Visual recognition

370: Intro to Computer Vision

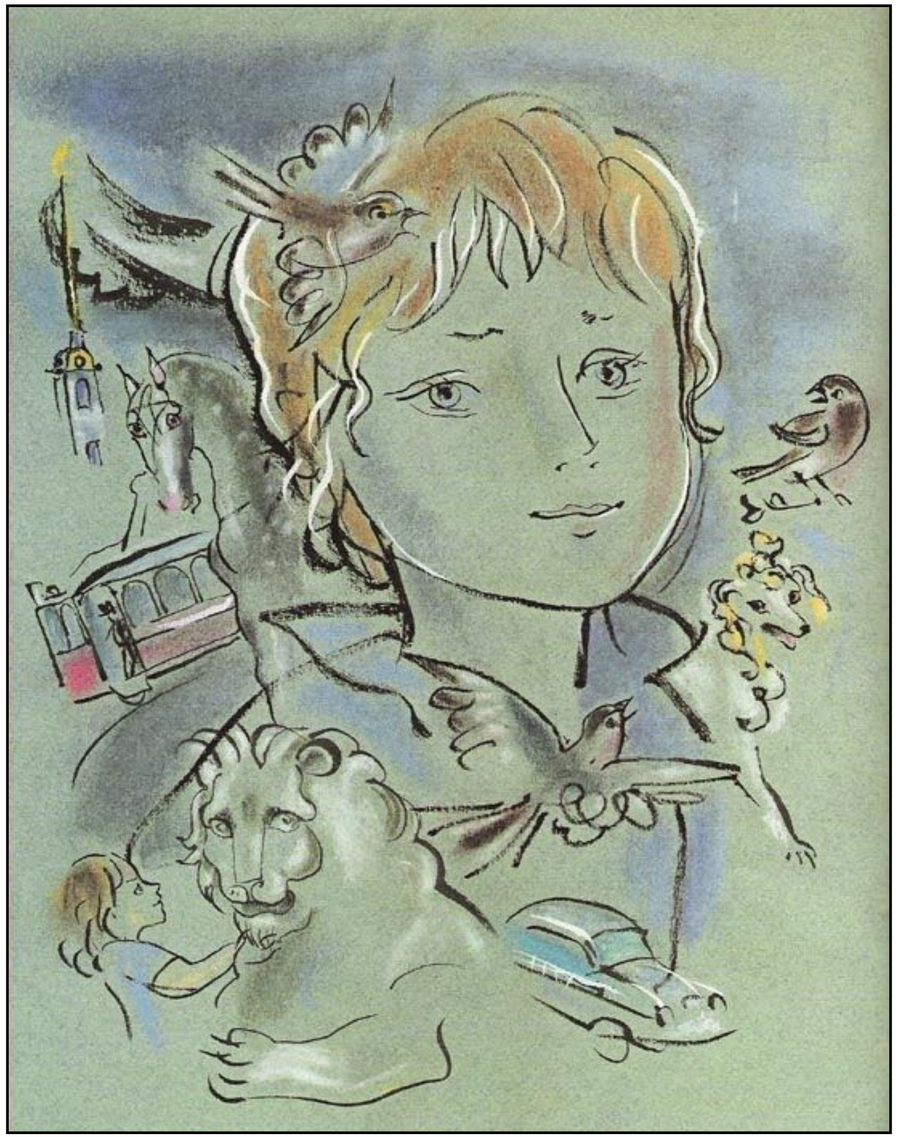
Subhransu Maji

April 10, 2025

College of
INFORMATION AND
COMPUTER SCIENCES

UMASS
AMHERST

A brief history of recognition …



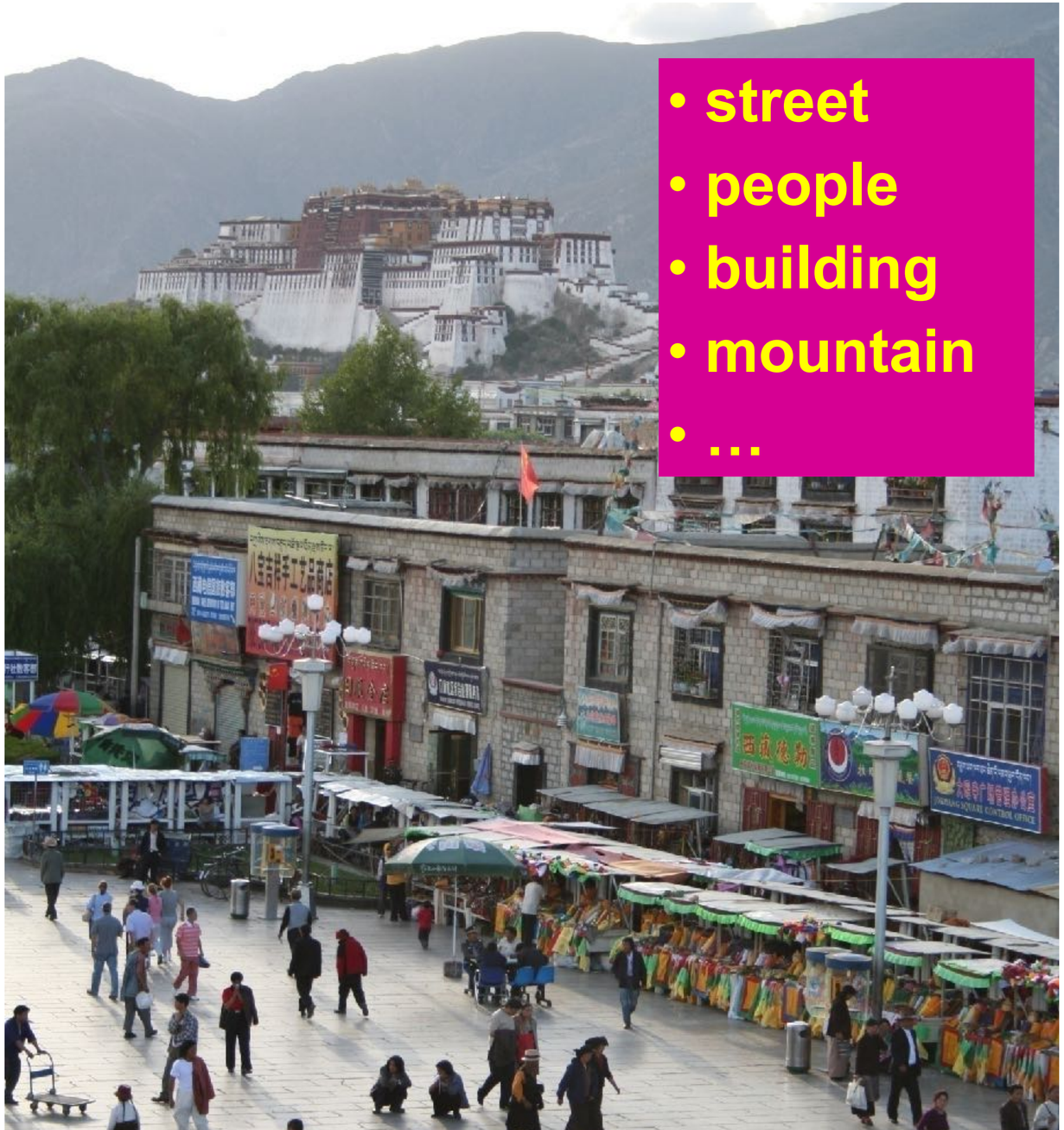
Slides adapted from Svetlana Lazebnik, Alex Berg, Fei-Fei Li, Rob Fergus, Antonio Torralba, and Jean Ponce

What does it mean to recognize?

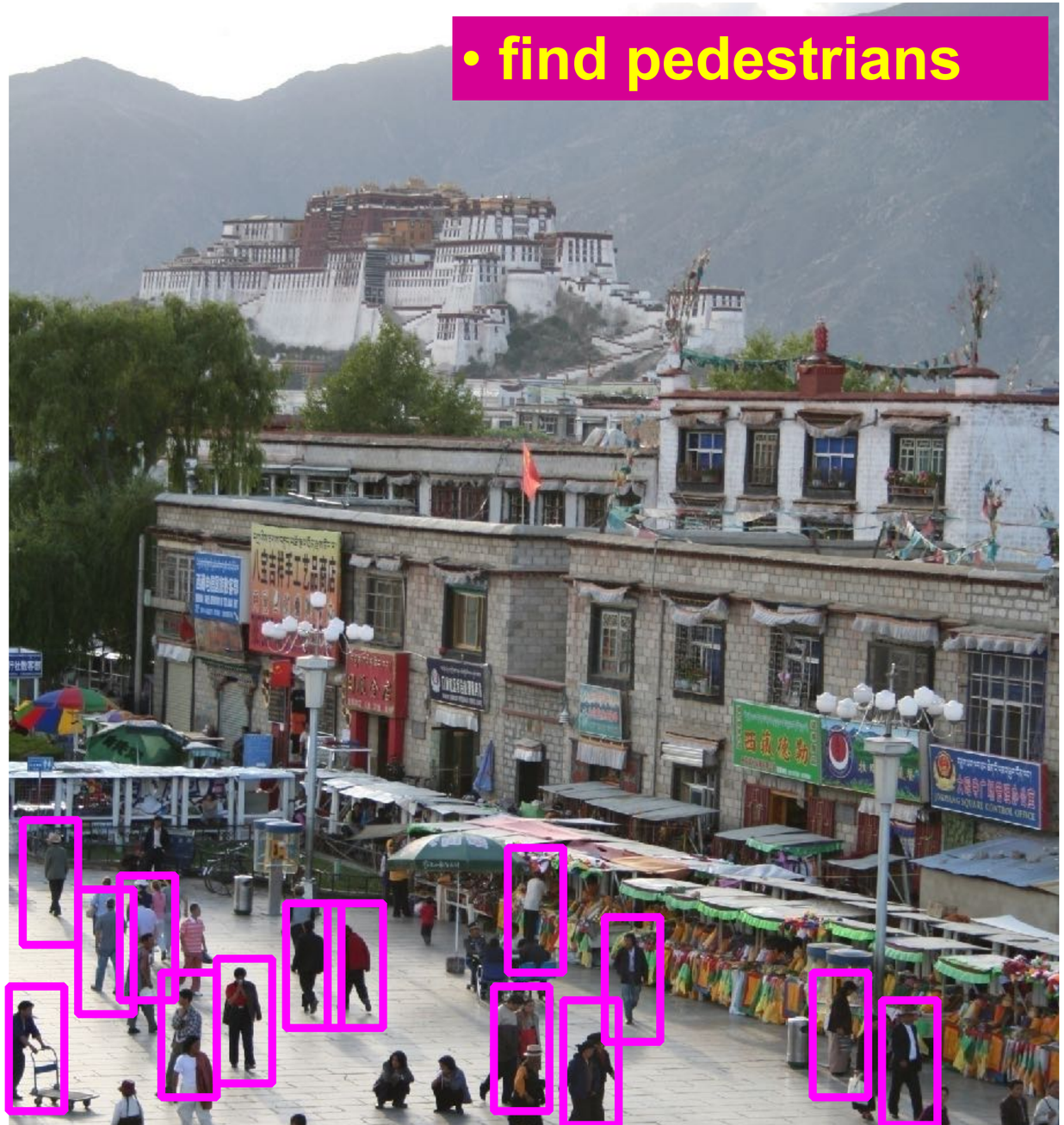
Scene categorization

- outdoor/indoor
- city/forest/factory/etc.

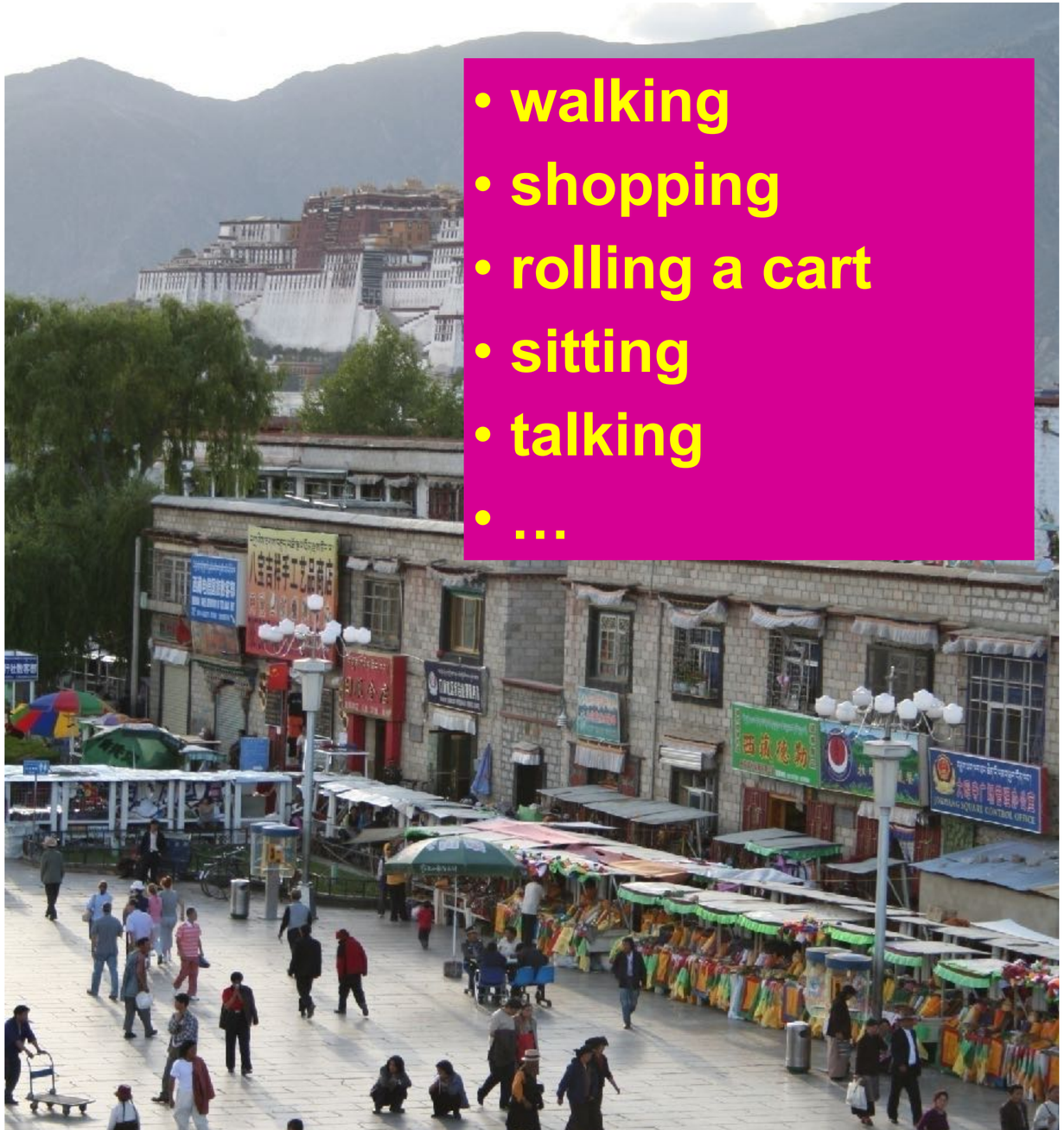
Image annotation or tagging



Object detection



Activity recognition

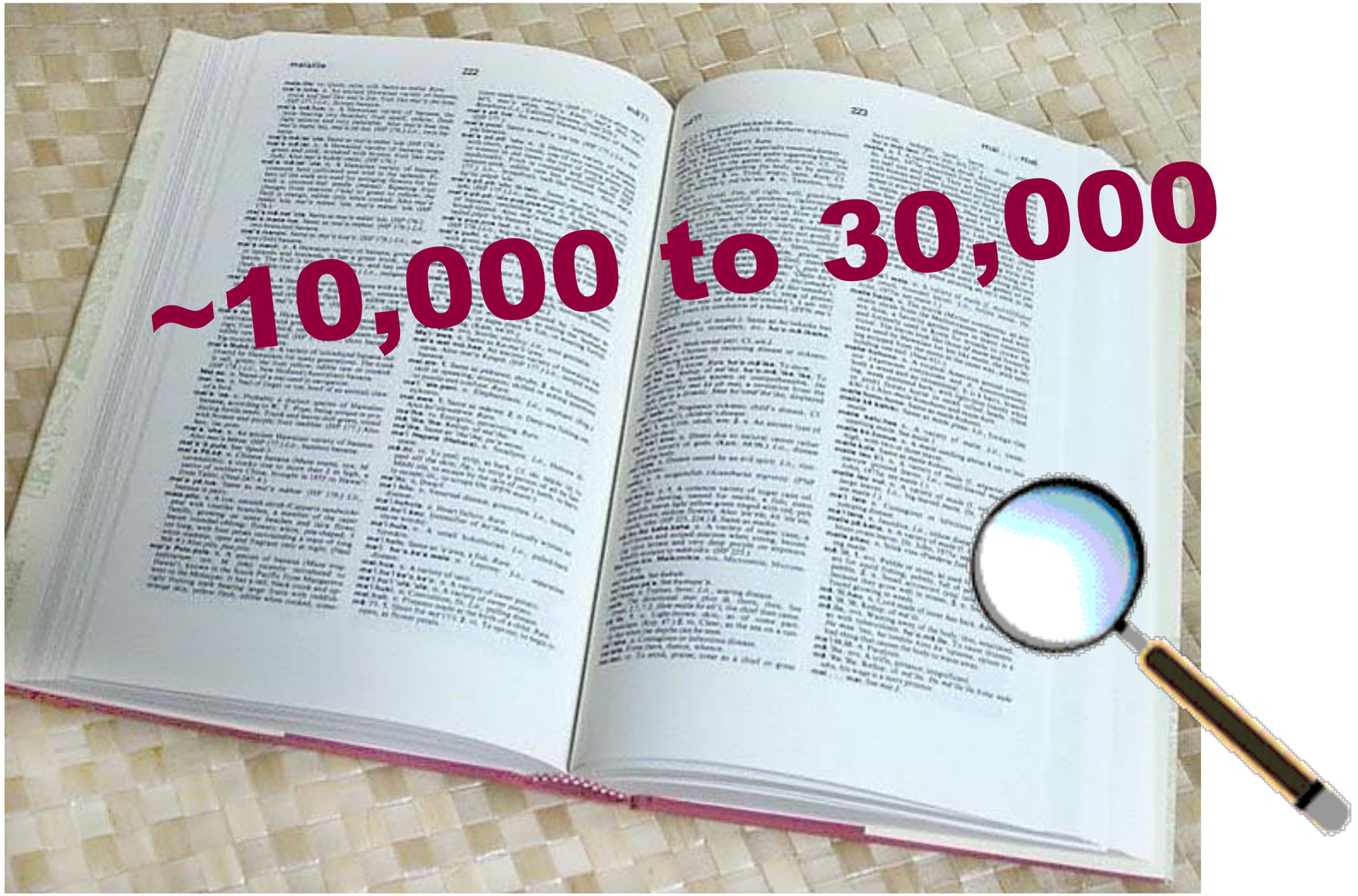


- walking
- shopping
- rolling a cart
- sitting
- talking
- …

Image parsing

Image understanding?

How many categories are there?



http://wexler.free.fr/library/files/biederman%20(1987)%20recognition-by-components.%20a%20theory%20of%20human%20image%20understanding.pdf

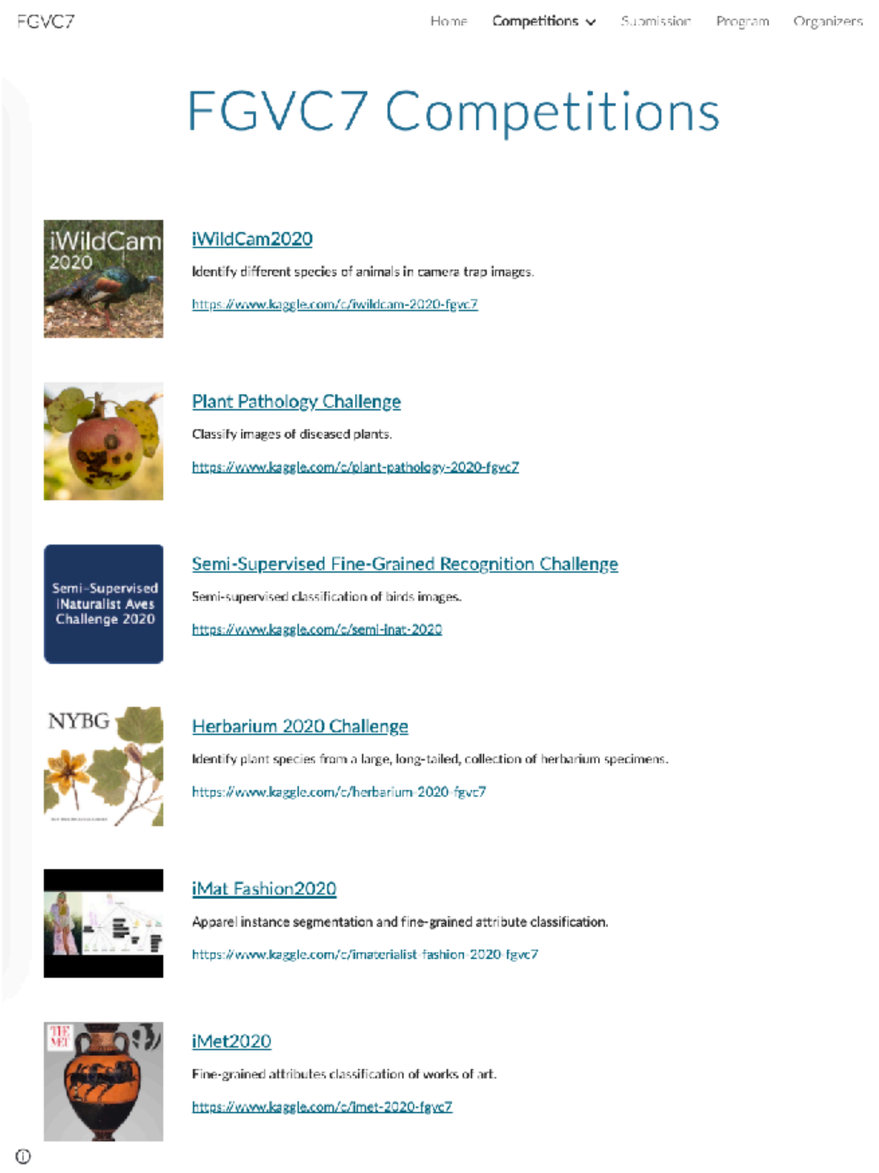
Biederman 1987


~10,000 to 30,000

How many categories?

https://hackercombat.com/free-facial-recognition-tool-to-track-people-on-social-media-sites/

https://osxdaily.com/2017/11/10/can-use-iphone-x-without-face-id/



How many faces can an average person recognize?

