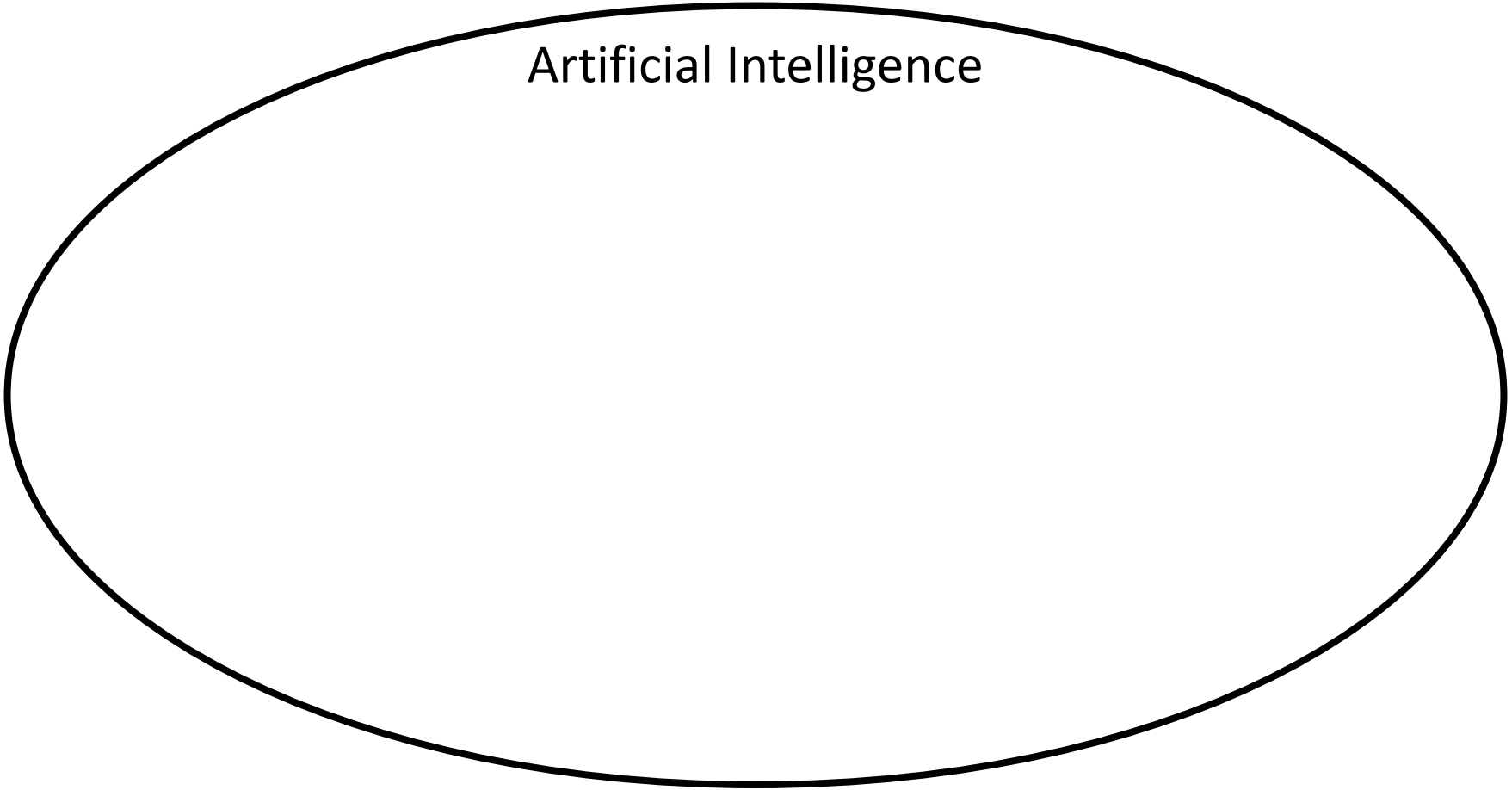
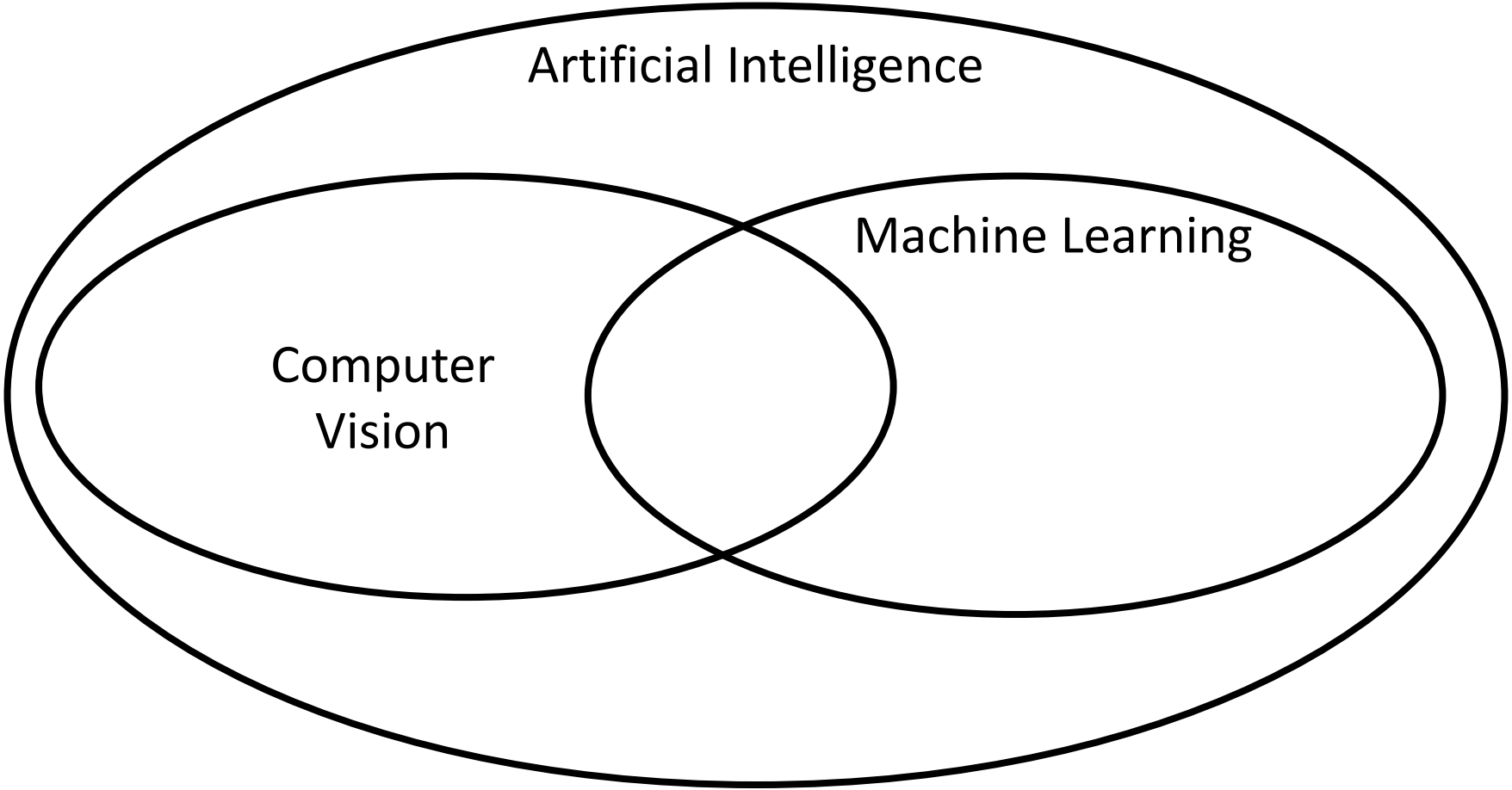
CS231n: Deep Learning for Computer Vision

