Announcements

- The final exam is in-class on Friday. It will be mainly multiple choice questions similar to the quizzes, and it will cover all of the assigned readings.
- I’m holding an extra set of office hours tomorrow from 1:30-3:30 in 3401 Walnut room 452C.
- HW5 is due on Tuesday, late days are OK to use.
- Today’s lecture on Computer Vision is just for fun – it’s not on the exam.
“Connect a television camera to a computer and get the machine to describe what it sees.”

Marvin Minsky
Turing award, 1969
Marvin Minsky, Pioneer in Artificial Intelligence, Dies at 88

By GLENN RIFKIN  JAN. 25, 2016

Marvin Minsky, who combined a scientist’s thirst for knowledge with a philosopher’s quest for truth as a pioneering explorer of artificial intelligence, work that helped inspire the creation of the personal computer and the Internet, died on Sunday night in Boston. He was 88.

His family said the cause was a cerebral hemorrhage.

Well before the advent of the microprocessor and the supercomputer, Professor Minsky, a revered computer science educator at M.I.T., laid the foundation for the field of artificial intelligence by demonstrating the possibilities of imparting common-sense reasoning to computers.

"Marvin was one of the very few people in computing whose visions and perspectives liberated the computer from the constraints of data and structure into the domain of meaning and inference," said John McCarthy, a co-founder of the field of artificial intelligence and a professor emeritus at Stanford University.

Minsky was a prodigy who, at 12, had earned a spot in the University of Chicago’s physics honors program. At 18, he joined the U.S. Navy. After World War II, he earned a doctorate in aeronautical engineering from the Massachusetts Institute of Technology, where he was an instructor in the computer science department.

In the 1950s, he became a leader in the development of computational neuroscience and the recognition of various aspects of the mind as computational systems. He founded the MIT Artificial Intelligence Laboratory in 1962 and served as its director until 1970.

"Minsky was one of the first computer scientists who had a mind for the future," said Stuart Kauffman, a co-founder of the Santa Fe Institute.

Minsky is survived by his wife; two children, John and Anne; and five grandchildren.

"Marvin, the archetypal mad scientist," said John McCarthy. "He was always a little bit crazy, always an iconoclast, always a seeker of ideas that were extreme, that were beyond what the current thinking was."
Stages of Visual Representation, David Marr, 1970

Input Image

Perceived intensities

Primal Sketch

Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves, boundaries.

2½-D Sketch

Local surface orientation and discontinuities in depth and in surface orientation

3-D Model Representation

3-D models hierarchically organised in terms of surface and volumetric primitives
The representation and matching of pictorial structures,
Fischler and Elschlager, 1973
1980’s

AI winter... ...back to basics

A Computational Approach to Edge Detection, Canny 1986
1984

Perceptual Organization and Visual Recognition,
David Lowe, 1984
Perceptual organization and the representation of natural form,
Alex Pentland, 1986
1989

Backpropagation applied to handwritten zip code recognition,
Lecun et al., 1989
Filters

Input

<table>
<thead>
<tr>
<th>80322</th>
<th>4129</th>
<th>80206</th>
</tr>
</thead>
<tbody>
<tr>
<td>40004</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>35460</td>
<td>4209</td>
<td></td>
</tr>
</tbody>
</table>

\[\begin{array}{ccc}
-1 & 0 & +1 \\
-2 & 0 & +2 \\
-1 & 0 & +1 \\
\end{array}\]

\[\begin{array}{ccc}
+1 & +2 & +1 \\
0 & 0 & 0 \\
-1 & -2 & -1 \\
\end{array}\]
Backpropagation applied to handwritten zip code recognition,
Lecun et al., 1989
Backpropagation applied to handwritten zip code recognition,
Lecun et al., 1989
Neural Network-Based Face Detection, Rowley at al., PAMI 1998
Sliding window in real time!
Boosting + Cascade = Speed

Rapid Object Detection using a Boosted Cascade of Simple Features,
Viola and Jones, CVPR 2001
No more sliding windows (interest points)
Better features (use more computation)

SIFT Matching

[SIFT: Lowe, 2004]
Interest points
Object Class Recognition by Unsupervised Scale-Invariant Learning, Fergus et al., CVPR 2003.
2003

Constellation model (redux)

The representation and matching of pictorial structures, Fischler and Elschlager, 1973
Interest points used to find parts:

Smaller number of candidate parts allows for more complex spatial models.
Why it fails

Interest points don’t work for category recognition
Too many springs...
Classification Vs. Detection

✓✓Cat

Cat

Cat
Histograms of oriented gradients for human detection,
Dalal and Triggs, CVPR 2005.
Pedestrians

- Defined by their contours

- Cluttered backgrounds

- Significant variance in texture

Interest points won’t work...