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Jenkins R, Dowsett AJ, Burton AM. 2018 How many faces do people know? *Proc. R. Soc. B* **285**: 20181319.
http://dx.doi.org/10.1098/rspb.2018.1319

Received: 13 June 2018
Accepted: 17 September 2018

How many faces do people know?

R. Jenkins[1], A. J. Dowsett[2] and A. M. Burton[1]

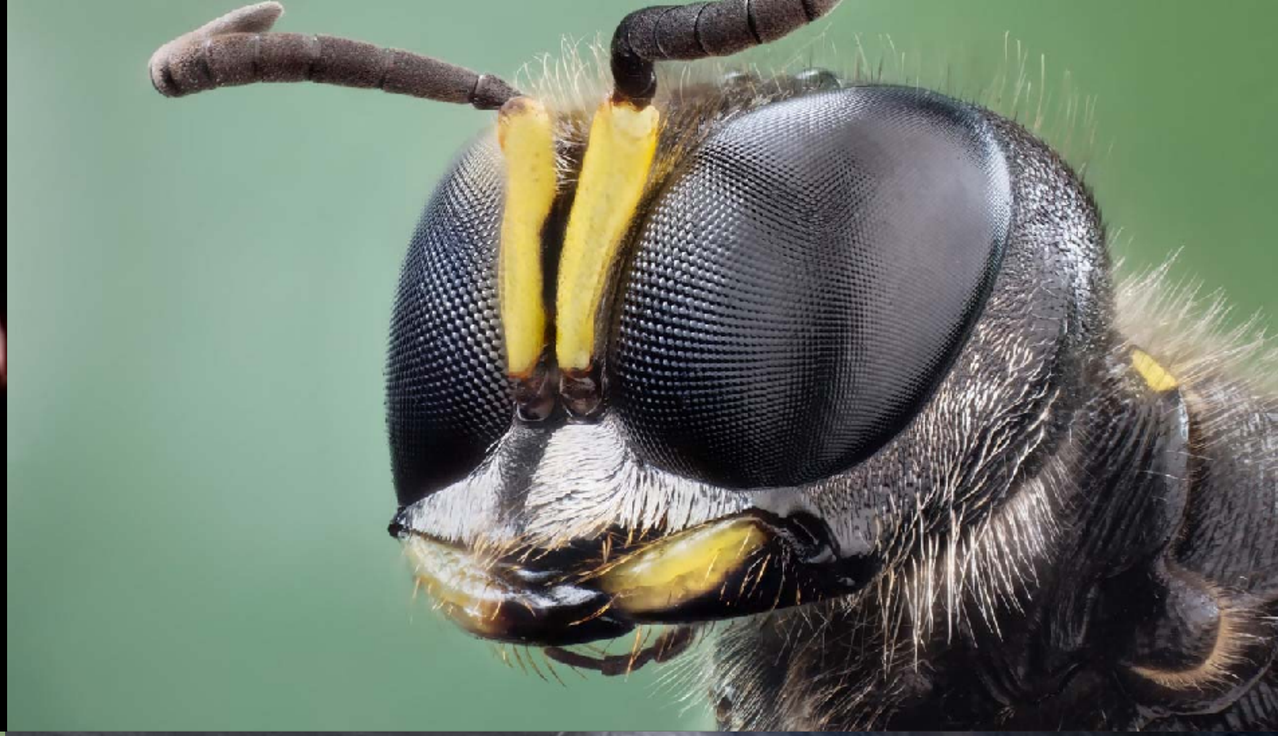
[1]Department of Psychology, University of York, York YO10 5DD, UK
[2]School of Psychology, University of Aberdeen, Aberdeen, UK

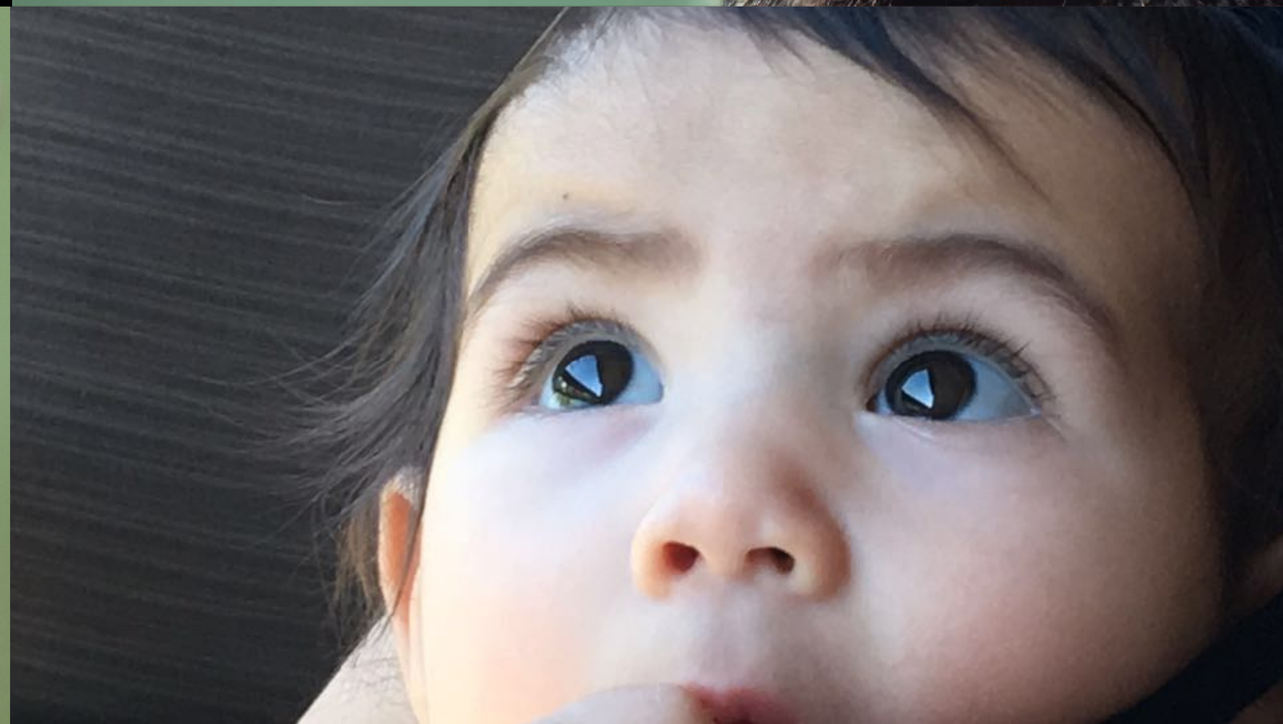
RJ, 0000-0003-4793-0435

Over our species history, humans have typically lived in small groups of under a hundred individuals. However, our face recognition abilities appear to equip us to recognize very many individuals, perhaps thousands. Modern society provides access to huge numbers of faces, but no one has established how many faces people actually know. Here, we describe a method for estimating this number. By combining separate measures of recall and recognition, we show that people know about 5000 faces on average and that individual differences are large. Our findings offer a possible explanation for large variation in identification performance. They also provide constraints on understanding the qualitative differences between perception of familiar and unfamiliar faces—a distinction that underlies all current theories of face recognition.

Looking back ….

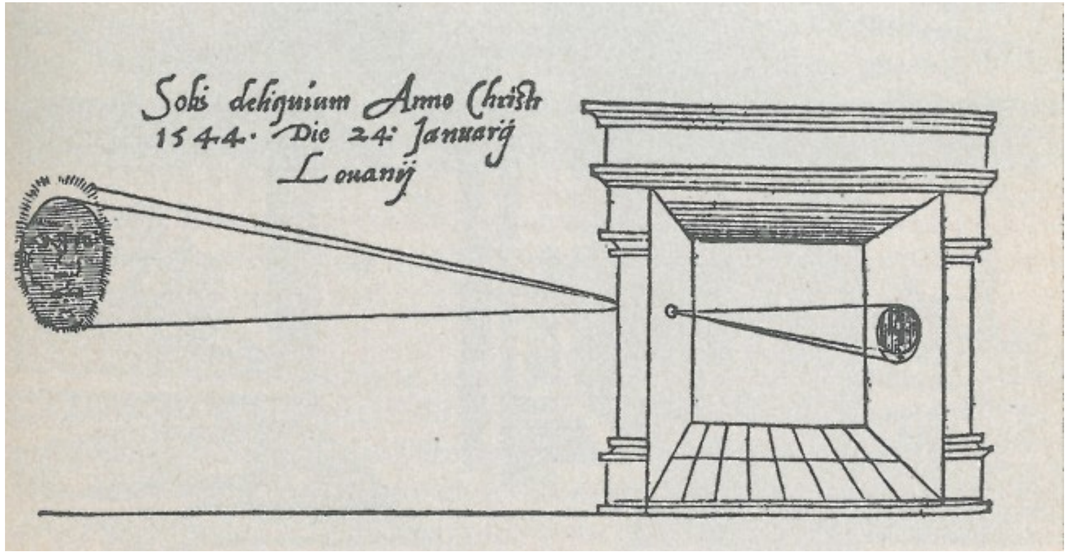
Evolution's Big Bang:
Cambrian Explosion, 530-540million years, B.C.

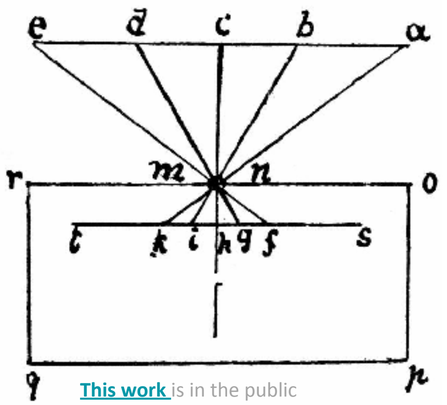




Camera Obscura

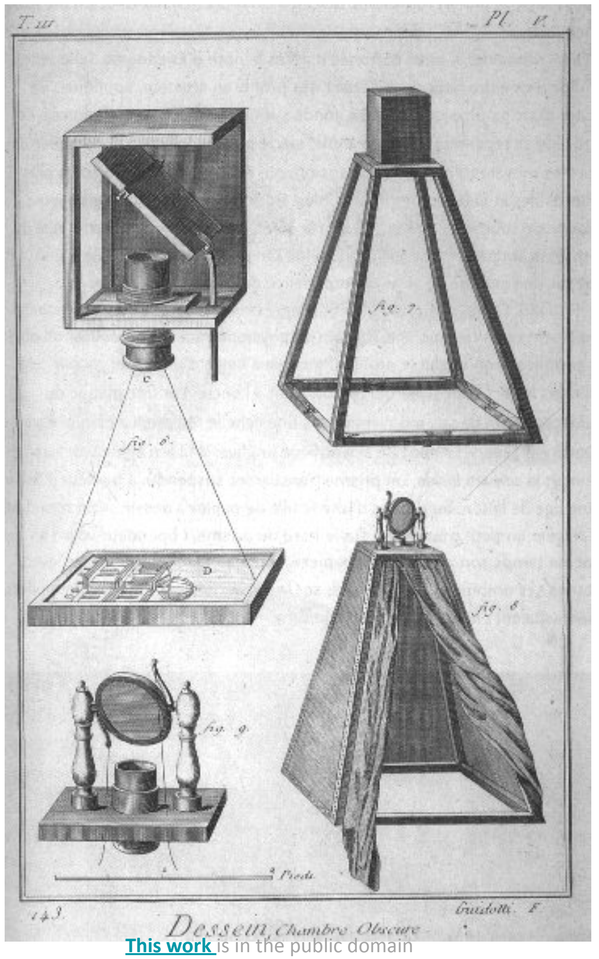
Gemma Frisius, 1545





Leonardo da Vinci,
16th Century AD

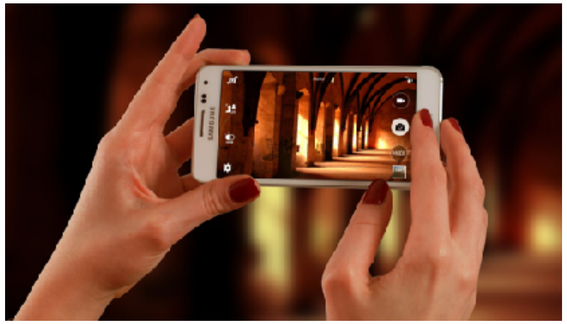
Encyclopedia, 18th Century



Computer Vision is everywhere!

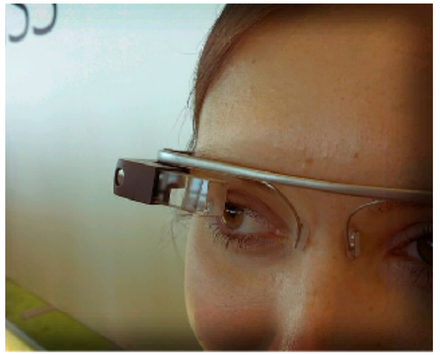



Left to right:
Image by Roger H Goun is licensed under CC BY 2.0
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain




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

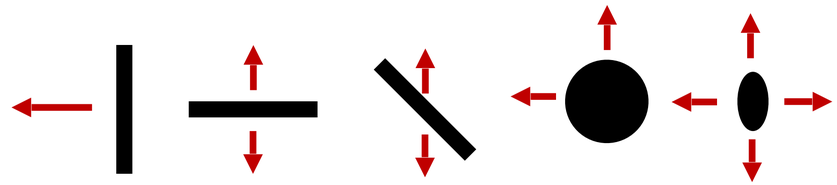


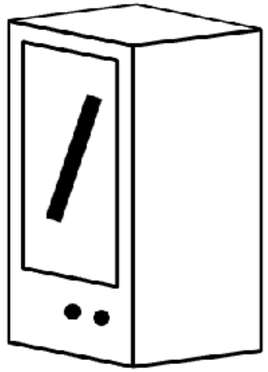
Bottom row, left to right
Image is CC0 1.0 public domain
Image by Derek Keats is licensed under CC BY 2.0; changes made
Image is public domain
Image is licensed under CC-BY 2.0; changes made

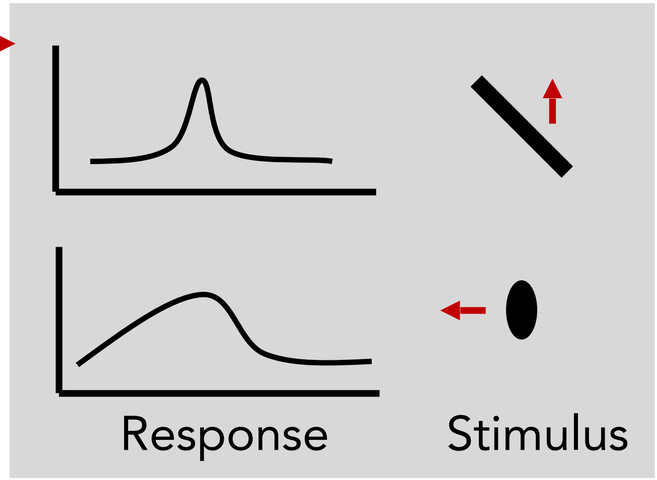
Where did we come from?

Hubel and Wiesel, 1959



Measure brain activity

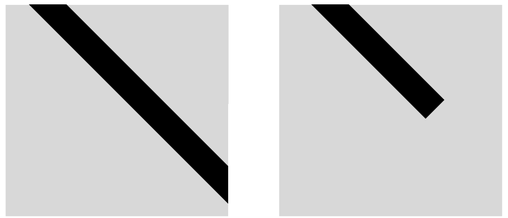




1959
Hubel & Wiesel

Simple cells:
Response to specific rotation and orientation

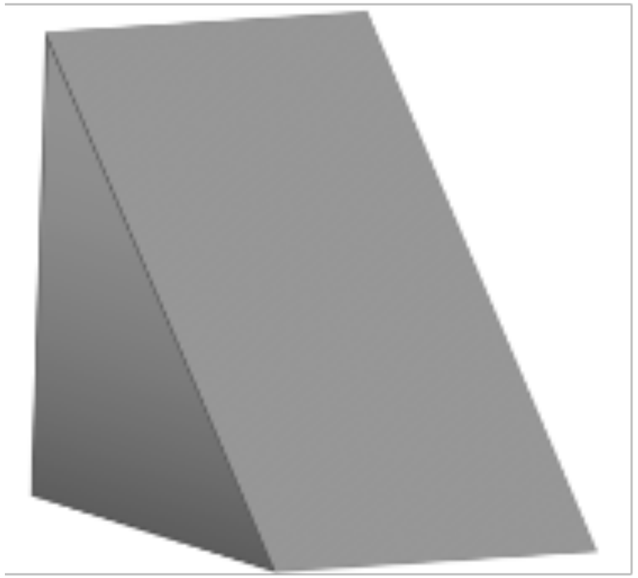
Complex cells:
Response to light orientation and movement, some translation invariance



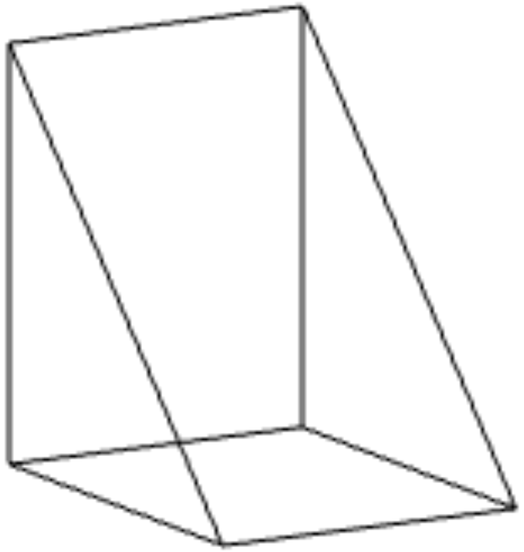
No response

Slide inspiration: Justin Johnson

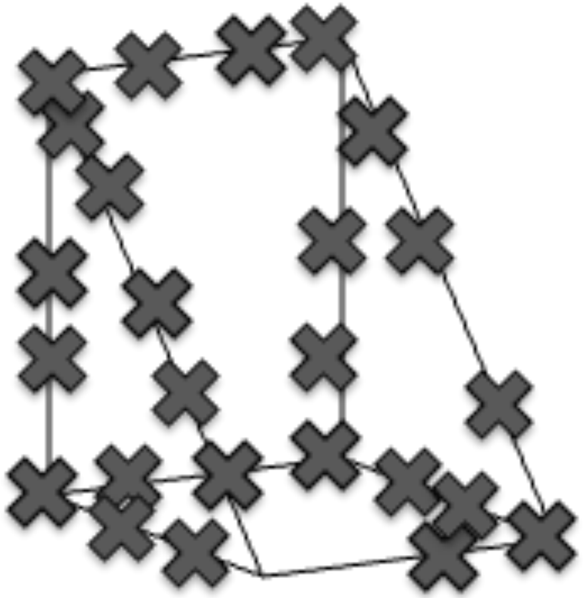
Larry Roberts, 1963



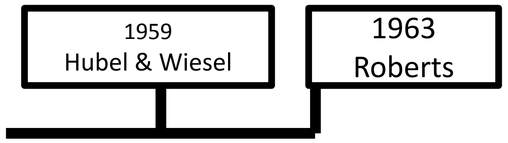
(a) Original picture



(b) Differentiated picture



(c) Feature points selected



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

Slide inspiration: Justin Johnson

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

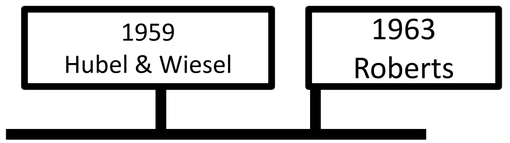
PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