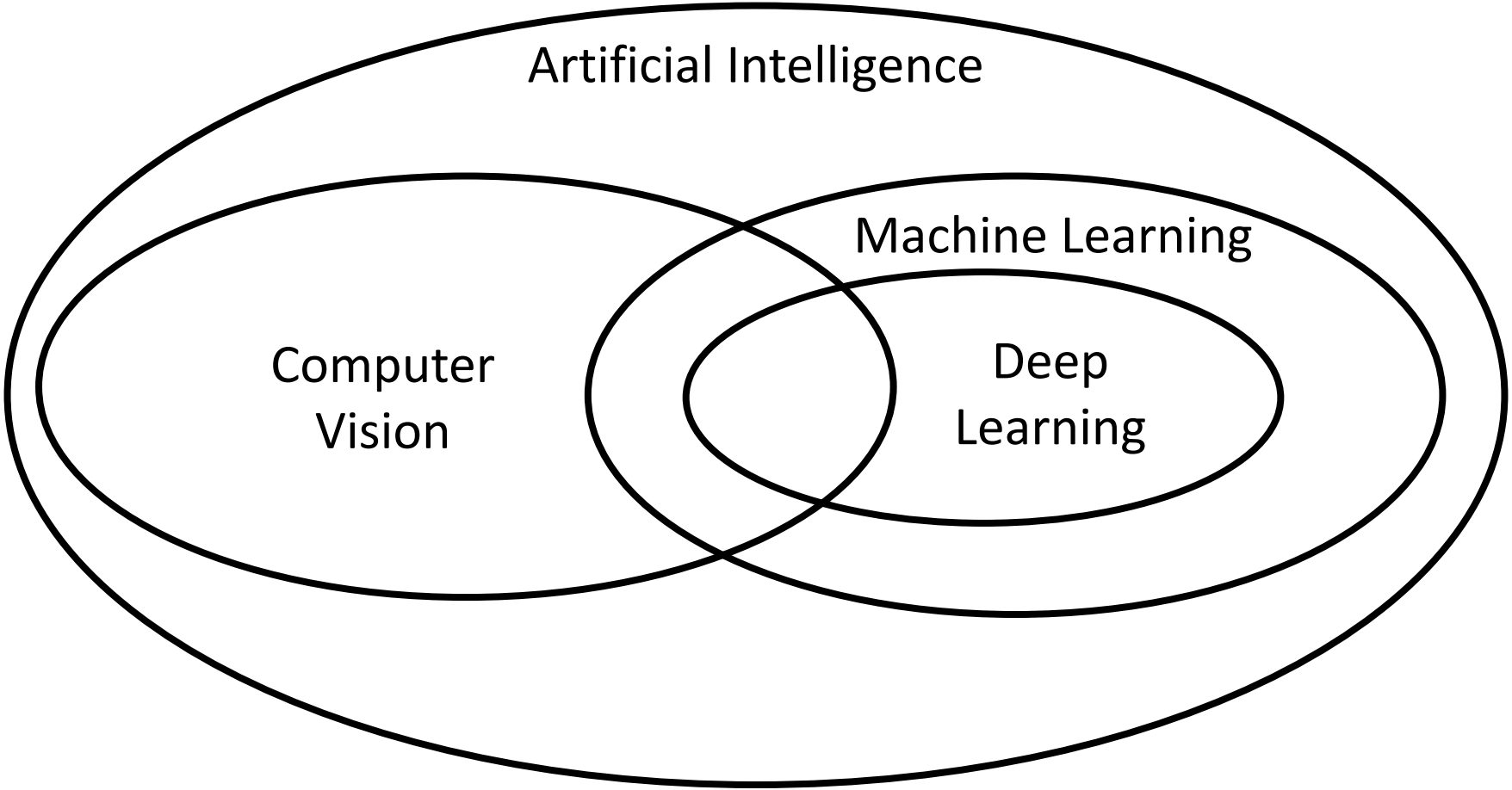
Lecture 1: Introduction



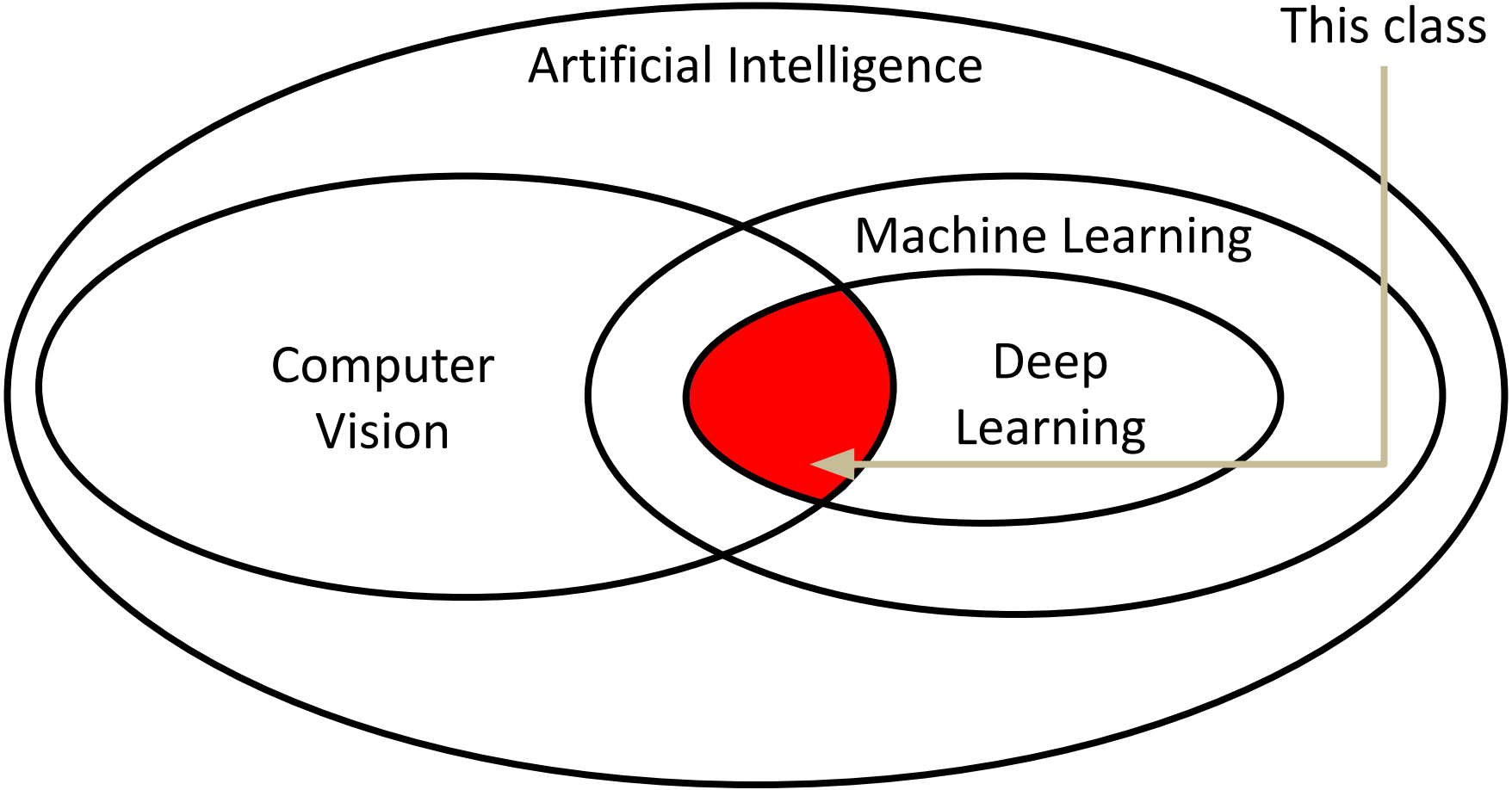
Slide inspiration: Justin Johnson



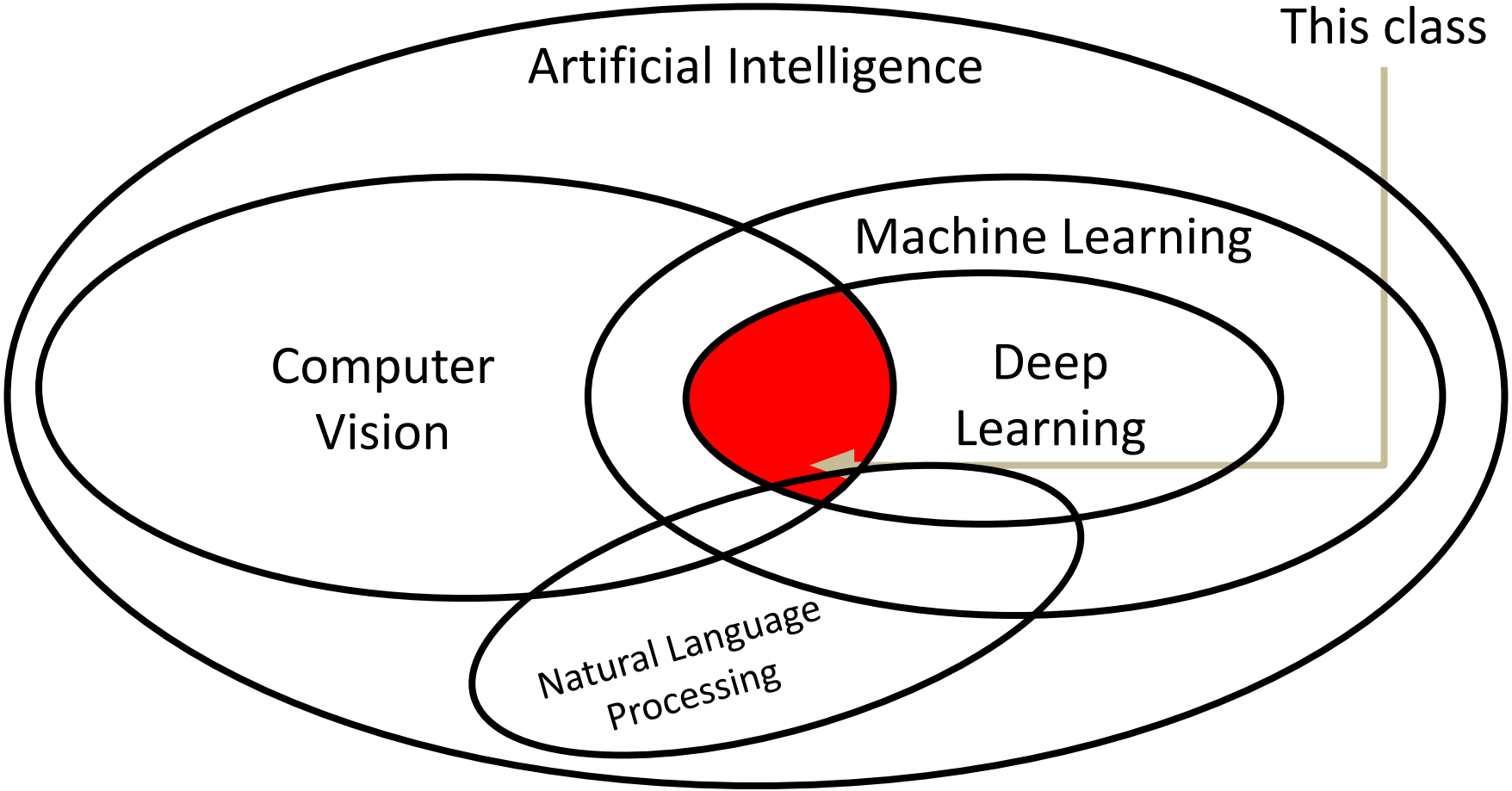
Slide inspiration: Justin Johnson



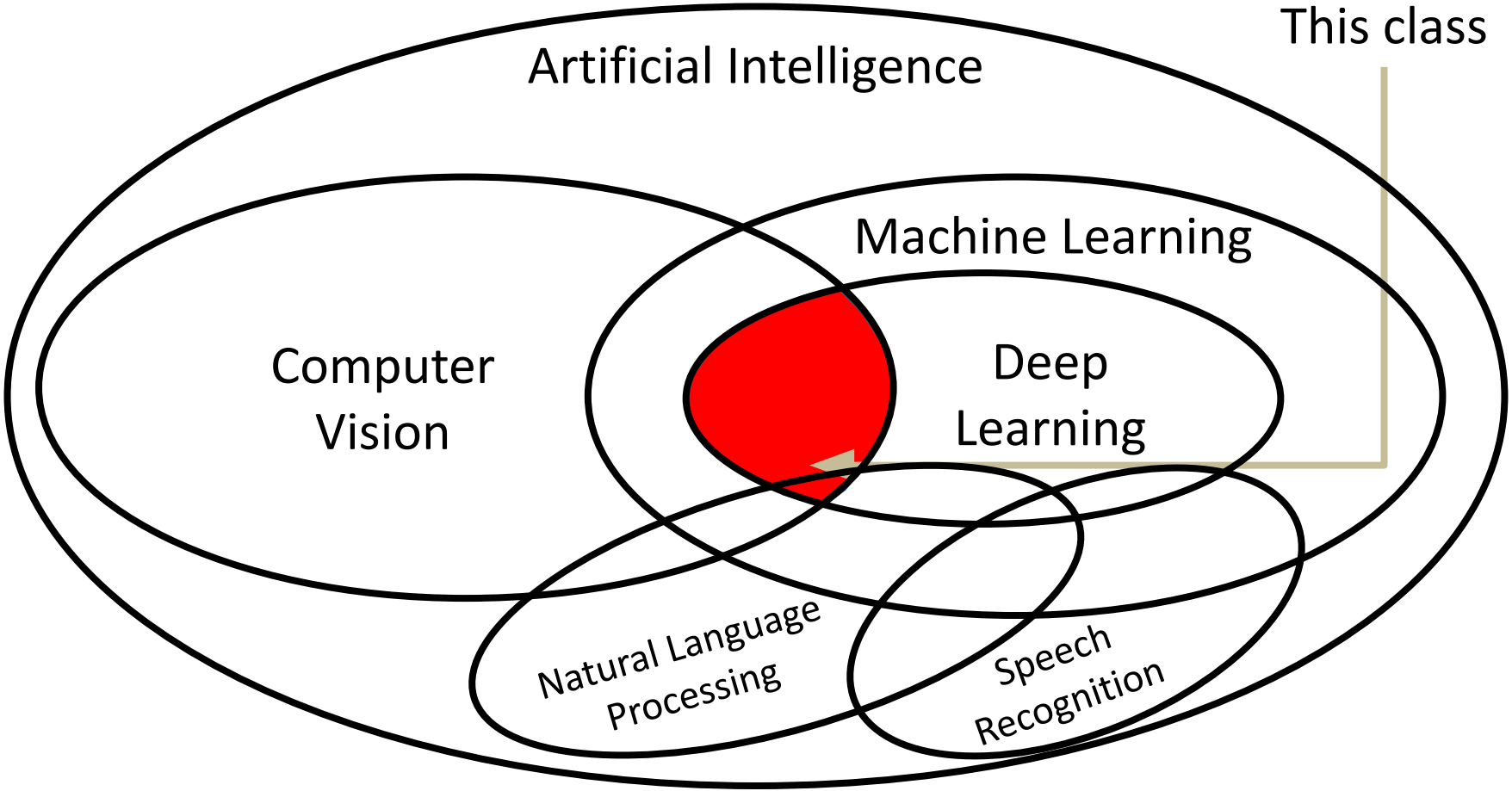
Slide inspiration: Justin Johnson



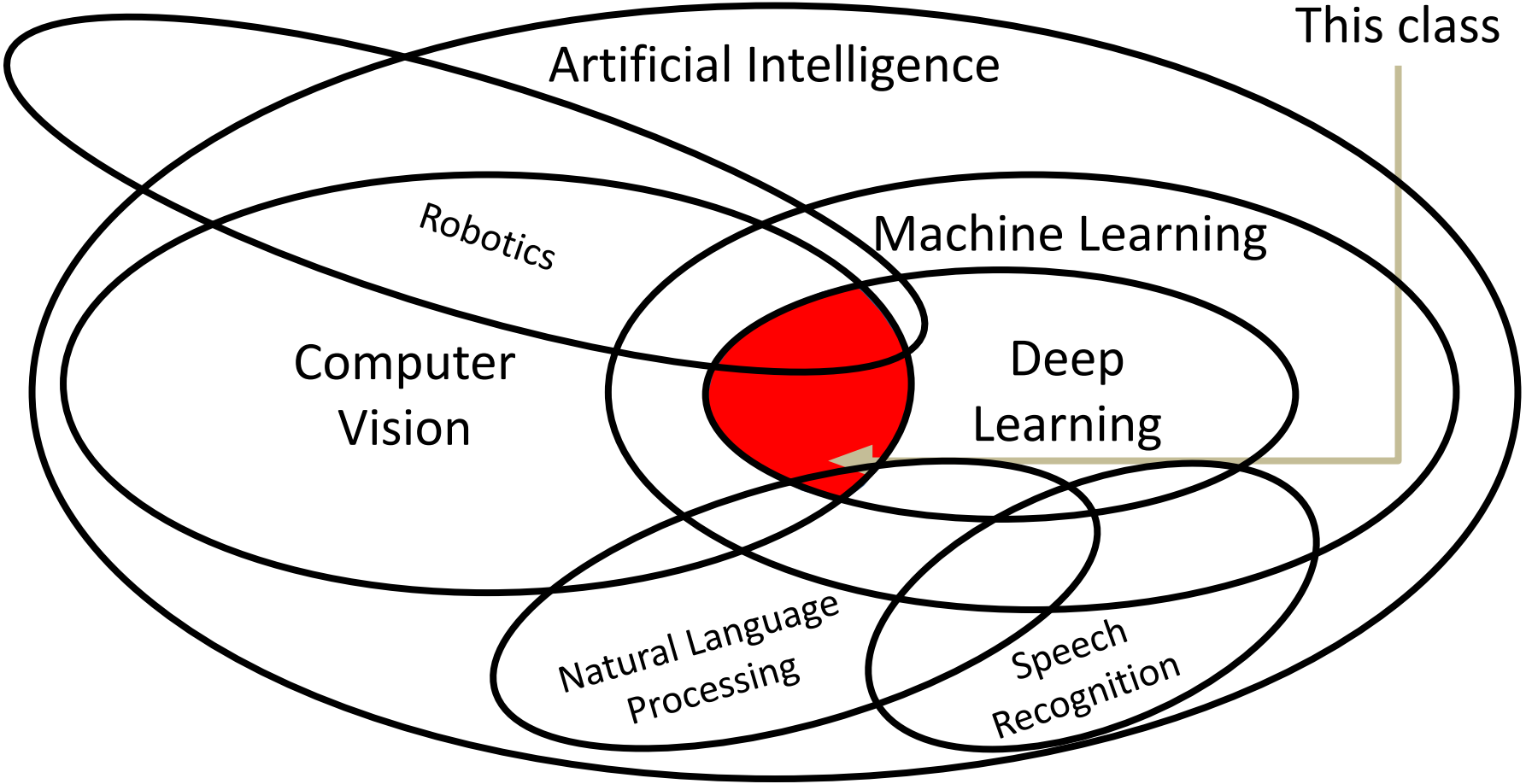
Slide inspiration: Justin Johnson



Slide inspiration: Justin Johnson



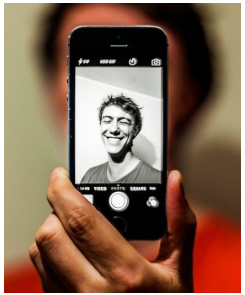
Slide inspiration: Justin Johnson

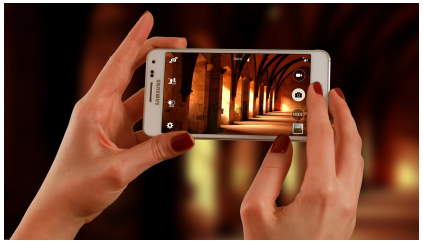


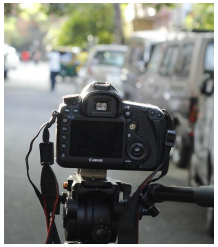
Slide inspiration: Justin Johnson

Computer Vision is everywhere!

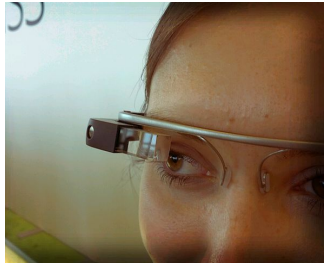






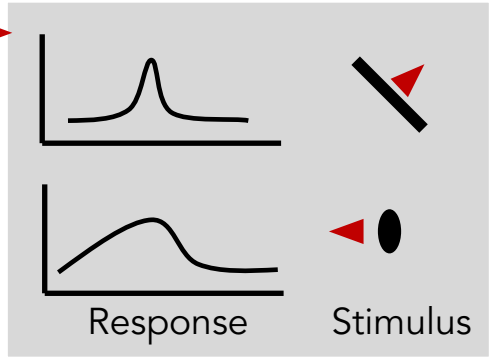



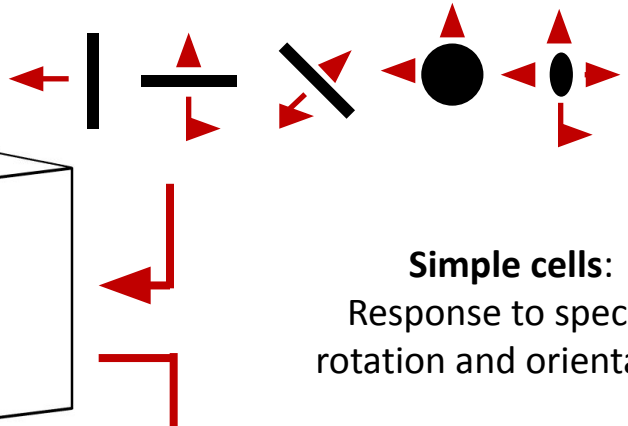
Where did we come from?

