CS 444: Deep Learning for Computer Vision

D. Hockney, Pool with two figures, 1972

https://slazebni.cs.illinois.edu/spring23/
Lecture overview

- About the class
- Milestones of deep learning
- Recent successes and origins
  - Visual recognition
  - Natural language understanding
  - Generative modeling
  - Games
  - Robotics
- Topics to be covered in class
A few historical milestones

- 1958: Rosenblatt’s perceptron

NEW NAVY DEVICE
LEARN 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 newspeople.

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.

LEARN 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.
A few historical milestones

• 1958: Rosenblatt’s perceptron
• 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
A few historical milestones

• 1958: Rosenblatt’s perceptron
• 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)
A few historical milestones

- 1958: Rosenblatt’s perceptron
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima’s Neocognitron
- 1986: Back-propagation
  - Popularized by Rumelhart, Hinton & Williams (1986)
A few historical milestones

- 1958: Rosenblatt’s perceptron
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima’s Neocognitron
- 1986: Back-propagation
  - LeNet to LeNet-5

---

**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.
A few historical milestones

- 1958: Rosenblatt’s perceptron
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima’s Neocognitron
- 1986: Back-propagation
- 2012: AlexNet

Photo source
A few historical milestones

• 1958: Rosenblatt’s perceptron
• 1969: Minsky and Papert Perceptrons book
• 1980: Fukushima’s Neocognitron
• 1986: Back-propagation
• 2012: AlexNet
  • Fascinating reading: The secret auction that set off the race for AI supremacy, Wired, 3/16/2021
A few historical milestones

- 1958: Rosenblatt’s perceptron
- 1969: Minsky and Papert Perceptrons book
- 1980: Fukushima’s Neocognitron
- 1986: Back-propagation
- 2012: AlexNet
- 2012 – present: deep learning explosion

Source, via J. Johnson
Lecture overview

• About the class
• Milestones of deep learning
• Progress in the last decade
  • Visual recognition
  • Natural language understanding
  • Generative modeling
  • Games
  • Robotics
Recognition: ImageNet Challenge

**ILSVRC**

Before deep learning

<table>
<thead>
<tr>
<th>Year</th>
<th>Method</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>Lin et al</td>
<td>28.2</td>
</tr>
<tr>
<td>2011</td>
<td>Sanchez &amp; Perronnin</td>
<td>25.8</td>
</tr>
</tbody>
</table>

Convolutional architectures

<table>
<thead>
<tr>
<th>Year</th>
<th>Method</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>Krizhevsky et al (AlexNet)</td>
<td>16.4</td>
</tr>
<tr>
<td>2013</td>
<td>Zeiler &amp; Fergus</td>
<td>11.7</td>
</tr>
<tr>
<td>2014</td>
<td>Simonyan &amp; Zisserman (VGG)</td>
<td>7.3</td>
</tr>
<tr>
<td>2014</td>
<td>Szegedy et al (GoogleNet)</td>
<td>6.7</td>
</tr>
<tr>
<td>2015</td>
<td>He et al (ResNet)</td>
<td>3.6</td>
</tr>
<tr>
<td>2016</td>
<td>Shao et al</td>
<td>3.0</td>
</tr>
<tr>
<td>2017</td>
<td>Hu et al (SENet)</td>
<td>2.3</td>
</tr>
<tr>
<td>Human</td>
<td></td>
<td>5.1</td>
</tr>
</tbody>
</table>

**Figure source**
ImageNet is obsolete?

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.

Object instance segmentation

K. He, G. Gkioxari, P. Dollar, and R. Girshick, \textit{Mask R-CNN}, ICCV 2017 (Best Paper Award)
Recognition on my iPhone

**Siri Knowledge**

**Iris**
- Iris is a flowering plant genus of 310 accepted species with showy flowers. As well as being the scientific name, it is often used as a verb for the act of being in love.
- [Wikipedia](https://en.wikipedia.org/wiki/Iris)

**Iris versicolor**
- Iris versicolor, the blue flag, is a perennial herb with blue flowers. Its scientific name refers to the Latin word for "blue flag".
- [Wikipedia](https://en.wikipedia.org/wiki/Iris_versicolor)

**Sphinx**
- Sphinx is a genus of insects belonging to the family Sphingidae. They are known for their long proboscises and are often mistaken for bees.
- [Wikipedia](https://en.wikipedia.org/wiki/Sphinx)

**Mydas**
- Mydas is a genus of large flies in the family Mydidae. They are characterized by their distinctive proboscides.
- [Wikipedia](https://en.wikipedia.org/wiki/Mydas)

**Self-Portrait**
- Painting by Jan Lievens (Leiden JL-105)
- [Wikipedia](https://en.wikipedia.org/wiki/Jan_Lievens)
Recognition on my iPhone

Is this Daria?

