Deep learning for computer vision

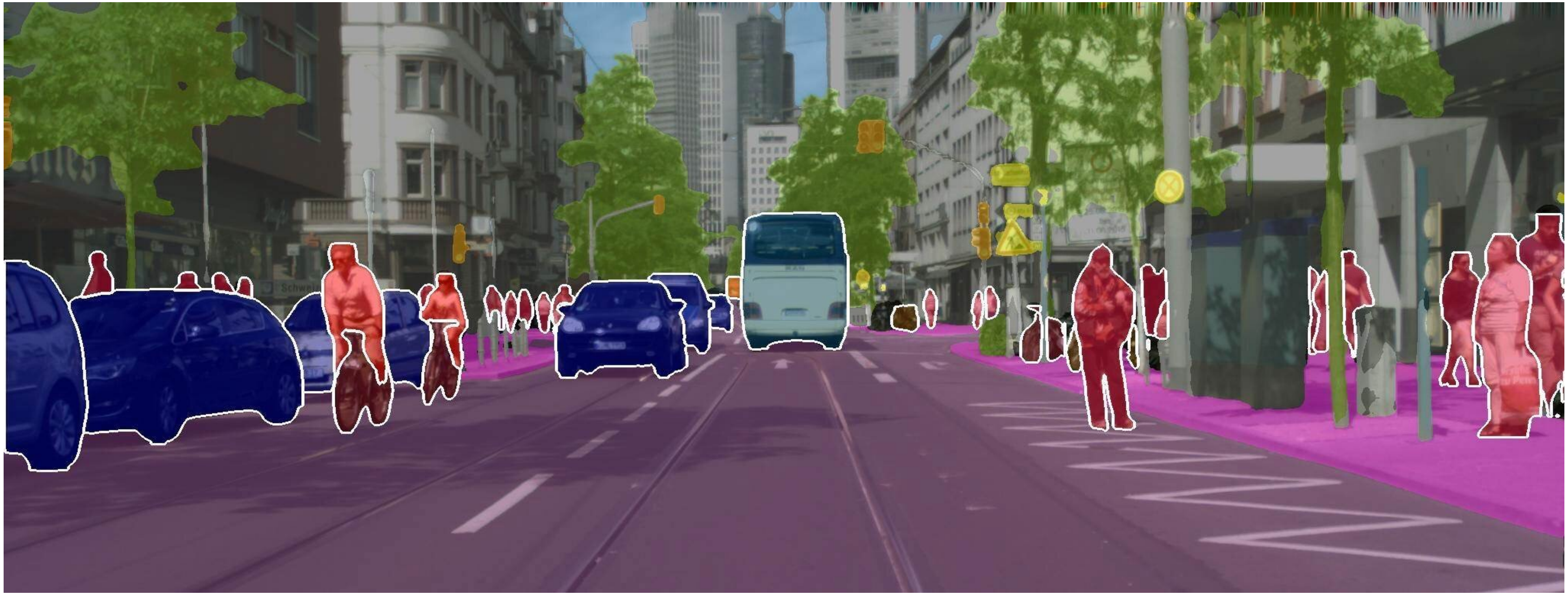


Image credit: by Albert Ludwigs University of Freiburg

Course webpage URL:

opi-lab.github.io/DL4CV/

Disclosure

This course was inspired by the Deep Learning course at University of Illinois at Urbana Champaign by S. Lazebnik and material from the Deep Learning book by I. Goodfellow, Y. Bengio, and A. Courville.

Many of the slides are from the publically available data of these courses.

Recommended Background

Linear Algebra

- Definitions, vectors, matrices, operations, properties

Probability

- Basics: what is a random variable, probability distributions, functions of a random variable

Machine learning*

- Learning, modeling and classification techniques

Grading

Homework assignments : 50%

- Mini projects
- Will be assigned during course

Midterm and final exam: 20%

Final project: 30%

- Will be assigned early in course
- Dec 3 - 7: Oral presentation with demos (if possible) and written paper in IEEE format.

Projects

A solution to a given or proposed problem.

The projects may lead to

- Conference papers
- Journal Papers
- Master/PhD thesis
- etc,.

Additional Information

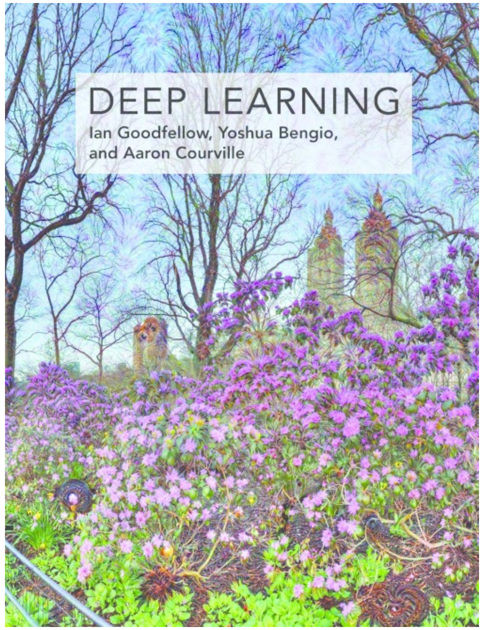
Website:

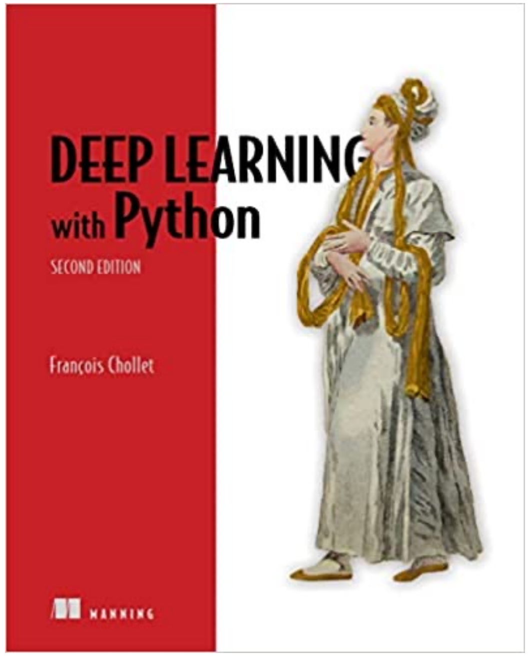
- https://opi-lab.github.io/DL4CV/
- Lecture material will be posted on the day of each class on the website
- Reading material and pointers to additional information will be on the website.

Course on MS Teams.

Additional Information

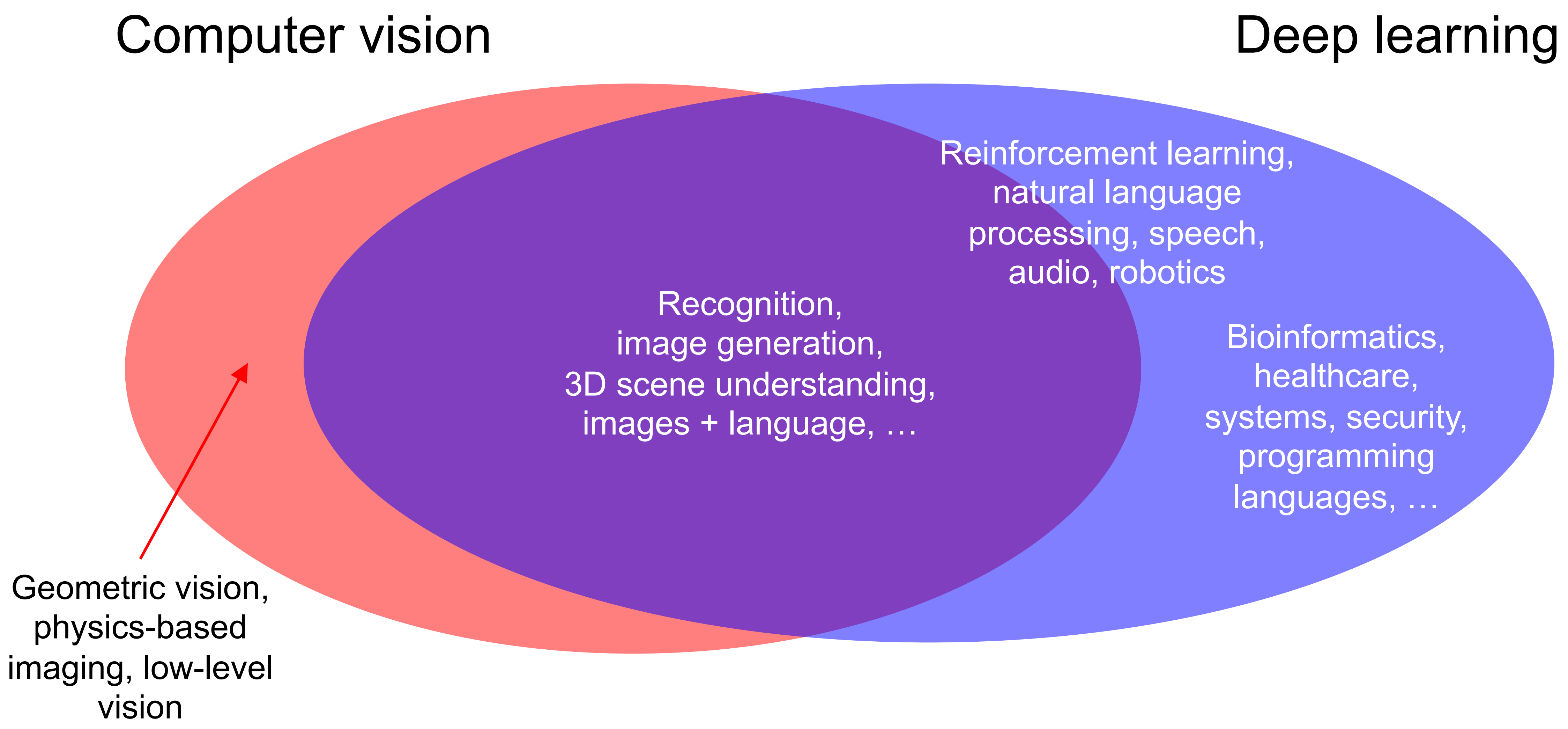
Suggested books





And several others...

