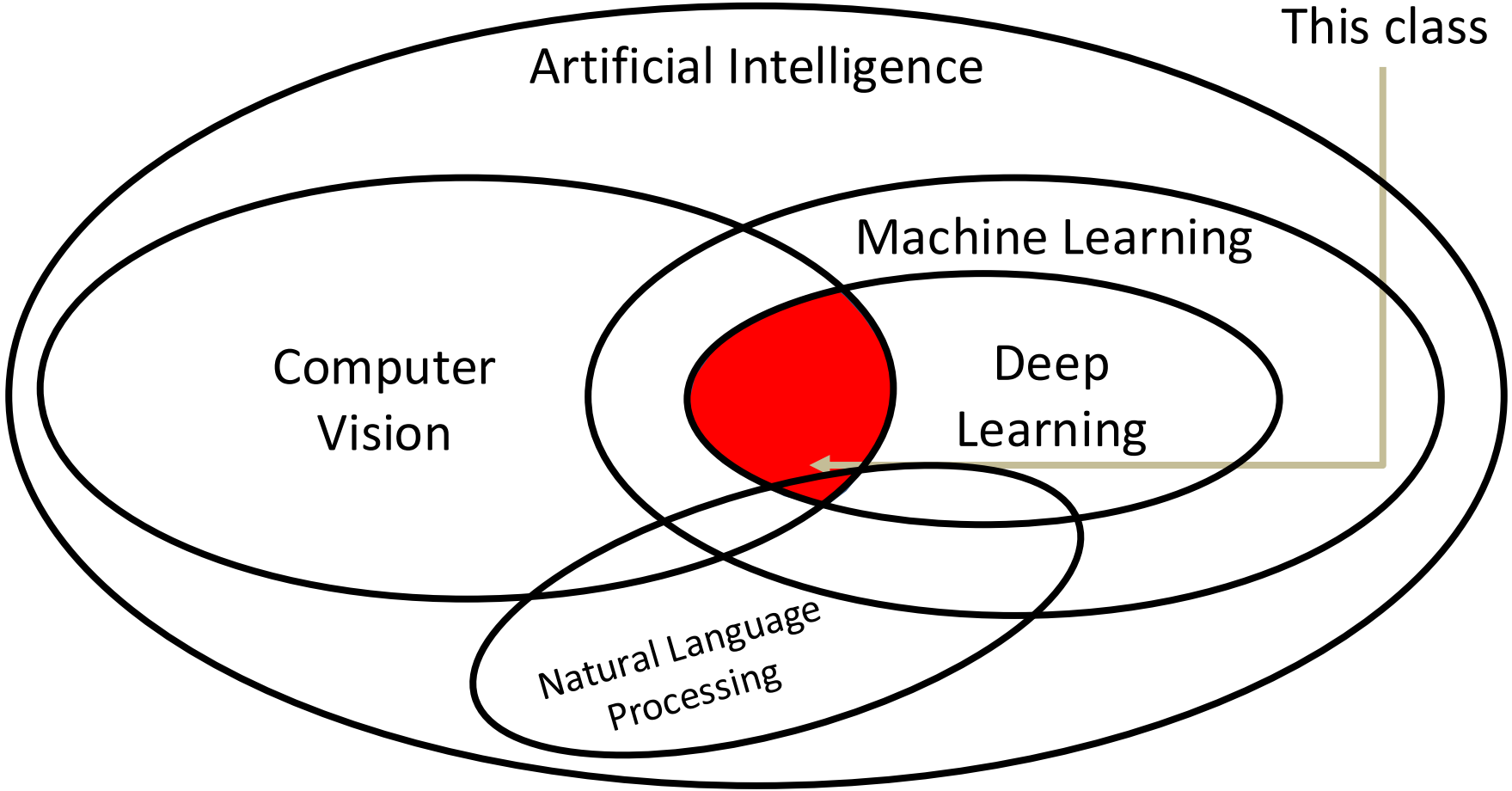
Slide inspiration: Justin Johnson



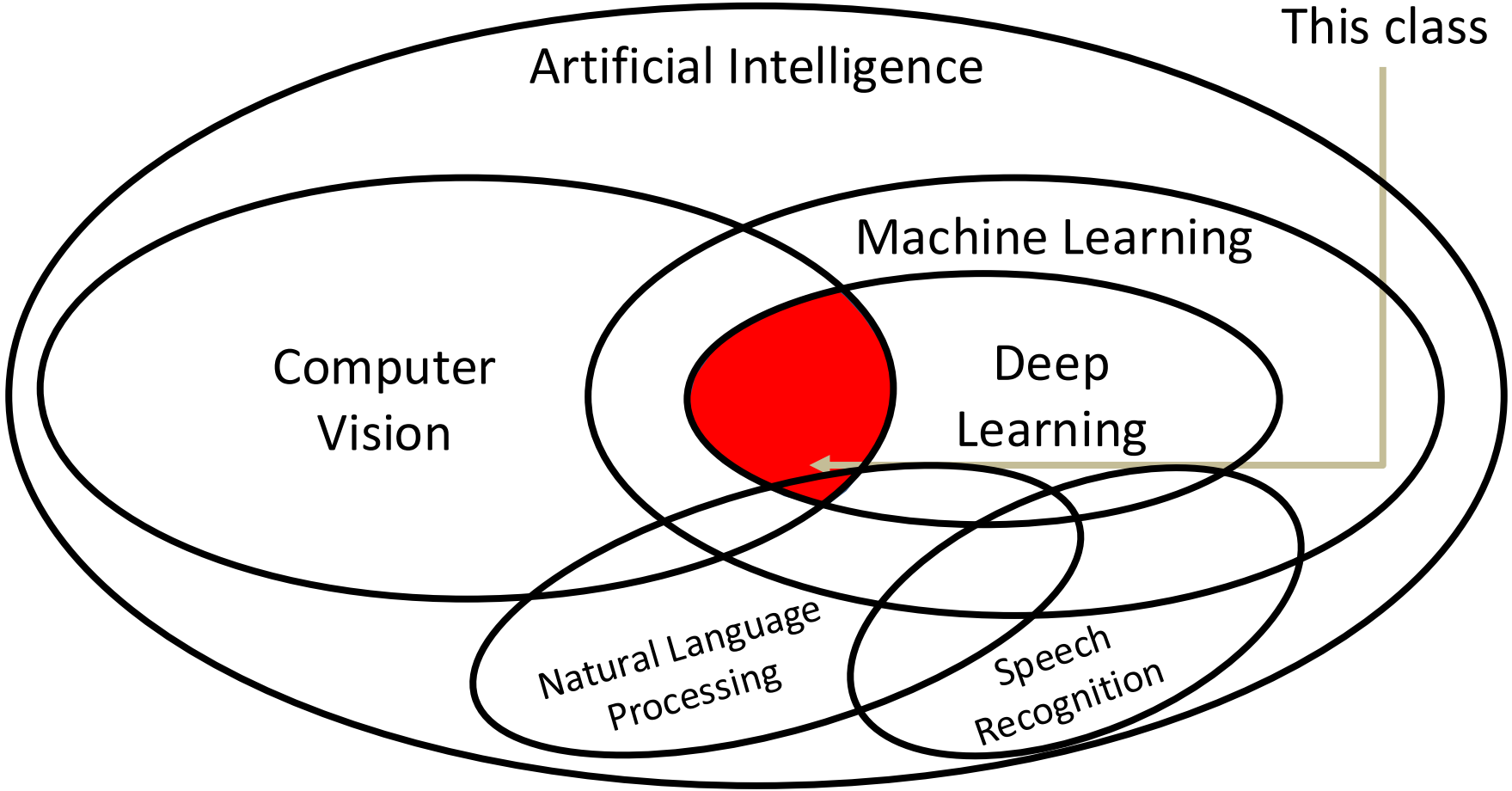
Slide inspiration: Justin Johnson



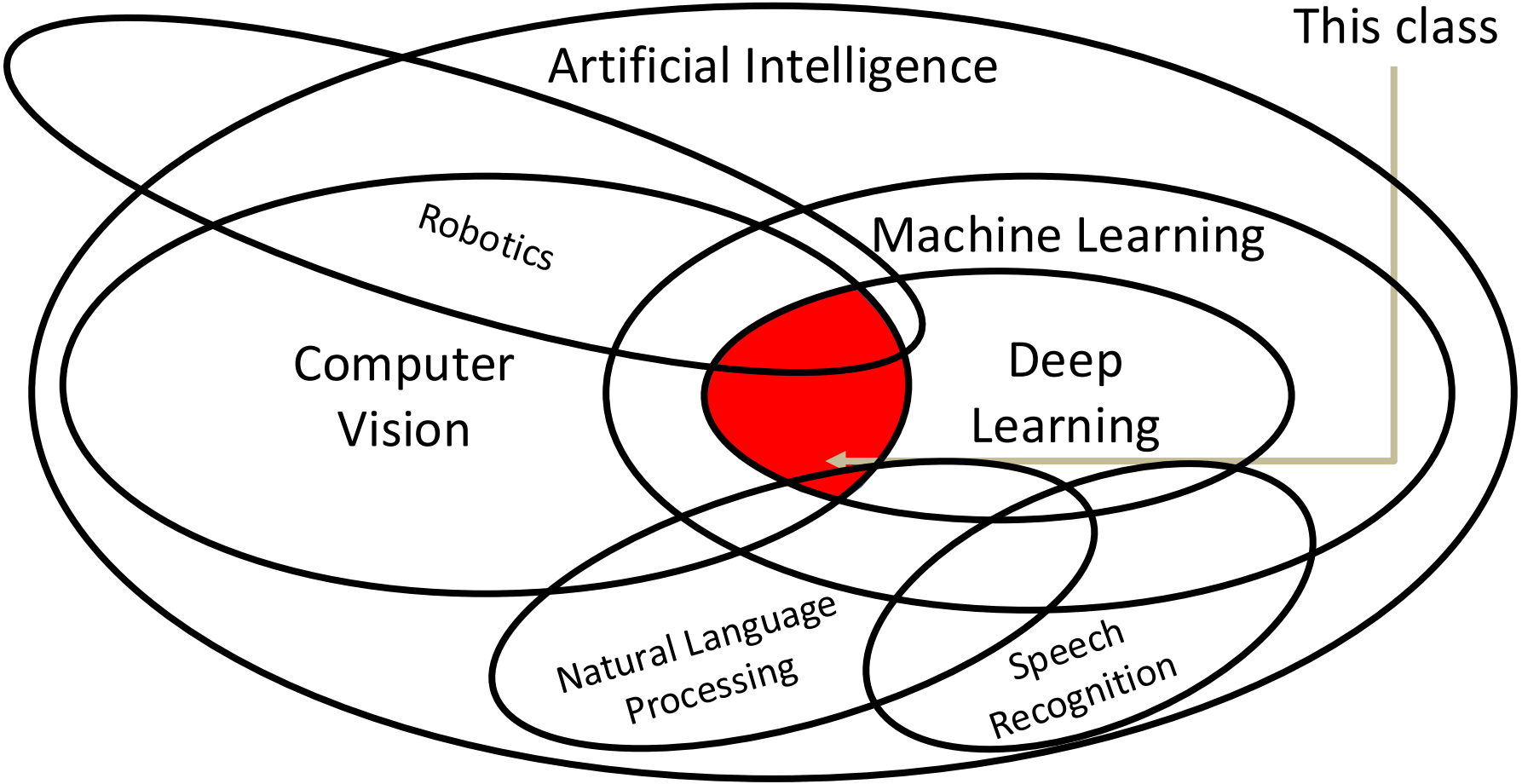
Slide inspiration: Justin Johnson



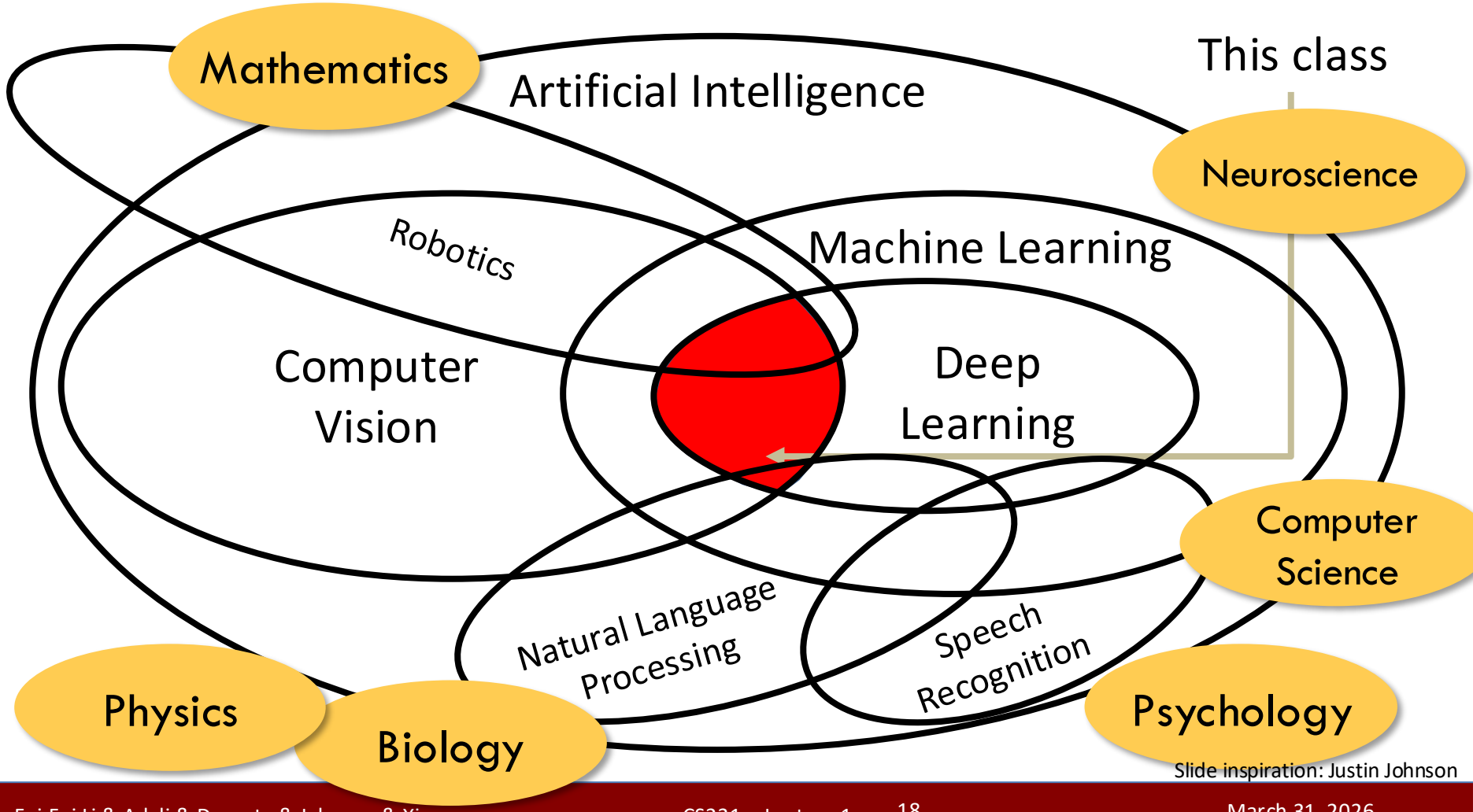
Slide inspiration: Justin Johnson



Slide inspiration: Justin Johnson



Slide inspiration: Justin Johnson

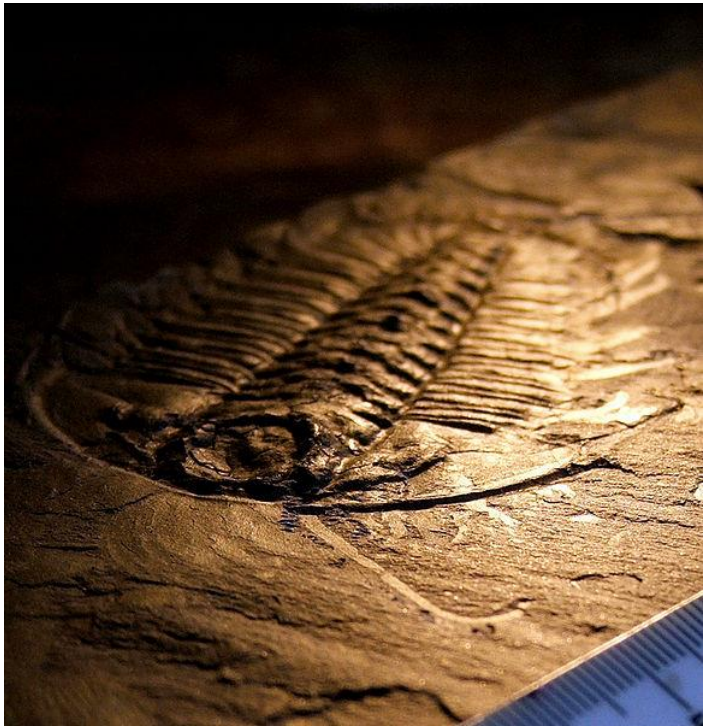


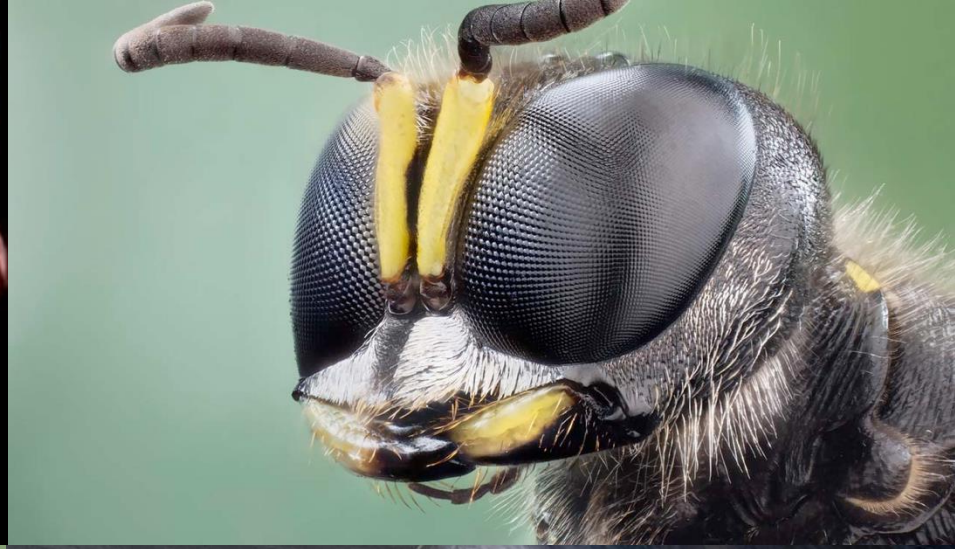
Slide inspiration: Justin Johnson

Today's agenda

- A brief history of computer vision and deep learning
- CS231n overview

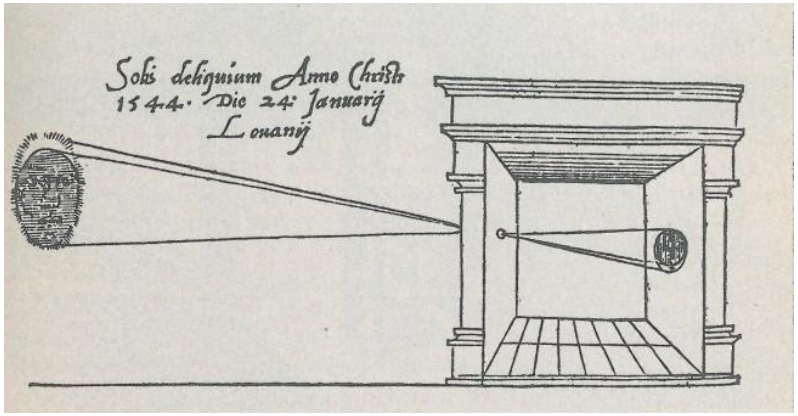
Evolution's Big Bang:
Cambrian Explosion, 530-540million years, B.C.

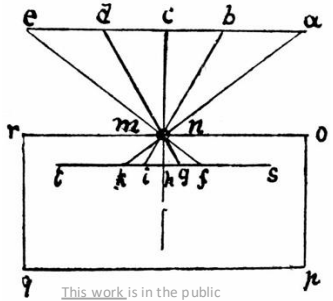




Camera Obscura

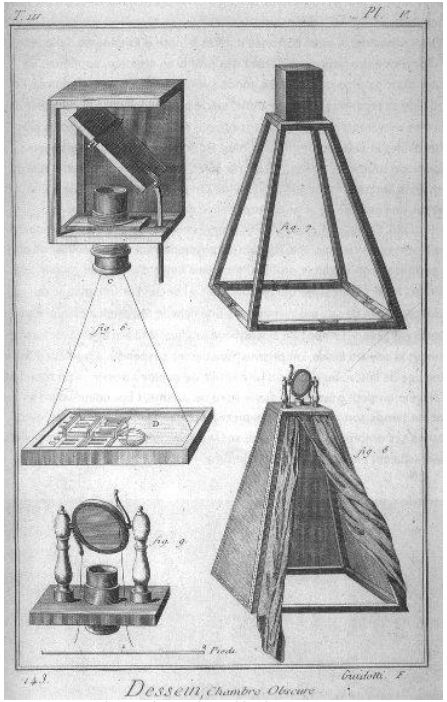
Gemma Frisius, 1545





Leonardo da Vinci, 16th Century AD

Encyclopedia, 18th Century



Computer Vision is everywhere!







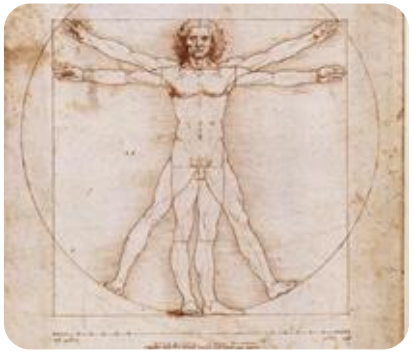
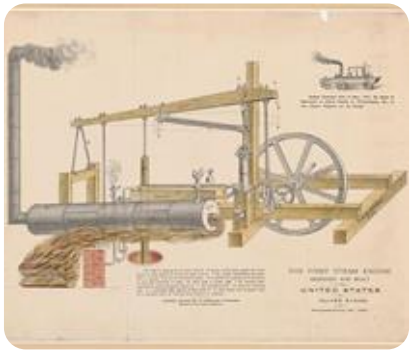



Human development depends on visual intelligence





Humans build civilization using visual intelligence

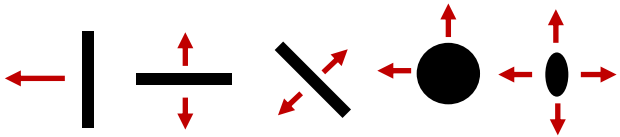




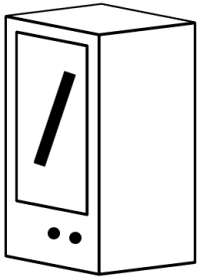


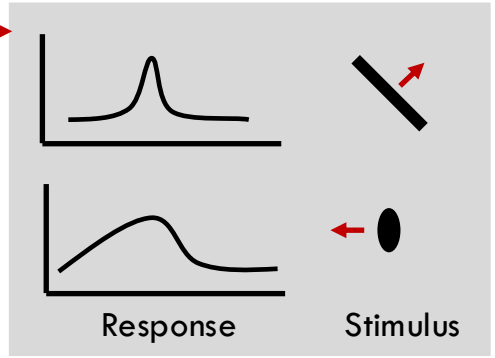

Where did we come from?

Hubel and Wiesel, 1959

Measure brain activity



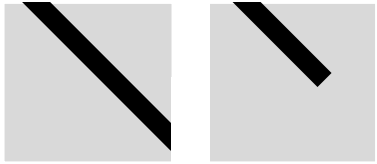




