



CS231n: Deep Learning for Computer Vision

Lecture 1: Introduction

Welcome to CS231n



Welcome to CS231n

Course Instructors



Fei-Fei Li Andrej Karpathy

Teaching Assistants



2015

Course Instructors



Fei-Fei Li Andrej Karpathy

Teaching Assistants



Sumeet Singh Sudheendra Dua Bhargav Reddy Albert Huang Namrata Anand Lane Morrison Catherine Dong

2016

Instructors



Fei-Fei Li Andrej Karpathy

Teaching Assistants



Johnson Harshit Bhatia Iwan Belli Namrata Anand Lane Morrison Catherine Dong

2018



Fei-Fei Li Andrej Karpathy



Albert Haque (Head TA) Chaitanya Asawa Josh Beal Vincent Chen Edward Chou John Clow



Justin Johnson Manik Dhar Jim (Linx) Fan Alexander (Kaiyi) Fu Pedro Pablo Garzon Michelle Guo Jingwei Huang



Serena Yeung Nishith Khandwala Carolyn Kim Winnie Lin Bingbin Liu Xingyu Liu Ajay Mandekar

Amani Peddada Mike Roberts Praty Sharma Fei Xia Danfei Xu Ben Zhang

Instructors



Fei-Fei Li



Justin Johnson



Serena Yeung

Teaching Assistants



Albert Haque (Head TA)



Rohit Bedi



Shyamal Buch



Zhao (Joe) Chen



Timnit Gebru

Instructors



Fei-Fei Li



Justin Johnson



Agrim Gupta

Teaching Assistants



De-An Huang



Russell Kaplan



Leo Keselman



Nishith Khandwala

Instructors



Fei-Fei Li



Ranjay Krishna

Teaching Assistants



William Shen (Head TA)



Jonathan Bratza

Instructors



Fei-Fei Li



Andrew Kondrich

Teaching Assistants



Fang-Yu Lin



Damian Mrowca



Boxiao Pan

2020

Nishant Rai

Instructors



Fei-Fei Li



Danfei Xu

Teaching Assistants



Ranjay Krishna



Andrew Kondrich



Fang-Yu Lin



Damian Mrowca



Boxiao Pan



Nishant Rai

Instructors



Fei-Fei Li



Amelie Byun

Teaching Assistants



Brent Yi



Christina Yuan



Kevin Zakra



Yiheng Zhang

2021

2021

Instructors



Fei-Fei Li



Justin Johnson



Serena Yeung

Teaching Assistants



Winnie Lin (Head TA)