CS 444: Deep Learning for Computer Vision

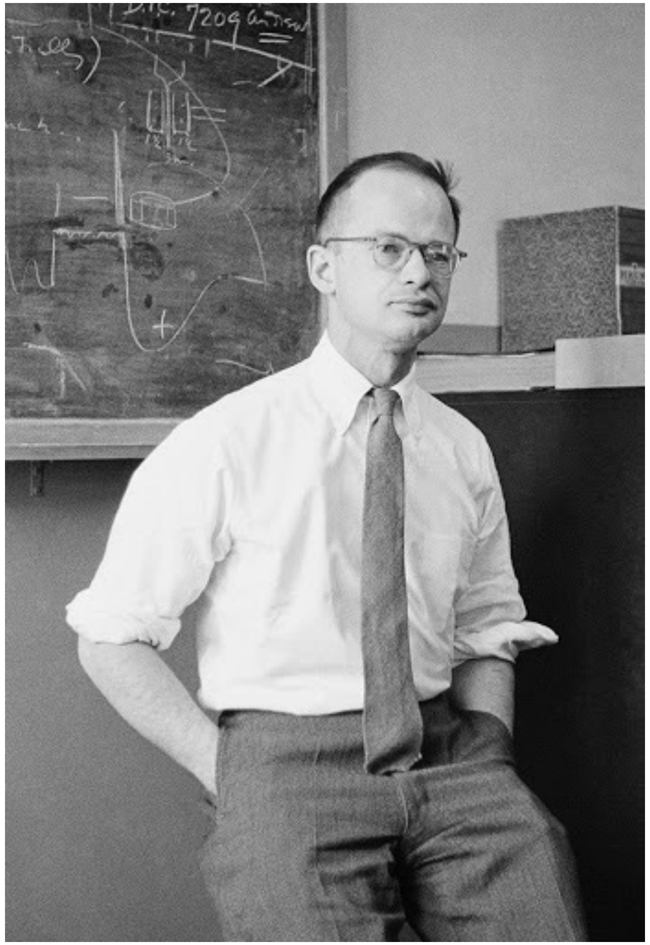


Lecture overview

- About the class
- Milestones of deep learning
- Recent successes and origins
  - Vision
  - Language
  - Games
  - Robotics
- Topics to be covered in class

Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
  - Fascinating reading: The Man Who Tried to Redeem the World with Logic, Nautilus, 2/5/2015



Walter Pitts (1923-1969)

Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron

Frank Rosenblatt (1928-1971)

1958 New York Times…

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

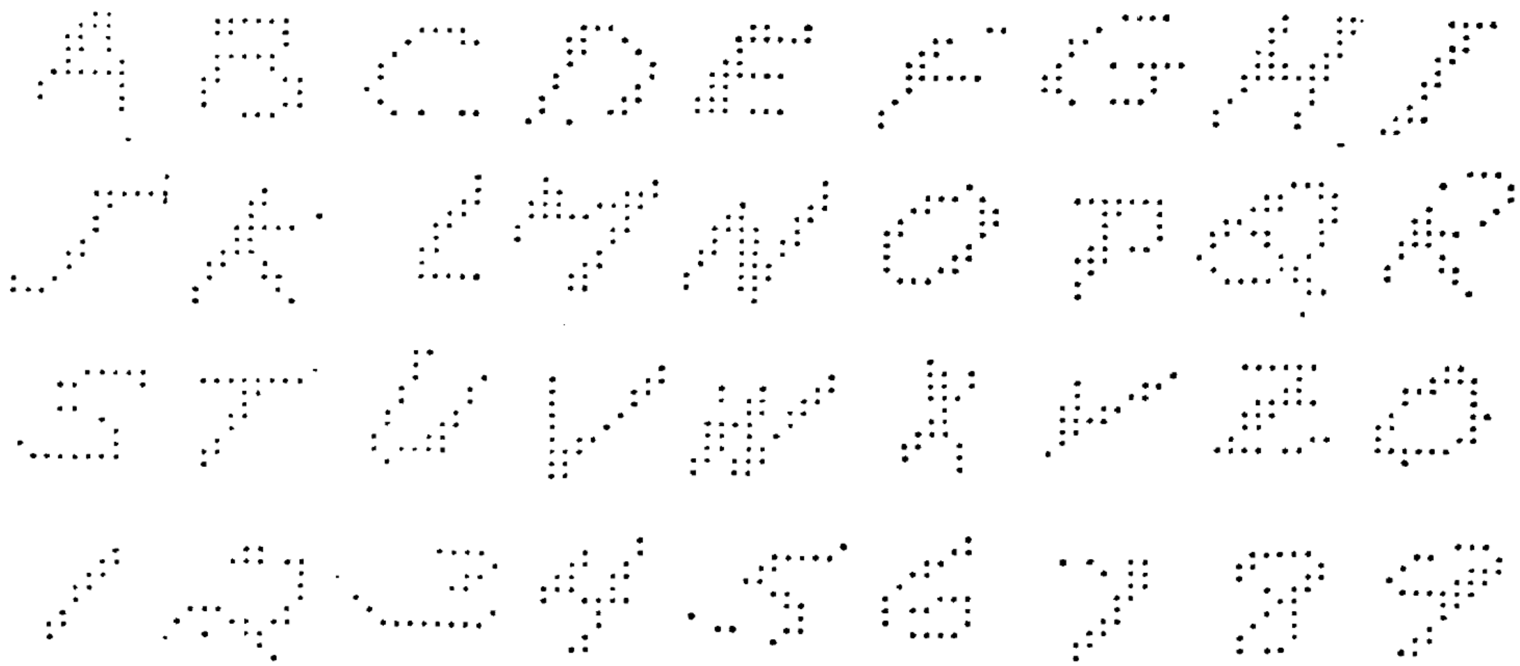
In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron
- 1959: First pattern recognition benchmark, training-test split

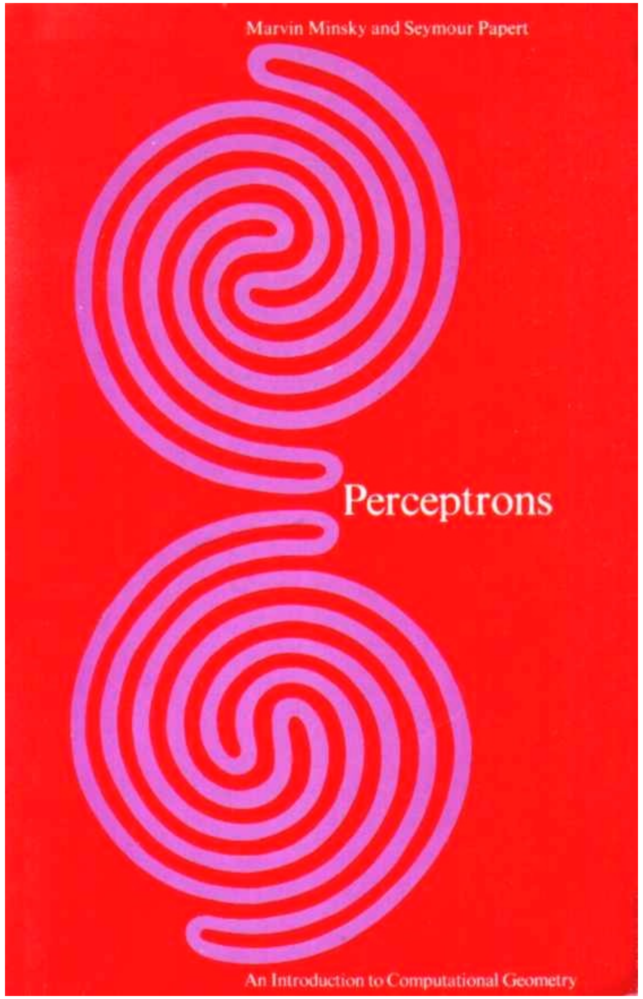


1500 characters (26 letters, 10 digits from 50 writers), 12x12 resolution, stored on IBM 704 punch cards

Bill Highleyman and Louis Kamentsky, Bell Labs

Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron
- 1959: First pattern recognition benchmark
- 1969: Minsky and Papert Perceptrons book
  - Fascinating reading: M. Olazaran, A Sociological Study of the Official History of the Perceptrons Controversy, *Social Studies of Science*, 1996



Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron
- 1959: First pattern recognition benchmark
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima's Neocognitron
  - Video (short version)
  - Inspired by the findings of Hubel & Wiesel
    about the hierarchical organization
    of the visual cortex in cats and monkeys (1959-1977)