Yes

No
Recognition: Concerns

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
Lecture overview

• About the class
• Milestones of deep learning
• Progress in the last decade
  • Visual recognition
  • Natural language understanding
Neural machine translation

Google Neural Machine Translation (GNMT)

https://mobile.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

Transformers

Previous system (before deep learning):
PBMT (2014): 37 BLEU

A. Vaswani et al. [Attention is all you need](https://arxiv.org/abs/1706.03762). NeurIPS 2017

Figure source

English French Translation Quality (BLEU score)
Large language models: Google BERT

- **Self-supervised pre-training** task: masked token prediction

J. Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. EMNLP 2018
Large language models: OpenAI GPT

- **Self-supervised pre-training** task: next token prediction

GPT: A. Radford et al. [Improving language understanding with unsupervised learning](https://openai.com/blog/language-unsupervised/). 2018

GPT-2 (1.5B parameters): A. Radford et al. [Language models are unsupervised multitask learners](https://openai.com/blog/language-unsupervised/). 2019

GPT-3 (175B parameters): T. Brown et al. [Language models are few-shot learners](https://openai.com/blog/language-unsupervised/). NeurIPS 2020 (Best Paper Award)
Stochastic parrots or sentient entities?*

*Asking either question will get you fired from Google

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

E. Bender et al., On the dangers of stochastic parrots: Can language models be too big? FAccT 2021

InstructGPT and ChatGPT

Reinforcement Learning with Human Feedback (RLHF)

L. Ouyang et al. Training language models to follow instructions with human feedback. NeurIPS 2022
https://openai.com/blog/chatgpt/
Write a poem explaining ChatGPT for an Introduction to Deep Learning class

ChatGPT is a model that’s deep and grand,
Built with machine learning in hand,
It processes language with such ease,
Giving answers that often please.

It’s based on the transformer architecture,
Which allows for seamless interaction,
It’s trained on vast amounts of data,
Giving it knowledge that’s second to none, one could say.

It can generate text that’s brand new,
Or complete the sentence you started, too,
It can answer questions with great accuracy,
And even write poetry with a certain fluidity.

But ChatGPT is more than just a tool,
It’s a glimpse into the future, that’s cool,
With each improvement and update,
We inch closer to machines that think and debate.

So let ChatGPT be your guide,
As you explore the world of deep learning inside,
And who knows, maybe one day you’ll build,
A model that’s even greater still.
Write a Biblical verse in the style of the King James Bible explaining ChatGPT to an Introduction to Deep Learning class

"And lo, in the land of machine learning, a model was created, deep and grand, yea, verily, it was named ChatGPT, and it was trained upon great amounts of data, and it was given the knowledge of language and the power to process it with ease. And the people did marvel at its ability to generate new text, answer questions and complete sentences, yea, they did call it a tool of great accuracy, and a glimpse into the future of artificial intelligence. And the Lord did say, 'let ChatGPT be thy guide as thou explore the depths of deep learning, for it is through understanding the creation that thou shalt become creators thyself'."

Note: This verse is a fictional creation, inspired by the style and language of the King James Bible. The King James Version is the first translation of the Bible and has literary style, in this verse I tried to mimic this style in a fun way that describes the concept of ChatGPT.
ChatGPT is a computer program that can understand and write in the English language. It's like having a super smart friend who can help you write stories, answer questions and even make up new words. It works by looking at a lot of examples of how people talk and write and then it can use what it has learned to help you. It's kind of like how you learn new words and sentences by listening to people talk and reading books. And it's fun to use because you can ask it to write stories or make up jokes and it can come up with some pretty funny stuff!
ChatGPT: Concerns

Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach

With the rise of the popular new chatbot ChatGPT, colleges are restructuring some courses and taking preventive measures.