Saahil Agrawal



Malavika Bindhi



Hayre Cai

Instructors



Fei-Fei Li



Danfei Xu

Teaching Assistants



Ranjay Krishna



Sean Liu



Mandy Lu



Nishant Rai



Geet Sethi



Lin Shao

Artificial Intelligence

Slide inspiration: Justin Johnson

Artificial Intelligence

Machine Learning

Computer
Vision

Slide inspiration: Justin Johnson

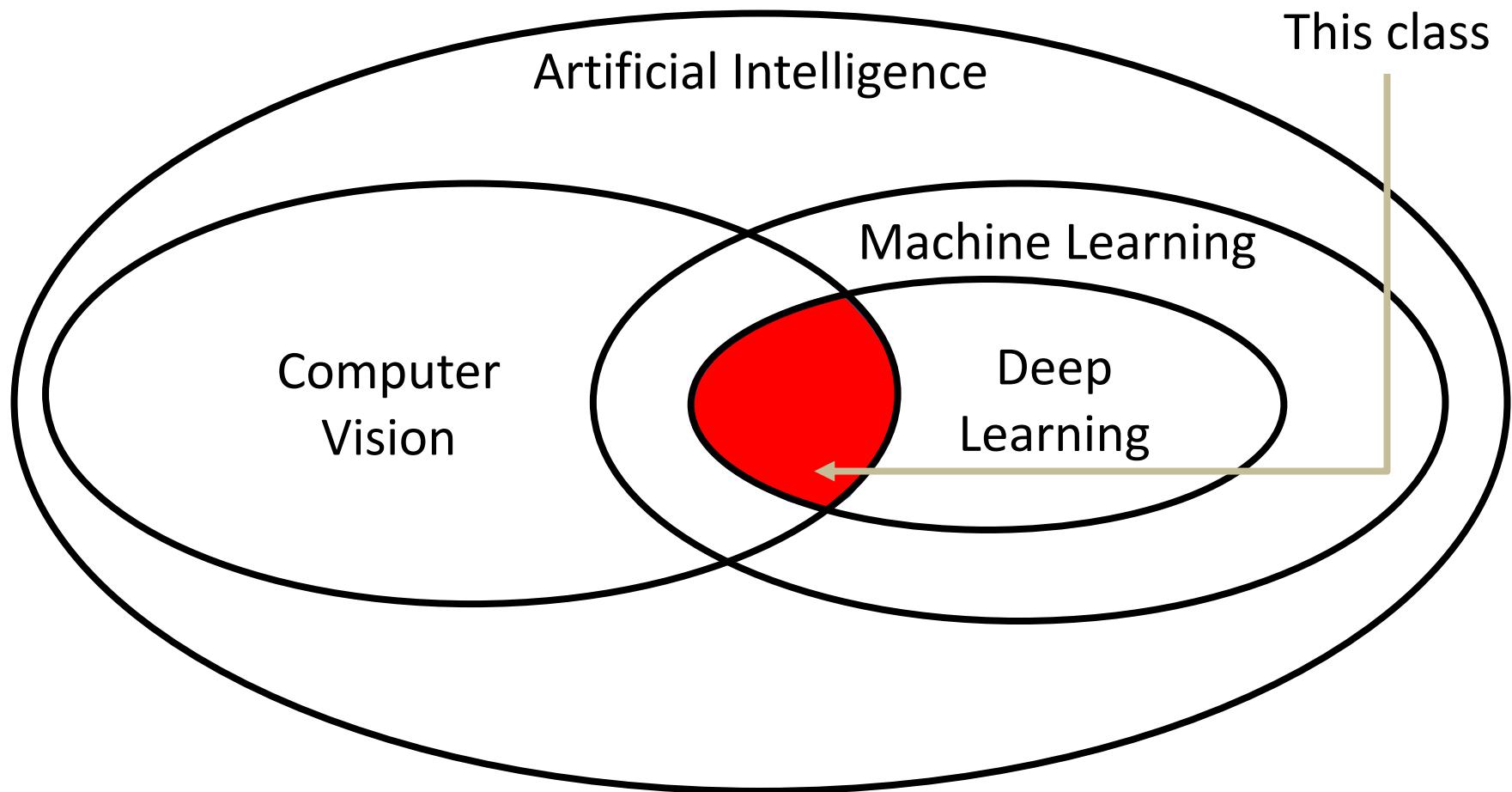
Artificial Intelligence

Machine Learning

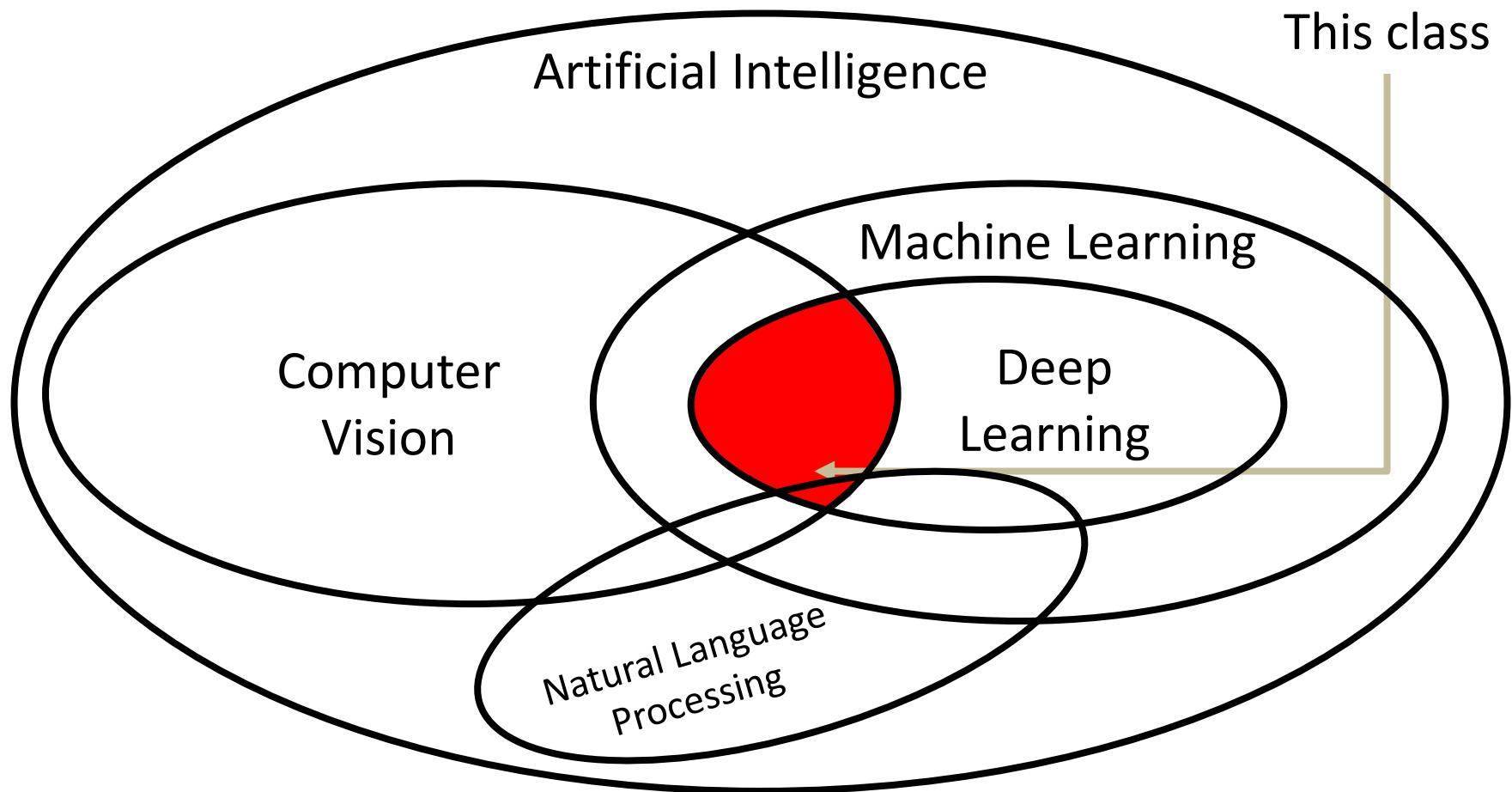
Computer
Vision

Deep
Learning

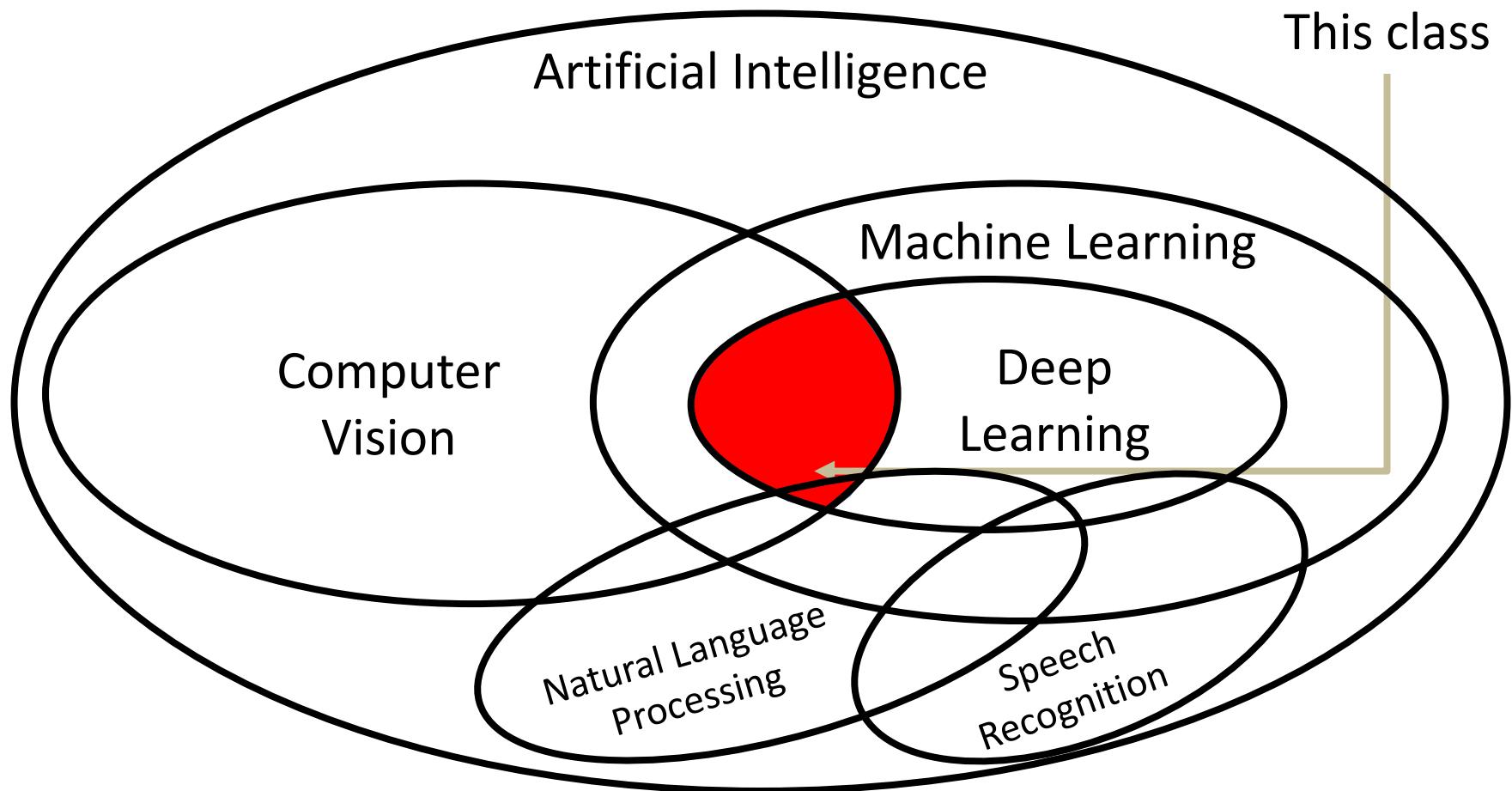
Slide inspiration: Justin Johnson



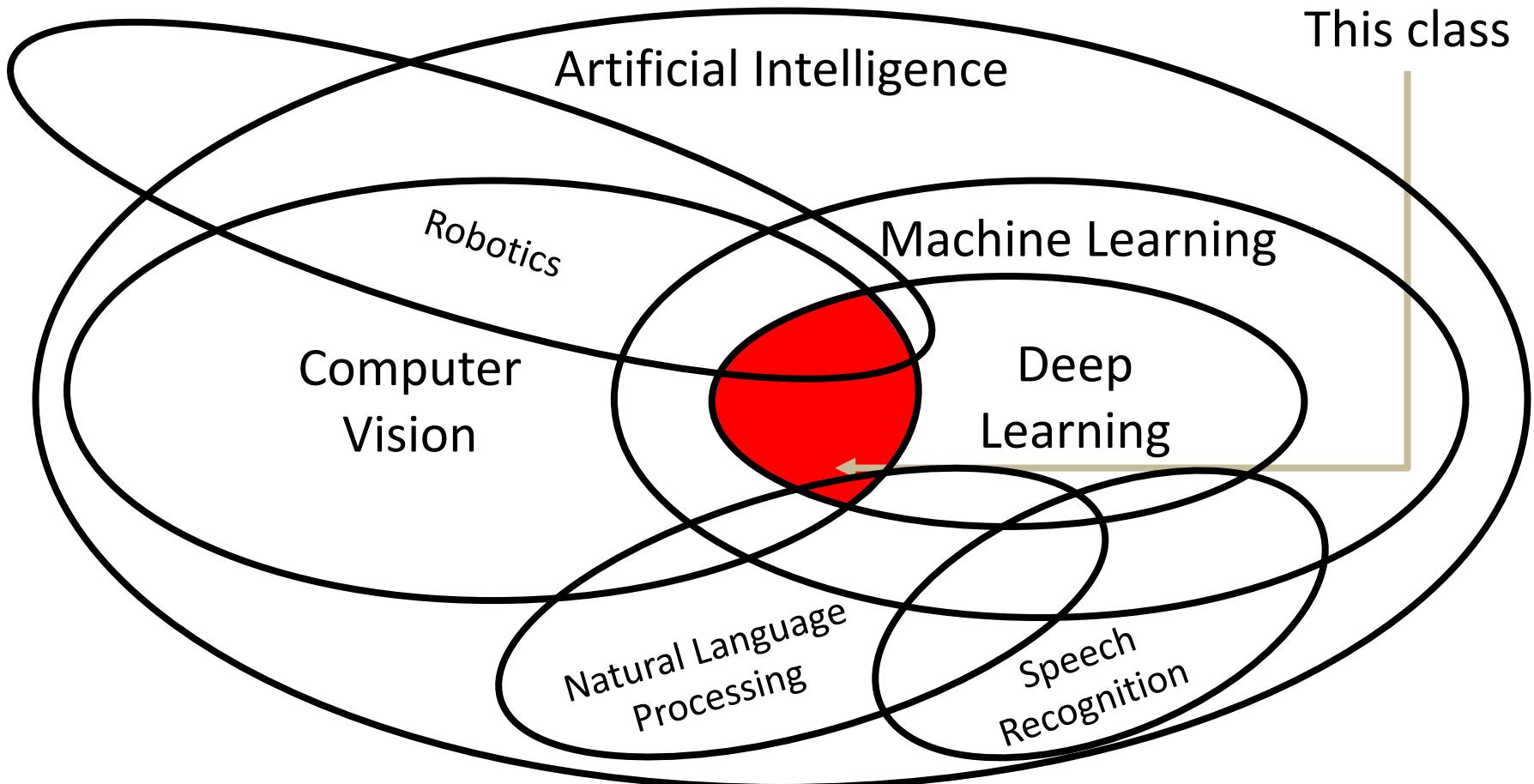
Slide inspiration: Justin Johnson



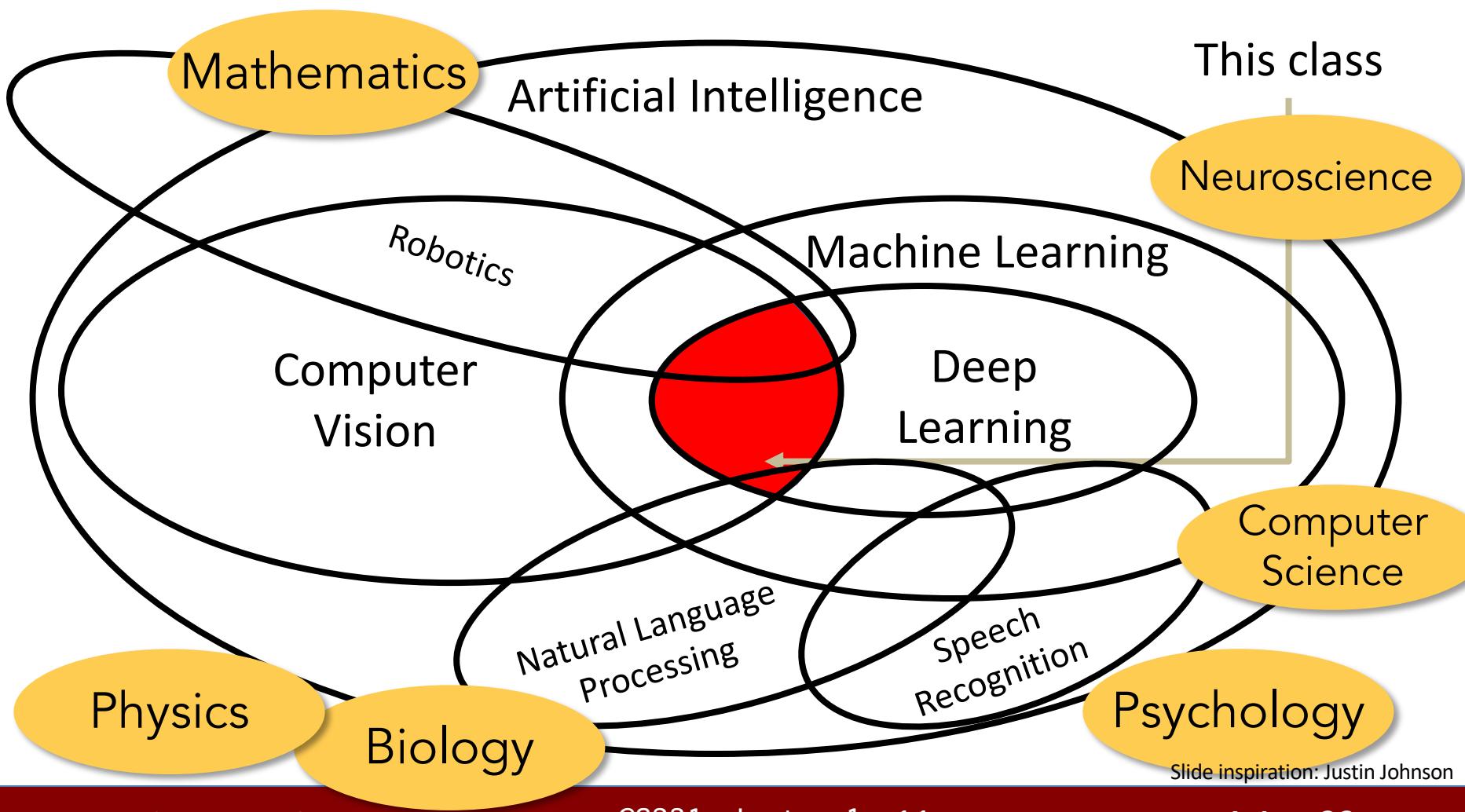
Slide inspiration: Justin Johnson



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.



[This image is licensed under CC-BY 2.5](#)



[This image is licensed under CC-BY 2.5](#)

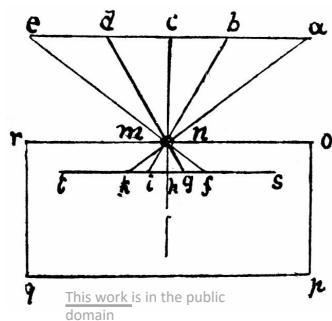
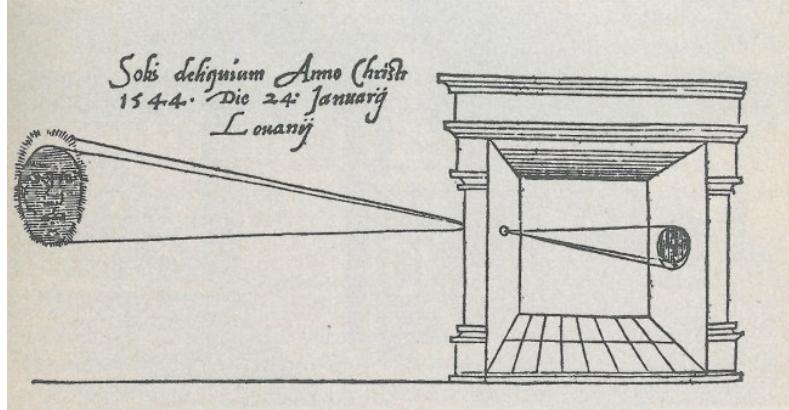


[This image is licensed under CC-BY 3.0](#)



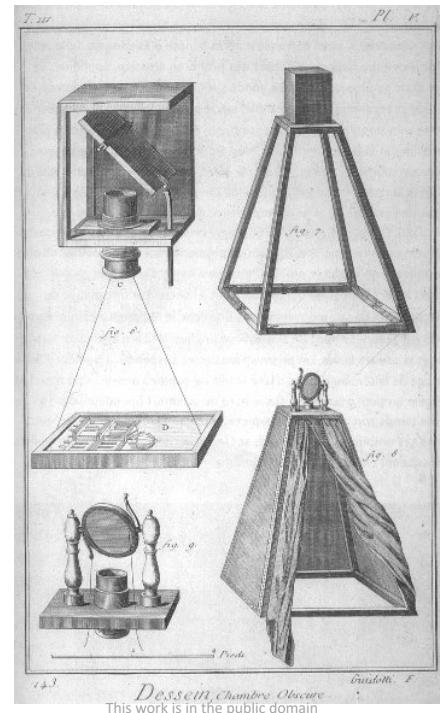
Camera Obscura

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?

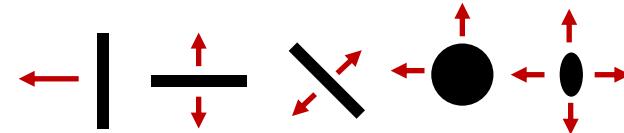
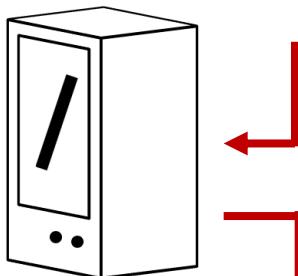
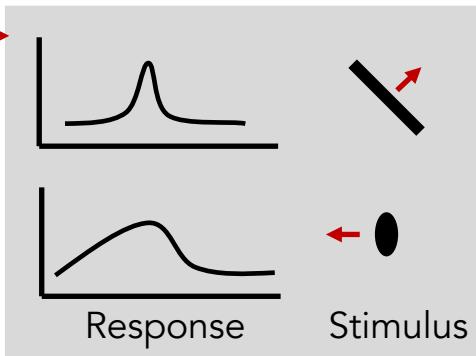
Hubel and Wiesel, 1959

Measure
brain activity



Cat image by CNX OpenStax is licensed under CC BY 4.0; changes made

1959
Hubel & Wiesel

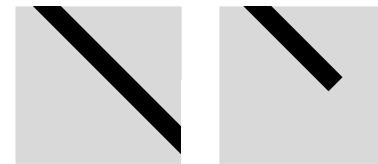


Simple cells:

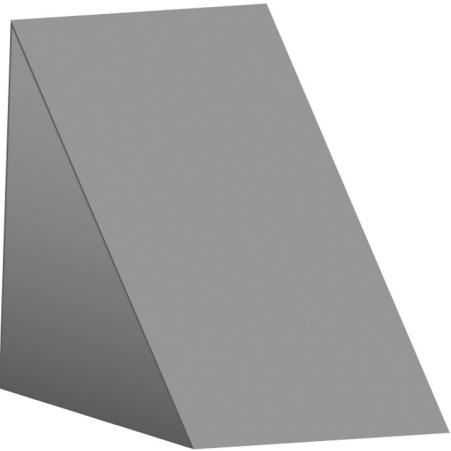
Response to specific
rotation and orientation

Complex cells:

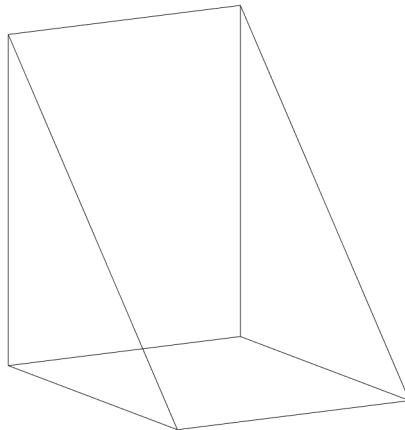
Response to light
orientation and
movement, some
translation invariance