1959
Hubel & Wiesel

Simple cells:
Response to specific rotation and orientation

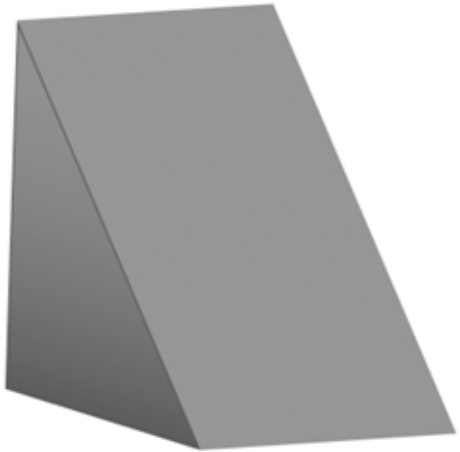
Complex cells:
Response to light orientation and movement, some translation invariance



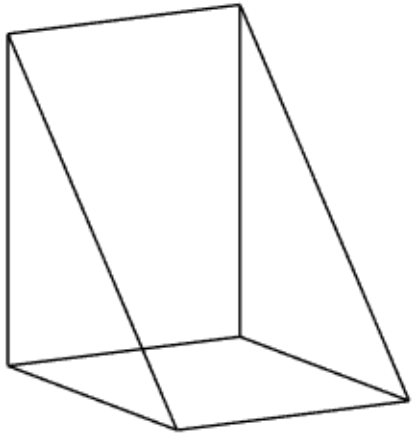
No response

Slide inspiration: Justin Johnson

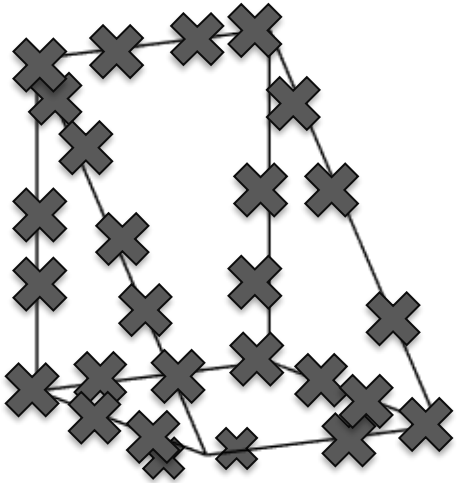
Larry Roberts, 1963



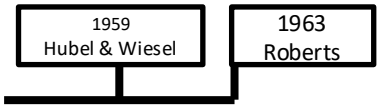
(a) Original picture



(b) Differentiated picture



(c) Feature points selected



Lawrence Gilman Roberts, "Machine Perception of Three-Dimensional Solids", 1963

Slide inspiration: Justin Johnson

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

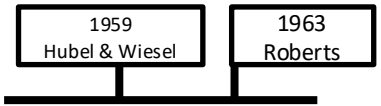
Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

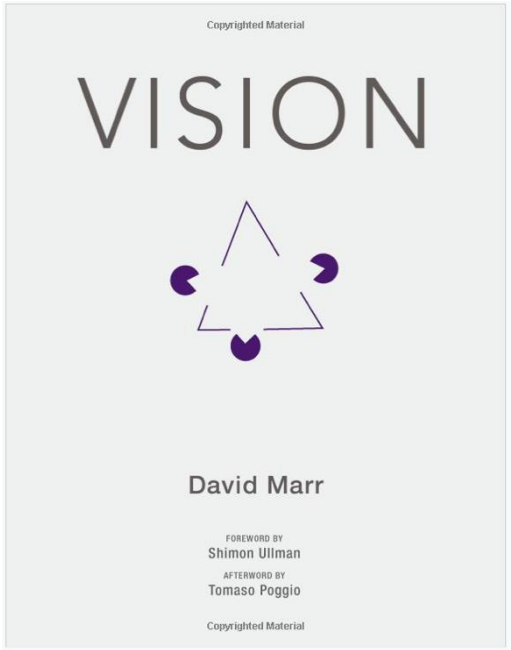
Seymour Papert

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".



https://dspace.mit.edu/handle/1721.1/6125

Slide inspiration: Justin Johnson



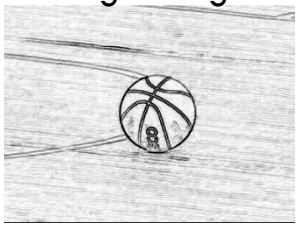
1959 Hubel & Wiesel

1963 Roberts

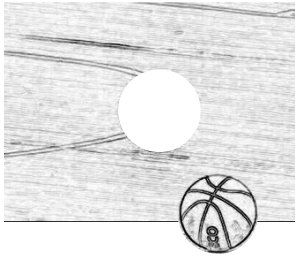
1970s David Marr

Input image

Edge image



2 ½-D sketch



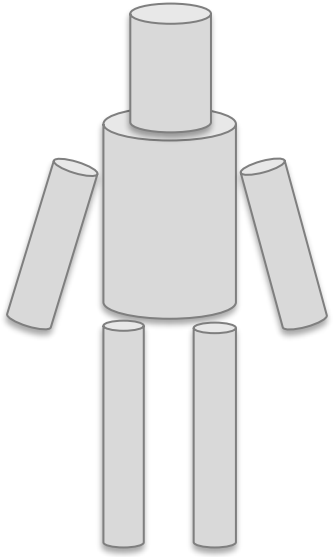
3-D model

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

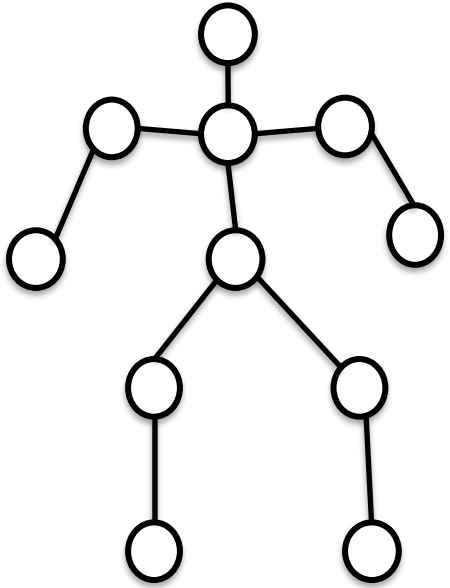
Stages of Visual Representation, David Marr, 1970s

