CS231n: Deep Learning for Computer Vision

Lecture 1: Introduction
Welcome to CS231n
Welcome to CS231n

CS231n: Lecture 1 - 3

April 2, 2024
Artificial Intelligence

Computer Vision

Machine Learning

Deep Learning

Slide inspiration: Justin Johnson
Artificial Intelligence

Machine Learning

Deep Learning

Computer Vision

This class

Slide inspiration: Justin Johnson
This class

Artificial Intelligence

Machine Learning

Deep Learning

Computer Vision

Natural Language Processing

Slide inspiration: Justin Johnson
Today’s agenda

• A brief history of computer vision and deep learning

• CS231n overview
Evolution’s Big Bang: Cambrian Explosion, 530-540 million years, B.C.
Camera Obscura

Gemma Frisius, 1545

Encyclopedia, 18th Century

Leonardo da Vinci, 16th Century AD
Computer Vision is everywhere!
Where did we come from?
Hubel and Wiesel, 1959

Measure brain activity

Simple cells:
Response to specific rotation and orientation

Complex cells:
Response to light orientation and movement, some translation invariance

1959
Hubel & Wiesel

Response
Stimulus
Larry Roberts, 1963

(a) Original picture
(b) Differentiated picture
(c) Feature points selected


Slide inspiration: Justin Johnson
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".
Stages of Visual Representation, David Marr, 1970s

- **Input Image**
  - Perceived intensities

- **Primal Sketch**
  - Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves, boundaries

- **2 ½-D Sketch**
  - Local surface orientation and discontinuities in depth and in surface orientation

- **3-D Model Representation**
  - 3-D models hierarchically organized in terms of surface and volumetric primitives

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr

Slide inspiration: Justin Johnson
Recognition via Parts (1970s)

- **1959** Hubel & Wiesel
- **1963** Roberts
- **1970s** David Marr
- **1979** Generalized Cylinders, Brooks and Binford

- **1973** Pictorial Structures, Fischler and Elshlager

Slide inspiration: Justin Johnson
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.
In the meantime... seminal work in cognitive and neuroscience
Perceiving Real-World Scenes

Irving Biederman

Rapid Serial Visual Perception (RSVP)
Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot

Neural correlates of object & scene recognition


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

Normalized Cuts, Shi and Malik, 1997

Slide inspiration: Justin Johnson
Recognition via Matching (2000s)

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
1986 Norm. Cuts
1997 SIFT

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

Slide inspiration: Justin Johnson
Caltech 101 images

PASCAL Visual Object Challenge

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

Slide inspiration: Justin Johnson
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

---

- 1958: Perceptron
  - Frank Rosenblatt, ~1957
- 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

---

Slide inspiration: Justin Johnson
Minsky and Papert, 1969

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>F(x,y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

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
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

Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1986 Canny
1985 Backprop
1997 Norm. Cuts
1999 SIFT
2001 V&J
2004, 2007 Caltech101; PASCAL

1958 Perceptron
1969 Minsky & Papert
1980 Neocognitron

Al Winter

Slide inspiration: Justin Johnson
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
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
The Image Classification Challenge:
1,000 object classes
1,431,167 images

Deng et al, 2009
Russakovsky et al. IJCV 2015

Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle
CS231n: Lecture 1 - 43

April 2, 2024

Large Scale Visual Recognition Challenge

- Hubel & Wiesel (1959)
- Roberts (1963)
- David Marr (1970s)
- Gen. Cylinders (1979)
- Canny (1986)
- Norm. Cuts (1997)
- V&J (1999)
- SIFT (1999)
- V&J (2001)
- Caltech101; PASCAL (2004, 2007)
- ImageNet (2009)
- LeNet (1985)
- Neocognitron (1980)
- Backprop (1985)
- Deep Learning (2006)
- ImageNet (2009)
- Al Winter

- Minsky & Papert (1958)
- Perceptron (1969)
- Neocognitron (1980)
- Backprop (1985)
- ImageNet (2009)

Fei-Fei Li & Ehsan Adeli
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
2009 ImageNet
1958 Perceptron
1969 Minsky & Papert
1980 Neocognitron
1985 Backprop
1998 LeNet
2006 Deep Learning
2012 AlexNet

AlexNet, 2012

Lin et al
Sanchez & Perronnin
Krizhevsky et al (AlexNet)
Zeiler & Fergus
Simonyan & Zisserman (VGG)
Szegedy et al (GoogLeNet)
He et al (ResNet)
Shao et al
Hu et al (SENet)
Russakovsky et al

ImageNet Large Scale Visual Recognition Challenge

Fei-Fei Li & Ehsan Adeli

CS231n: Lecture 1 - 44
April 2, 2024
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**

- Subm...
- Acce...

