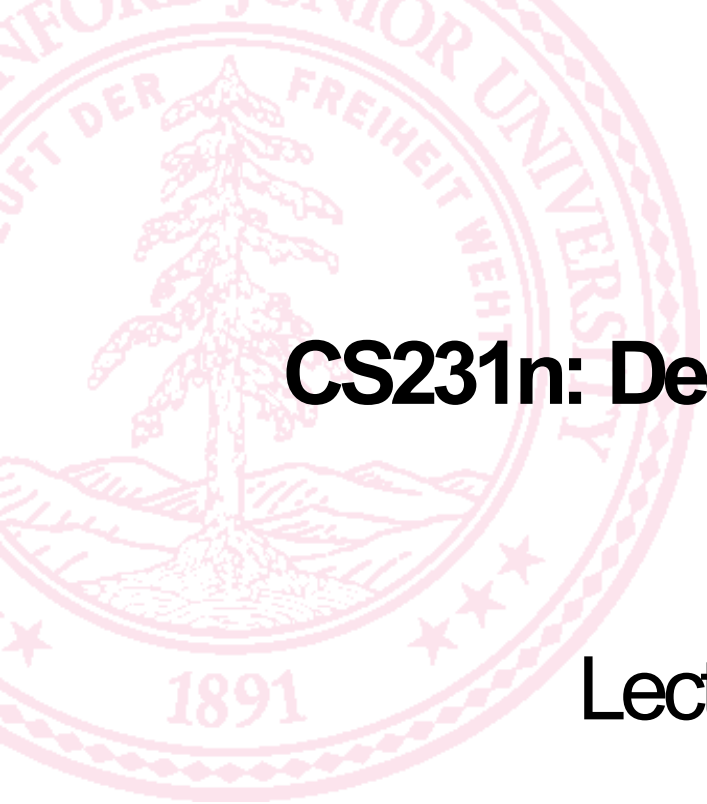
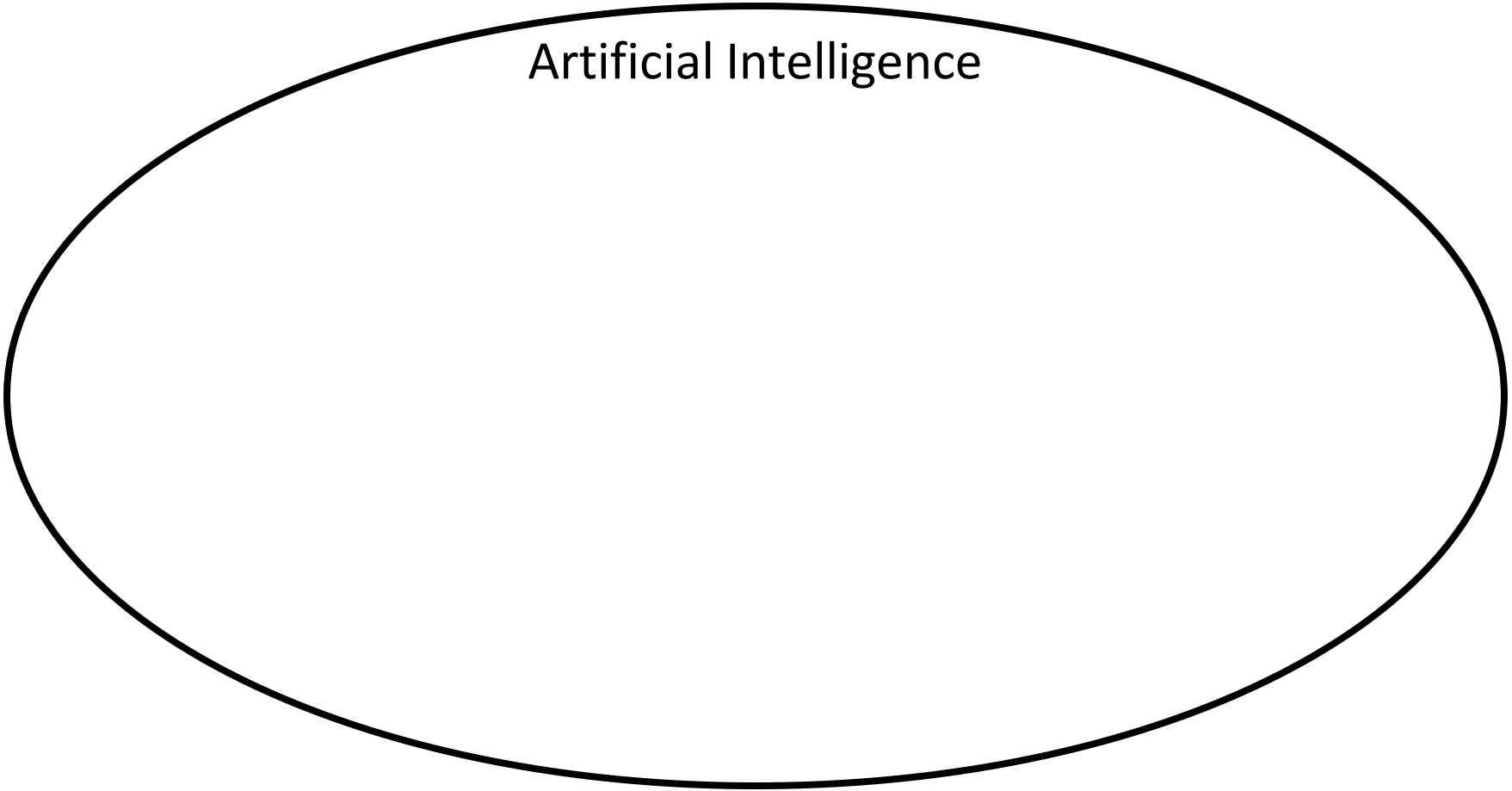
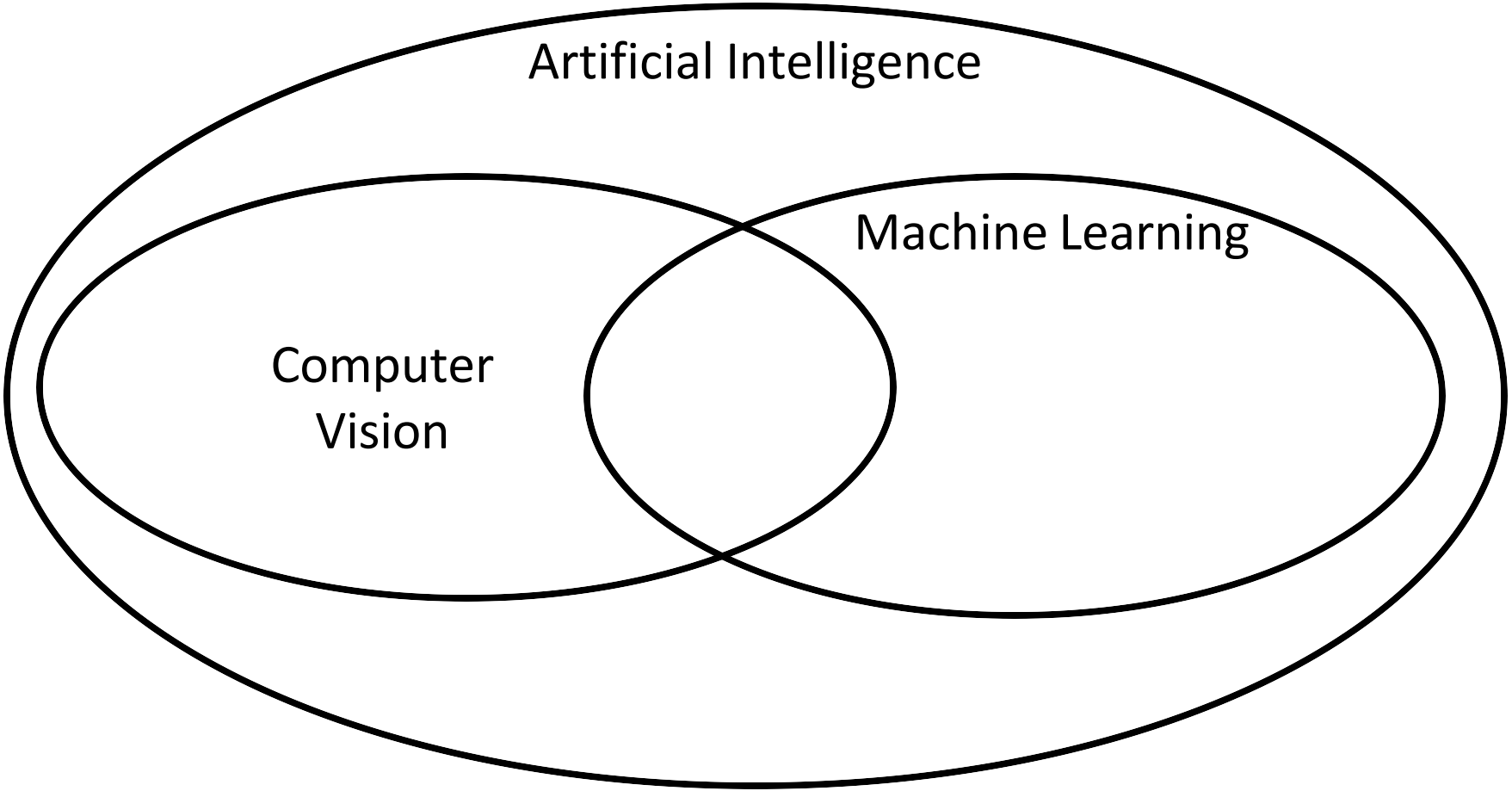