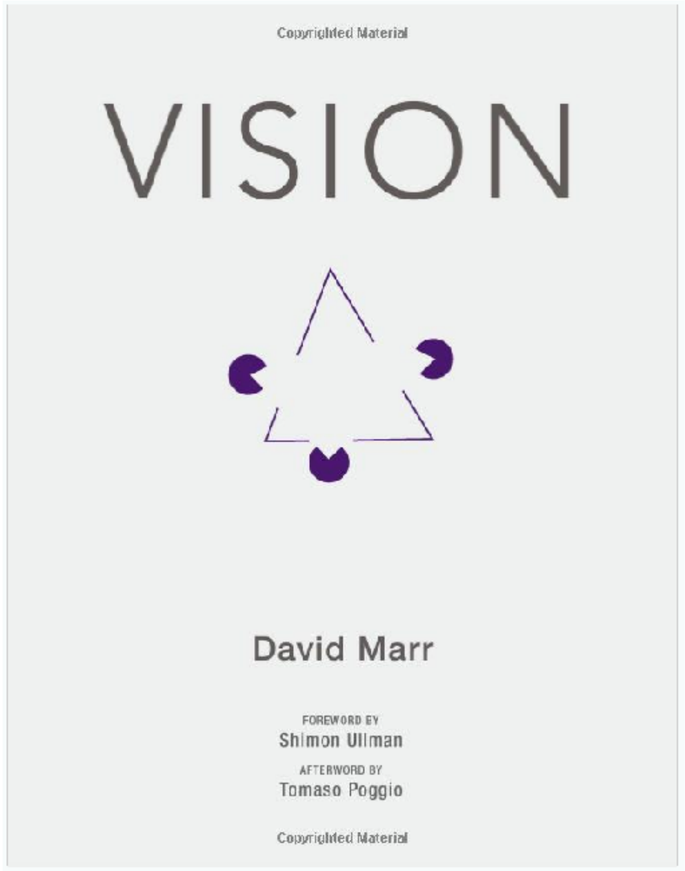
Seymour Papert

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

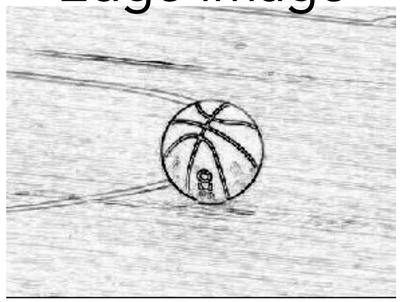


https://dspace.mit.edu/handle/1721.1/6125

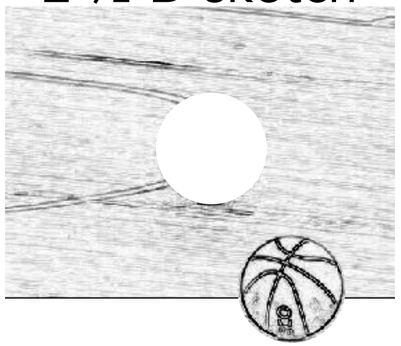
Slide inspiration: Justin Johnson



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

1959 Hubel & Wiesel

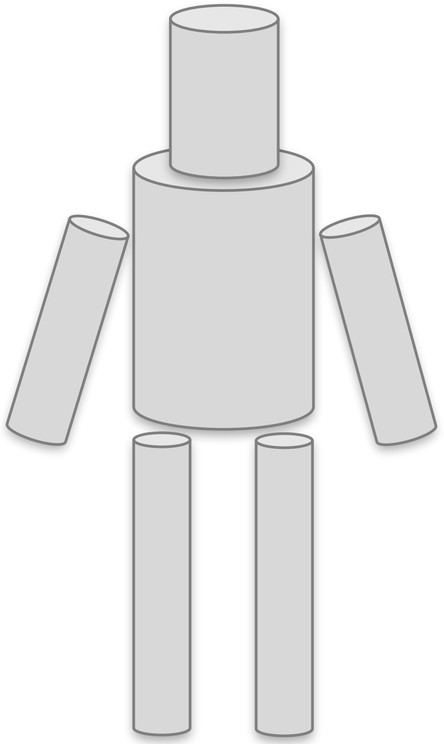
1963 Roberts

1970s David Marr

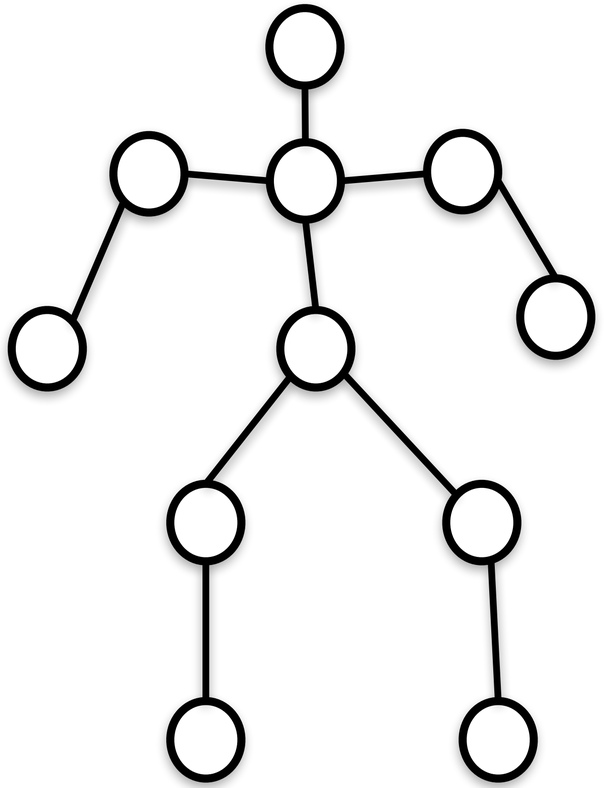
Stages of Visual Representation, David Marr, 1970s

Slide inspiration: Justin Johnson

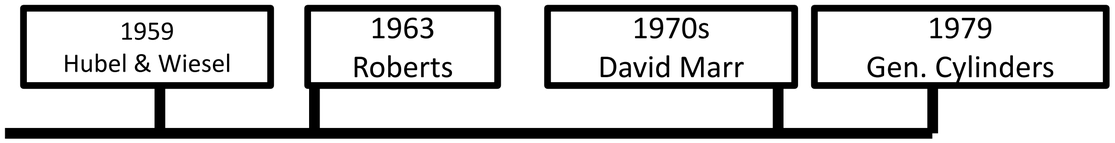
Recognition via Parts (1970s)



Generalized Cylinders, Brooks and Binford, 1979

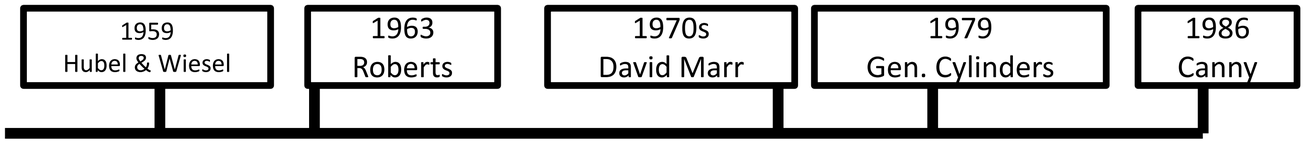


Pictorial Structures, Fischler and Elshlager, 1973



Slide inspiration: Justin Johnson

Recognition via Edge Detection (1980s)

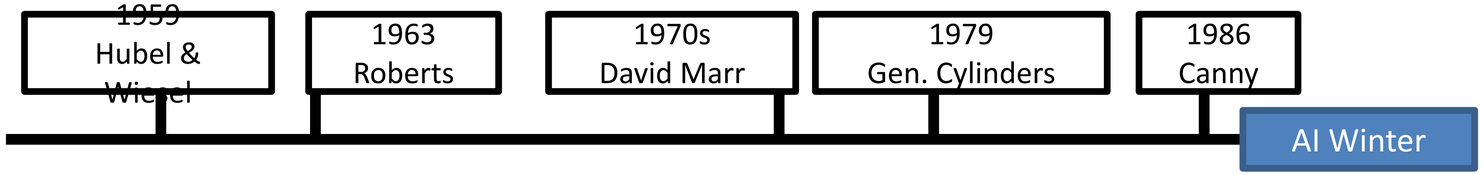


John Canny, 1986
David Lowe, 1987

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.

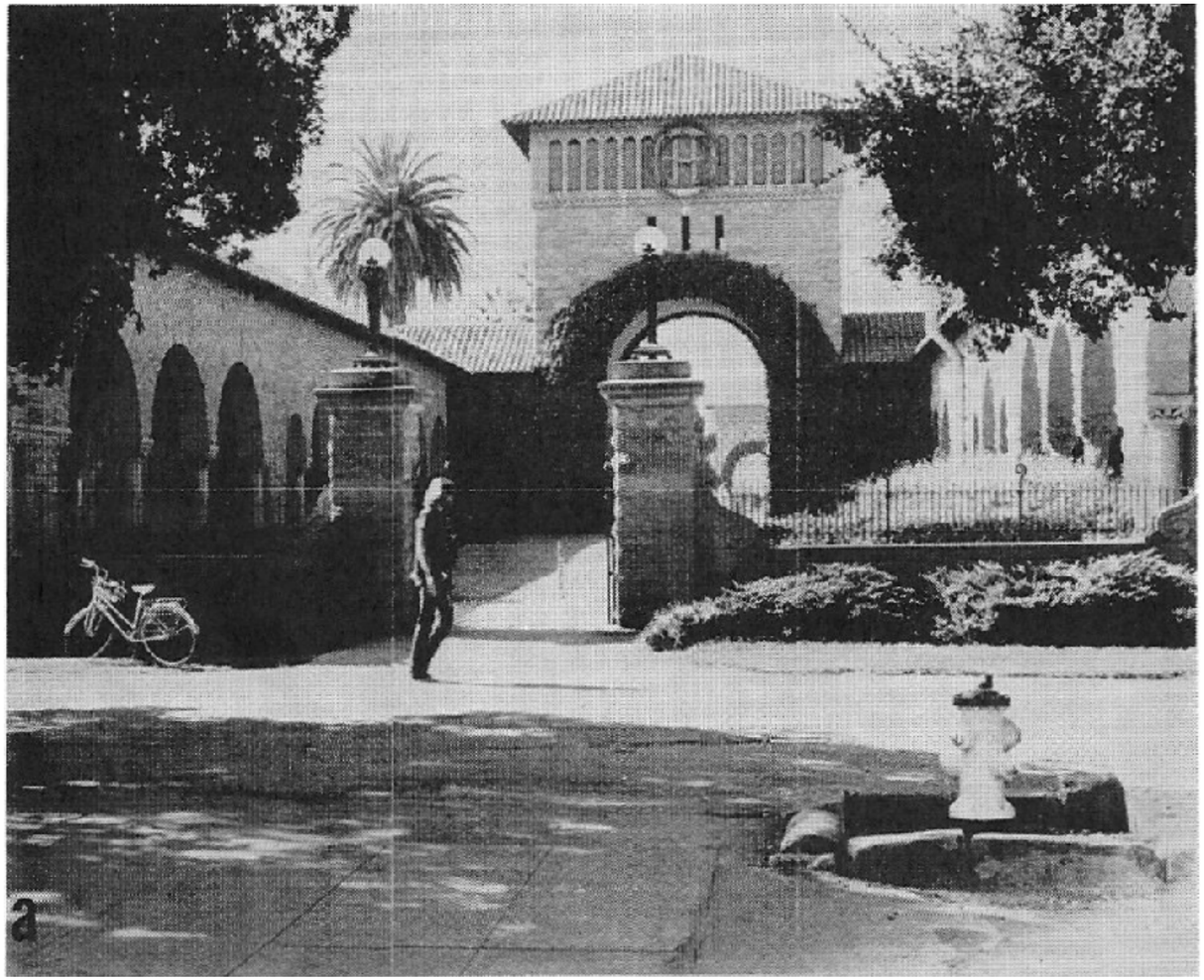


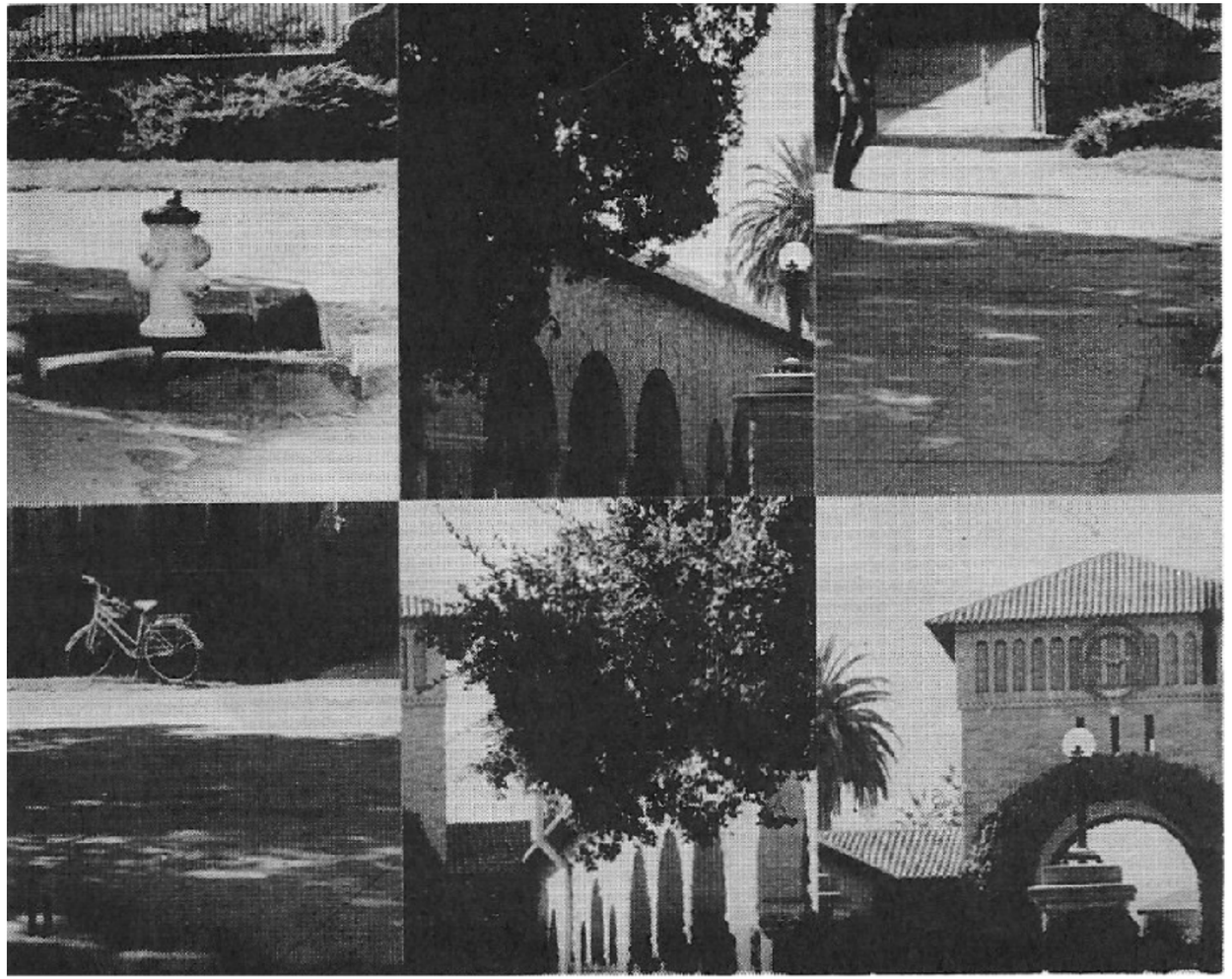
Slide inspiration: Justin Johnson

In the meantime…seminal work in cognitive and neuroscience

Perceiving Real-World Scenes

Irving Biederman





I. Biederman, *Science*, 1972

Rapid Serial Visual Perception (RSVP)

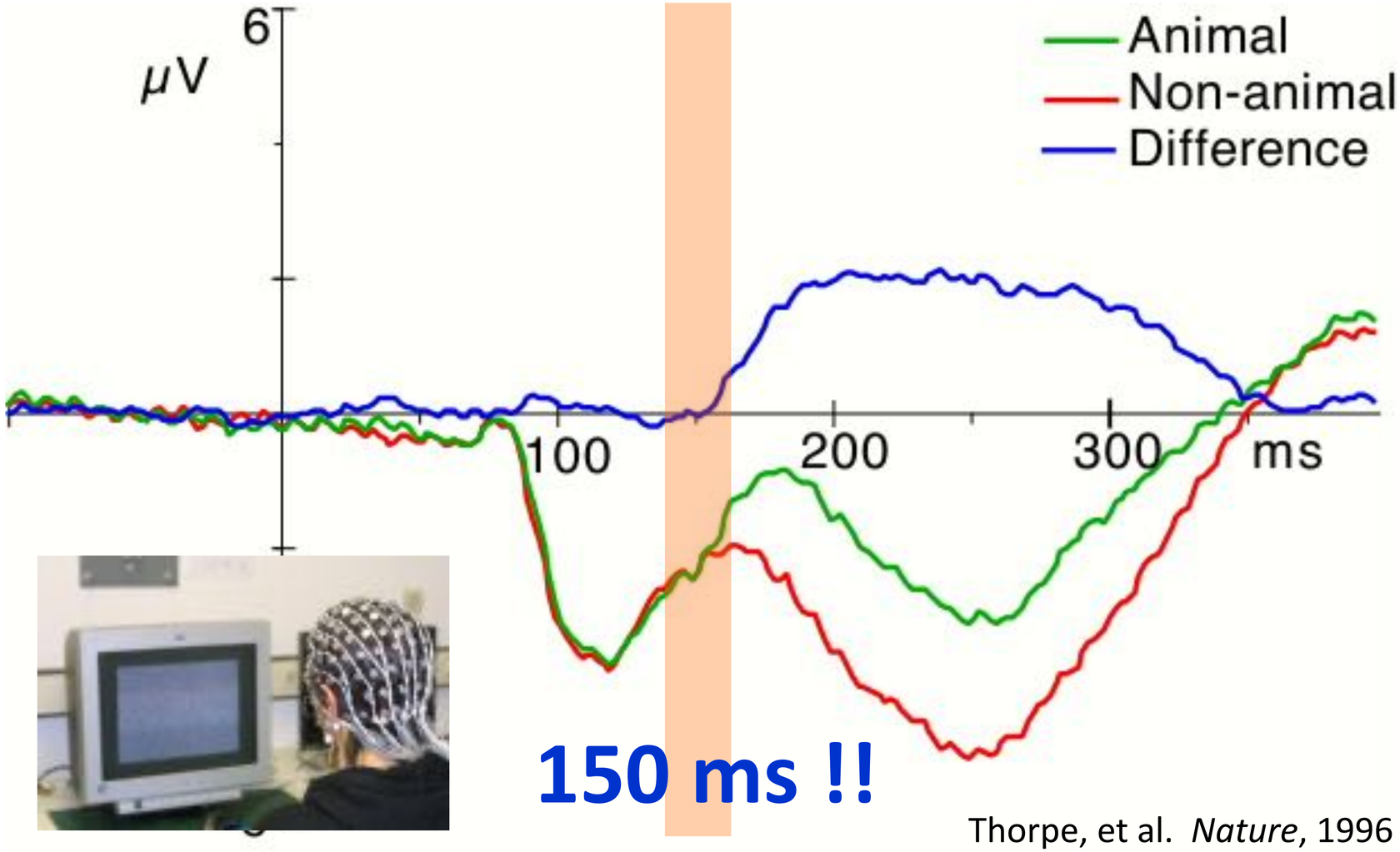
Potter, etc. 1970s






Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot



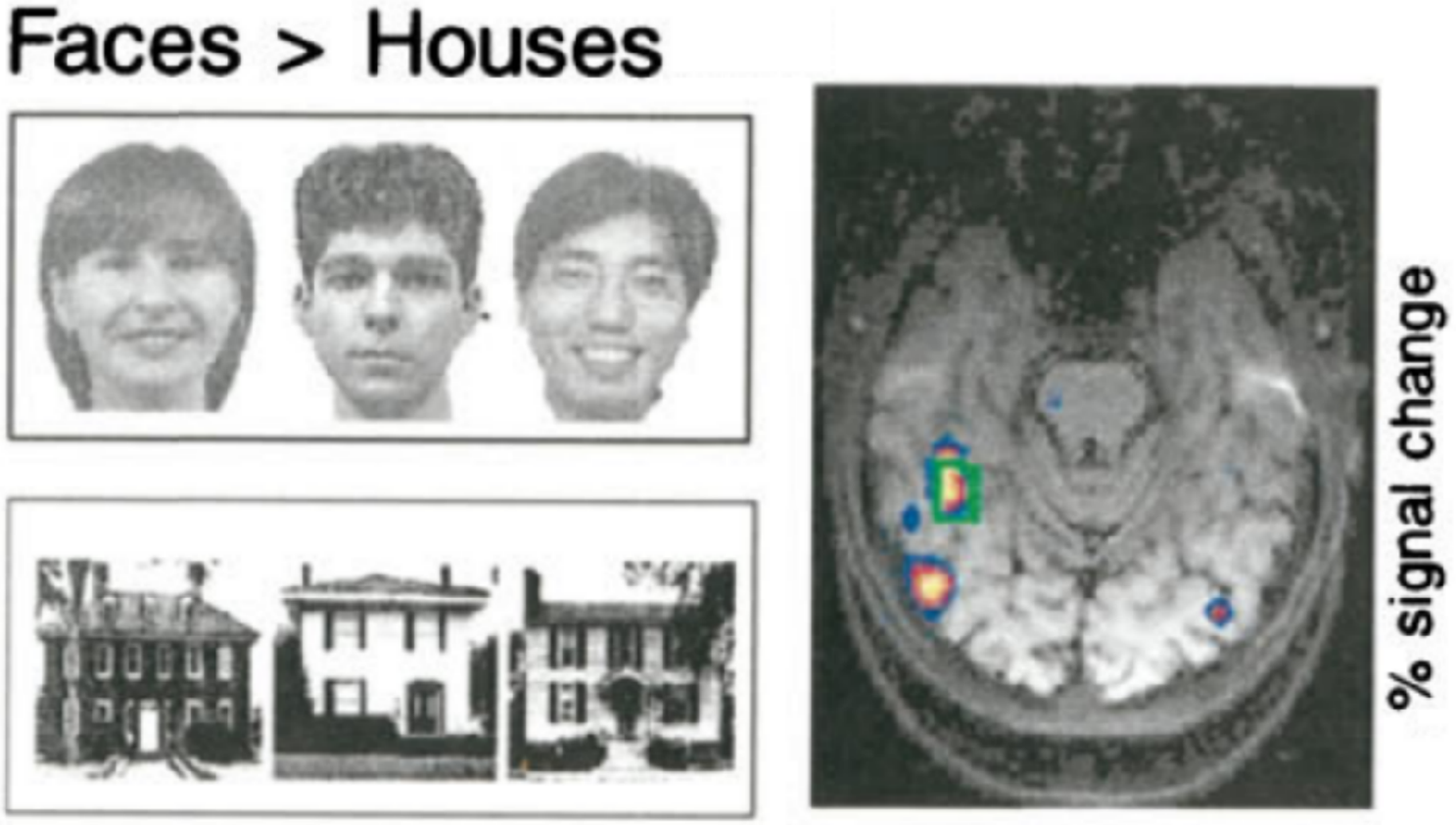
150 ms !!

