Attributes and More
Crowdsourcing

Computer Vision
CS 143, Brown

James Hays

Many slides from Derek Hoiem
Recap: Human Computation

• Active Learning: Let the classifier tell you where more annotation is needed.

• Human-in-the-loop recognition: Have a human and computer cooperate to do recognition.

• Mechanical Turk is powerful but noisy
  – Determine which workers are trustworthy
  – Find consensus over multiple annotators
  – “Gamify” your task to the degree possible
Recap: Data Sets

- ImageNet
  - Huge, Crowdsourced, Hierarchical, *Iconic* objects
- PASCAL VOC
  - *Not* Crowdsourced, bounding boxes, 20 categories.
- SUN Scene Database
  - *Not* Crowdsourced, 397 (or 720) scene categories
- LabelMe (Overlaps with SUN)
  - Sort of Crowdsourced, Segmentations, Open ended
- SUN *Attribute* database (Overlaps with SUN)
  - Crowdsourced, 102 attributes for every scene
PASCAL VOC Progress
Pascal VOC 2007 Average Precision

[Bar chart showing average precision (AP) for various categories such as person, train, car, aeroplane, horse, boat, motorbike, bicycle, bus, cat, bird, dining table, chair, tv monitor, sofa, dog, sheep, cow, potted plant, and bottle. The chart includes Max, Median, and Chance categories.]
Pascal VOC 2012 Average Precision
Describing Objects by their Attributes

Ali Farhadi, Ian Endres, Derek Hoiem, David Forsyth
CVPR 2009
What do we want to know about this object?
What do we want to know about this object?

Object recognition expert: “Dog”
What do we want to know about this object?

Object recognition expert: “Dog”

Person in the Scene: “Big pointy teeth”, “Can move fast”, “Looks angry”
Our Goal: Infer Object Properties

Can I **poke with it?**

Is it **alive?**

Can I **put stuff in it?**

What **shape** is it?

Is it **soft?**

Does it have a **tail?**

Will it **blend?**
Why Infer Properties

1. We want detailed information about objects

“Dog” vs. “Large, angry animal with pointy teeth”
Why Infer Properties

2. We want to be able to infer something about unfamiliar objects
Why Infer Properties

2. We want to be able to infer something about unfamiliar objects

If we can infer category names…

Familiar Objects  New Object

Cat  Horse  Dog  ???
Why Infer Properties

2. We want to be able to infer something about unfamiliar objects

If we can infer properties...

Familiar Objects

- Has Stripes
- Has Ears
- Has Eyes
- Has Four Legs
- Has Mane
- Has Tail
- Has Snout

Brown
Muscular

New Object

- Has Stripes (like cat)
- Has Mane and Tail (like horse)
- Has Snout (like horse and dog)
Why Infer Properties

3. We want to make comparisons between objects or categories

What is unusual about this dog?

What is the difference between horses and zebras?
Strategy 1: Category Recognition

Object Image

classifier

Category

“Car”

associated properties

Has Wheels
Used for Transport
Made of Metal
Has Windows
…

Category Recognition: PASCAL 2008
Category → Attributes: ??
Strategy 2: Exemplar Matching

Object Image

Similar Image

Has Wheels
Used for Transport
Made of Metal
Old
...

similarity function
associated properties

Malisiewicz Efros 2008
Hays Efros 2008
Efros et al. 2003
Strategy 3: Infer Properties Directly

Object Image

classifier for each attribute

No Wheels
Old
Brown
Made of Metal
...

See also Lampert et al. 2009
Gibson’s affordances
The Three Strategies

Object Image

Category
“Car”
associated properties

Similar Image

Has Wheels
Used for Transport
Made of Metal
Has Windows
Old
No Wheels
Brown
…

Direct
classifier for each attribute

classifier

similarity function

associated properties
Our attributes

- Visible parts: “has wheels”, “has snout”, “has eyes”

- Visible materials or material properties: “made of metal”, “shiny”, “clear”, “made of plastic”