CS231n: Deep Learning for Computer Vision

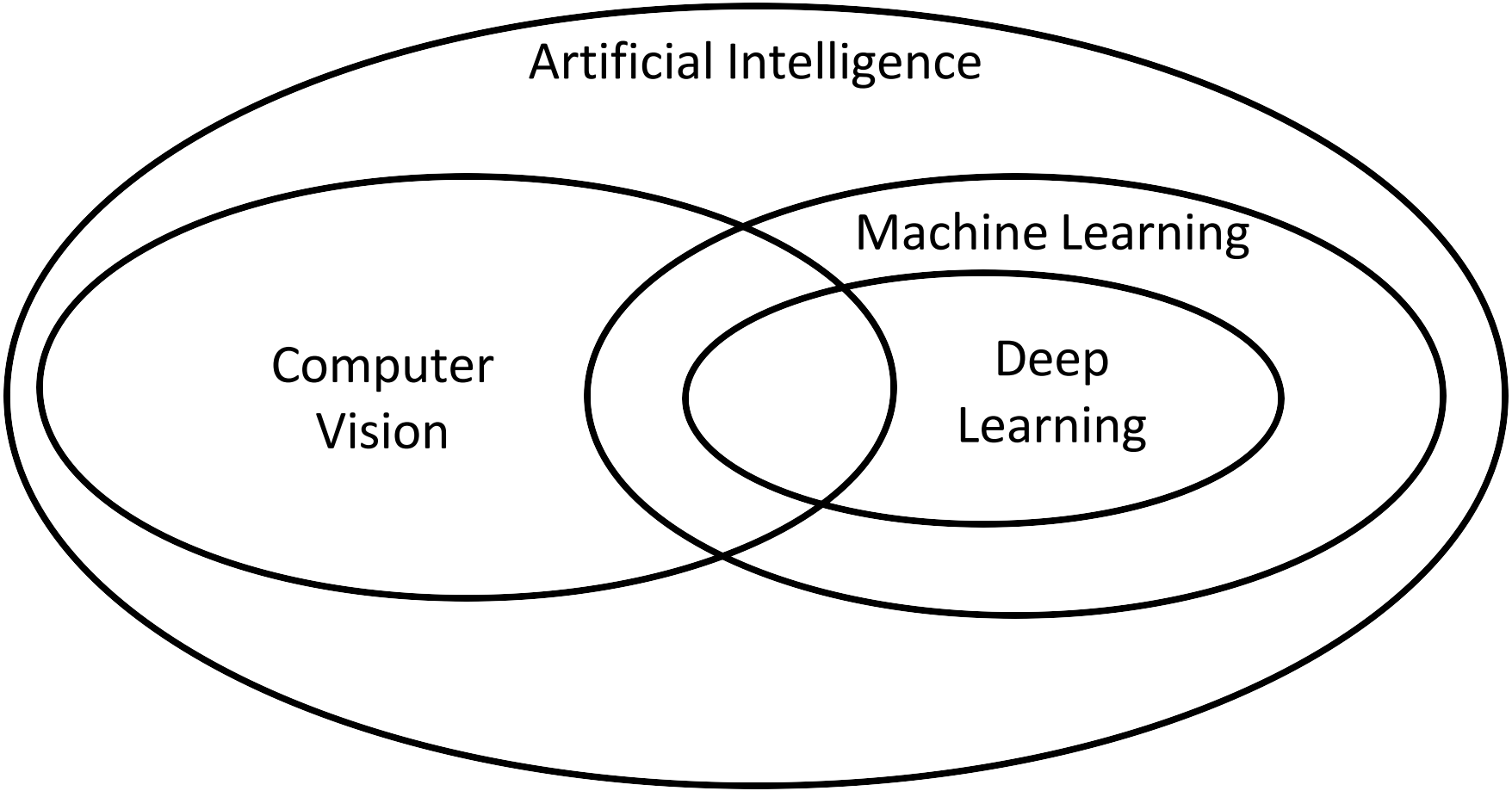
Lecture 1: Introduction



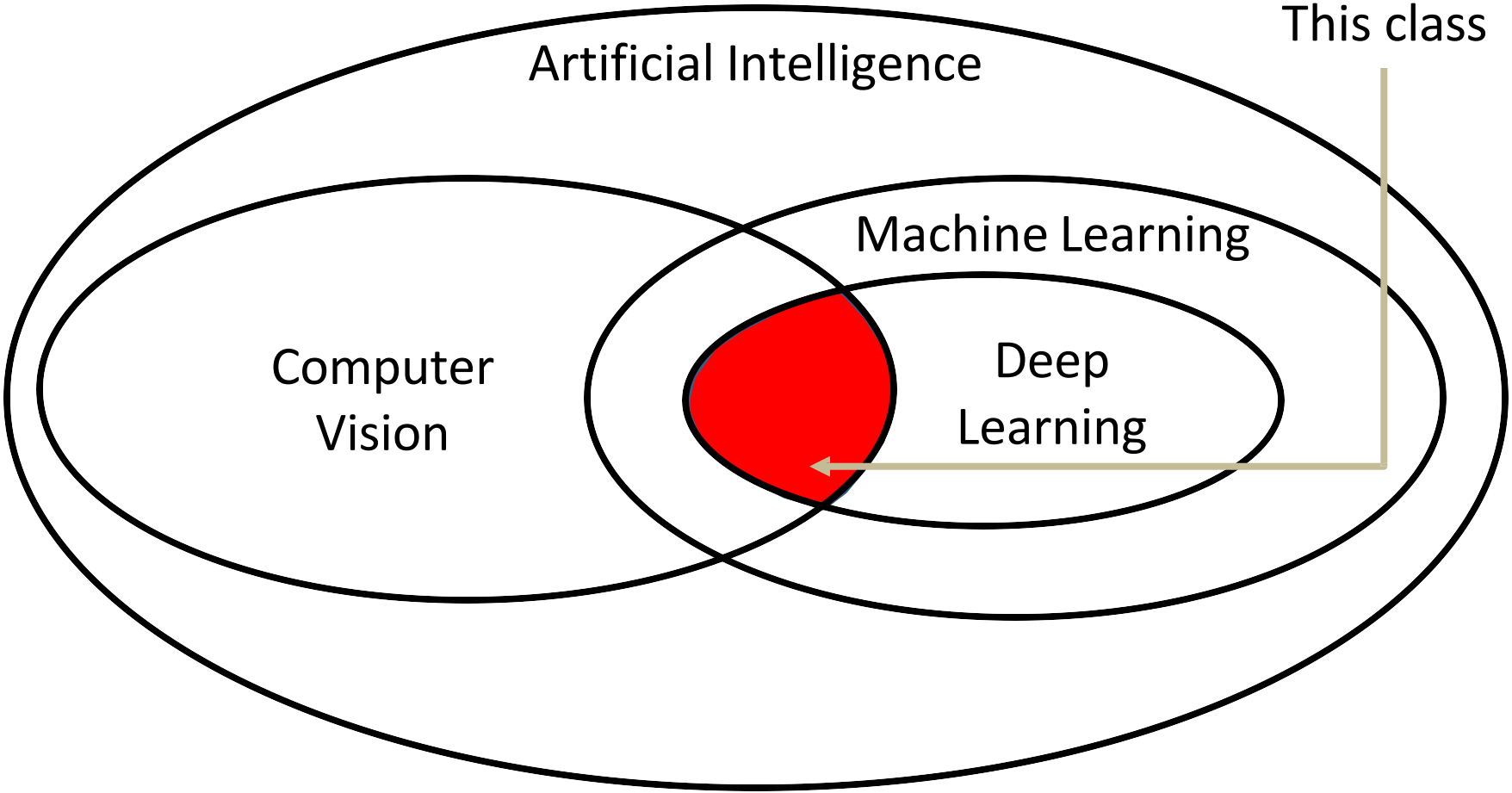
Slide inspiration: Justin Johnson



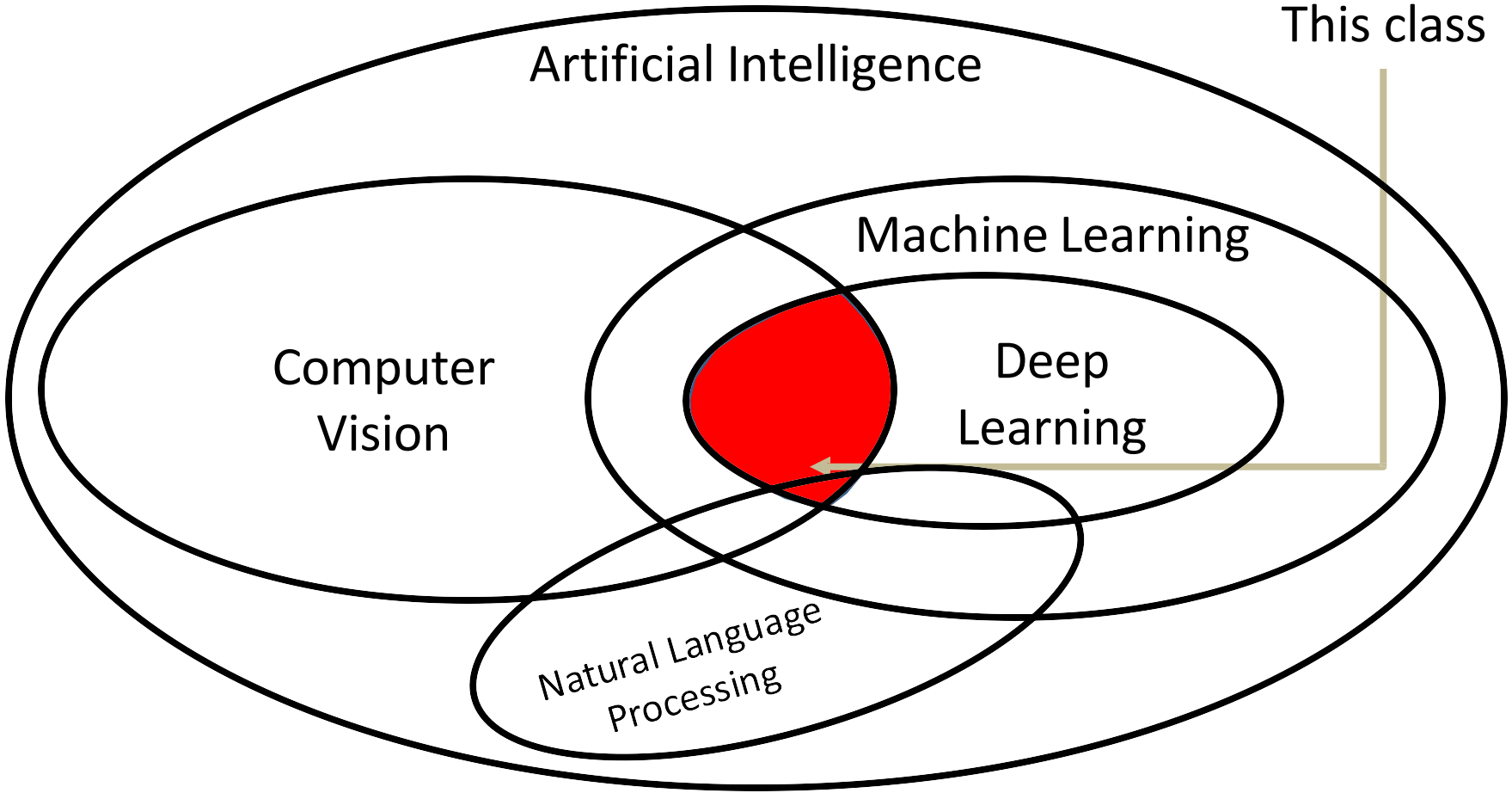
Slide inspiration: Justin Johnson



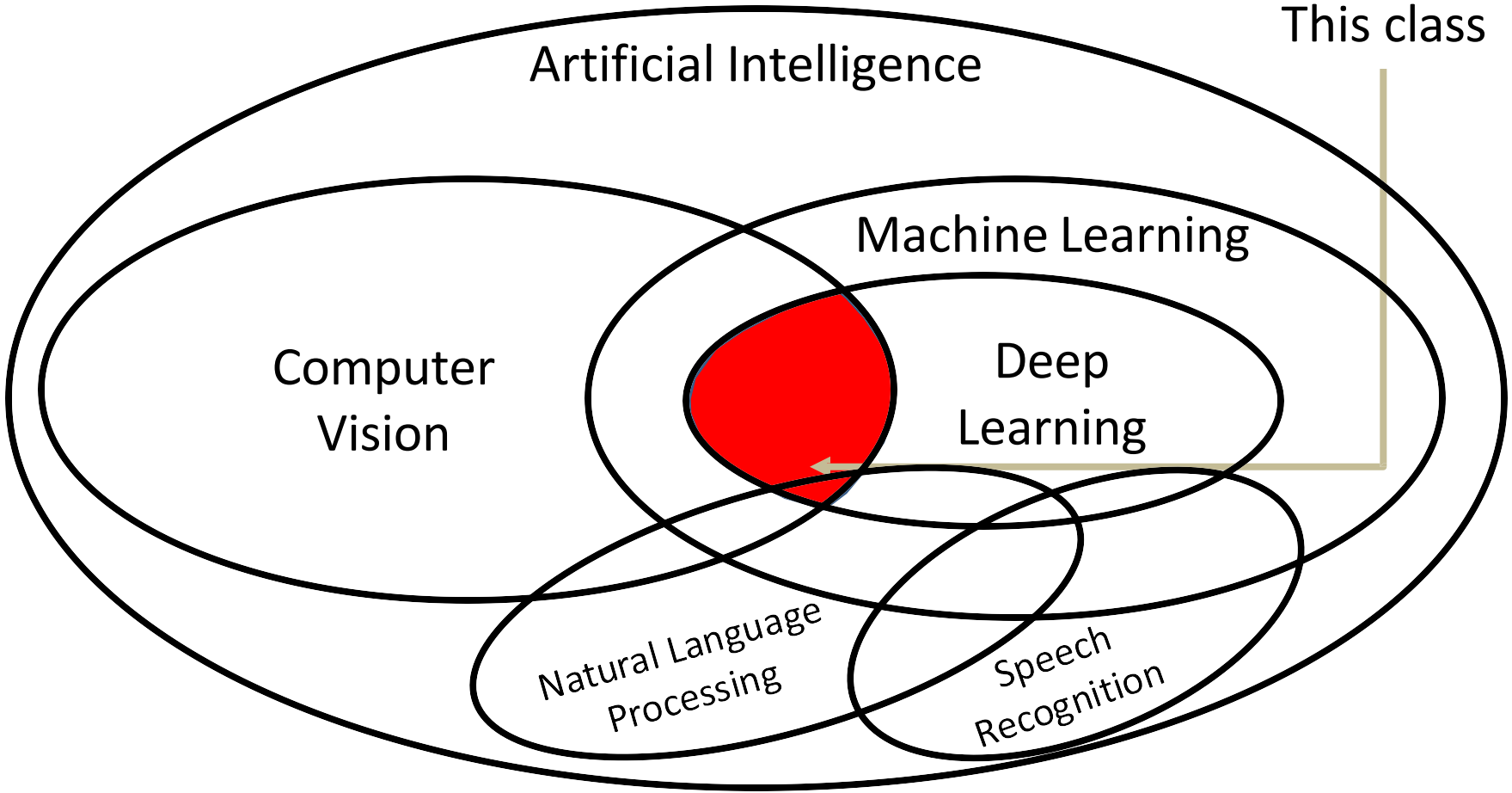
Slide inspiration: Justin Johnson



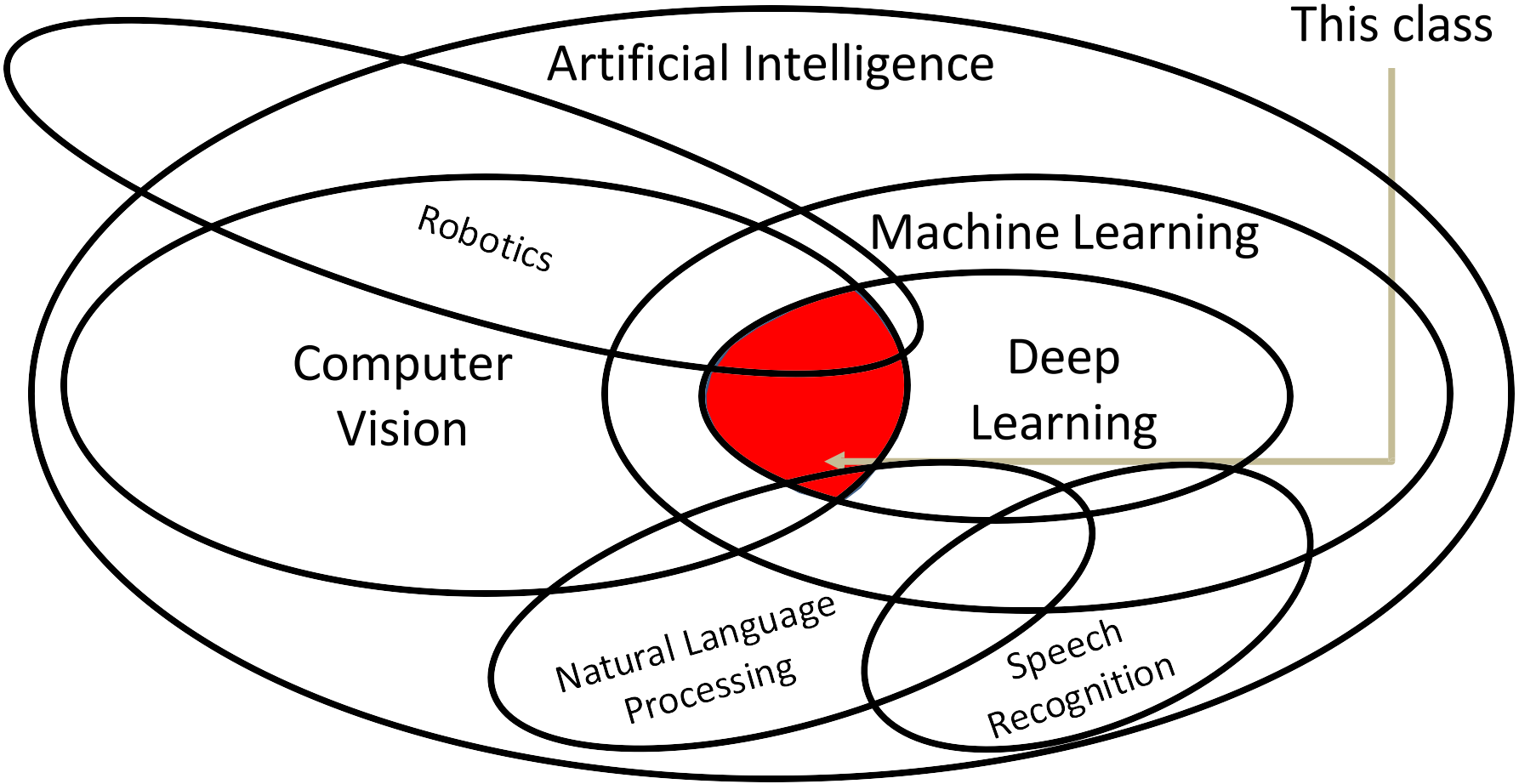
Slide inspiration: Justin Johnson



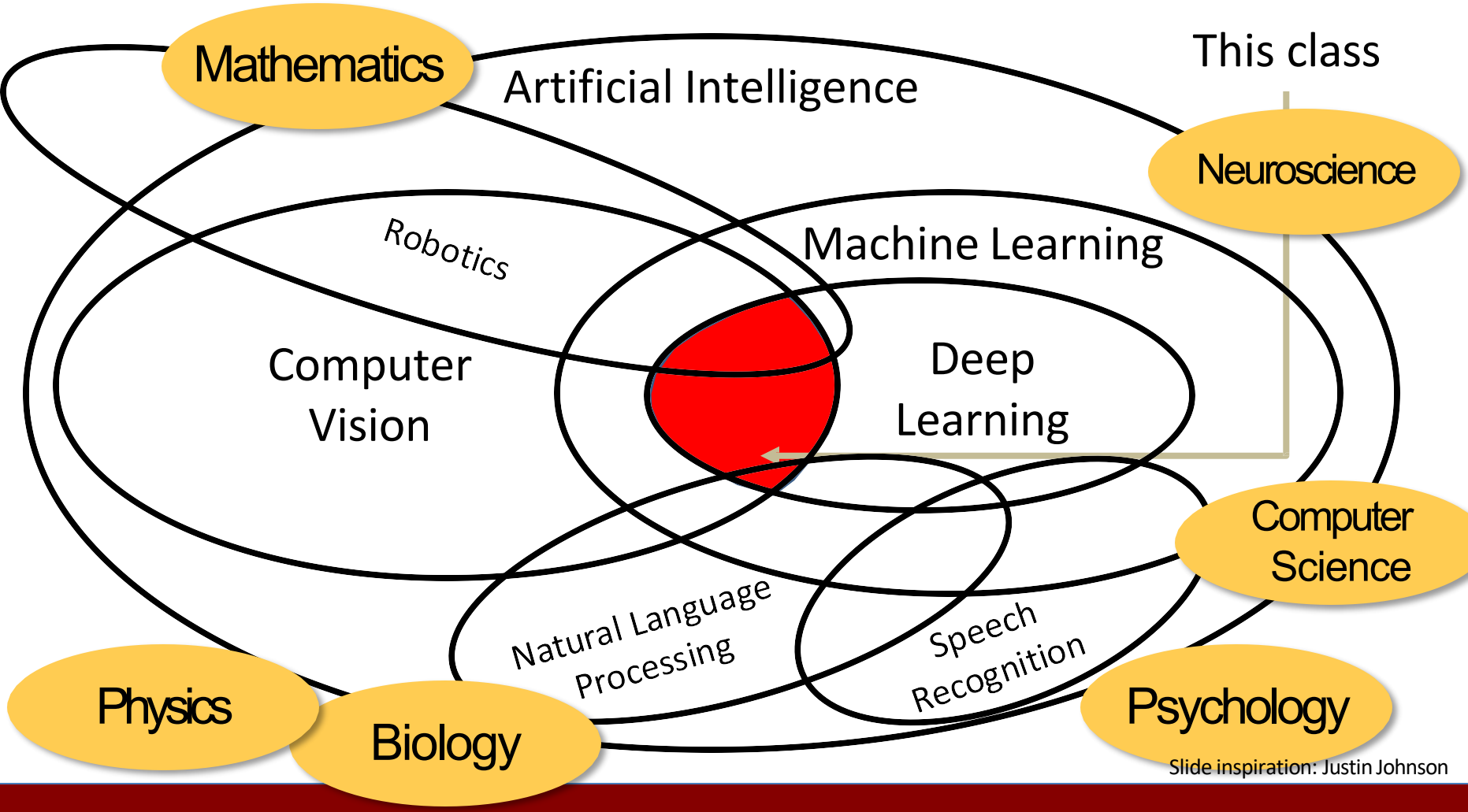
Slide inspiration: Justin Johnson



Slide inspiration: Justin Johnson



Slide inspiration: Justin Johnson

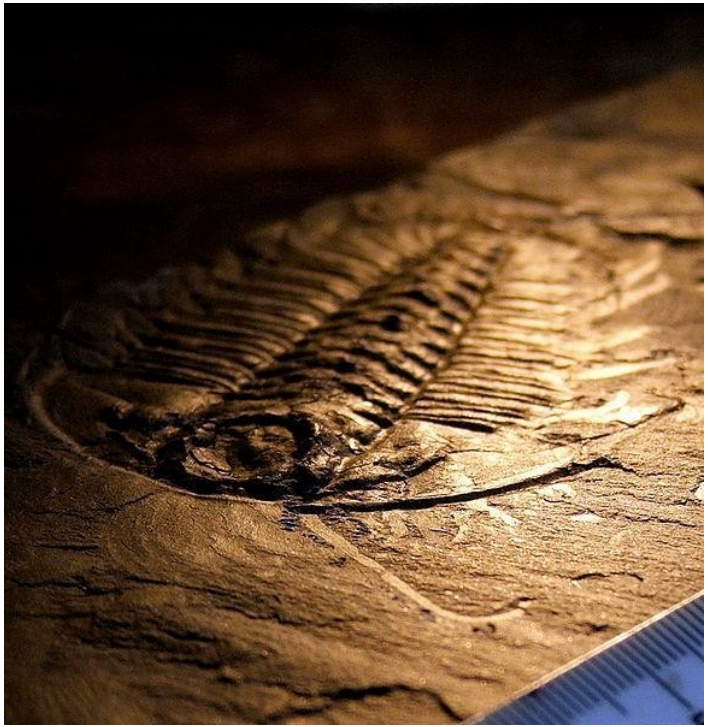


Slide inspiration: Justin Johnson

Today's agenda

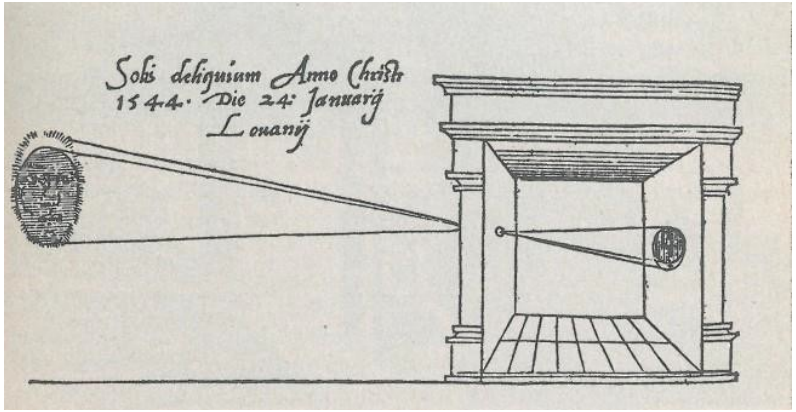
- A brief history of computer vision and deep learning
- CS231n overview

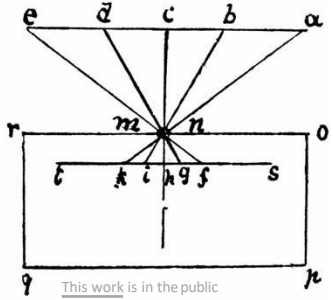
Evolution's Big Bang:
Cambrian Explosion, 530-540million years, B.C.



Camera Obscura

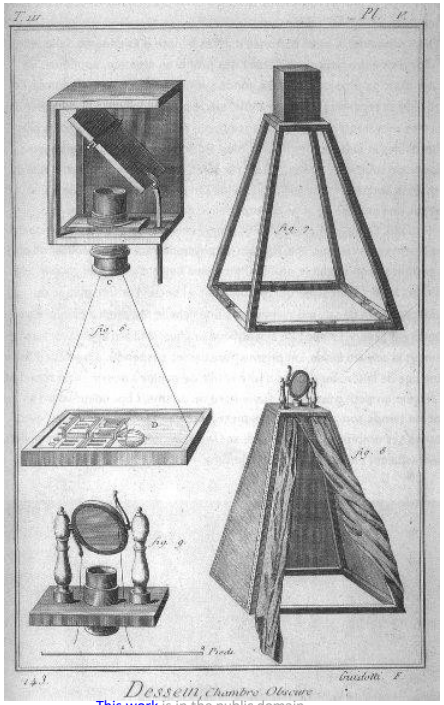
Gemma Frisius, 1545





Leonardo da Vinci, 16th Century AD

Encyclopedia, 18th Century



Computer Vision is everywhere!




Left to right:
Image by Roger H Goun is licensed under CC BY 2.0
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain




Left to right:
Image is free to use
Image is CC0 1.0 public domain
Image by NASA is licensed under CC BY 2.0
Image is CC0 1.0 public domain

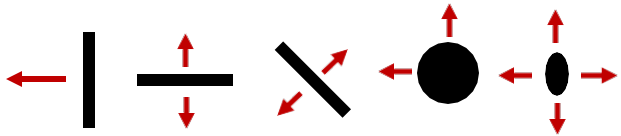


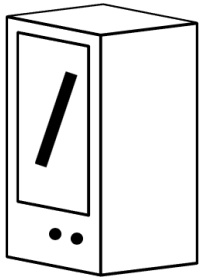

Bottom row, left to right
Image is CC0 1.0 public domain
Image by Derek Keats is licensed under CC BY 2.0; changes made
Image is public domain
Image is licensed under CC BY 2.0; changes made

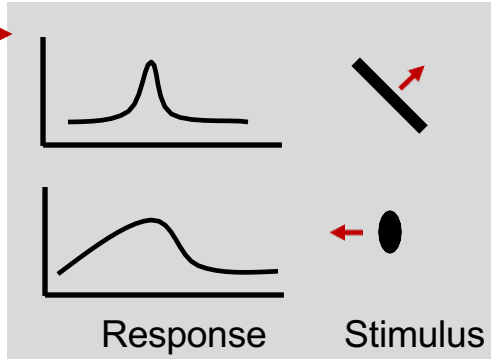
Where did we come from?

