I,Robot
Isaac Asimov
1950
Future Vision

Eye
Robot
CSCI
1430
2017 MWF 1pm 368
Computer Vision
Things to remember

• Sliding window for search

• Features based on differences of intensity (gradient, wavelet, etc.)
  – Excellent results require careful feature design

• Boosting for feature selection

• Integral images, cascade for speed

• Bootstrapping to deal with many, many negative examples
Starting point: sliding window classifiers

- Detect objects by testing each subwindow
  - Reduces object detection to binary classification
  - Dalal & Triggs: HOG features + linear SVM classifier
  - Previous state of the art for detecting people

Feature vector
\[ x = [..., ..., ..., ..., ...] \]
Histogram of Gradient (HOG) features

- Image is partitioned into 8x8 pixel blocks
- In each block we compute a histogram of gradient orientations
  - 
  - Invariant to changes in lighting, small deformations, etc.
- Compute features at different resolutions (pyramid)
HOG Filters

- Array of weights for features in subwindow of HOG pyramid
- Score is dot product of filter and feature vector

Score of $F$ at position $p$ is

$$ F \cdot \phi(p, H) $$

$\phi(p, H) =$ concatenation of HOG features from subwindow specified by $p$
Dalal & Triggs: HOG + linear SVMs

There is much more background than objects
Start with random negatives and repeat:
1) Train a model
2) Harvest false positives to define “hard negatives”
Discriminative part-based models

- root filters
  - coarse resolution
- part filters
  - finer resolution
- deformation models

Each component has a root filter $F_0$ and $n$ part models $(F_i, v_i, d_i)$
Discriminative part-based models

Root filter  Part filters  Deformation weights

P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Object Detection with Discriminatively Trained Part Based Models, PAMI 32(9), 2010
Car model

Component 1

Component 2
Person model
Bottle model
The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

• Twenty object categories (aeroplane to TV/monitor)

• Three challenges:
  – Classification challenge (is there an X in this image?)
  – Detection challenge (draw a box around every X)
  – Segmentation challenge
Images downloaded from **flickr**
- 500,000 images downloaded and random subset selected for annotation
Dataset: Annotation

- “Complete” annotation of all objects
- Annotated over web with written guidelines
  - High quality (?)
Dataset: Annotation

- “Complete” annotation of all objects
- Annotated over web with written guidelines
  - High quality (?)

20 classes.
- Train / validation data has 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations.
Classification Challenge