Hubel and Wiesel, 1959

Measure brain activity

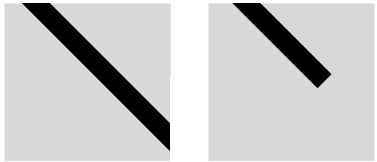
1959
Hubel & Wiesel





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

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

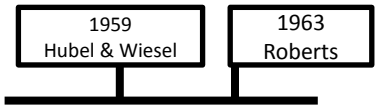
Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

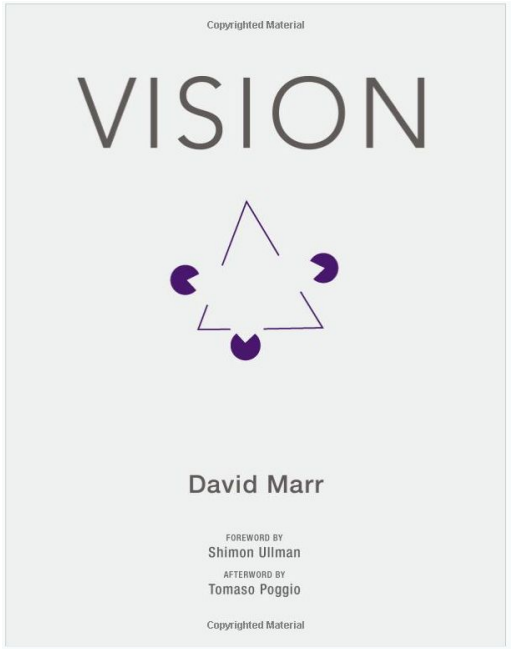
Seymour Papert

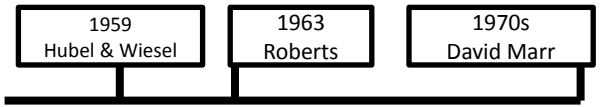
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the 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

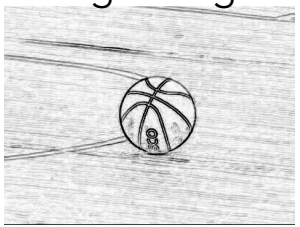
Slide inspiration: Justin Johnson



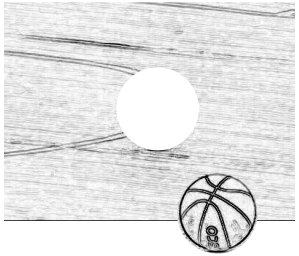


Input image

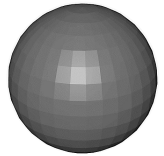
Edge image

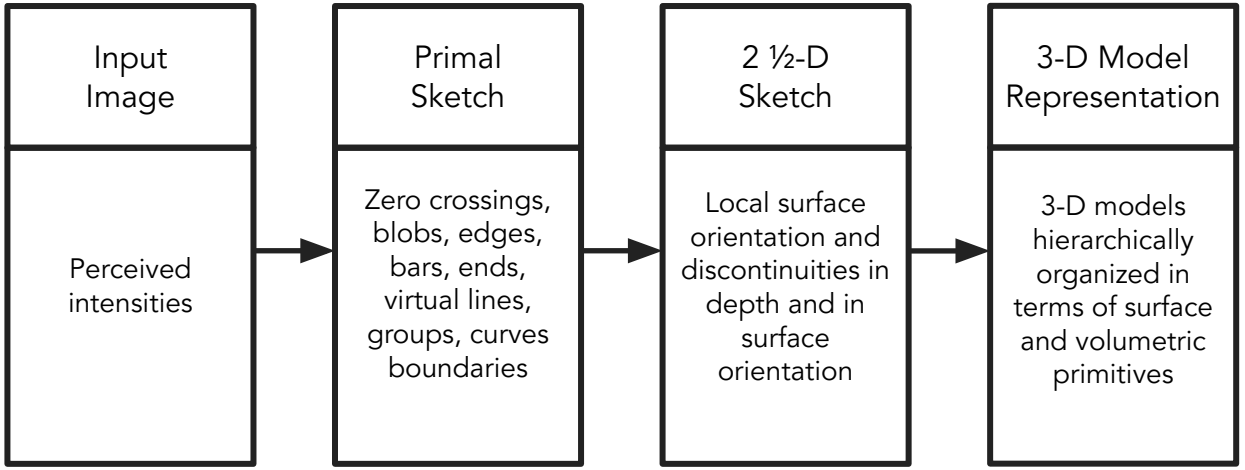


2 ½-D sketch



3-D model

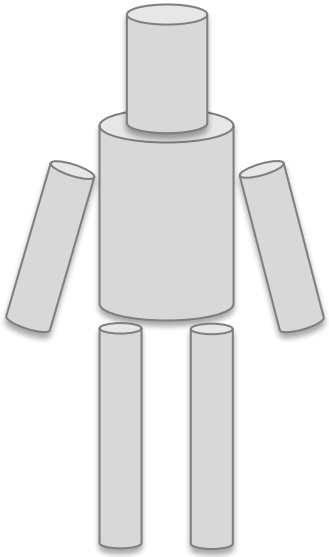




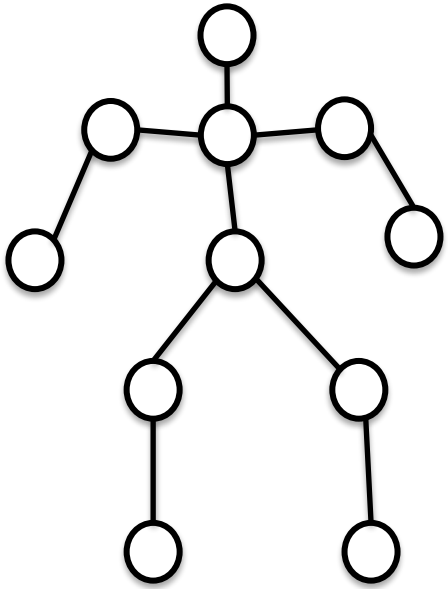
Stages of Visual Representation, David Marr, 1970s

Slide inspiration: Justin Johnson

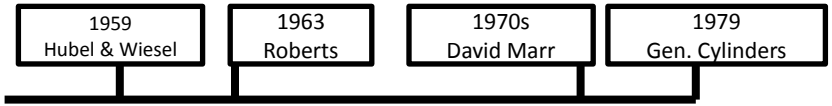
Recognition via Parts (1970s)



Generalized Cylinders,
Brooks and Binford,
1979

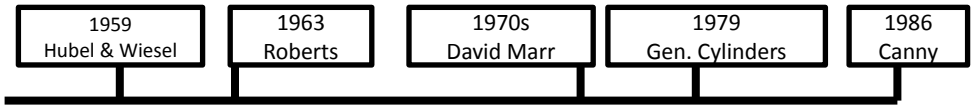


Pictorial Structures,
Fischler and Elshlager, 1973



Slide inspiration: Justin Johnson

Recognition via Edge Detection (1980s)

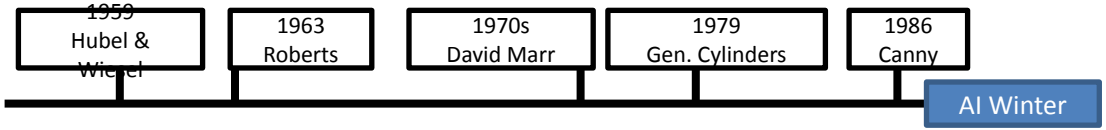


John Canny, 1986
David Lowe, 1987

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.

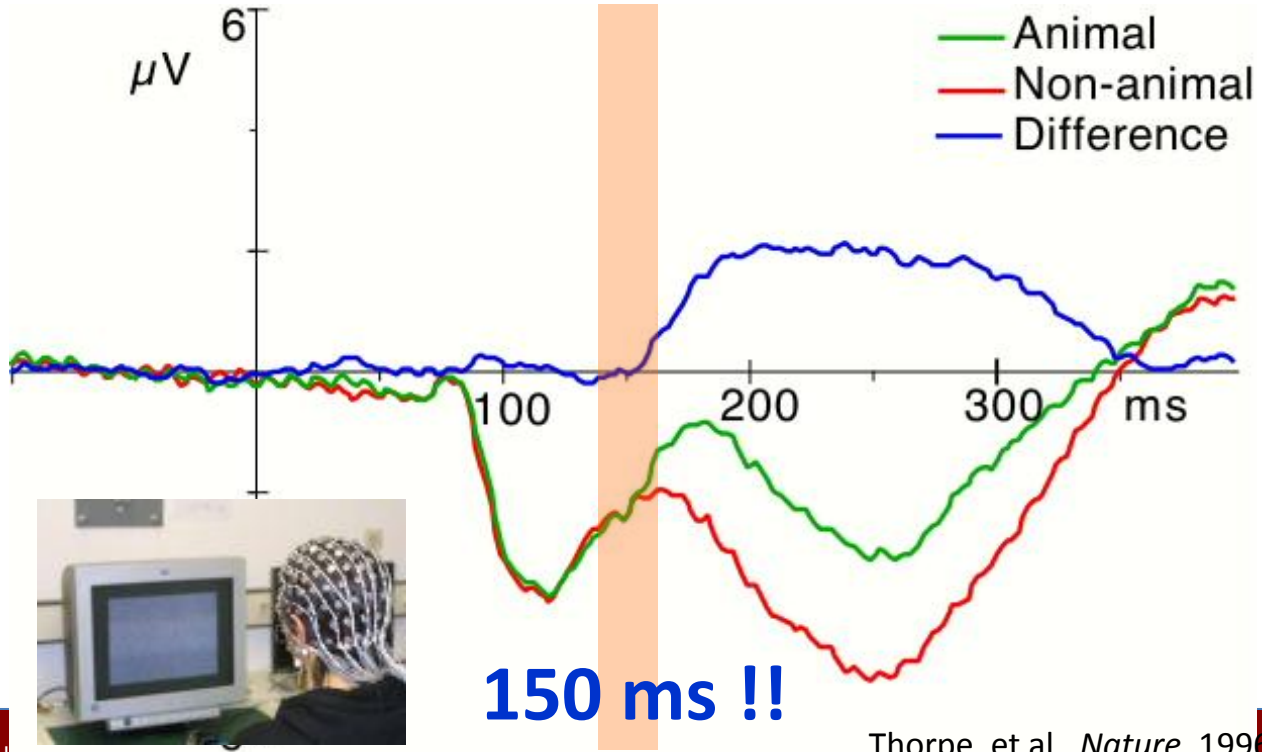


Slide inspiration: Justin Johnson

In the meantime…seminal work in cognitive and neuroscience

Speed of processing in the human visual system

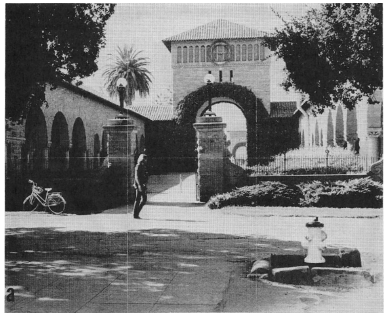
Simon Thorpe, Denis Fize & Catherine Marlot

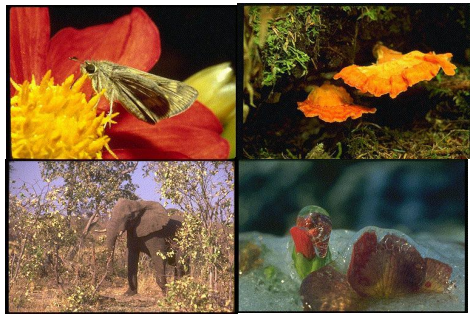


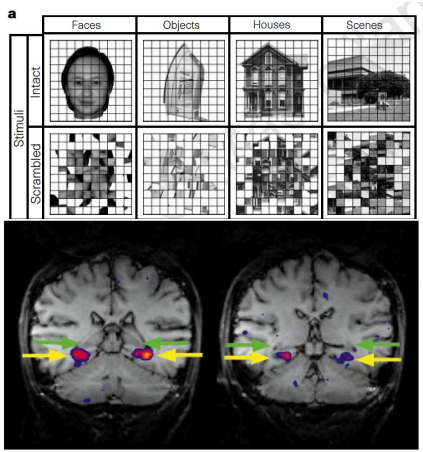
150 ms !!

Thorpe, et al. *Nature*, 1996

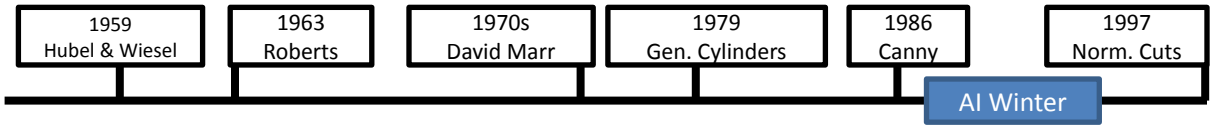
Visual recognition is a fundamental task for visual intelligence







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

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
AI Winter
1997 Norm. Cuts
1999 SIFT

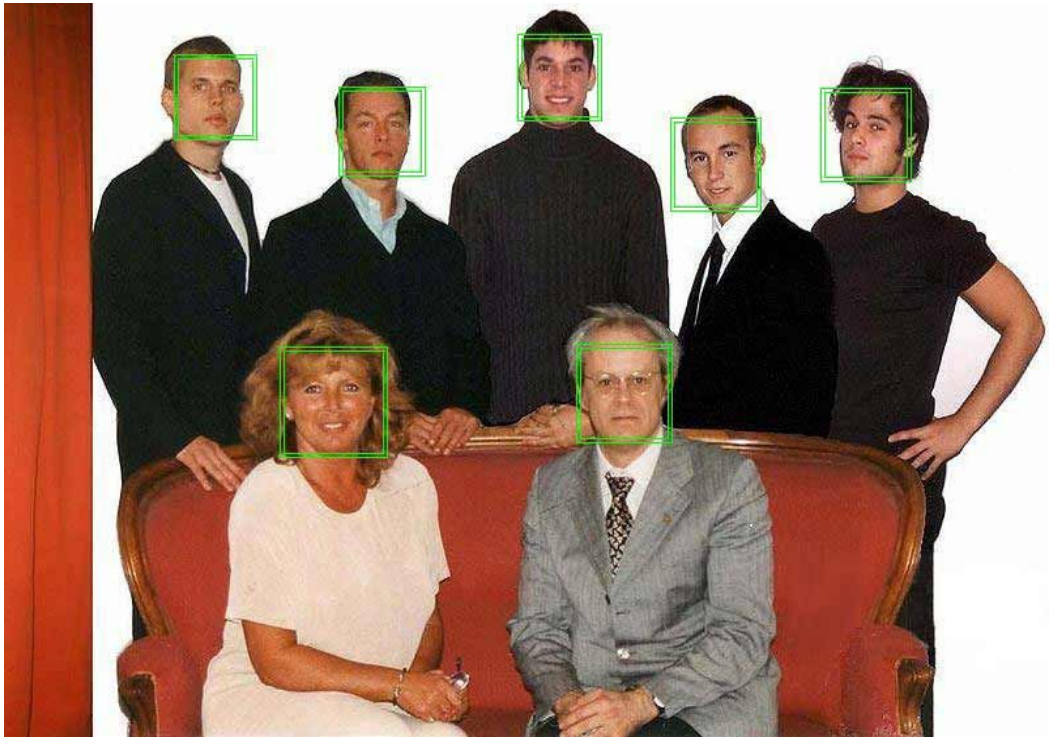
SIFT, David Lowe, 1999

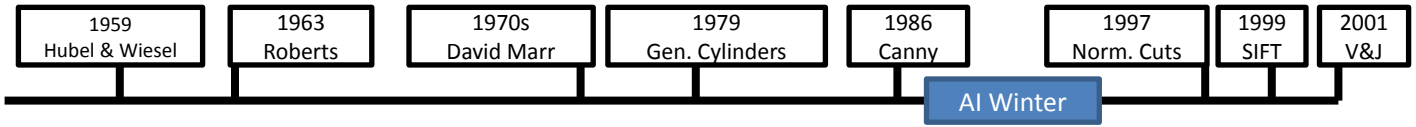
Slide inspiration: Justin Johnson

Face Detection

Viola and Jones, 2001

One of the first successful applications of machine learning to vision

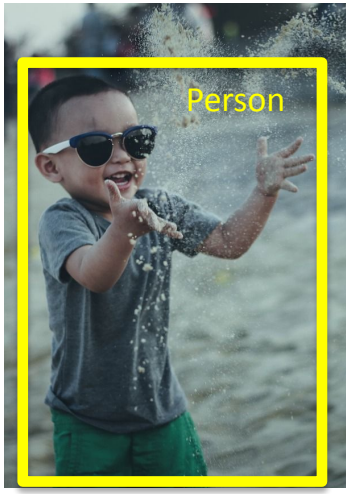




Slide inspiration: Justin Johnson

Caltech 101 images

PASCAL Visual Object Challenge

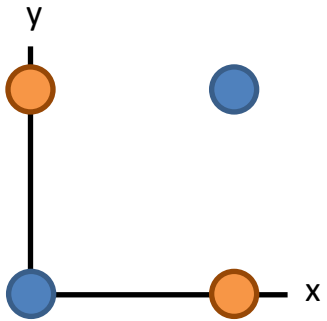


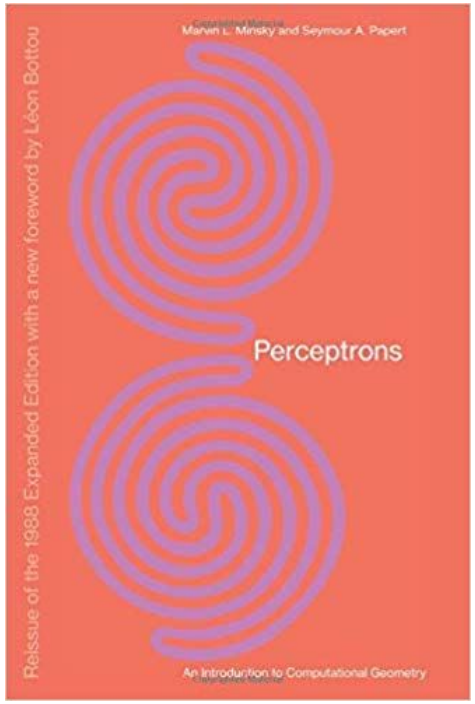
1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
AI Winter
1997 Norm. Cuts
1999 SIFT
2001 V&J
2004, 2007 Caltech101; PASCAL