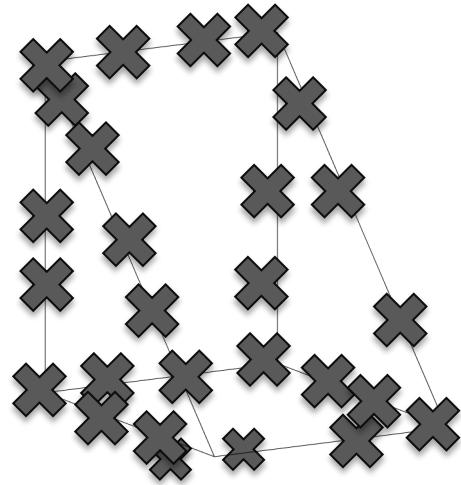
Larry Roberts, 1963



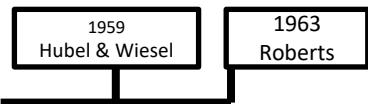
(a) Original picture



(b) Differentiated picture



(c) Feature points selected



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

Slide inspiration: Justin Johnson

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

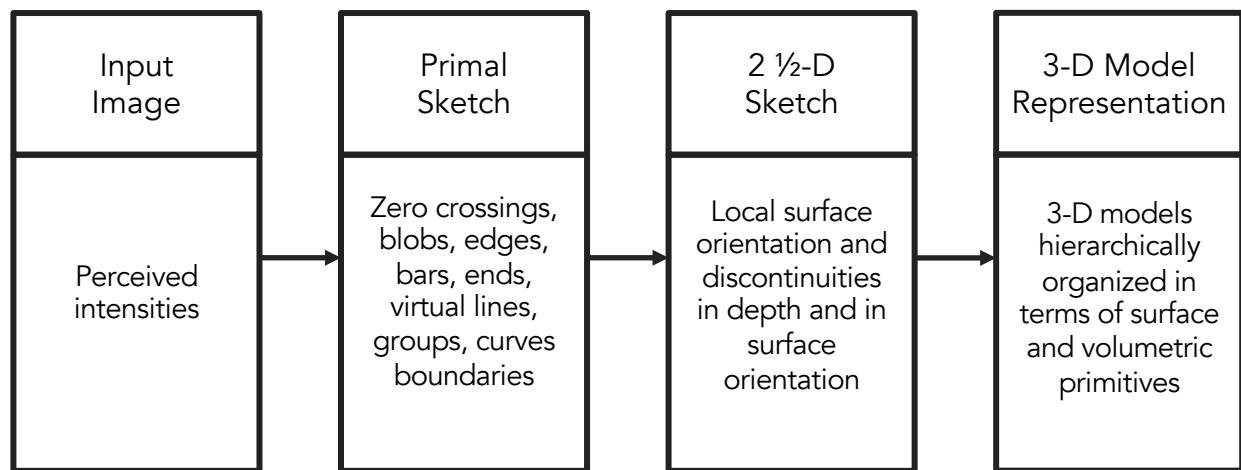
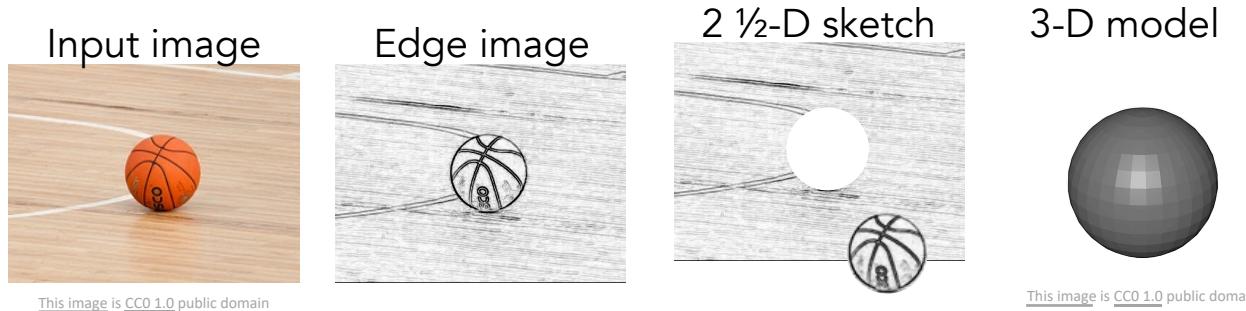
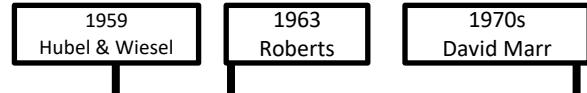
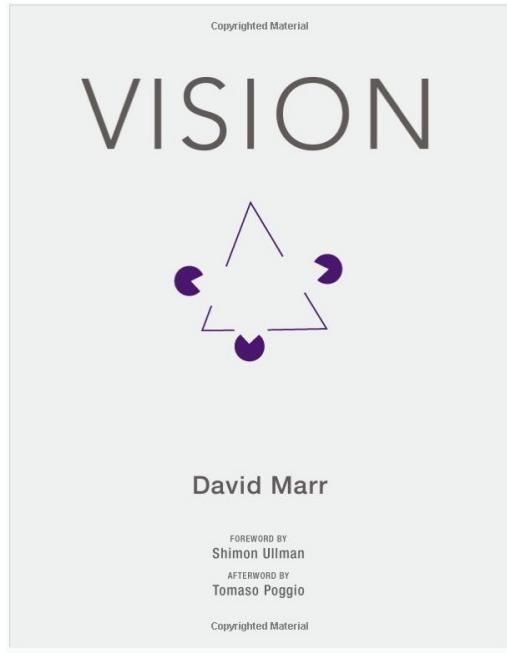
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

1959
Hubel & Wiesel

1963
Roberts

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

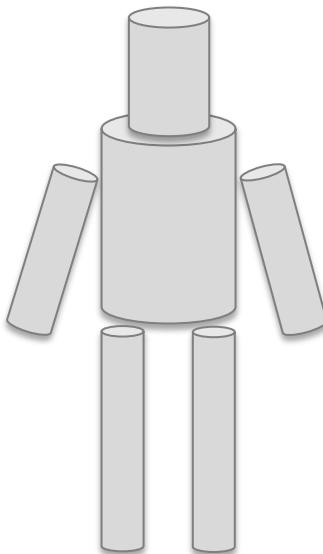
Slide inspiration: Justin Johnson



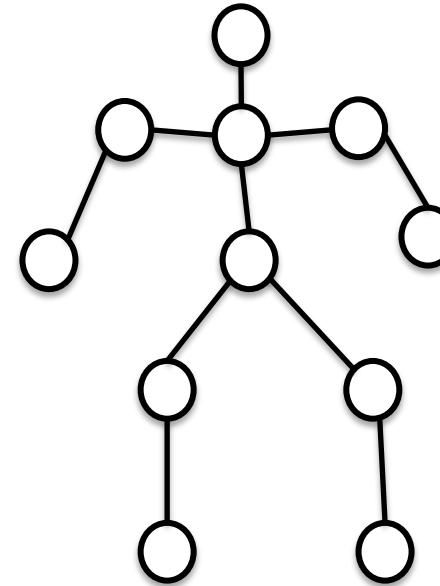
Stages of Visual Representation, David Marr, 1970s

Slide inspiration: Justin Johnson

Recognition via Parts (1970s)



Generalized Cylinders,
Brooks and Binford,
1979



Pictorial Structures,
Fischler and Elshlager, 1973



Recognition via Edge Detection (1980s)



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

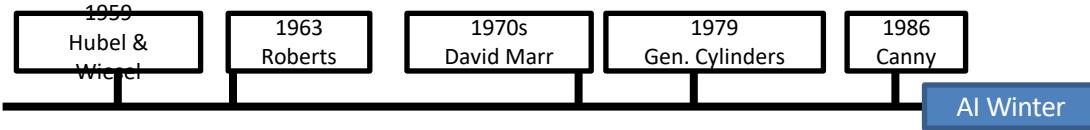
1986
Canny

John Canny, 1986
David Lowe, 1987

Image is CC0 1.0 public domain

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.



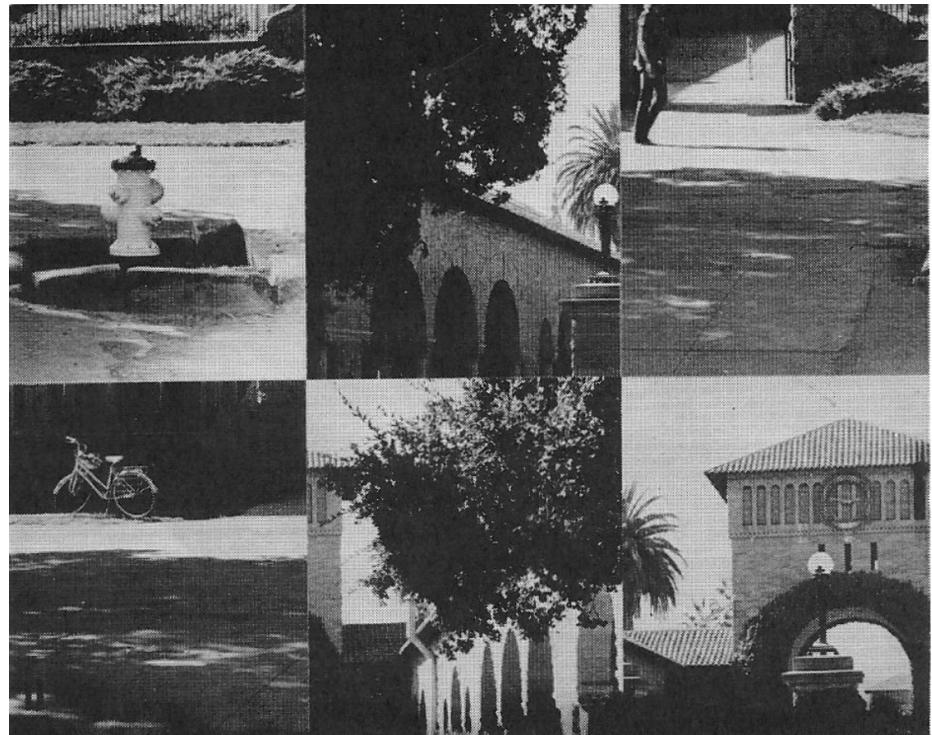
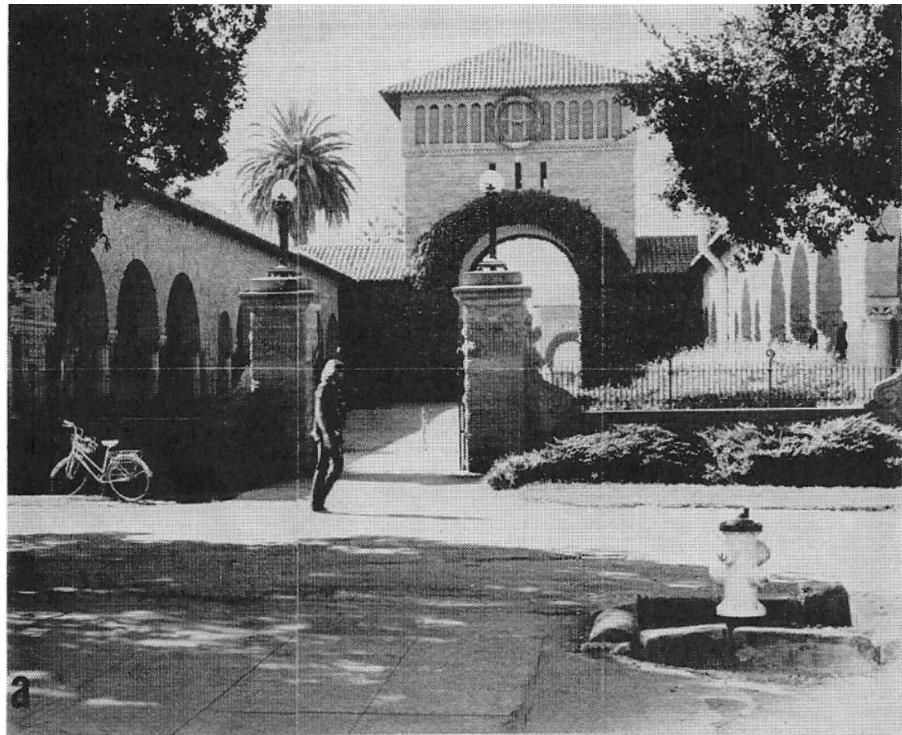
Left Image is CC BY 3.0
Middle Image is public domain

Right Image is CC-BY 2.0; changes made

In the meantime...seminal work in
cognitive and neuroscience

Perceiving Real-World Scenes

Irving Biederman



I. Biederman, *Science*, 1972

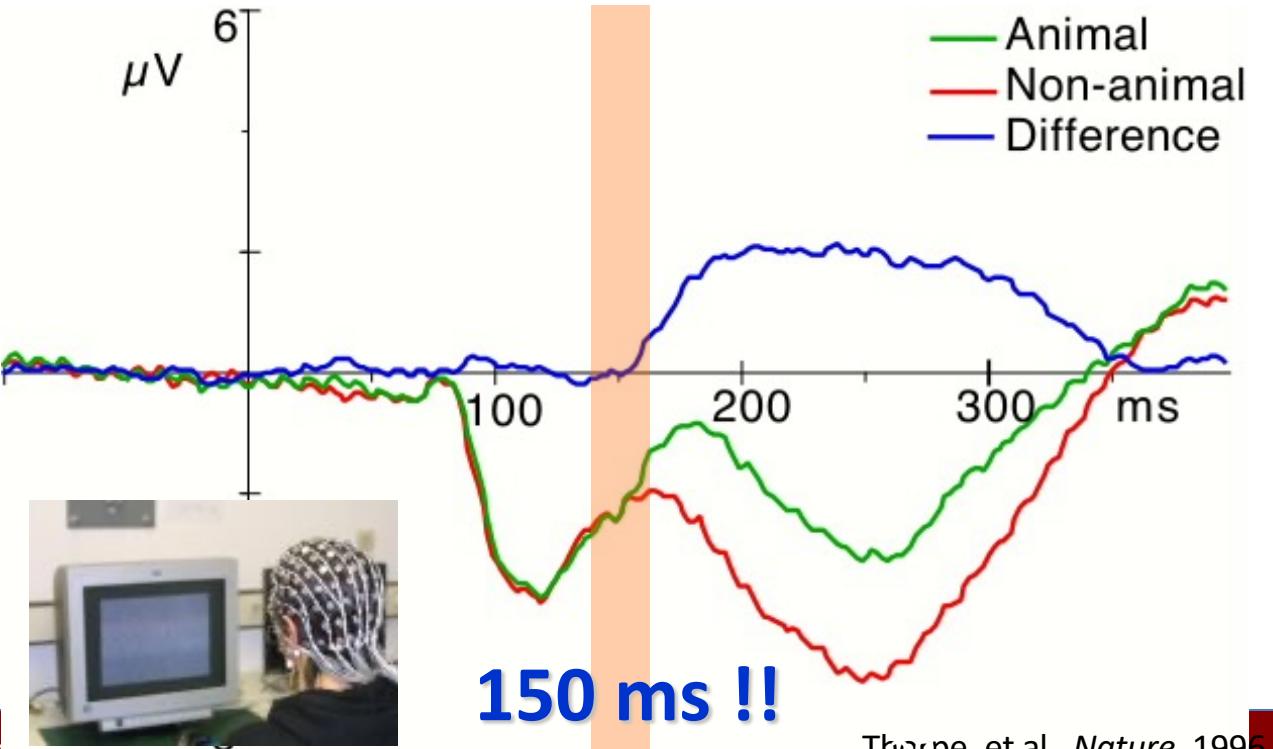
Rapid Serial Visual Perception (RSVP)



Potter, etc. 1970s

Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot



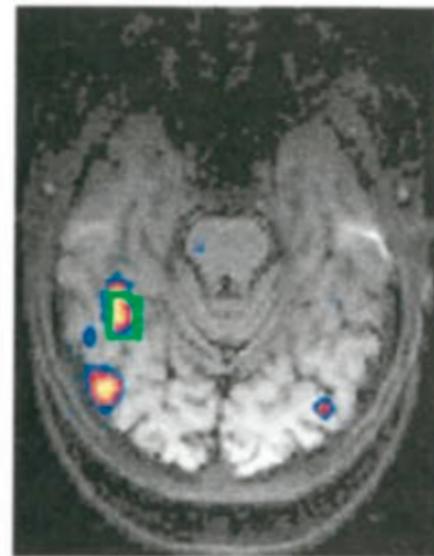
Thorpe, et al. *Nature*, 1996

28

4-Apr 23

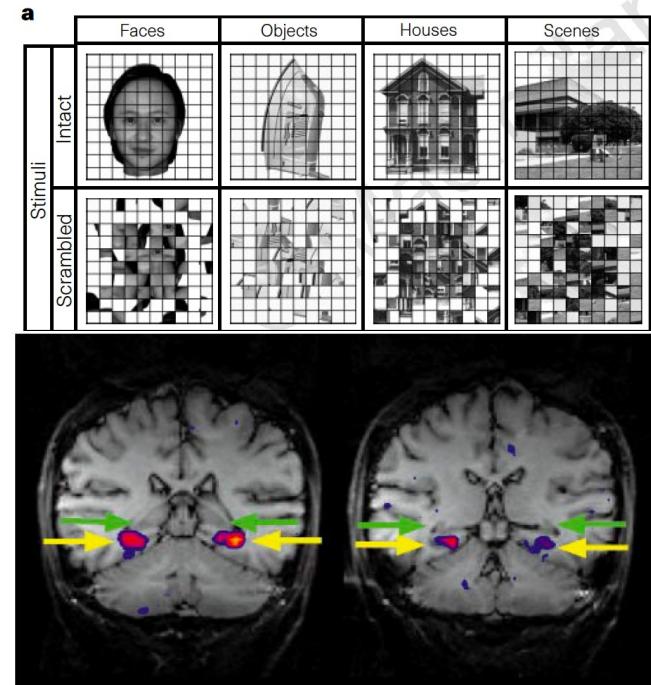
Neural correlates of object & scene recognition