Hubel and Wiesel, 1959

Measure brain activity



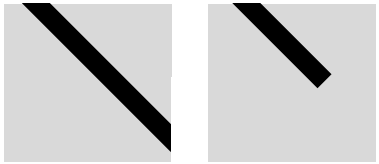




1959
Hubel & Wiesel

Simple cells:
Response to specific rotation and orientation

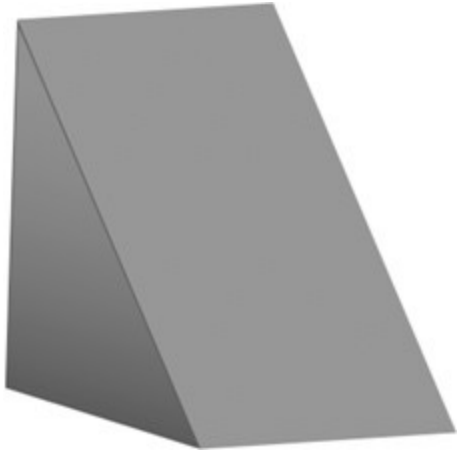
Complex cells:
Response to light orientation and movement, some translation invariance



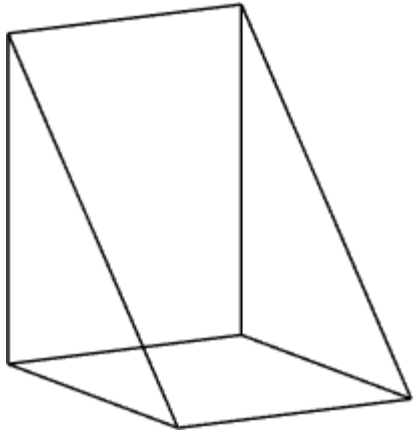
No response

Slide inspiration: Justin Johnson

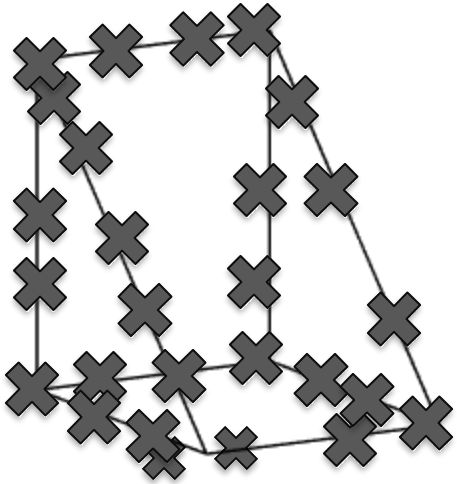
Larry Roberts, 1963



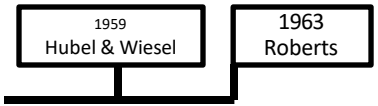
(a) Original picture



(b) Differentiated picture



(c) Feature points selected

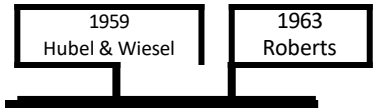


Lawrence Gilman Roberts, “Machine Perception of Three-Dimensional Solids”, 1963

Slide inspiration: Justin Johnson

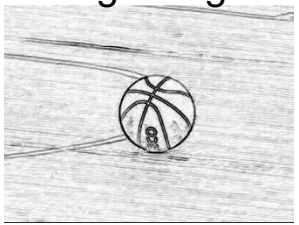
The Summer Vision Project- Seymour Paper

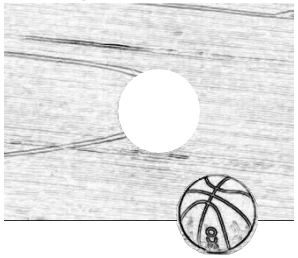
https://dspace.mit.edu/handle/1721.1/6125



Slide inspiration: Justin Johnson

This image is CC0 1.0 public domain





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

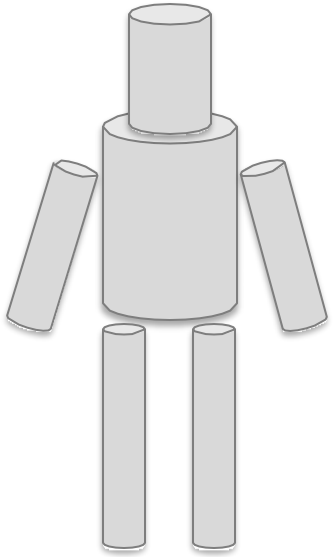
1959 Hubel & Wiesel

1963 Roberts

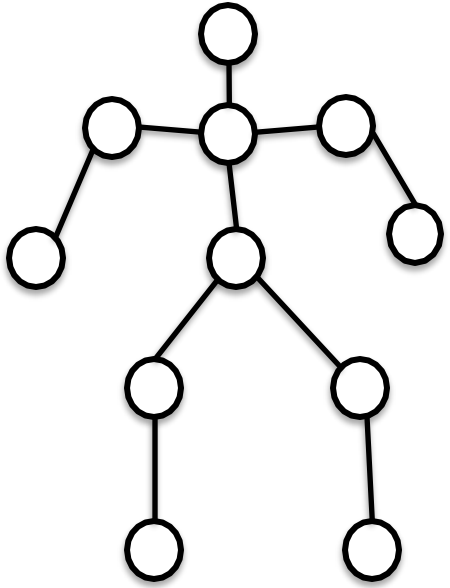
1970s David Marr

Stages of Visual Representation, David Marr, 1970s

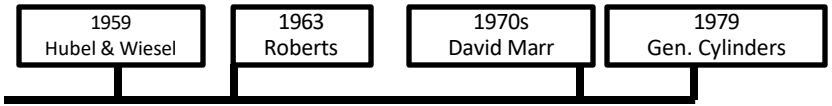
Recognition via Parts (1970s)



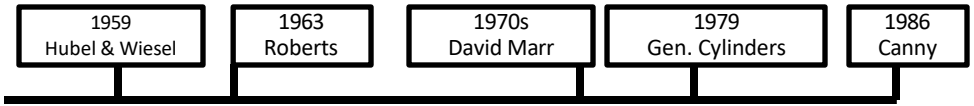
Generalized Cylinders,
Brooks and Binford,
1979



Pictorial Structures,
Fischler and Elshlager, 1973



Recognition via Edge Detection (1980s)

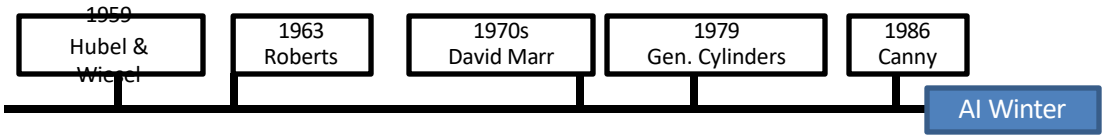


John Canny, 1986
David Lowe, 1987

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.

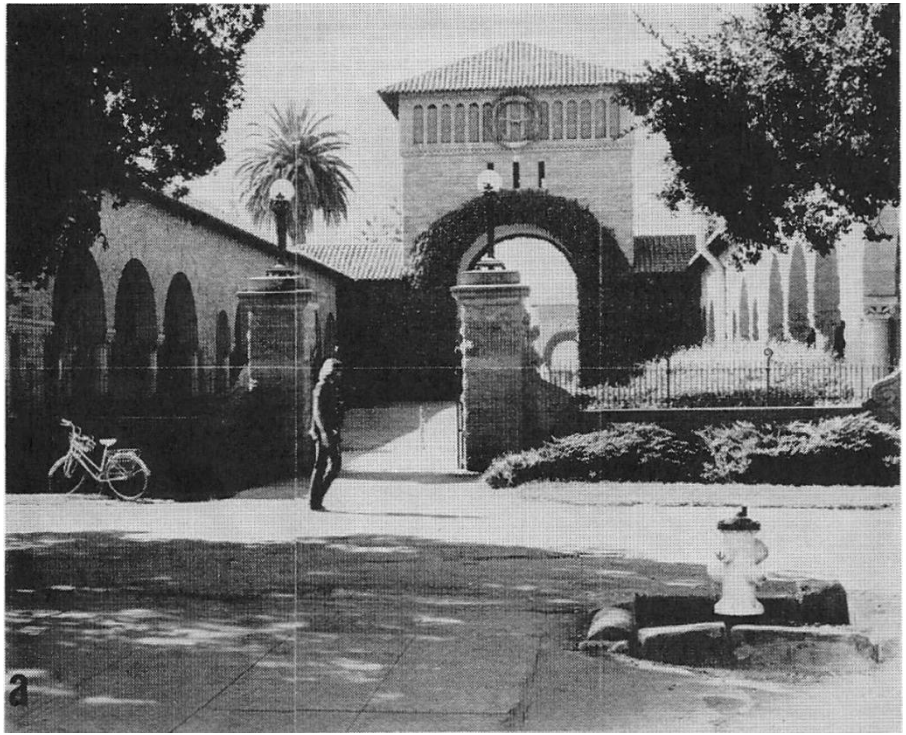


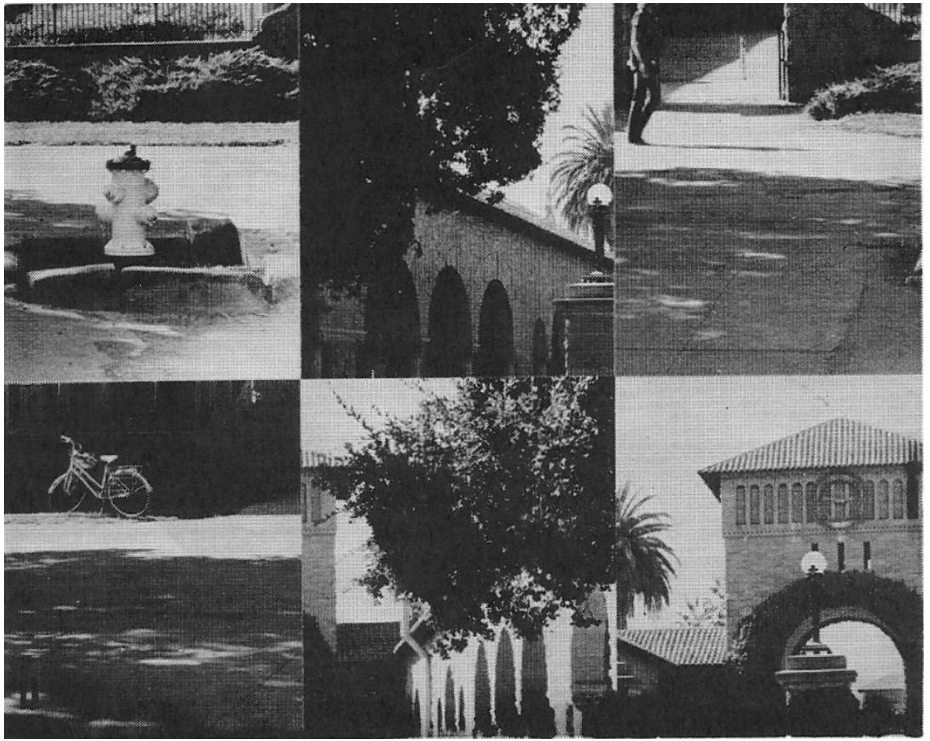
Slide inspiration: Justin Johnson

In the meantime…seminal work in cognitive and neuroscience

Perceiving Real-World Scenes

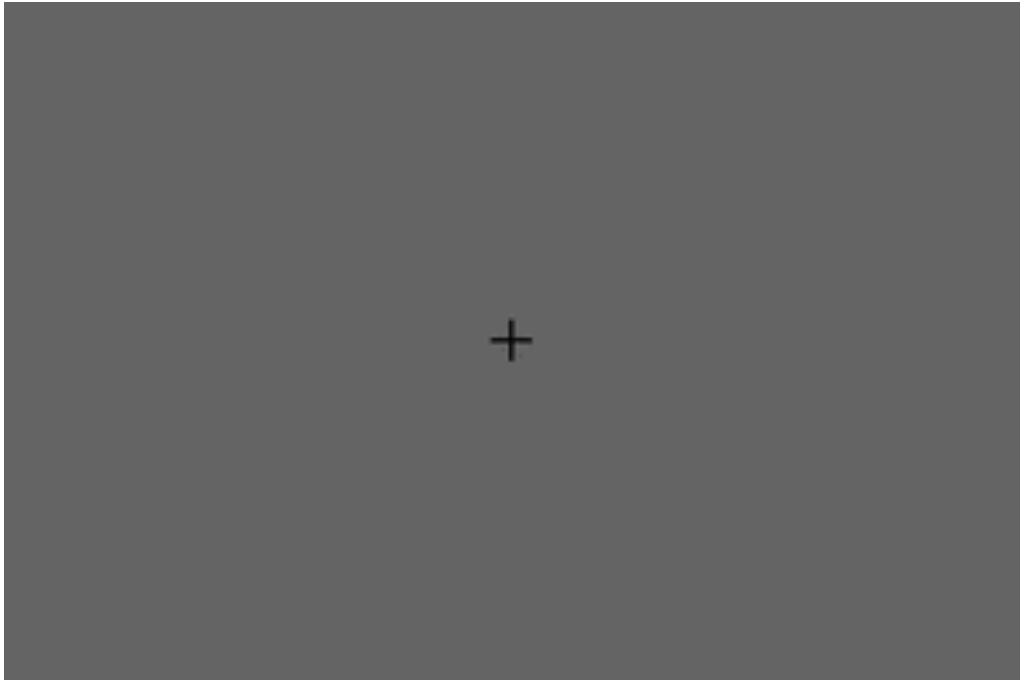
Irving Biederman





I. Biederman, *Science*, 1972

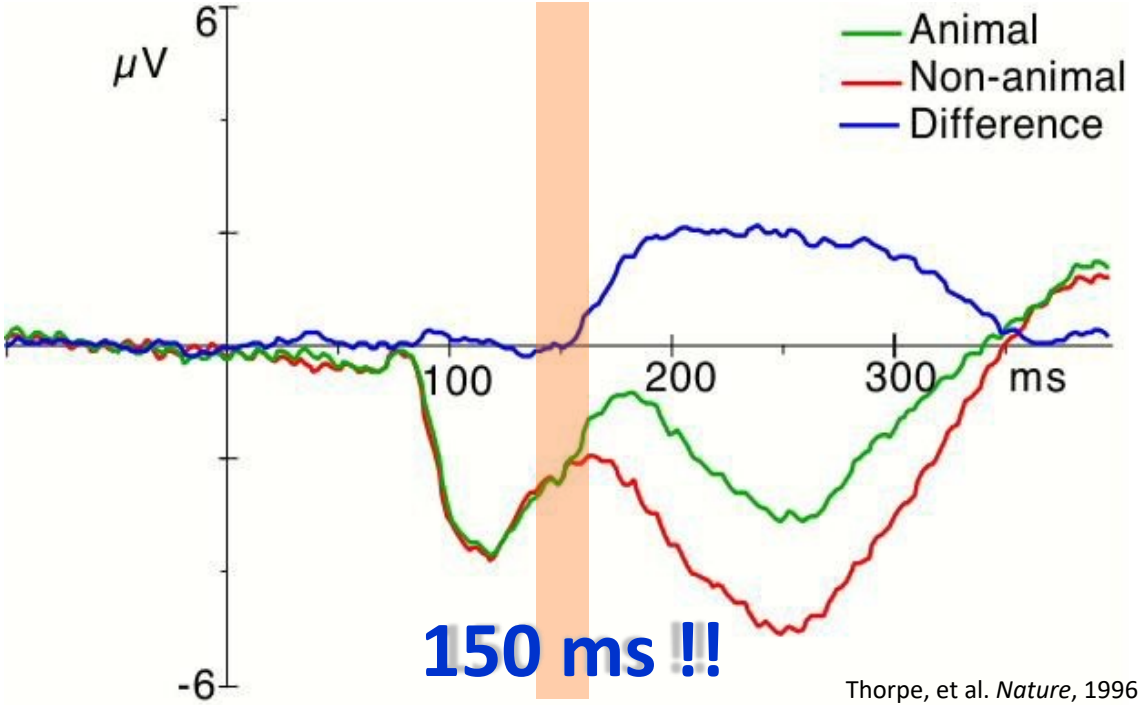
Rapid Serial Visual Perception (RSVP)



Potter, etc. 1970s

Speed of processing in the human visual system

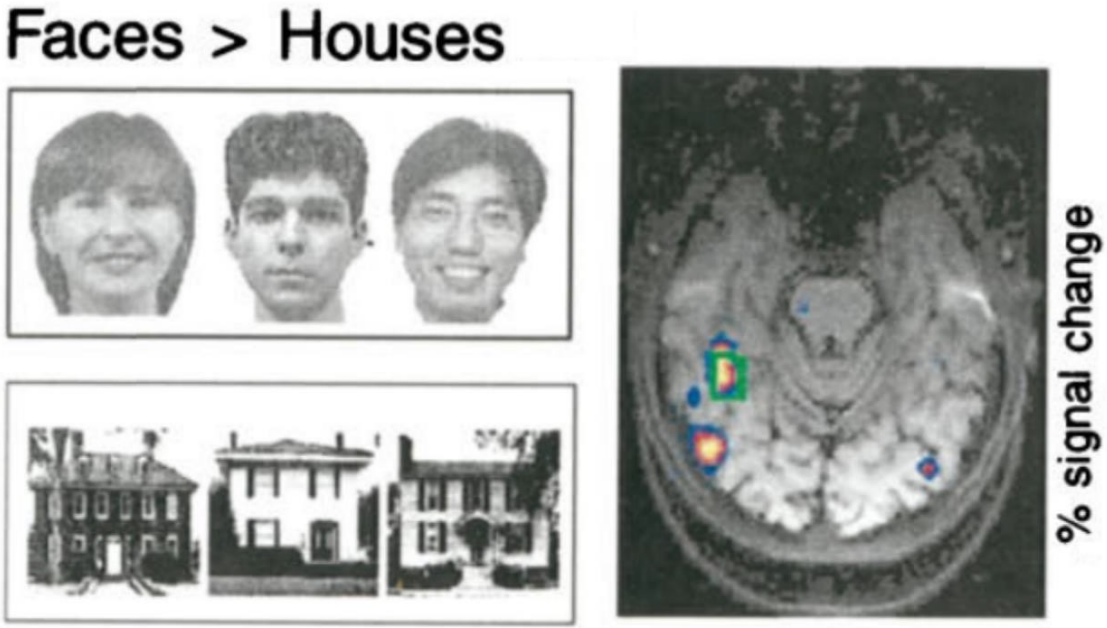
Simon Thorpe, Denis Fize & Catherine Marlot



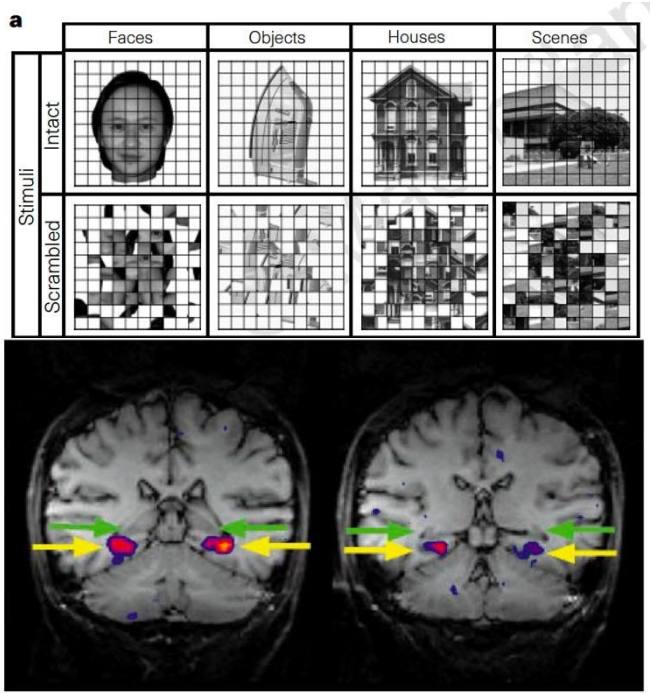
150 ms !!