Slide inspiration: Justin Johnson

Minsky and Papert, 1969

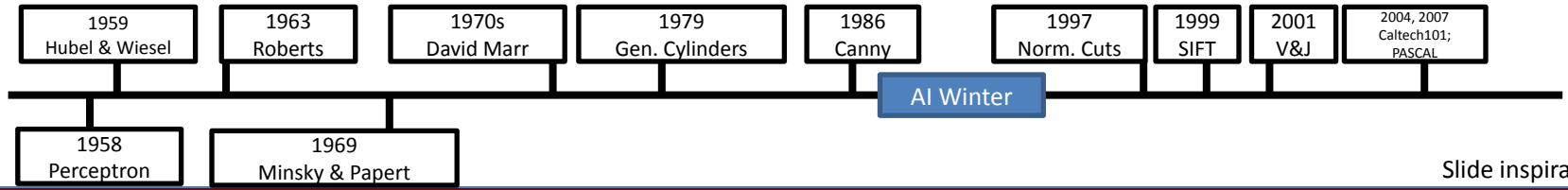
X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0





Showed that Perceptrons could not learn the XOR function

Caused a lot of disillusionment in the field



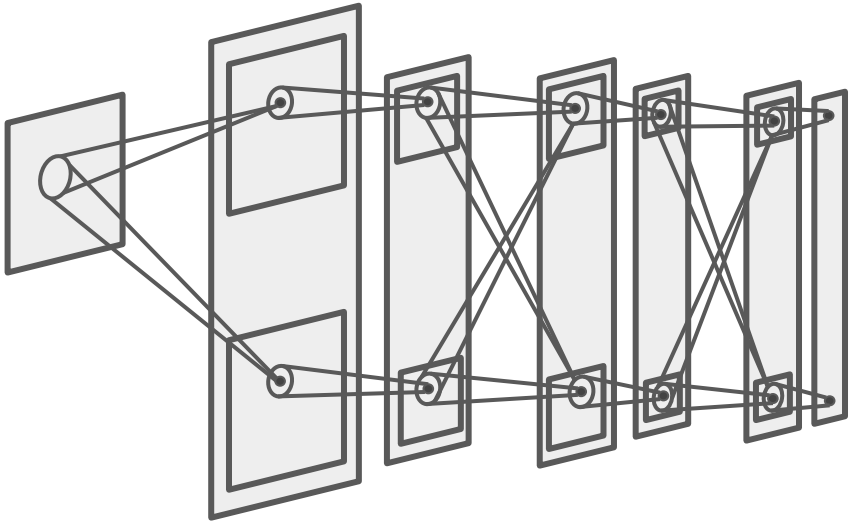
Slide inspiration: Justin Johnson

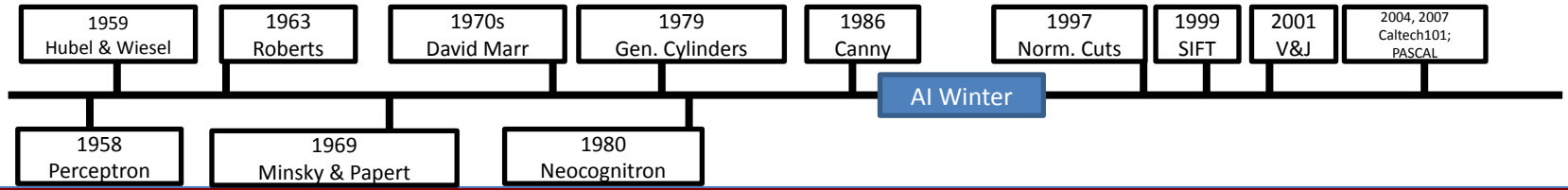
Neocognitron: Fukushima, 1980

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

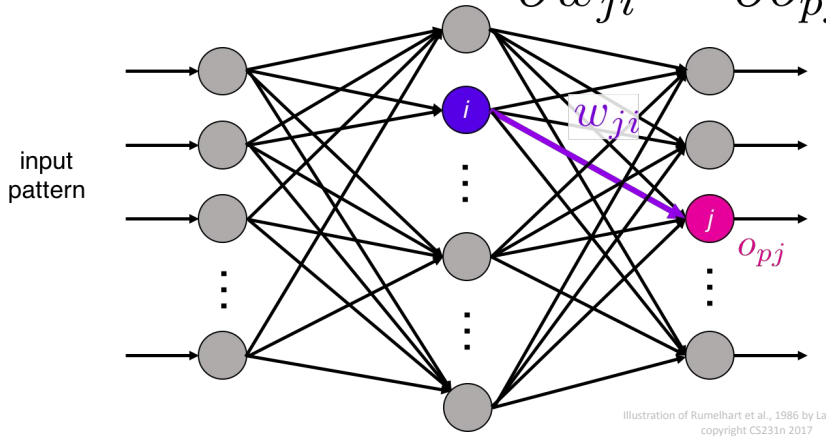
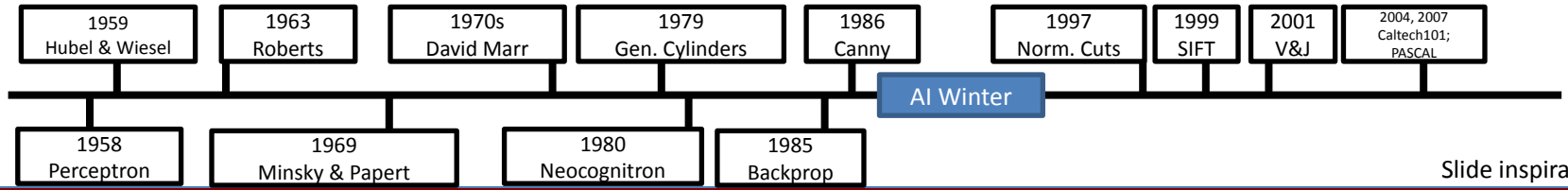
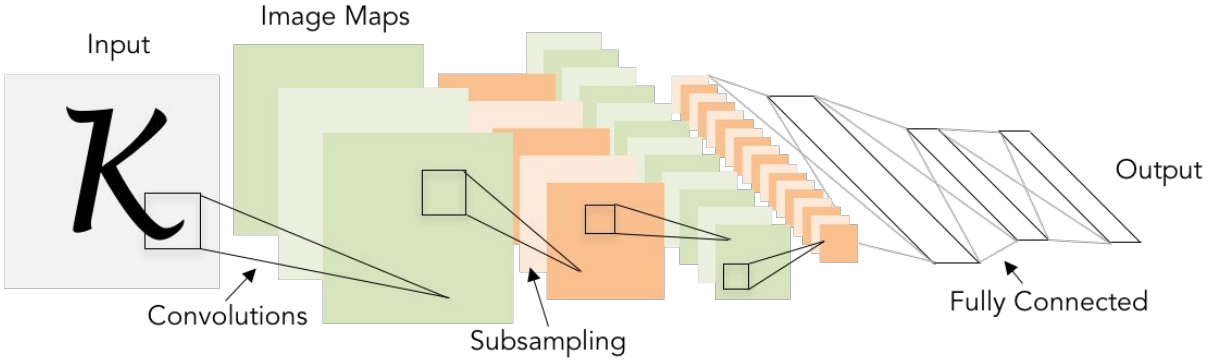


Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017

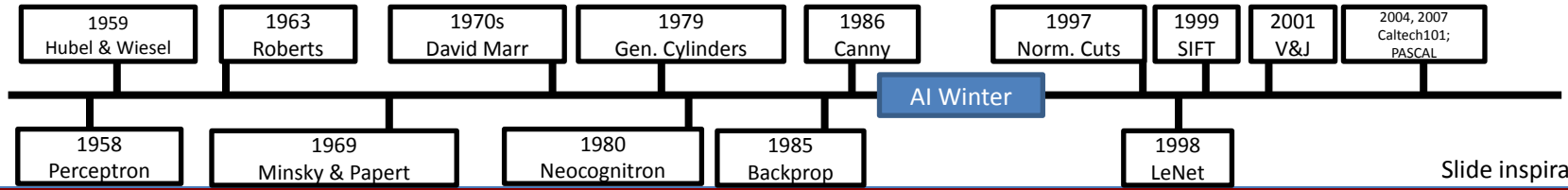


Slide inspiration: Justin Johnson

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!



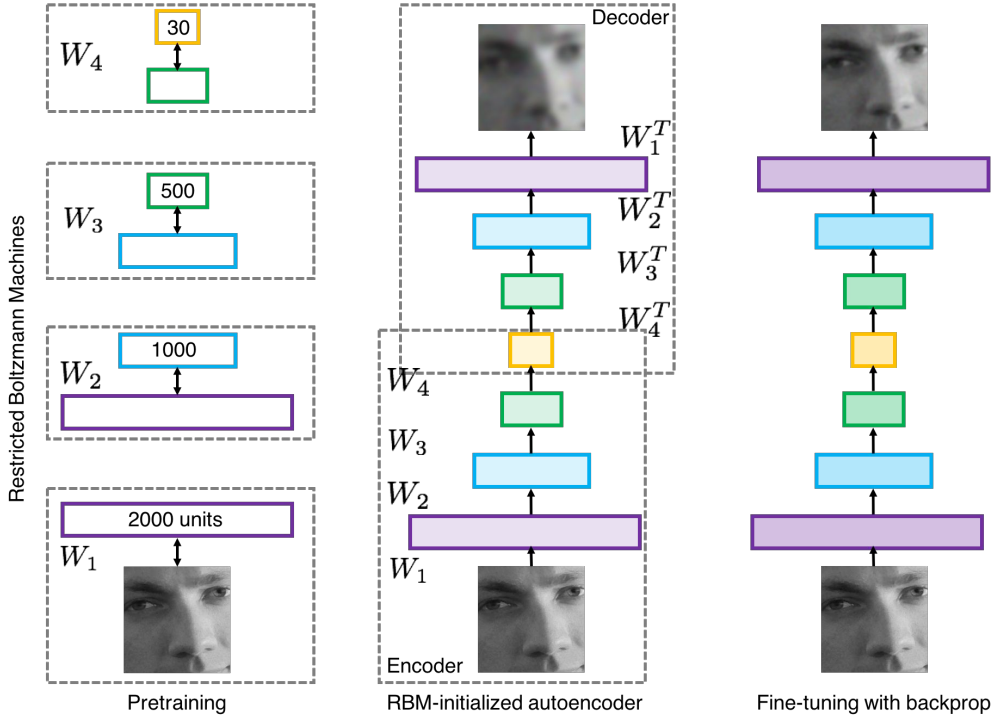
Slide inspiration: Justin Johnson

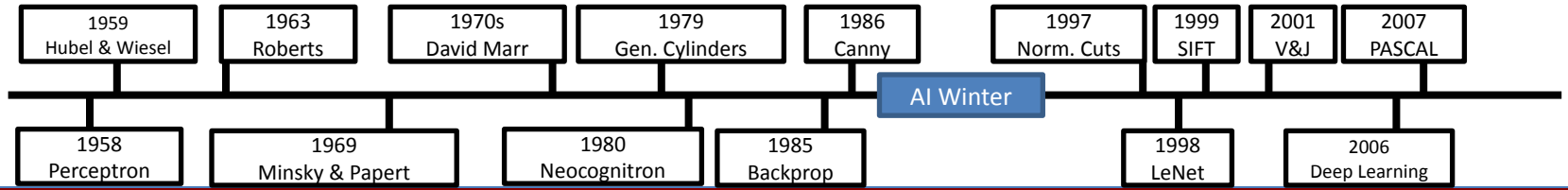
2000s: “Deep Learning”

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

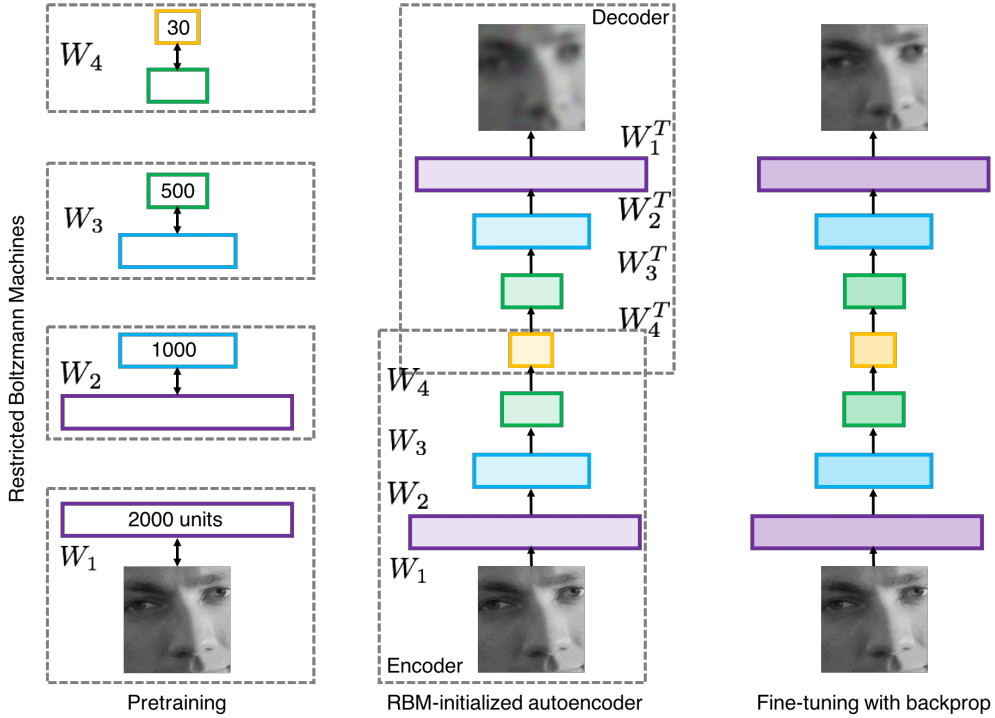
2000s: “Deep Learning”