Slide inspiration: Justin Johnson

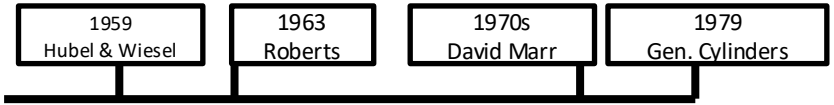
Recognition via Parts (1970s)



Generalized Cylinders,
Brooks and Binford,
1979

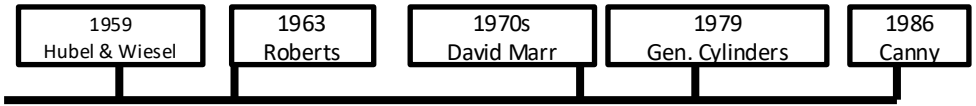


Pictorial Structures,
Fischler and Elshlager, 1973



Slide inspiration: Justin Johnson

Recognition via Edge Detection (1980s)

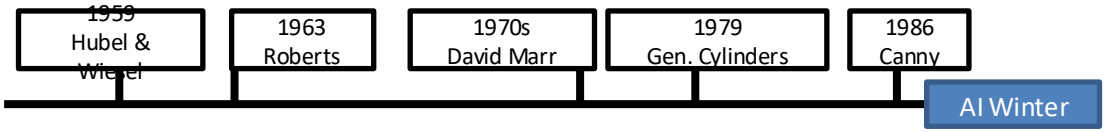


John Canny, 1986
David Lowe, 1987

Slide inspiration: Justin Johnson

Arriving at an “AI winter”

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

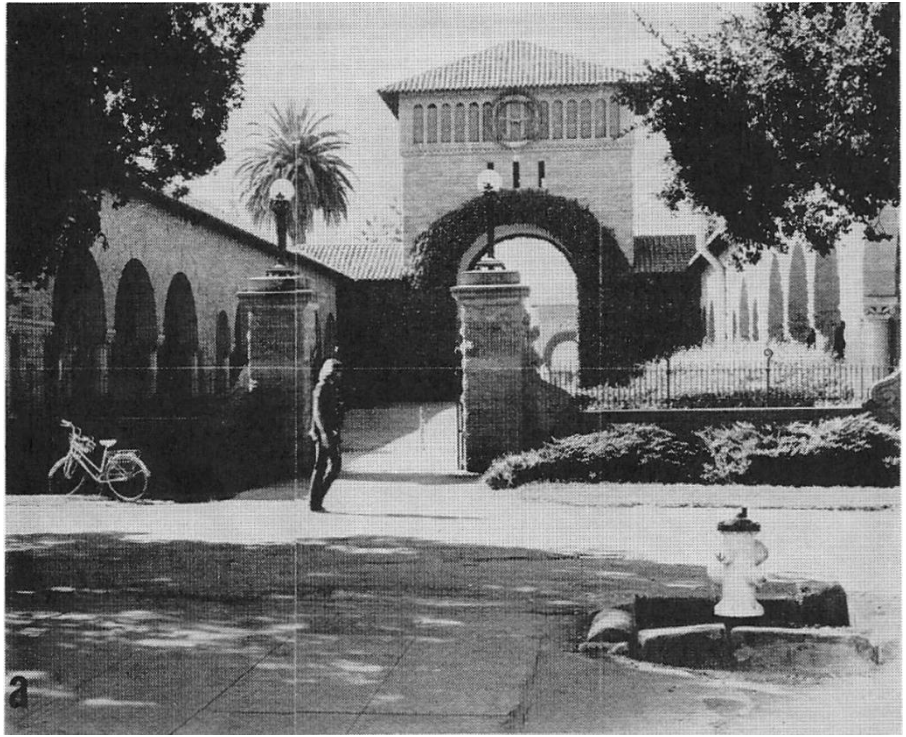


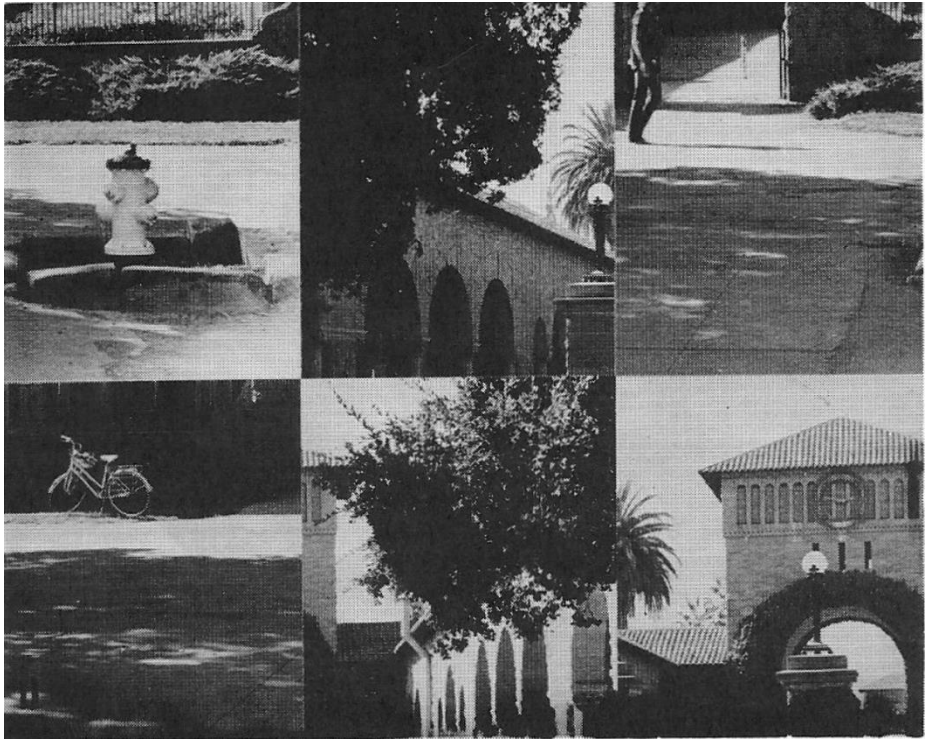
Slide inspiration: Justin Johnson

In the meantime…seminal work in cognitive and neuroscience

Perceiving Real-World Scenes

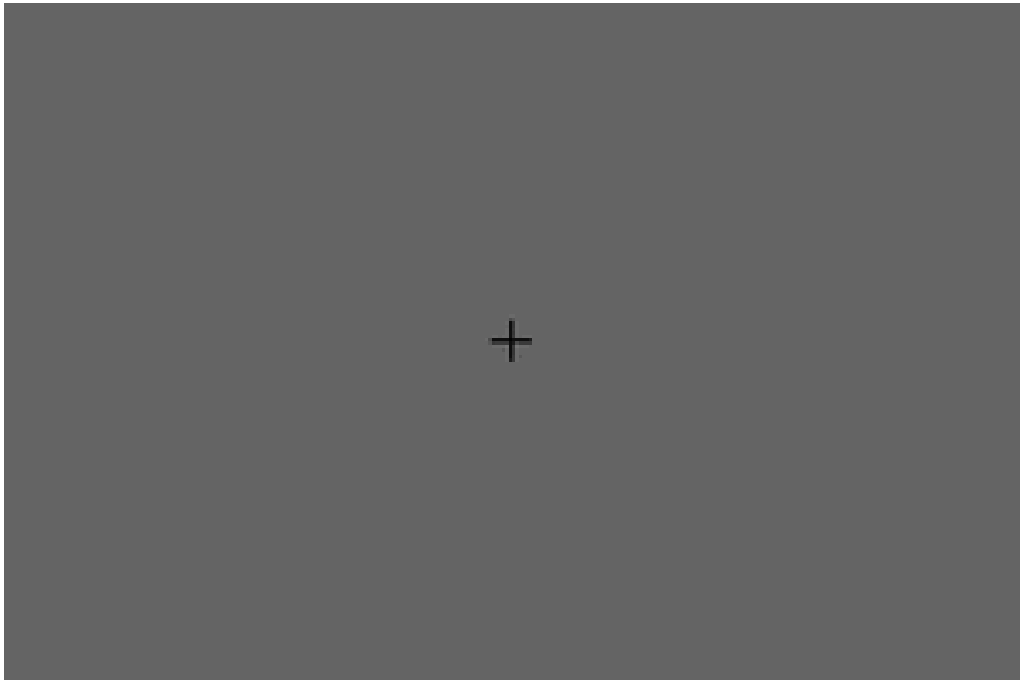
Irving Biederman





I. Biederman, *Science*, 1972

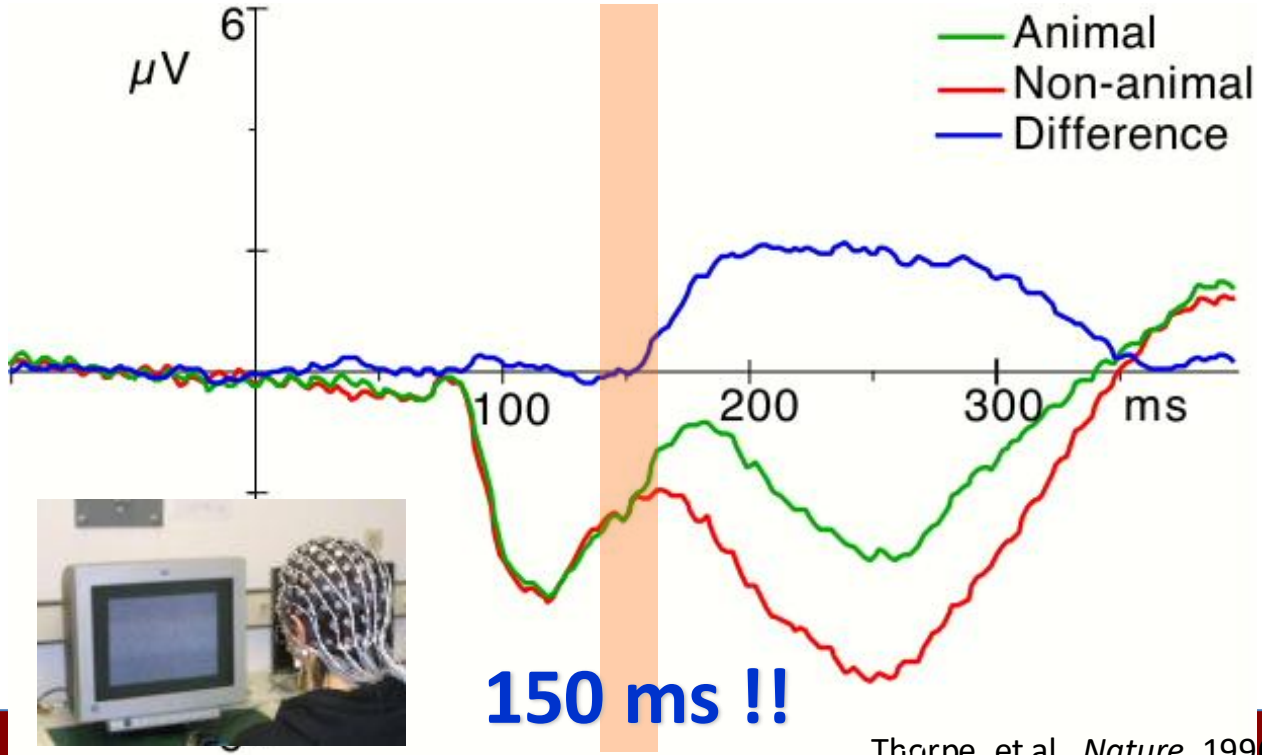
Rapid Serial Visual Perception (RSVP)



Potter, etc. 1970s

Speed of processing in the human visual system

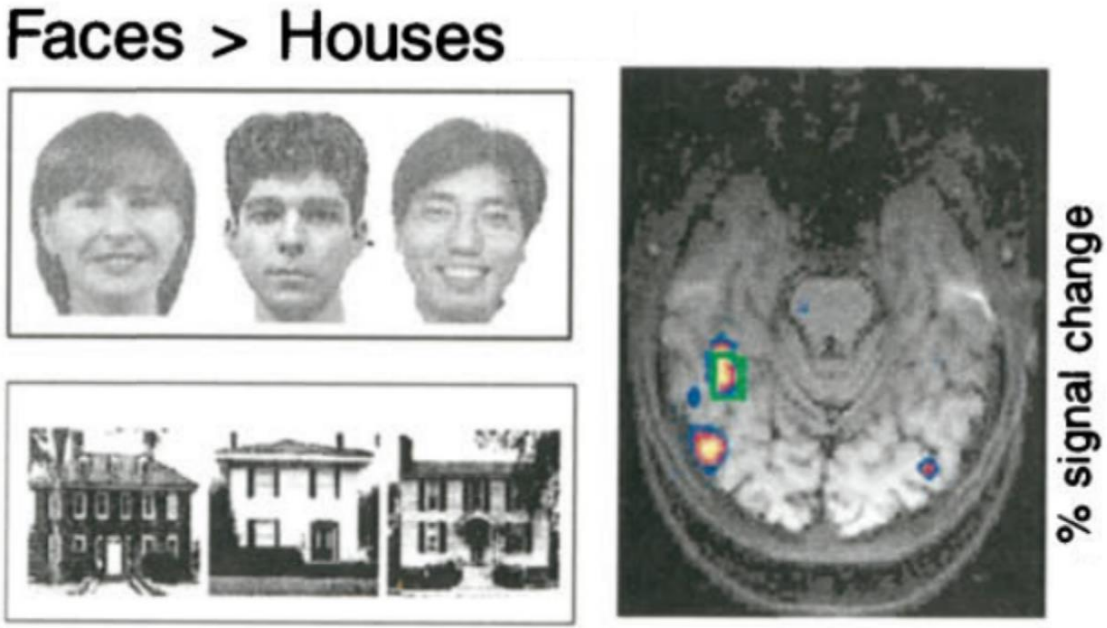
Simon Thorpe, Denis Fize & Catherine Marlot



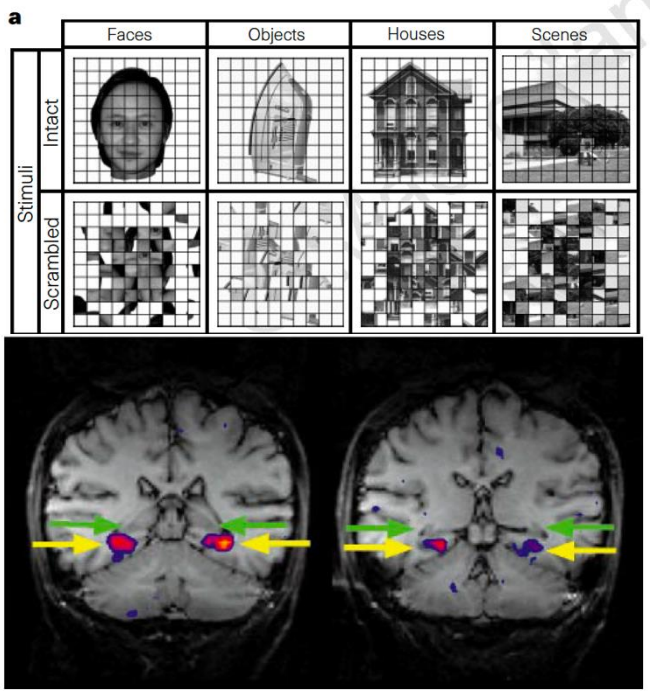
150 ms !!