Faces > Houses



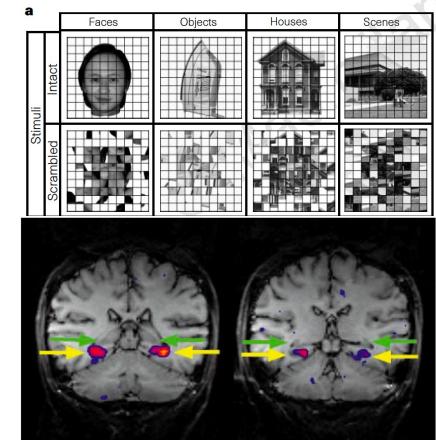
% signal change

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

1997
Norm. Cuts

AI Winter

Normalized Cuts, Shi and Malik, 1997

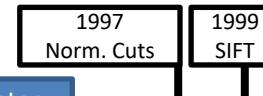
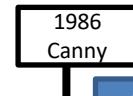
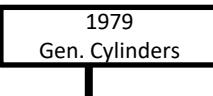
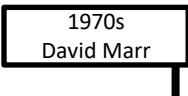
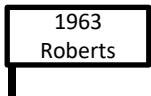
Recognition via Matching (2000s)



[Image](#) is public domain



[Image](#) is public domain



AI Winter

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



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

1999
SIFT

2001
V&J

AI Winter

Slide inspiration: Justin Johnson

Caltech 101 images



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

1999
SIFT

2001
V&J

2004, 2007
Caltech101;
PASCAL

AI Winter

PASCAL Visual Object Challenge

Image is CC0 1.0 public domain



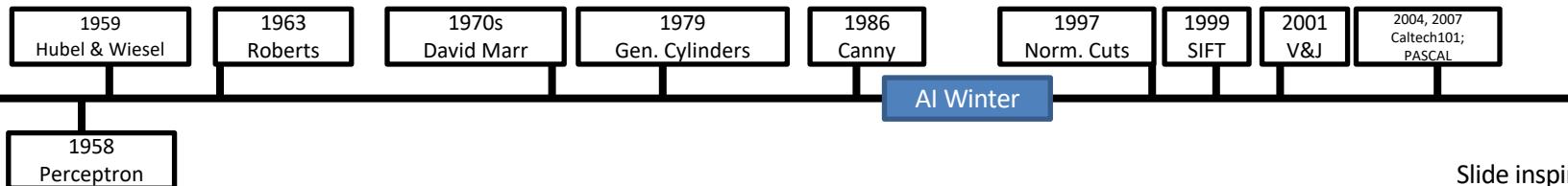
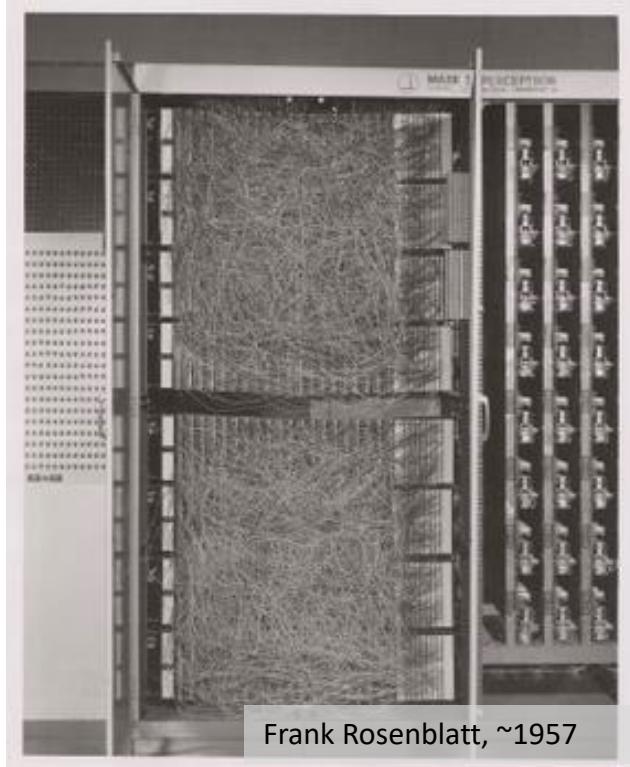
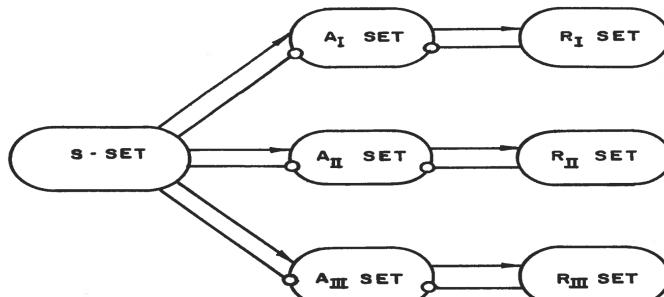
Image is CC0 1.0 public domain

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

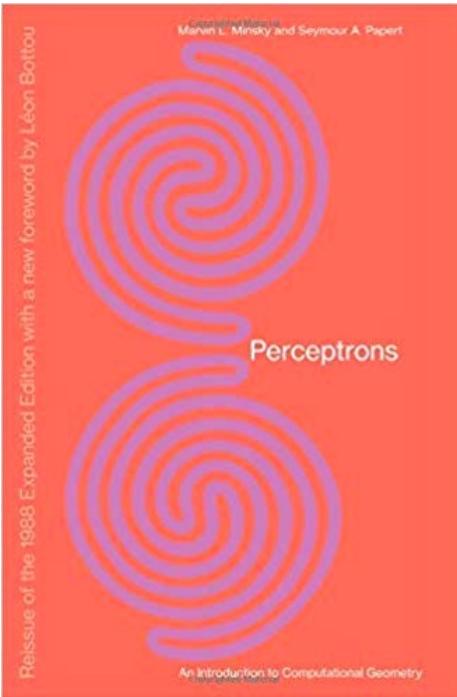
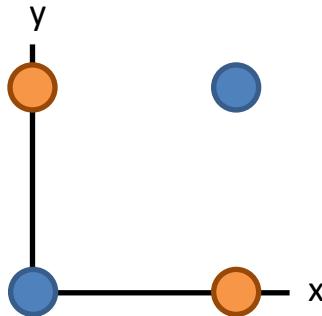
† Department of Computer Science, Carnegie-Mellon University,
Pittsburgh, Philadelphia 15213, USA



Slide inspiration: Justin Johnson

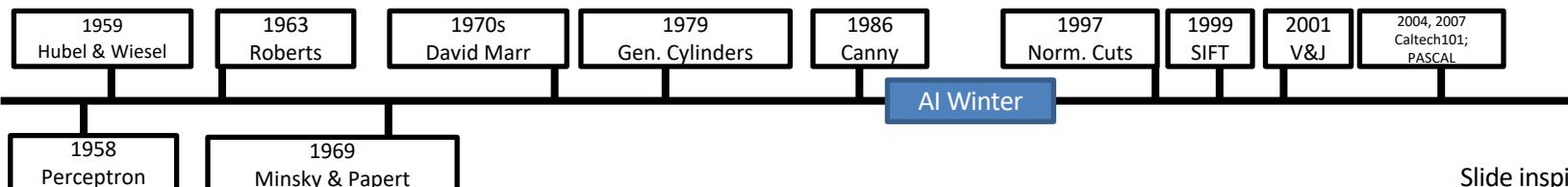
Minsky and Papert, 1969

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



Showed that Perceptrons could not learn the XOR function

Caused a lot of disillusionment in the field



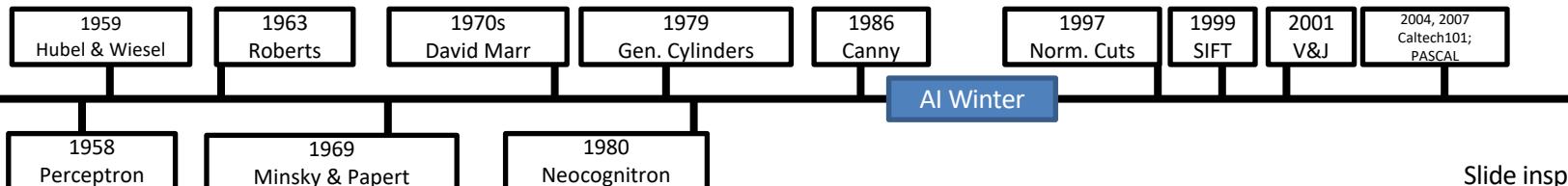
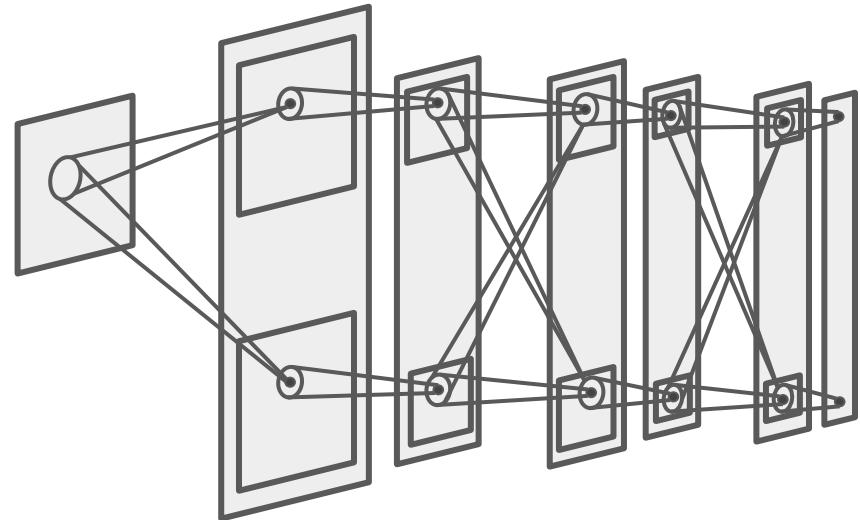
Slide inspiration: Justin Johnson

Neocognitron: Fukushima, 1980

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

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

No practical training algorithm

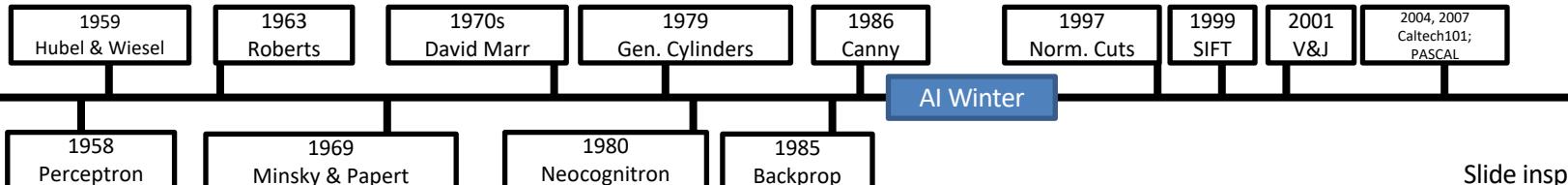
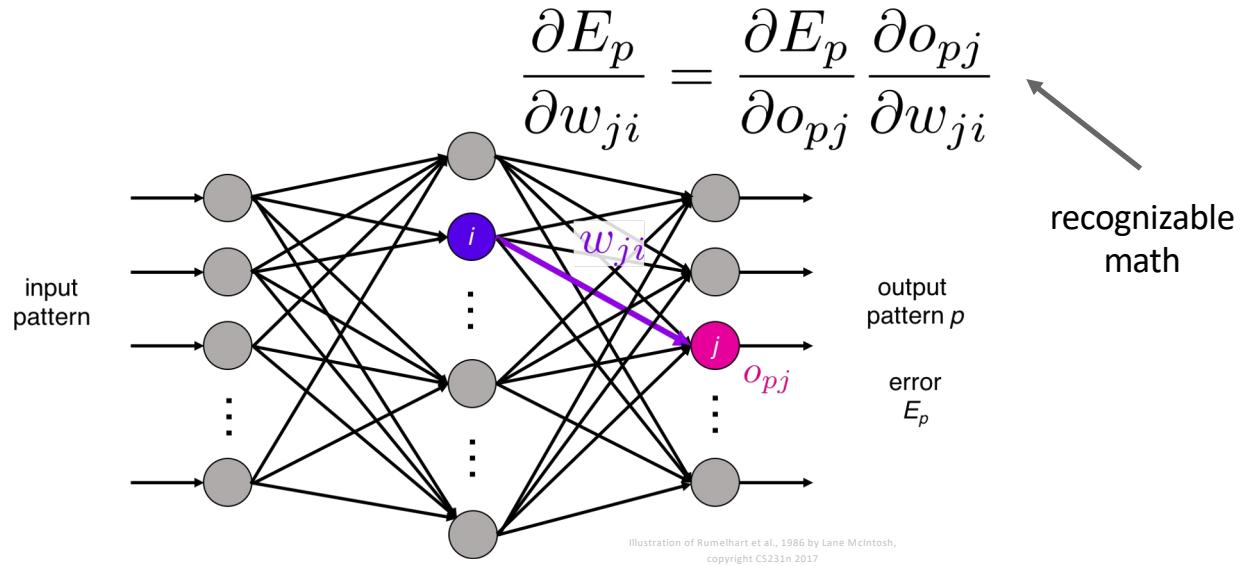


Slide inspiration: Justin Johnson

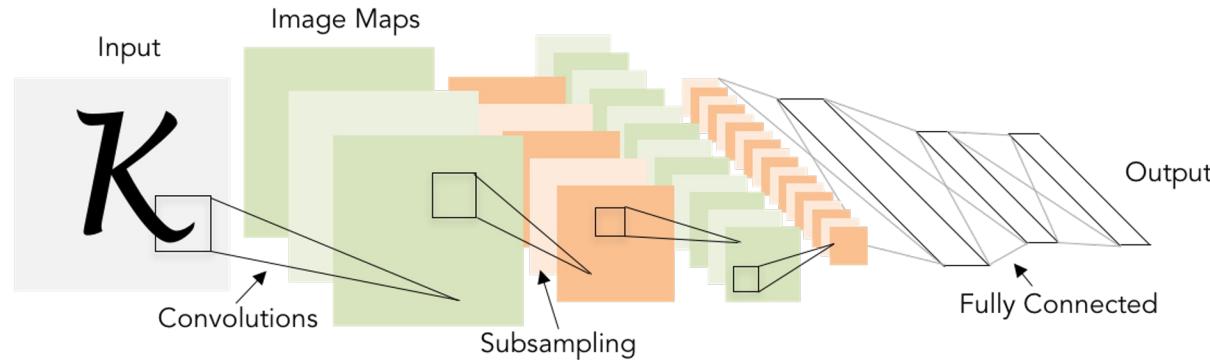
Backprop: Rumelhart, Hinton, and Williams, 1986

Introduced backpropagation for computing gradients in neural networks

Successfully trained perceptrons with multiple layers



Convolutional Networks: LeCun et al, 1998

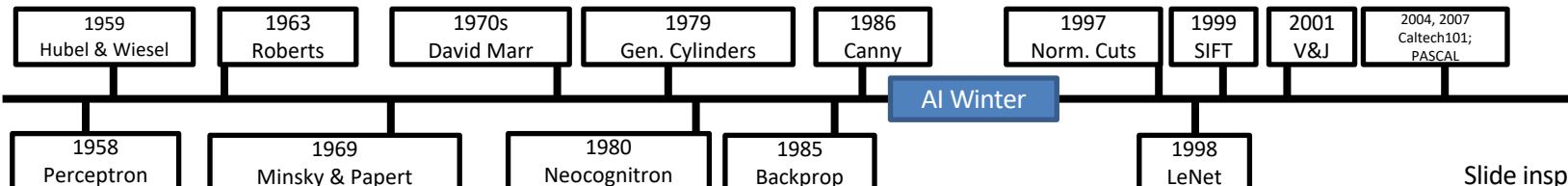


Applied backprop algorithm to a Neocognitron-like architecture

Learned to recognize handwritten digits

Was deployed in a commercial system by NEC, processed handwritten checks

Very similar to our modern convolutional networks!