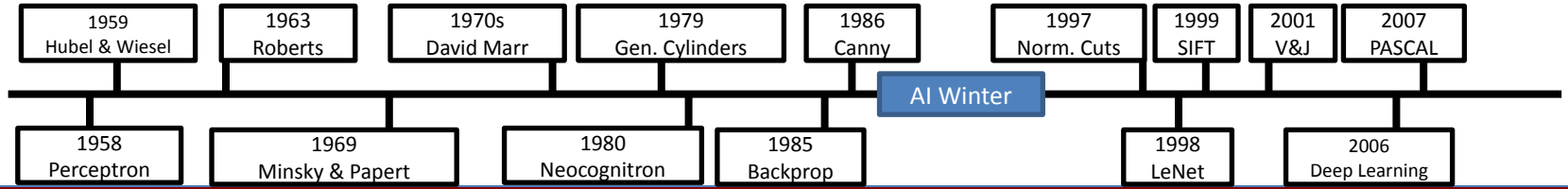
People tried to train neural networks that were deeper and deeper

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





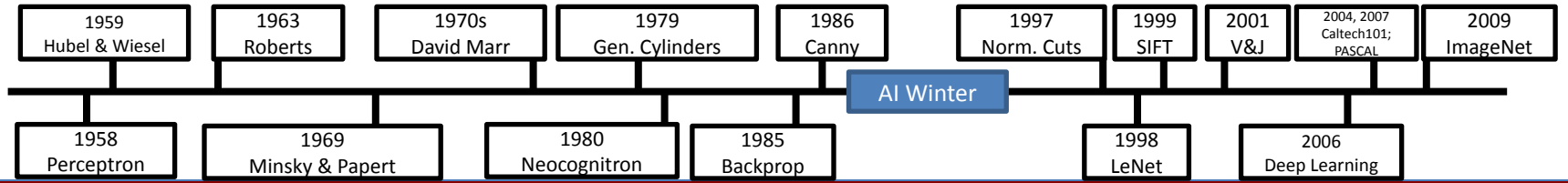
Slide inspiration: Justin Johnson

IMAGENET Large Scale Visual Recognition Challenge

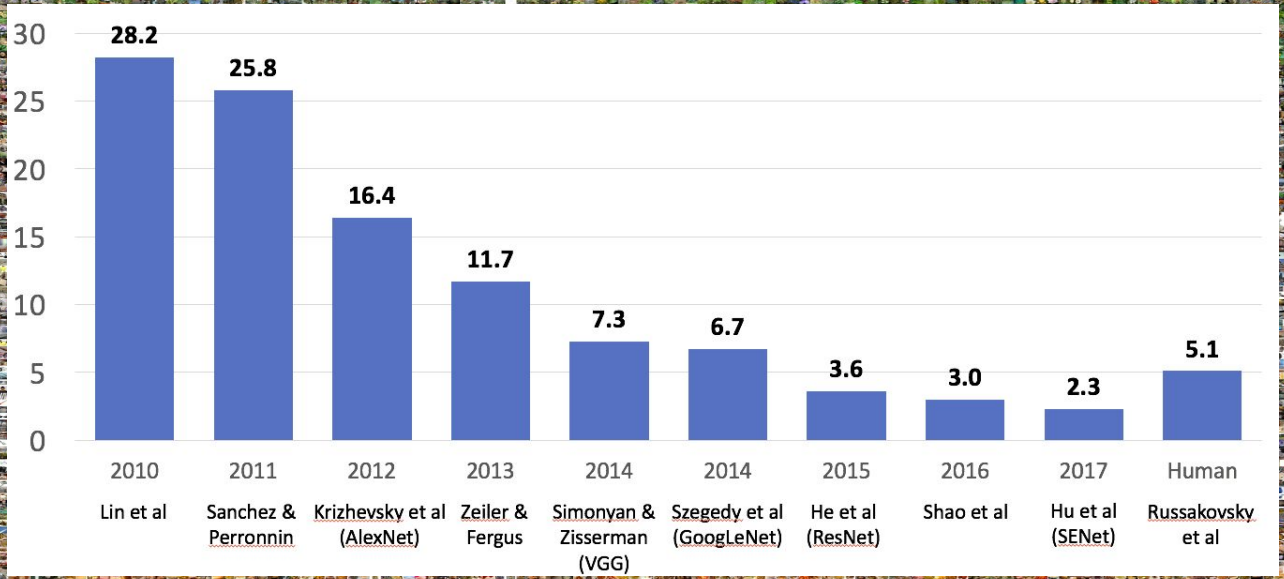
The Image Classification Challenge:
1,000 object classes
1,431,167 images

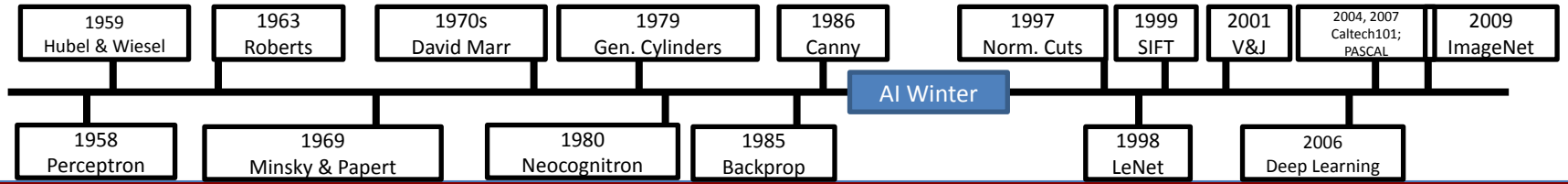
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

Deng et al, 2009
Russakovsky et al. IJCV 2015

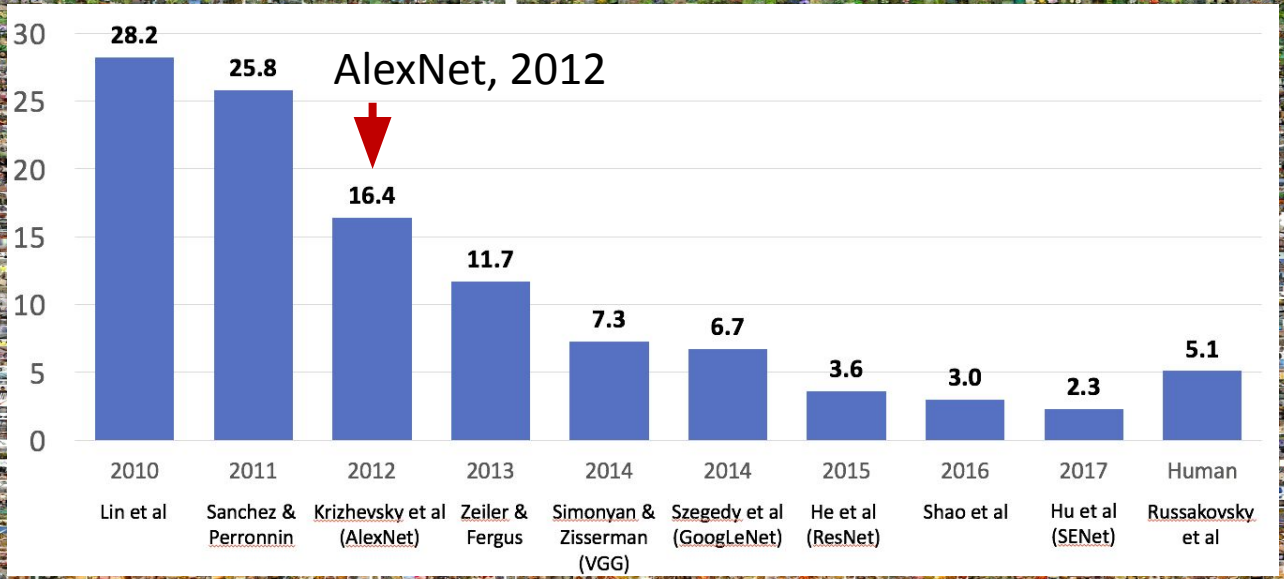


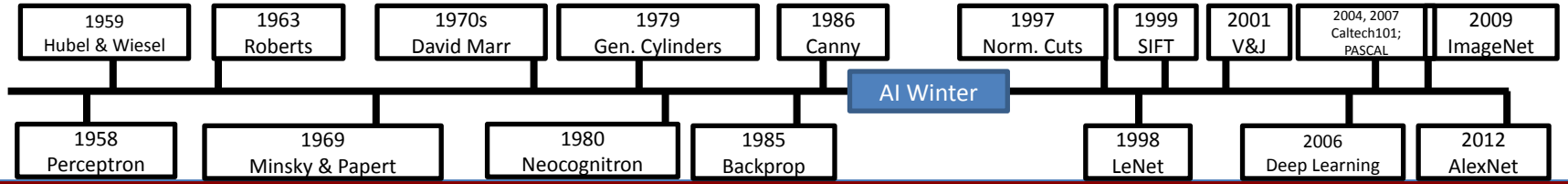
IMAGENET Large Scale Visual Recognition Challenge



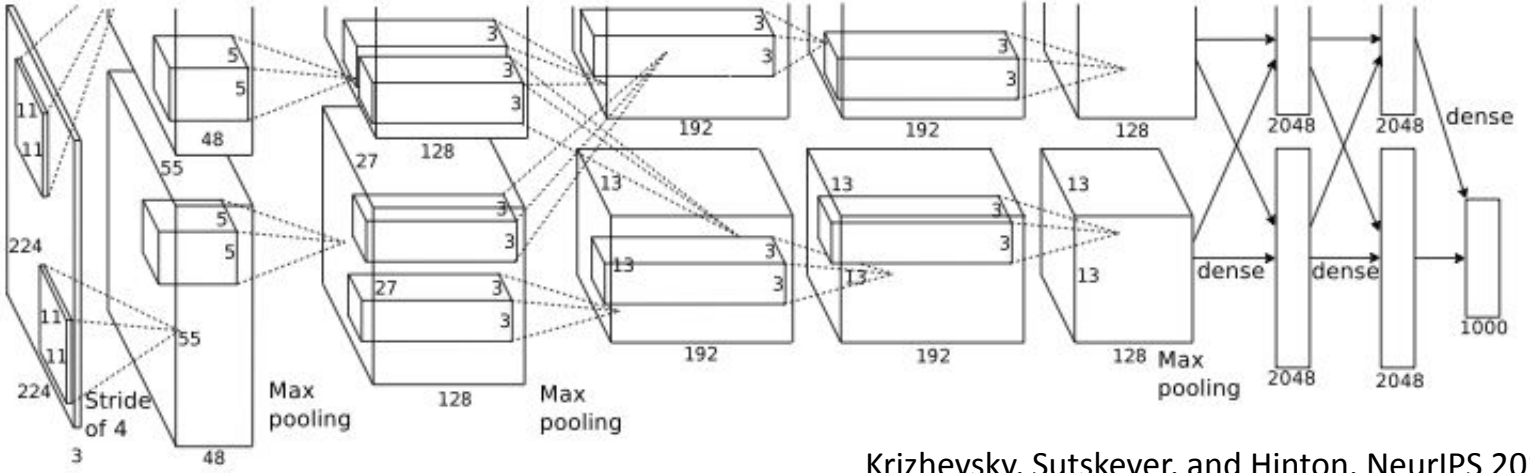


IMAGENET Large Scale Visual Recognition Challenge

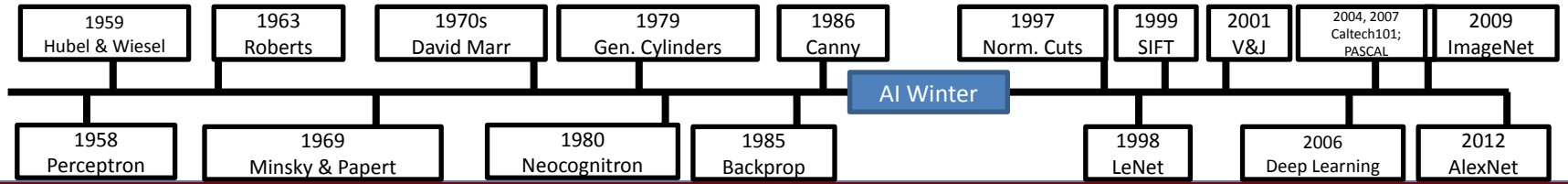




AlexNet: Deep Learning Goes Mainstream

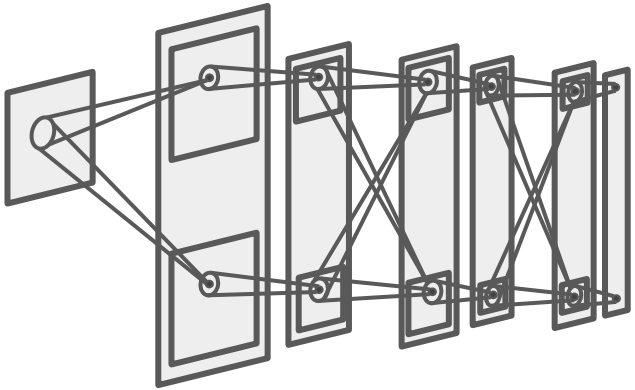


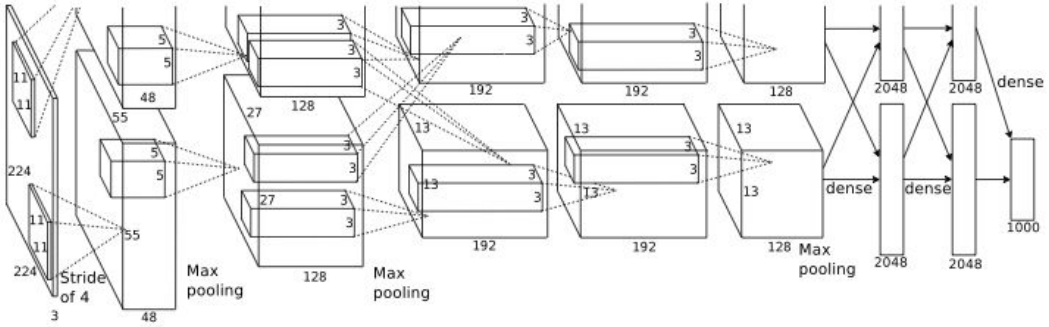
Krizhevsky, Sutskever, and Hinton, NeurIPS 2012