Thorpe, et al. *Nature*, 1996

Neural correlates of object & scene recognition

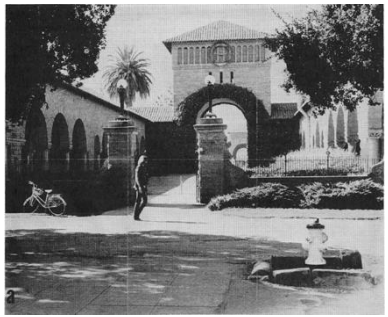


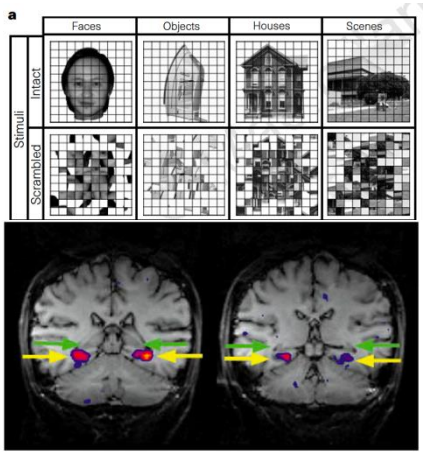
Kanwisher et al. J. Neuro. 1997



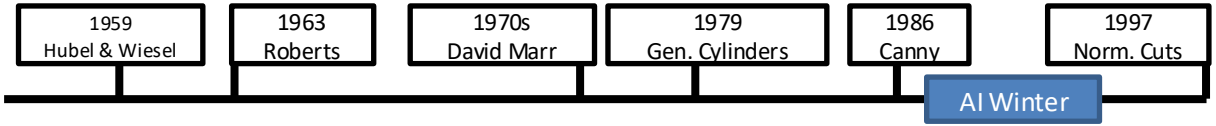
Epstein & Kanwisher, Nature, 1998

Visual recognition is a fundamental task for visual intelligence





Recognition via Grouping (1990s)



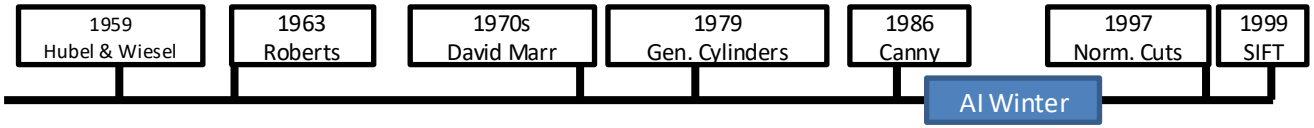
Normalized Cuts, Shi and Malik, 1997

Slide inspiration: Justin Johnson

Recognition via Matching (2000s)

Image is public domain

Image is public domain



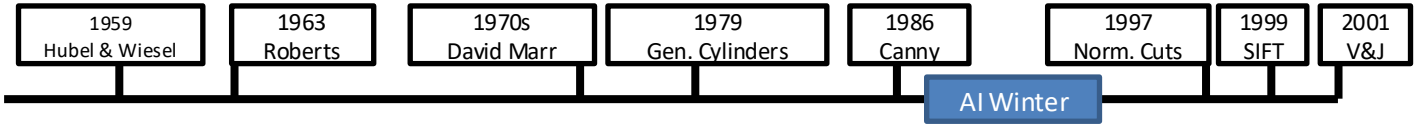
SIFT, David Lowe, 1999

Slide inspiration: Justin Johnson

Face Detection

Viola and Jones, 2001

One of the first successful applications of machine learning to vision



Slide inspiration: Justin Johnson

Caltech 101 images

PASCAL Visual Object Challenge

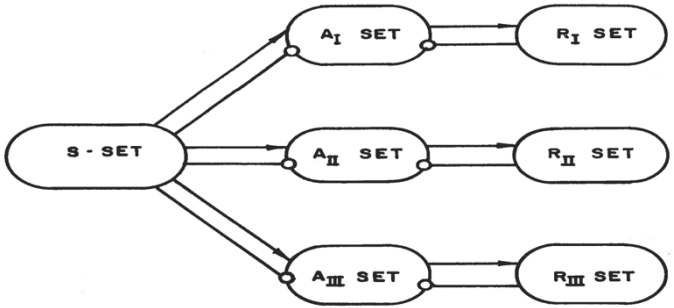
1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
AI Winter
1997 Norm. Cuts
1999 SIFT
2001 V&J
2004, 2007 Caltech101; PASCAL

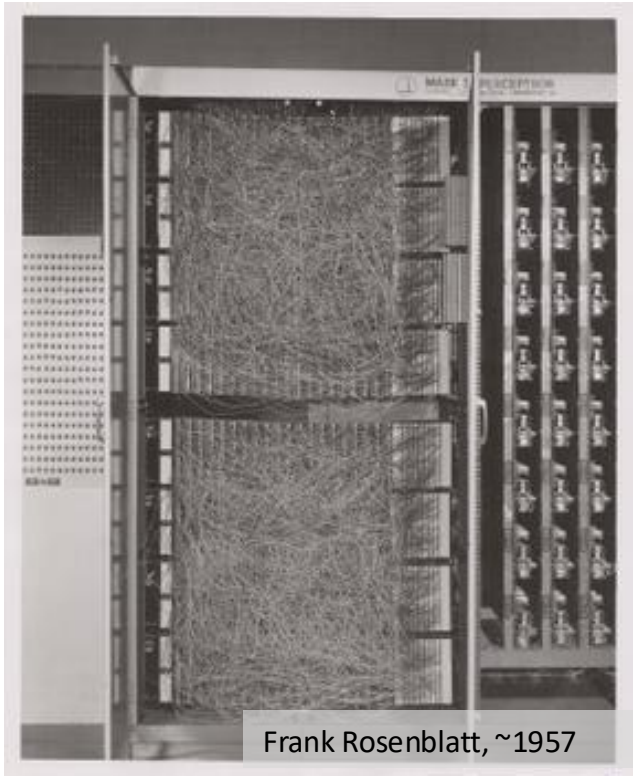
Slide inspiration: Justin Johnson

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA



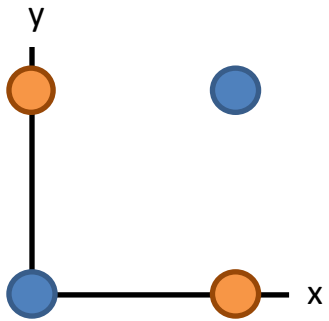

Frank Rosenblatt, ~1957

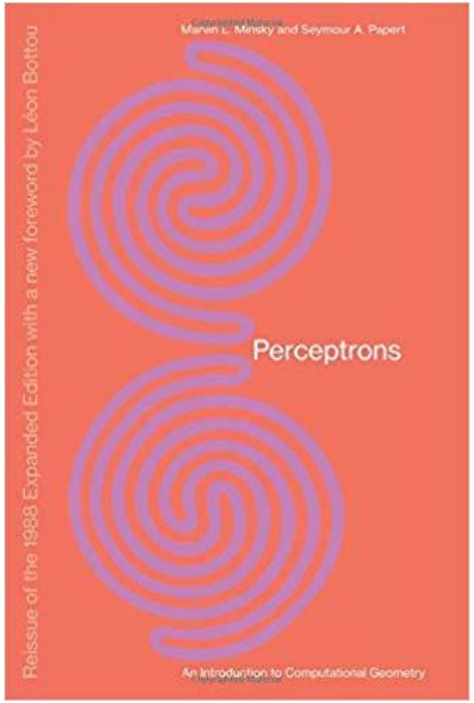
1959 Hubel & Wiesel
1958 Perceptron
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
AI Winter
1997 Norm. Cuts
1999 SIFT
2001 V&J
2004, 2007 Caltech101; PASCAL

Slide inspiration: Justin Johnson

Minsky and Papert, 1969

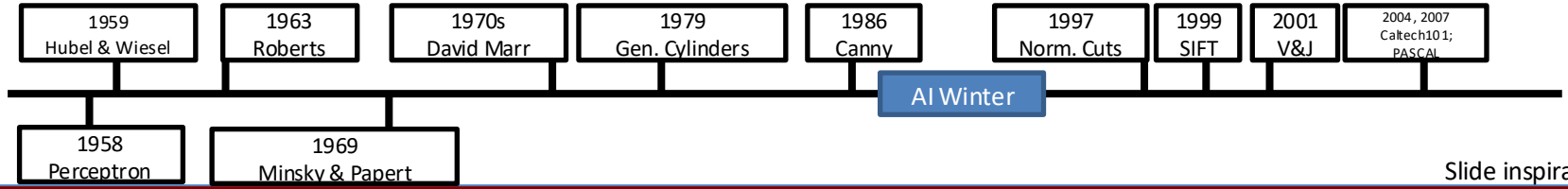
X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0





Showed that Perceptrons could not learn the XOR function

Caused a lot of disillusionment in the field



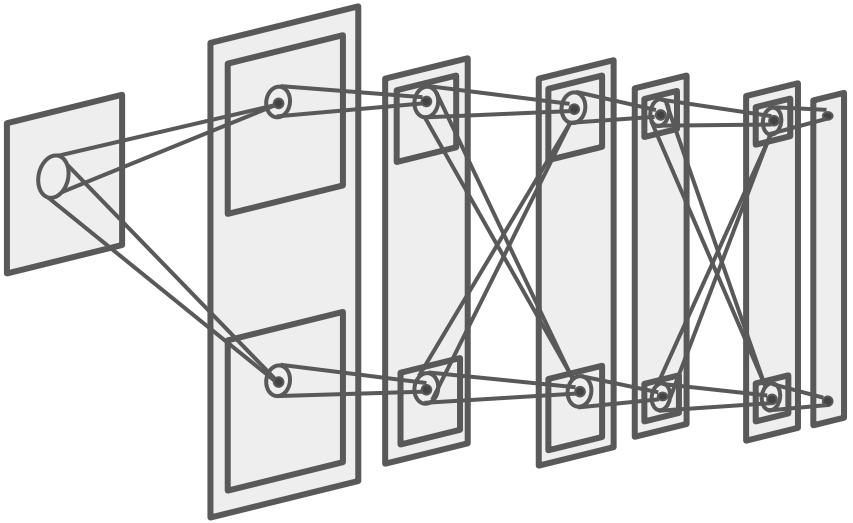
Slide inspiration: Justin Johnson

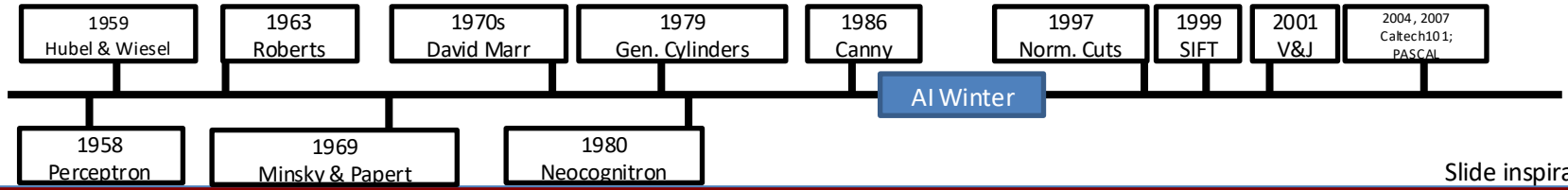
Neocognitron: Fukushima, 1980

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





Slide inspiration: Justin Johnson

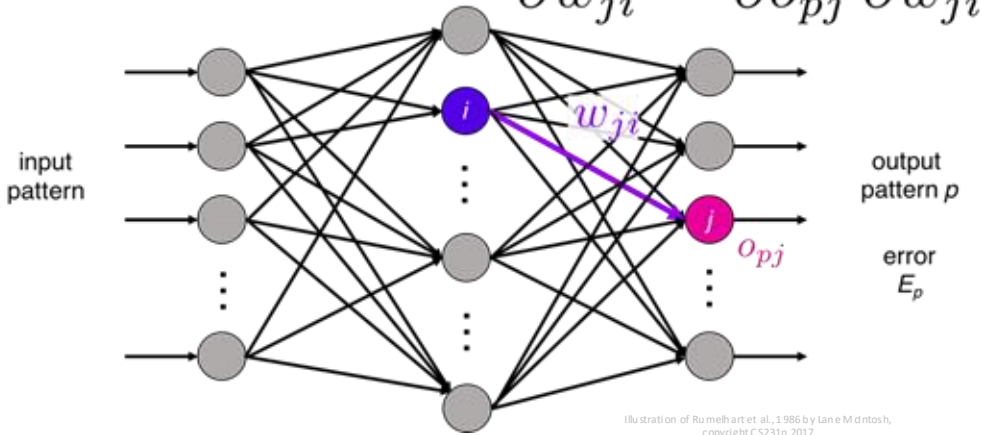
Backprop: Rumelhart, Hinton, and Williams, 1986

