

# CMSC 373 Artificial Intelligence

## Fall 2023

### 15-Object Recognition

Deepak Kumar  
Bryn Mawr College

1

## Recent AI Milestones

- 1998 [Furby](#) (over 40 million sold!)
- 2002 [Microsoft Clippy](#)
- Mid 2000's  
Google Translate launched  
Voice Assistants (Siri, Alexa)  
YouTube auto subtitles  
Skype Voice Translation  
Facebook Facial Recognition
- 2011 Watson wins Jeopardy
- 2012 Google's cat recognition
- 2012 Microsoft Voice Translation
- 2016 AlphaGo beats World Champion



It looks like you're writing a letter.

Would you like help?

- Get help with writing the letter
- Just type the letter without help
- Don't show me this tip again



WEIRD STAFF SCIENCE JUN 26, 2012 11:15 AM

### Google's Artificial Brain Learns to Find Cat Videos

When computer scientists at Google's mysterious X lab built a neural network of 16,000 computer processors with one billion connections and let it browse YouTube, it did what many web users might do -- it began to look for cats.

2

2

## Furby (1998)



2023 Furby reintroduced!

3

3

## Google Translate, 2006

- Machine Translation
- See: <https://translate.google.com/>

4

4

# Google Translate (My own Tests)



11/6/2023

5

5

# Machine Translation: Progress

## Automatically translated text:

Lentil important not to be required to bathroom. We recommend cleaning a finger.

December 2009

Important lentil should not be kept in the bathroom. Finger cleaning is recommended.



Suggest an edit

September 2018

Importante la lenticchia non va tenuta a bagno. Si consiglia la pulitura a "dito".



82 / 5,000

Important lentils should not be kept in the water. Finger cleaning is recommended.



Send feedback

August 2022

Important: the lentil should not be soaked. Finger cleaning is recommended.

[Look up details](#)

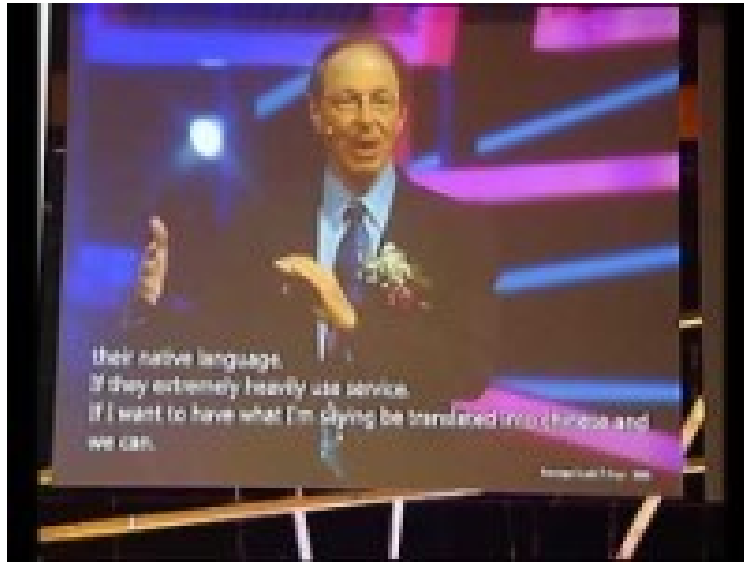

November 2023

11/6/2023

6

6

## Voice Translatrion Demo, 2012



7

7

## Characterizing AI (Modern)

- Narrow AI
- General AI
- The Singularity
- Neural Engineering
- Artificial Intelligence = Deep Learning

8

8

## “Christopher Columbus Ploy”

- “They all laughed at Christopher Columbus”  
underestimating the the progress and impact of technology
- “I think there is a world market for maybe five computers.”
- “There’s no reason for individuals to have a computer in their home.”
- 640,000 bytes of memory ought to be enough for anybody.”