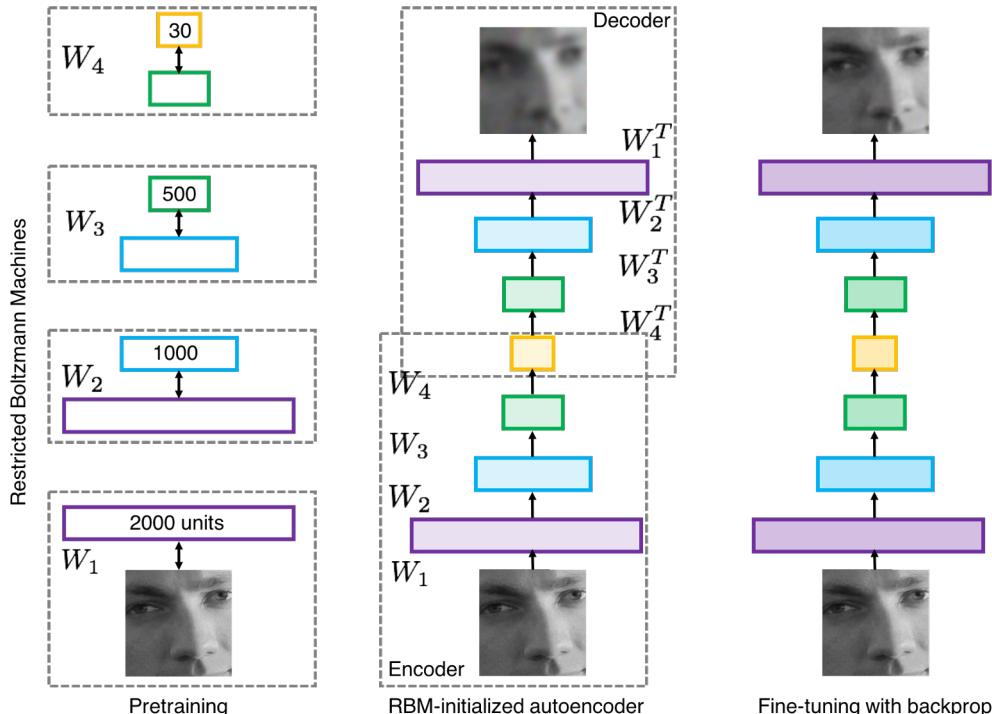
Slide inspiration: Justin Johnson

2000s: “Deep Learning”

People tried to train neural networks that were deeper and deeper

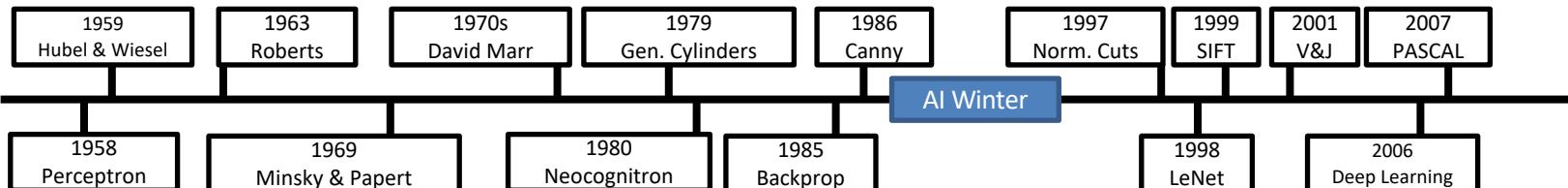
Not a mainstream research topic at this time

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



Fine-tuning with backprop

Slide inspiration: Justin Johnson



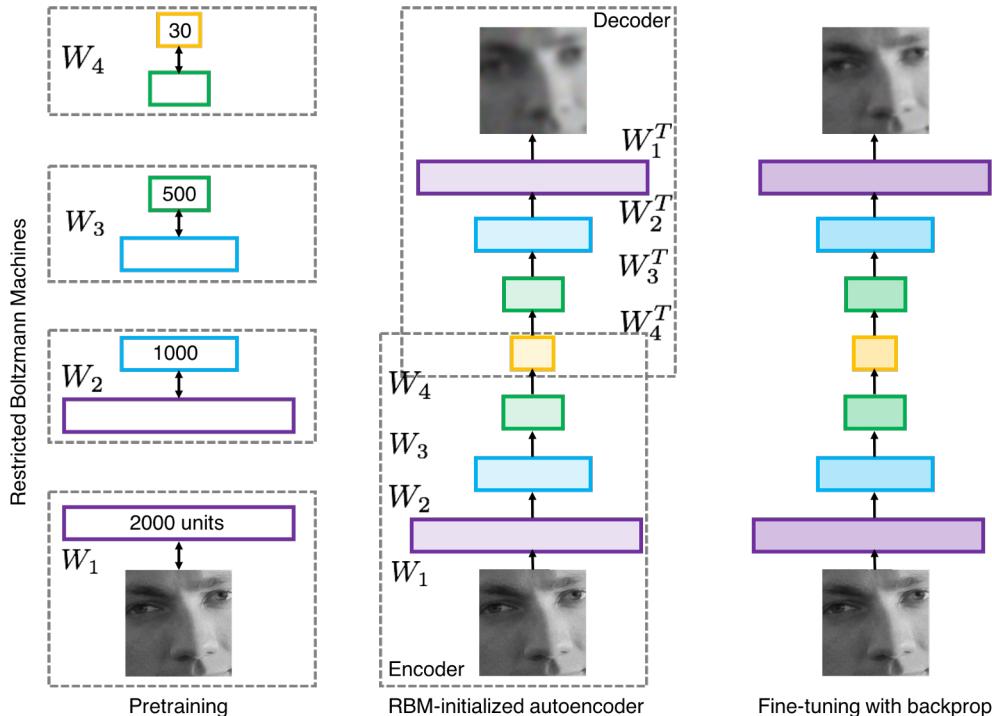
2000s: “Deep Learning”

People tried to train neural networks that were deeper and deeper

Not a mainstream research topic at this time

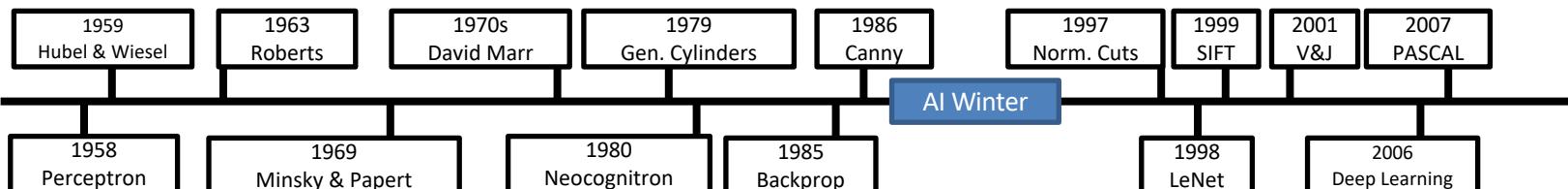
No good dataset to work on

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



Fine-tuning with backprop

Slide inspiration: Justin Johnson



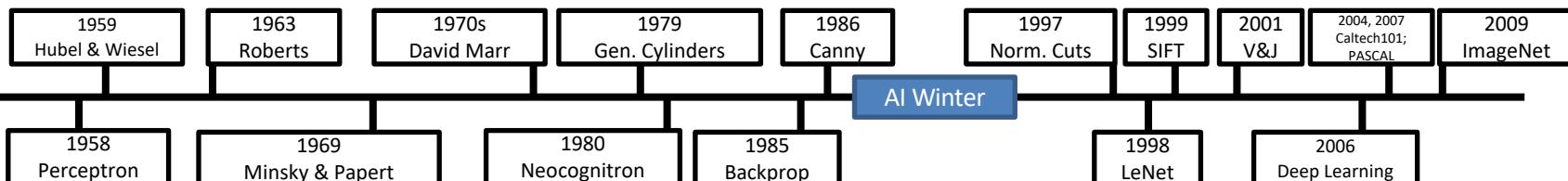
IMAGENET Large Scale Visual Recognition Challenge

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

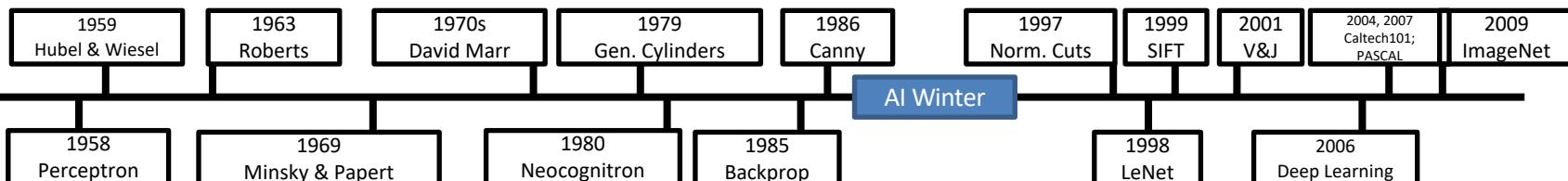
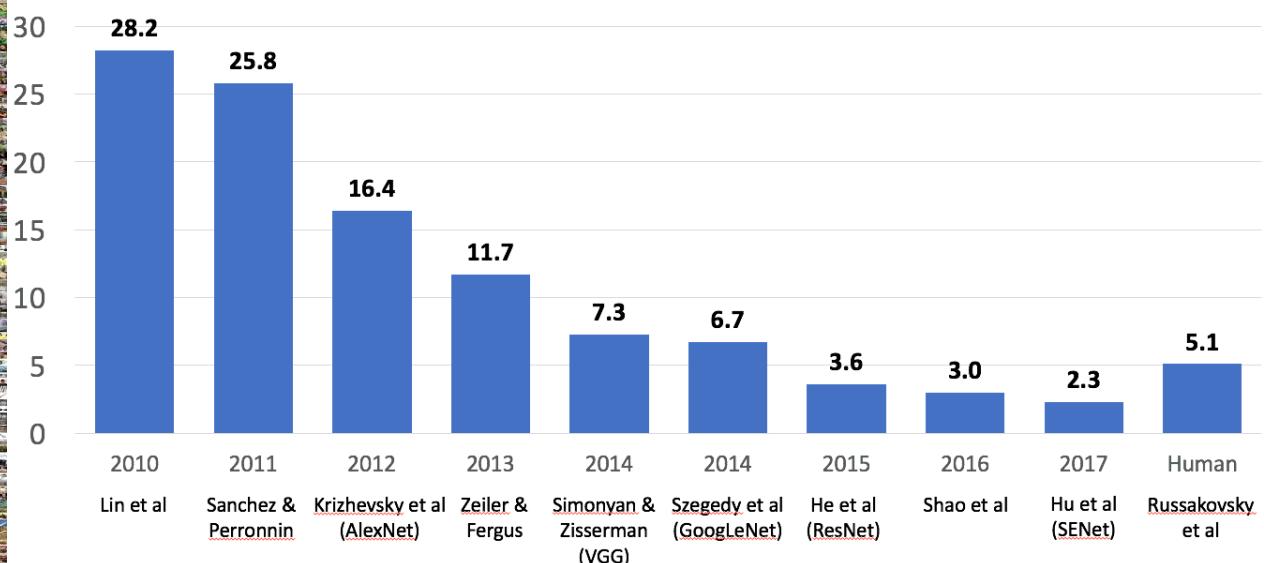


Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

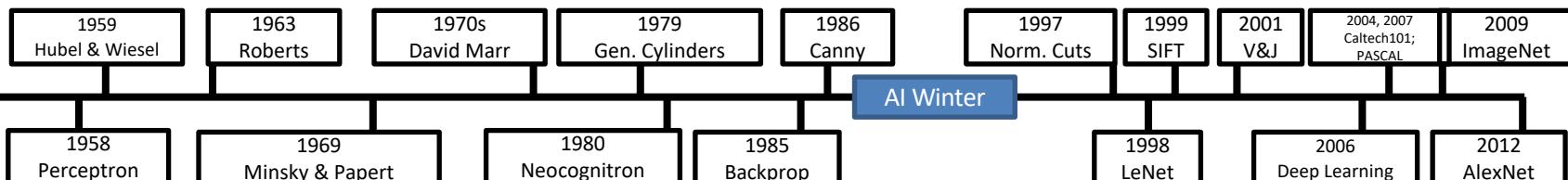
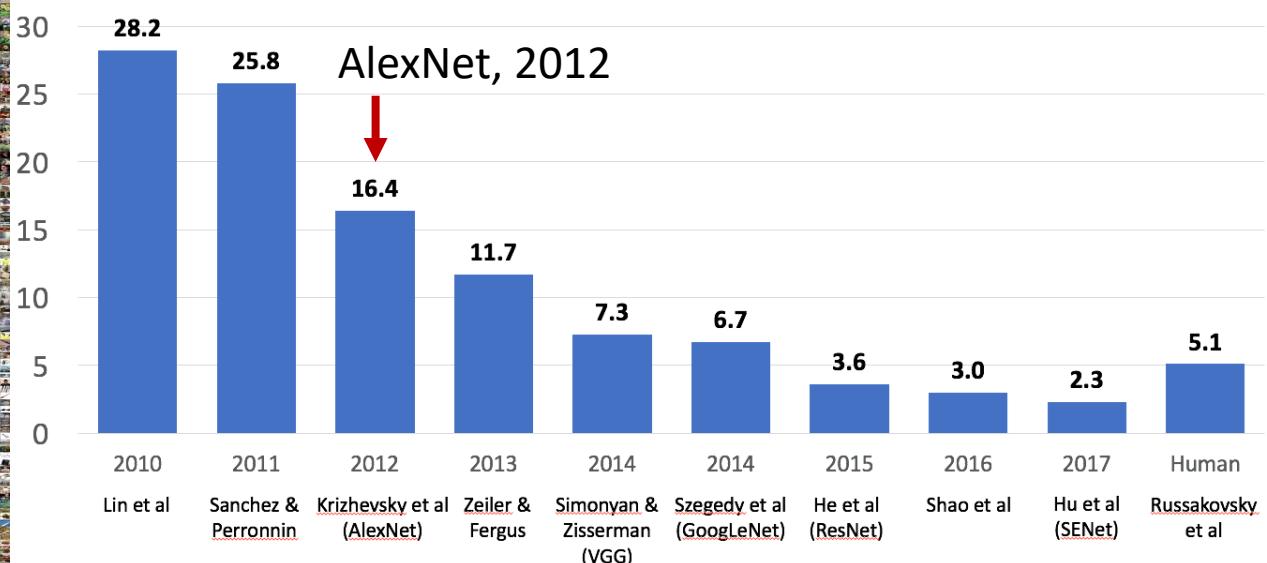
Deng et al, 2009
Russakovsky et al. IJCV 2015