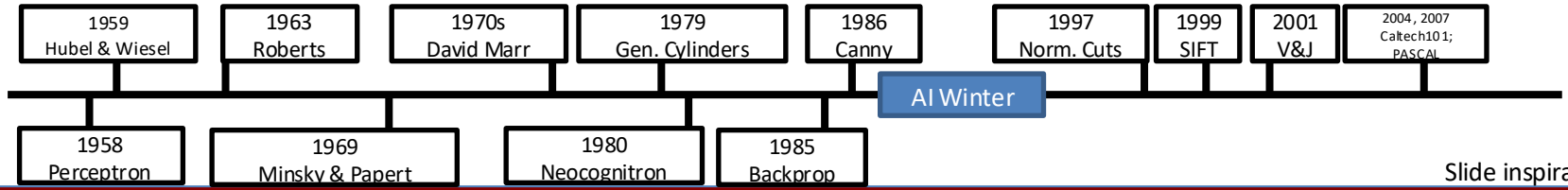
Introduced backpropagation for computing gradients in neural networks

Successfully trained perceptrons with multiple layers

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial w_{ji}}$$

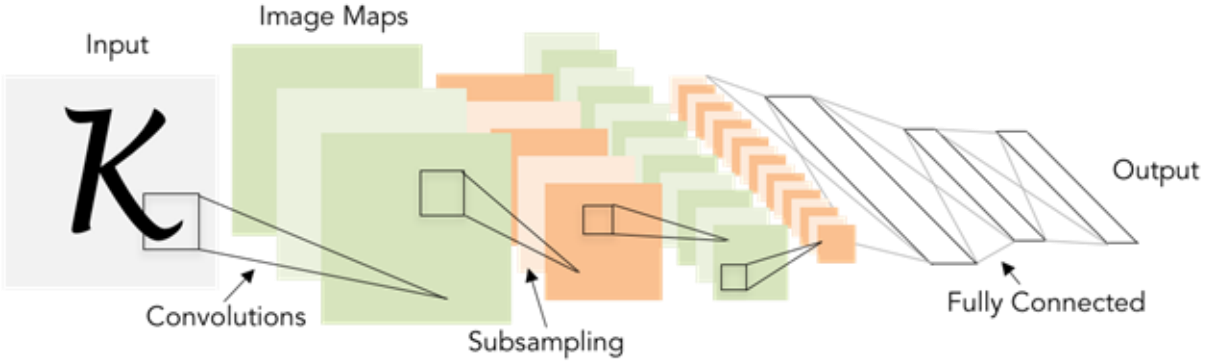
recognizable math



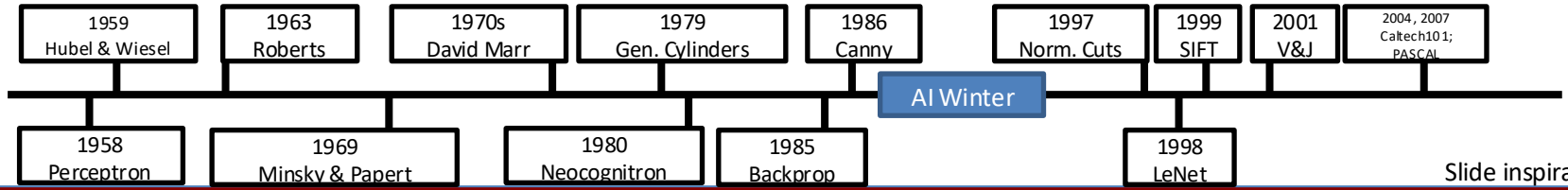


Slide inspiration: Justin Johnson

Convolutional Networks: LeCun et al, 1998



Applied backprop algorithm to a Neocognitron-like architecture
Learned to recognize handwritten digits
Was deployed in a commercial system by NEC, processed handwritten checks
Very similar to our modern convolutional networks!



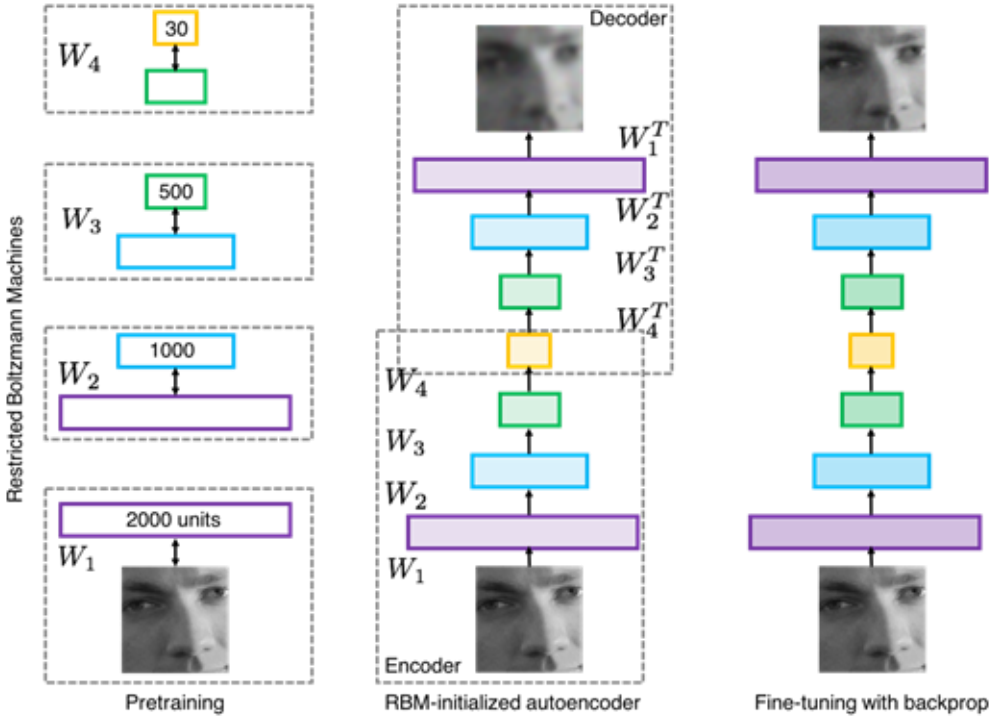
Slide inspiration: Justin Johnson

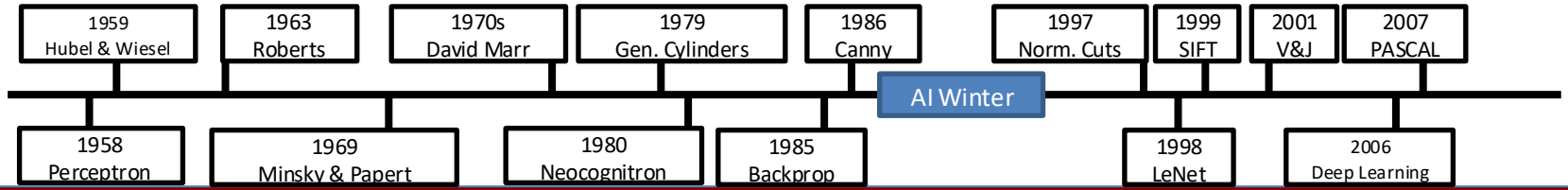
2000s: “Deep Learning”

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





Slide inspiration: Justin Johnson

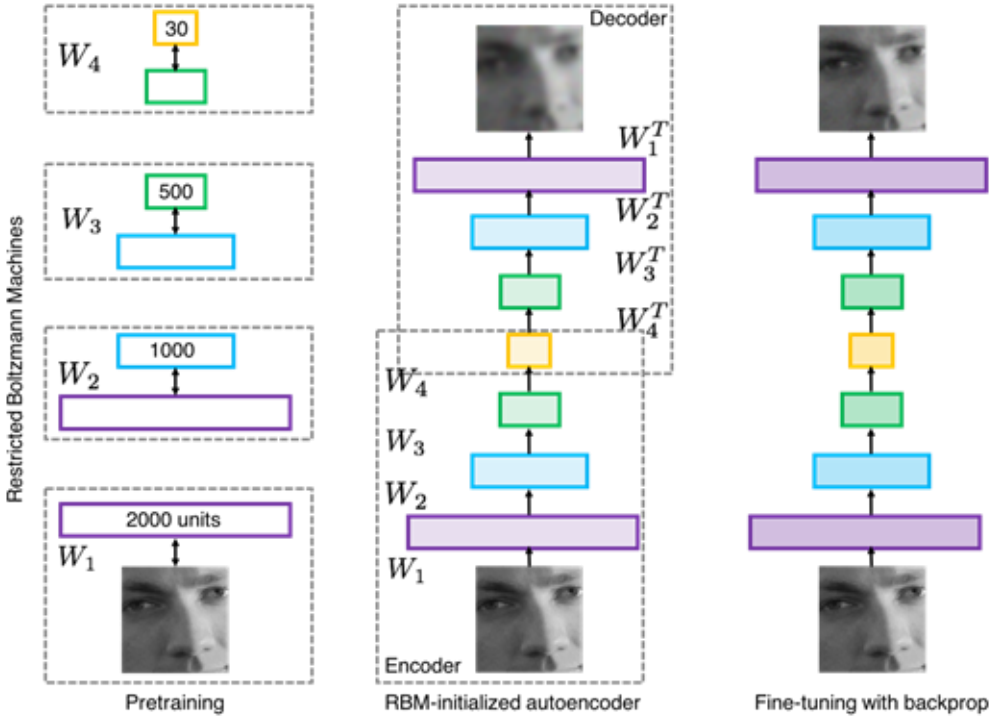
2000s: “Deep Learning”

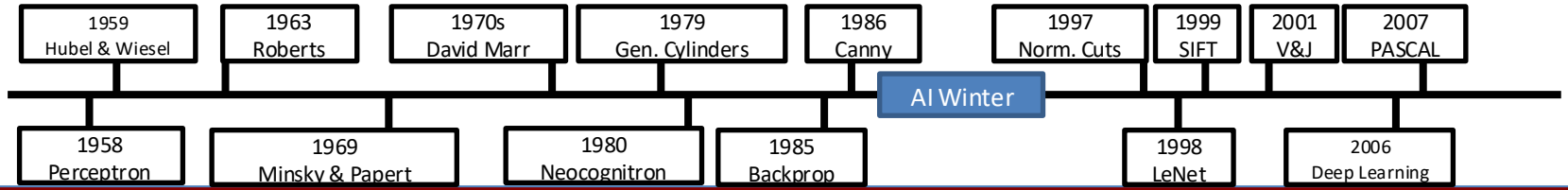
People tried to train neural networks that were deeper and deeper

Not a mainstream research topic at this time

No good dataset to work on

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





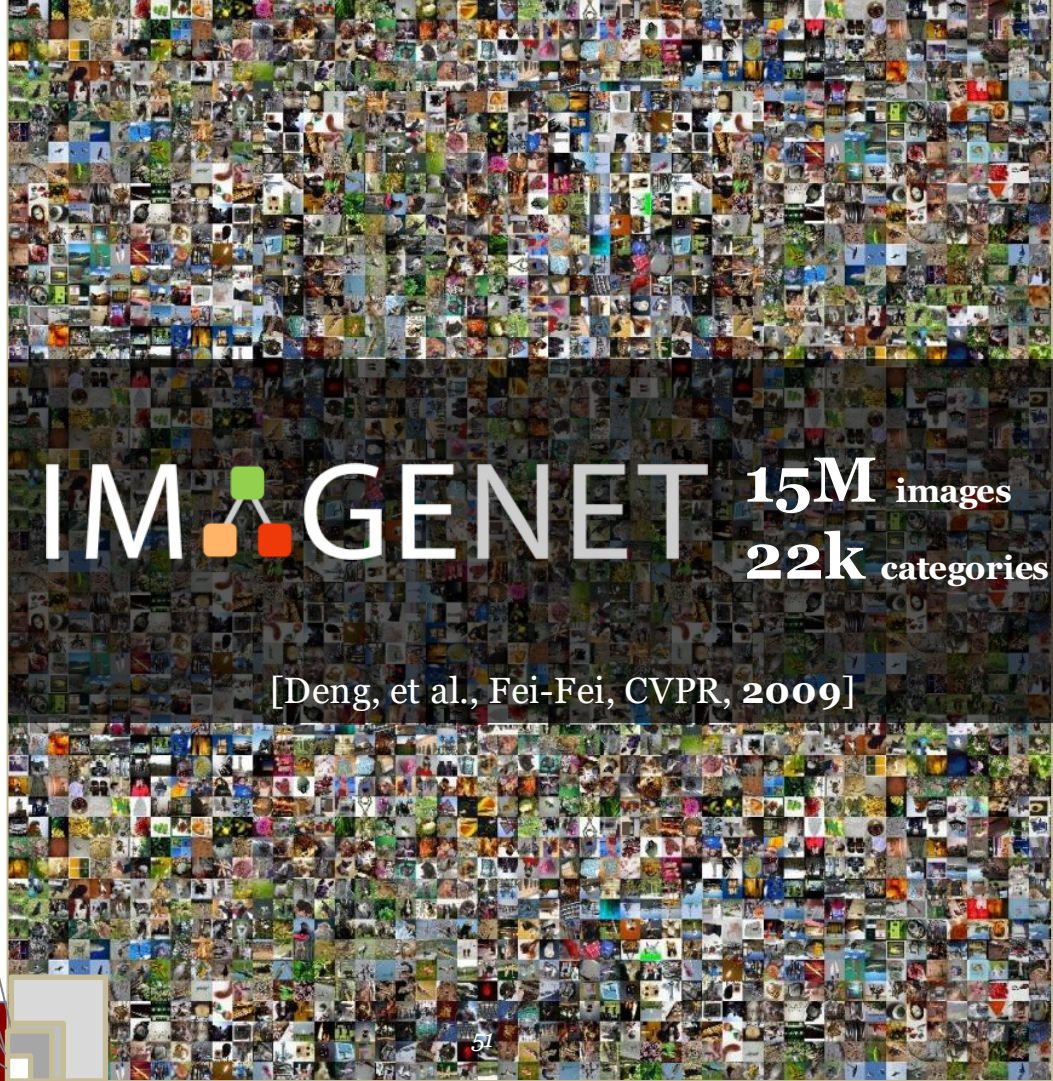
Slide inspiration: Justin Johnson

SUN, 131K images
*[Xiao et al. **2010**]*

LabelMe, 37K images
*[Russell et al. **2007**]*

PASCAL VOC, 30K images
*[Everingham et al. **2006-2012**]*

Caltech101, 9K images
*[Fei-Fei, Fergus, Perona, **2003**]*



IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images

Output:

Scale

T-shirt

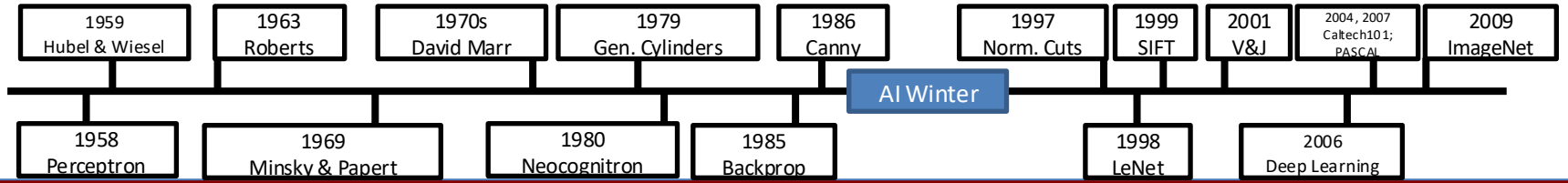
Steel drum

Drumstick

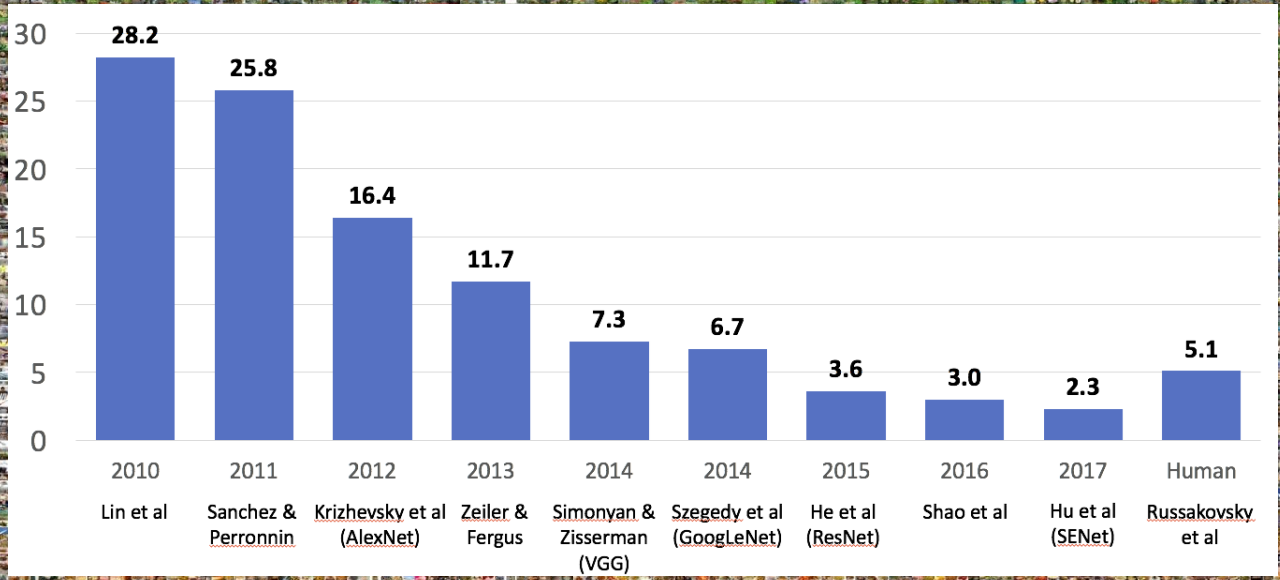
Mud turtle

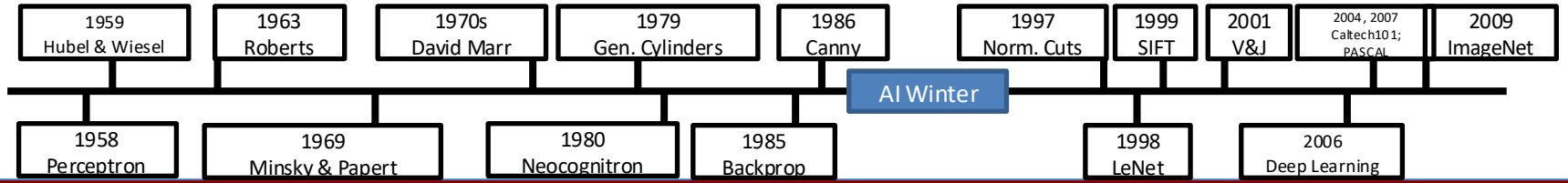
Deng et al, 2009

Russakovsky et al. IJCV 2015

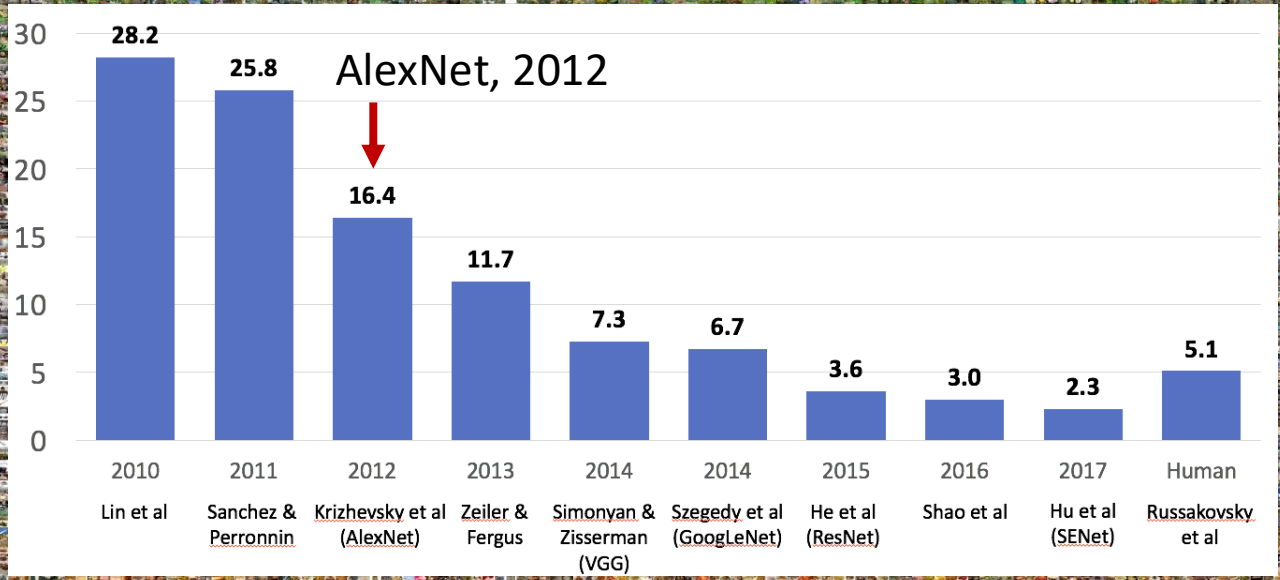


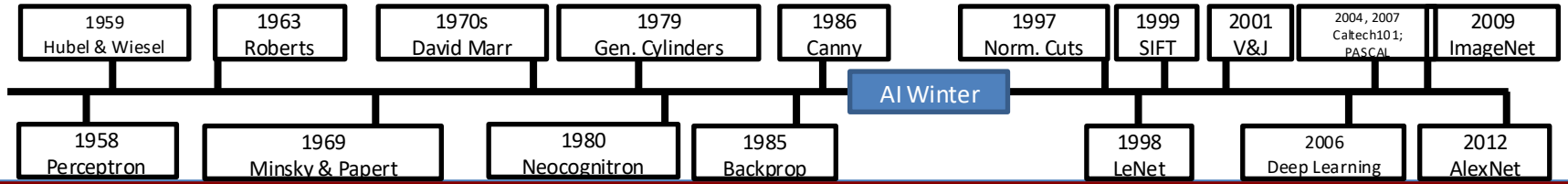
IMAGENET Large Scale Visual Recognition Challenge



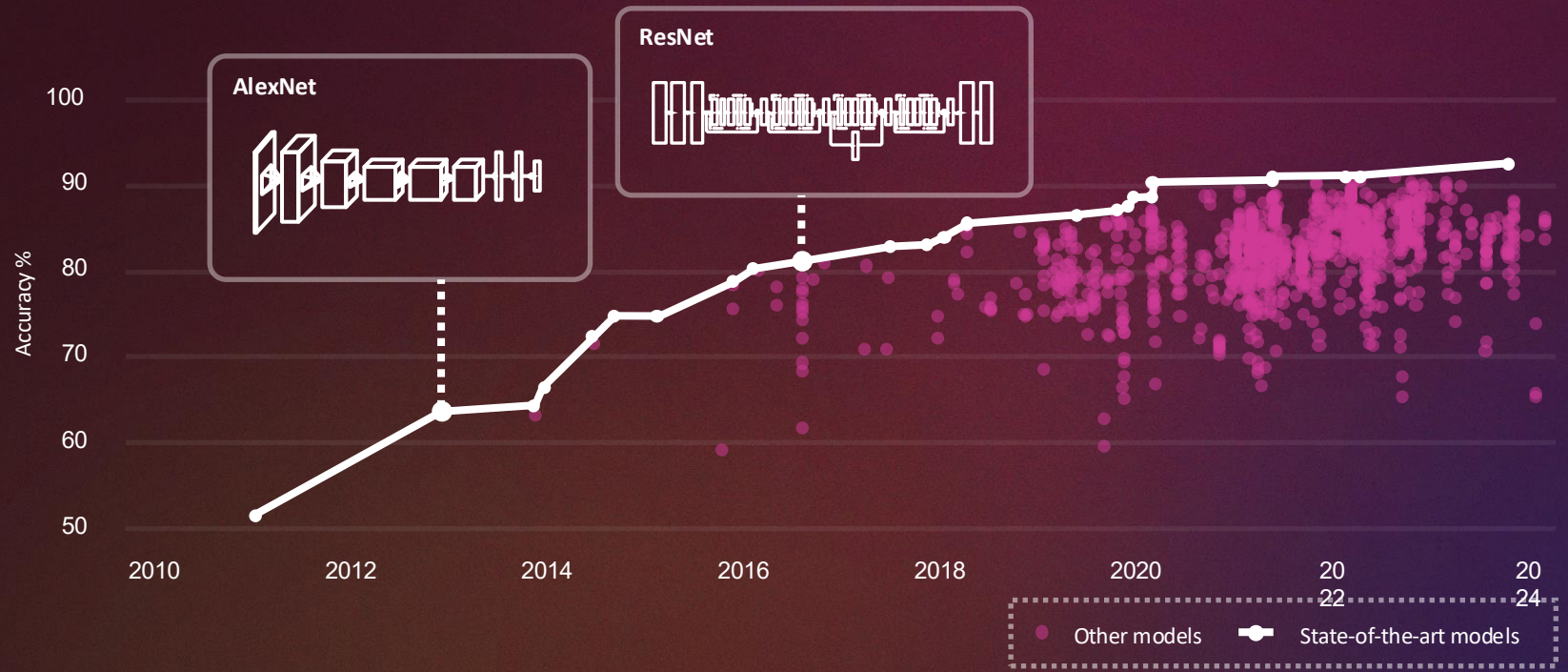


IMAGENET Large Scale Visual Recognition Challenge



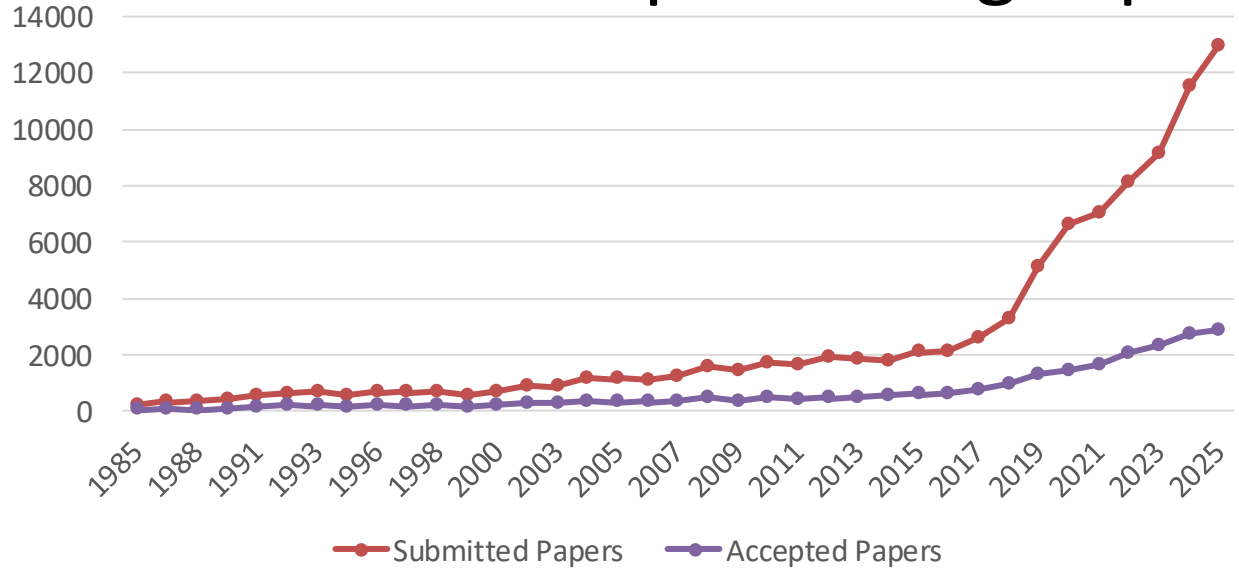


Top 1 Accuracy Model in Image Classification on IMAGENET Each Year

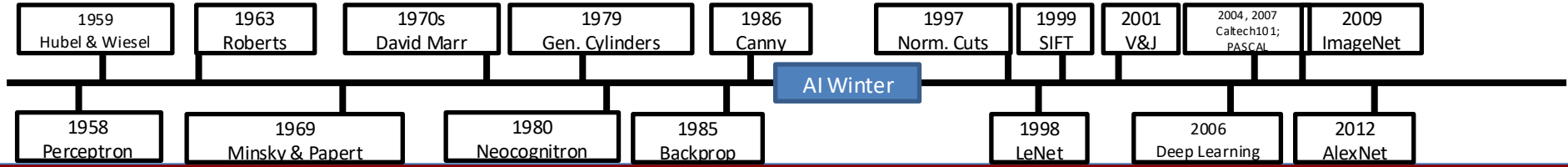


https://paperswithcode.com/sota/image-classification-on-imagenet

2012 to Present: Deep Learning Explosion



Publications at top Computer Vision conference

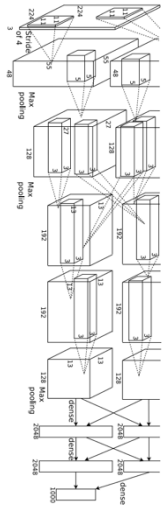


Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Year 2012

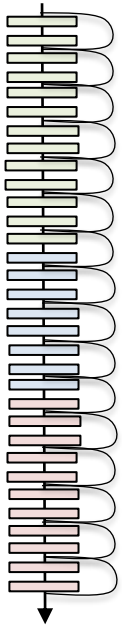
AlexNet



[Krizhevsky NIPS 2012]