Thorpe, et al. *Nature*, 1996

Neural correlates of object & scene recognition



Kanwisher et al. J. Neuro. 1997
Top image on left source: CC-BY-NC-SA 4.0
Bottom image on left source: Wikimedia public domain



Epstein & Kanwisher, Nature, 1998

Visual recognition is a fundamental task for visual intelligence

Recognition via Grouping (1990s)

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
1997 Norm. Cuts
AI Winter

Normalized Cuts, Shi and Malik, 1997

Slide inspiration: Justin Johnson

Recognition via Matching (2000s)

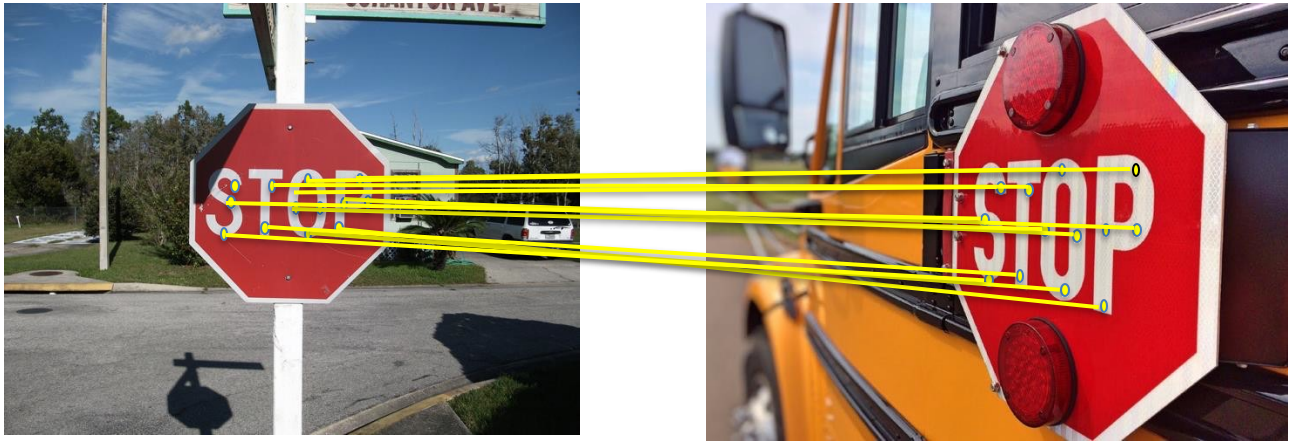
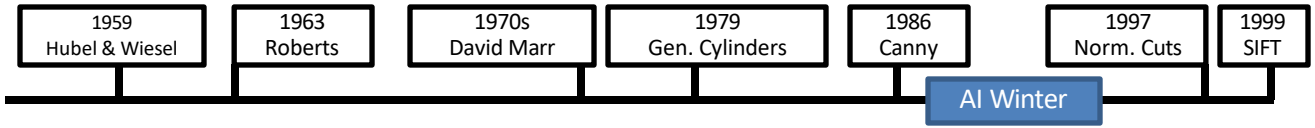


Image is public domain

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

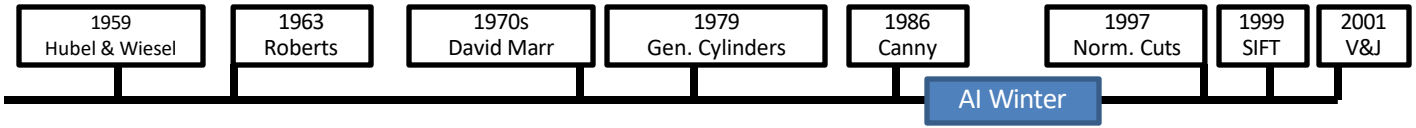
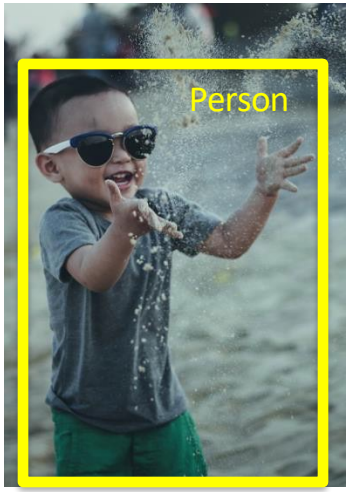
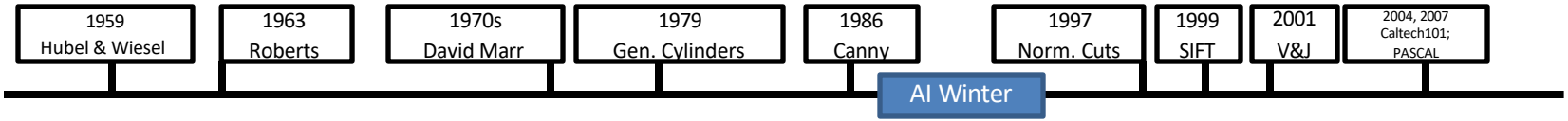


Image source: Wikimedia public domain

Caltech 101 images

PASCAL Visual Object Challenge



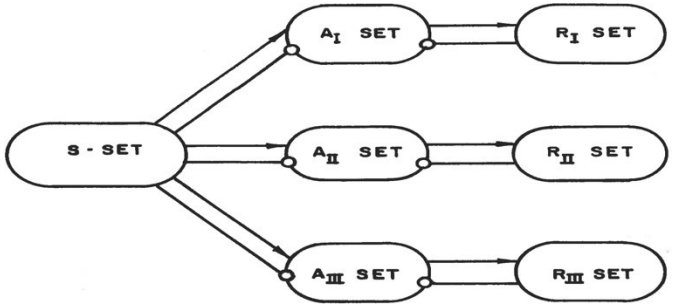


Perceptron

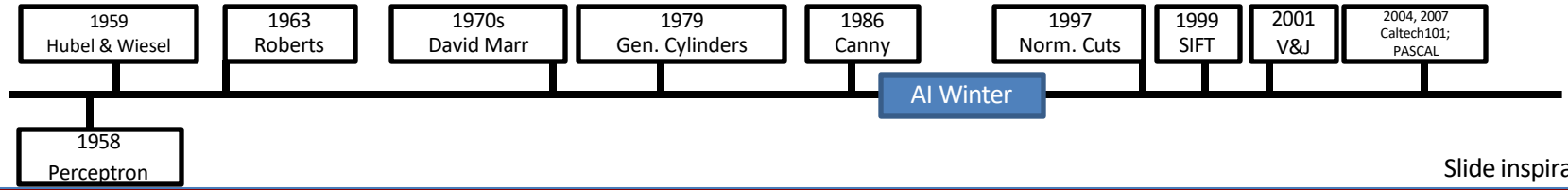
Learning representations by back-propagating errors

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



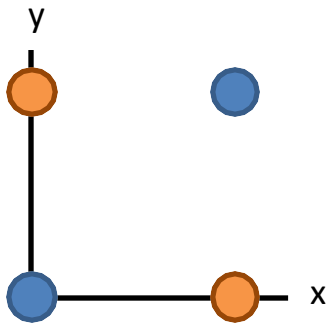
Frank Rosenblatt, ~1957



Slide inspiration: Justin Johnson

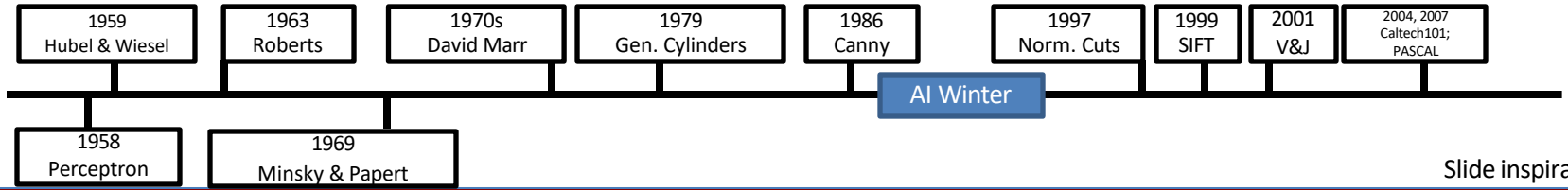
Minsky and Papert, 1969

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



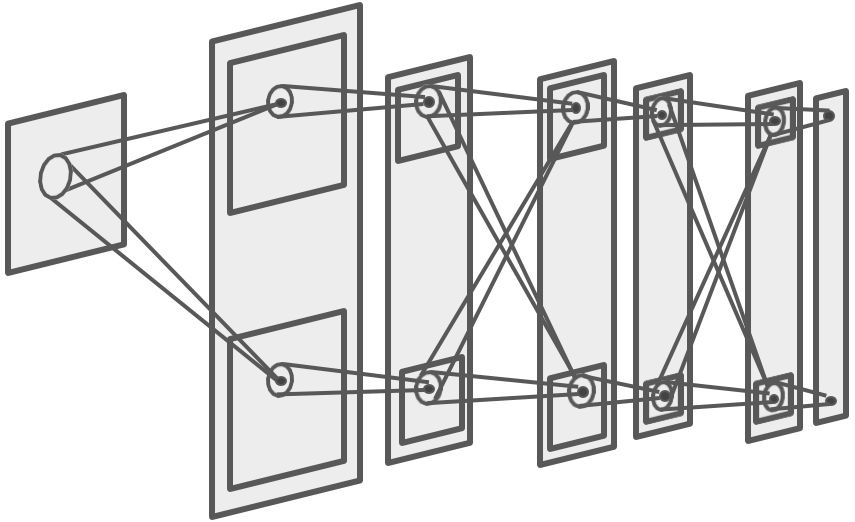
Slide inspiration: Justin Johnson

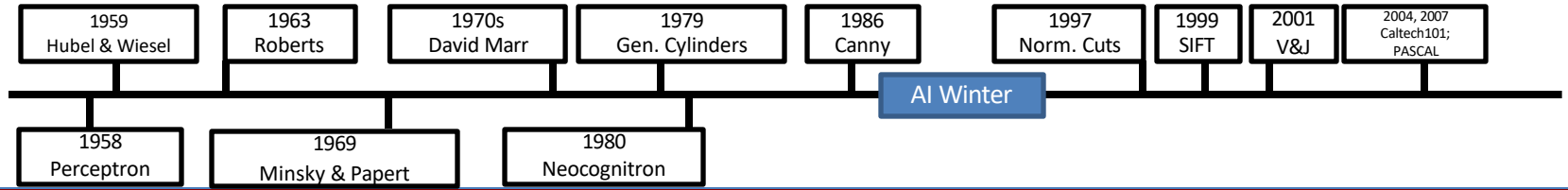
Neocognitron: Fukushima, 1980

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





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

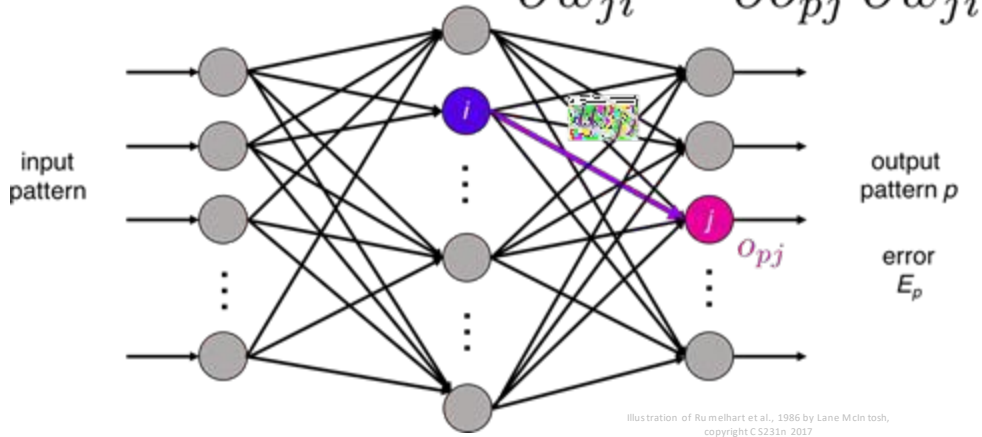
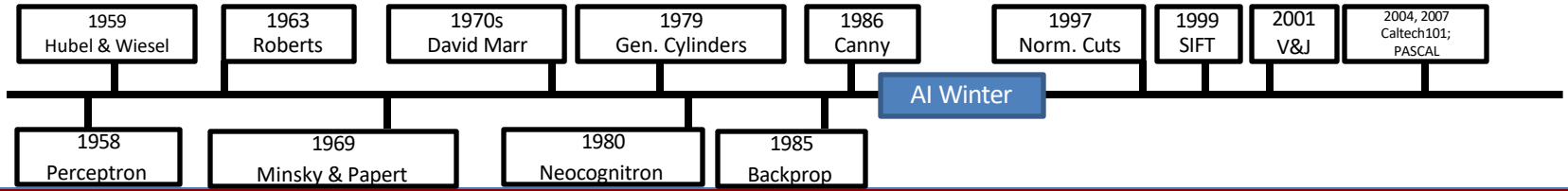
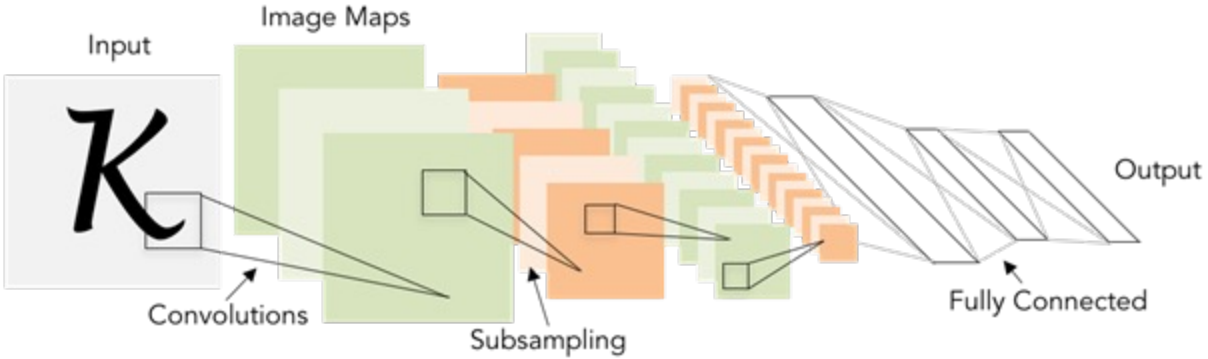


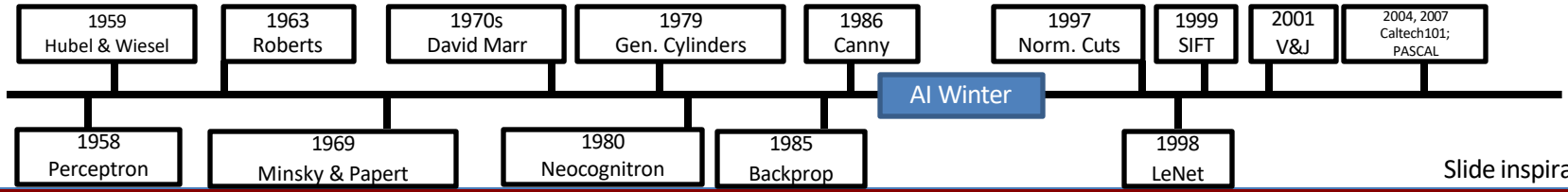
Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017



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!



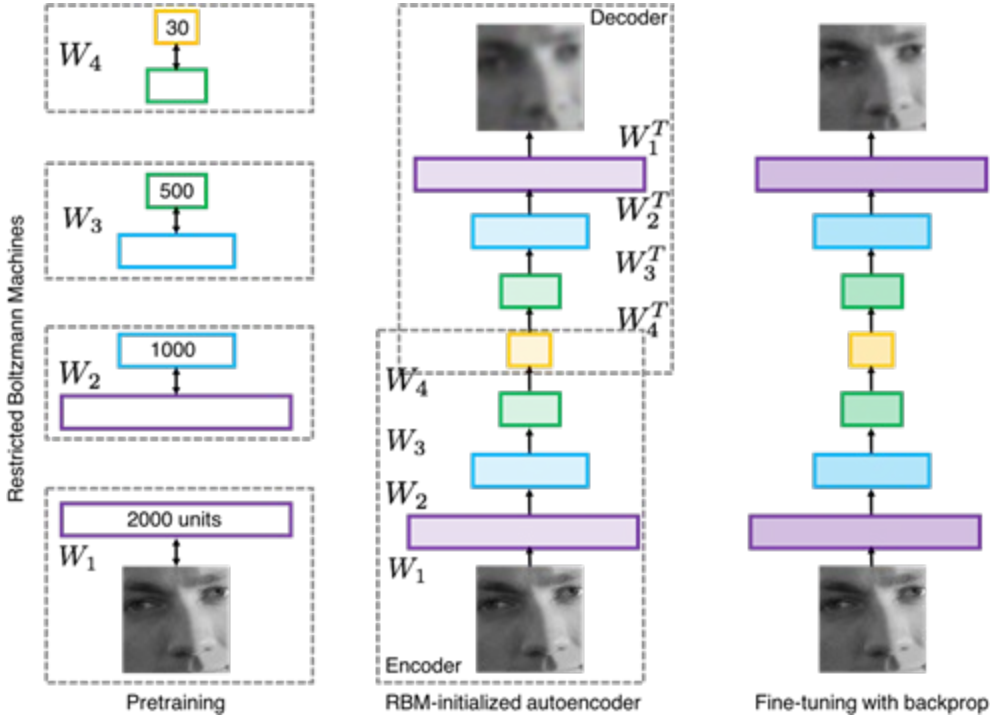
Slide inspiration: Justin Johnson

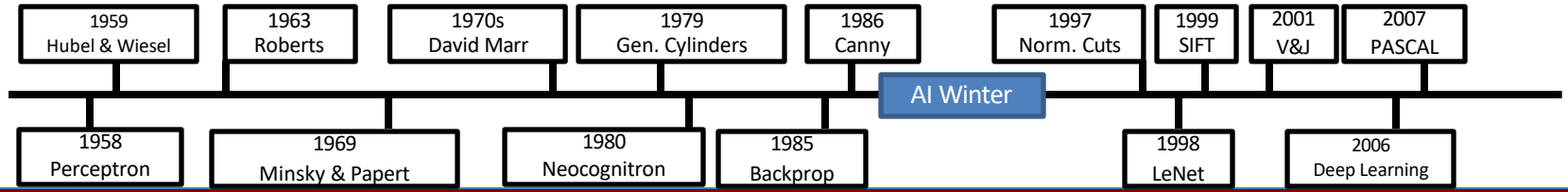
2000s: “Deep Learning”

