Deep Learning for Computer Vision

Lecture 1: Introduction

Deep Learning for Computer Vision

Deep Learning for Computer Vision

Building artificial systems
that process, perceive, and
reason about visual data

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

Deep Learning for Computer Vision

Building artificial systems that
learn from data and experience

Deep Learning for Computer Vision

Hierarchical learning algorithms with many “layers”, (very) loosely inspired by the brain

Artificial Intelligence

Artificial Intelligence
Machine Learning
Computer Vision

Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning

This class
Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning

This class
Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning
Natural Language Processing

This class
Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning
Natural Language Processing
Speech Recognition

This class
Artificial Intelligence
Robotics
Machine Learning
Computer Vision
Deep Learning
Natural Language Processing
Speech Recognition

Today's Agenda

- A brief history of computer vision and deep learning
- Course overview and logistics

Hubel and Wiesel, 1959

Measure brain activity

Simple cells:
Response to light orientation

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

1959
Hubel & Wiesel

No response

Hubel and Wiesel, 1959

Measure brain activity

Simple cells: Response to light orientation

Complex cells: Response to light orientation and movement

No response

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

1959
Hubel & Wiesel

Larry Roberts, 1963

(a) Original picture

(b) Differentiated picture

(c) Feature points selected

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

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

1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

Input image

This image is CC0 1.0 public domain

Edge image

2 ½-D sketch

3-D model

This image is CC0 1.0 public domain

Input Image	Primal Sketch	2 ½-D Sketch	3-D Model Representation
Perceived intensities	Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries	Local surface orientation and discontinuities in depth and in surface orientation	3-D models hierarchically organized in terms of surface and volumetric primitives

Stages of Visual Representation, David Marr, 1970s

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

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

Left Image is CC BY 3.0 Middl Image is public domain Right Image is CC-BY 2.0; changes made

Recognition via Matching (2000s)

Image is public domain

Image is public domain

SIFT, David Lowe, 1999

Face Detection

Viola and Jones, 2001

One of the first successful applications of machine learning to vision

PASCAL Visual Object Challenge

Image is CC0 1.0 public domain

Image is CC0 1.0 public domain

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

AlexNet: Deep Learning Goes Mainstream

Krizhevsky, Sutskever, and Hinton, NeurIPS 2012

Perceptron

One of the earliest algorithms that could learn from data

Implemented in hardware! Weights stored in potentiometers, updated with electric motors during learning

Connected to a camera that used 20x20 cadmium sulfide photocells to make a 400-pixel image

Could learn to recognize letters of the alphabet

Today we would recognize it as a **linear classifier**

Frank Rosenblatt, ~1957

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

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

Neocognitron: Fukushima, 1980

Looks a lot like AlexNet more than 32 years later!

Backprop: Rumelhart, Hinton, and Williams, 1986

Introduced backpropagation for computing gradients in neural networks

Successfully trained perceptrons with multiple layers

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial w_{ji}}$$

recognizable math

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!

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

2012 to Present: Deep Learning Explosion

Publications at top Computer Vision conference

arXiv papers per month (source)

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

Galaxy Classification

Dieleman et al, 2014

From left to right: public domain by NASA, usage permitted by ESA/Hubble, public domain by NASA, and public domain.

Whale recognition

Kaggle Challenge

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

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

Image Captioning

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

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

All images are CC0 Public domain:
https://pixabay.com/en/luggage-antique-cat-1643010/
https://pixabay.com/en/teddy-plush-bears-cute-teddy-bear-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

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

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

Mordvinsev et al, 2015

Gatys et al, 2016

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

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

2012 to Present: Deep Learning is Everywhere

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

Algorithms
Data
Computation

GFLOP per Dollar

GFLOP per Dollar

2018 Turing Award

Despite our success, 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

This image is copyright-free United States government work

Example credit:
Andrej Karpathy

Computer Vision Technology Can Better Our Lives

Outside border images, clockwise, starting from top left:
Image by Tucania is licensed under CC BY 2.0; changes made
Image by the US Government is in the public domain
Image by the US Government is in the public domain
Image by Flickr is licensed under CC BY-SA 2.0; changes made
Image by Genius is licensed under CC BY 3.0; changes made
Image by the US Government is in the public domain

Inside image, clockwise, starting from top left:
Image by Tucania is licensed under CC BY 2.0; changes made
Image by the US Government is in the public domain
Image by Intuitive Surgical, Inc. is licensed under CC BY-SA 3.0; changes made
Image by Ovundari Zorigtbaatar is licensed under CC BY-SA 4.0

Today's Agenda

- A brief history of computer vision and deep learning
- Course overview and logistics

Today's Agenda

- A brief history of computer vision and deep learning
- Course overview and logistics

Optional Textbook

- *Deep Learning* by Goodfellow, Bengio, and Courville
- Free online

Next time: Image Classification