ChatGPT: Concerns – and opportunities

- **Gizmodo**
  - iPhone ChatGPT App Charges $7.99 Per Week for Free AI Chatbot
  - The shady "ChatGPT Chat GPT AI With GPT-3" app is currently the fifth most popular download in the App Store's productivity category.
  - 1 hour ago

- **MarketWatch**
  - OpenAI's ChatGPT has awed the world. Expect a version of Google Duplex to strike back amid a wave of AI assistants
  - Since it was unveiled in 2018, Google Duplex has been making waves in the world of artificial intelligence and voice assistants...
  - 2 hours ago

- **Axios**
  - Hackers are using ChatGPT to write malware, build data-encryption tools, researchers find
  - Malicious hackers are already using the flashy new AI chatbot, ChatGPT, to create new low-level cyber tools, including malware and...
  - 10 hours ago

- **Business Insider**
  - ChatGPT is taking over the internet
  - If you've somehow missed the ChatGPT buzz so far, it's a chatbot equipped with the most advanced example of generative artificial...
  - 2 days ago

- **Dark Reading**
  - Attackers Are Already Exploiting ChatGPT to Write Malicious Code
  - The AI-based chatbot is allowing bad actors with absolutely no coding experience to develop malware.
  - 21 hours ago

Some Google search results as of 1/10/2023
Lecture overview

- About the class
- Milestones of deep learning
- Progress in the last decade
  - Vision
  - Language
  - Generative modeling
Progress in face generation

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

2014 2015 2016 2017 2018
Progress in general category generation

GAN-generated dogs in 2017

Source: EBGAN

GAN-generated dogs in 2018

Source: BigGAN
Text-to-image generation: OpenAI DALL-E

A. Ramesh et al., Zero-Shot Text-to-Image Generation, ICML 2021
https://openai.com/blog/dall-e/
Text-to-image generation: OpenAI DALL-E

- Underlying technology: autoregressive generation using a transformer decoder

A. Ramesh et al., Zero-Shot Text-to-Image Generation, ICML 2021
https://openai.com/blog/dall-e/
Text-to-image generation: OpenAI DALL-E 2

A. Ramesh et al. Hierarchical text-conditional image generation with CLIP latents. 2022
Diffusion models

- Idea: convert noise to an image in multiple passes


Diffusion models

• Idea: convert noise to an image in multiple passes
• Proliferation of models: Imagen, Stable Diffusion, Midjourney, …
• Text-to-video, text-to-3D, …
Diffusion models: The next gold rush?

Why Venture Capital Investors Are Betting on Generative AI

19 December 2022 | Blog
Author(s): Louis Lehot, Shabbi S. Khan, Natasha Allen
Published To: Foley Ignite, Innovative Technology Insights

Generative modeling: Concerns

- Deepfakes
- Biases, toxic content
- AI replacing artists?

A deepfake of Ukrainian President Volodymyr Zelensky calling on his soldiers to lay down their weapons was reportedly uploaded to a hacked Ukrainian news website today, per @Shayan86

DALL-E 2 images of lawyers, flight attendants (source)

AI-generated work wins first prize at art fair

https://www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/
Lecture overview

• About the class
• Milestones of deep learning
• Progress in the last decade
  • Vision
  • Language
  • Generative modeling
• Games
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
Lecture overview

• About the class
• Milestones of deep learning
• Progress in the last decade
  • Vision
  • Language
  • Generative modeling
  • Games
  • 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
Sensorimotor learning

Figure 1: Our robot can traverse a variety of challenging terrain in indoor and outdoor environments, urban and natural settings during day and night using a single front-facing depth camera. The robot can traverse curbs, stairs and moderately rocky terrain. Despite being much smaller than other commonly used legged robots, it is able to climb stairs and curbs of a similar height. Videos at https://vision-locomotion.github.io

Lecture overview

- About the class
- Milestones of deep learning
- Progress in the last decade
  - Vision
  - Language
  - Generative modeling
  - 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, dense prediction**

**Self-supervised learning**

**Generative models (GANs, image-to-image translation, diffusion models)**

**Recurrent models**

**Transformers, large language models, transformers for vision**

**Deep reinforcement learning**
Fascinating historical reading

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

Walter Pitts (1923-1969)
Fascinating historical reading

• 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