Thorpe, et al. *Nature*, 1996

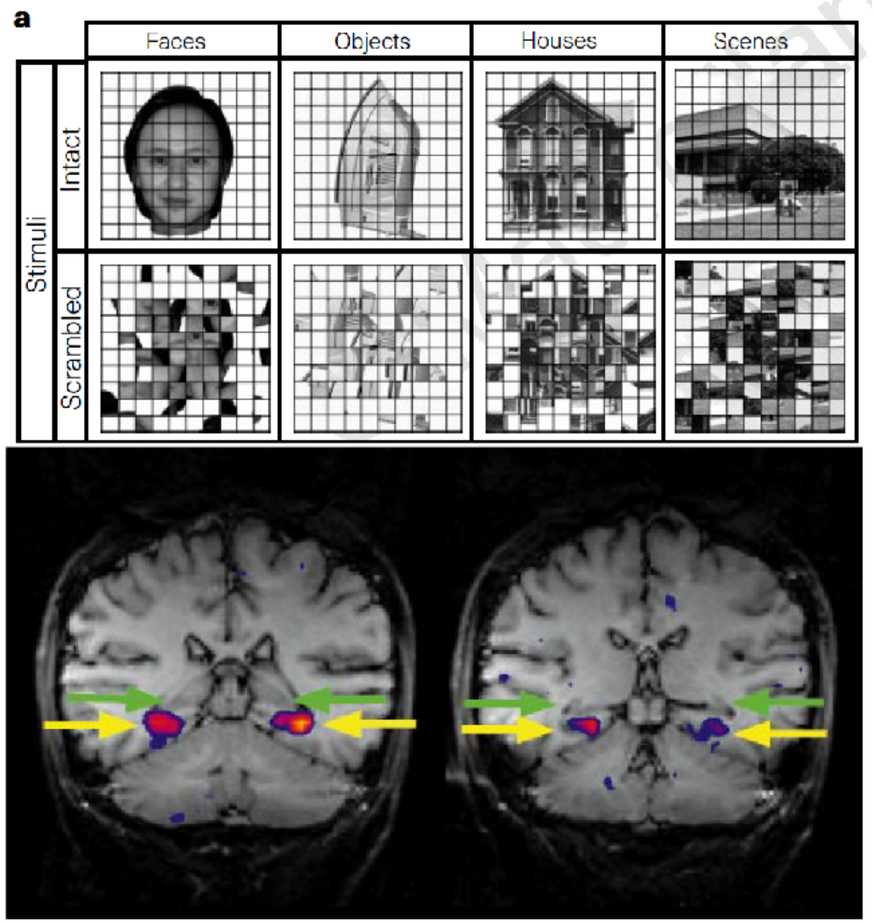




Neural correlates of object & scene recognition

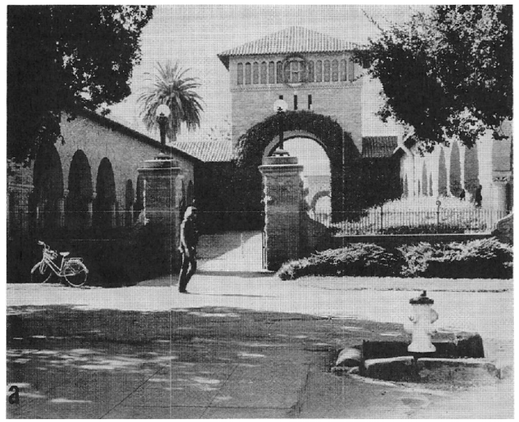


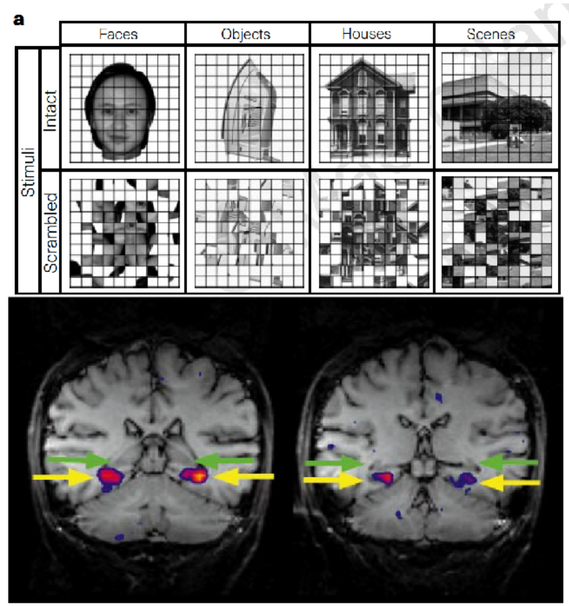
Kanwisher et al. J. Neuro. 1997



Epstein & Kanwisher, Nature, 1998

Visual recognition is a fundamental task for visual intelligence





Recognition via Grouping (1990s)

1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

AI Winter

1997 Norm. Cuts

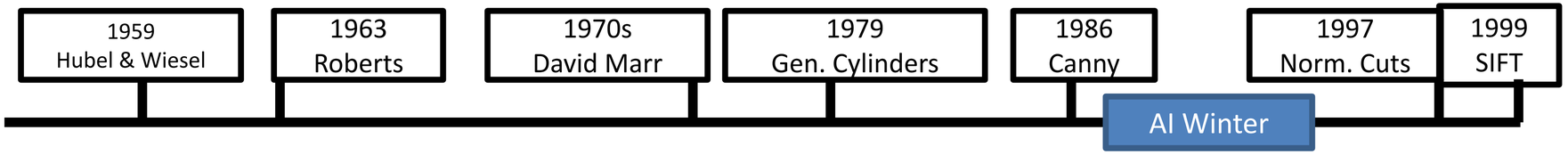
Normalized Cuts, Shi and Malik, 1997

Slide inspiration: Justin Johnson

Recognition via Matching (2000s)

Image is public domain

Image is public domain



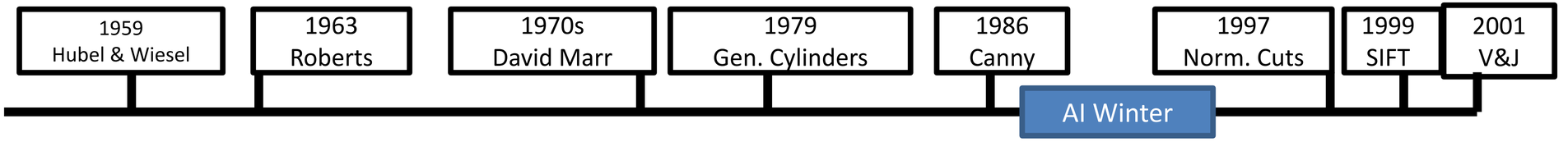
SIFT, David Lowe, 1999

Slide inspiration: Justin Johnson

Face Detection

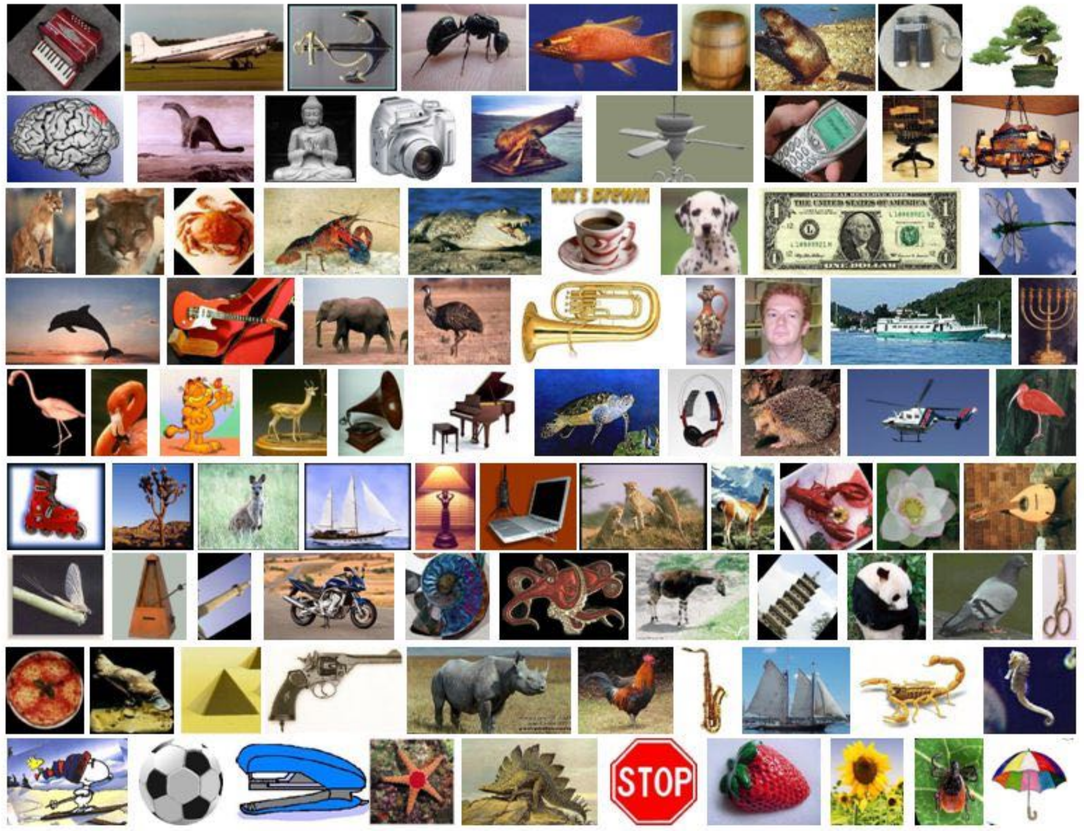
Viola and Jones, 2001

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

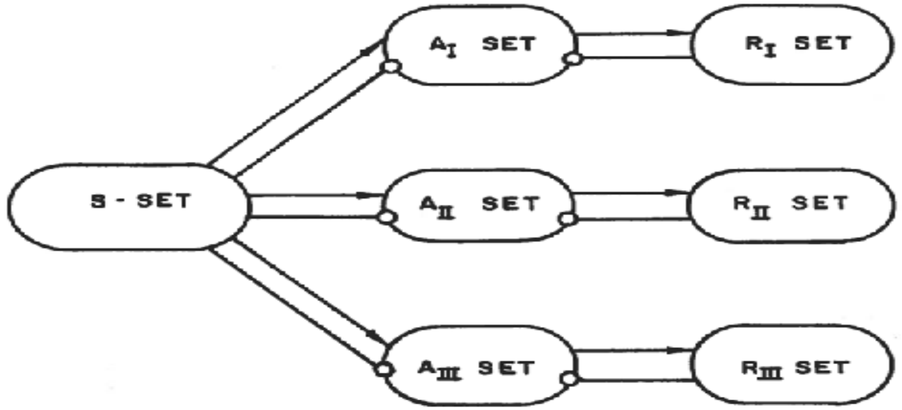
Slide inspiration: Justin Johnson

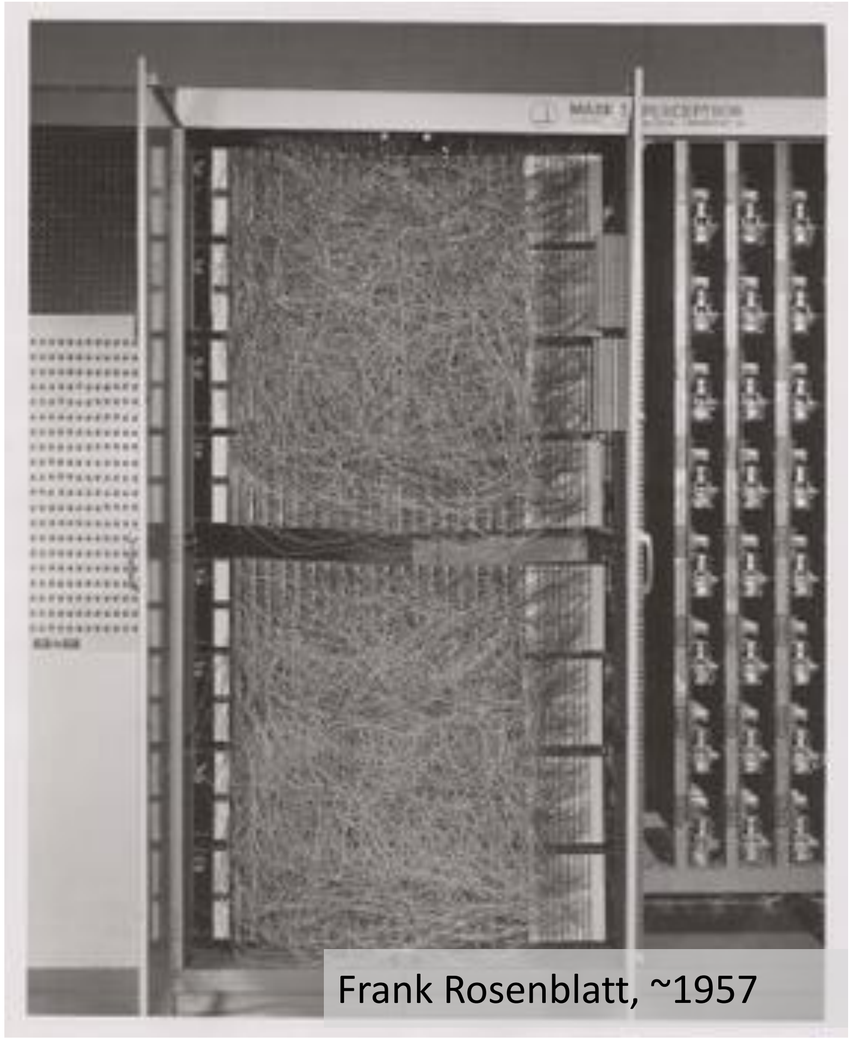
Perceptron

Learning representations by back-propagating errors

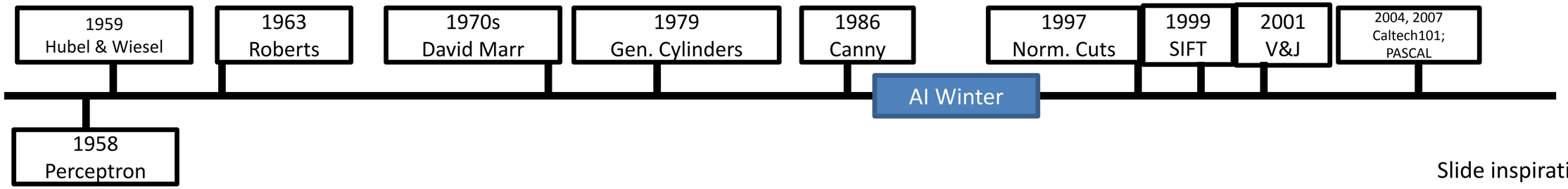
**David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams***

* 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





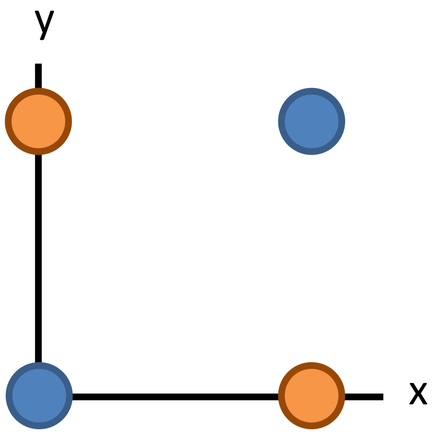
Frank Rosenblatt, ~1957

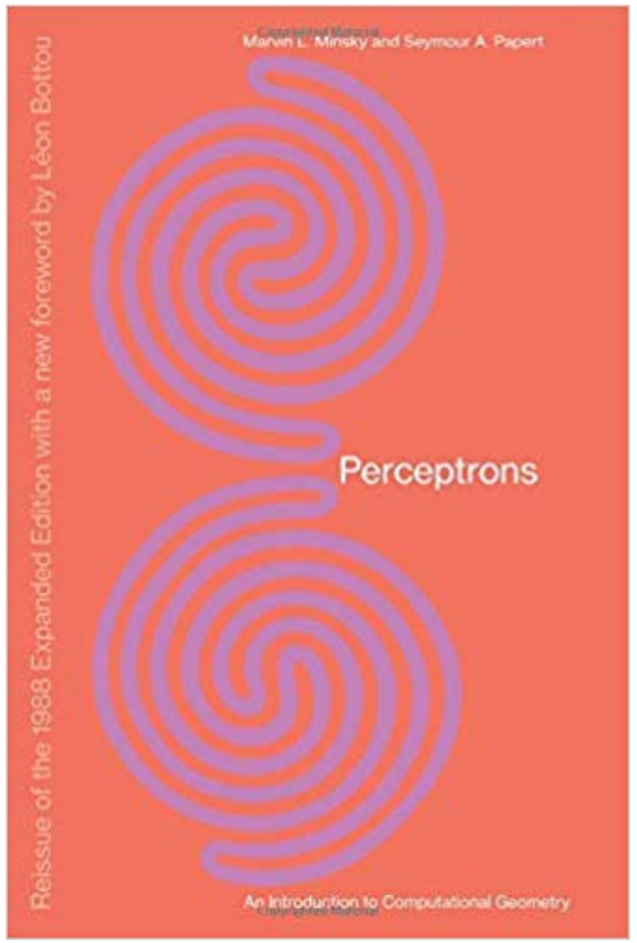


Slide inspiration: Justin Johnson

Minsky and Papert, 1969

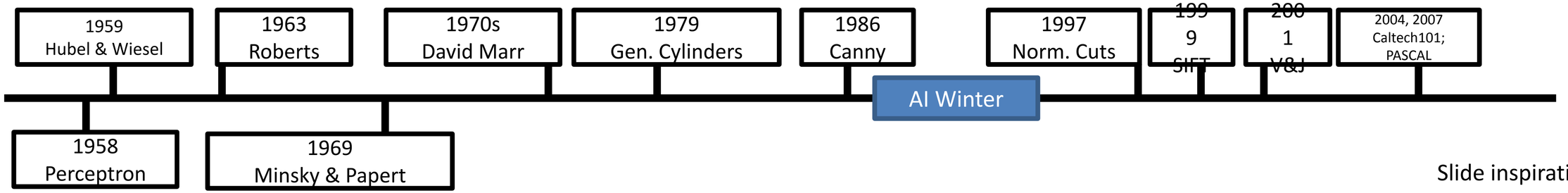
X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0