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

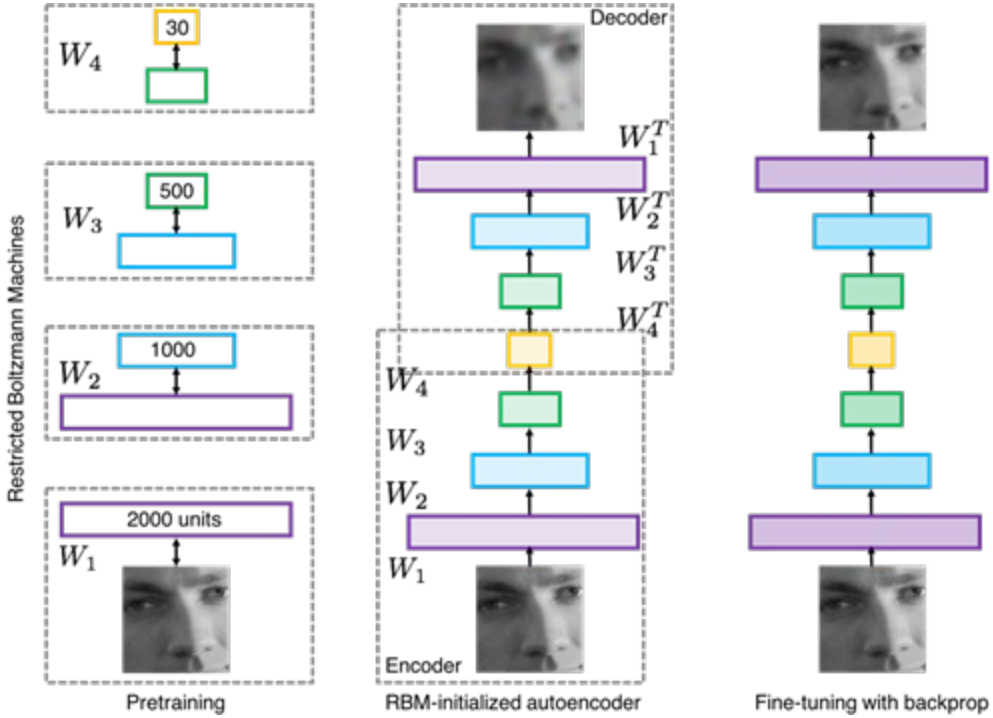
2000s: “Deep Learning”

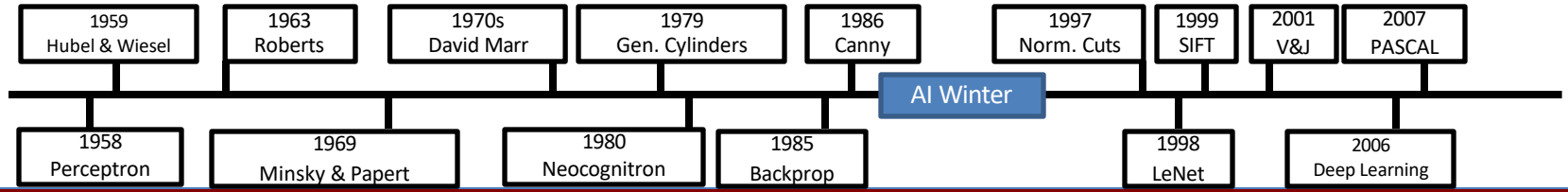
People tried to train neural networks that were deeper and deeper

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

IMAGENET Large Scale Visual Recognition Challenge

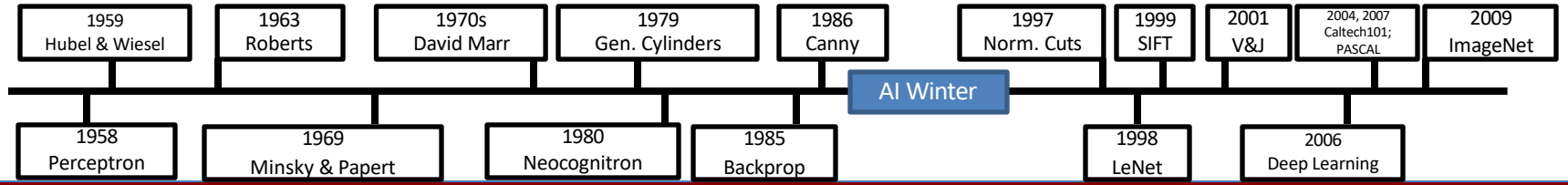
The Image Classification Challenge:

1,000 object classes

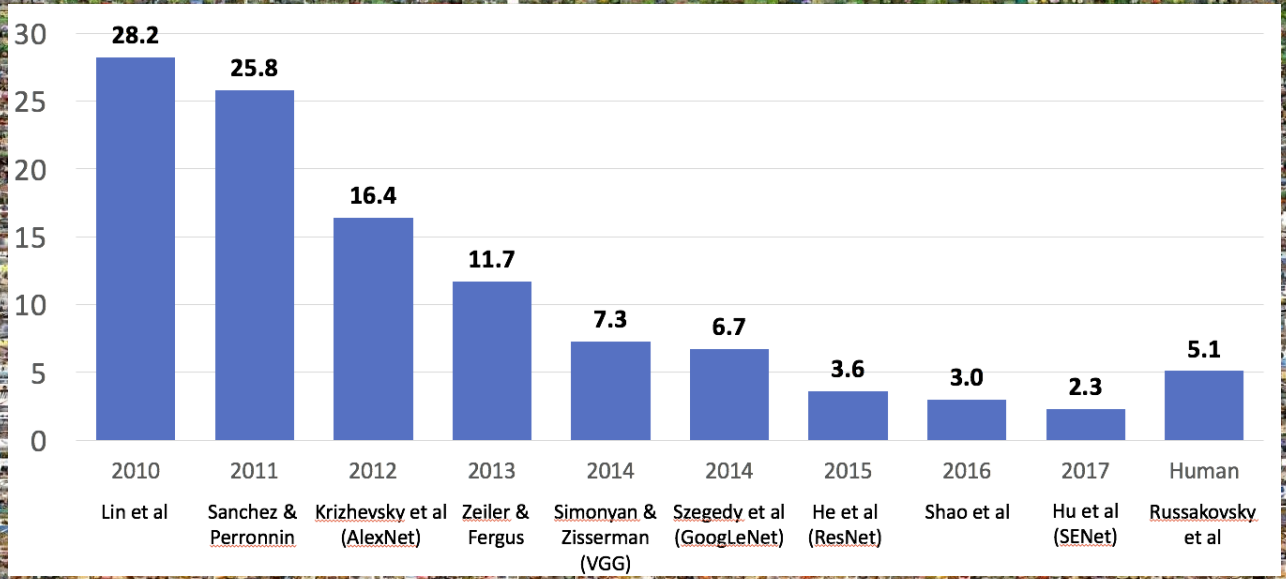
1,431,167 images

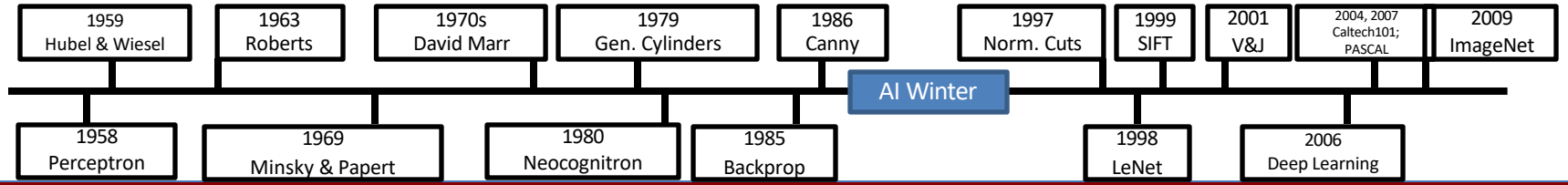
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

Deng et al, 2009
Russakovsky et al. IJCV 2015

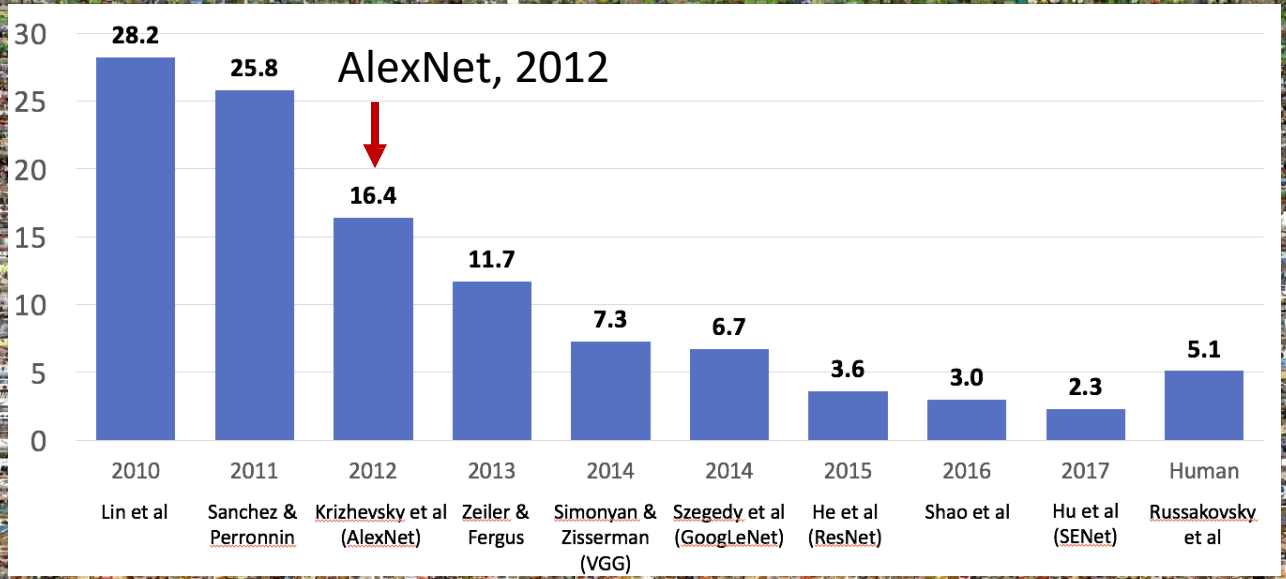


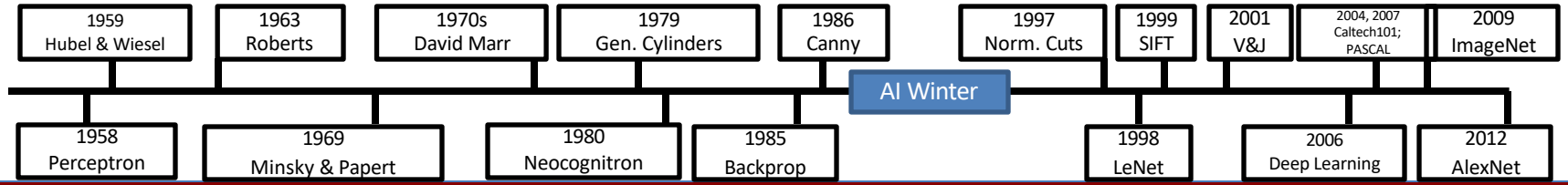
IMAGENET Large Scale Visual Recognition Challenge



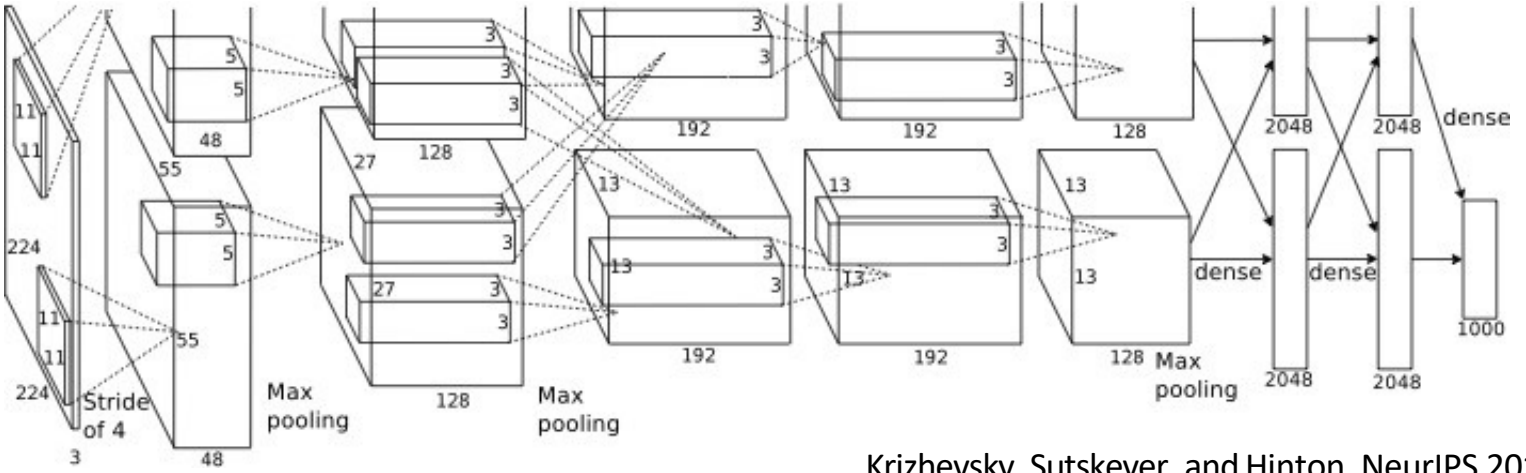


IMAGENET Large Scale Visual Recognition Challenge

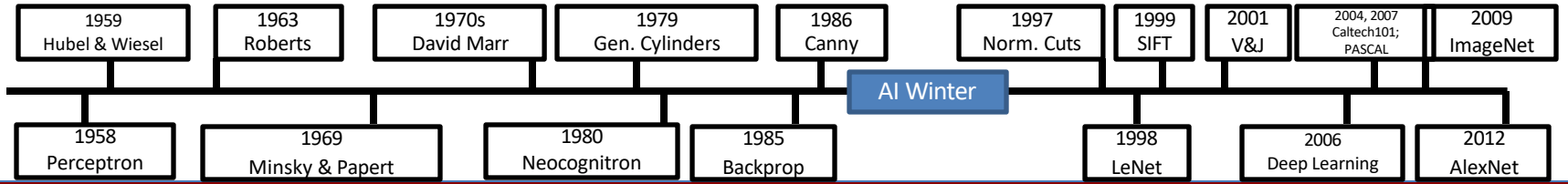




AlexNet: Deep Learning Goes Mainstream

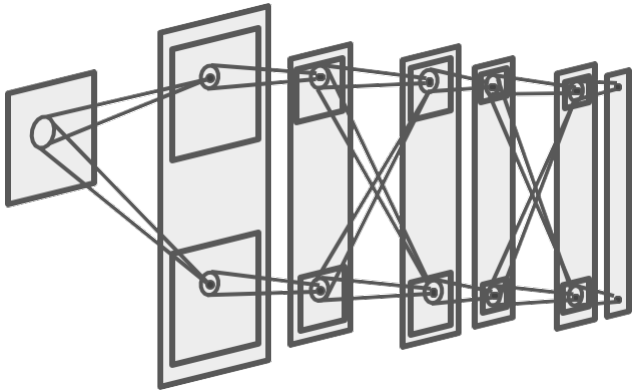


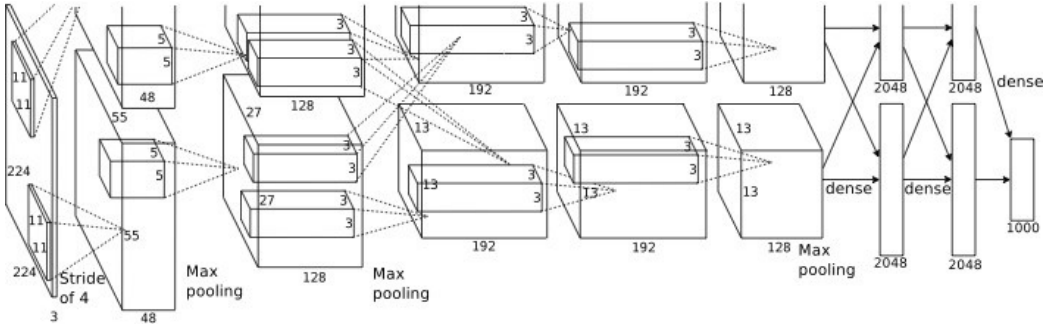
Krizhevsky, Sutskever, and Hinton, NeurIPS 2012



Slide inspiration: Justin Johnson

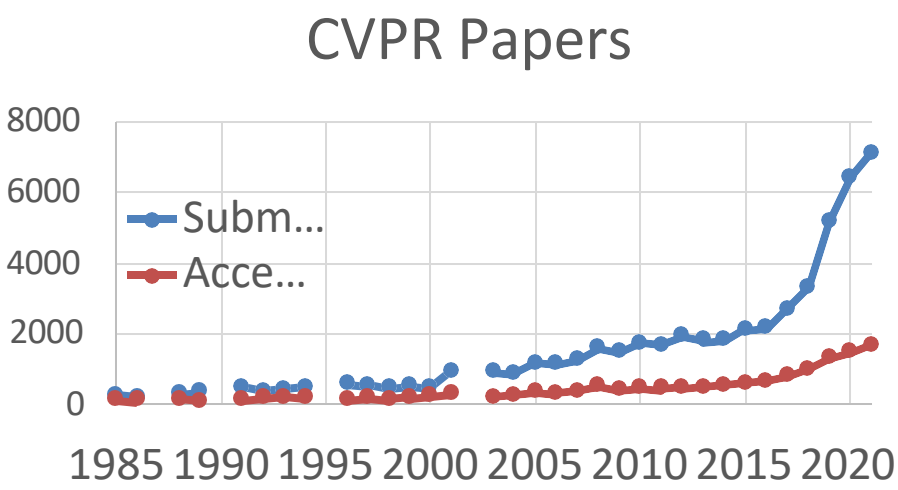
AlexNet vs. Neocognitron: 32 years apart