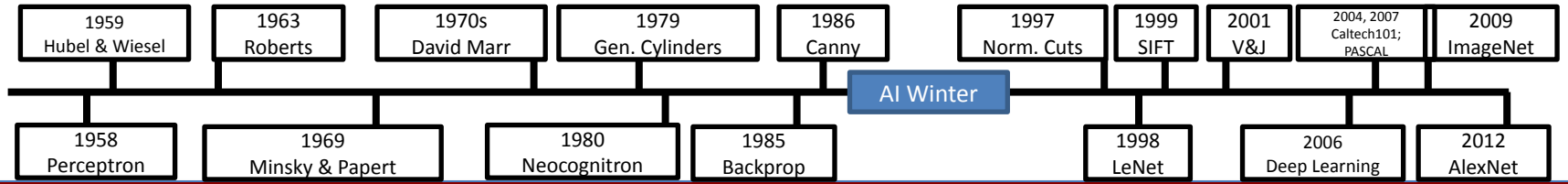


Slide inspiration: Justin Johnson

AlexNet vs. Neocognitron: 32 years apart



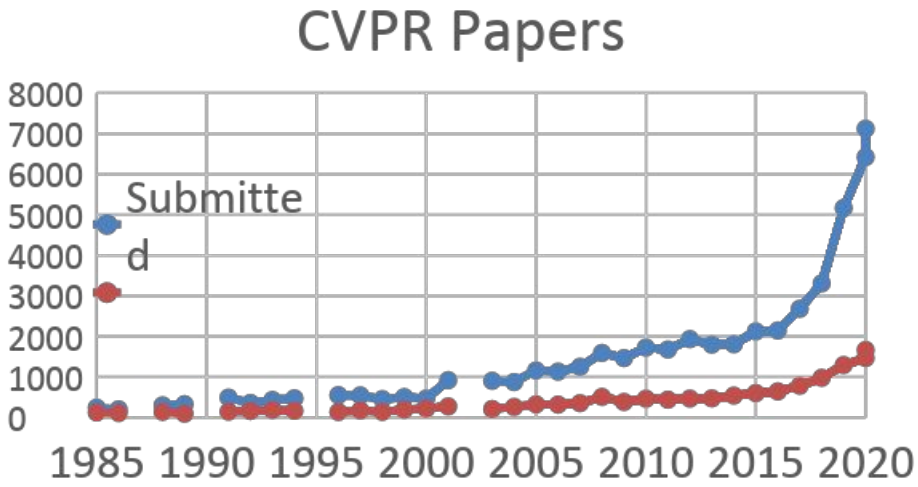




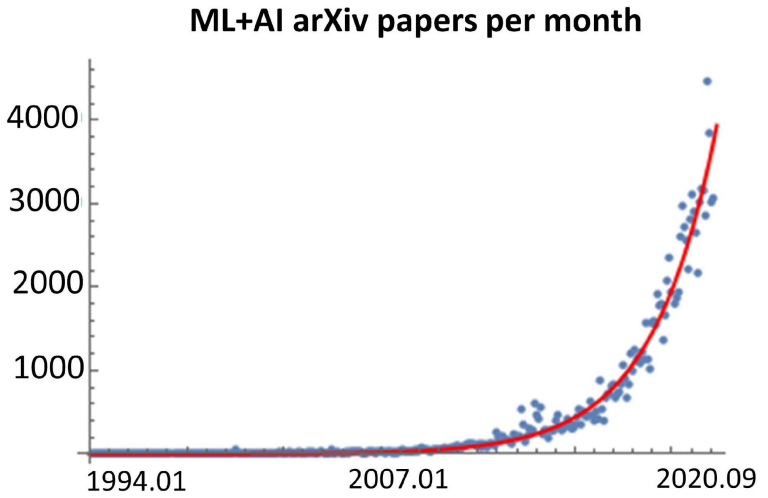
Slide inspiration: Justin Johnson

The AI winter thawed,
deep learning revolution arrived

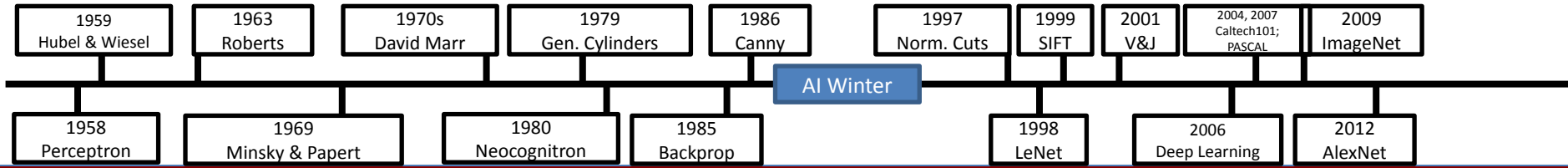
2012 to Present: Deep Learning Explosion



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

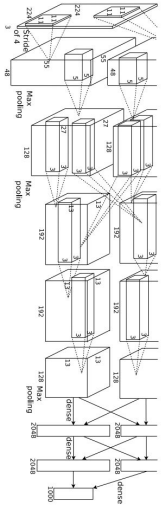
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

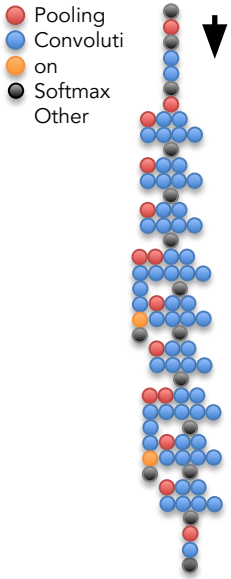
SuperVision



[Krizhevsky NIPS 2012]

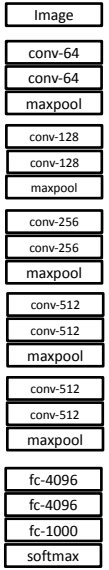
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

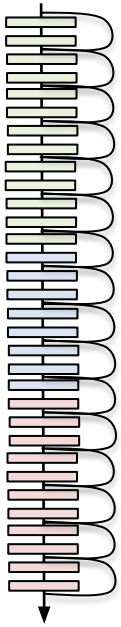
VGG



[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

2012 to Present: Deep Learning is Everywhere

Image Classification

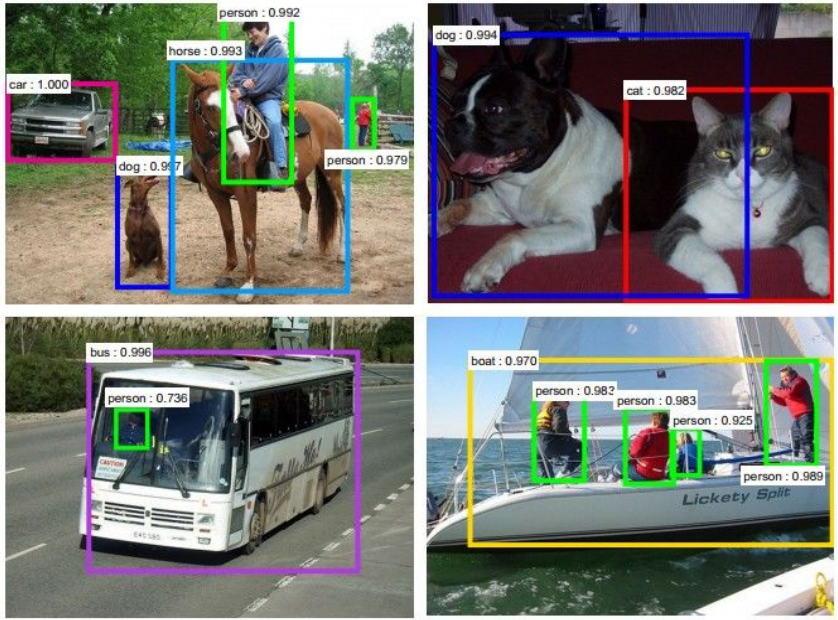
Image Retrieval

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

Slide inspiration: Justin Johnson

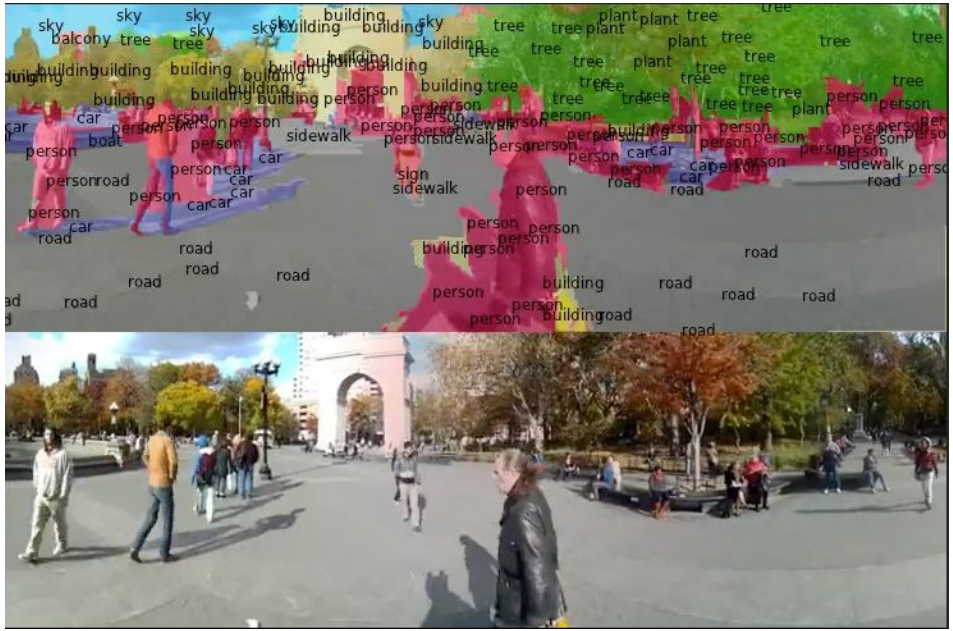
2012 to Present: Deep Learning is Everywhere

Object Detection



Ren, He, Girshick, and Sun, 2015

Image Segmentation

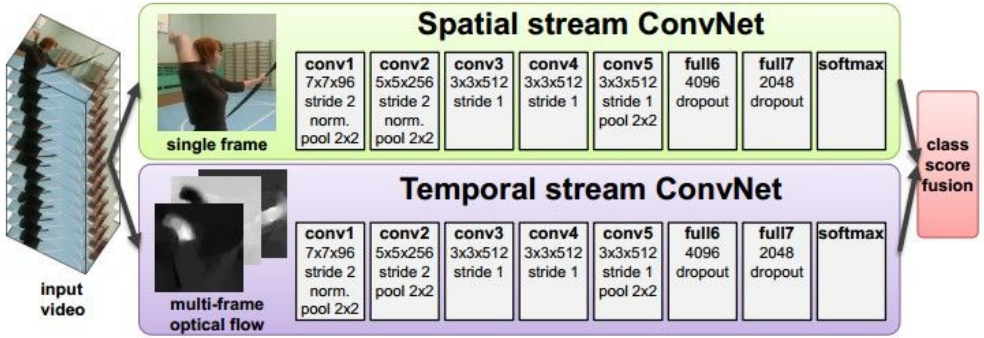


Fabaret et al, 2012

Slide inspiration: Justin Johnson

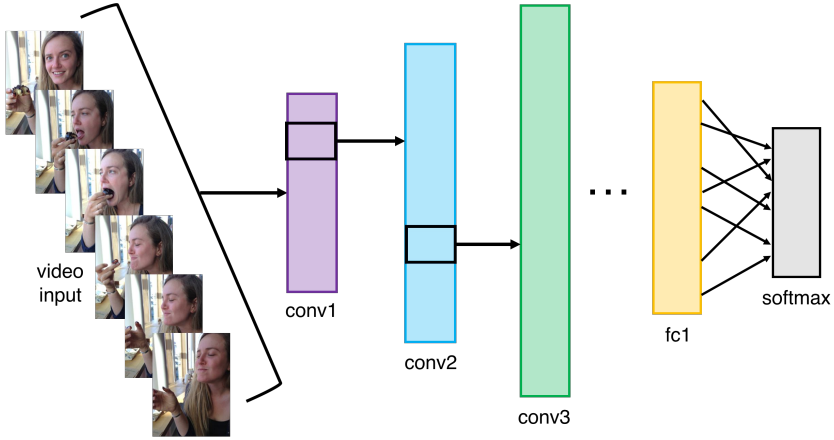
2012 to Present: Deep Learning is Everywhere

Video Classification



Simonyan et al, 2014

Activity Recognition

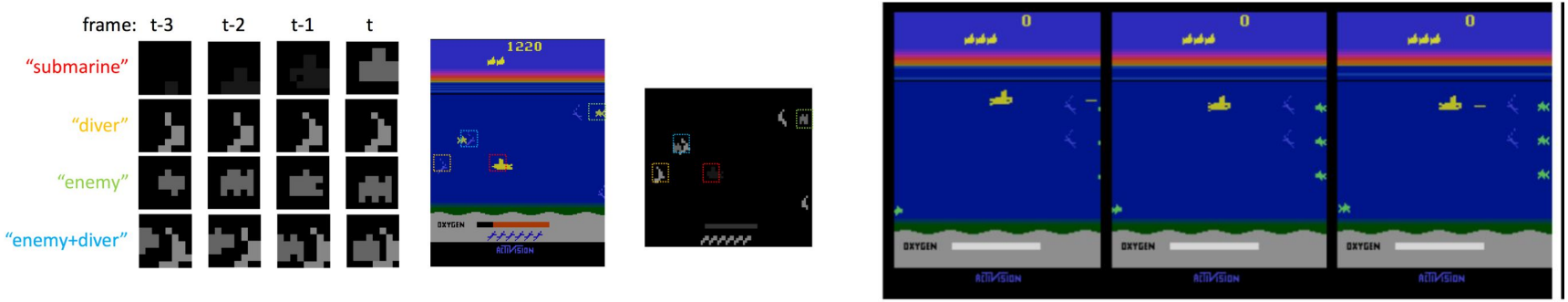


Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Pose Recognition (Toshev and Szegedy, 2014)

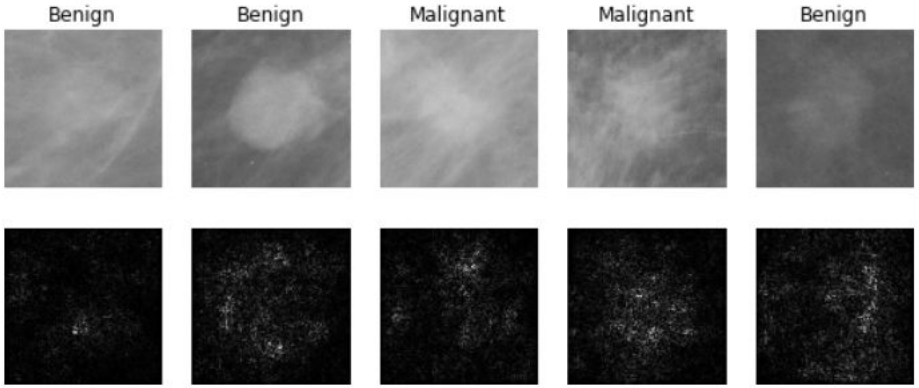
Playing Atari games (Guo et al, 2014)



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 Fei-Fei, 2015

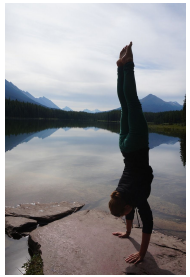
A white teddy bear sitting in the grass

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

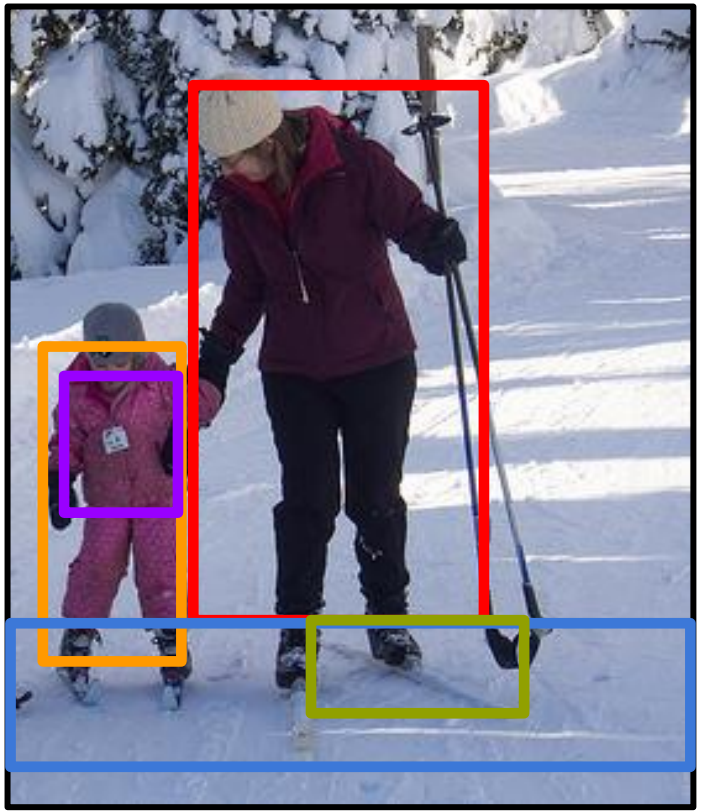


A woman standing on a beach holding a surfboard

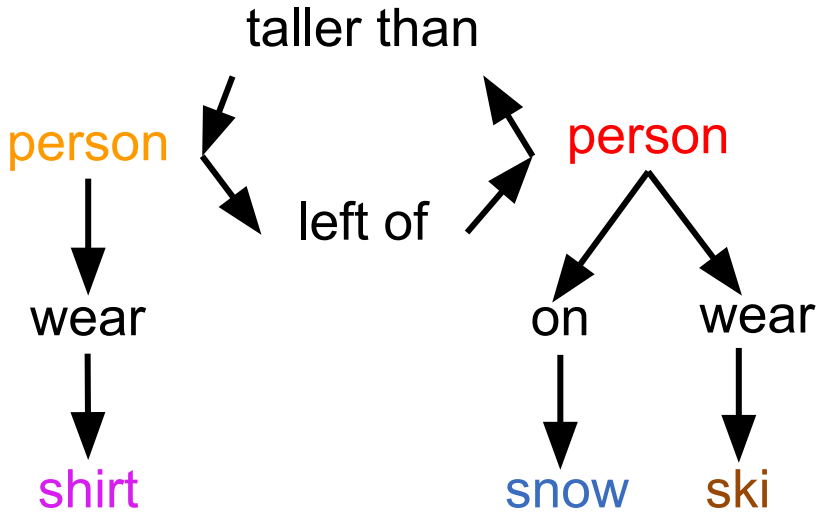
Captions generated by Justin Johnson using Neuraltalk2