Showed that Perceptrons could not learn the XOR function

Caused a lot of disillusionment in the field



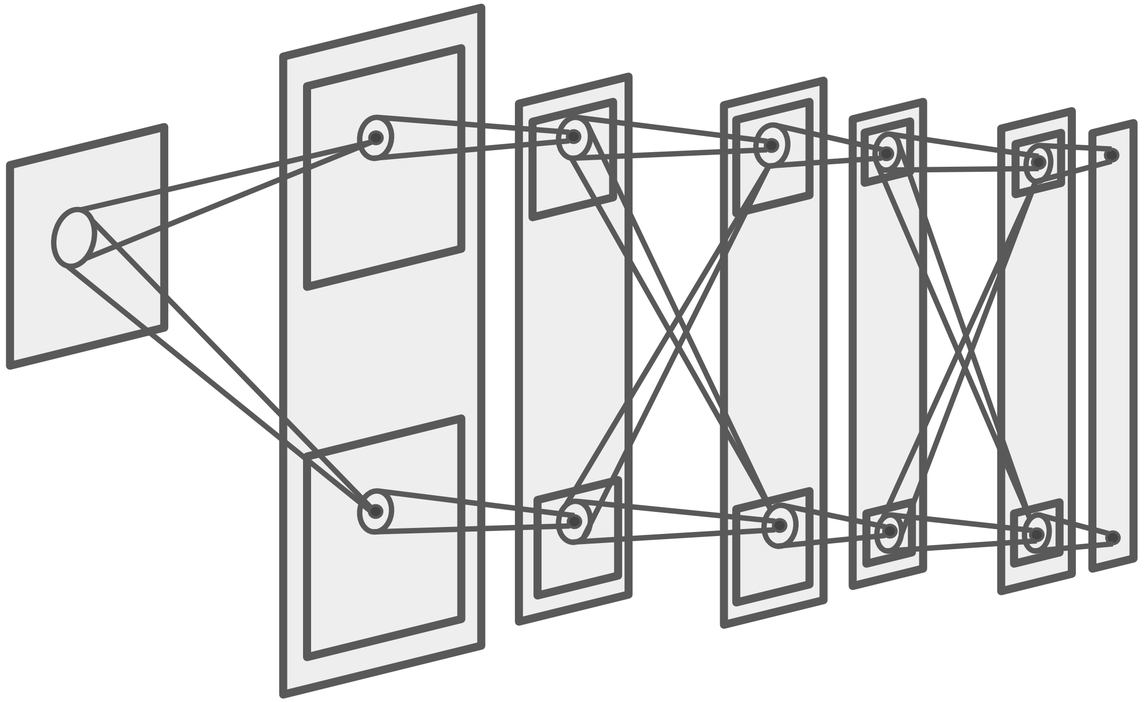
Slide inspiration: Justin Johnson

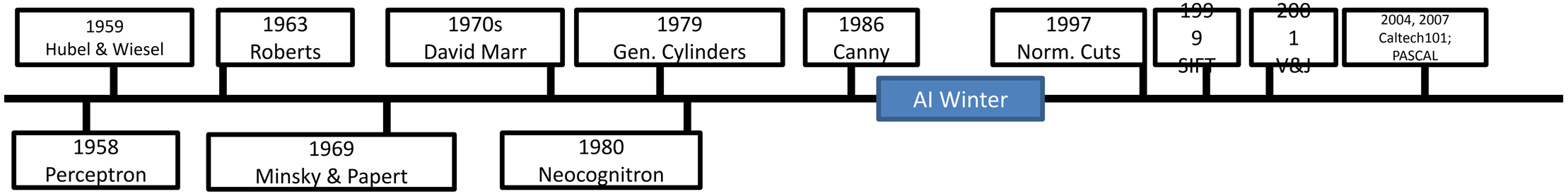
Neocognitron: Fukushima, 1980

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

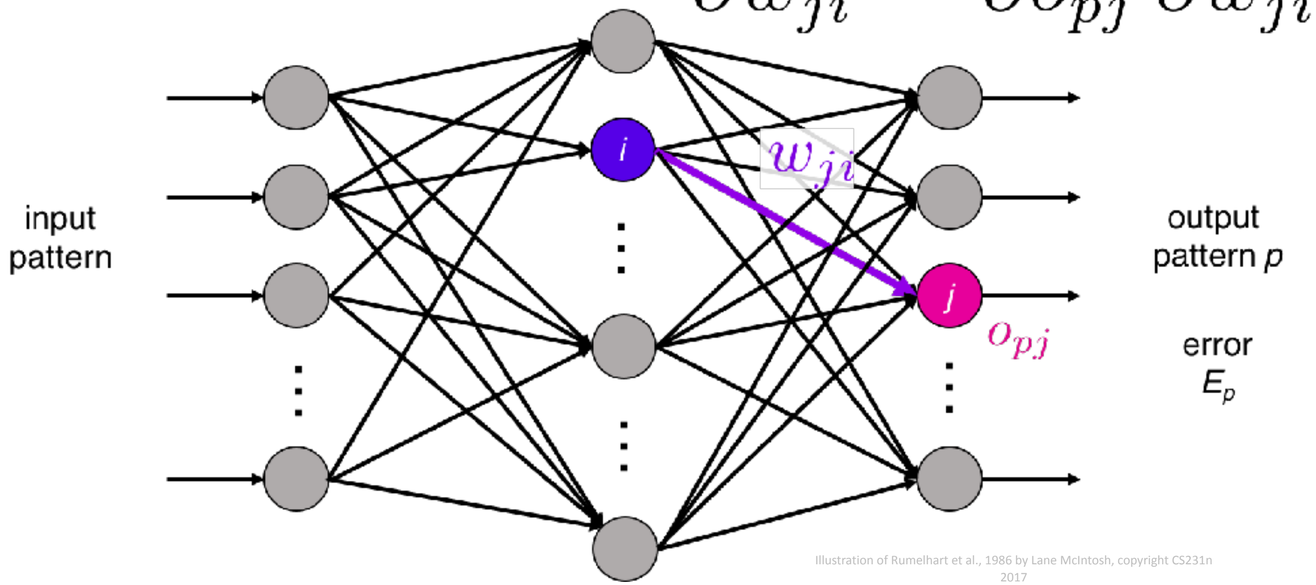
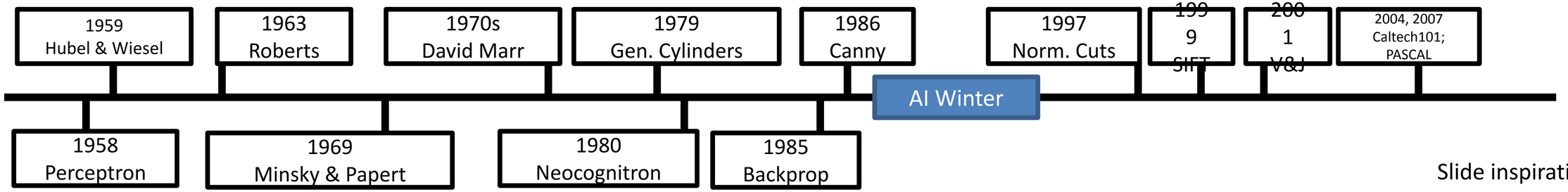
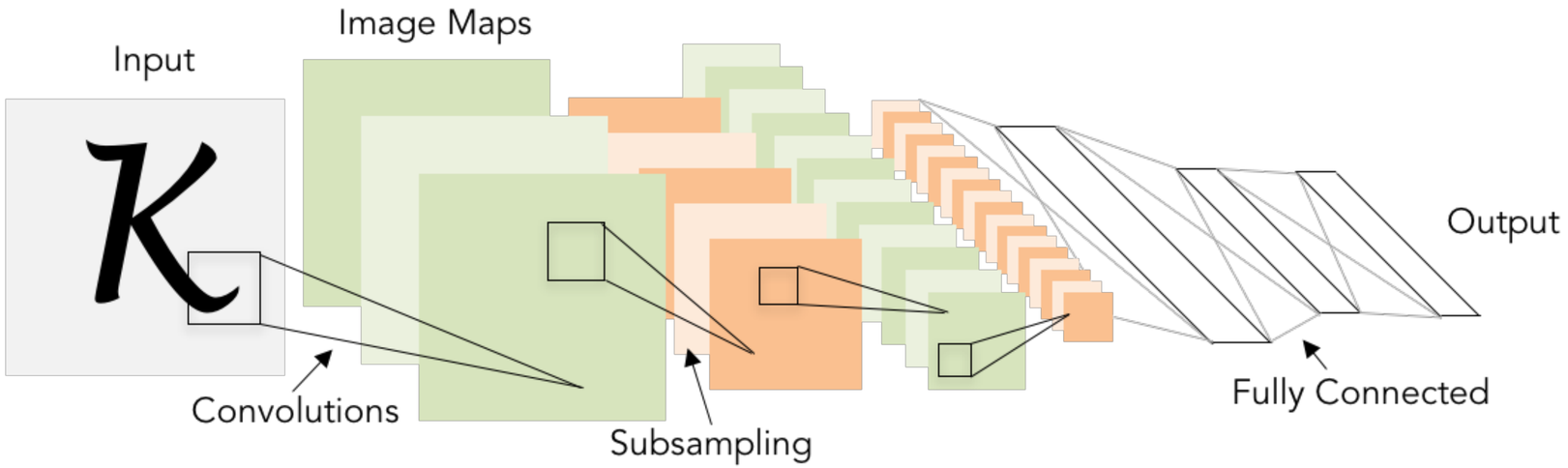


Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017

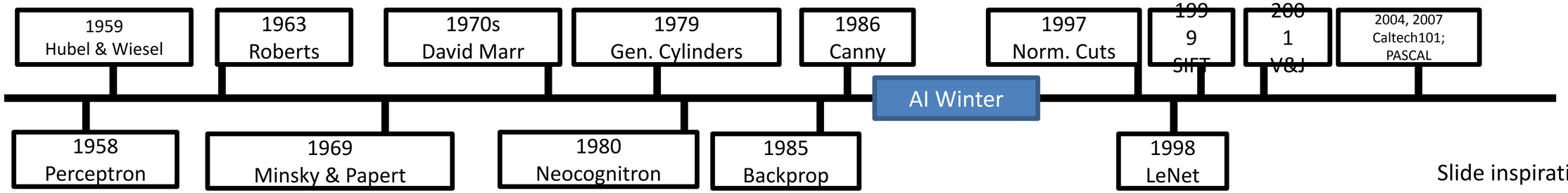


Slide inspiration: Justin Johnson

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!



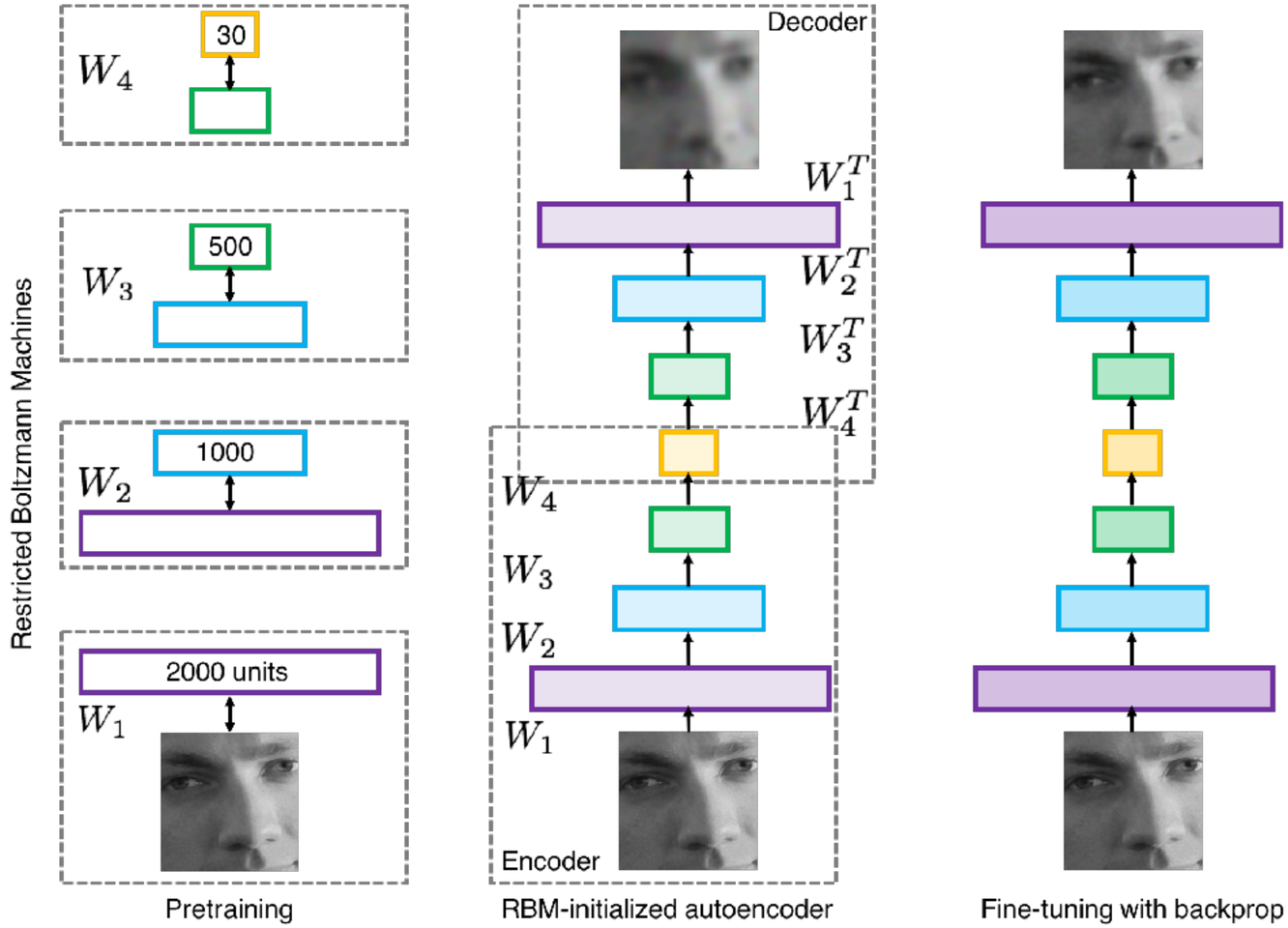
Slide inspiration: Justin Johnson

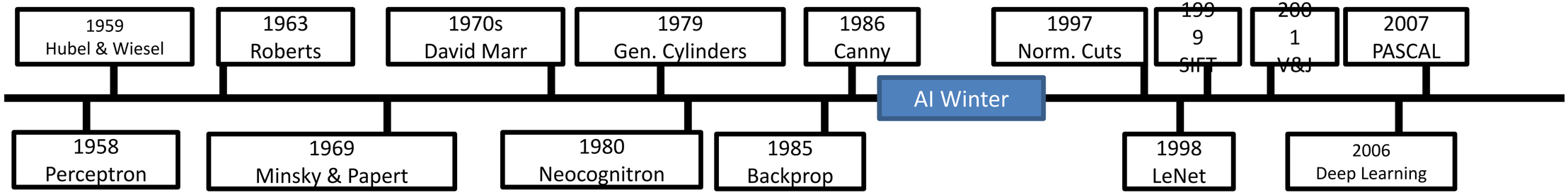
2000s: “Deep Learning”

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

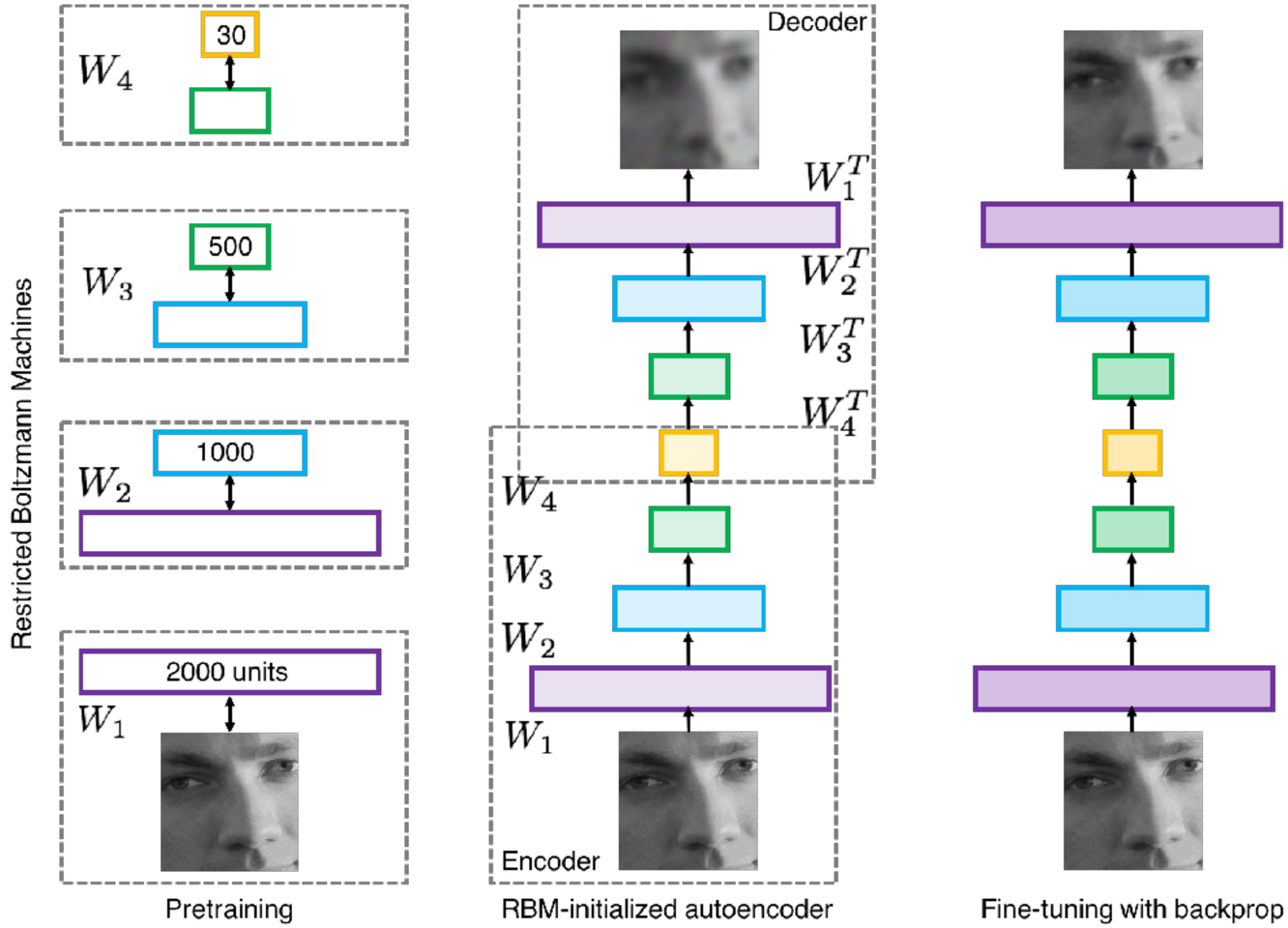
2000s: “Deep Learning”

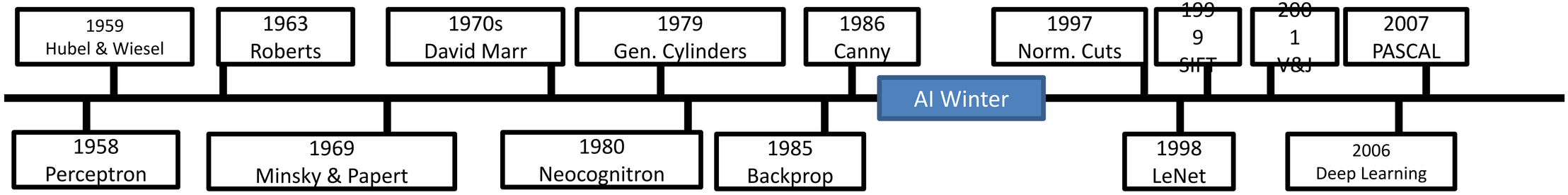
People tried to train neural networks that were deeper and deeper

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

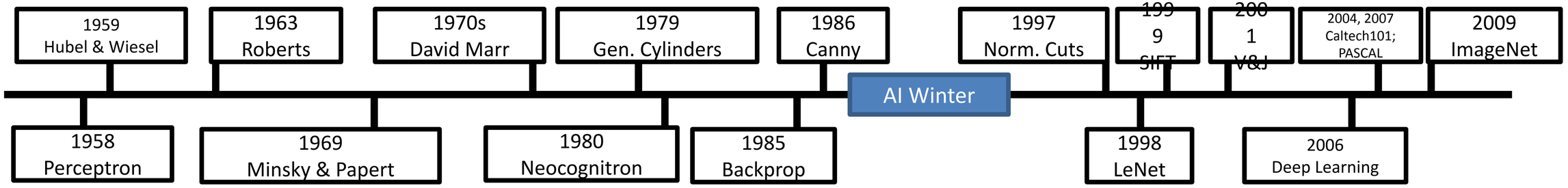
The Image Classification Challenge:

1,000 object classes

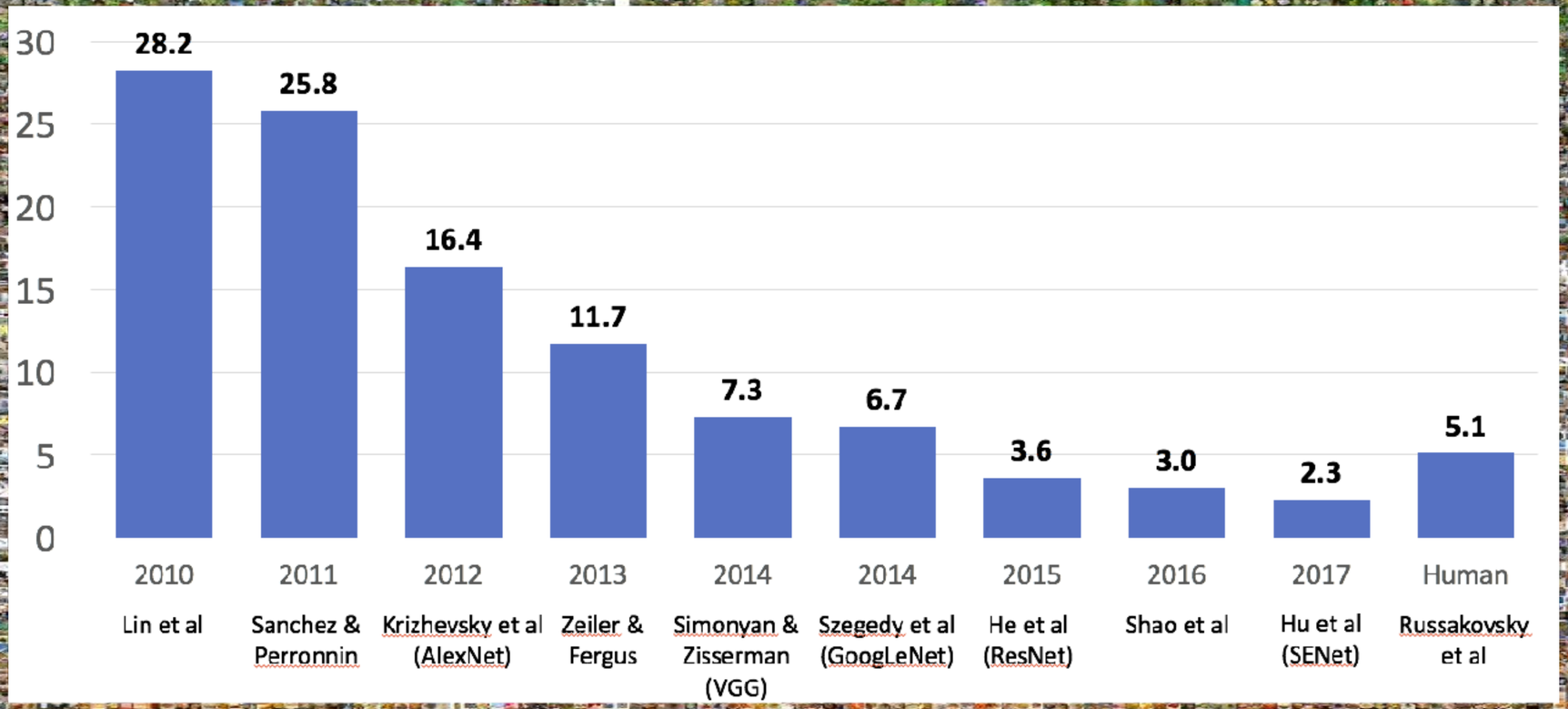
1,431,167 images

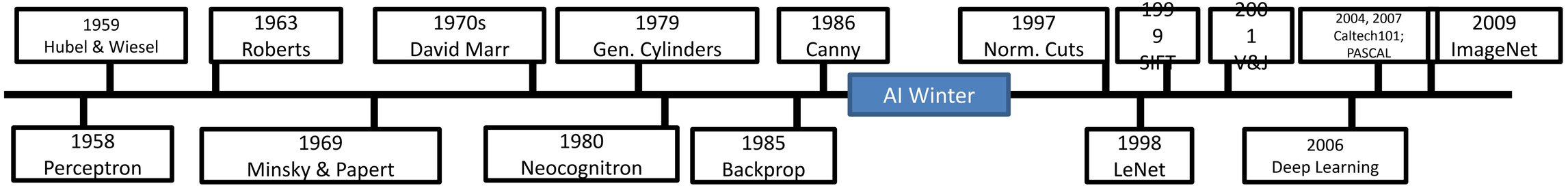
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

Deng et al, 2009
Russakovsky et al. IJCV 2015

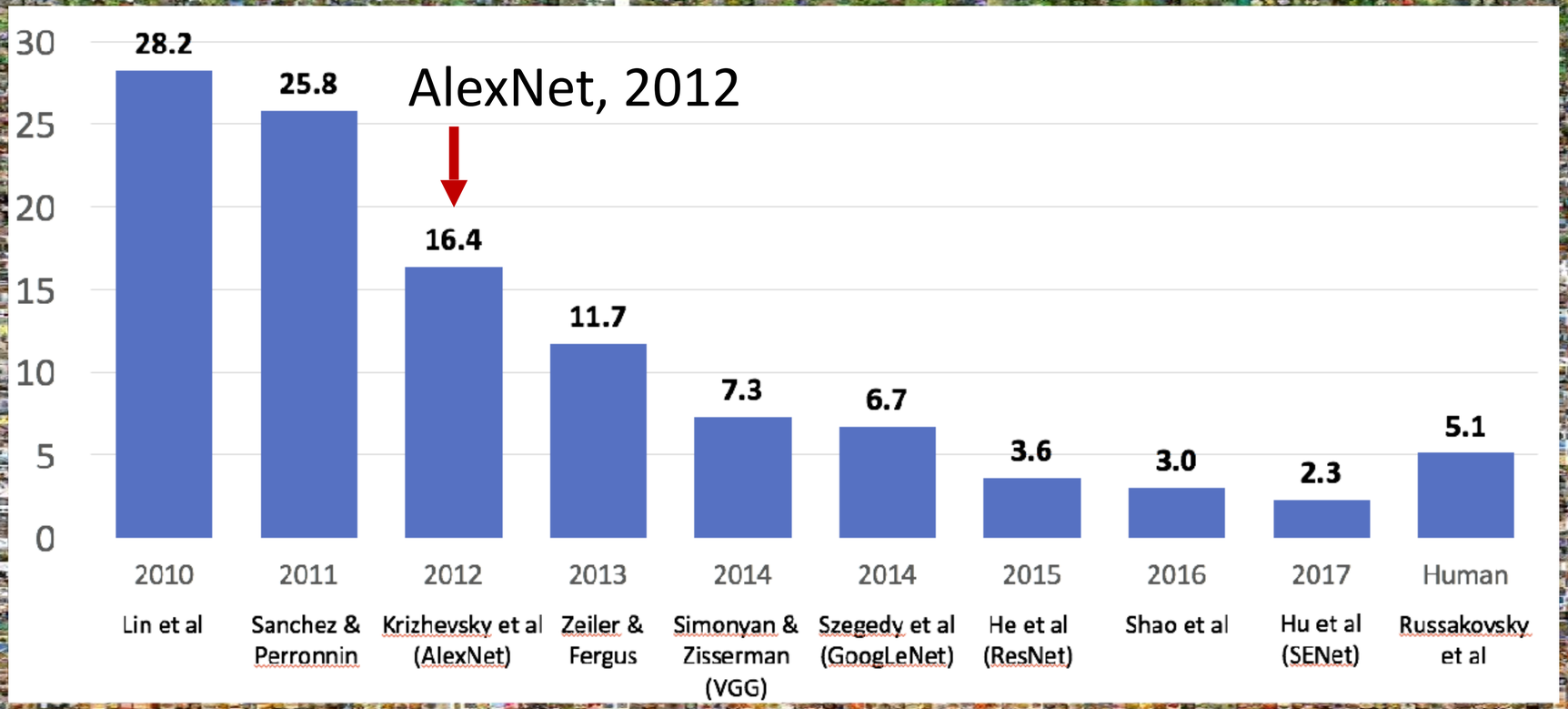


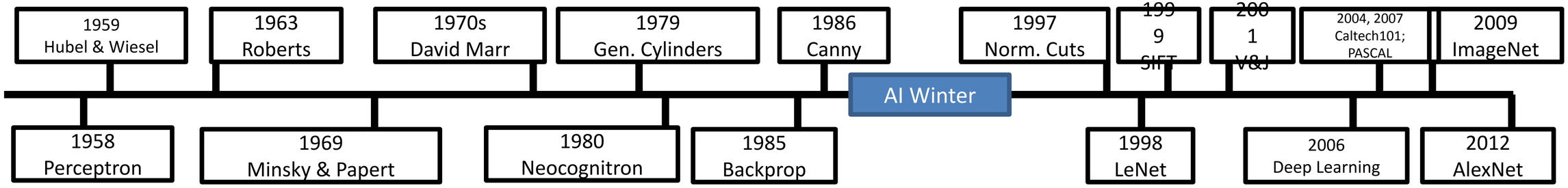
IMAGENET Large Scale Visual Recognition Challenge



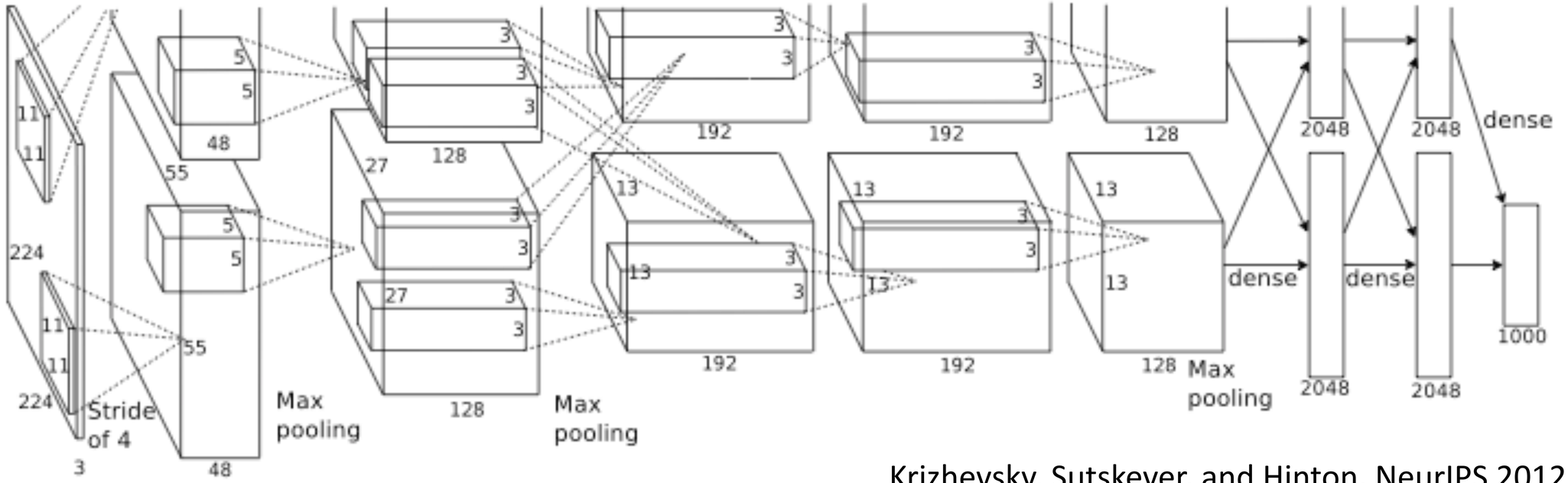


IMAGENET Large Scale Visual Recognition Challenge

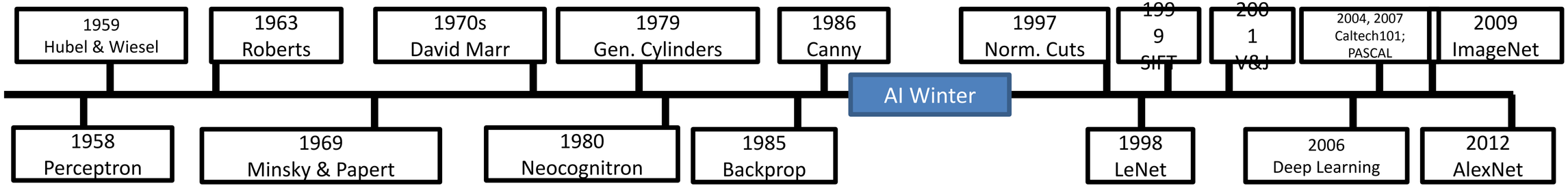




AlexNet: Deep Learning Goes Mainstream

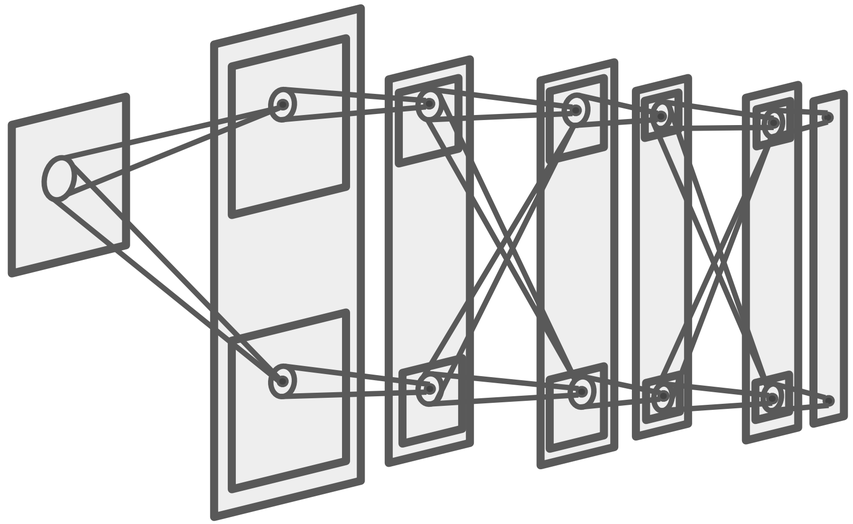


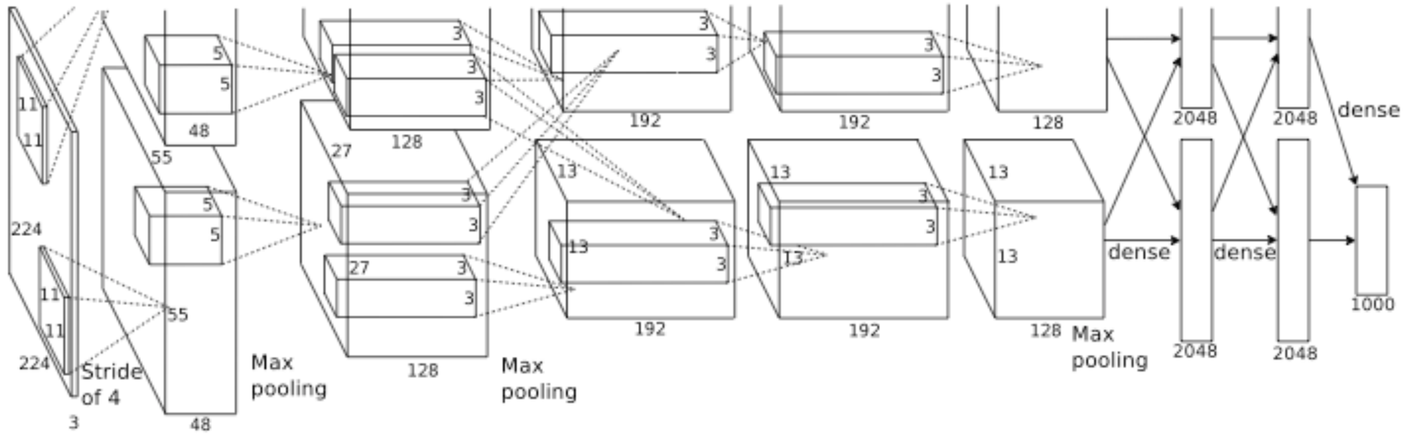
Krizhevsky, Sutskever, and Hinton, NeurIPS 2012

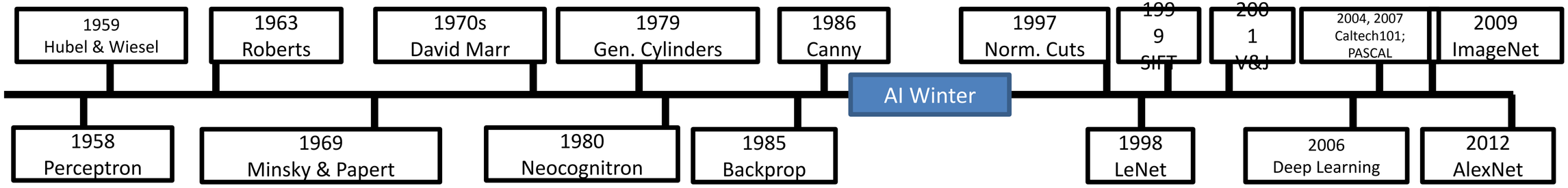


Slide inspiration: Justin Johnson

AlexNet vs. Neocognitron: 32 years apart



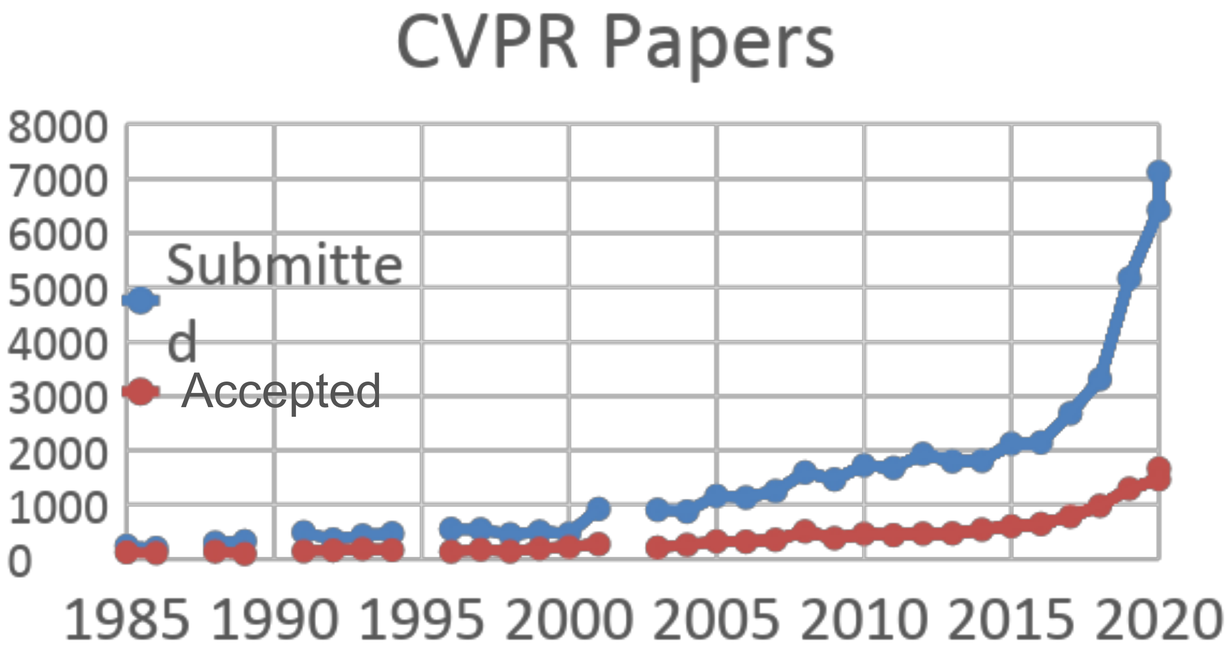




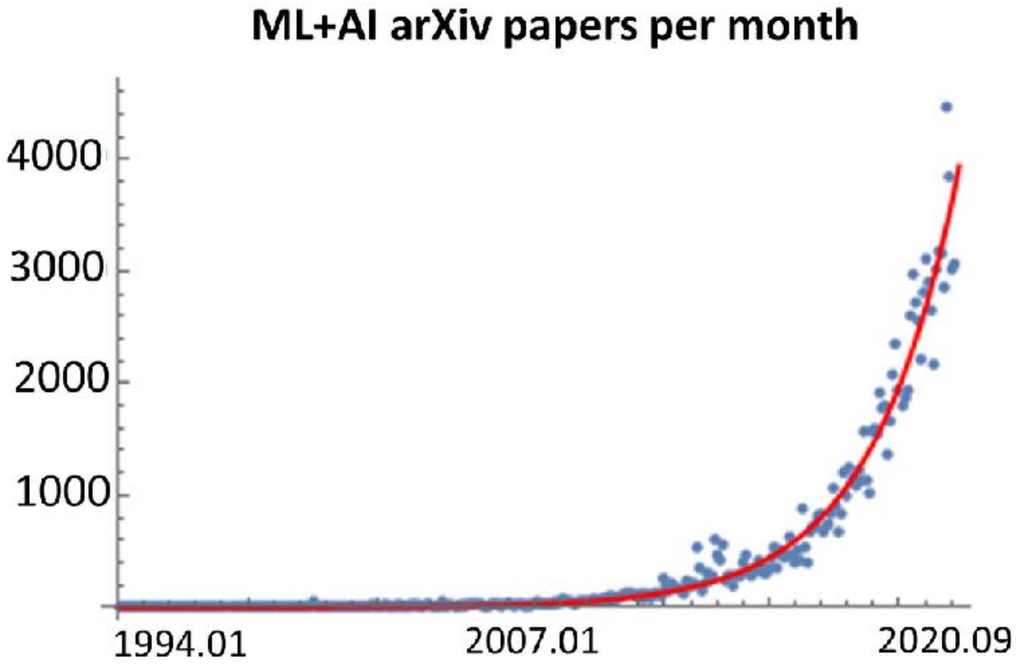
Slide inspiration: Justin Johnson

The AI winter thawed, deep learning revolution arrived

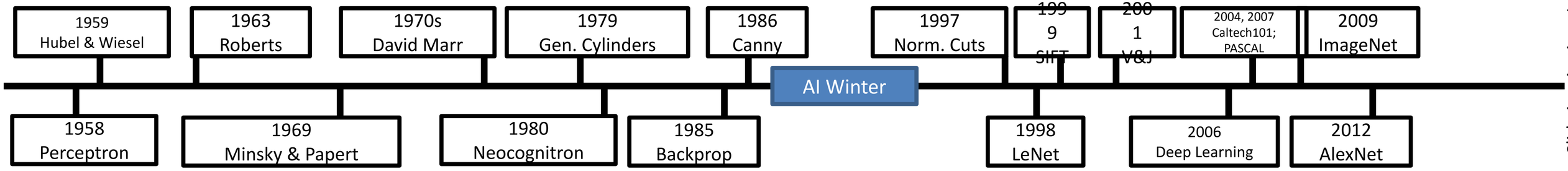
2012 to Present: Deep Learning Explosion