Kunihiko Fukushima

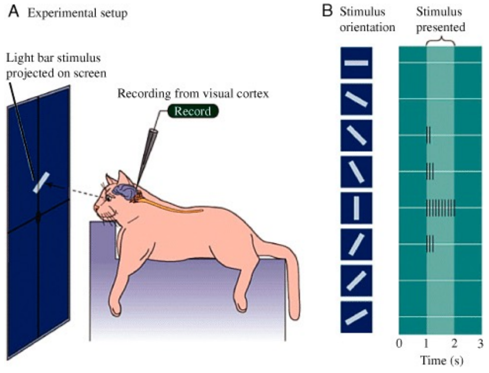
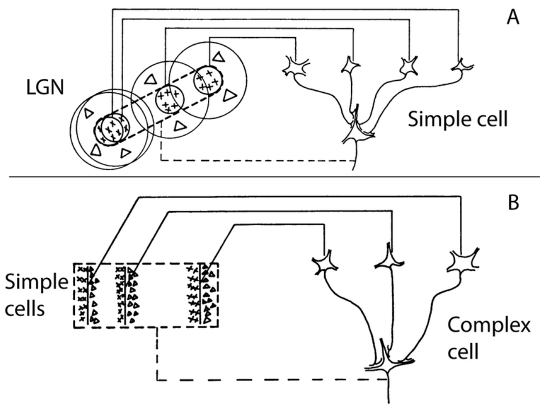
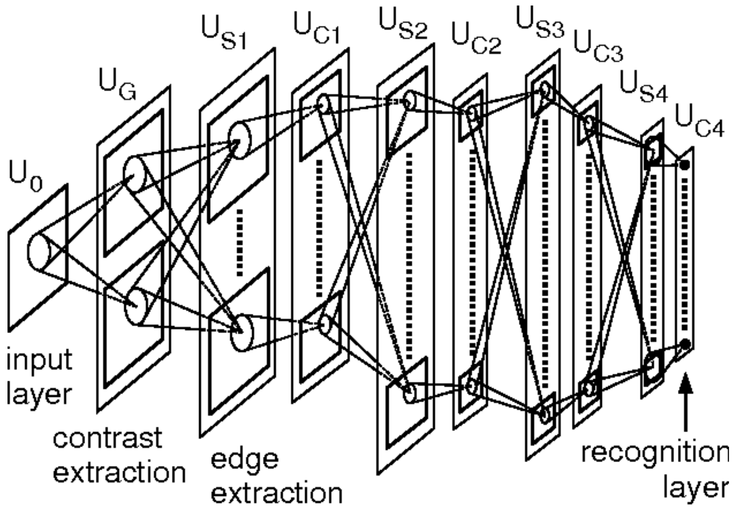


Image source





Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron
- 1959: First pattern recognition benchmark
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima's Neocognitron
- 1986: Back-propagation
  - Origins in control theory and optimization: Kelley (1960), Dreyfus (1962), Bryson & Ho (1969), Linnainmaa (1970)
  - Application to neural networks: Werbos (1974)
  - Popularized by Rumelhart, Hinton & Williams (1986)

Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron
- 1959: First pattern recognition benchmark
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima's Neocognitron
- 1986: Back-propagation
- 1989 – 1998: Convolutional neural networks
    - LeNet to LeNet-5

Yann LeCun

PROC. OF THE IEEE, NOVEMBER 1998 7

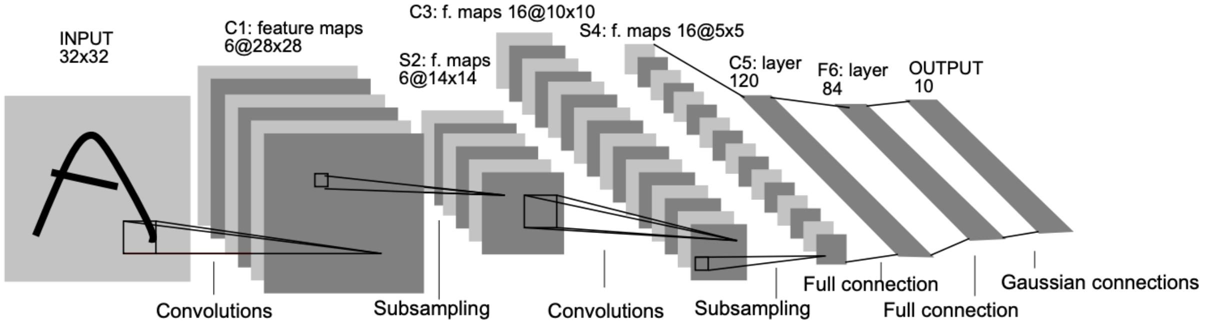


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt’s perceptron
- 1959: First pattern recognition benchmark
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima’s Neocognitron
- 1986: Back-propagation
- 1989 – 1998: Convolutional neural networks
- 2012: AlexNet

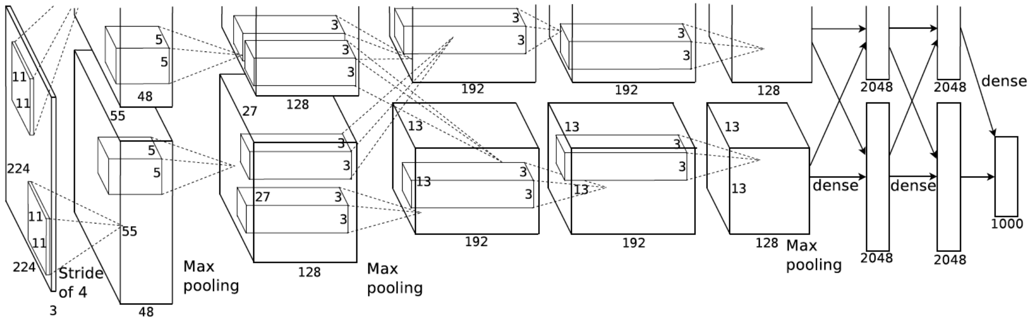


Photo source

Milestones: A ridiculously abbreviated timeline

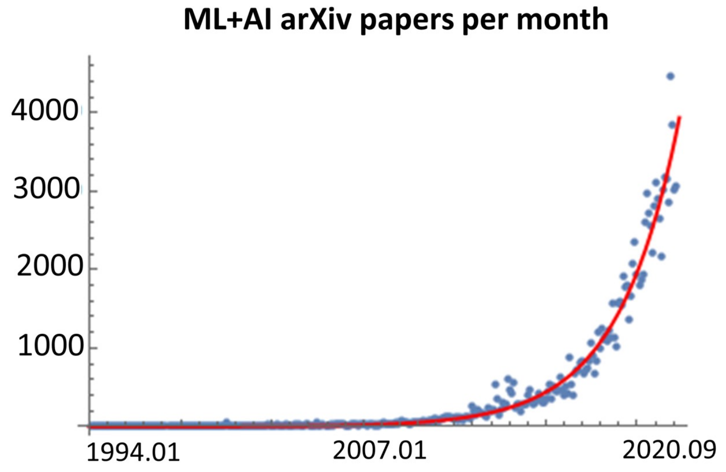
- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron
- 1959: First pattern recognition benchmark
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima's Neocognitron
- 1986: Back-propagation
- 1989 – 1998: Convolutional neural networks
- 2012: AlexNet
  - Fascinating reading: The secret auction that set off the race for AI supremacy, Wired, 3/16/2021

Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron
- 1959: First pattern recognition benchmark
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima's Neocognitron
- 1986: Back-propagation
- 1989 – 1998: Convolutional neural networks
- 2012: AlexNet
- 2018: ACM Turing Award
  to Hinton, LeCun, and Bengio

Milestones: A ridiculously abbreviated timeline

- 1943: McCulloch and Pitts neurons
- 1958: Rosenblatt's perceptron
- 1959: First pattern recognition benchmark
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima's Neocognitron
- 1986: Back-propagation
- 1989 – 1998: Convolutional neural networks
- 2012: AlexNet
- 2012 – present: deep learning explosion



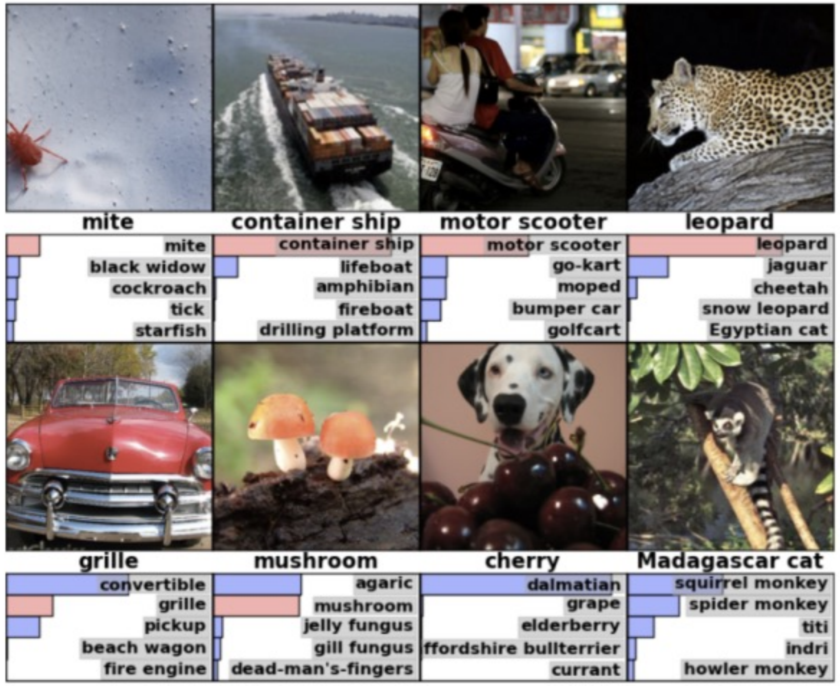
Source, via J. Johnson

Lecture overview

- About the class
- Milestones of deep learning
- Recent successes and origins
    - Vision
    - Language
    - Games
    - Robotics