- Predict whether at least one object of a given class is present in an image

is there a cat?
### Results: AP by Method and Class

<table>
<thead>
<tr>
<th>Method</th>
<th>aero</th>
<th>bicycle</th>
<th>bird</th>
<th>boat</th>
<th>bottle</th>
<th>bus</th>
<th>car</th>
<th>cat</th>
<th>chair</th>
<th>cow</th>
<th>dining</th>
<th>table</th>
<th>dog</th>
<th>horse</th>
<th>motor</th>
<th>bike</th>
<th>person</th>
<th>potted</th>
<th>plant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tv/monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVC_FLAT</td>
<td>85.3</td>
<td>57.8</td>
<td>66.0</td>
<td>66.1</td>
<td>36.2</td>
<td>70.6</td>
<td>60.6</td>
<td>63.5</td>
<td>55.1</td>
<td>44.6</td>
<td>53.4</td>
<td>49.1</td>
<td>64.4</td>
<td>66.8</td>
<td>84.8</td>
<td>37.4</td>
<td>44.1</td>
<td>47.9</td>
<td>81.9</td>
<td>67.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVC_FLAT-HOG-ESS</td>
<td>86.3</td>
<td>60.7</td>
<td>66.4</td>
<td>65.3</td>
<td>41.0</td>
<td>71.7</td>
<td>64.7</td>
<td>63.9</td>
<td>55.5</td>
<td>40.1</td>
<td>51.3</td>
<td>45.9</td>
<td>65.2</td>
<td>68.9</td>
<td>85.0</td>
<td>40.8</td>
<td>49.0</td>
<td>49.1</td>
<td>81.8</td>
<td>68.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVC_PLUS</td>
<td>86.6</td>
<td>58.4</td>
<td>66.7</td>
<td>67.3</td>
<td>34.8</td>
<td>70.4</td>
<td>60.0</td>
<td>64.2</td>
<td>52.5</td>
<td>43.0</td>
<td>50.8</td>
<td>46.5</td>
<td>64.1</td>
<td>66.8</td>
<td>84.4</td>
<td>37.5</td>
<td>45.1</td>
<td>45.4</td>
<td>82.1</td>
<td>67.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIRSTNIKON_AVGSRKD</td>
<td>83.3</td>
<td>59.3</td>
<td>62.7</td>
<td>65.3</td>
<td>30.2</td>
<td>71.6</td>
<td>58.2</td>
<td>62.2</td>
<td>54.3</td>
<td>40.7</td>
<td>49.2</td>
<td>50.0</td>
<td>66.6</td>
<td>62.9</td>
<td>83.3</td>
<td>34.2</td>
<td>48.2</td>
<td>46.1</td>
<td>83.4</td>
<td>65.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIRSTNIKON_AVGSVM</td>
<td>83.8</td>
<td>58.2</td>
<td>62.6</td>
<td>65.2</td>
<td>32.0</td>
<td>69.8</td>
<td>57.7</td>
<td>61.1</td>
<td>54.5</td>
<td>44.0</td>
<td>50.3</td>
<td>49.6</td>
<td>64.6</td>
<td>61.7</td>
<td>83.2</td>
<td>33.4</td>
<td>46.5</td>
<td>48.0</td>
<td>81.6</td>
<td>65.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIRSTNIKON_BOOSTSRKD</td>
<td>83.0</td>
<td>59.2</td>
<td>61.4</td>
<td>64.6</td>
<td>33.2</td>
<td>71.1</td>
<td>57.5</td>
<td>61.0</td>
<td>54.8</td>
<td>40.7</td>
<td>48.3</td>
<td>50.0</td>
<td>65.5</td>
<td>63.4</td>
<td>82.8</td>
<td>32.8</td>
<td>47.0</td>
<td>47.1</td>
<td>83.3</td>
<td>64.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIRSTNIKON_BOOSTSVM</td>
<td>83.5</td>
<td>56.8</td>
<td>61.8</td>
<td>65.5</td>
<td>33.2</td>
<td>69.7</td>
<td>57.3</td>
<td>60.5</td>
<td>54.6</td>
<td>43.1</td>
<td>48.3</td>
<td>50.3</td>
<td>64.3</td>
<td>62.4</td>
<td>82.3</td>
<td>32.9</td>
<td>46.9</td>
<td>48.4</td>
<td>82.0</td>
<td>64.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEAR_CHI-SVM-MULT_LOC</td>
<td>79.5</td>
<td>55.5</td>
<td>54.5</td>
<td>63.9</td>
<td>43.7</td>
<td>70.3</td>
<td>66.4</td>
<td>56.5</td>
<td>54.4</td>
<td>38.8</td>
<td>44.1</td>
<td>46.2</td>
<td>58.5</td>