Year 2015

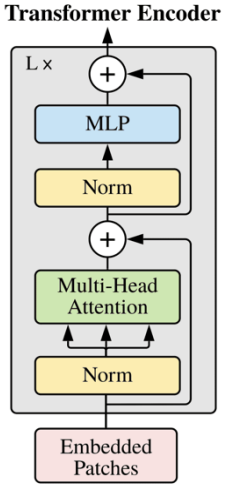
ResNet



[He ICCV 2015]

Year 2021

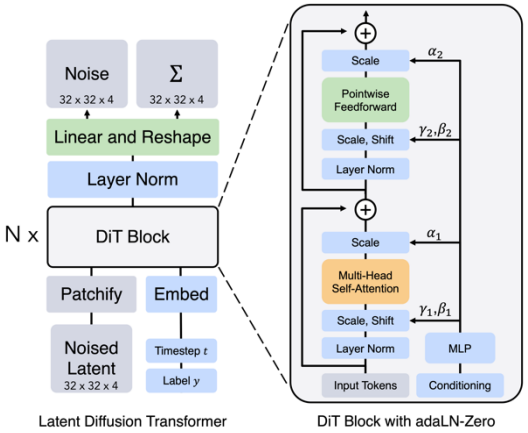
Vision Transformer (ViT)



[Dosovitskiy ICLR 2021]

Year 2023

Diffusion Transformer (DiT)



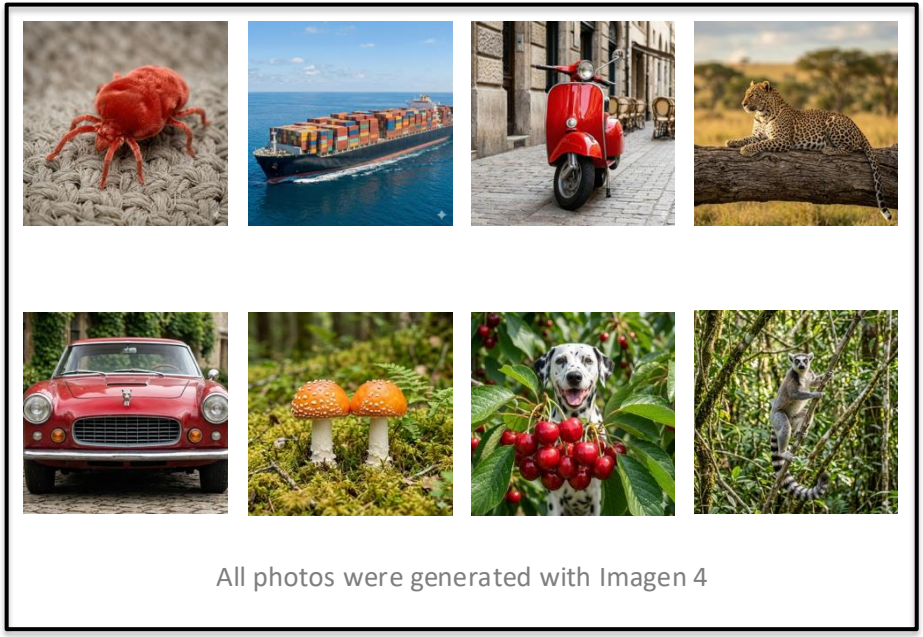
[Peebles ICCV 2023]

2012 to Present: Deep Learning is Everywhere

Image Classification (circa 2012)

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

Image Generation (2026)

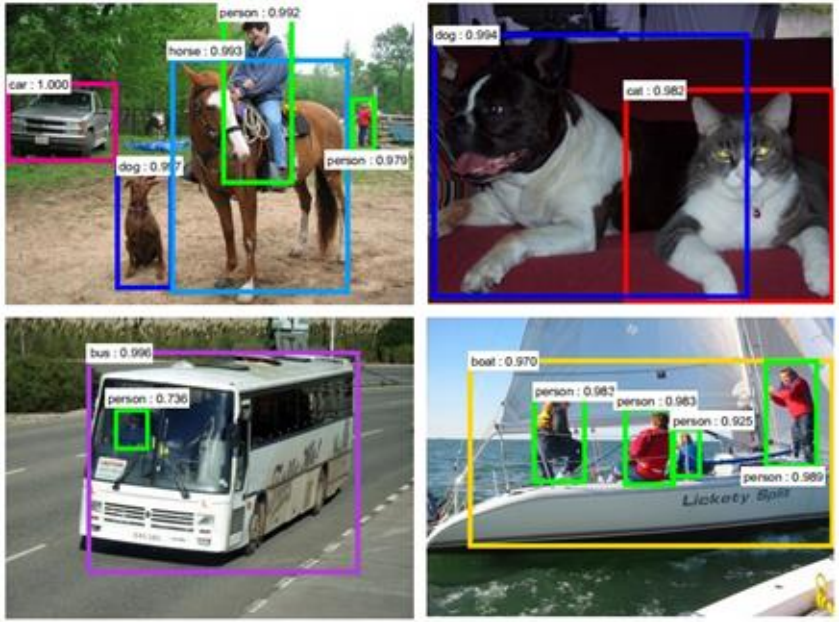


Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Object Detection



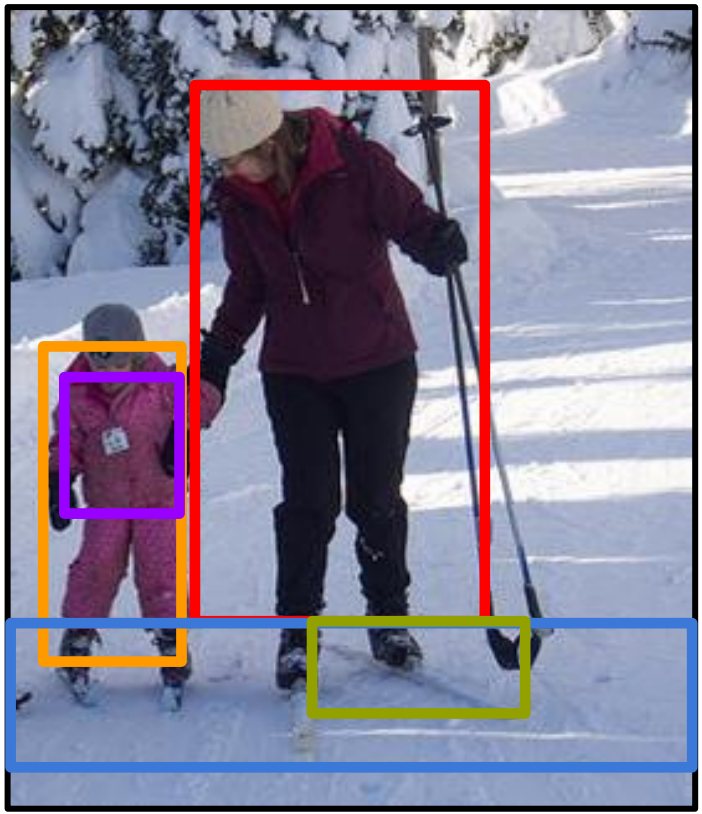
Faster R-CNN
Ren, He, Girshick, and Sun, NeurIPS 2016

Image Segmentation

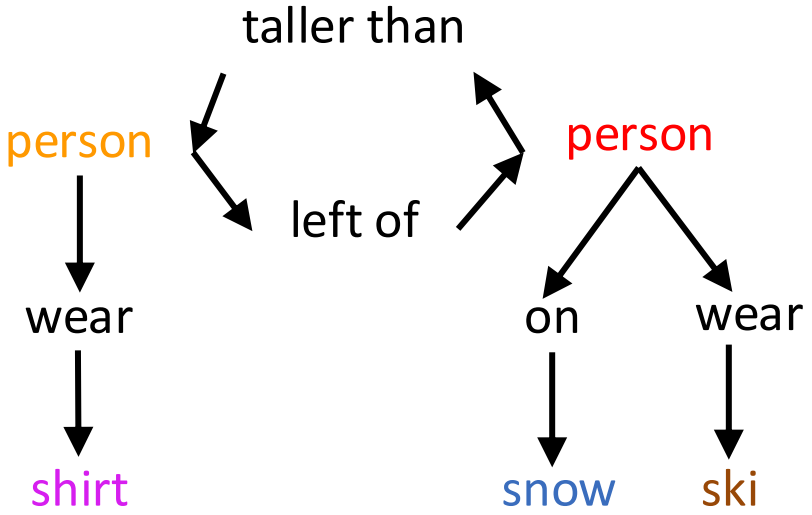
Segment Anything
Kirillov et al, ICCV 2023

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere



Results:
spatial, comparative, asymmetrical, verb, prepositional



Krishna*, Lu*, Bernstein, Fei-Fei, *ECCV* 2016

2012 to Present: Deep Learning is Everywhere

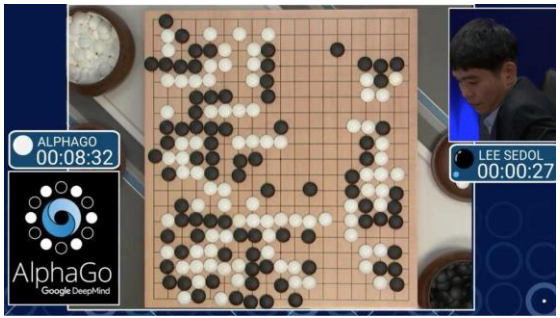
Video Understanding and Activity Recognition

A. Gupta, J. W, J. Deng & L. Fei-Fei, *NeurIPS*, 2023, Z Luo, Z. Durante, L. Li, W, Xie, R. Liu, E. Jin, Z. Huang, L. Y. Li, J. Wu, J. C. Niebles, E. Adeli, & L. Fei-Fei, *NeurIPS*, 2022

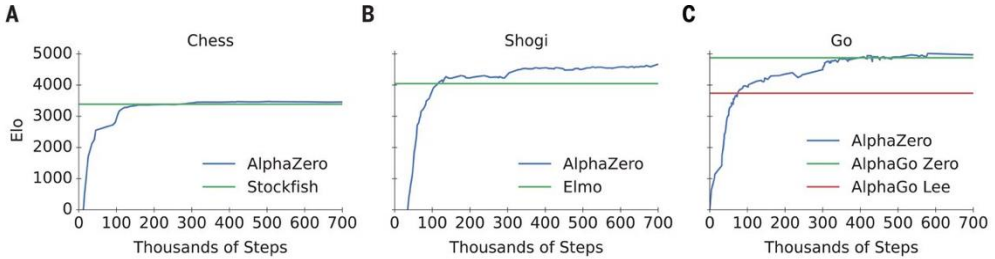
2012 to Present: Deep Learning is Everywhere

Human Mesh Reconstruction (Goel et al, 2023)

Game Playing (AlphaGo/AlphaZero, Silver et al, 2016 & 2017)



Source: https://www.bbc.com/news/technology-35785875



Source: https://arxiv.org/pdf/1712.01815

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Early 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

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Modern Image Captioning – can be much more detailed!

A white teddy bear sitting in the grass

Gemini 3 (2026):

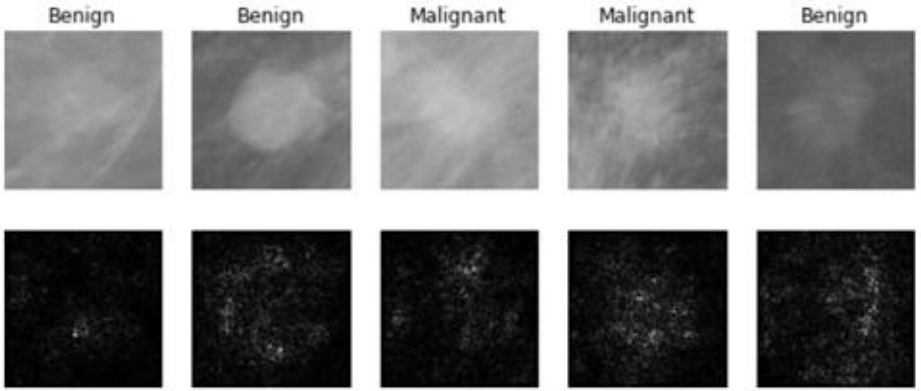
The image features a charming, **cream-colored teddy bear** sitting peacefully in a sunlit grassy field.

Key Details:
The Bear: It has soft, fluffy fur and is wearing a **hand-knitted white sweater** or dress. Its expression is neutral yet sweet, with dark eyes and a stitched brown nose.
The Setting: The bear is positioned in lush green grass dotted with small white **daisies**. One particular daisy appears to be "growing" or resting right on top of the bear's head, giving it a whimsical look.
Lighting: The scene is bright and warm, suggesting a clear, sunny day. The sunlight creates soft shadows and highlights the texture of both the bear's fur and the knitted yarn of its outfit.
The overall mood is serene, nostalgic, and gentle.

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Medical Imaging

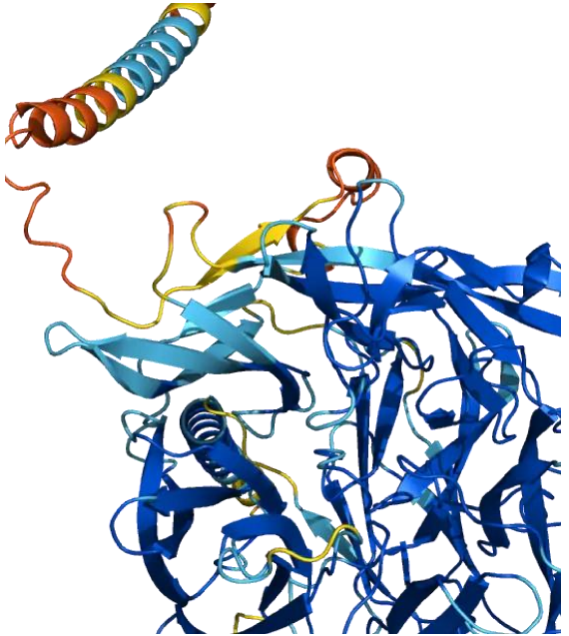


Levy et al, 2016 Figure reproduced with permission

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.