Slide inspiration: Justin Johnson

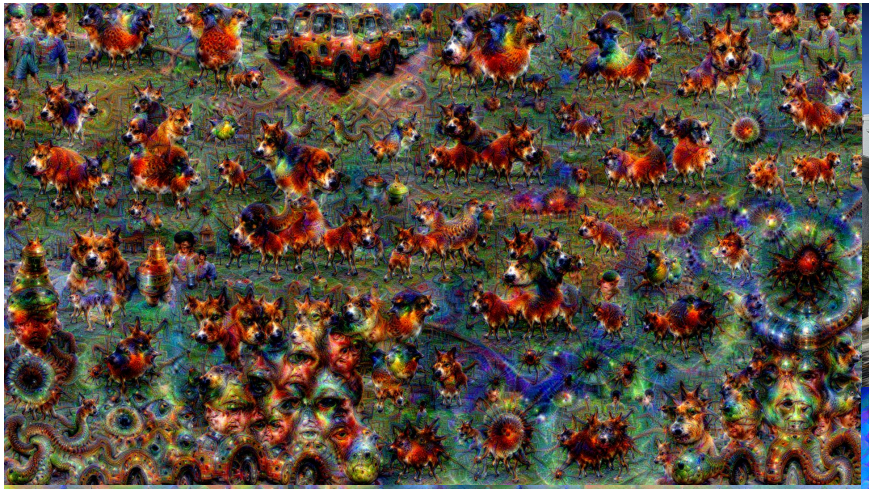
2012 to Present: Deep Learning is Everywhere

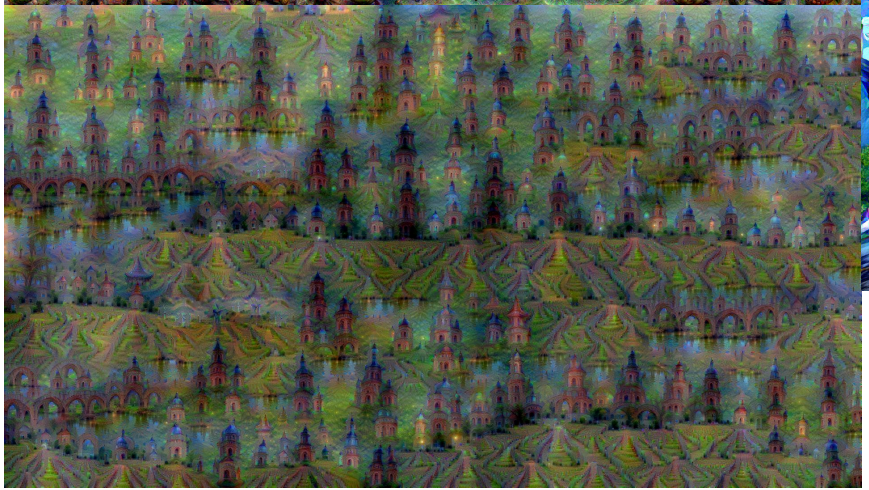


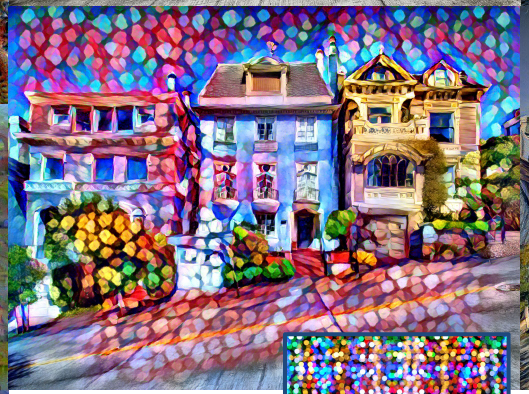
Results:
spatial, comparative, asymmetrical, verb, prepositional



Krishna*, Lu*, Bernstein, Fei-Fei, *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 public domain
Stylized images copyright Justin Johnson, 2017;
reproduced with permission

Mordvinsev et al, 2015
Gatys et al, 2016

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

Slide inspiration: Justin Johnson

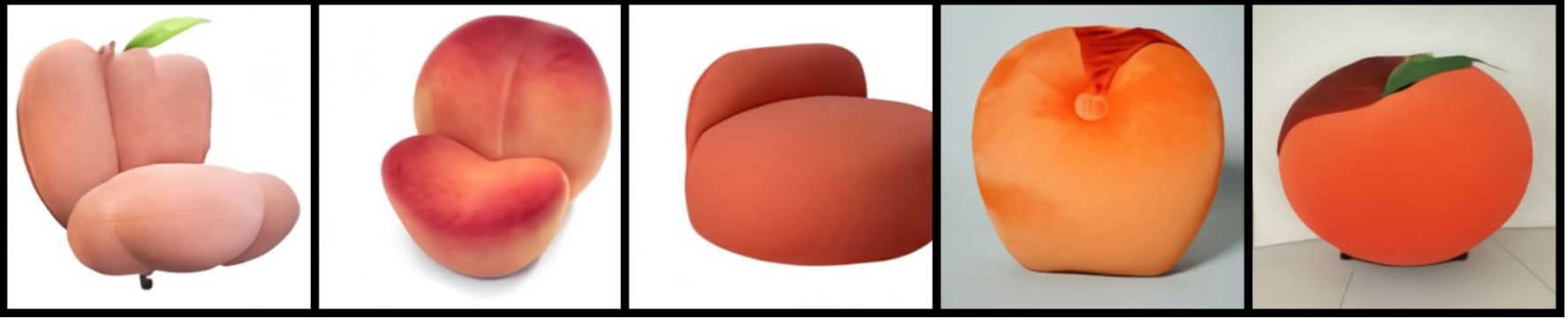
Ramesh et al, "DALL·E: Creating Images from Text", 2021. https://openai.com/blog/dall-e/

2012 to Present: Deep Learning is Everywhere

TEXT PROMPT

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

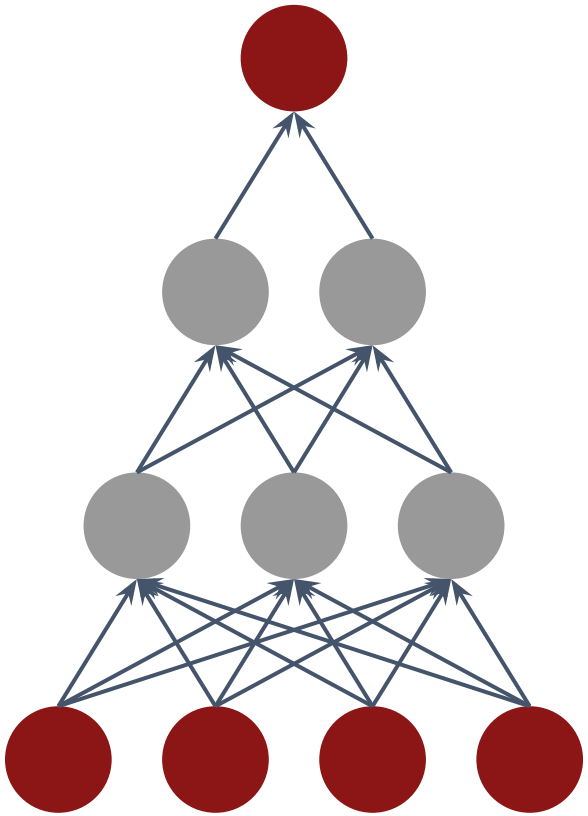
AI-GENERATED IMAGES



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

Slide inspiration: Justin Johnson

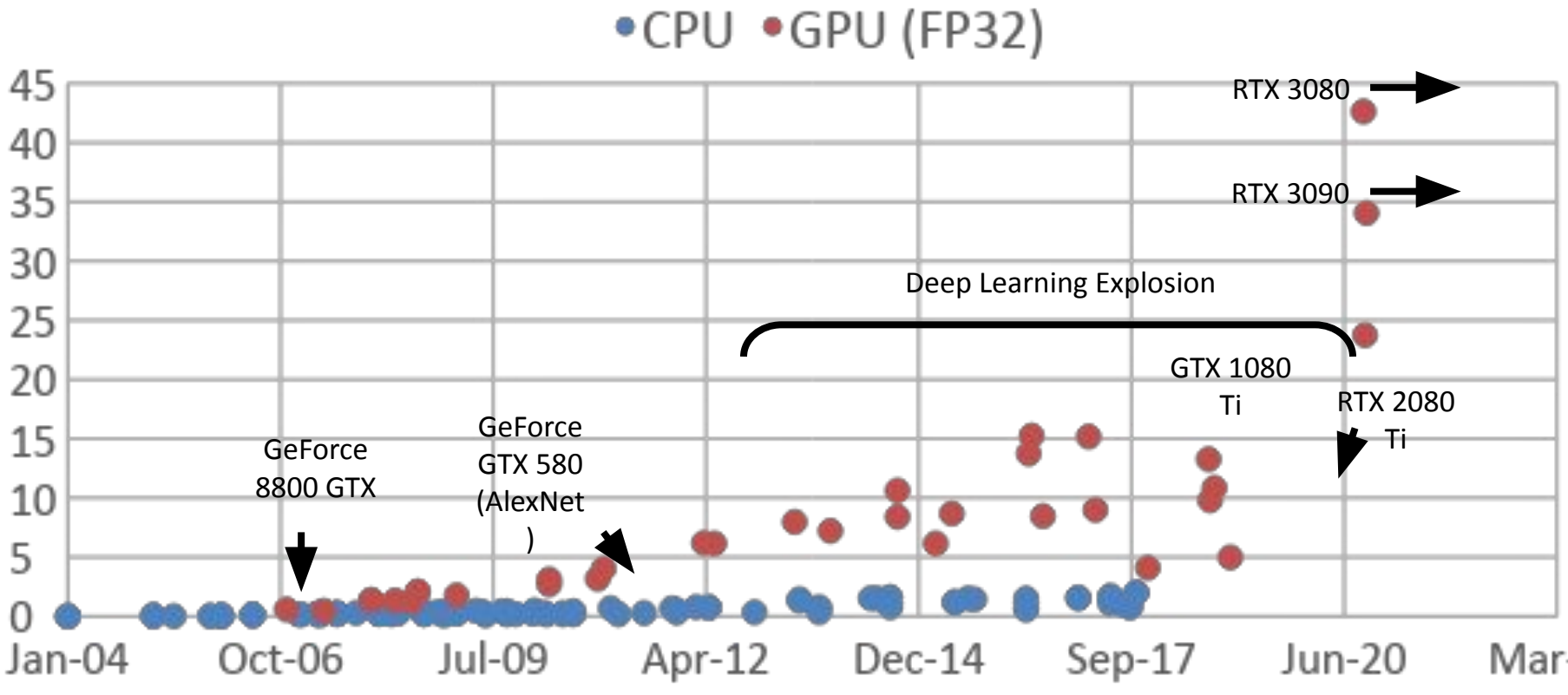
Computation



Algorithms

Data

GFLOP per Dollar

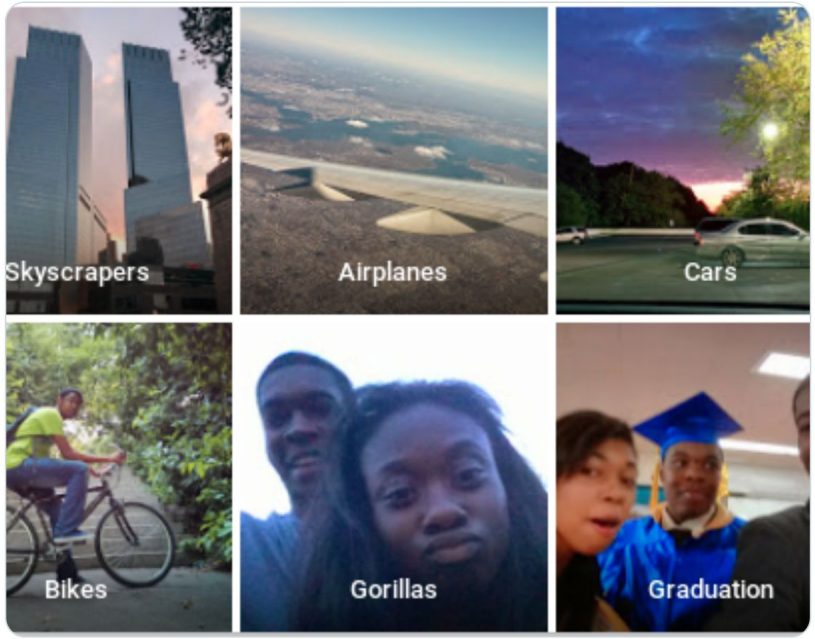


Slide inspiration: Justin Johnson

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

Computer Vision Can Cause Harm

Harmful Stereotypes



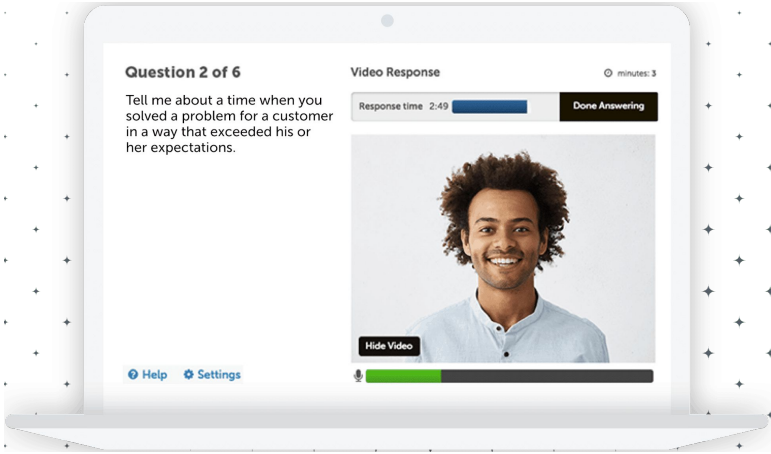
Barocas et al, "The Problem With Bias: Allocative Versus Representational Harms in Machine Learning", SIGCIS 2017
Kate Crawford, "The Trouble with Bias", NeurIPS 2017 Keynote
Source: https://twitter.com/jackyalcine/status/615329515909156865 (2015)

Affect people's lives

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

Slide inspiration: Justin Johnson

CS231n:
Deep Learning for Computer Vision

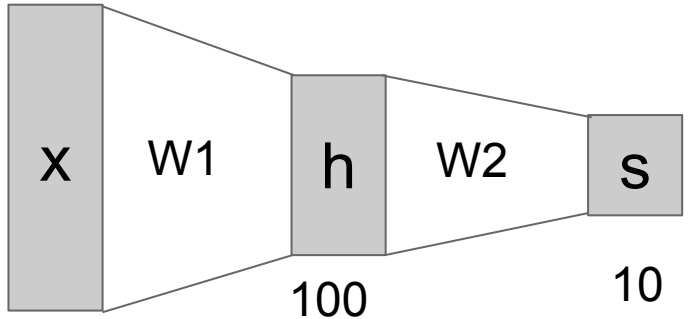
Lecture 1 - Overview

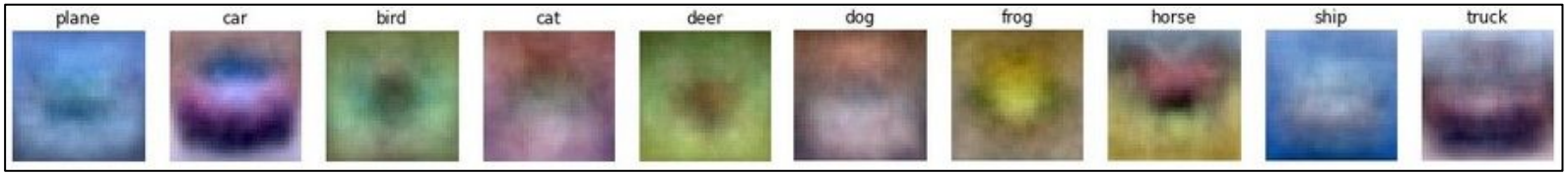
There are many visual recognition problems that are related to image classification, such as object detection, image captioning, image segmentation, visual question answering, visual instruction navigation, video understanding, etc.