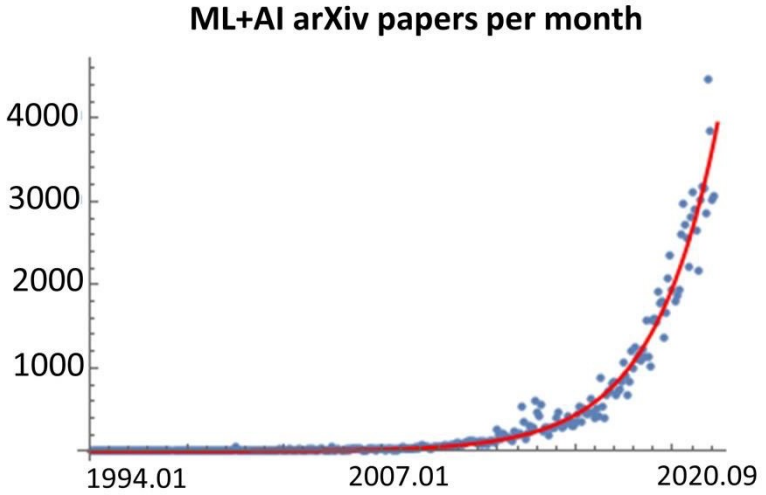


Slide inspiration: Justin Johnson

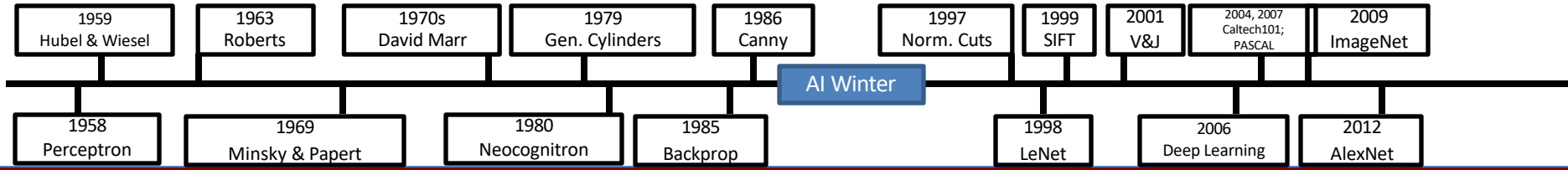
2012 to Present: Deep Learning Explosion



Publications at top Computer Vision conference



arXiv papers per month (source)



Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

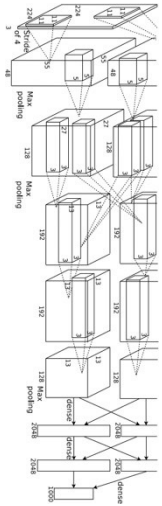
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

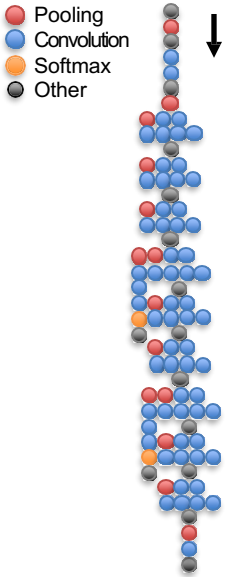
SuperVision



[Krizhevsky NIPS 2012]

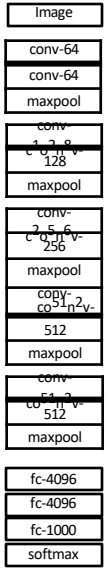
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

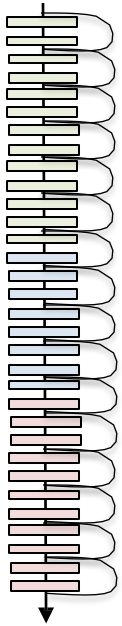
VGG



[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

2012 to Present: Deep Learning is Everywhere

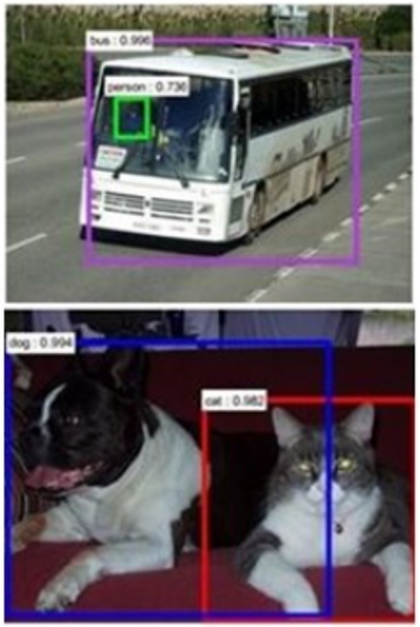
Image Classification

Image Retrieval

Slide inspiration: Justin Johnson

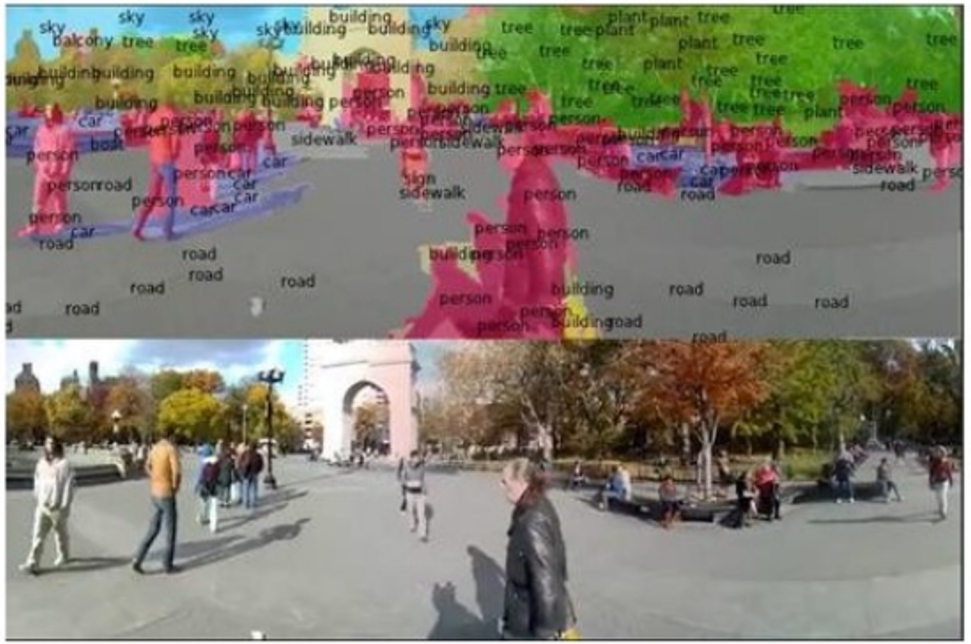
2012 to Present: Deep Learning is Everywhere

Object Detection



Figures copyright Shaoqing Ren, Kaiming He, Ross Girschick, Jian Sun, 2015. Reproduced with permission.
Images adapted from above source

Image Segmentation



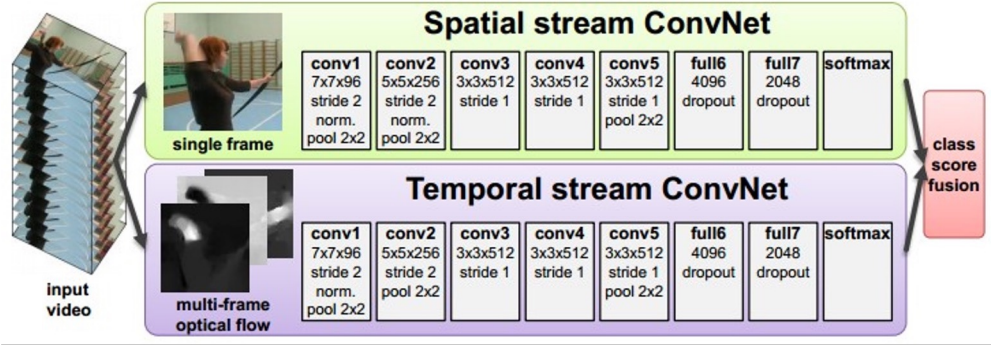
Figures copyright Clement Farabet, 2012. Reproduced with permission.

[Farabet et al., 2012]

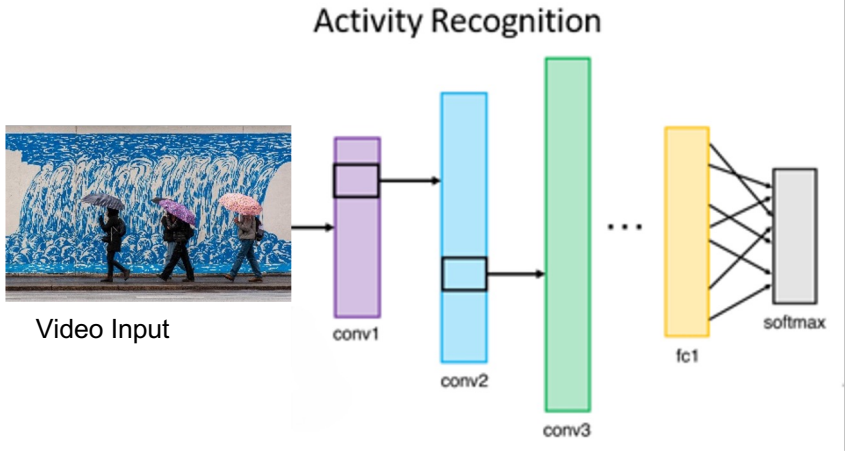
Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Video Classification



Simonyan et al, 2014

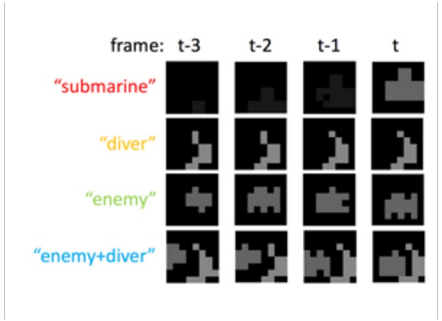


Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Pose Recognition (Toshev and Szegedy, 2014)

Playing Atari games (Guo et al, 2014)



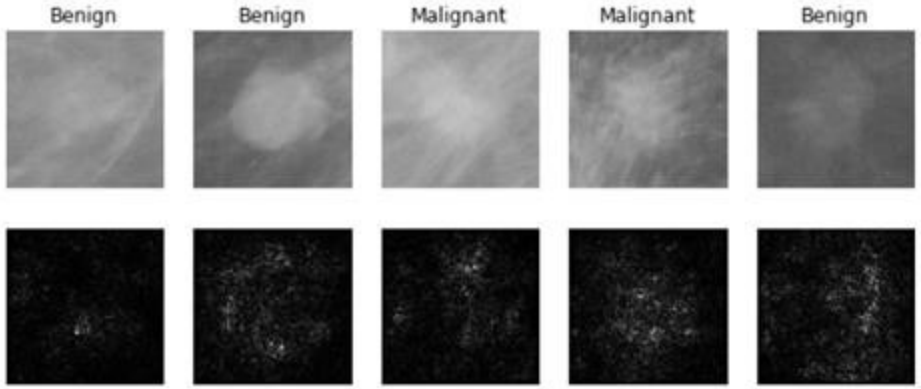
ATARI®

Logo source: Wikimedia Commons Public Domain

Slide inspiration: Justin Johnson

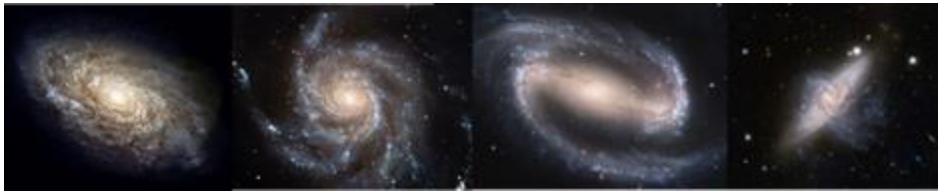
2012 to Present: Deep Learning is Everywhere

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

Whale recognition

Kaggle Challenge

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

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Image Captioning

Vinyals et al, 2015
Karpathy and , 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

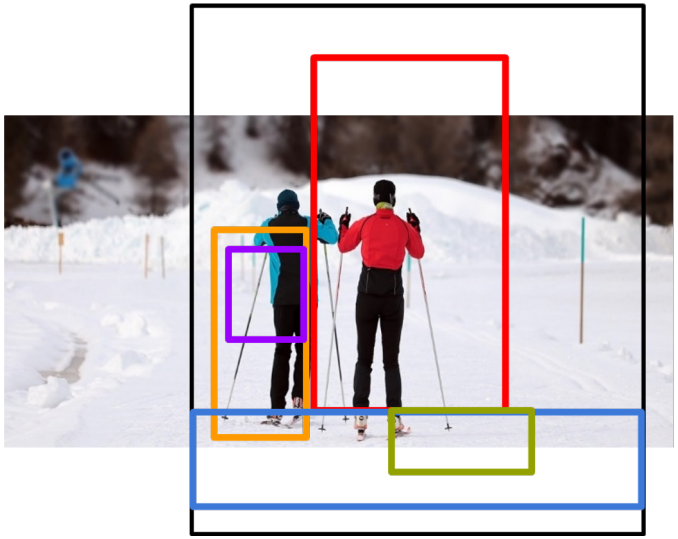
A cat sitting on a suitcase on the floor

A woman standing on a beach holding a surfboard

Captions generated by Justin Johnson using Neuraltalk2

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere



Results:
spatial, comparative, asymmetrical, verb, prepositional

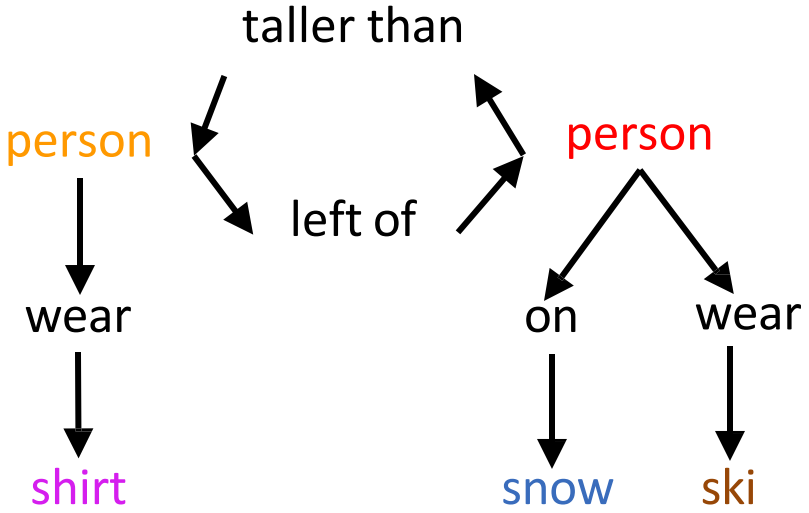
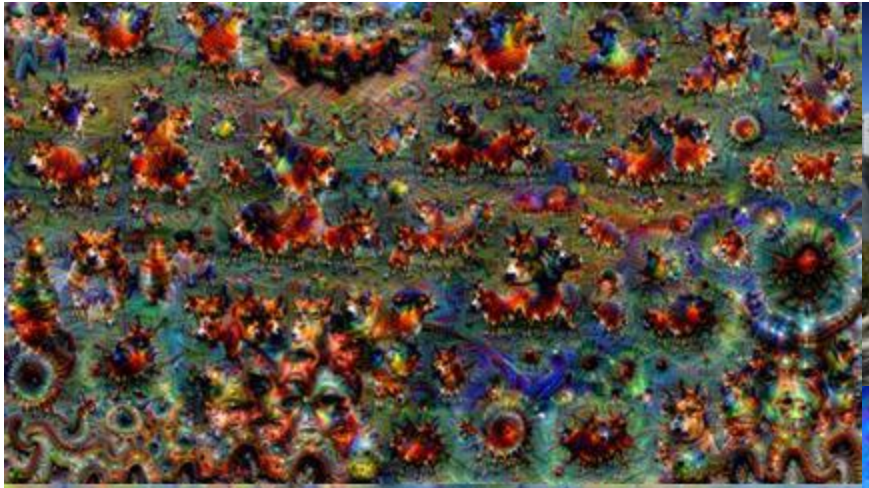
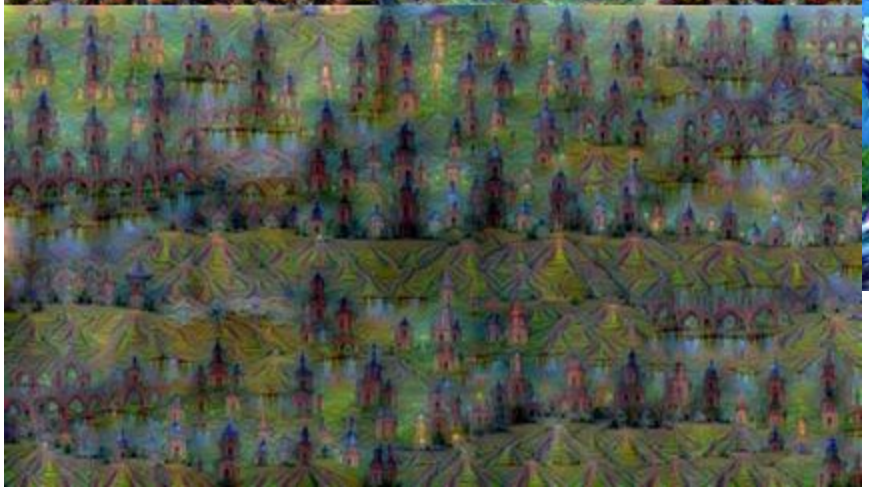


Image has been adapted

Krishna*, Lu*, Bernstein, , *ECCV* 2016





Original image is CC0 public domain
Starry Night and Tree Roots by Van Gogh are in the public domain
Bokeh image is in the publicd omain
Stylized images copyright Justin Johnson, 2017;
reproduced with per mission

Mordvinsev et al, 2015

Gatys et al, 2016

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

TEXT PROMPT

an armchair in the shape of an avocado. an armchair imitating an avocado.

AI-GENERATED IMAGES

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

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

TEXT PROMPT

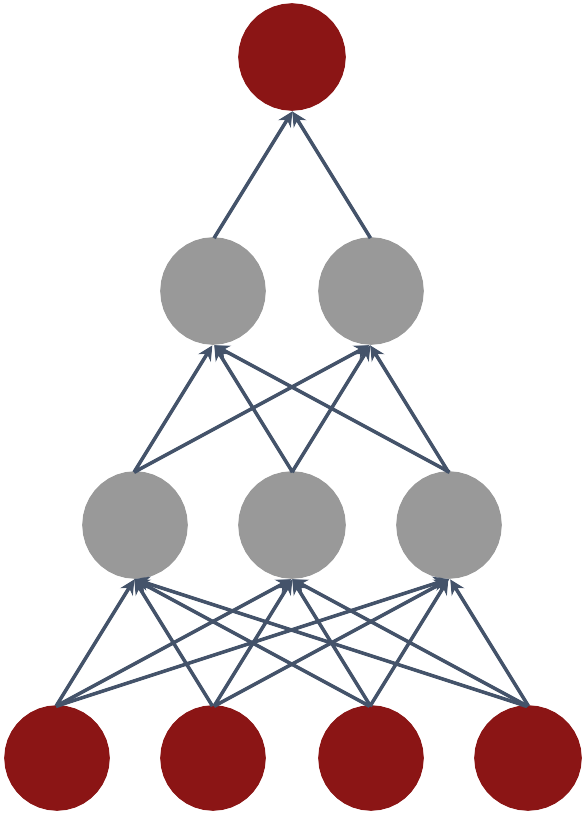
an armchair in the shape of a peach. an armchair imitating a peach.