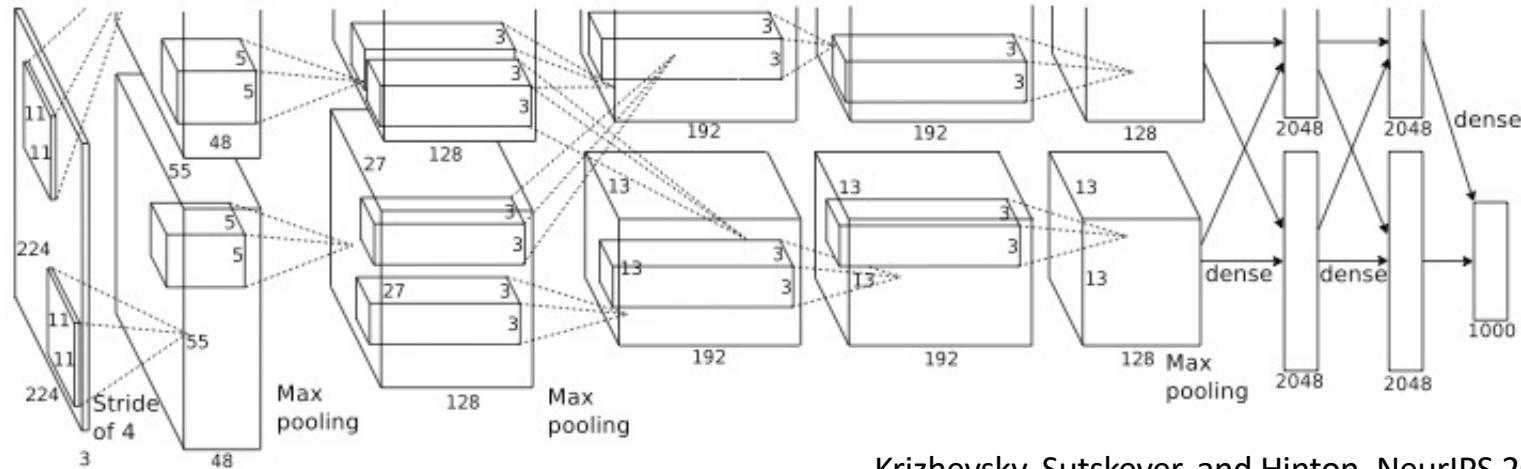
IMAGENET Large Scale Visual Recognition Challenge



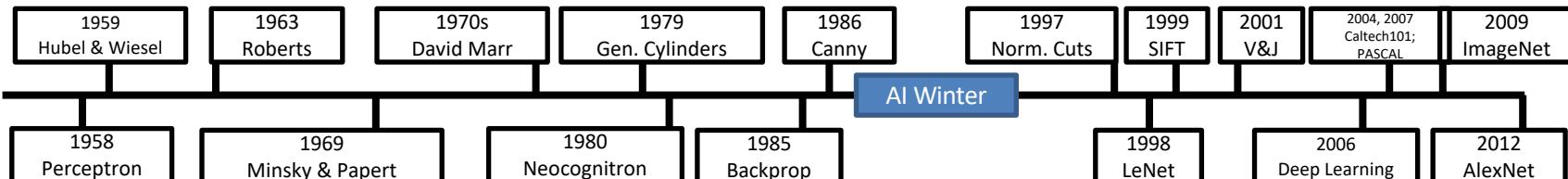
IMAGENET Large Scale Visual Recognition Challenge



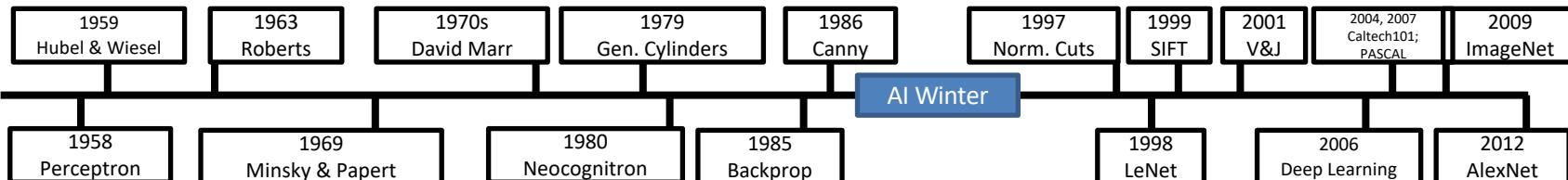
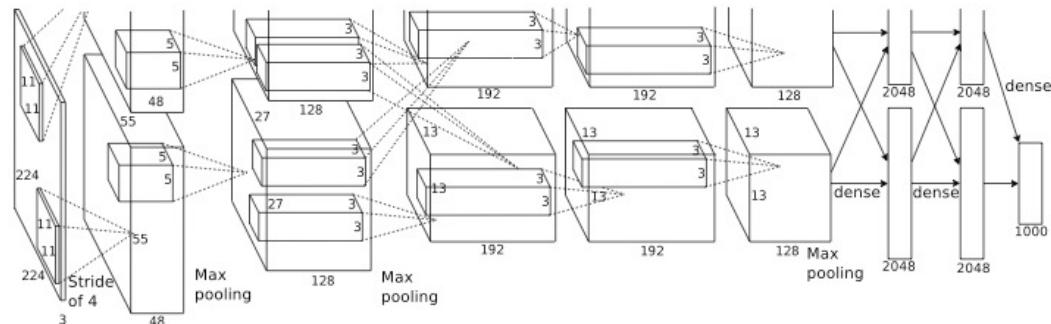
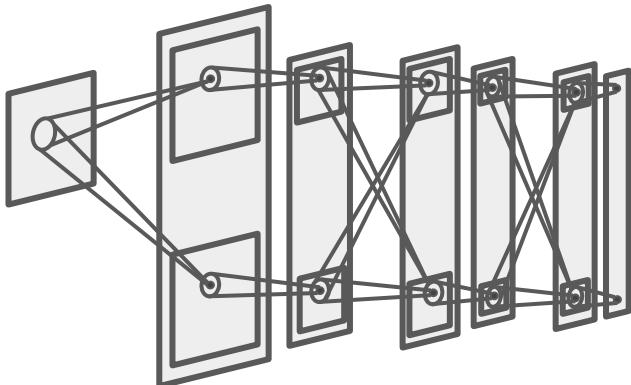
AlexNet: Deep Learning Goes Mainstream



Krizhevsky, Sutskever, and Hinton, NeurIPS 2012

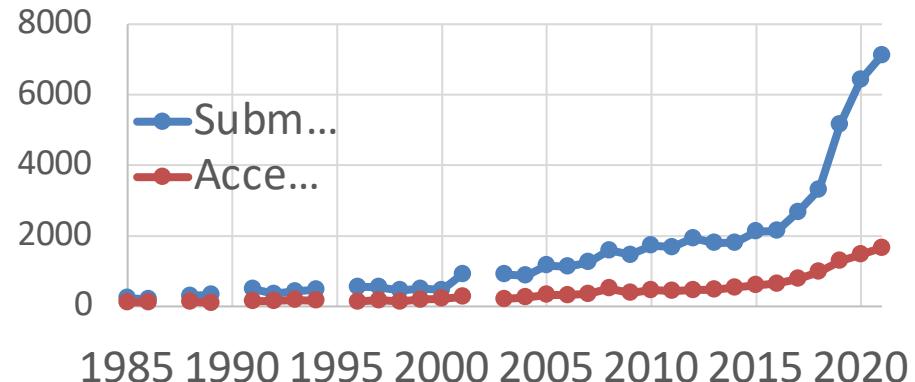


AlexNet vs. Neocognitron: 32 years apart

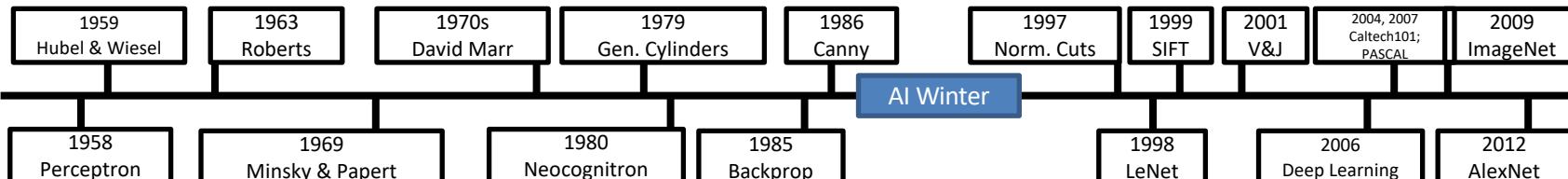


2012 to Present: Deep Learning Explosion

CVPR Papers



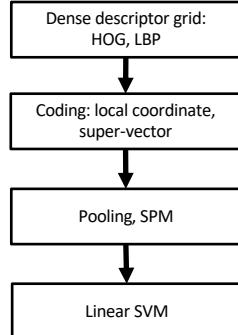
Publications at top Computer Vision conference



2012 to Present: Deep Learning is Everywhere

Year 2010

NEC-UIUC

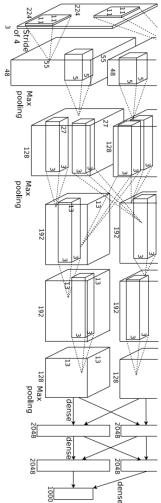


[Lin CVPR 2011]

Lion image by Swissfrog
is
licensed under CC-BY 3.0

Year 2012

SuperVision



[Krizhevsky NIPS 2012]

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

Year 2014

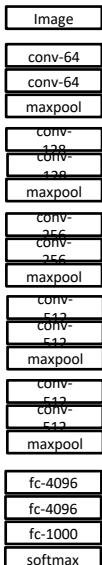
GoogLeNet

● Pooling
● Convolution
● Softmax
● Other



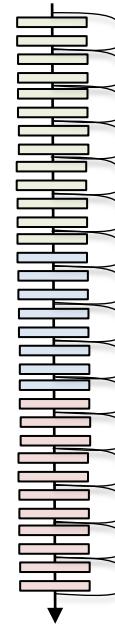
[Szegedy arxiv 2014] [Simonyan arxiv 2014]

VGG



Year 2015

MSRA



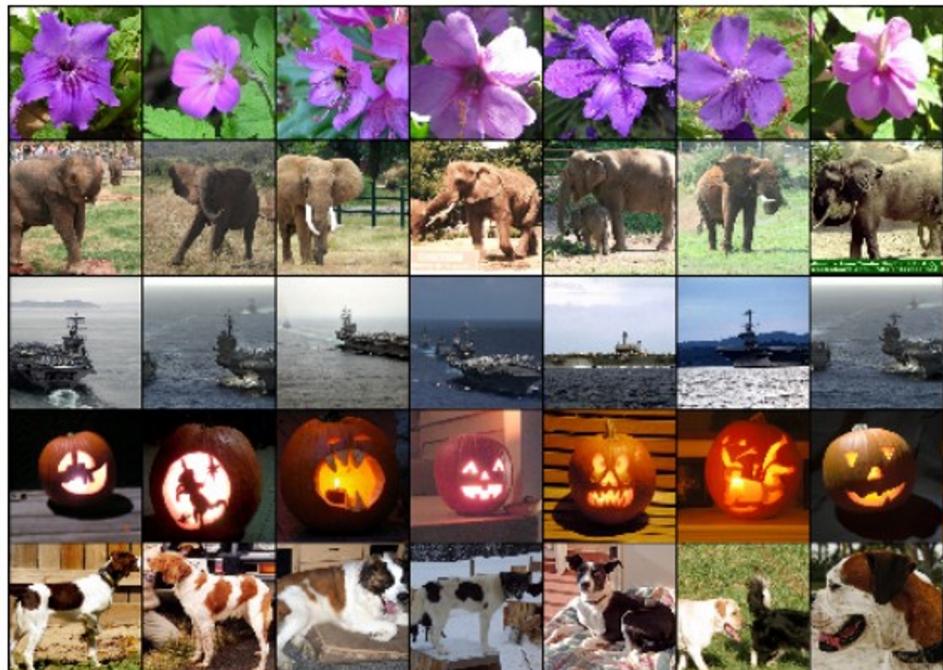
[He ICCV 2015]

2012 to Present: Deep Learning is Everywhere

Image Classification



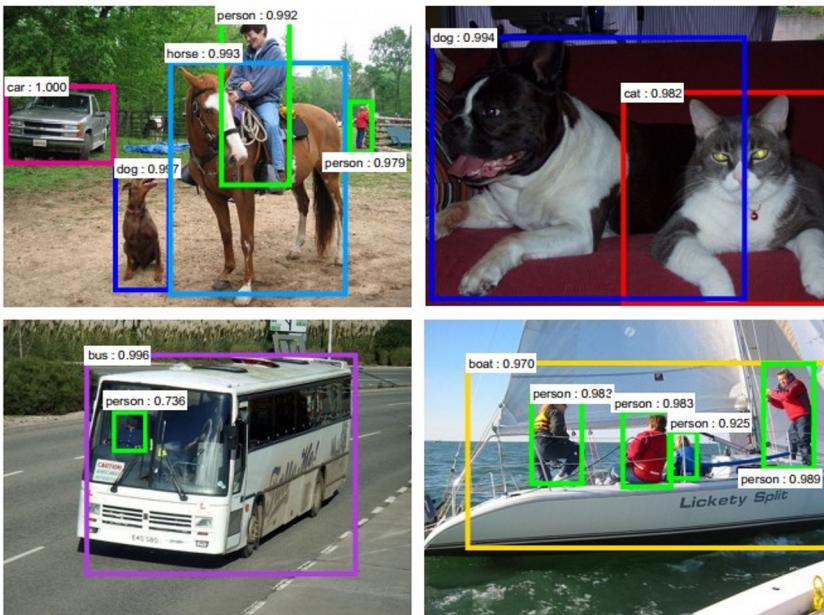
Image Retrieval



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

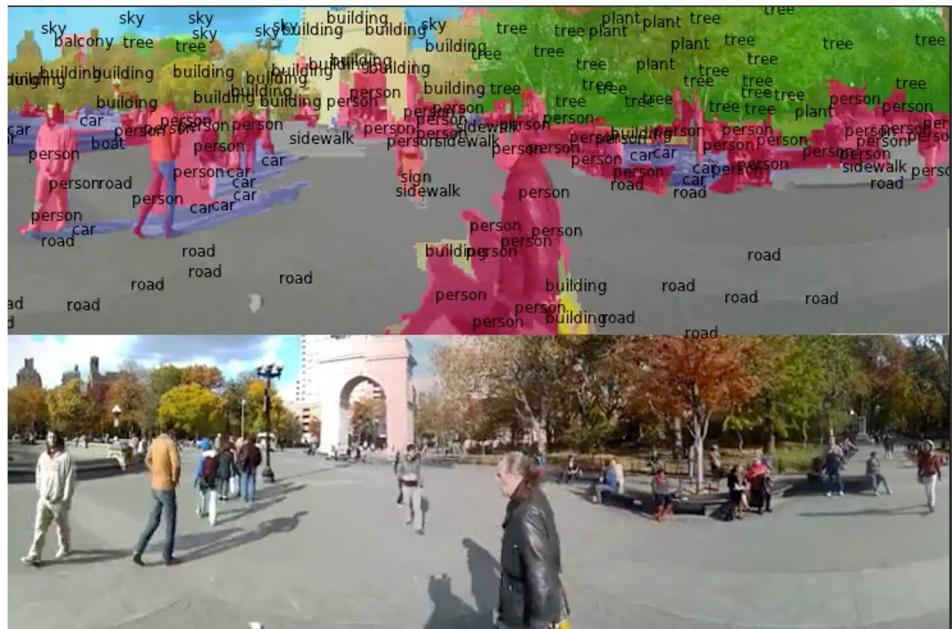
2012 to Present: Deep Learning is Everywhere