<td>64.2</td>
<td>82.2</td>
<td>39.1</td>
<td>41.3</td>
<td>39.8</td>
<td>73.6</td>
<td>66.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NECUIUC_CDCV</td>
<td>88.1</td>
<td>66.0</td>
<td>68.0</td>
<td>72.5</td>
<td>41.0</td>
<td>78.9</td>
<td>70.4</td>
<td>70.4</td>
<td>58.1</td>
<td>53.4</td>
<td>55.7</td>
<td>59.3</td>
<td>73.1</td>
<td>71.3</td>
<td>84.5</td>
<td>32.3</td>
<td>53.3</td>
<td>56.7</td>
<td>86.0</td>
<td>66.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NECUIUC_CLS-DTCT</td>
<td>88.0</td>
<td>68.6</td>
<td>67.9</td>
<td>72.9</td>
<td>44.2</td>
<td>79.5</td>
<td>72.5</td>
<td>70.8</td>
<td>59.5</td>
<td>53.6</td>
<td>57.5</td>
<td>59.0</td>
<td>72.6</td>
<td>72.3</td>
<td>85.3</td>
<td>36.6</td>
<td>56.9</td>
<td>57.9</td>
<td>85.9</td>
<td>68.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NECUIUC_LL-CDCV</td>
<td>87.1</td>
<td>67.4</td>
<td>65.8</td>
<td>72.3</td>
<td>40.9</td>
<td>78.3</td>
<td>69.7</td>
<td>69.7</td>
<td>58.5</td>
<td>50.1</td>
<td>55.1</td>
<td>56.3</td>
<td>71.8</td>
<td>70.8</td>
<td>84.1</td>
<td>31.4</td>
<td>51.5</td>
<td>55.1</td>
<td>84.7</td>
<td>65.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NECUIUC_LN-CDCV</td>
<td>87.7</td>
<td>67.8</td>
<td>68.1</td>
<td>71.1</td>
<td>39.1</td>
<td>78.5</td>
<td>70.6</td>
<td>70.7</td>
<td>57.4</td>
<td>51.7</td>
<td>53.3</td>
<td>59.2</td>
<td>71.6</td>
<td>70.6</td>
<td>84.0</td>
<td>30.9</td>
<td>51.7</td>
<td>55.9</td>
<td>85.9</td>
<td>66.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UVASURREY_BASELINE</td>
<td>84.1</td>
<td>59.2</td>
<td>62.7</td>
<td>65.4</td>
<td>35.7</td>
<td>70.6</td>
<td>59.8</td>
<td>61.3</td>
<td>56.7</td>
<td>45.3</td>
<td>52.4</td>
<td>50.6</td>
<td>66.1</td>
<td>66.6</td>
<td>83.7</td>
<td>34.8</td>
<td>47.2</td>
<td>47.7</td>
<td>80.8</td>
<td>65.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UVASURREY_MKFDABOW</td>
<td>84.7</td>
<td>63.9</td>
<td>66.1</td>
<td>67.3</td>
<td>37.9</td>
<td>74.1</td>
<td>63.2</td>
<td>64.0</td>
<td>57.1</td>
<td>46.2</td>
<td>54.7</td>
<td>53.5</td>
<td>68.1</td>
<td>70.6</td>
<td>85.2</td>
<td>38.5</td>
<td>47.2</td>
<td>49.3</td>
<td>83.2</td>
<td>68.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UVASURREY_TUNCOLORKERNELSEL</td>
<td>85.0</td>
<td>62.8</td>
<td>65.1</td>
<td>66.5</td>
<td>37.6</td>
<td>73.5</td>
<td>62.1</td>
<td>62.0</td>
<td>57.4</td>
<td>45.1</td>
<td>54.5</td>
<td>52.5</td>
<td>67.7</td>
<td>69.8</td>
<td>84.8</td>
<td>39.1</td>
<td>46.8</td>
<td>49.9</td>
<td>82.9</td>
<td>68.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UVASURREY_TUNCOLORSPECKDA</td>
<td>84.6</td>
<td>62.4</td>
<td>65.6</td>
<td>67.2</td>
<td>39.4</td>
<td>74.0</td>
<td>63.4</td>
<td>62.8</td>
<td>56.7</td>
<td>43.8</td>
<td>54.7</td>
<td>52.7</td>
<td>67.3</td>
<td>70.6</td>
<td>85.0</td>
<td>38.8</td>
<td>46.9</td>
<td>50.0</td>
<td>82.2</td>
<td>66.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Only methods in 1st, 2nd or 3rd place by group shown
- Groups: CVC, FIRST/Nikon, NEC/UIUC, UVA/Surrey
Max AP: 88.1% (aeroplane) ... 40.8% (potted plant)
Set threshold on ‘detection’ to create one pair of precision / recall values.