AI-GENERATED IMAGES

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

Slide inspiration: Justin Johnson

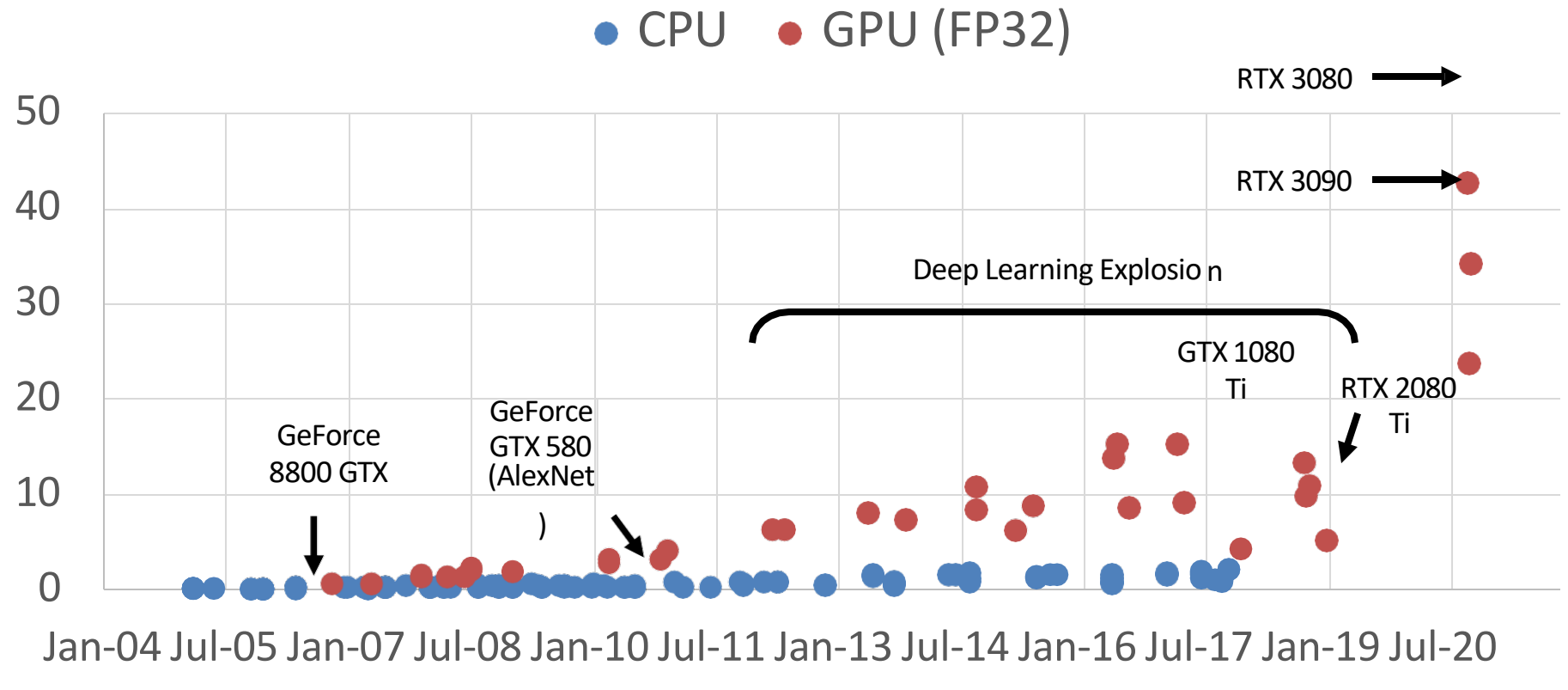
Computation



Data

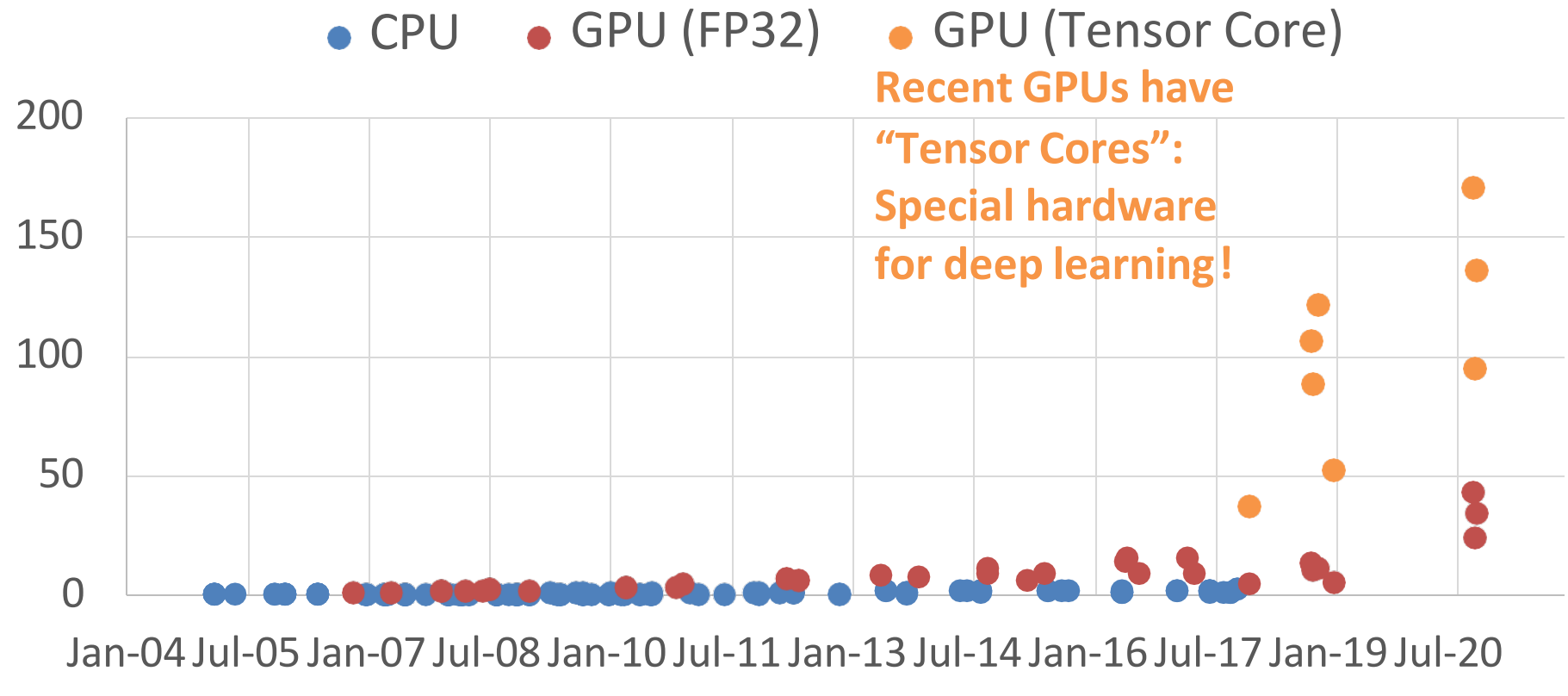
Algorithms

GFLOP per Dollar



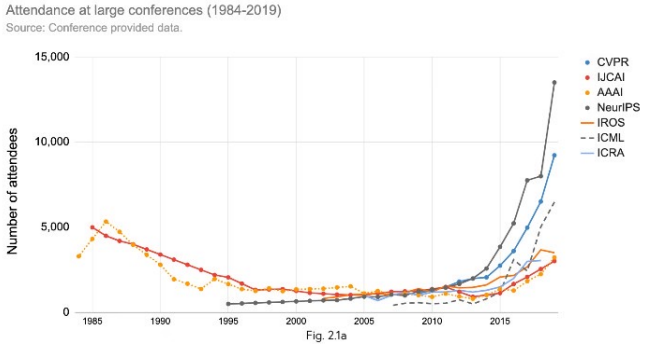
Slide inspiration: Justin Johnson

GFLOP per Dollar



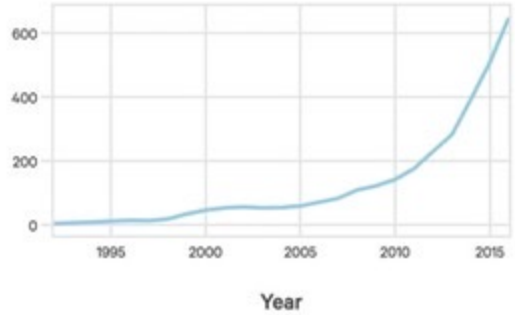
Slide inspiration: Justin Johnson

AI's Explosive Growth & Impact



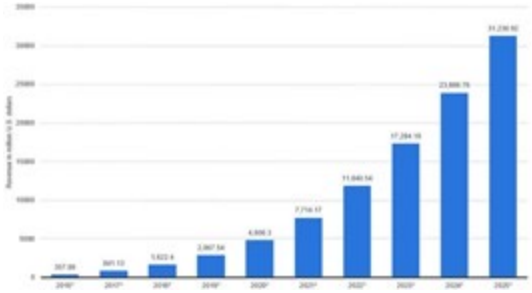
Number of attendance At AI conferences

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

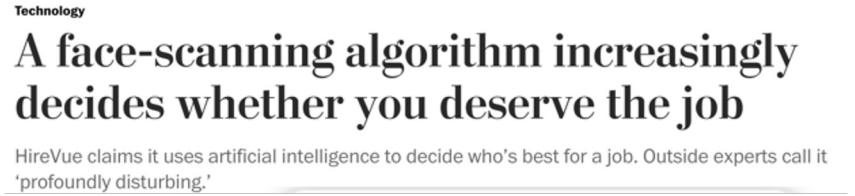
Computer Vision Can Cause Harm

Harmful Stereotypes

https://www.youtube.com/watch?app=desktop&v=fMym_BKWQzk

The Trouble with Bias

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.'

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://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

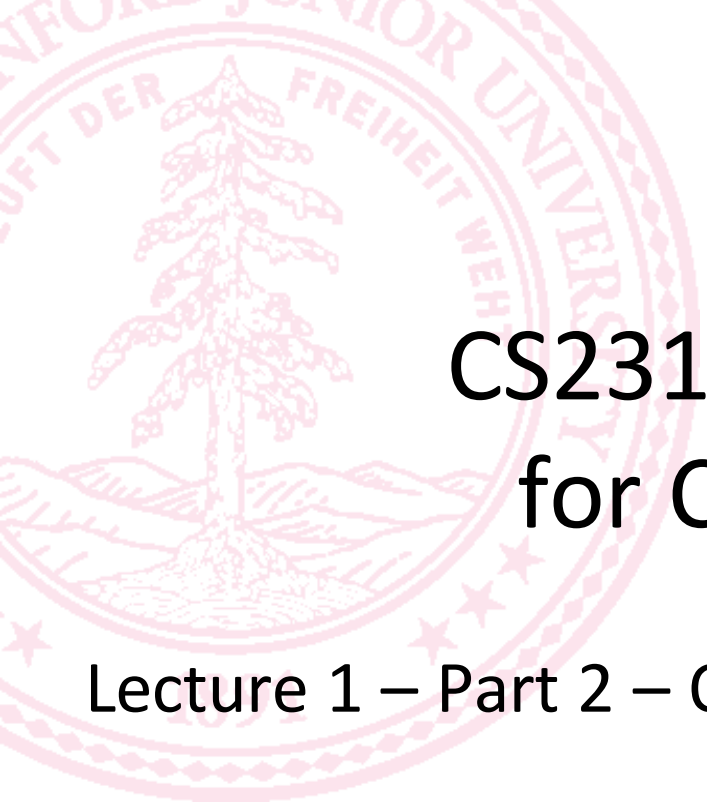
Example Credit: Timnit Gebru

Computer Vision Can Save Lives

Slide inspiration: Justin Johnson

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

Slide inspiration: Andrej Karpathy



CS231n: Deep Learning for Computer Vision

Lecture 1 – Part 2 – Overview

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Deep Learning Basics

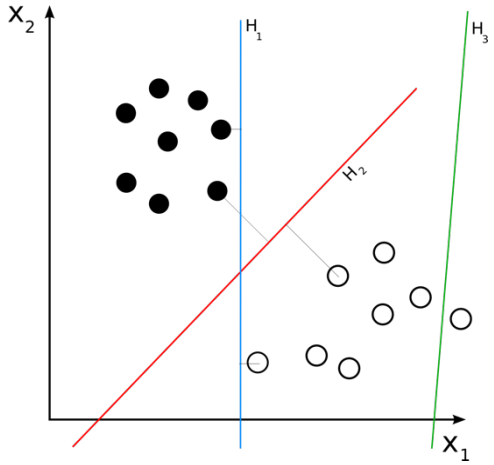
- Image Classification: A core task in Computer Vision

→ cat

Deep Learning Basics

- Image Classification: A core task in Computer Vision

cat

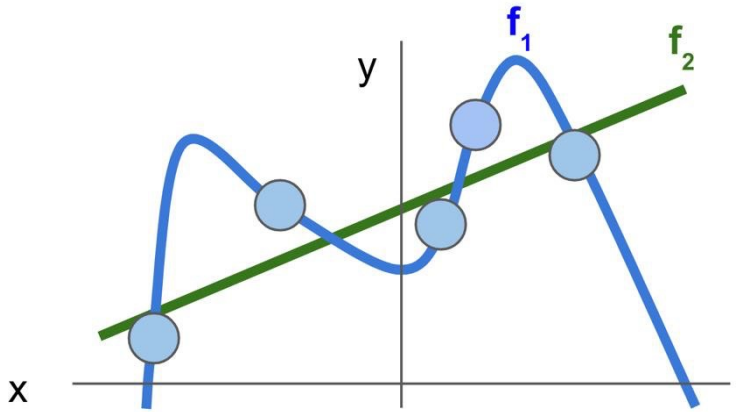


Linear Classifier

Deep Learning Basics

- Image Classification: A core task in Computer Vision

cat

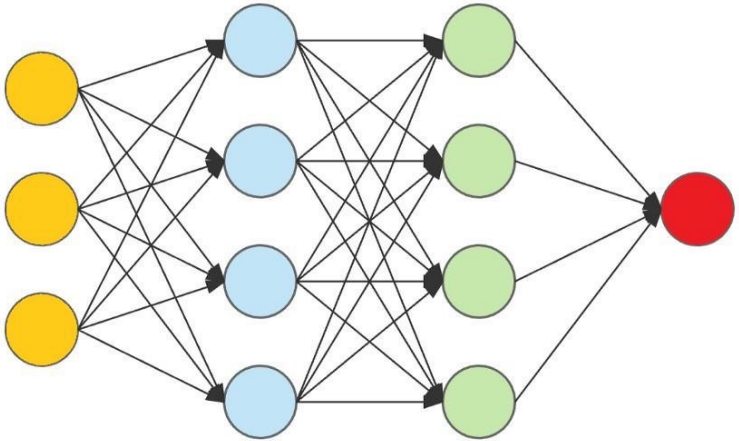


Regularization & Optimization

Deep Learning Basics

- Image Classification: A core task in Computer Vision

cat



Neural Networks

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Perceiving and Understanding the Visual World

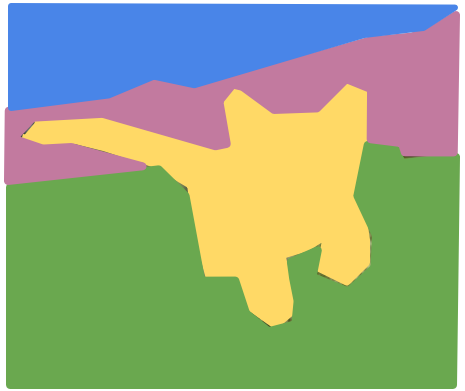
Tasks Beyond Image Classification

Classification

CAT

No spatial extent

Semantic Segmentation



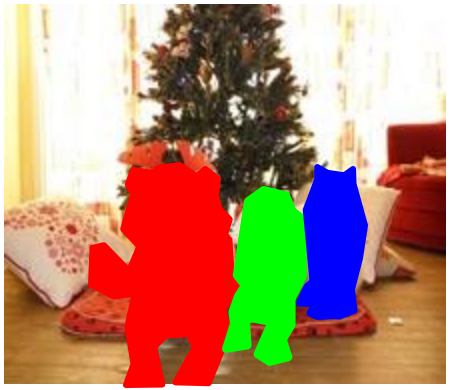
GRASS, CAT, TREE, SKY

No objects, just pixels

Object Detection

DOG, DOG, CAT

Instance Segmentation



DOG, DOG, CAT

Multiple Object

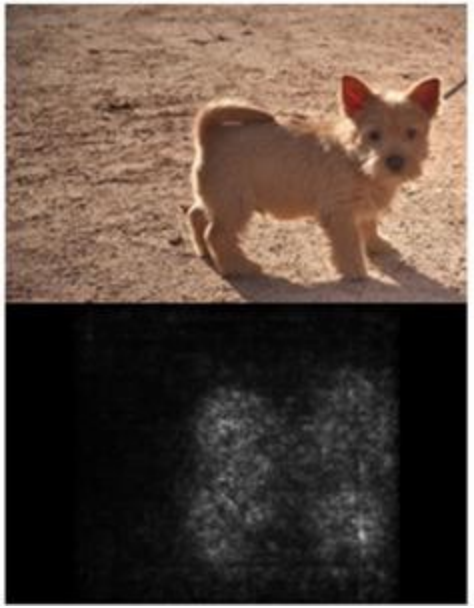
Tasks Beyond Image Classification

Video Classification

Running? Jumping?