- Shape: “3D boxy”, “round”
Attribute Examples

**Shape:** Horizontal Cylinder
**Part:** Wing, Propeller, Window, *Wheel*
**Material:** *Metal*, Glass

**Shape:**
**Part:** Window, *Wheel*, Door, Headlight, Side Mirror
**Material:** *Metal*, Shiny
Attribute Examples

**Shape:**
**Part:** Head, Ear, Nose, Mouth, Hair, Face, Torso, Hand, Arm
**Material:** Skin, Cloth

**Shape:**
**Part:** Head, Ear, Snout, Eye
**Material:** Furry

**Shape:**
**Part:** Head, Ear, Snout, Eye, Torso, Leg
**Material:** Furry
Our approach

Feature extraction → Feature Selection

Attribute Predictions

Category Models

Bird

Has Beak, Has Eye, Has foot, Has Feather

Attribute Classifiers
Features

Strategy: cover our bases

• Spatial pyramid histograms of quantized
  – Color and texture for materials
  – Histograms of gradients (HOG) for parts
  – Canny edges for shape
Learning Attributes

• Learn to distinguish between things that have an attribute and things that do not
• Train one classifier (linear SVM) per attribute
Experiments

• Predict attributes for unfamiliar objects

• Identify what is unusual about an object
Describing Objects by their Attributes

No examples from these object categories were seen during training
No examples from these object categories were seen during training
Identifying Unusual Attributes

• Look at predicted attributes that are not expected given class label
Absence of typical attributes

- Aeroplane: No "wing"
- Car: No "window"
- Boat: No "sail"
- Aeroplane: No "jet engine"
- Motorbike: No "side mirror"
- Car: No "door"
- Sheep: No "wool"

752 reports
68% are correct
Presence of atypical attributes

Motorbike "cloth"
People "label"
Bird "Leaf"
Bus "face"

Aeroplane "beak"
Sofa "wheel"
Bike "Horn"

951 reports
47% are correct
Two Crowdsourced Recognition Databases

SUN Attribute Database

Sketched Object Database
Space of Scenes
Space of Scenes

- Ice Cave
- Cavern
- Forest
- Volcano
- Savanna
- Beach
- Canyon
- River
- Dentist’s Office
- Classroom
- Subway
- Village
- Highway
- Fountain
- Railroad
- Canal
Space of Scenes

[Diagram showing different environments connected by lines: Forest, Cavern, Volcano, Savanna, Beach, Classroom, Village, Railroad, Ice Cave, Savanna, Ray]

[Images of a forest, a savanna, and a tree]
Space of Scenes

- Ice Cave
- Cavern
- Forest
- Savanna
- Beach
- Railroad
- Canyon

- Village
- River
- Beach
Space of Scenes

Dentist’s Office
Classroom
Cavern
Subway
Highway
Fountain
Canyon
Canal
Railroad
River

?
Space of Scenes
Big Picture

• Scenes don’t fit neatly into categories.
  – Objects often do!

• Categories aren’t expressive enough.

• We should reason about scene *attributes* instead of (or in addition to) scene categories.
Attribute-based Visual Understanding

Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer.
Lampert, Nickisch, and Harmeling. CVPR 2009.

Describing Objects by their Attributes.
Farhadi, Endres, Hoiem, Forsyth. CVPR 2009.

Attribute and Simile Classifiers for Face Verification.
Kumar, Berg, Belhumeur, Nayar. ICCV 2009.

Numerous more recent works on activity, texture, 3d models, etc.
• Spatial layout: large, enclosed
• Affordances / functions: can fly, park, walk
• Materials: shiny, black, hard
• Object presence: has people, ships
• Simile: looks like Star Trek
• Emotion: scary, intimidating
Which Scene Attributes are Relevant?

Inspired by the “splitting” task of Oliva and Torralba and “ESP game” by von Ahn and Blum.

Which attributes distinguish the scenes on the left from the scenes on the right?

rock, warm, barren, natural
102 Scene Attributes
Scene Attribute Labeling

Click on the scenes below that contain the following lighting or material:

**camping**: Either an actual camp site, or scene in wilderness suitable enough for humans to make a tent and/or sleep.

These HITs are reviewed before being approved or rejected.

*For further instructions Click Here!*

This task can be very subjective. If you are not sure about which images should be selected, please **skip this HIT** or email us to ask for clarification. There are more HITs with less subjective attributes.