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

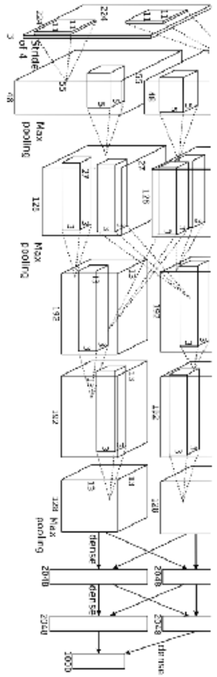
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

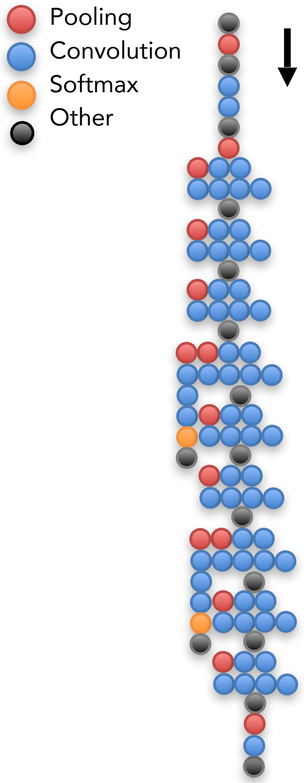
SuperVision



[Krizhevsky NIPS 2012]

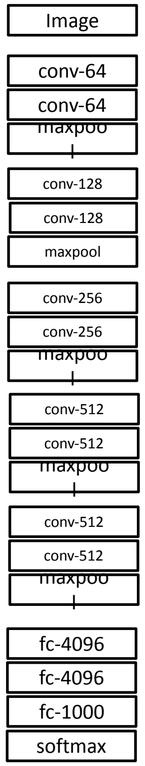
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

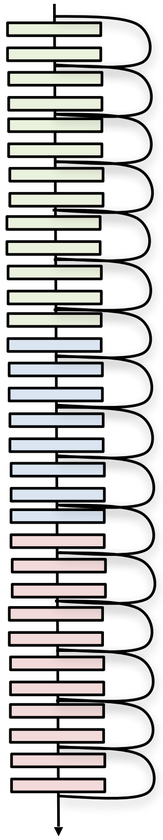
VGG



[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

2012 to Present: Deep Learning is Everywhere

Image Classification

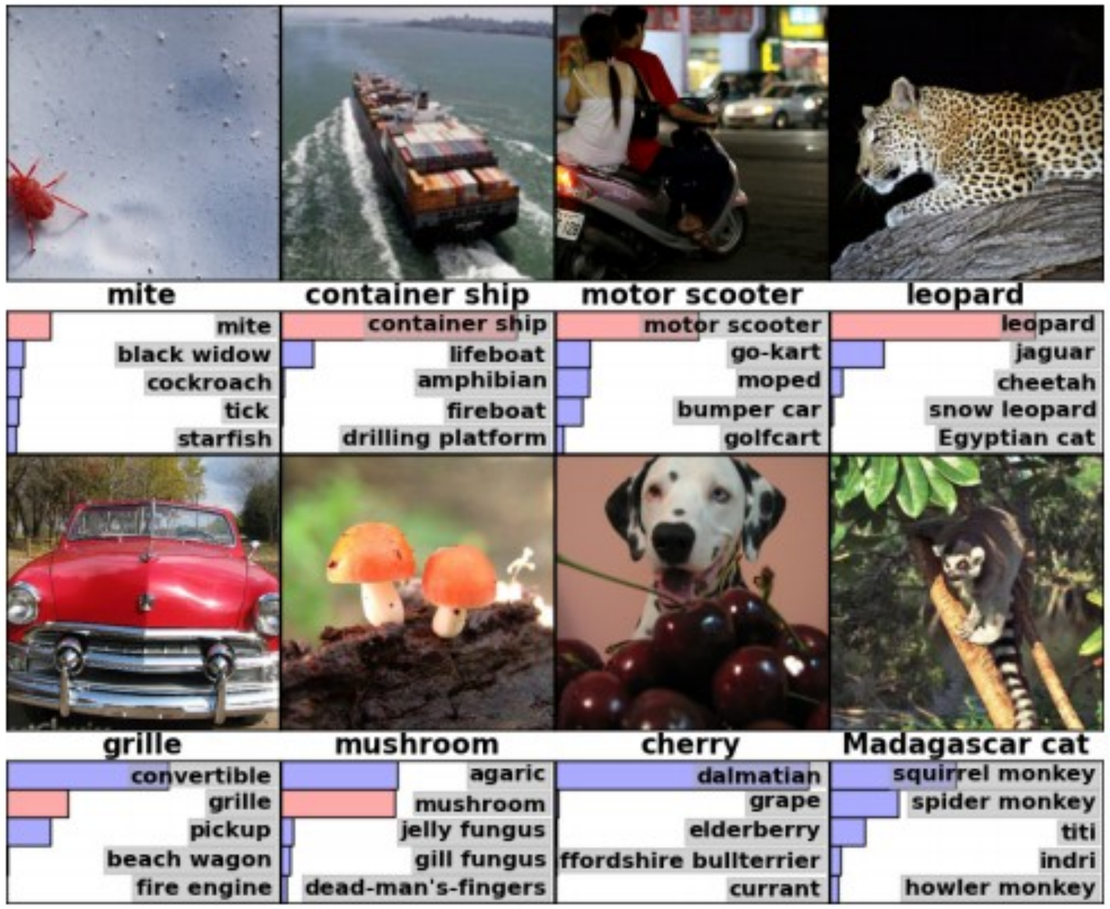


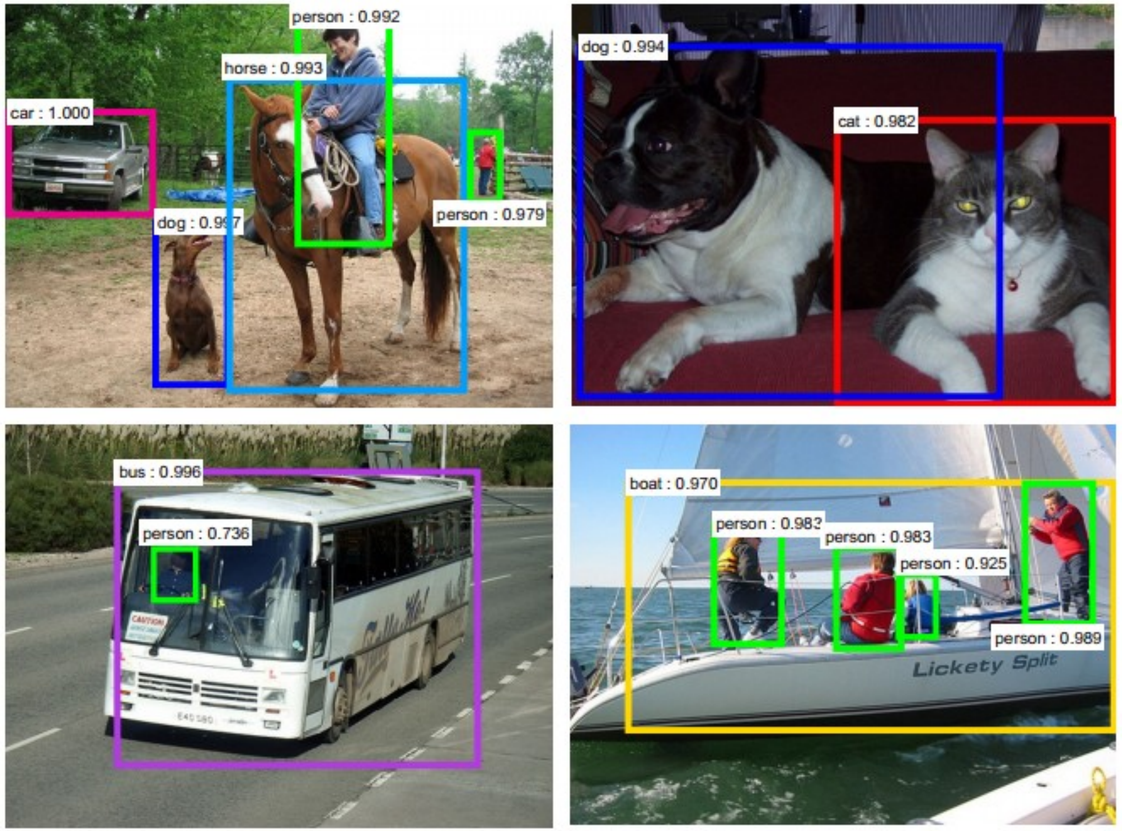
Image Retrieval

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

Slide inspiration: Justin Johnson

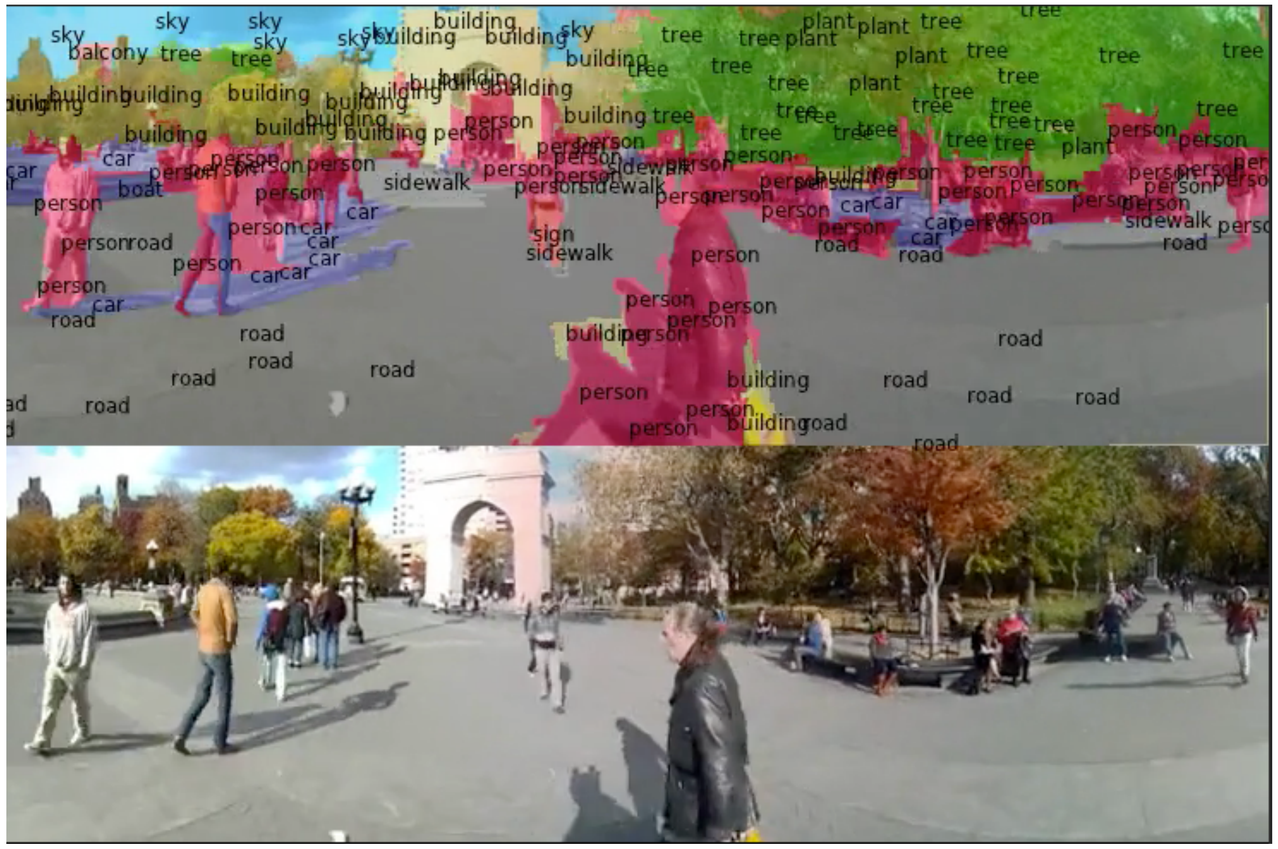
2012 to Present: Deep Learning is Everywhere

Object Detection



Ren, He, Girshick, and Sun, 2015

Image Segmentation

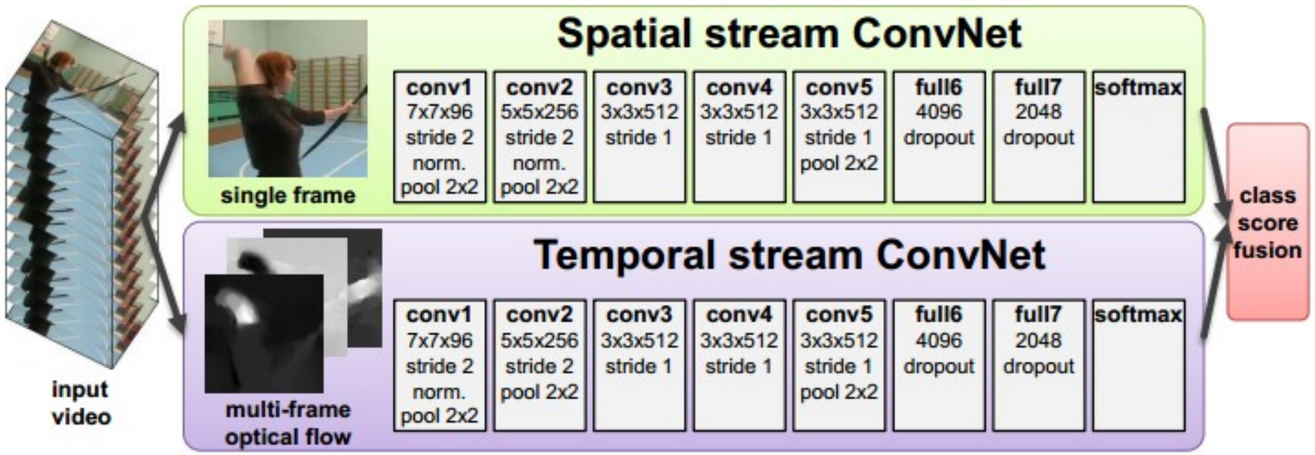


Fabaret et al, 2012

Slide inspiration: Justin Johnson

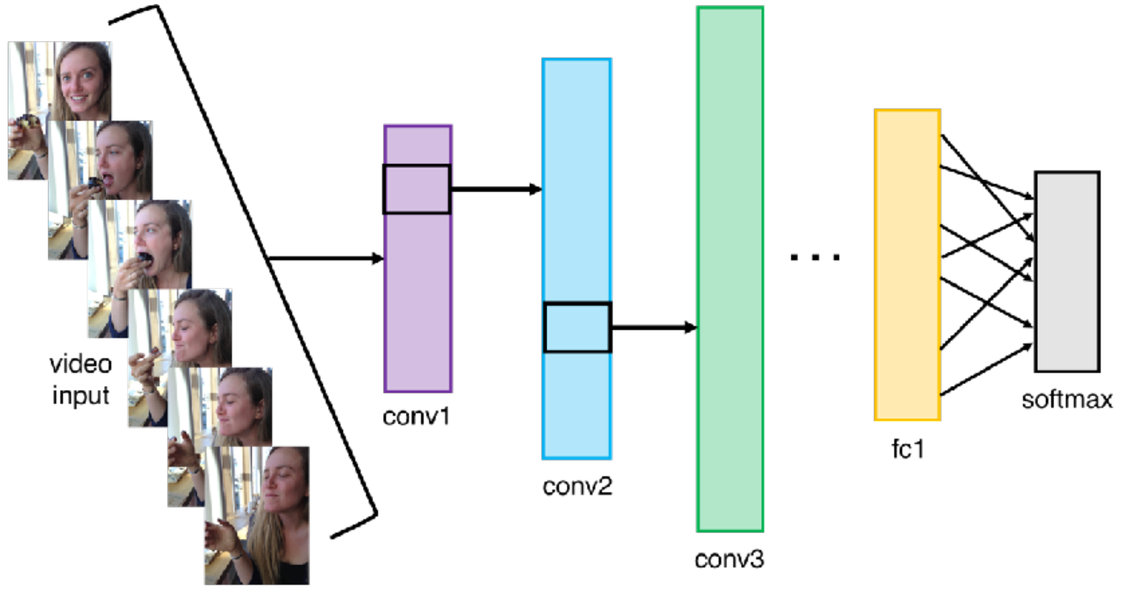
2012 to Present: Deep Learning is Everywhere

Video Classification



Simonyan et al, 2014

Activity Recognition

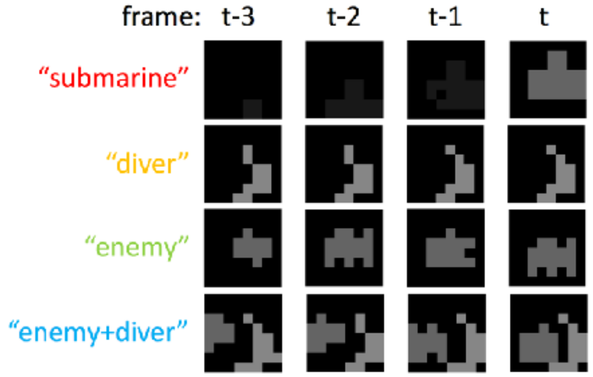


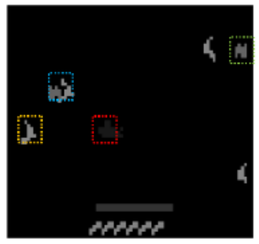
Slide inspiration: Justin Johnson

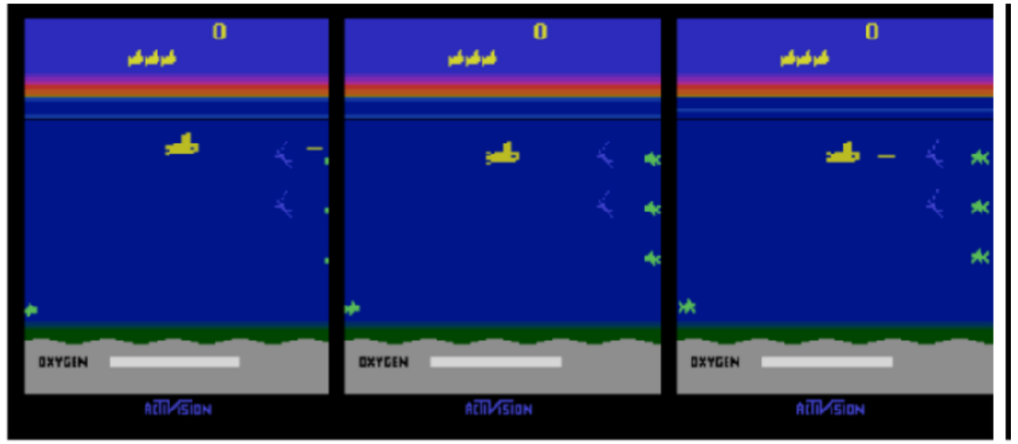
2012 to Present: Deep Learning is Everywhere

Pose Recognition (Toshev and Szegedy, 2014)

Playing Atari games (Guo et al, 2014)



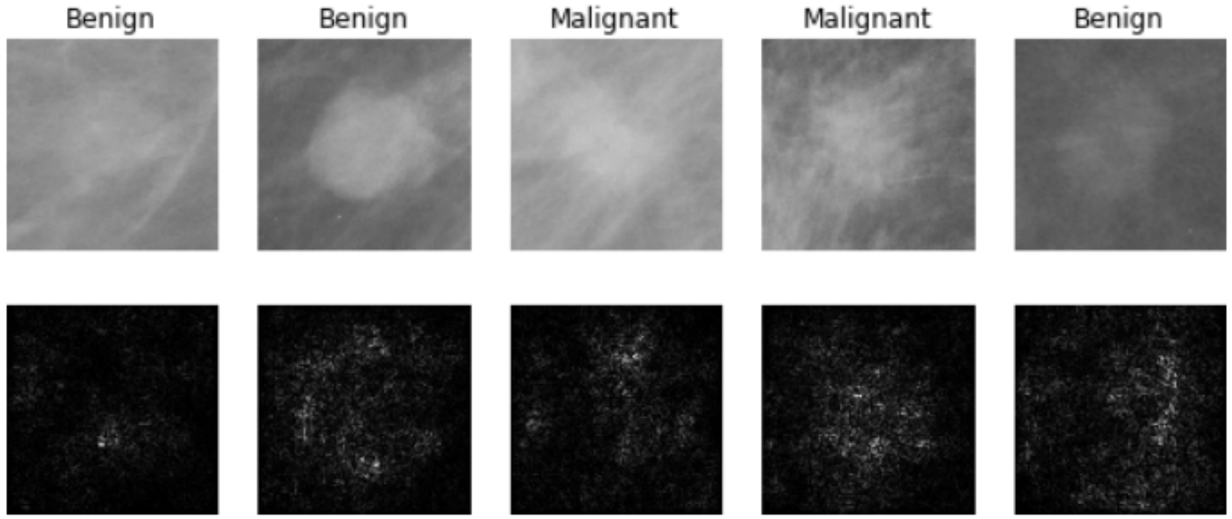




Slide inspiration: Justin Johnson

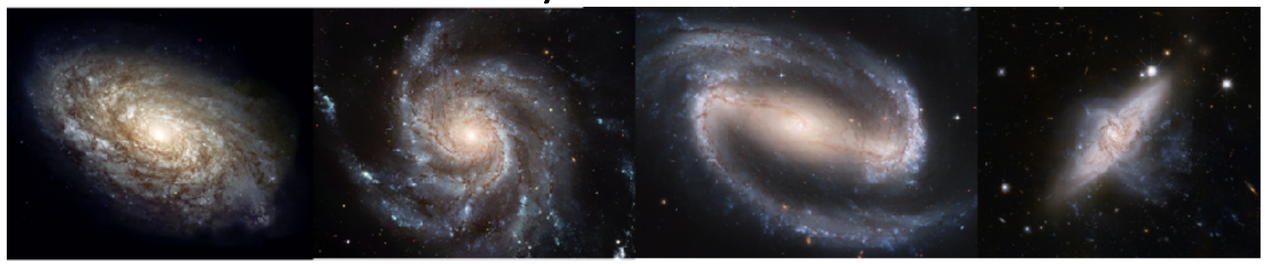
2012 to Present: Deep Learning is Everywhere

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.