Vary threshold across all values to generate precision / recall curves:

\[
\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}
\]

\[
\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}
\]
Precision/Recall: Aeroplane (All)
Precision/Recall: Potted plant (Top 10 by AP)
Ranked Images: Aeroplane

- Class images:
  Highest ranked
Ranked Images: Chair

- Class images: Highest ranked
Detection Challenge

- Predict the bounding boxes of all objects of a given class in an image (if any)
Evaluation

- **Average Precision [TREC]** averages precision over the entire range of recall
  - Curve interpolated to reduce influence of “outliers”

- A good score requires both high recall and high precision
  - Application-independent
  - Penalizes methods giving high precision but low recall
Evaluating Bounding Boxes

- Area of Overlap (AO) Measure

\[ \text{AO}(B_{gt}, B_p) = \frac{|B_{gt} \cap B_p|}{|B_{gt} \cup B_p|} \]

- Need to define a threshold $t$ such that $\text{AO}(B_{gt}, B_p)$ implies a correct detection: 50%
AP by Class

Chance essentially 0
Precision/Recall - Car
Precision/Recall – Potted plant
True Positives - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP

NECUIUC_CLS-DTCT
False Positives - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP

NECUIUC_CLS-DTCT
“Near Misses” - Person
True Positives - Bicycle

UoCTTI_LSVM-MDPM

OXFORD_MKL

NECUIUC_CLS-DTCT
False Positives - Bicycle

UoCTTI_L SVM-MDPM

OXFORD_MKL

NECUIUC_CLS-DTCT
PASCAL VOC: 2010-2014

NEC LLC
2009: 66.5%
2010: 73.8%
2011: 78.7%
2012: 82.2%
2013: 79.0%

Deep feature
2012: 82.2%
Deep feature
2013: 79.0%

Sub-category
2014: 83.2%

HCP
2014: 91.4%

Shuicheng Yan
Opportunities of Scale

Computer Vision

James Hays

Many slides from James Hays, Alyosha Efros, and Derek Hoiem

Graphic from Antonio Torralba
Computer Vision so far

- The geometry of image formation
  - Ancient / Renaissance
- Signal processing / Convolution
  - 1800, but really the 50’s and 60’s
- Hand-designed Features for recognition, either instance-level or categorical
  - 1999 (SIFT), 2003 (Video Google), 2005 (Dalal-Triggs), 2006 (spatial pyramid)
- Learning from Data
  - 1991 (EigenFaces) but late 90’s to now especially
What has changed in the last decade?

- The Internet
- Crowdsourcing
- Learning representations from the data these sources provide (deep learning)
Google and massive data-driven algorithms

A.I. for the postmodern world:

– all questions have already been answered...many times, in many ways
– Google is dumb, the “intelligence” is in the data
The Unreasonable Effectiveness of Data

Peter Norvig
Google
If a machine can convincingly simulate an intelligent conversation, does it necessarily understand? In the experiment, Searle imagines himself in a room, acting as a computer by manually executing a program that convincingly simulates the behavior of a native Chinese speaker.

Most of the discussion consists of attempts to refute it. "The overwhelming majority," notes BBS editor Stevan Harnad, "still think that the Chinese Room Argument is dead wrong." The sheer volume of the literature that has grown up around it inspired Pat Hayes to quip that the field of cognitive science ought to be redefined as "the ongoing research program of showing Searle's Chinese Room Argument to be false."
Questions from the piece:
Q1. Does the Chinese Room argument prove the impossibility of machine consciousness?
A1: Hell no. ... See More

Can Machines Become Moral?
The question is heard more and more often, both from those who think that machines cannot become moral, and who think that to believe otherwise is a dangerous illusion, and from those who think that machines must become moral....

BIGQUESTIONSONLINE.COM | BY DON HOWARD
Big Idea

• Do we need computer vision systems to have strong AI-like reasoning about our world?
• What if invariance / generalization isn’t actually the core difficulty of computer vision?
• What if we can perform high level reasoning with brute-force, data-driven algorithms?
Image Completion Example

[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]

http://graphics.cs.cmu.edu/projects/scene-completion/
What should the missing region contain?
Which is the original?

(a)

(b)

(c)
How it works

• Find a similar image from a large dataset
• Blend a region from that image into the hole
Hopefully, if you have enough images, the dataset will contain very similar images that you can find with simple matching methods.
How many images is enough?
Nearest neighbors from a collection of 20 thousand images
Nearest neighbors from a collection of 2 million images
Image Data on the Internet

• Flickr (as of Sept. 19th, 2010)
  – 5 billion photographs
  – 100+ million geotagged images

• Facebook (as of 2009)
  – 15 billion

Image Data on the Internet

• Flickr (as of Nov 2013)
  – 10 billion photographs
  – 100+ million geotagged images
  – 3.5 million a day

• Facebook (as of Sept 2013)
  – 250 billion+
  – 300 million a day

• Instagram
  – 55 million a day
Image completion: how it works

[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]
The Algorithm
Scene Matching
Scene Descriptor
Scene Descriptor

Scene Gist Descriptor (Oliva and Torralba 2001)
Scene Descriptor

Scene Gist Descriptor
(Oliva and Torralba 2001)
2 Million Flickr Images
Context Matching
Result Ranking

We assign each of the 200 results a score which is the sum of:

- The scene matching distance
- The context matching distance (color + texture)
- The graph cut cost
... 200 scene matches
Which is the original?