9

9

## “Christopher Columbus Ploy”

- “They all laughed at Christopher Columbus”  
underestimating the progress and impact of technology
- “I think there is a world market for maybe five computers.”  
[Thomas J. Watson, Chairman of IBM, 1943.](#)
- “There’s no reason for individuals to have a computer in their home.”  
[Ken Olson, Cofounder of DEC, 1977.](#)
- 640,000 bytes of memory ought to be enough for anybody.”  
[Bill Gates, Founder of Microsoft, 1981.](#)

10

10

## “Christopher Columbus Ploy”

- “They all laughed at Christopher Columbus”  
underestimating the progress and impact of technology
- “I think there is a world market for maybe five computers.”  
Thomas J. Watson, Chairman of IBM, 1943.
- “There’s no reason for individuals to have a computer in their home.”  
Ken Olson, Cofounder of DEC, 1977.
- 640,000 bytes of memory ought to be enough for anybody.”  
Bill Gates, Founder of Microsoft, 1981.

“By 2000 computers will defeat the world Chess champion.”  
Raymod Kurzweil, 1990.

[For more Kurzweil predictions, ee: [https://medium.com/@singularity\\_41680/ray-kurzweils-mind-boggling-predictions-for-the-next-25-years-ce3c9163588b](https://medium.com/@singularity_41680/ray-kurzweils-mind-boggling-predictions-for-the-next-25-years-ce3c9163588b)]

11

11

## Image Understanding



David Schwimmer as Ross Geller, Jennifer Aniston as Rachel Green, Courteney Cox as Monica Geller, Matthew Perry as Chandler Bing, Lisa Kudrow as Phoebe Buffay, Matt LeBlanc as Joey Tribbiani in "Friends."  
Getty Images / Warner Bros. Television via Getty Images

12

12

## Easy things are hard 😊



1966, “Marvin Minsky hired a first-year undergraduate and assigned him a problem to solve over the summer: connect a television camera to a computer and get the machine to describe what it sees.”

- **Computer Vision**

Still do not have a program that can look at and describe an image the way humans do. Still far out of reach.

A simpler task:

**Object recognition** (see any dogs?)  
using **Deep Learning**.



13

13

## Deep Learning Revolution

- **Deep Learning** refers to deep neural networks (i.e. many hidden layers)
- The “deep” in Deep learning is NOT “learning that is deep” (i.e. meaningful or sophisticated learning!!)
- The “deep” ONLY refers to the “**depth in layers**” of the neural network.
- **Convolution Networks** are a kind of Deep Neural network.

14

14

## Taking Inspiration from the Visual Cortex

- Hubel & Wiesel's experiments with cats and primates
  - Visual cortex is a hierarchical series of layers of neurons. Layers communicate back and forth extensively.
  - Layers act as feature detectors (edges, shapes, objects, etc).
  - Each neuron receives input corresponding to a specific small region of the visual scene (neuron's receptive field).
  - Neurons activate only if their receptive field contains a particular kind of edge/feature (e.g. horizontal edge, vertical edge, angular edge, etc)
  - Lower-level neurons feed into higher level layers of the visual cortex for detecting shapes, objects, faces, etc.
- This is still a gross simplification. The brain is much more complex!



Image From: <https://www.news-medical.net/life-sciences/How-do-Visual-Neurons-Work.aspx>