Object Detection



Ren, He, Girshick, and Sun, 2015

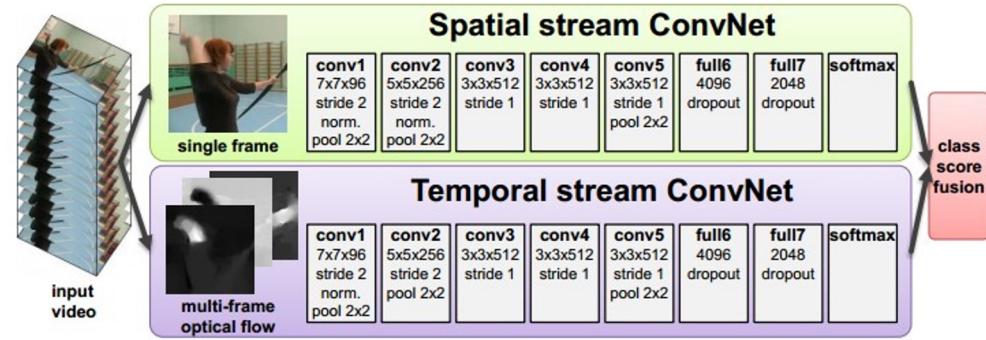
Image Segmentation



Fabaret et al, 2012

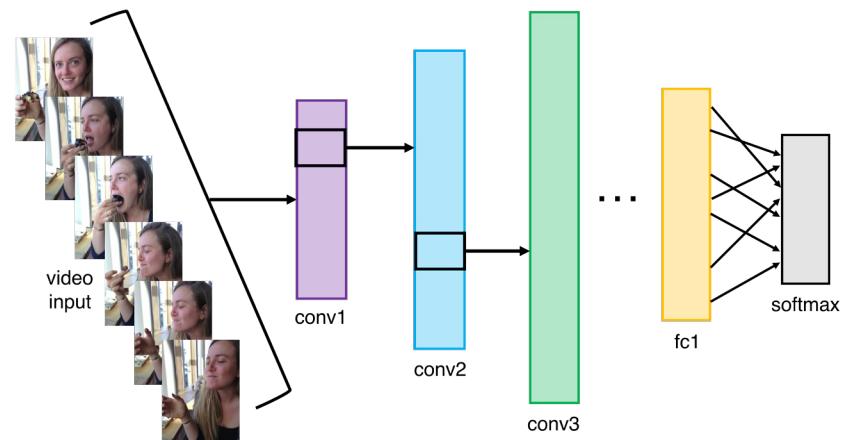
2012 to Present: Deep Learning is Everywhere

Video Classification



Simonyan et al, 2014

Activity Recognition

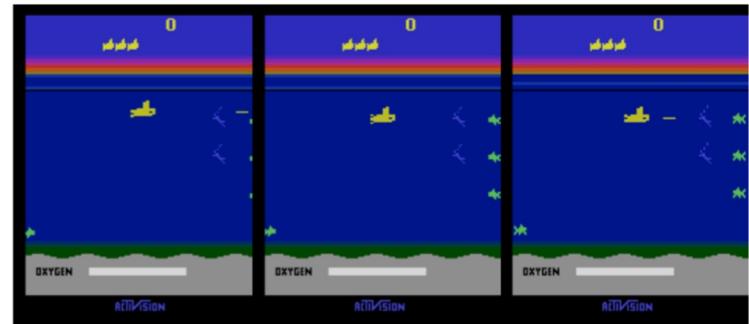
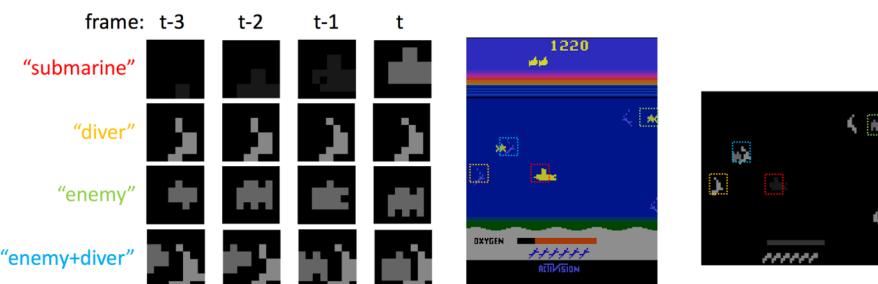


2012 to Present: Deep Learning is Everywhere

Pose Recognition (Toshev and Szegedy, 2014)

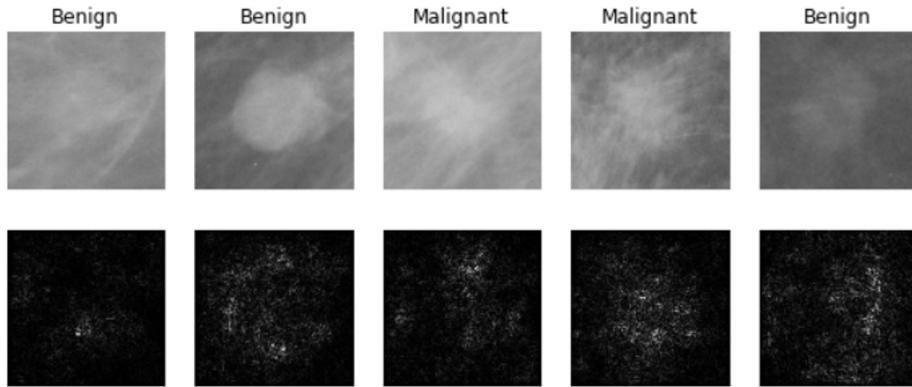


Playing Atari games (Guo et al, 2014)



2012 to Present: Deep Learning is Everywhere

Medical Imaging



Levy et al, 2016

Figure reproduced with permission

Whale recognition



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

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

Kaggle Challenge

2012 to Present: Deep Learning is Everywhere



*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*



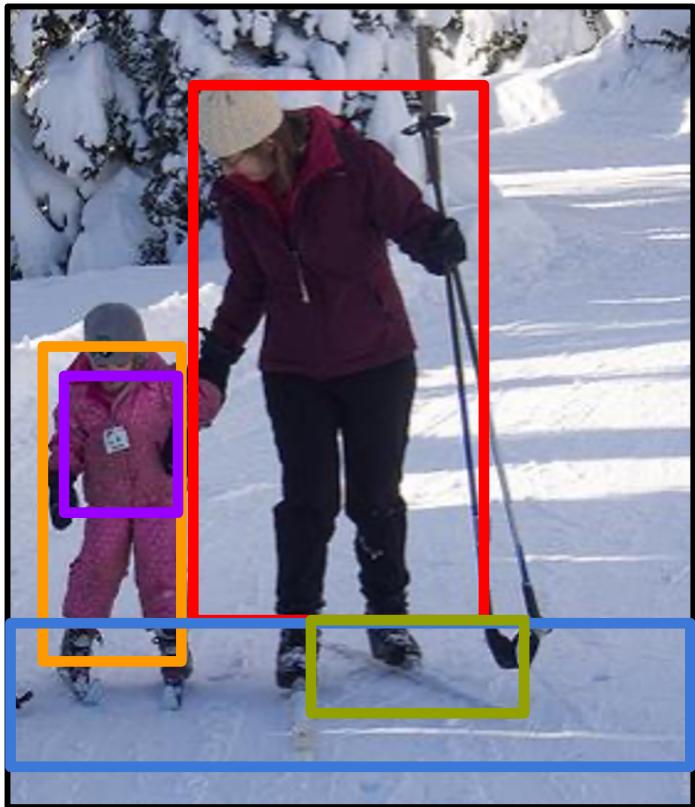
*A woman standing on a
beach holding a surfboard*

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

All images are CC0 Public domain:
<https://pixabay.com/en/teddy-plush-bear-cute-teddy-bear-1623010/>
<https://pixabay.com/en/teal-plush-teddy-bear-toy-1623436/>
<https://pixabay.com/en/surf-wave-summer-sport-litoral-1668716/>
<https://pixabay.com/en/woman-female-model-portrait-adult-983967/>
<https://pixabay.com/en/handstand-lake-meditation-496008/>
<https://pixabay.com/en/baseball-player-shortstop-infield-1045263/>

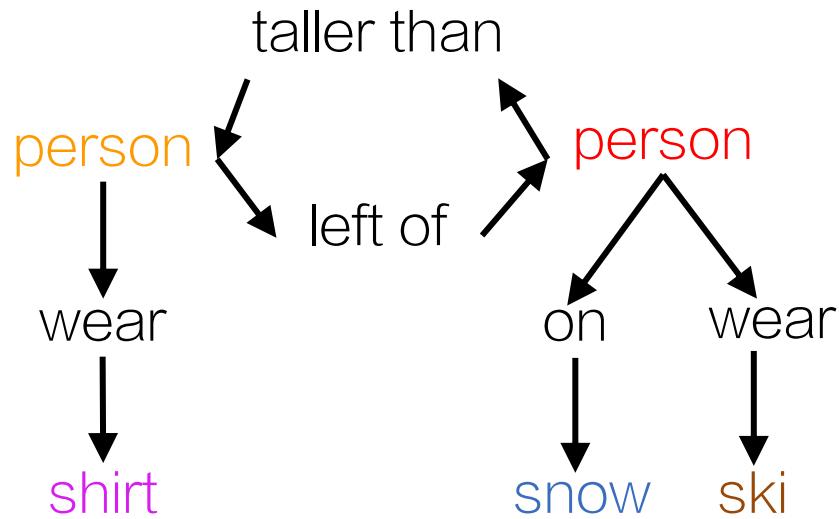
Captions generated by Justin Johnson using [Neuraltalk2](#)

2012 to Present: Deep Learning is Everywhere

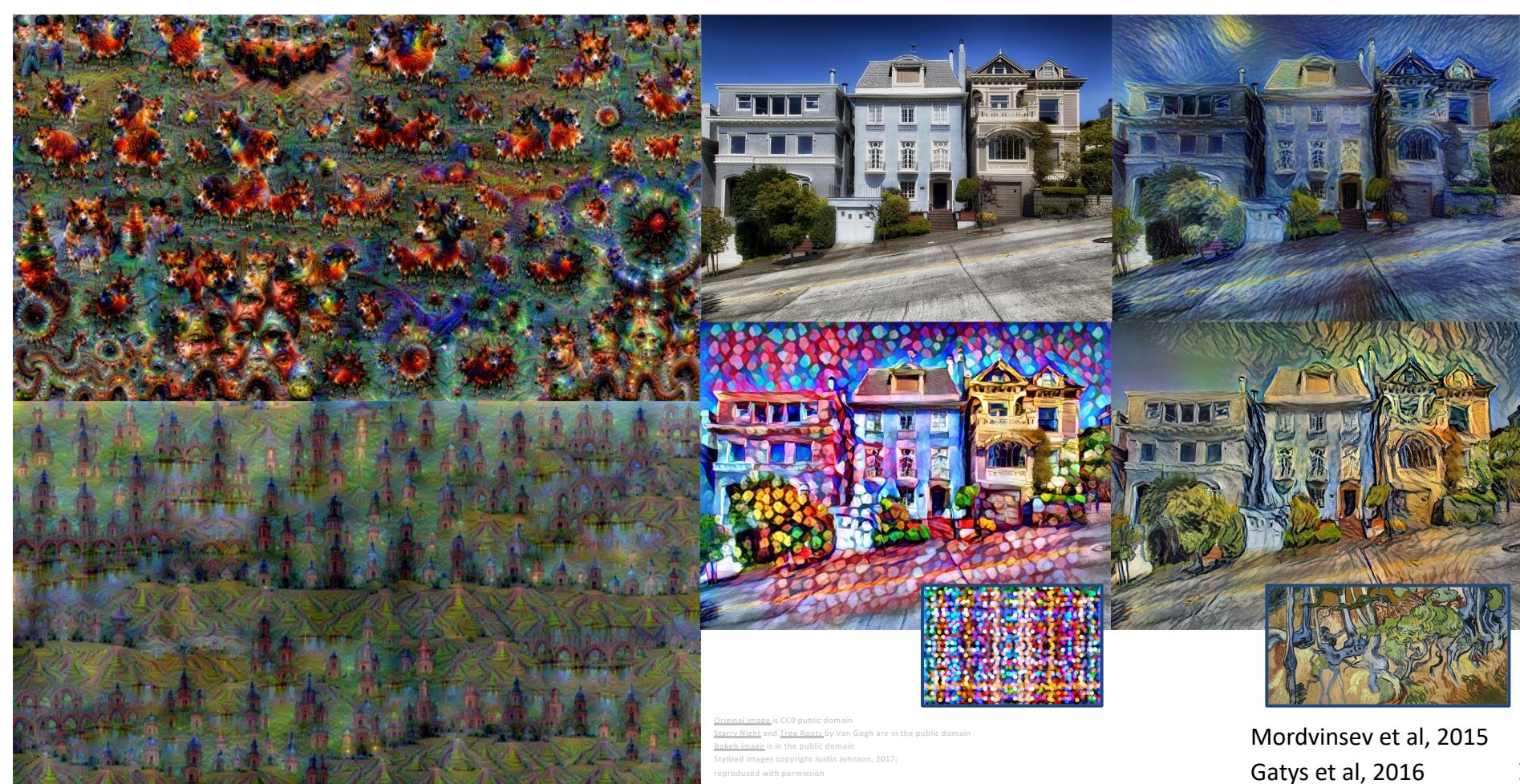


Results:

spatial, comparative, asymmetrical, verb,
prepositional



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



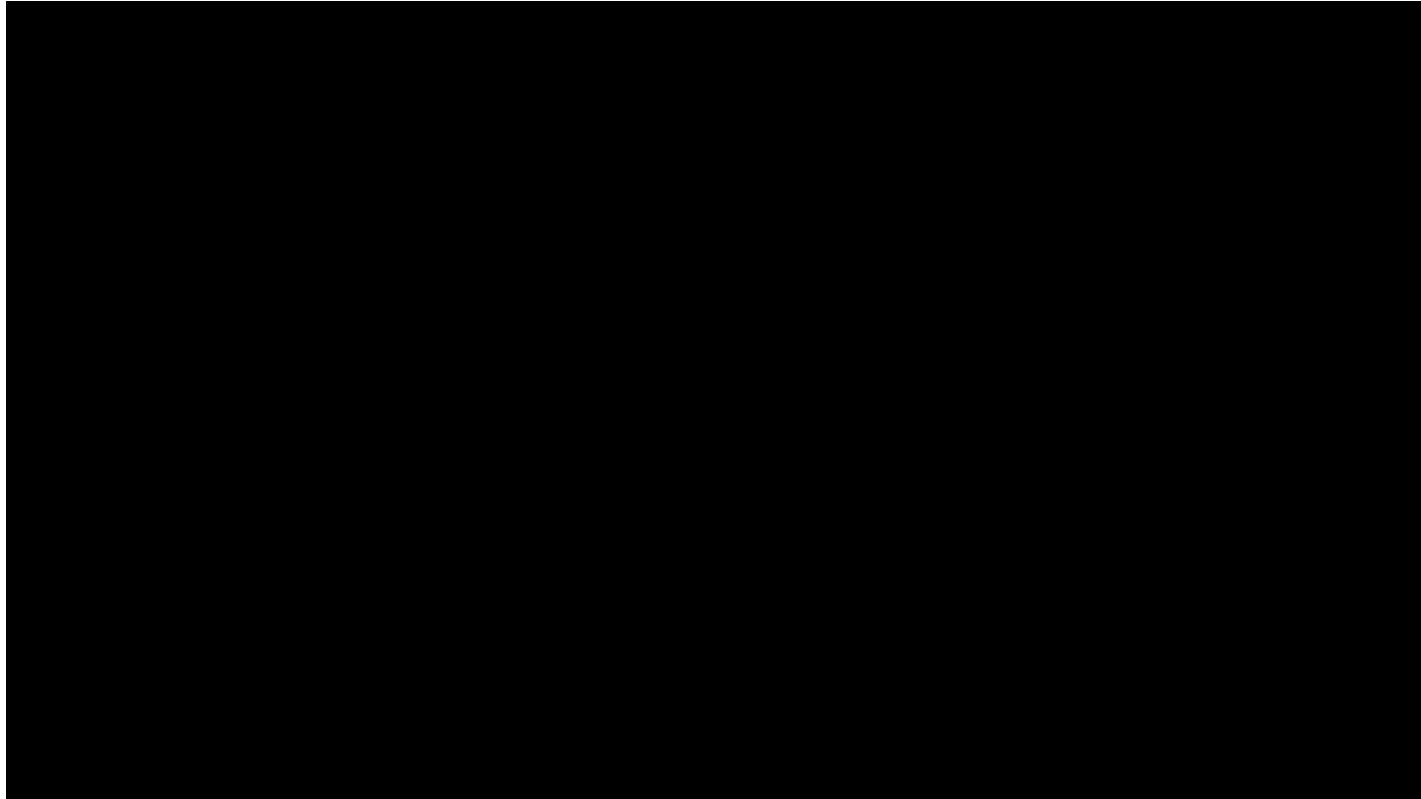
[Original Image](#) is CC0 public domain
[Starry Night](#) and [Tree Roots](#) by Van Gogh are in the public domain
Bokeh image is in the public domain
Stylized images copyright Justin Johnson, 2017;
reproduced with permission