**Publications at top Computer Vision conference**

- 1959 Hubel & Wiesel
- 1963 Roberts
- 1970s David Marr
- 1979 Gen. Cylinders
- 1986 Canny
- 1958 Perceptron
- 1969 Minsky & Papert
- 1980 Neocognitron
- 1985 Backprop

**arXiv papers per month (source)**

- 1994.01
- 2007.01
- 2020.09

- 1959 Hubel & Wiesel
- 1963 Roberts
- 1970s David Marr
- 1979 Gen. Cylinders
- 1986 Canny
- 1958 Perceptron
- 1969 Minsky & Papert
- 1980 Neocognitron
- 1985 Backprop

**ML+AI arXiv papers per month**

- 1997 Norm. Cuts
- 1999 SIFT
- 2001 V&J
- 2004, 2007 Caltech101; PASCAL
- 2009 ImageNet
- 1998 LeNet
- 2006 Deep Learning
- 2012 AlexNet

Slide inspiration: Justin Johnson
2012 to Present: Deep Learning is Everywhere

Year 2010
NEC-UIUC
- Dense descriptor grid: HOG, LBP
- Coding: local coordinate, super-vector
- Pooling, SPM
- Linear SVM

[Lin CVPR 2011]

Year 2012
SuperVision

[Krizhevsky NIPS 2012]

Year 2014
GoogLeNet

[Simonyan arxiv 2014]

VGG

[He ICCV 2015]

Year 2015
MSRA

Image
conv-64
conv-64
maxpool
conv
conv
maxpool
conv
conv
maxpool
conv
maxpool
fc-4096
fc-4096
fc-1000
softmax

Pooling
Convoluti
on
Softmax
Other

Other

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

Lion image by Swissfrog
Licensed under CC BY 3.0

Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.
2012 to Present: Deep Learning is Everywhere

Image Classification

Image Retrieval

2012 to Present: Deep Learning is Everywhere

Object Detection

Image Segmentation

Ren, He, Girshick, and Sun, 2015

Fabaret et al, 2012

Slide inspiration: Justin Johnson
2012 to Present: Deep Learning is Everywhere

Video Classification

Activity Recognition

Simonyan et al, 2014
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

Galaxy Classification

Dieleman et al, 2014

Whale recognition

Kaggle Challenge

---

Figure reproduced with permission.
2012 to Present: Deep Learning is Everywhere

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

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

Captions generated by Justin Johnson using Neuraltalk2

All images are CC0 Public domain:

CS231n: Lecture 1 - 54
April 2, 2024
Fei-Fei Li & Ehsan Adeli
2012 to Present: Deep Learning is Everywhere

Results:
spatial, comparative, asymmetrical, verb, prepositional

taller than

person

left of

person

wear

on

wear

shirt

snow

ski

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

TEXT PROMPT

an **armchair** in the **shape** of an **avocado**. an **armchair** imitating an **avocado**.

AI-GENERATED IMAGES

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

GFLOP per Dollar

- CPU
- GPU (FP32)

GFLOP per Dollar chart showing the performance improvements over time with various GeForce GPUs. The chart highlights the 'Deep Learning Explosion' period, with significant improvements from GTX 8800 to RTX 3090.
Recent GPUs have "Tensor Cores": Special hardware for deep learning!
AI’s Explosive Growth & Impact

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

Affect people’s lives

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.’

Example Credit: Timnit Gebru

Kate Crawford, “The Trouble with Bias”, NeurIPS 2017 Keynote


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

April 2, 2024
And there is a lot we don’t know how to do
Today’s agenda

• A brief history of computer vision & deep learning

• CS231n overview