15

15

## Convolution Networks – Short History

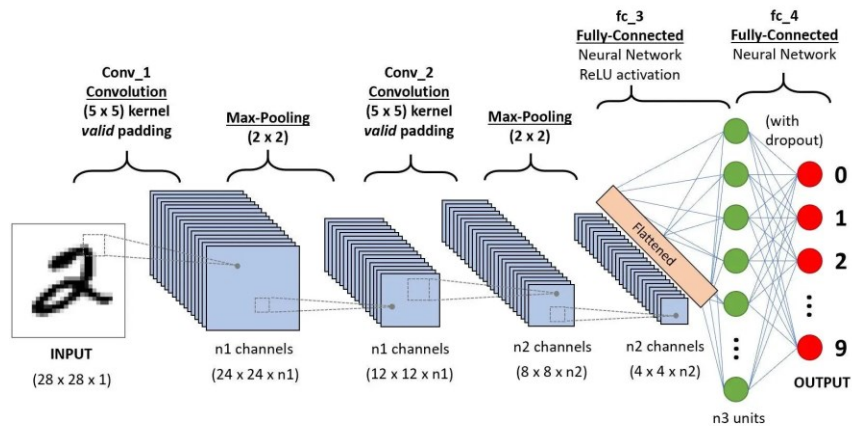
- Taking inspiration from Hubel & Wiesel...
- From Cognitron, to NeoCognitron (Fukushima, 1970s)  
Showed how a hierarchical network (using Relu!) could learn using unsupervised means.
- To Convolutional Neural Networks (1989) used for hand-written zipcodes. Developed by Yann LeCun (at AT&T Bell Labs). In 1995, LeCun et al developed LeNet-5 to classify handwritten digits (32x32 pixel images). Used to recognize numbers on checks by banks.

16

16



# ConvNet Architecture



From: <https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/>

17

17

## ConvNets in 1980s

- While successfully used for handwritten zip code recognition, and numbers on bank checks, they did not scale up to more complex vision tasks.
- By mid-1990s ConvNets fell out of favor.
- But, LeCun “...carried the torch through the dark ages.”
- They believed that improved, larger versions of ConvNets would enable better computer vision if they could be trained with enough data.

18

18

# ImageNet

- 2005-2010 PASCAL Visual Object Classes Competitions  
~15,000 photographs from Flickr, 20 human labelled categories
- However, the dataset was not large enough. And would not scale up beyond the 20 categories.
- Fei Fei Li (Princeton) created an image database structured around the nouns in WordNet. Each noun would be linked to a large number of images with examples of that noun. [ImageNet]
- Used **Amazon Mechanical Turk** (“Artificial Artificial Intelligence”) to pay tens of thousands of human workers to sort out irrelevant images for WordNet terms. Took two years and more than 3 million images were labelled.
- 2012 Launch of **ImageNet Visual Recognition Challenge**. Contained 1.2 million labelled training images. Task was to predict the correct category of each input image. There were 1000 image categories. There were 150,000 test images.

19

19

# Amazon Mechanical Turk

- “Potemkin AI”
- AI’s hidden human labor.
- Global crowdsourcing marketplace.
- Primarily used for human data labelling tasks needed for ML.

Amazon Mechanical Turk (MTurk) is a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually. This could include anything from conducting simple data validation and research to more subjective tasks like survey participation, content moderation, and more. MTurk enables companies to harness the collective intelligence, skills, and insights from a global workforce to streamline business processes, augment data collection and analysis, and accelerate machine learning development.

While technology continues to improve, there are still many things that human beings can do much more effectively than computers, such as moderating content, performing data deduplication, or research. Traditionally, tasks like this have been accomplished by hiring a large temporary workforce, which is time-consuming, expensive and difficult to scale, or have gone undone. Crowdsourcing is a good way to break down a manual, time-consuming project into smaller, more manageable tasks to be completed by distributed workers over the Internet (also known as ‘microtasks’).

## Benefits

### Optimize efficiency

MTurk is well-suited to take on simple and repetitive tasks in your workflows which need to be handled manually. Using MTurk to outsource microtasks ensures that work gets done quickly, while freeing up time and resources for the company—so internal staff can focus on higher value activities.

### Increase flexibility

Scaling up and down a workforce isn't the easiest undertaking. With access to a global, on-demand, 24x7 workforce, MTurk enables businesses and organizations to get work done easily and quickly when they need it—without the difficulty associated with dynamically scaling your in-house workforce.