Mordvinsev et al, 2015
Gatys et al, 2016

Slide inspiration: Justin Johnson

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

2012 to Present: Deep Learning is Everywhere



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

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

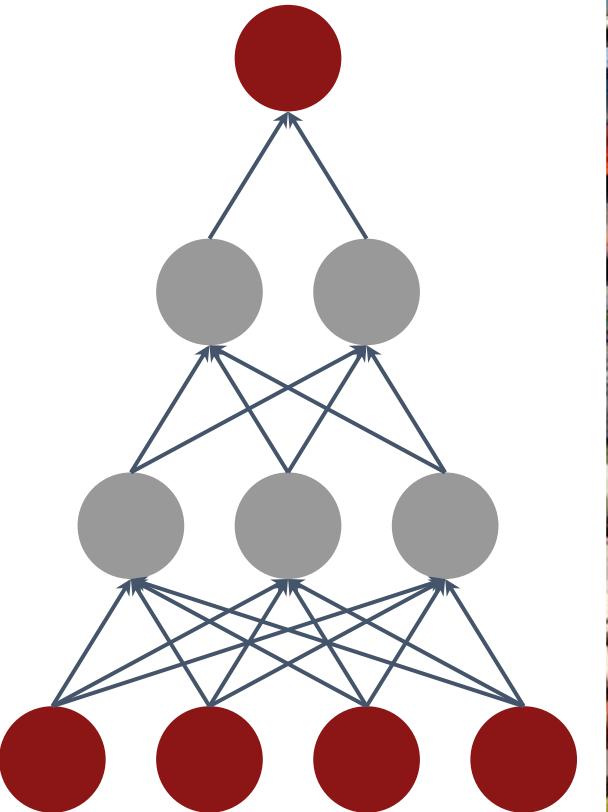


Ramesh et al., "DALL-E: Creating Images from Text", 2021. <https://openai.com/blog/dall-e/>



Computation

4 Apr 23



Algorithms



Data

60

GFLOP per Dollar

● CPU ● GPU (FP32)

RTX 3080 →

RTX 3090 →

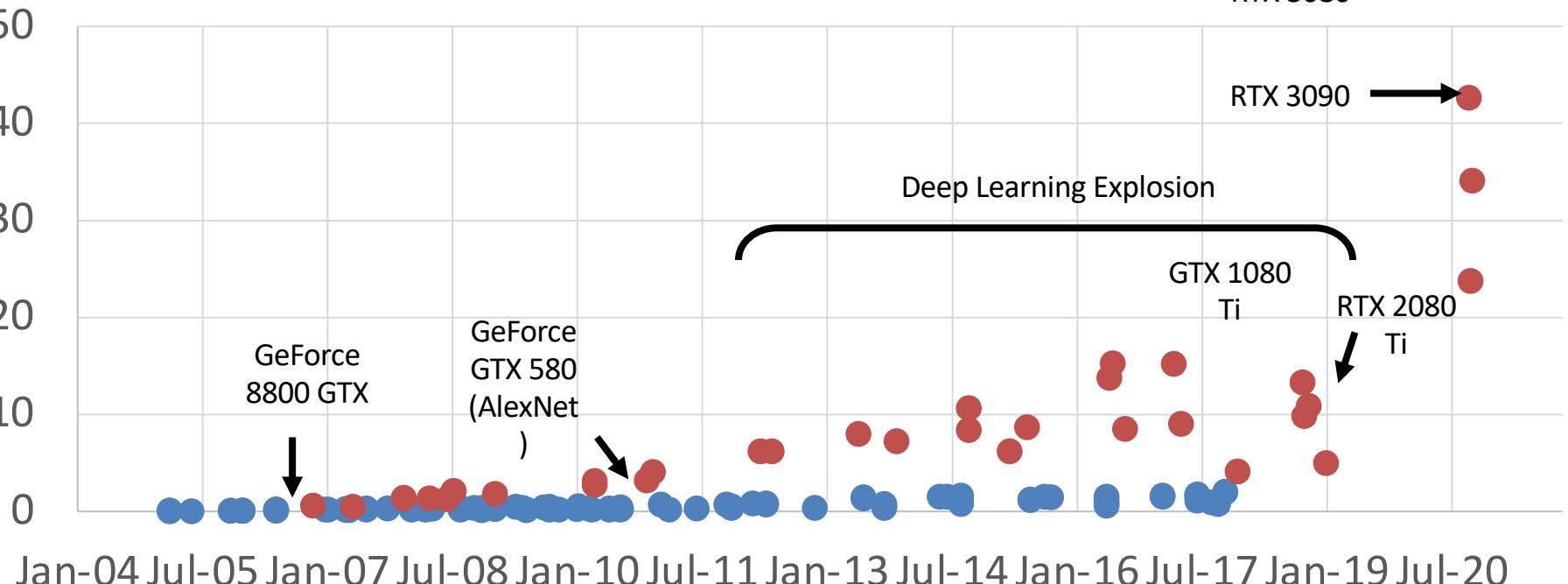
Deep Learning Explosion

GTX 1080
Ti

RTX 2080
Ti

GeForce
8800 GTX

GeForce
GTX 580
(AlexNet
)



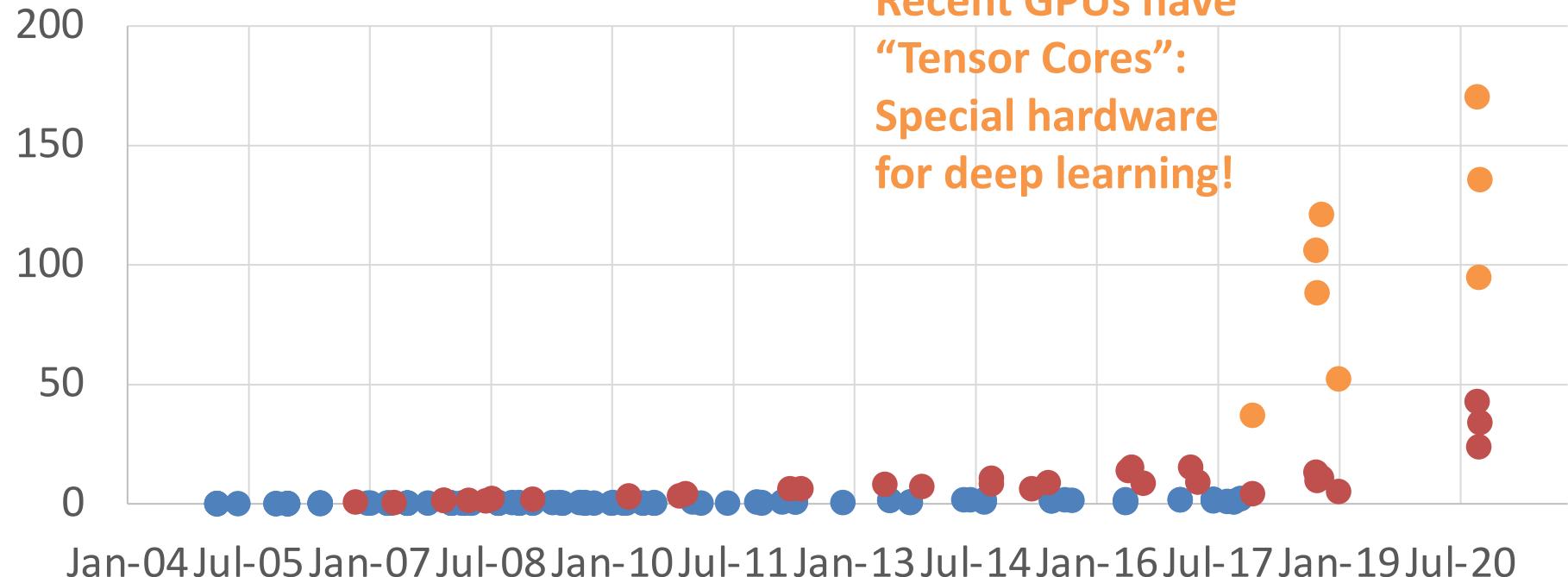
GFLOP per Dollar

● CPU

● GPU (FP32)

● GPU (Tensor Core)

Recent GPUs have
“Tensor Cores”:
Special hardware
for deep learning!



AI's Explosive Growth & Impact

Attendance at large conferences (1984-2019)
Source: Conference provided data.

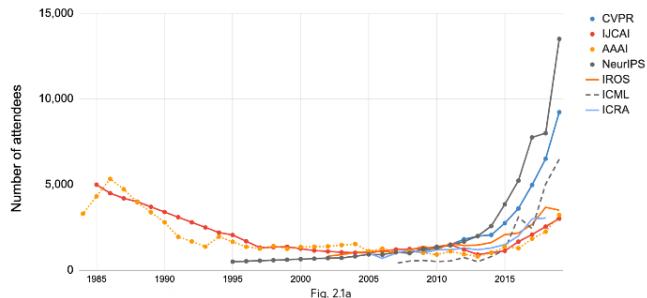
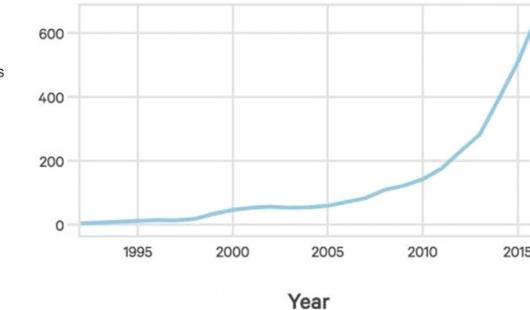


Fig. 2.1a

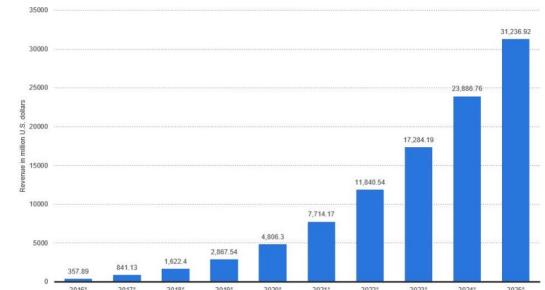
**Number of attendance
At AI conferences**

Source: The Gradient



**Startups Developing AI
Systems**

Source: Crunchbase, VentureSource, Sand
Hill Econometrics



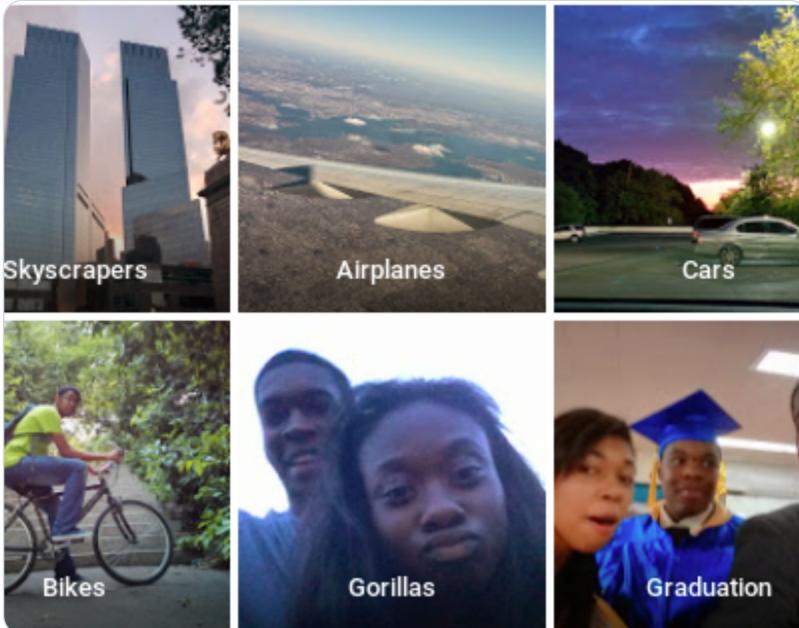
**Enterprise Application AI
Revenue**

Source: Statista

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

Computer Vision Can Cause Harm

Harmful Stereotypes



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

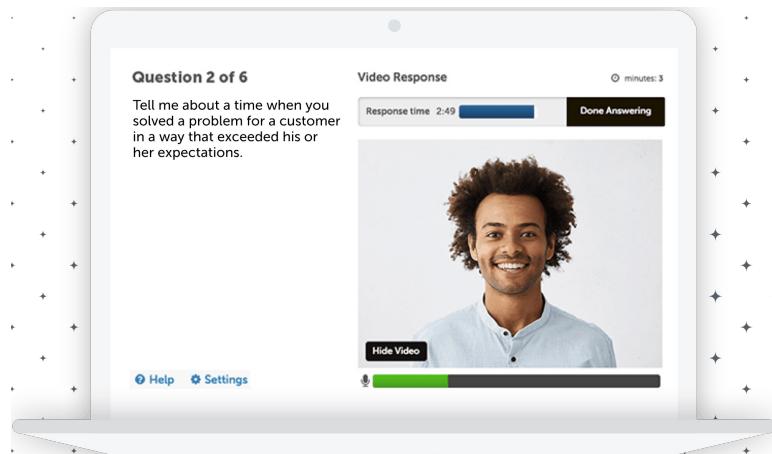
Source: <https://twitter.com/jackyalcine/status/615329515909156865> (2015)

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

**How to take care of seniors
while keeping them safe?**



Early Symptom Detection
of COVID-19



Monitor Patients with
Mild Symptoms



Manage Chronic Conditions

Versatile



Mobility



Infection



Sleep



Diet



Scalable



Low-cost



Burden-free



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

Today's agenda

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