...back to sliding window.
2005 HOG (histograms of oriented gradients)
2005 HOG (histograms of oriented gradients)

Presence > Magnitude

✔️ Normalization by a local window
What’s this?

[Dalal and Triggs, 2005]
Yup.
2007 PASCAL VOC

20 classes

Why it is hard
2008 DPM (Deformable parts model)

Object Detection with Discriminatively Trained Part Based Model,
Felzenszwalb, Girshick, McAllester and Ramanan, PAMI, 2010
Object Detection with Discriminatively Trained Part Based Model, Felzenszwalb, Girshick, McAllester and Ramanan, PAMI, 2010
Multiple components
Problems with **Visual** Categories

- A lot of categories are functional

- World is too varied

- Categories are 3D, but images are 2D
Do We Need More Training Data or Better Models for Object Detection?
**IMAGENET**

**22K categories and 14M images**

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
  - Food
  - Materials
- Structures
  - Artifact
  - Tools
  - Appliances
  - Structures
- Person
  - Scenes
  - Indoor
  - Geological Formations
  - Sport Activities

*Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009*
2009 ImageNet

22K categories, 14M images

Corgi

Orb weaving spider

ImageNet: A Large-Scale Hierarchical Image Database, Deng, Dong, Socher, Li, Li and Fei-Fei, CVPR, 2009
Algorithms

2009
- Pooling
- SVM
- HOG
- Image

2012
- Dense
- Convolution
- Dense
- Convolution
- Dense
- Convolution
- Dense
- Convolution
- Dense
- Convolution
- Image
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]
IMAGENET Large Scale Visual Recognition Challenge

**Year 2010**
- NEC-UIUC
  - Dense grid descriptor: HOG, LBP
  - Coding: local coordinate, super-vector
  - Pooling, SPM
  - Linear SVM
  - [Lin CVPR 2011]

**Year 2012**
- SuperVision

**Year 2014**
- GoogLeNet
- VGG
- MSRA
  - [Szegedy arxiv 2014]
  - [Simonyan arxiv 2014]
  - [He arxiv 2014]
Revolution of Depth

ILSVRC'15 ResNet: 3.57%
ILSVRC'14 GoogleNet: 6.7%
ILSVRC'14 VGG: 7.3%
ILSVRC'13: 11.7%
ILSVRC'12 AlexNet: 16.4%
ILSVRC'11: 25.8%
ILSVRC'10: 28.2%

ImageNet Classification top-5 error (%)

Revolution of Depth

AlexNet, 8 layers (ILSVRC 2012)
- 11x11 conv, 96, /4, pool/2
- 5x5 conv, 256, pool/2
- 3x3 conv, 384
- 3x3 conv, 384
- 3x3 conv, 256, pool/2
- fc, 4096
- fc, 4096
- fc, 1000

VGG, 19 layers (ILSVRC 2014)
- 3x3 conv, 64
- 3x3 conv, 64, pool/2
- 3x3 conv, 128
- 3x3 conv, 128, pool/2
- 3x3 conv, 256
- 3x3 conv, 256
- 3x3 conv, 256
- 3x3 conv, 256, pool/2
- 3x3 conv, 512
- 3x3 conv, 512
- 3x3 conv, 512
- 3x3 conv, 512, pool/2
- 3x3 conv, 512
- 3x3 conv, 512
- 3x3 conv, 512
- fc, 4096
- fc, 4096
- fc, 1000

GoogleNet, 22 layers (ILSVRC 2014)

Revolution of Depth

AlexNet, 8 layers (ILSVRC 2012)

VGG, 19 layers (ILSVRC 2014)

ResNet, 152 layers (ILSVRC 2015)

The winner of the 2014 ImageNet competition had 4 million parameters, while the winner of the 2017 challenge had 145.8 million parameters - a 36X increase in three years. source: @jackclarkSF Shall we increase parameters by another 36X, or solve more interesting problems?
“Connect a television camera to a computer and get the machine to describe what it sees.”
“Connect a television camera to a computer and get the machine to describe what it sees.”
Going beyond categorization...

“Connect a television camera to a computer and get the machine to describe what it sees.”

two girls sitting at a table smiling and eating and drinking. a woman is eating a doughnut and drinking beer. there are two woman drinking beers and eating food a woman leaning into another woman as she holds a sandwich towards her. two ladies are enjoying beer and treats at the table.
Going beyond categorization...
VQA: Visual Question Answering

Stanislaw Antol*, Aishwarya Agrawal*, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, Devi Parikh

- What color are her eyes? What is the mustache made of?
- How many slices of pizza are there? Is this a vegetarian pizza?
- Is this person expecting company? What is just under the tree?
- Does it appear to be rainy? Does this person have 20/20 vision?
“Connect a television camera to a computer and get the machine to describe what it sees.”

Marvin Minsky
Turing award, 1969