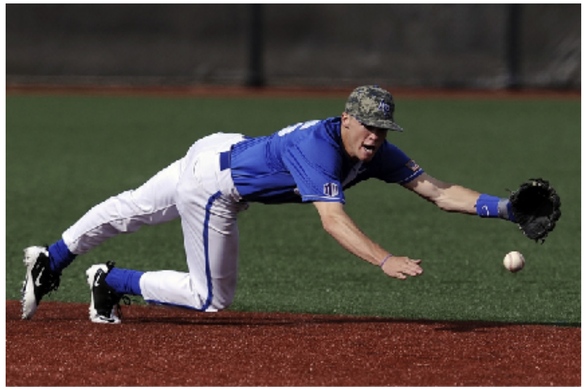
Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

Image Captioning

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

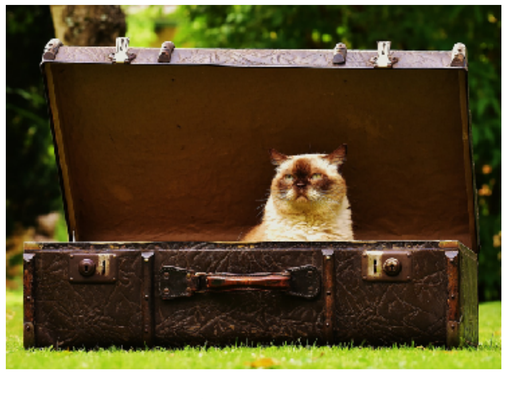
A white teddy bear sitting in the grass



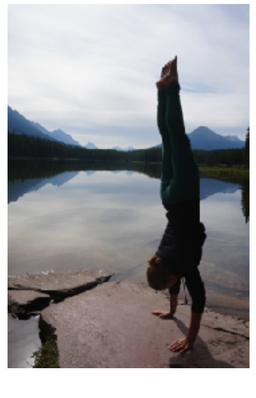
A man in a baseball uniform throwing a ball

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

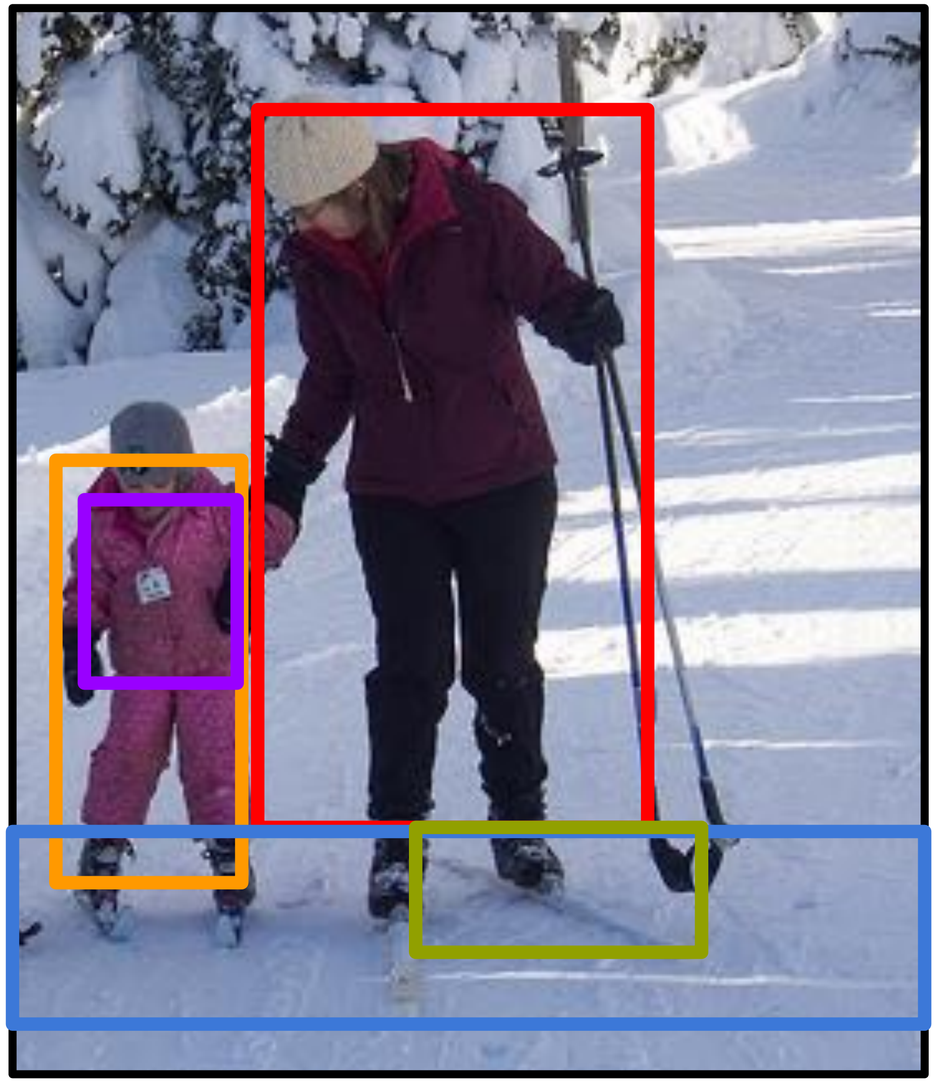


A woman standing on a beach holding a surfboard

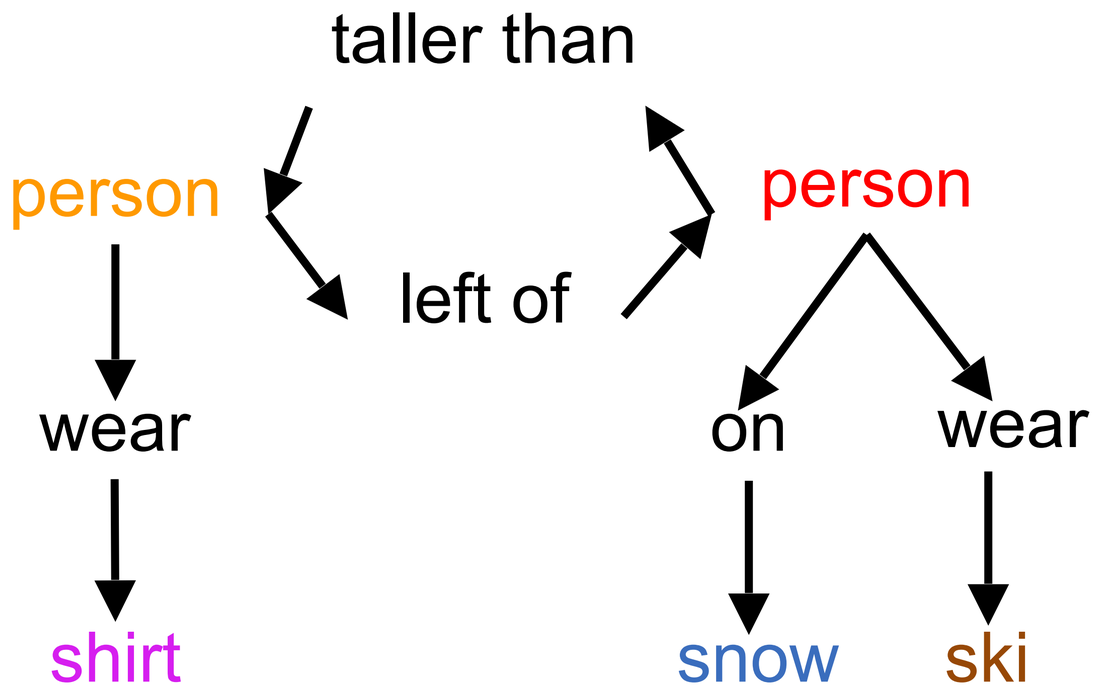
Captions generated by Justin Johnson using Neuraltalk2

Slide inspiration: Justin Johnson

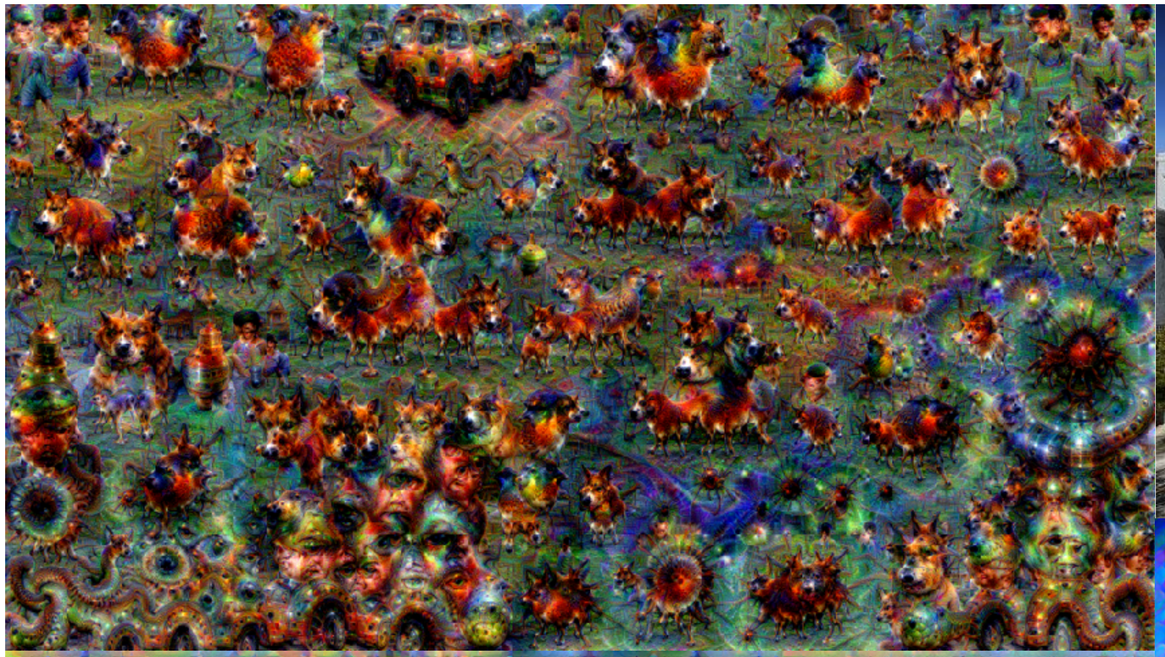
2012 to Present: Deep Learning is Everywhere

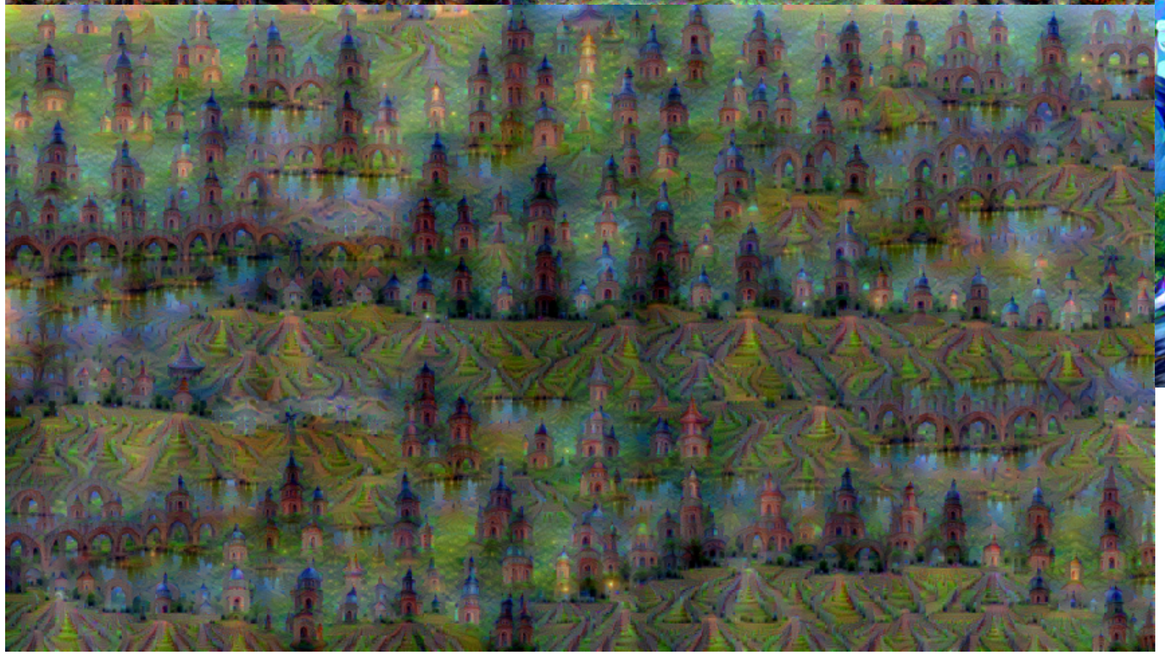


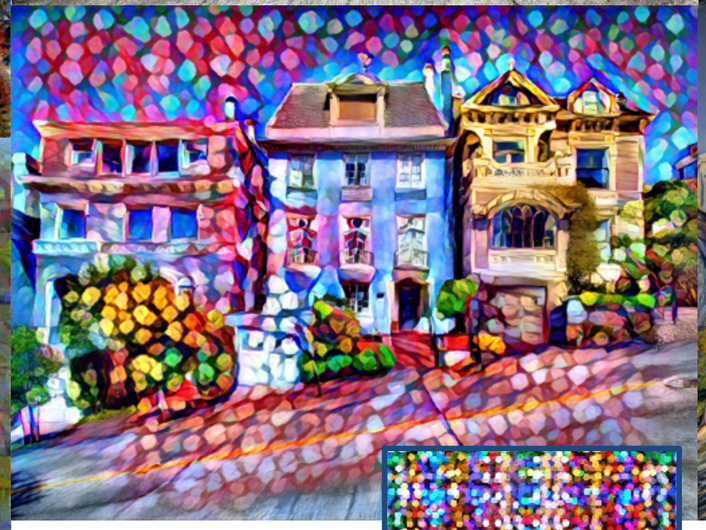
Results:
spatial, comparative, asymmetrical, verb, prepositional



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







Figures copyright Justin Johnson, 2015. Reproduced with permission. Generated using the Inceptionism approach from a **blog post** by Google Research.

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

A Dutch still life of an arrangement of tulips in a fluted vase. The lighting is subtle, casting gentle highlights on the flowers and emphasizing their delicate details and natural beauty.

Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

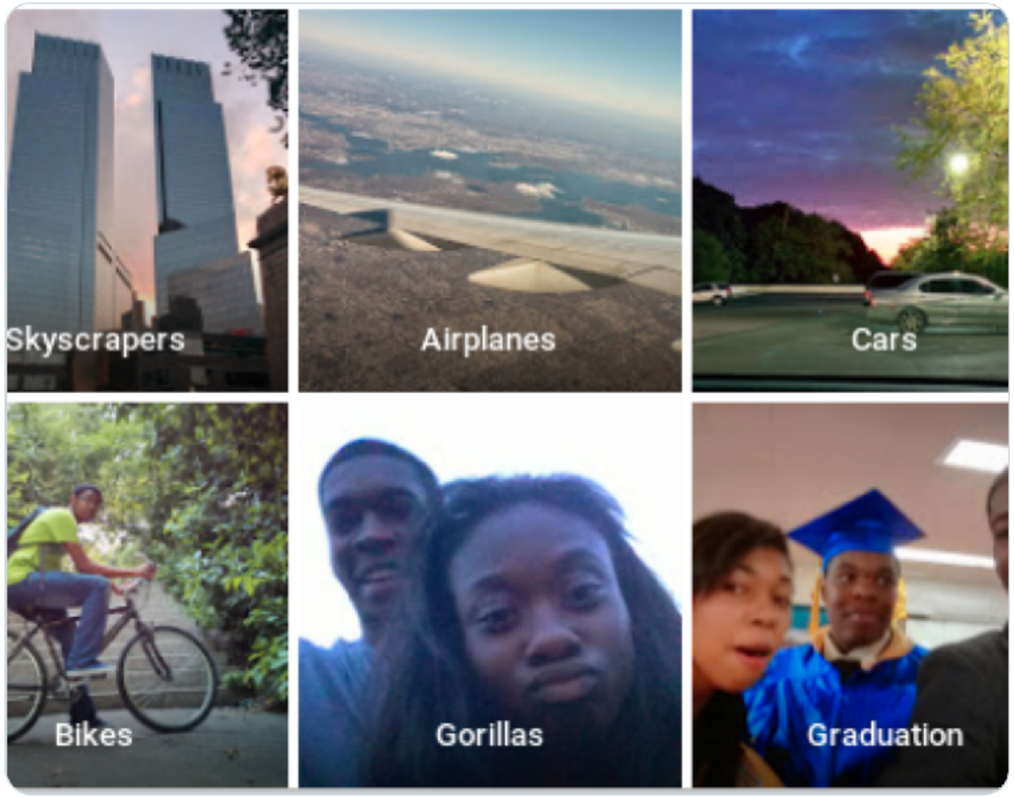
A 2D animation of a folk music band composed of anthropomorphic autumn leaves, each playing traditional bluegrass instruments, amidst a rustic forest setting dappled with the soft light of a harvest moon.

Slide inspiration: Justin Johnson

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

Computer Vision Can Cause Harm

Harmful Stereotypes



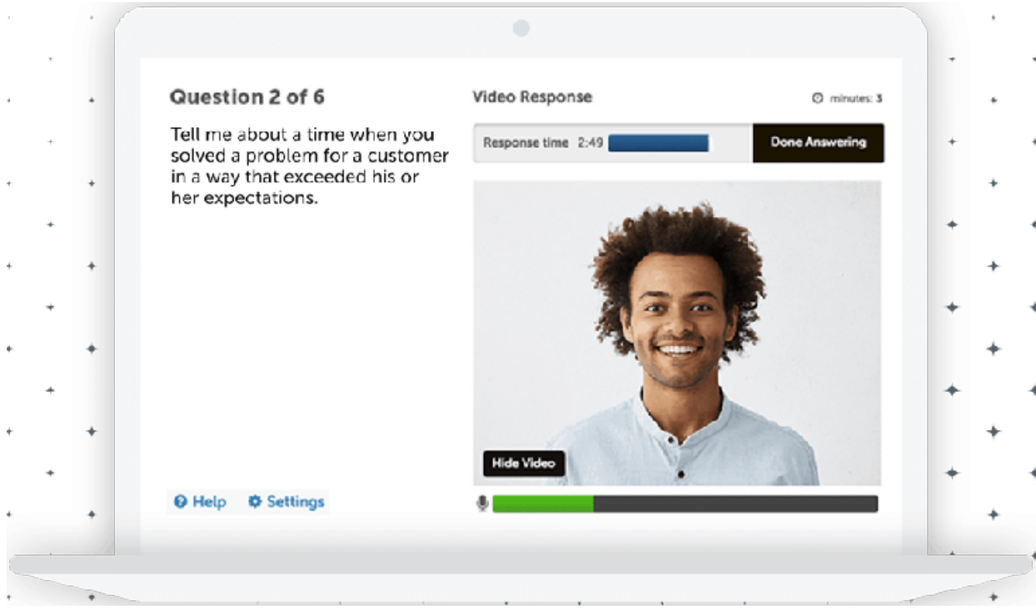
Barocas et al, “The Problem With Bias: Allocative Versus Representational Harms in Machine Learning”, SIGCIS 2017
Kate Crawford, “The Trouble with Bias”, NeurIPS 2017 Keynote
Source: https://twitter.com/jackyalcine/status/615329515909156865 (2015)

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

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

Further thoughts and readings

Chapter 14, Szeliski's book

Think of the applications of computer vision around you

Some slides kindly provided by Fei-Fei Li, Jiajun Wu, Erik Learned-Miller