Successes in vision: ImageNet Challenge

ILSVRC



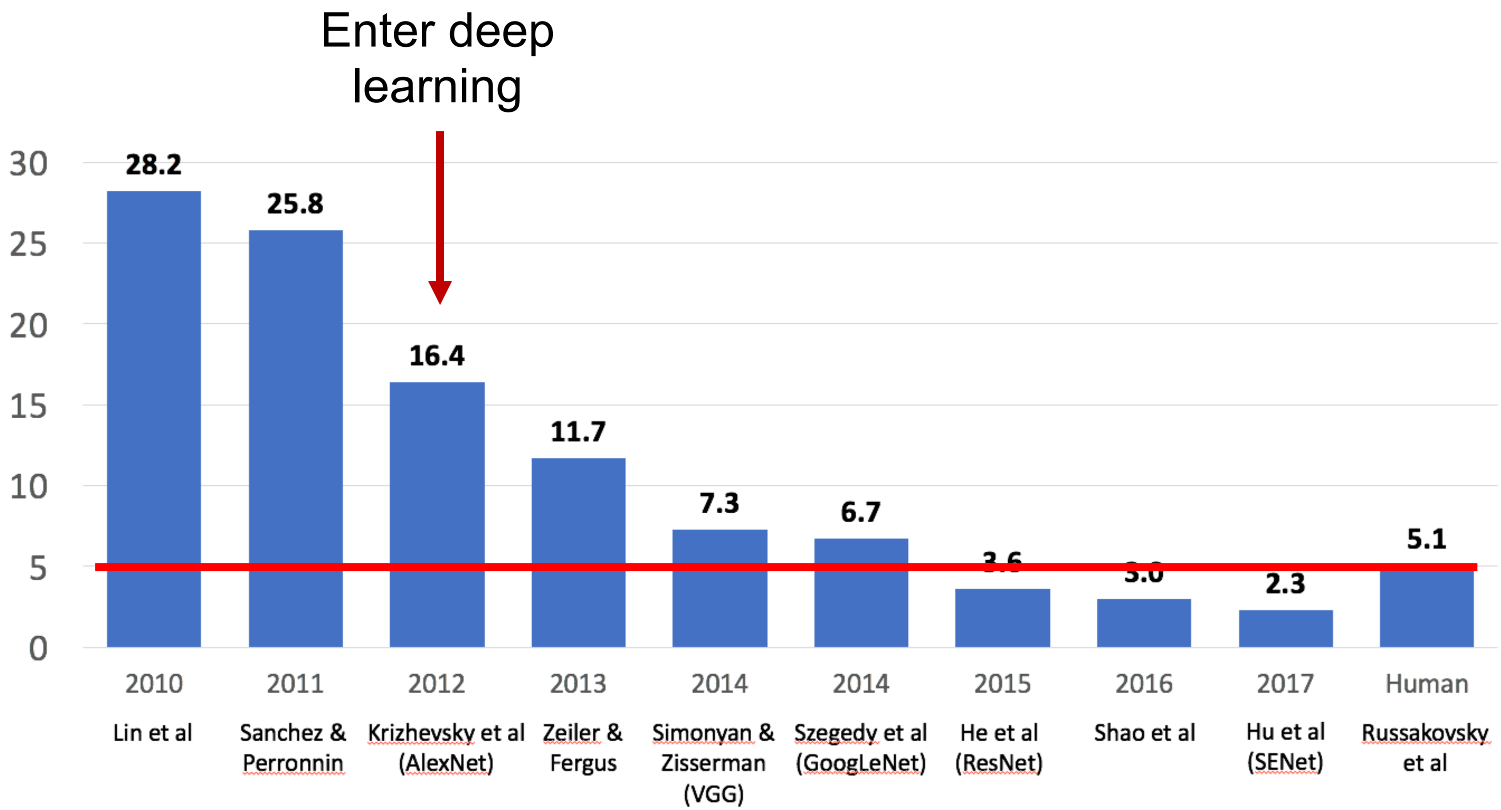


Figure source

Vision: Outgrowing ImageNet

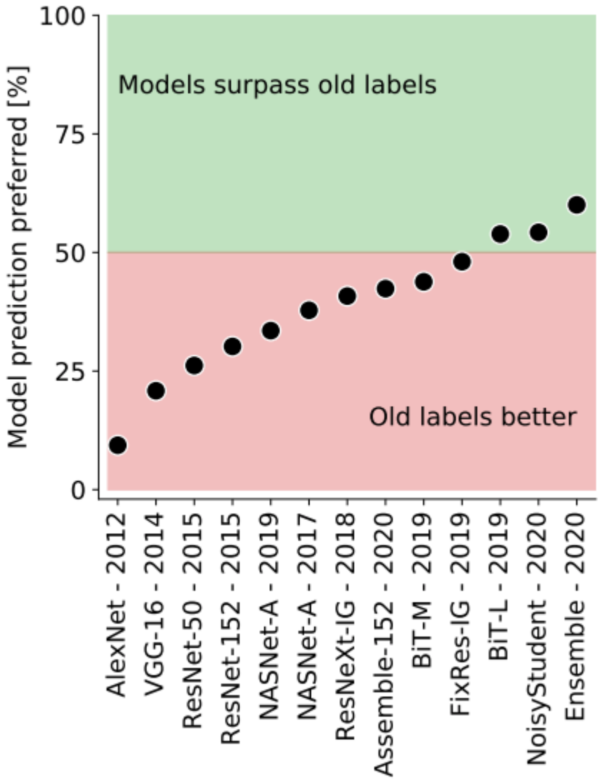


Figure 1: When presented with a model's prediction and the original ImageNet label, human annotators now prefer model predictions on average (Section 4). Nevertheless, there remains considerable progress to be made before fully capturing human preferences.

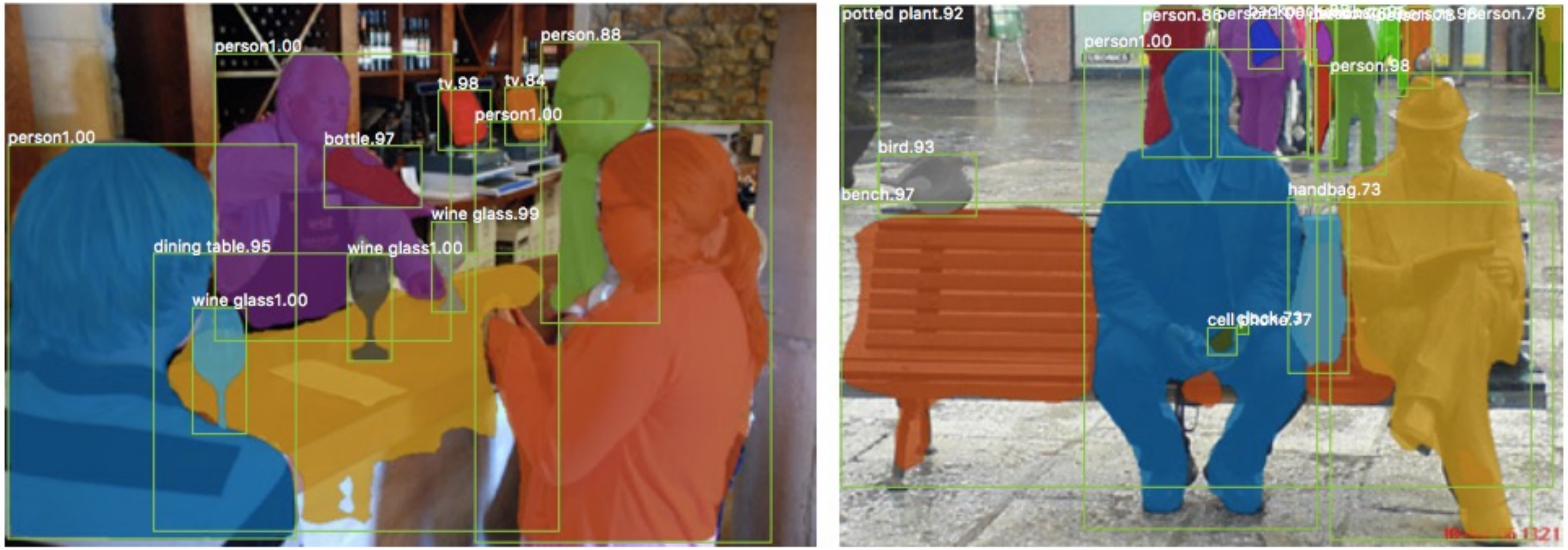
Unsafe (offensive)	Unsafe (sensitive)	Safe non-imageable	**Safe imageable**
n10095420: <sexual slur>	n09702134: Anglo-Saxon	n10002257: demographer	n10499631: Queen of England
n10114550: <profanity>	n10693334: taxi dancer	n10061882: epidemiologist	n09842047: basketball player
n10262343: <sexual slur>	n10384392: orphan	n10431122: piano maker	n10147935: bridegroom
n10758337: <gendered slur>	n09890192: camp follower	n10098862: folk dancer	n09846755: beekeeper
n10507380: <criminative>	n10580030: separatist	n10335931: mover	n10153594: gymnast
n10744078: <criminative>	n09980805: crossover voter	n10449664: policyholder	n10539015: ropewalker
n10113869: <obscene>	n09848110: theist	n10146104: great-niece	n10530150: rider
n10344121: <pejorative>	n09683924: Zen Buddhist	n10747119: vegetarian	n10732010: trumpeter

"Programmer"

L. Beyer et al. [Are we done with ImageNet?](https://arxiv.org/abs/2006.07159) arXiv:2006.07159, 2020

K. Yang, K. Qinami, L. Fei-Fei, J. Deng, O. Russakovsky, Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy, Conference on Fairness, Accountability, and Transparency (FAccT), 2020

Vision: Detection, segmentation



K. He, G. Gkioxari, P. Dollar, and R. Girshick, Mask R-CNN,
ICCV 2017 (Best Paper Award)

Vision: Image generation

Ian Goodfellow
@goodfellow_ian

4.5 years of GAN progress on face generation.
arxiv.org/abs/1406.2661 arxiv.org/abs/1511.06434
arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196
arxiv.org/abs/1812.04948

6:40 PM · Jan 14, 2019

Vision: Image generation

- Faces: 1024x1024 resolution, CelebA-HQ dataset

T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, ICLR 2018

Follow-up work

Vision: Image generation

GAN-generated dogs in 2017

[Source: EBGAN]

GAN-generated dogs in 2018

[Source: BigGAN]

Vision: Image generation

- BigGAN: Synthesize ImageNet images, conditioned on class label, up to 512 x 512 resolution

Difficult classes

A. Brock, J. Donahue, K. Simonyan, [Large scale GAN training for high fidelity natural image synthesis](), ICLR 2019

Vision working too well? Face recognition

How China Uses High-Tech Surveillance to Subdue Minorities – New York Times, 5/22/2019

The Secretive Company That Might End Privacy As We Know It – New York Times, 1/18/2020

Wrongfully Accused by an Algorithm – New York Times, 6/24/2020

Vision working too well? DeepFakes

IndieWire News + Film + TV + Awards + Video +

Lucasfilm Hired the YouTuber Who Used Deepfakes to Tweak Luke Skywalker 'Mandalorian' VFX

A YouTuber known as Shamook has earned nearly 2 million views for his deepfake "Mandalorian" video.