Deep Learning for Computer Vision

Hierarchical computing systems with many “layers”, that are very loosely inspired by Neuroscience

Neural Networks

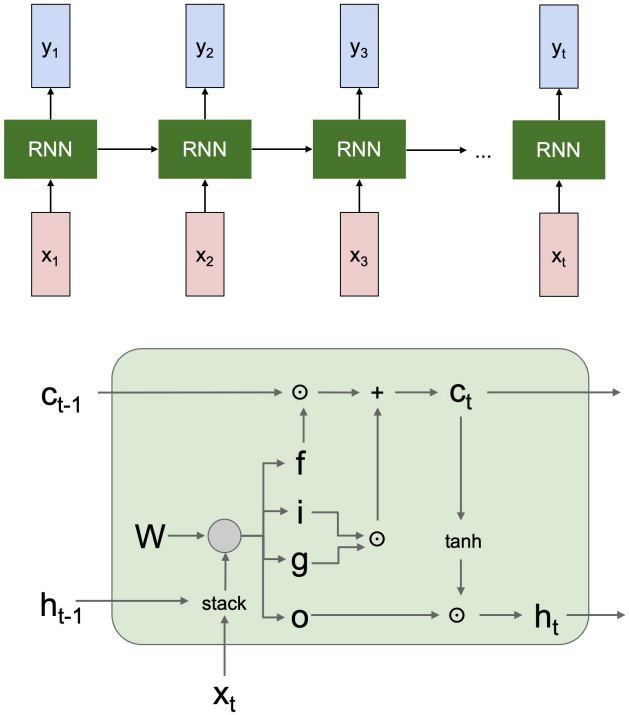




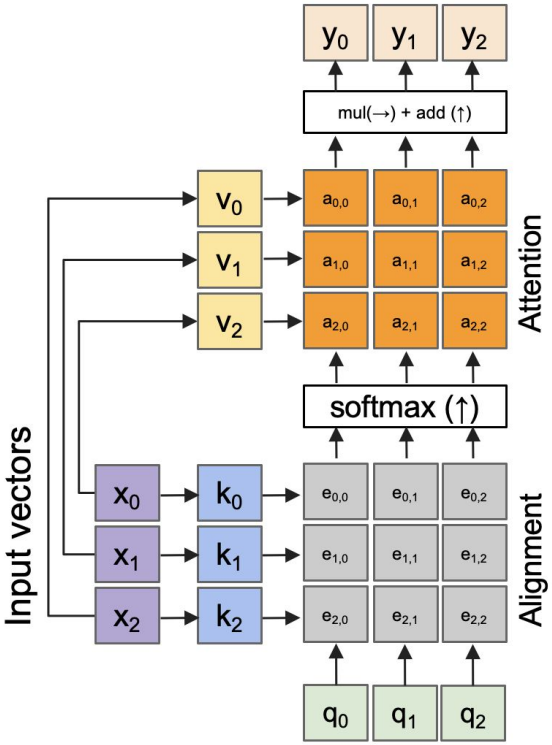
Convolutional Neural Networks for Visual Recognition

A class of Neural Networks that have become an important tool for visual recognition

Beyond Convolutional Neural Networks



Recurrent neural network



Attention mechanism / Transformers

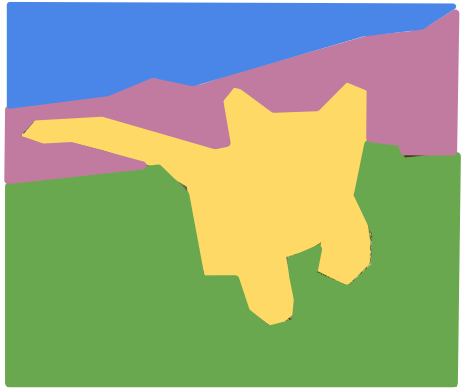
Beyond Image Classification

Classification

CAT

No spatial extent

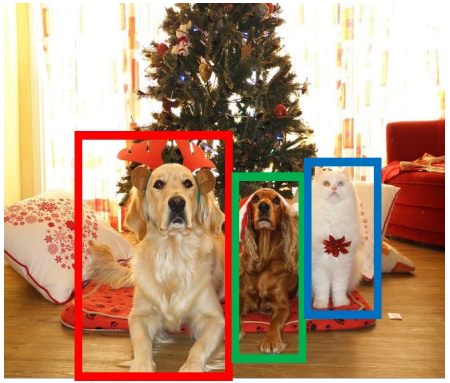
Semantic Segmentation



GRASS, **CAT**, **TREE**, **SKY**

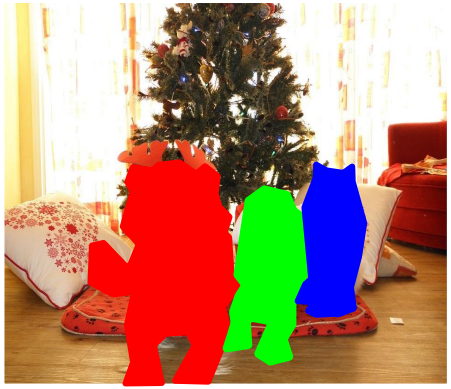
No objects, just pixels

Object Detection



DOG, **DOG**, **CAT**

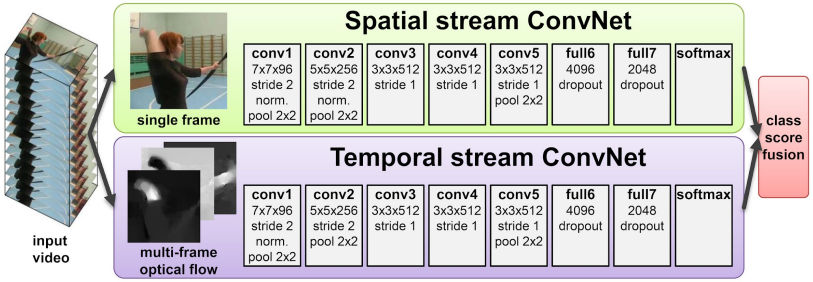
Instance Segmentation



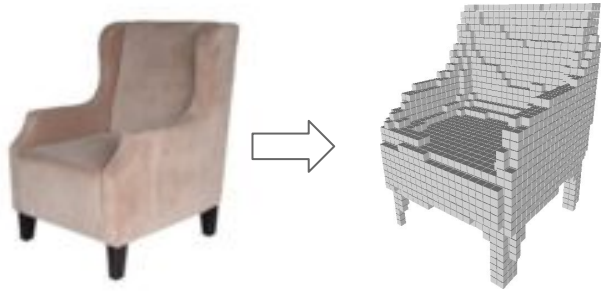
DOG, **DOG**, **CAT**

Multiple Object

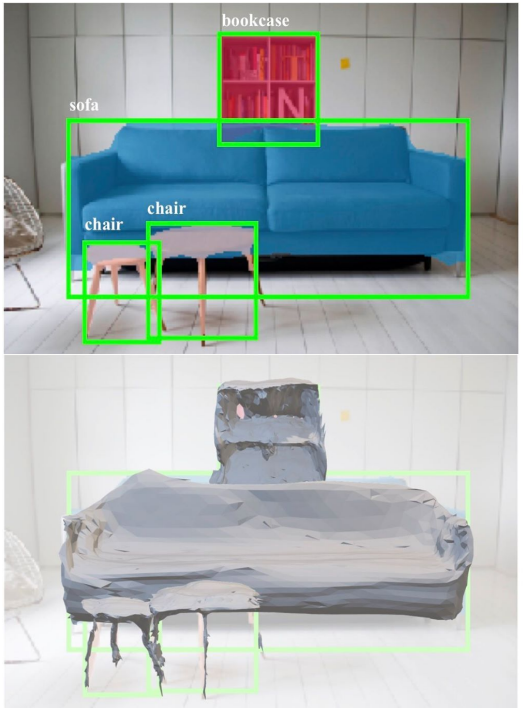
Beyond 2D Images



Simonyan and Zisserman, “Two-stream convolutional networks for action recognition in videos”, NeurIPS 2014

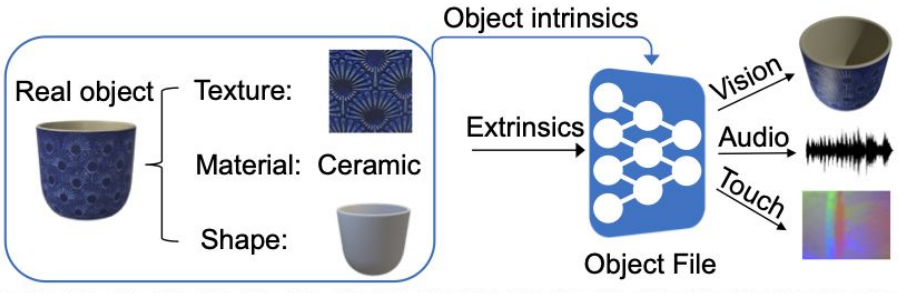


Choy et al., 3D-R2N2: Recurrent Reconstruction Neural Network (2016)



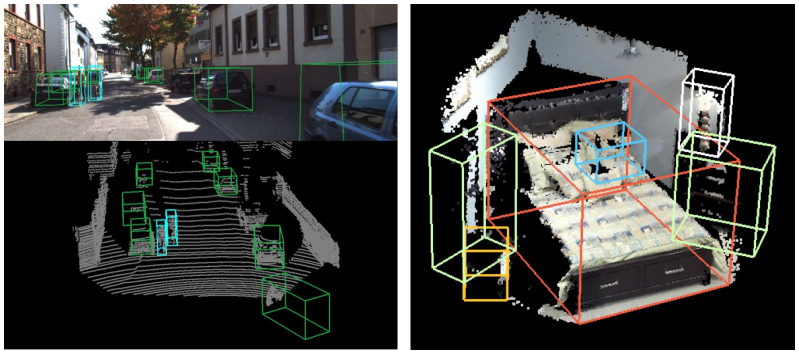
Gkioxari et al., “Mesh R-CNN”, ICCV 2019

Beyond Vision

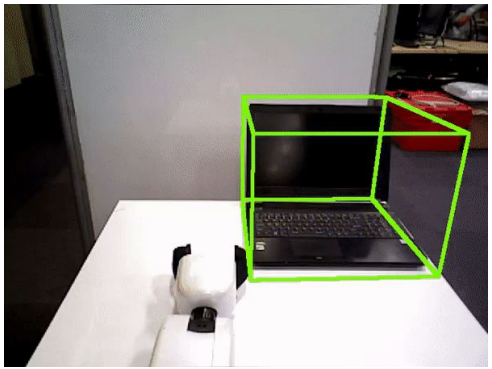


Gao et al., ObjectFolder 2.0: A Multisensory Object Dataset for Sim2Real Transfer (2022)

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



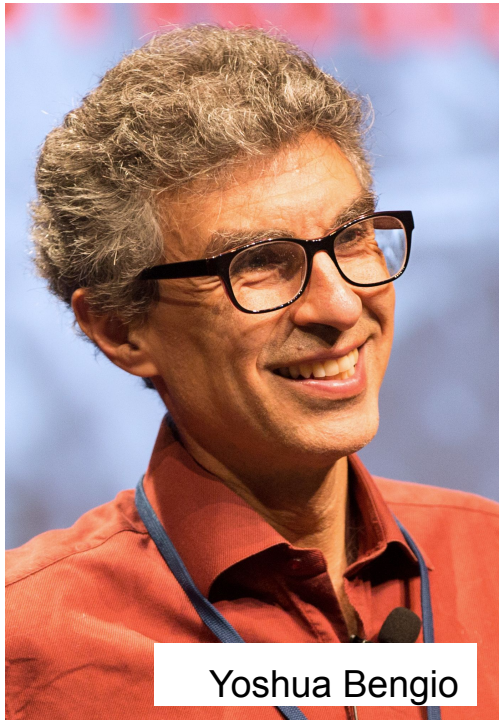
Xu et al., PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation (2018)

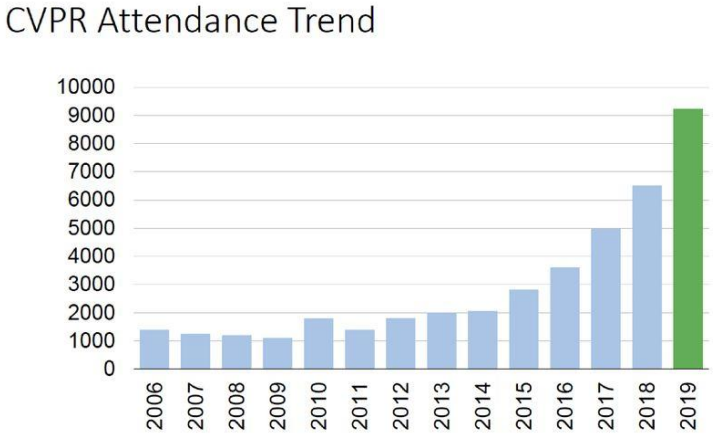


Wang et al., 6-PACK: Category-level 6D Pose Tracker with Anchor-Based Keypoints (2020)

2018 Turing Award for deep learning

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





>8k submissions, 2,067 accepted papers

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

Syllabus

Deep Learning Basics

Data-driven learning
Linear classification & kNN
Loss functions
Optimization
Backpropagation
Multi-layer perceptrons
Neural Networks

Convolutional Neural Networks

Convolutions
PyTorch / TensorFlow
Activation functions
Batch normalization
Transfer learning
Data augmentation
Momentum / RMSProp / Adam
Architecture design

Computer Vision Applications

RNNs / Attention / Transformers
Image captioning
Object detection and segmentation
Style transfer
Video understanding
Generative models
Self-supervised learning
3D vision
Human-centered AI
Fairness & ethics