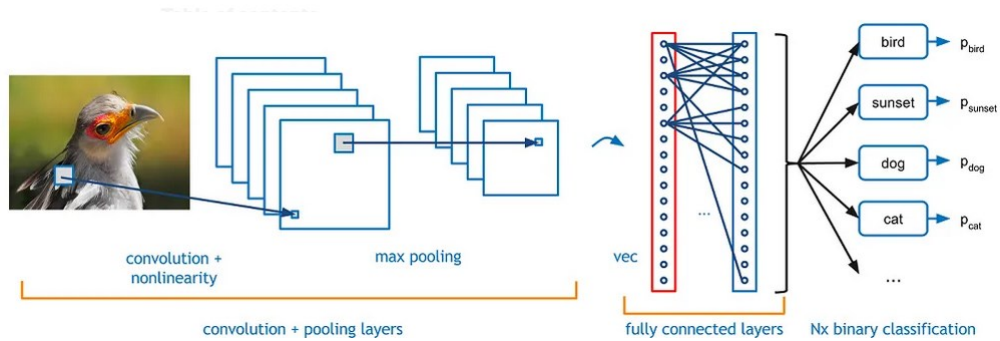
### Reduce cost

MTurk offers a way to effectively manage labor and overhead costs associated with hiring and managing a temporary workforce. By leveraging the skills of distributed Workers on a pay-per-task model, you can significantly lower costs while achieving results that might not have been possible with just a dedicated team.

20

20

## Convnet Architecture



From: <https://towardsdatascience.com/convolutional-neural-network-cb0883dd6529>

21

21

## ImageNet Competitions

- Programs were to use the “top-5” accuracy metric.
- 2010: Winning program used SVMs and obtained **72%** accuracy.
- 2011: Winning program used SVMs and obtained **74%** accuracy.
- 2012: A ConvNet (aka AlexNet) obtained **85%** accuracy. Used a scaled up version of LeCun’s LeNet and utilized **GPU’s** to improve computing power. **AlexNet** has 60 million weights and used backpropagation.
- Simultaneously, using ConvNets gained a significantly better performance on speech recognition.

22

22

## ImageNet Competitions

- 2015: Baidu's cheating scandal in ImageNet competition (obtaining 94.6% accuracy).
- 2015: Team from Microsoft announced **95.06%**.
- 2015: Team from Google obtained **95.18%** accuracy.
- 2015: Baidu announced **95.42%** accuracy (due to cheating!)
- 2017: Winning entry had **98%** accuracy.

23

23

## Headlines...

- *The Guardian*, May 13, 2015.
- *UPI Science News*, Feb 14, 2015.

SCIENCE NEWS FEB. 14, 2015 / 8:06 PM

### Microsoft has developed a computer system that can identify objects better than humans

It can identify objects with slightly fewer errors than humans.

From: [https://www.upi.com/Science\\_News/2015/02/14/Microsoft-has-developed-a-computer-system-that-can-identify-objects-better-than-humans/1171423959603/](https://www.upi.com/Science_News/2015/02/14/Microsoft-has-developed-a-computer-system-that-can-identify-objects-better-than-humans/1171423959603/)



### Computers now better than humans at recognising and sorting images

Baidu's Minwa supercomputer can sort a million images into a thousand predefined categories with an error rate less than the typical person



People talk in front of Baidu's company logo at its headquarters in Beijing, China. Jason Lee/Reuters Photograph: Jason Lee/Reuters

First, the robots beat us at assembling cars. Then, they beat us at playing chess. Now, they've also got better than us at sorting images into predefined categories.

From: <https://www.theguardian.com/global/2015/may/13/baidu-minwa-supercomputer-better-than-humans-recognising-images>

24

24

## Easy things are very hard

- “get a machine to describe what it sees” (in an image)
- Machines will need to recognize more than just objects.
- Machines will need to know something about living beings, and how they interact with the world, their actions, goals, emotions, etc.
- We are still very far from visual intelligence (Visual Intelligence is ***AI Complete***).



25

25

## References

- F. Chollet: *Deep Learning with Python, Second Edition*, Manning 2021.
- K. Crawford: *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.
- M. Mitchell: *Artificial Intelligence: A Guide For Thinking Humans*, Farrar, Straus, Giroux, 2019.
- M. Wooldridge: *A Brief History of Artificial Intelligence*. Flatiron Books, 2020.

26

26