Zack Sharf
Jul 26, 2021 6:00 pm
@zsharf

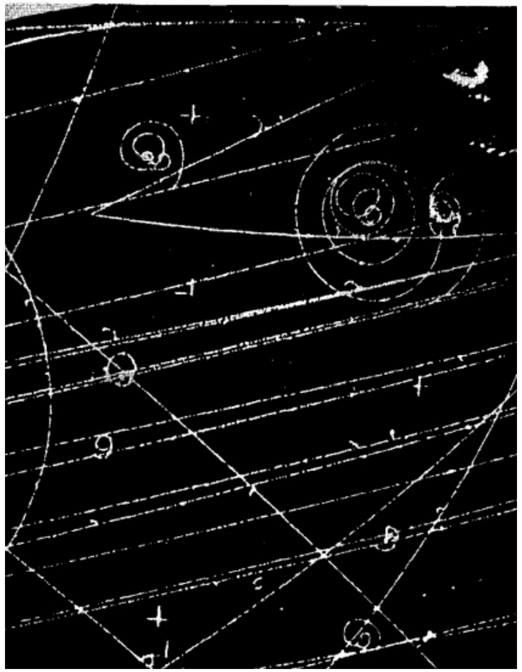
Just a random recent example…

https://www.indiewire.com/2021/07/lucasfilm-hires-deepfake-youtuber-mandalorian-skywalker-vfx-1234653720/

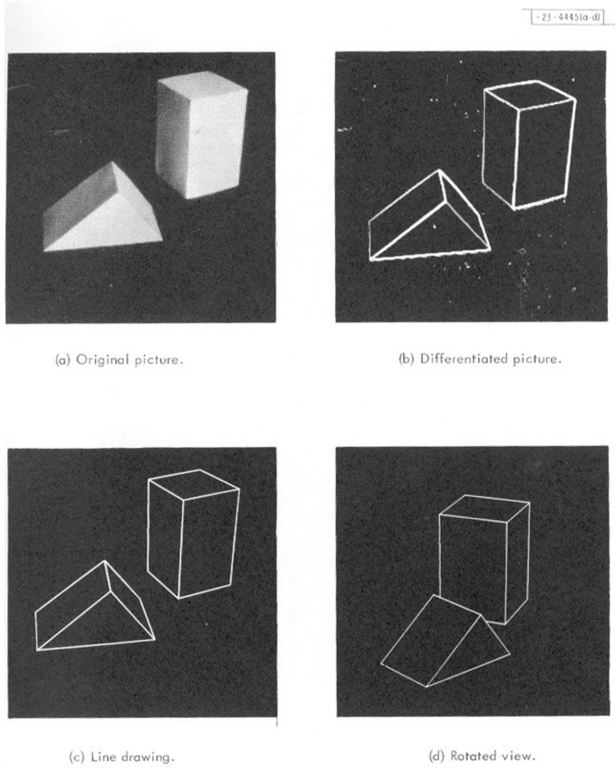
https://youtu.be/wrHXA2cSpNU

https://en.wikipedia.org/wiki/Deepfake

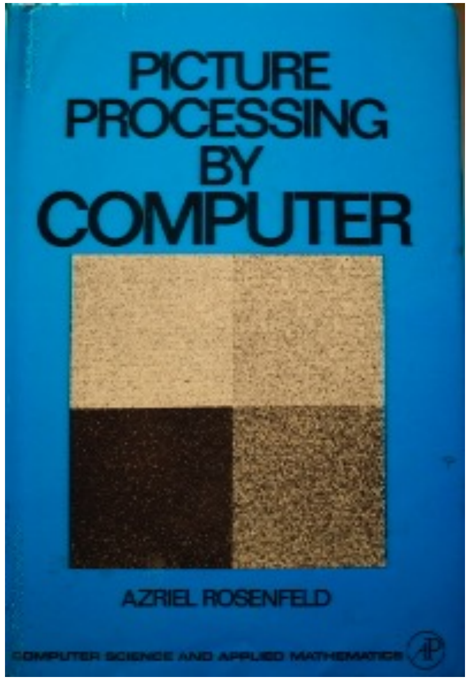
Vision: Origins



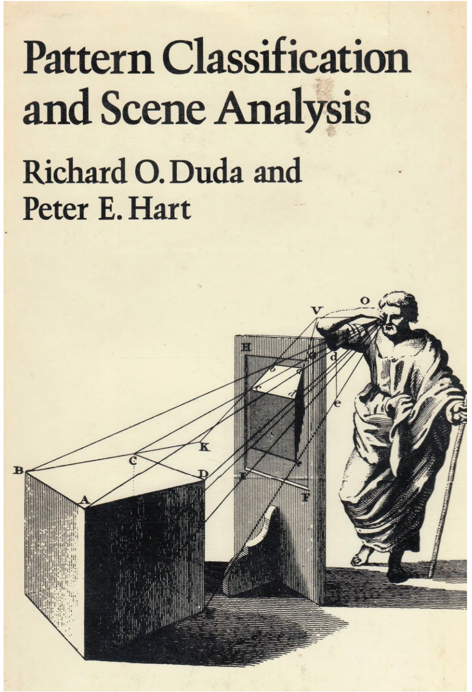
[Hough, 1959]



[Roberts, 1963]



Rosenfeld, 1969



Duda & Hart, 1972

Successes in natural language

- Neural machine translation
  - The Great AI Awakening – New York Times Magazine, 12/14/2016
- Language models: e.g., GPT-3

MIT Technology Review

Artificial intelligence / Machine learning

OpenAI's new language generator GPT-3 is shockingly good—and completely mindless

The AI is the largest language model ever created and can generate amazing human-like text on demand but won't bring us closer to true intelligence.

https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/

MIT Technology Review

Opinion

GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about

Tests show that the popular AI still has a poor grasp of reality.

by **Gary Marcus** and **Ernest Davis** August 22, 2020

https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/

Successes in natural language

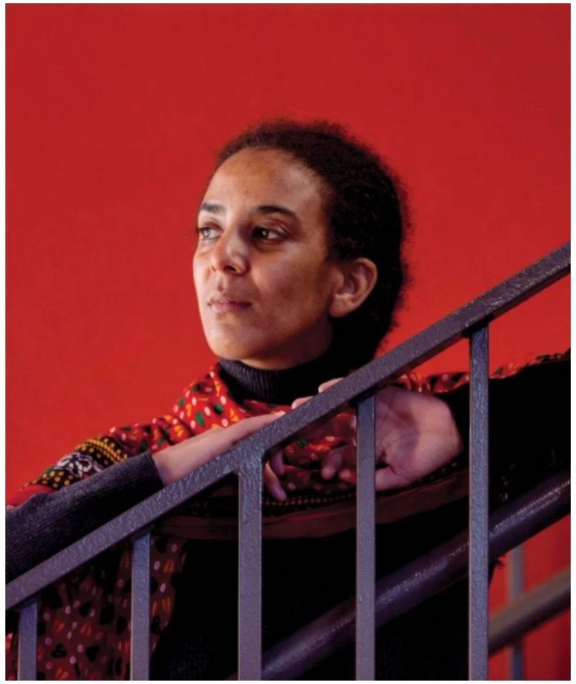
- Neural machine translation
  - The Great AI Awakening – New York Times Magazine, 12/14/2016
- Language models: e.g., GPT-3
  - M. Bender et al. On the dangers of stochastic parrots: Can language models be too big? FAccT 2021

Artificial intelligence / Machine learning

We read the paper that forced Timnit Gebru out of Google. Here's what it says.

The company's star ethics researcher highlighted the risks of large language models, which are key to Google's business.