Multimodal Video Understanding

Visualization & Understanding



Models Beyond Multi-Layer Perceptron

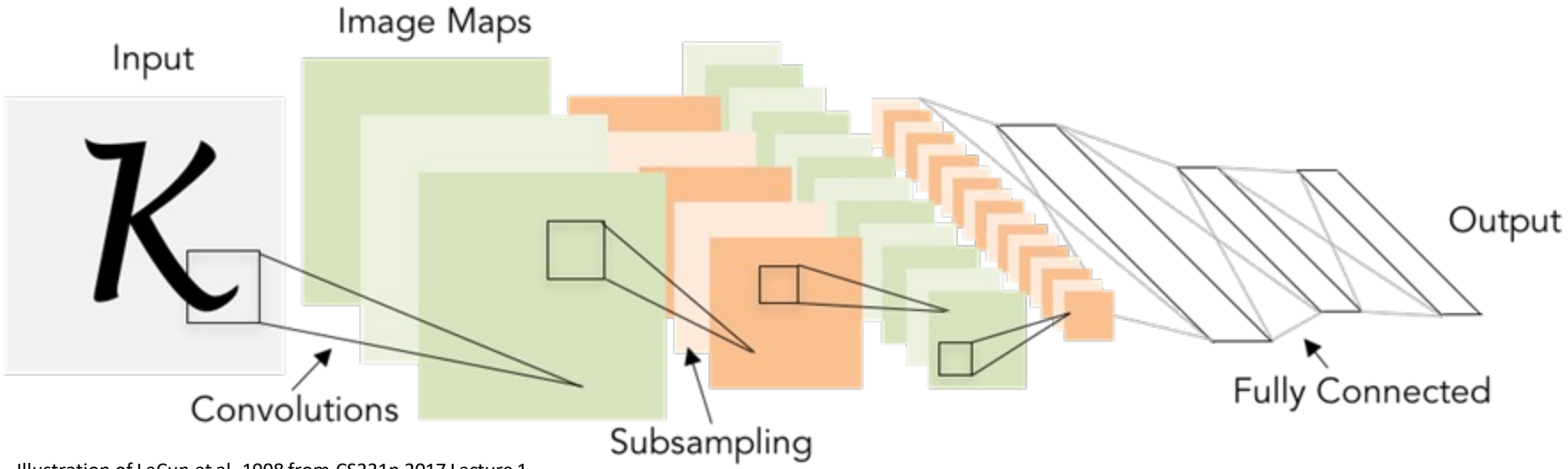
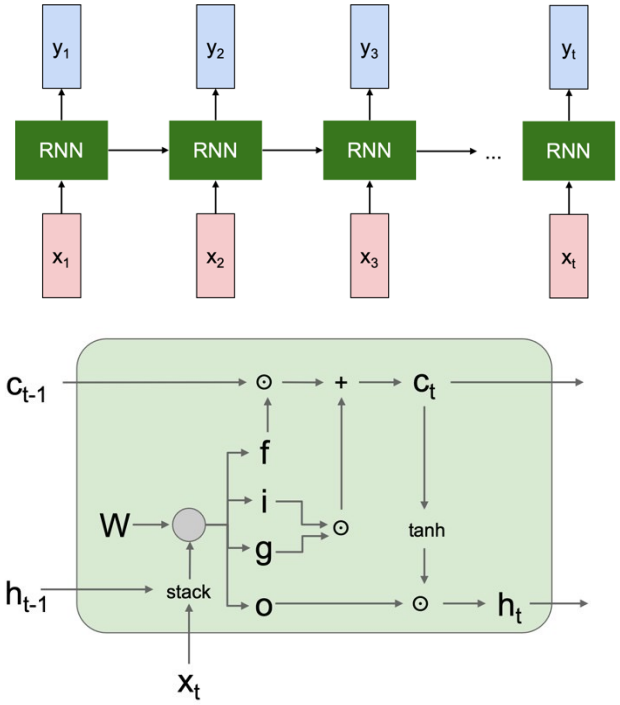


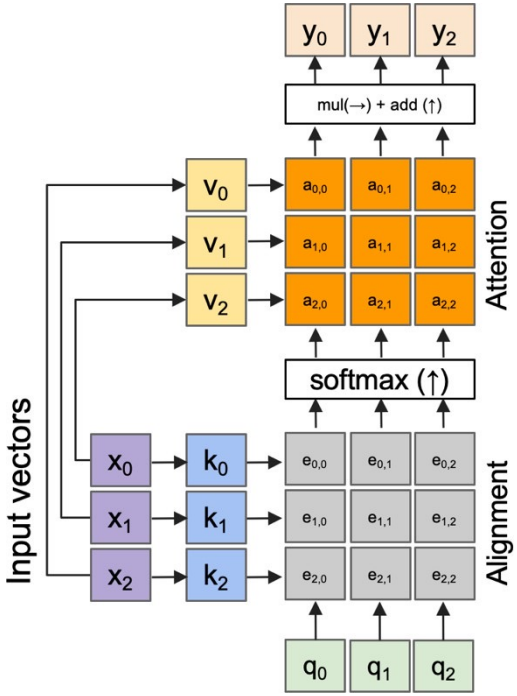
Illustration of LeCun et al. 1998 from CS231n 2017 Lecture 1

Convolutional neural network

Models Beyond Multi-Layer Perceptron



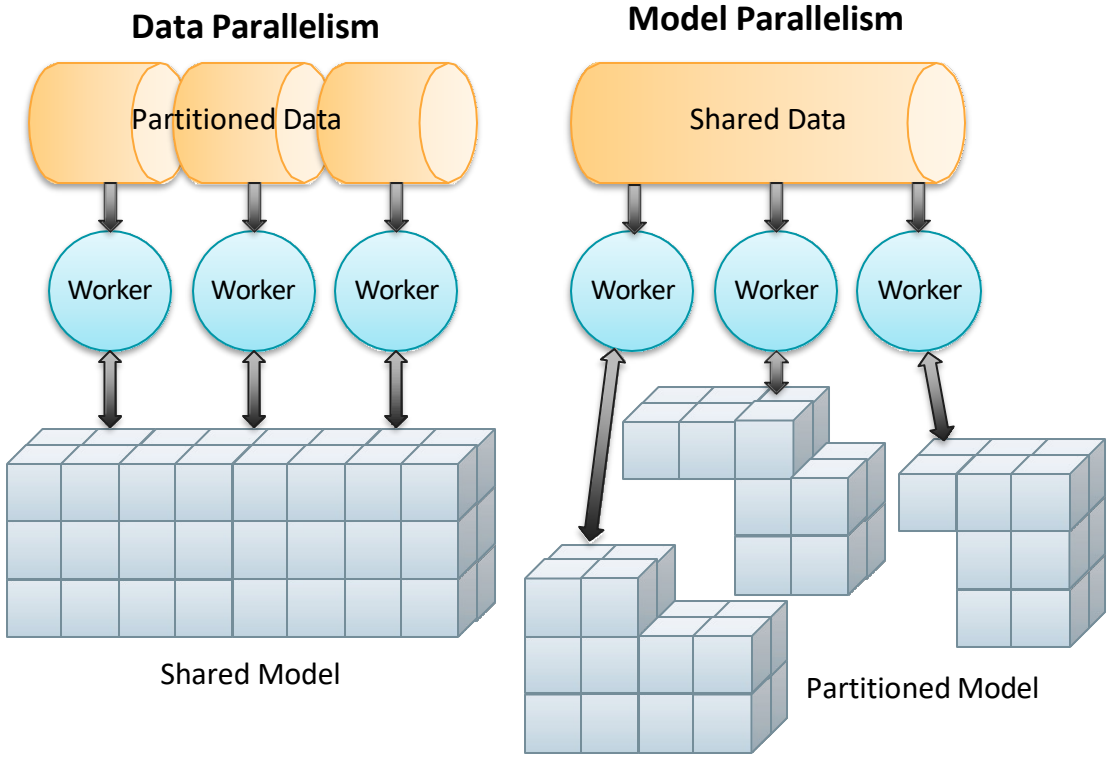
Recurrent neural network



Attention mechanism / Transformers

Large Scale Distributed Training

- Train Large Models on big datasets faster
- Scale beyond single GPU/machine limitations
- How?
  - Data Parallelism: Copy the model to all workers, split the data
  - Model Parallelism: Split model across devices
  - Synchronous vs. Asynchronous gradient updates



CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Beyond 2D Recognition

Beyond 2D Recognition: Self-supervised Learning

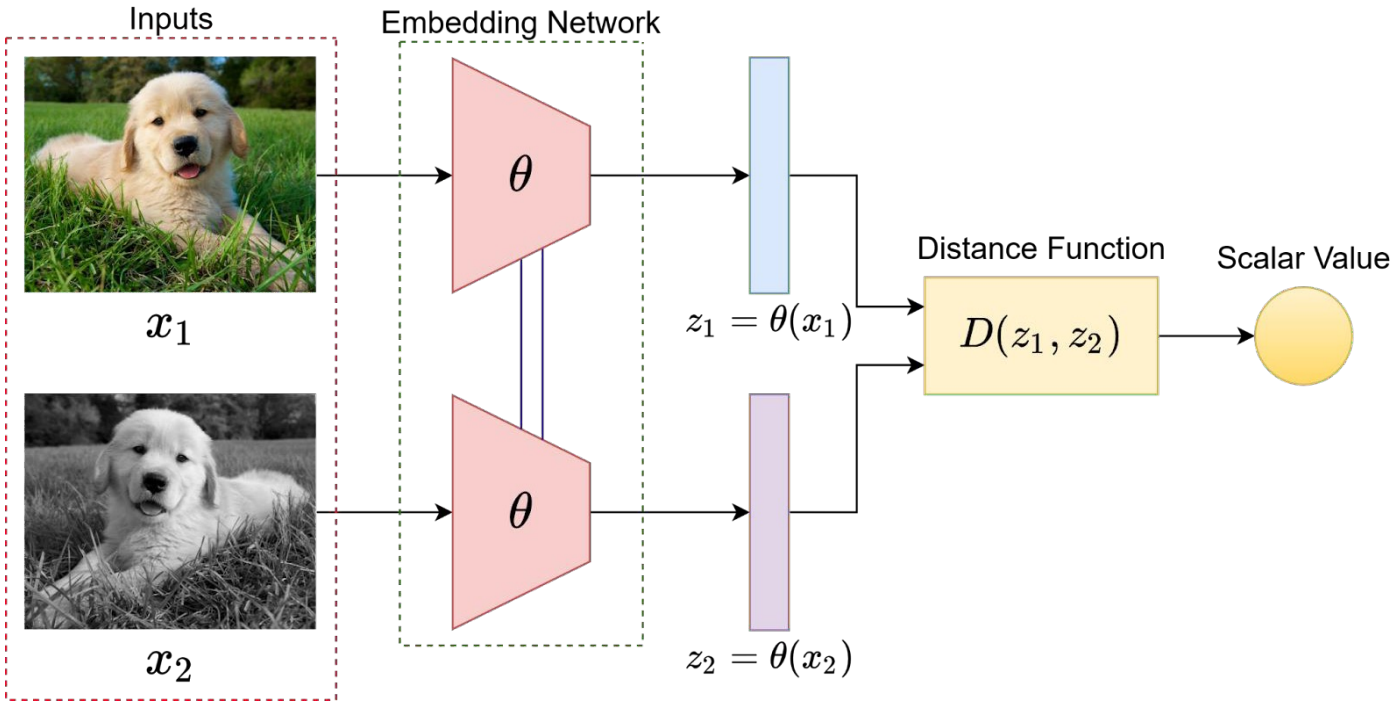


Image Courtesy of Rohit Kundu

Beyond 2D Recognition: Generative Modeling

Style Transfer

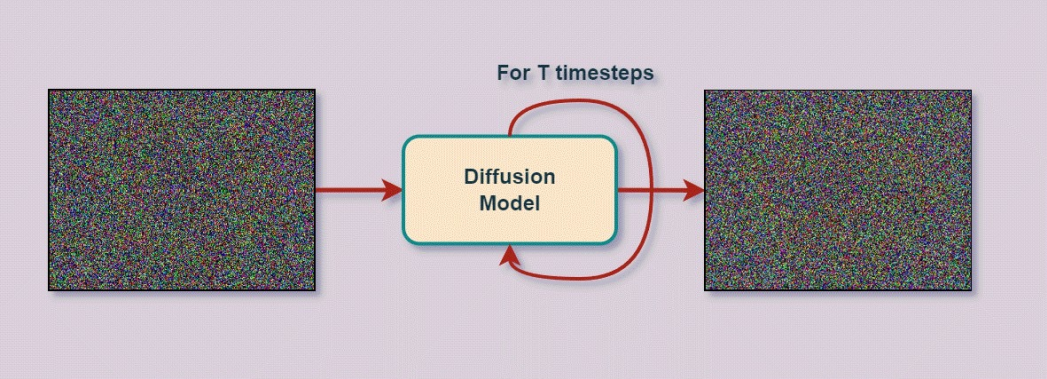
Beyond 2D Recognition: Generative Modeling

"Teddy bears working on new AI research underwater with 1990s technology"

DALL-E 2

Beyond 2D Recognition: Generative Modeling

Image Generation using Diffusion Models

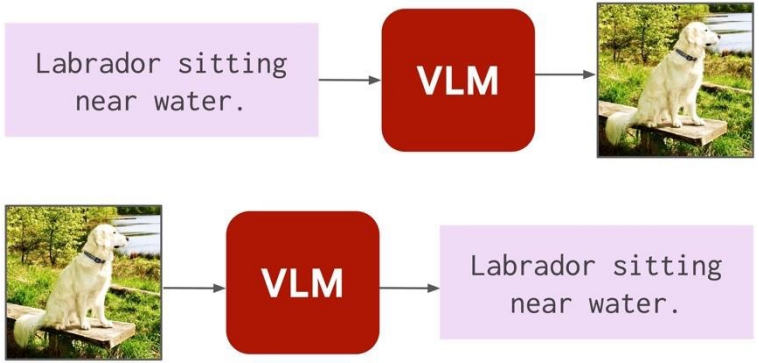


https://learnopencv.com/image-generation-using-diffusion-models/

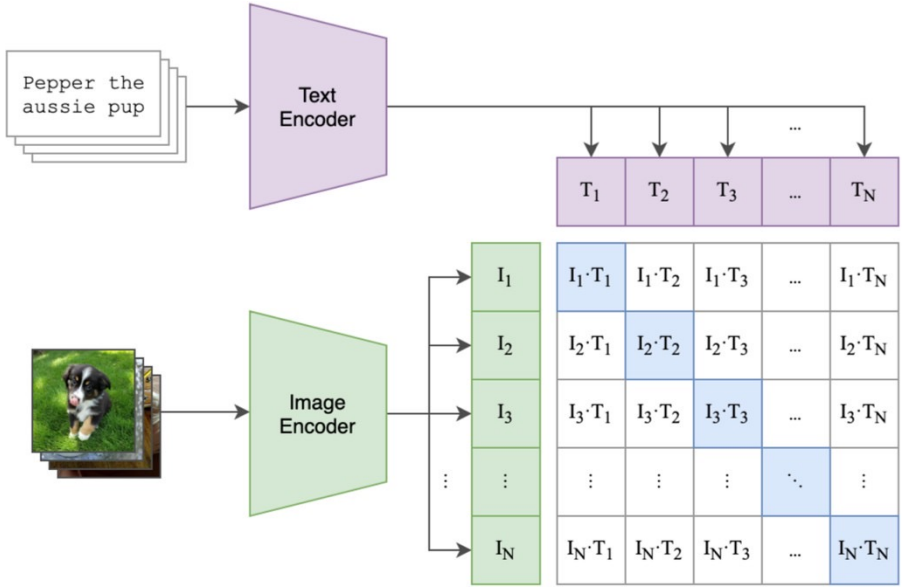
You will learn and implement a generative model in Assignment 3 that generates emojis from text inputs

Face with a cowboy hat

Beyond 2D Recognition: Vision Language Models



Yasunaga, Michihiro, et al. "Retrieval-augmented multimodal language modeling." arXiv preprint arXiv:2211.12561 (2022).

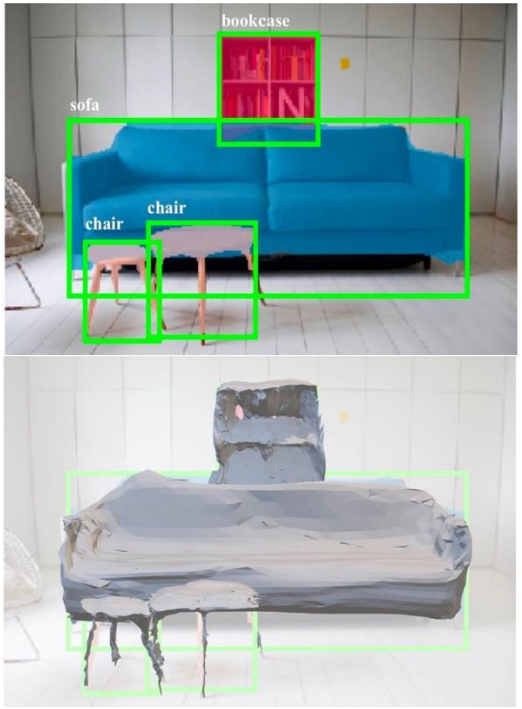


Contrastive pre-training in CLIP. The blue squares are the pairs for which we want to optimize the similarity. Image derived from https://github.com/openai/CLIP

Beyond 2D Recognition: 3D Vision

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

Zhou et al., 3D Shape Generation and Completion through Point-Voxel Diffusion (2021)



Gkioxari et al., "Mesh R-CNN", ICCV 2019

Beyond 2D Recognition: Embodied Intelligence

Li et al., BEHAVIOR-1K: A Benchmark for Embodied AI with 1,000 Everyday Activities and Realistic Simulation (2022)

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

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

2018 Turing Award for deep learning

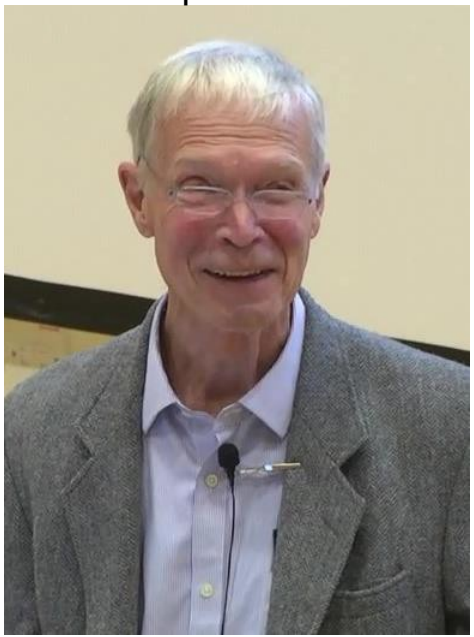
most prestigious technical award, is given for major contributions of lasting importance to computing.

2024 Nobel Prize in Physics

Hinton speaking at the Nobel Prize Lectures in Stockholm in 2024

This image is CC0 public domain

John Hopefield in 2016



This image is CC0 public domain

In 2024, he was jointly awarded the Nobel Prize in Physics with John Hopfield “for foundational discoveries and inventions that enable machine learning with artificial neural networks.”

Learning objectives

Formalize computer vision 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 vision models

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

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

- What new research has come out in the last 0-5 years?
- What are open research challenges?
- What ethical and societal considerations should we consider before deployment?

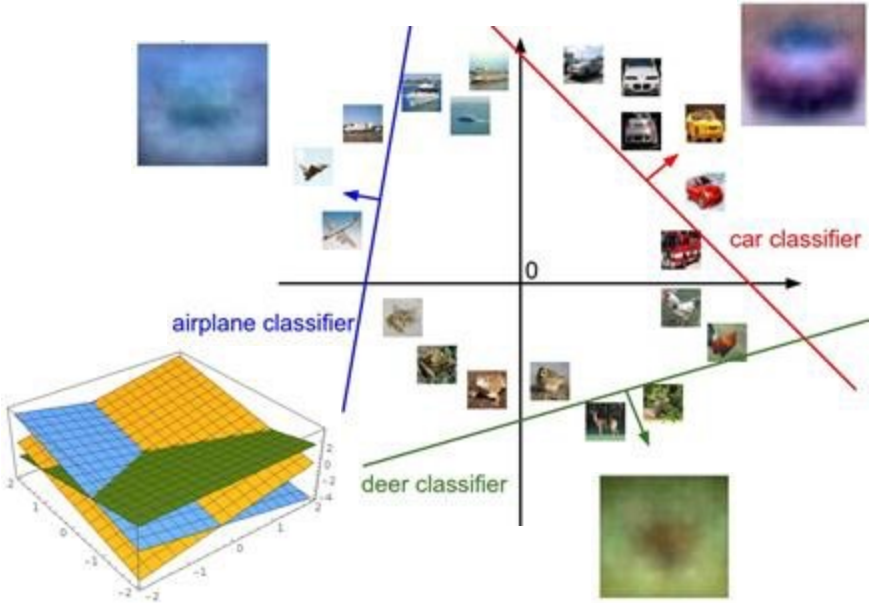
CS231n: Deep Learning for Computer Vision

- Deep Learning Basics (Lecture 2 – 4)
- Perceiving and Understanding the Visual World (Lecture 5 – 12)
- Reconstructing and Interacting with the Visual World (Lecture 13 – 17)
- Human-Centered Artificial Intelligence (Lecture 18)

Next time: Image classification with Linear Classifiers

k- nearest neighbor

Linear classification



Plot created using Wolfram Cloud