Protein Folding

(Jumper et al 2021)

This image is from the AlphaFold Protein Structure Database (alphafold.ebi.ac.uk) by DeepMind Technologies Limited and EMBL-EBI, licensed under CC-BY 4.0.

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

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

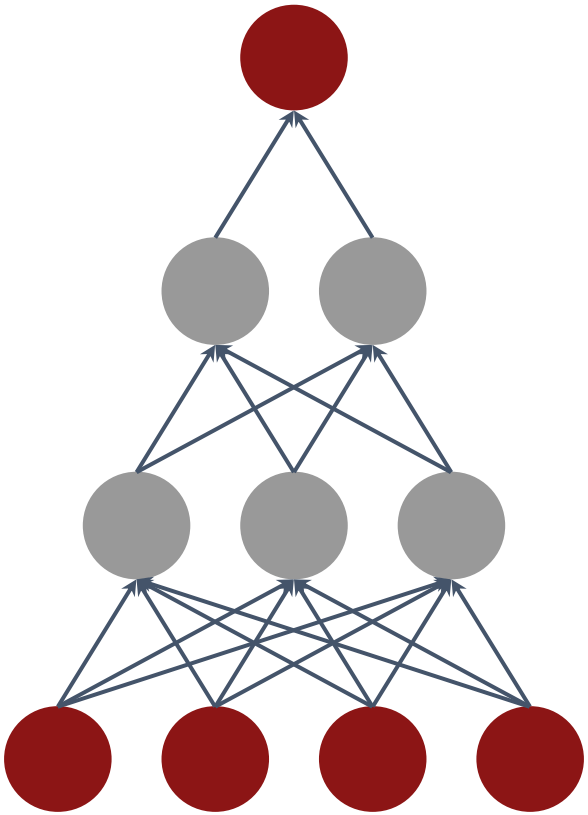
Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Ramesh et al, "DALL·E: Creating Images from Text", 2021. https://openai.com/blog/dall-e/
Remaining images generated by GPT-4o and Imagen 4

Slide inspiration: Justin Johnson

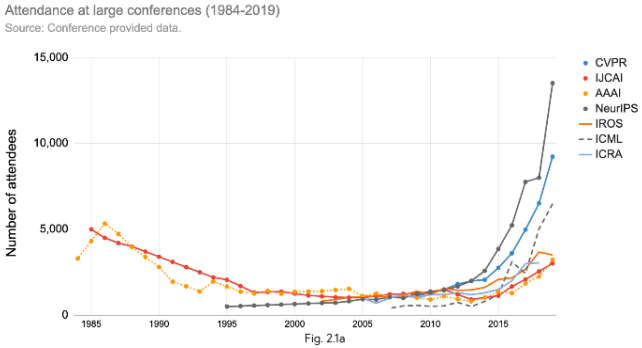
Computation



Algorithms

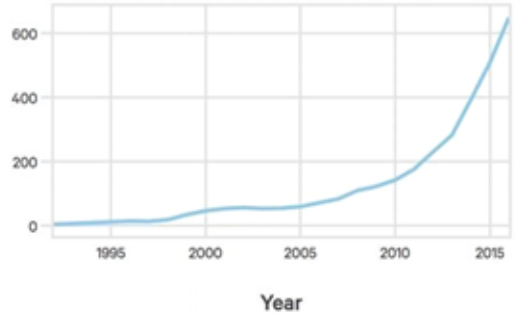
Data

AI's Explosive Growth & Impact



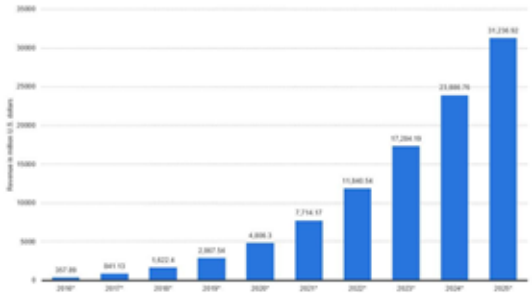
Number of attendance At AI conferences

Source: The Gradient



Startups Developing AI Systems

Source: Crunchbase, VentureSource, Sand Hill Econometrics



Enterprise Application AI Revenue

Source: Statista

Despite the successes, computer vision still has a long way to go

Visual Intelligence entails

Understanding **Reasoning** **Generation**

Data is critical for Deep Learning algorithms

Computer Vision Can Cause Harm

Harmful Stereotypes

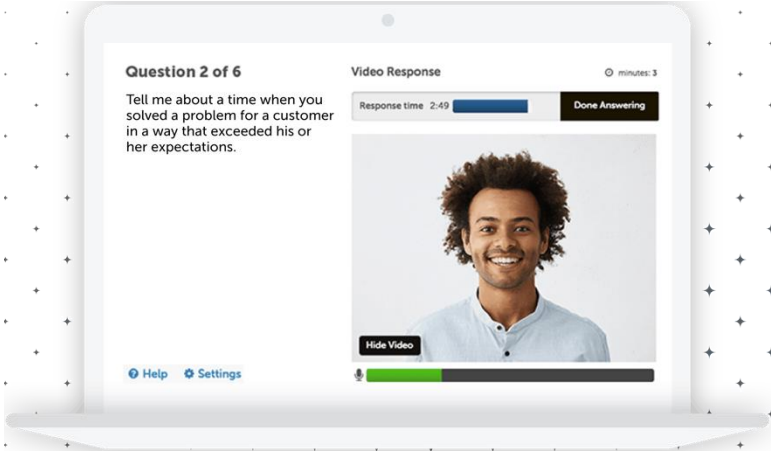
Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Affect people's lives

Technology

A face-scanning algorithm increasingly decides whether you deserve the job

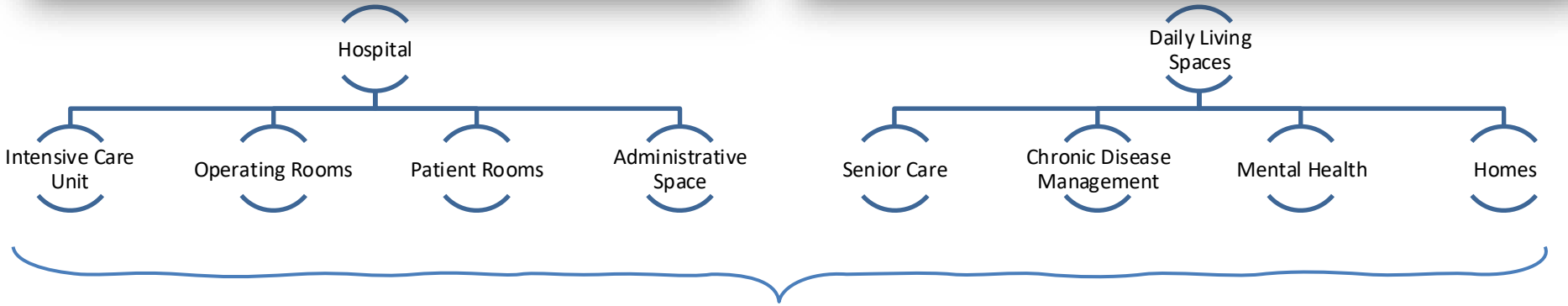
HireVue claims it uses artificial intelligence to decide who's best for a job. Outside experts call it 'profoundly disturbing.'



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

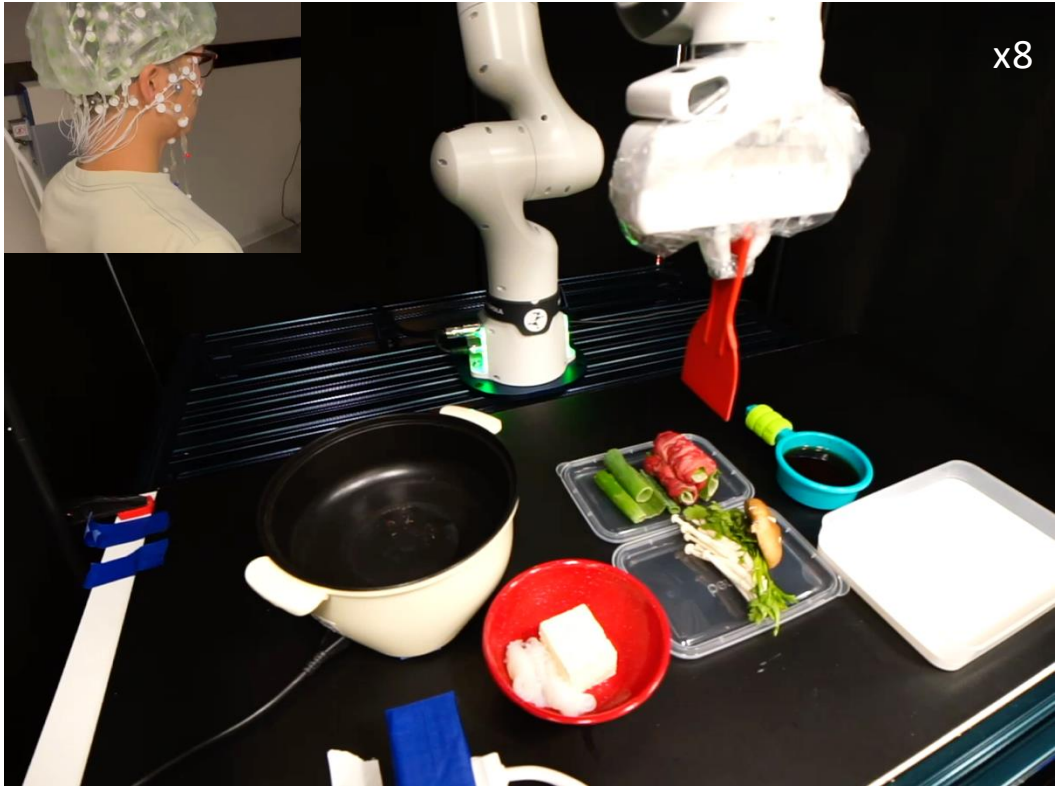
Computer Vision Can Save Lives



Recognizing Activities from Contactless Sensors using Ambient Intelligence Computer Vision Technology

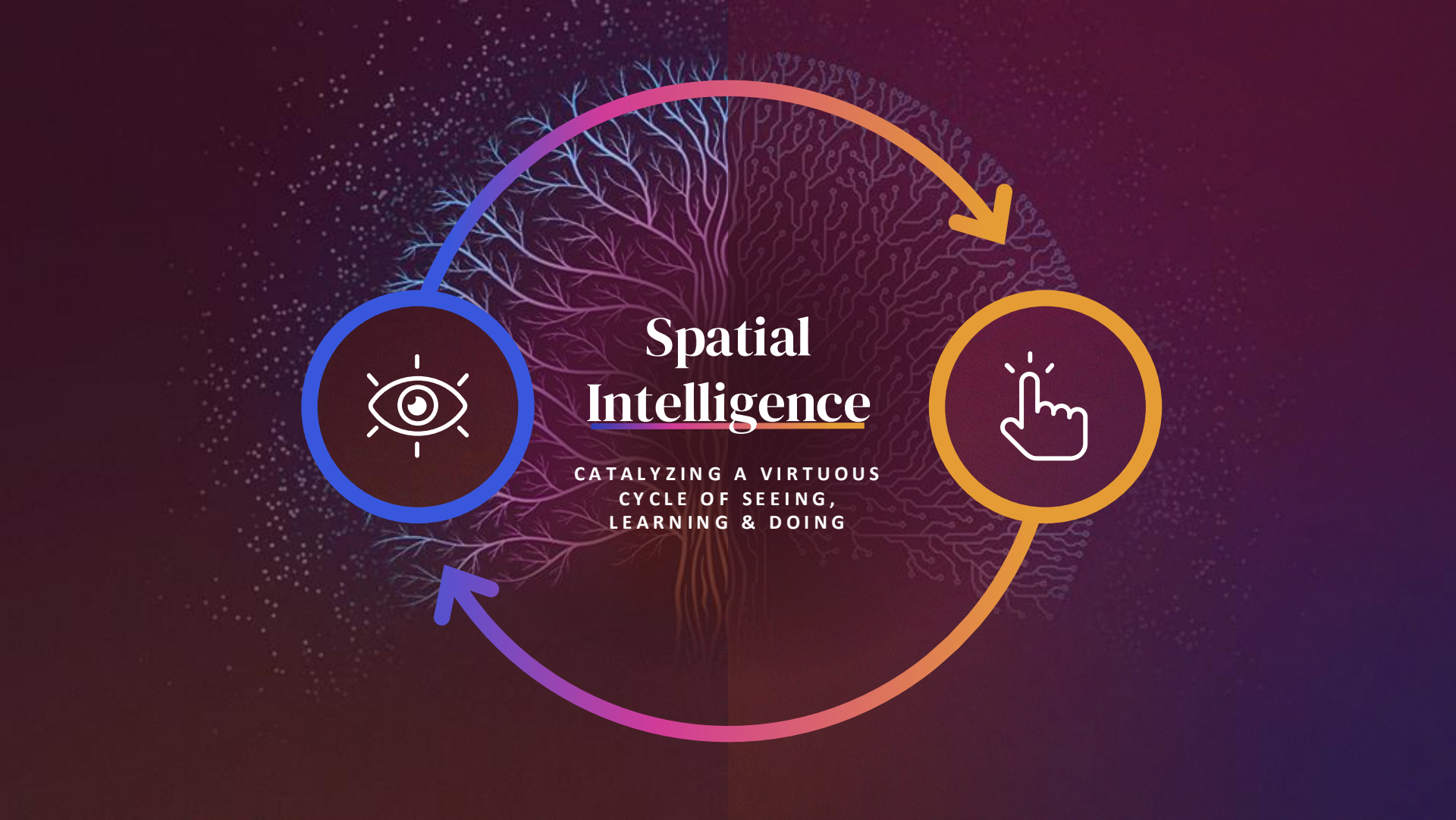
Computer Vision Can Save Lives

Slide inspiration: Justin Johnson



R. Zhang, S. Lee, M. Hwang, A. Hiranaka, C. Wang, W, Ai, J. J. Tan, S. Gupta, Y. Hao, G. Levine, R. Gao, A. Norcia, L. Fei-Fei, J. Wu, *CoRL*, 2023


ayano hiranaka


Spatial Intelligence
CATALYZING A VIRTUOUS CYCLE OF SEEING, LEARNING & DOING

World Modeling

R Wu, B. Mildenhall, P. Henzler, K. Park, R. Gao, D. Watson, P. Srinivasan, D. Verbin, J. T. Barron, B. Pool, A. Holynski. *arXiv*, 2023; *CVPR*, 2024

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

This image is copyright-free United States government work

Slide inspiration: Andrej Karpathy

Today's agenda

- A brief history of computer vision & deep learning

- CS231n overview