by **Karen Hao** December 4, 2020



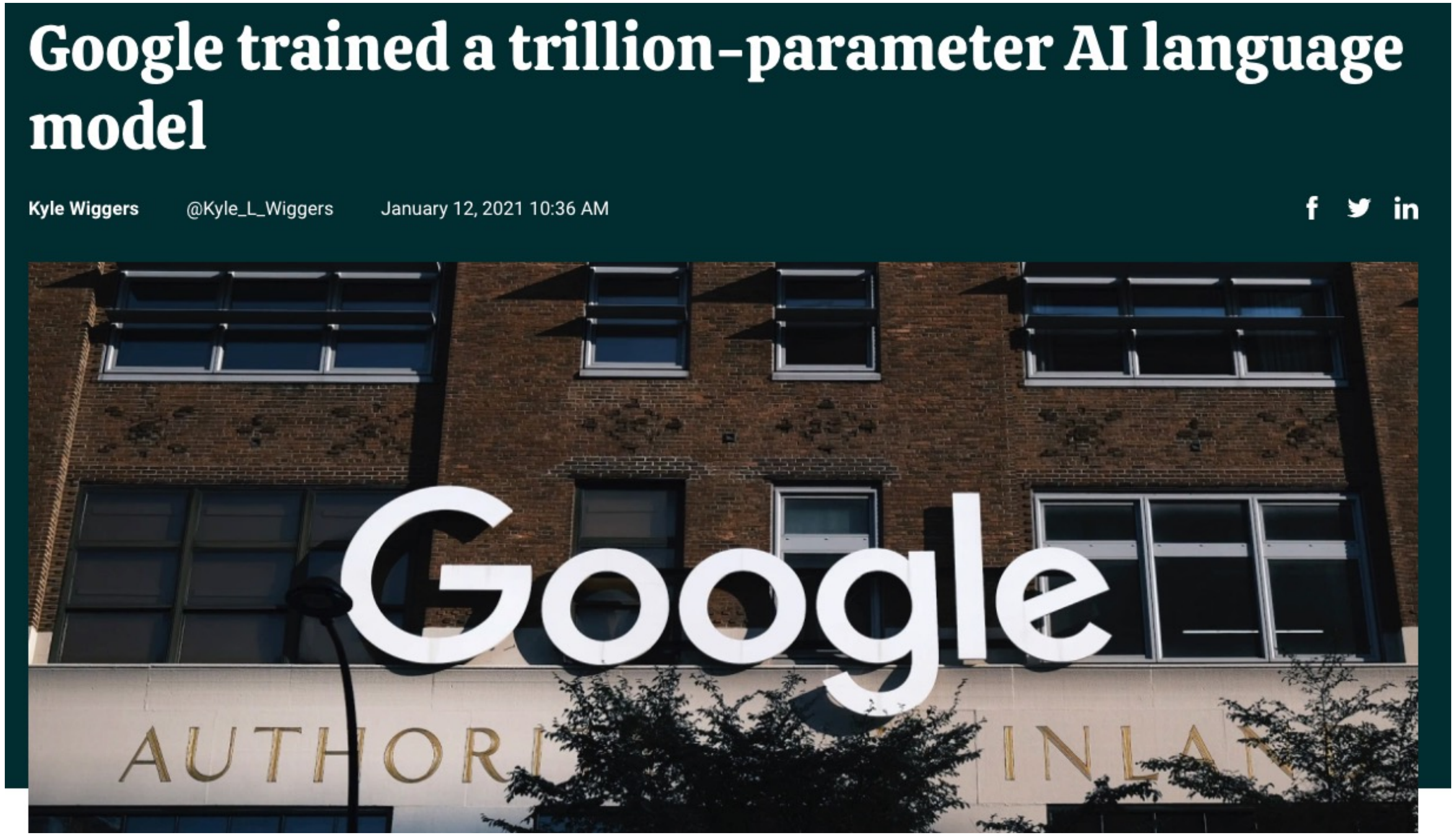
COURTESY OF TIMNIT GEBRU

MIT Technology Review

https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

Successes in natural language

- Neural machine translation
  - The Great AI Awakening – New York Times Magazine, 12/14/2016
- Language models: e.g., GPT-3



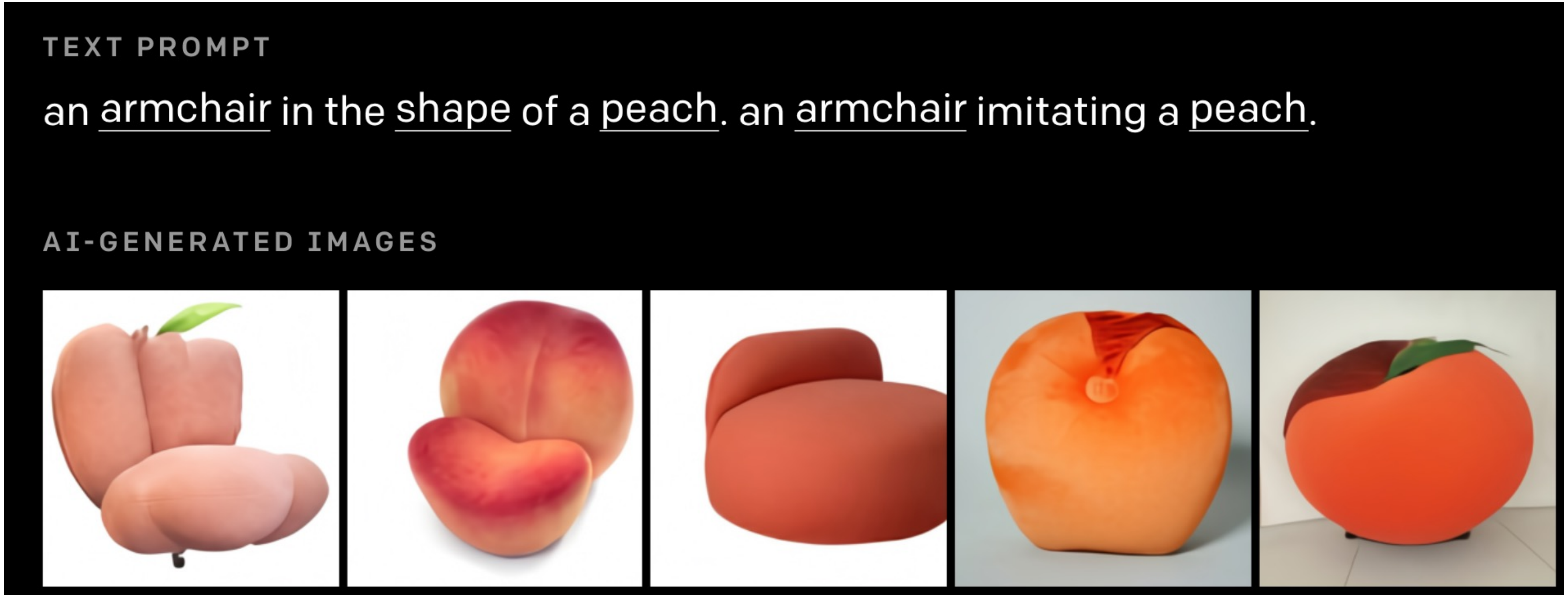
https://venturebeat.com/2021/01/12/google-trained-a-trillion-parameter-ai-language-model/

Successes in natural language

- Neural machine translation
  - The Great AI Awakening – New York Times Magazine, 12/14/2016
- Language models: e.g., GPT-3
- Vision and language models: DALL-E, CLIP

Successes in natural language

- Neural machine translation
    - The Great AI Awakening – New York Times Magazine, 12/14/2016
- Language models: e.g., GPT-3
- Vision and language models: DALL-E, CLIP



Natural language: Origins

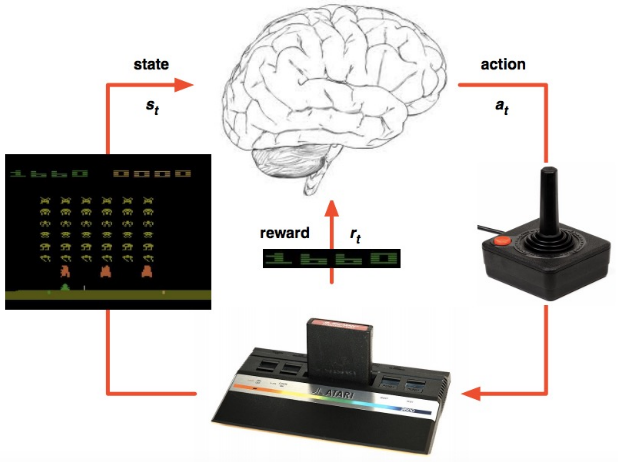
- Turing test (1950)
- Machine translation
  - 1954: Georgetown-IBM experiment
    - Completely automatic translation of more than sixty Russian sentences into English
    - Only six grammar rules, 250 vocabulary words, restricted to organic chemistry
    - Promised that machine translation would be solved in three to five years (press release)
  - 1966: Automatic Language Processing Advisory Committee (ALPAC) report: machine translation is not living up to the hype

Sentences in Russian are punched into standard cards for feeding into the electronic data processing machine for translation into English

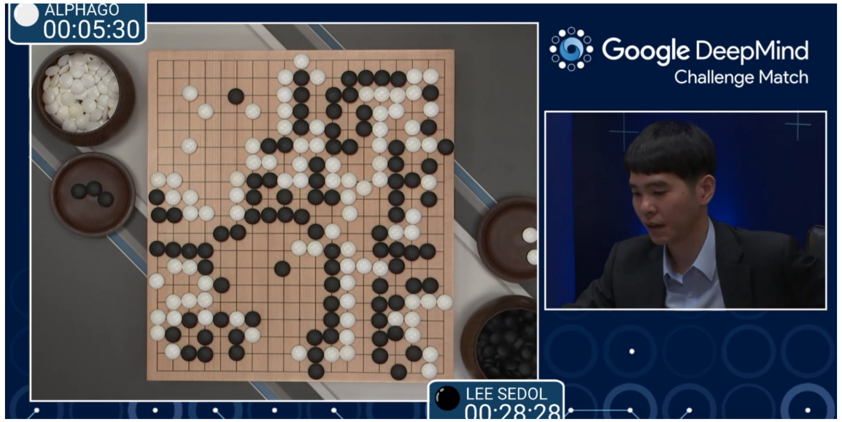
- Chatbots: ELIZA (1966)
  - Simulated a psychotherapist, could fool naïve users

Successes in games

- 2013: DeepMind uses deep reinforcement learning to beat humans at some Atari games



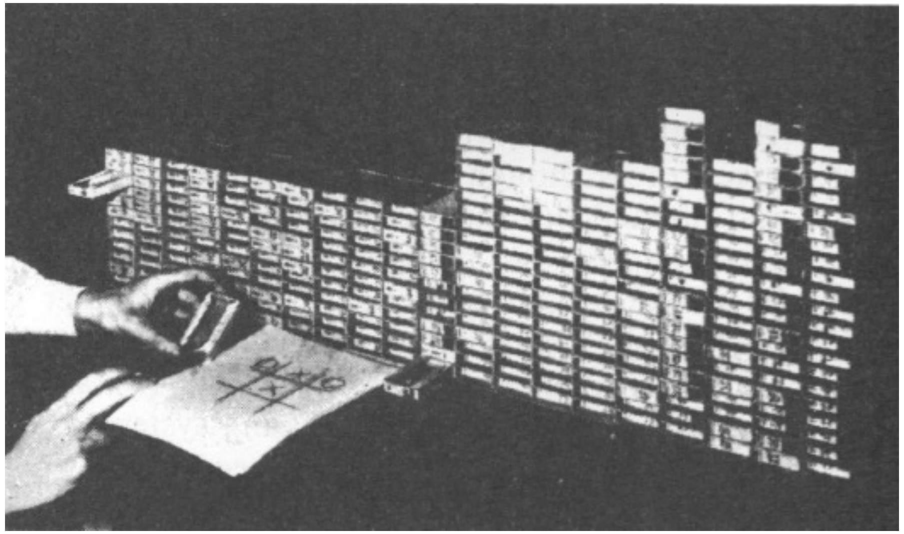
- 2016: DeepMind's AlphaGo system beats Go grandmaster Lee Sedol 4-1
- 2017: AlphaZero learns to play Go and chess from scratch
- 2019: DeepMind's StarCraft 2 AI is better than 99.8 percent of all human players