Images continued down the page...
SUN Attributes: A Large-Scale Database of Scene Attributes

http://www.cs.brown.edu/~gen/sunattributes.html

Global, binary attributes describing:
- Affordances / Functions (e.g. farming, eating)
- Materials (e.g. carpet, running water)
- Surface Properties (e.g. aged, sterile)
- Spatial Envelope (e.g. enclosed, symmetrical)

Statistics of database:
- 14,340 images from 717 scene categories
- 102 attributes
- 4 million+ labels
- good workers ~92% accurate
- pre-trained classifiers for download
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Images given 0 votes</th>
<th>Images given 1 vote</th>
<th>Images given 2 votes</th>
<th>Images given 3 votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camping</td>
<td><img src="image1.png" alt="Images" /></td>
<td><img src="image2.png" alt="Images" /></td>
<td><img src="image3.png" alt="Images" /></td>
<td><img src="image4.png" alt="Images" /></td>
</tr>
<tr>
<td>Diving</td>
<td><img src="image5.png" alt="Images" /></td>
<td><img src="image6.png" alt="Images" /></td>
<td><img src="image7.png" alt="Images" /></td>
<td><img src="image8.png" alt="Images" /></td>
</tr>
<tr>
<td>Medical Activity</td>
<td><img src="image9.png" alt="Images" /></td>
<td><img src="image10.png" alt="Images" /></td>
<td><img src="image11.png" alt="Images" /></td>
<td><img src="image12.png" alt="Images" /></td>
</tr>
<tr>
<td>Cluttered Space</td>
<td><img src="image13.png" alt="Images" /></td>
<td><img src="image14.png" alt="Images" /></td>
<td><img src="image15.png" alt="Images" /></td>
<td><img src="image16.png" alt="Images" /></td>
</tr>
</tbody>
</table>
102 dimensional attribute space reduced to 2d with t-SNE
Sailing
Instances of the “15 Scene” Categories
Average Precision of Attribute Classifiers
## Attribute Recognition

<table>
<thead>
<tr>
<th>Test Scene Images</th>
<th>Highest Confidence Attributes with Confidence Values</th>
<th>Lowest Confidence Attributes with Confidence Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Golf Course" /></td>
<td>0.74 vegetation</td>
<td>-1.33 studying</td>
</tr>
<tr>
<td></td>
<td>0.63 open area</td>
<td>-1.36 gaming</td>
</tr>
<tr>
<td></td>
<td>0.60 sunny</td>
<td>-1.38 fire</td>
</tr>
<tr>
<td></td>
<td>0.57 sports</td>
<td>-1.42 carpet</td>
</tr>
<tr>
<td></td>
<td>0.55 natural light</td>
<td>-1.60 tiles</td>
</tr>
<tr>
<td></td>
<td>0.52 no horizon</td>
<td>-1.60 smoke</td>
</tr>
<tr>
<td></td>
<td>0.51 foliage</td>
<td>-1.65 medical</td>
</tr>
<tr>
<td></td>
<td>0.49 competing</td>
<td>-1.67 cleaning</td>
</tr>
<tr>
<td></td>
<td>0.46 railing</td>
<td>-1.71 sterile</td>
</tr>
<tr>
<td></td>
<td>0.46 natural</td>
<td>-1.74 marble</td>
</tr>
<tr>
<td><img src="image2.png" alt="Restaurant" /></td>
<td>0.91 eating</td>
<td>-1.07 gaming</td>
</tr>
<tr>
<td></td>
<td>0.89 socializing</td>
<td>-1.11 running water</td>
</tr>
<tr>
<td></td>
<td>0.70 waiting in line</td>
<td>-1.19 tiles</td>
</tr>
<tr>
<td></td>
<td>0.51 cloth</td>
<td>-1.27 railroad</td>
</tr>
<tr>
<td></td>
<td>0.42 shopping</td>
<td>-1.35 waves/surf</td>
</tr>
<tr>
<td></td>
<td>0.42 reading</td>
<td>-1.36 building</td>
</tr>
<tr>
<td></td>
<td>0.39 stressful</td>
<td>-1.37 fire</td>
</tr>
<tr>
<td></td>
<td>0.39 congregating</td>
<td>-1.40 bathing</td>
</tr>
<tr>
<td></td>
<td>0.37 man-made</td>
<td>-1.50 ice</td>
</tr>
<tr>
<td></td>
<td>0.31 plastic</td>
<td>-1.63 smoke</td>
</tr>
<tr>
<td>Category</td>
<td>Confident Classifications</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td>Competing</td>
<td>![Competing Images]</td>
<td></td>
</tr>
<tr>
<td>Farming</td>
<td>![Farming Images]</td>
<td></td>
</tr>
<tr>
<td>Metal</td>
<td>![Metal Images]</td>
<td></td>
</tr>
<tr>
<td>Cold</td>
<td>![Cold Images]</td>
<td></td>
</tr>
<tr>
<td>Eating</td>
<td>![Eating Images]</td>
<td></td>
</tr>
</tbody>
</table>
Most Confident Classifications

Moist/Damp

Natural

Stressful

Vacationing

Praying