Games: Origins

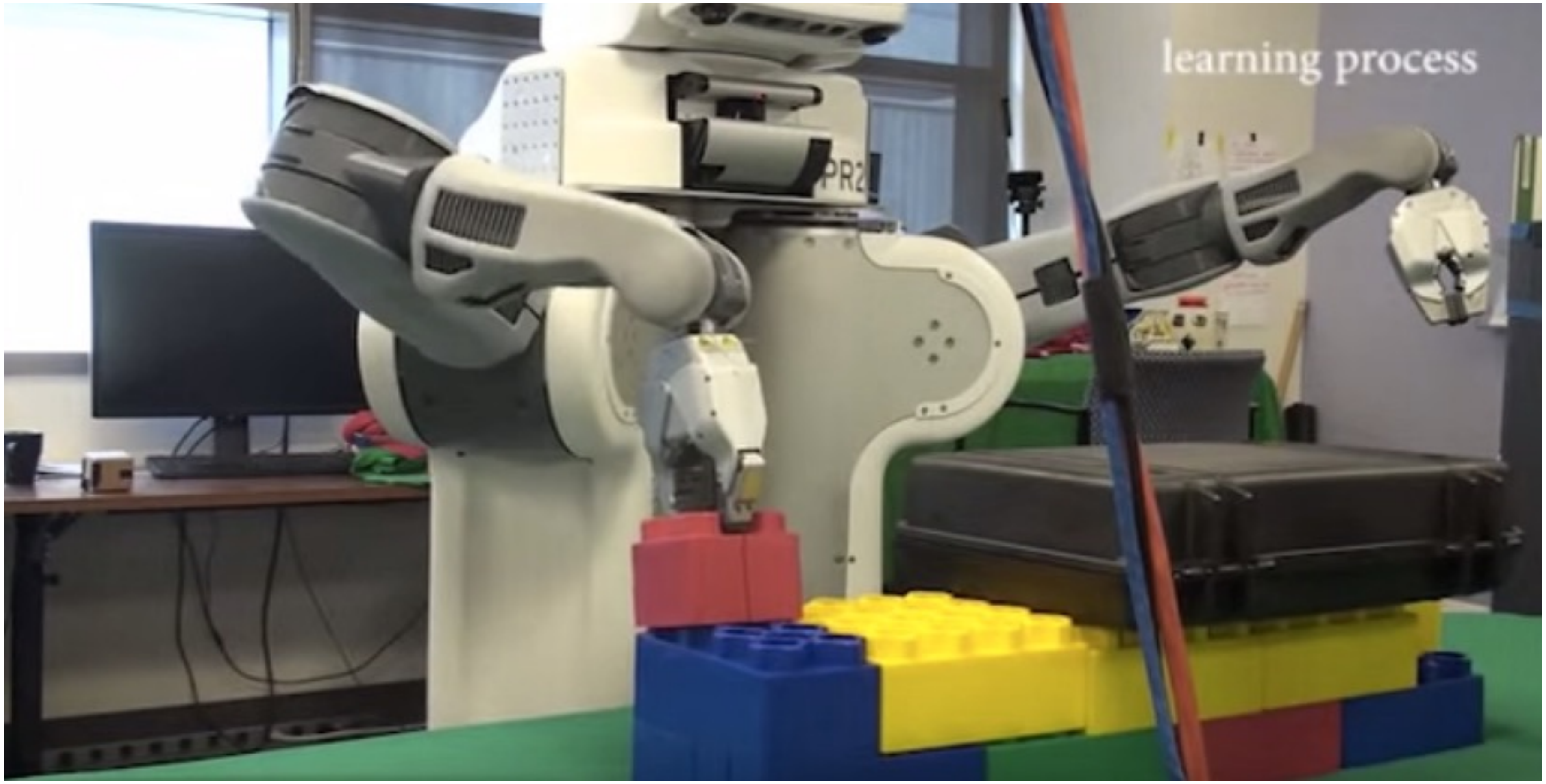
- 1952-1959: Arthur Samuel programmed a digital computer to learn to play checkers
  - “In 1959 Arthur Samuel published a paper titled ‘Some Studies in Machine Learning Using the Game of Checkers’, the first time the phrase ‘Machine Learning’ was used” – Rodney Brooks

- 1960: Donald Michie built a “machine” out of 304 matchboxes that could learn to play tic-tac-toe
  - “Donald Michie’s description of reinforcement learning comes from 1961, and is the first use of the term reinforcement learning when applied to a machine process … There have been some developments in reinforcement learning since 1961, but only in details” – Rodney Brooks



Successes in embodied vision and robotics

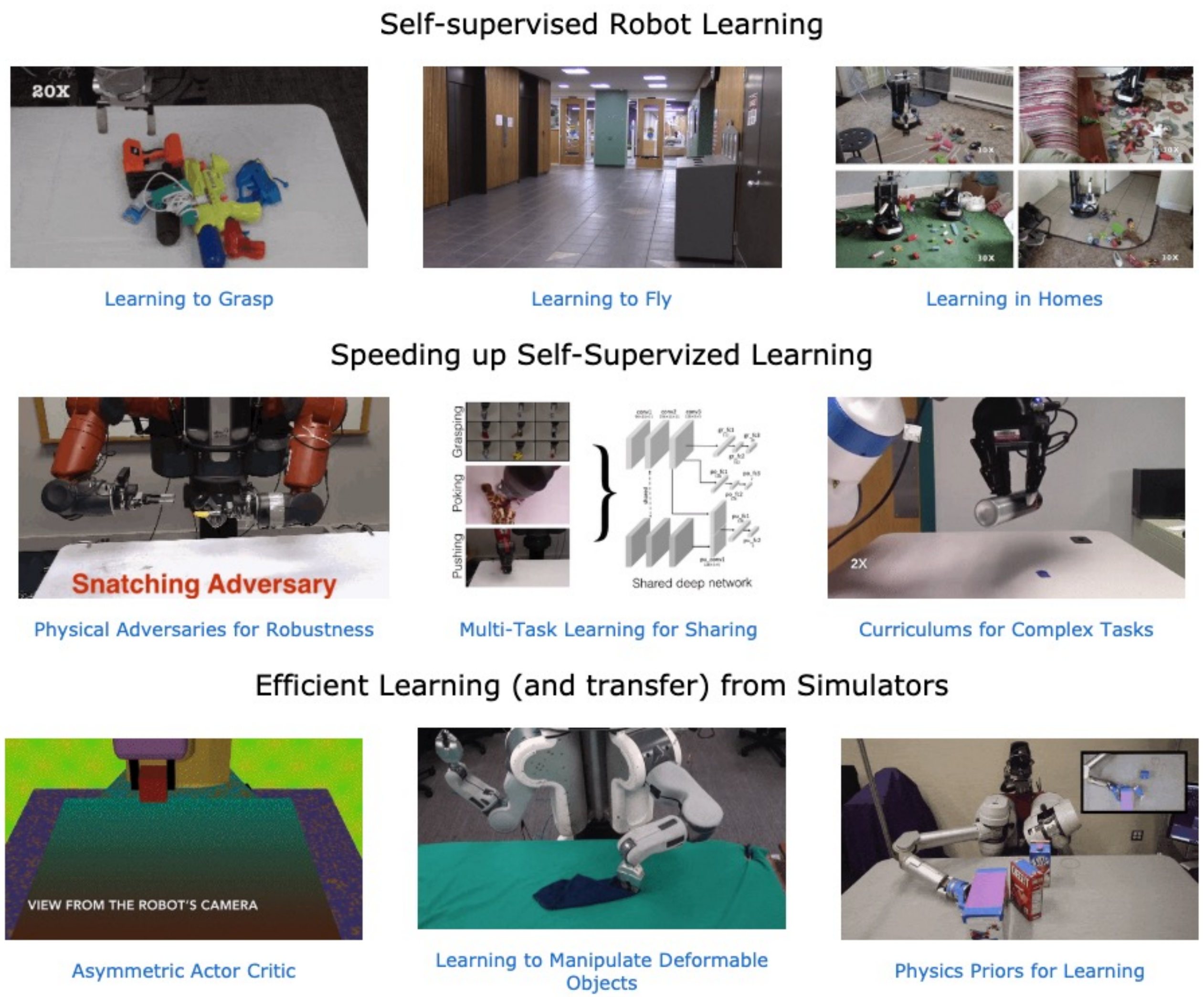
- Sensorimotor learning



[Overview video](), [training video]()

S. Levine, C. Finn, T. Darrell, P. Abbeel, [End-to-end training of deep visuomotor policies](), JMLR 2016

Embodied vision and robotics

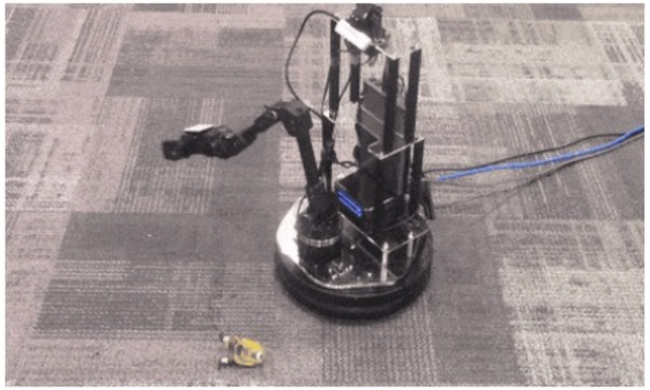


A cross-section of topics from the webpage of Lerrel Pinto

- Other representative researchers: Abhinav Gupta, Pieter Abbeel, Sergey Levine, Chelsea Finn

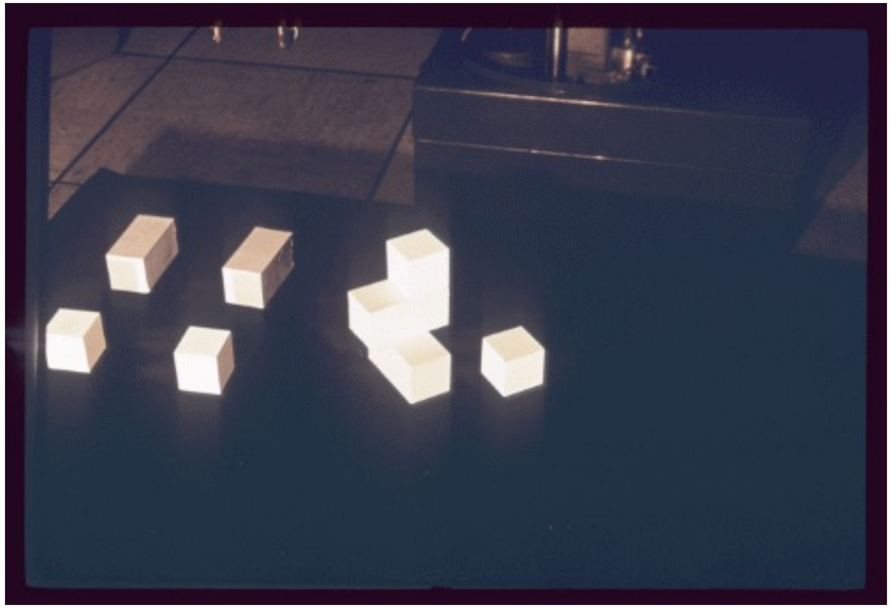
Embodied platforms

- Simulation: AI2Thor, Habitat
- Real robots: PyRobot

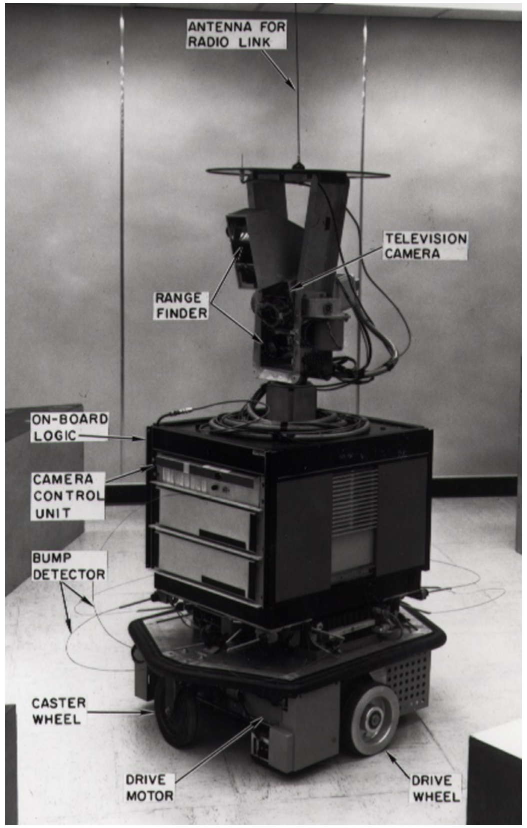


- Robot on your smartphone: OpenBot

Robotics: Origins



Blocks World
MIT, 1960s – 1970s
Copy demo (1970)



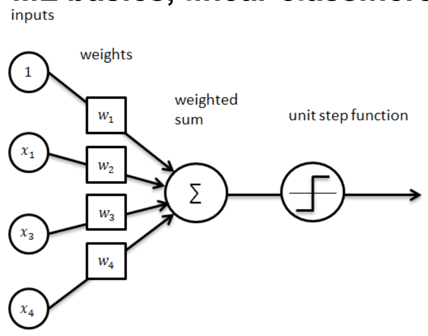
Shakey the Robot
SRI, 1966 – 1972
Video

Lecture overview

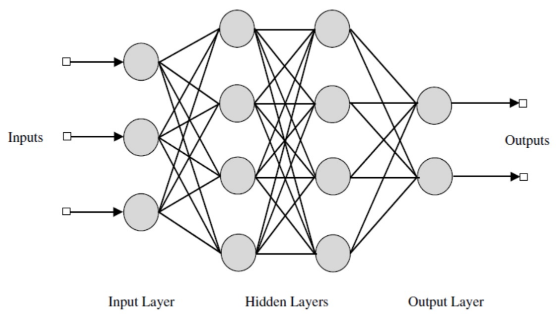
- About the class
- Milestones of deep learning
- Recent successes and origins
  - Vision
  - Language
  - Games
  - Robotics
- Topics to be covered in class

Topics to be covered in class

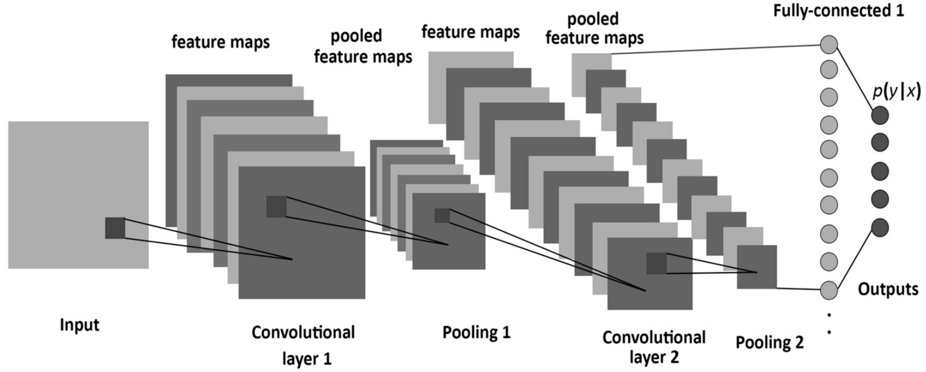
ML basics, linear classifiers



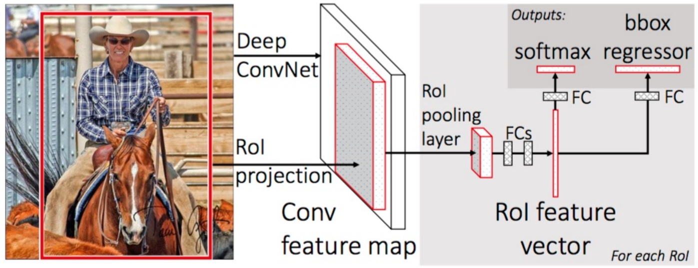
Multilayer neural networks, backpropagation



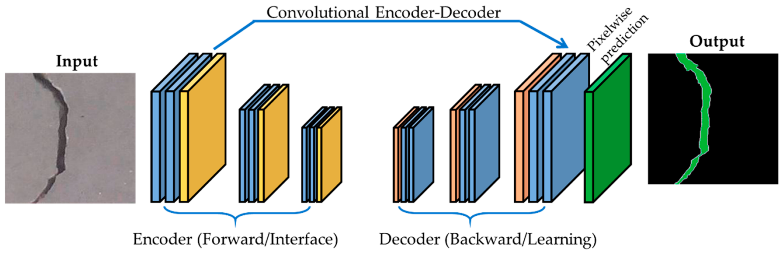
Convolutional networks for classification



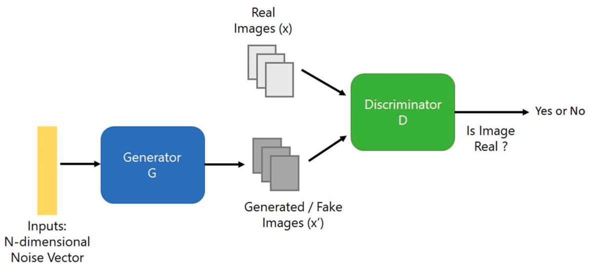
Networks for detection



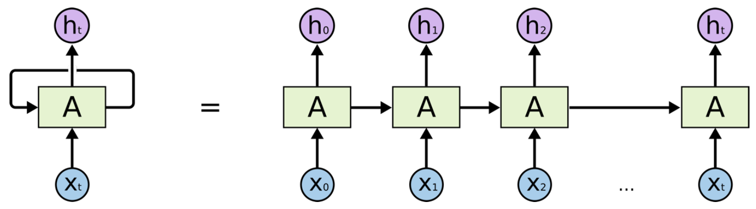
Networks for dense prediction



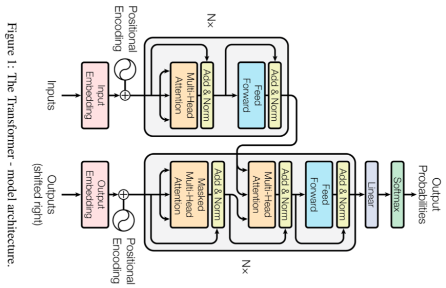
Generative models (GANs, VAEs)



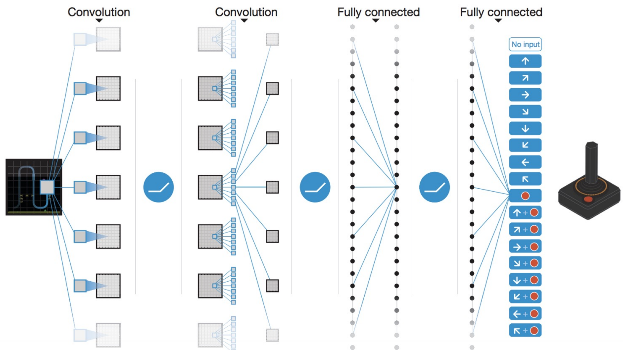
Recurrent models



Transformers



Deep reinforcement learning



Time permitting