Describing Objects by their Attributes

- A. Farhadi, I. Endres, D. Hoiem and D. Forsyth

Aashish Sheshadri

19th October 2012

Motivation

- What is Recognition ?
  - Is it identifying object names given a static frame ?
    - If yes, how do we decide on object categories ?
      - Reaching a consensus on object categories.
- Do we really need object categories ?
  - Maybe not!
- Changing perspective …
  - Traditional : Where is It ?
  - Recent : What is it like ? - Recognition by association.
  - This paper : What is it ? What can it be ? - Recognition by describing attributes.

Related Work

Recognition by Association via Learning Per-Exemplar Distances

- Tomasz Malisiewicz and Alexei A. Efros

Learning Visual Attributes

- Vittorio Ferrari and Andrew Zisserman

Natural Scene Retrieval based on a Semantic Modeling Step

- Julia Vogel and Bernt Schiele

Learning to Recognize Activities from the Wrong View Point

- Ali Farhadi and Mostafa Kamali Tabrizi

Why Attributes

To Re-Cognize

To make descriptions

To make inferences

"Cat"

vs.

"Large, angry animal with pointy teeth"

Inferring Attributes

Category

classifier

“Car”

associated properties

Object Image

Similar Image

similarity function

associated properties

Direct

classifier for each attribute

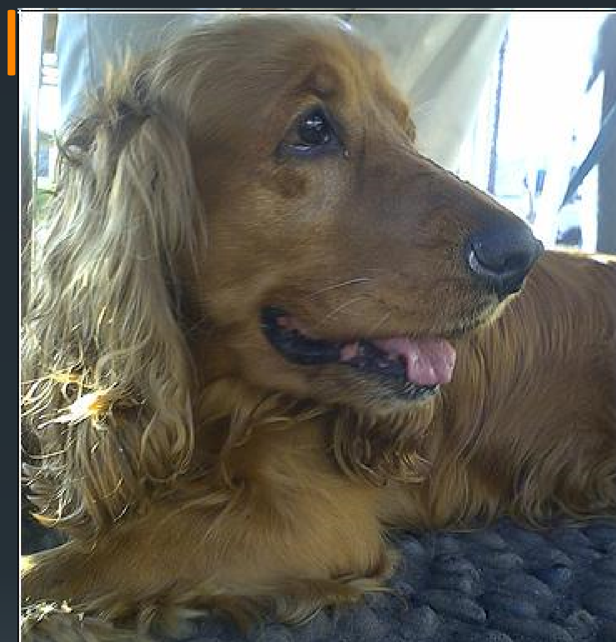
Has Wheels
Used for Transport
Made of Metal
Has Windows
Old
No Wheels
Brown
…

Original slides by Derek Hoiem - http://www.cs.illinois.edu/homes/dhoiem/

Attributes

- Semantic 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”
- Discriminative Attributes
  - Random Splits
  - Train by selecting subset of classes and features
    - Dogs vs. sheep using color
    - Cars and buses vs. motorbikes and bicycles using edges

antic Attribute Exampl

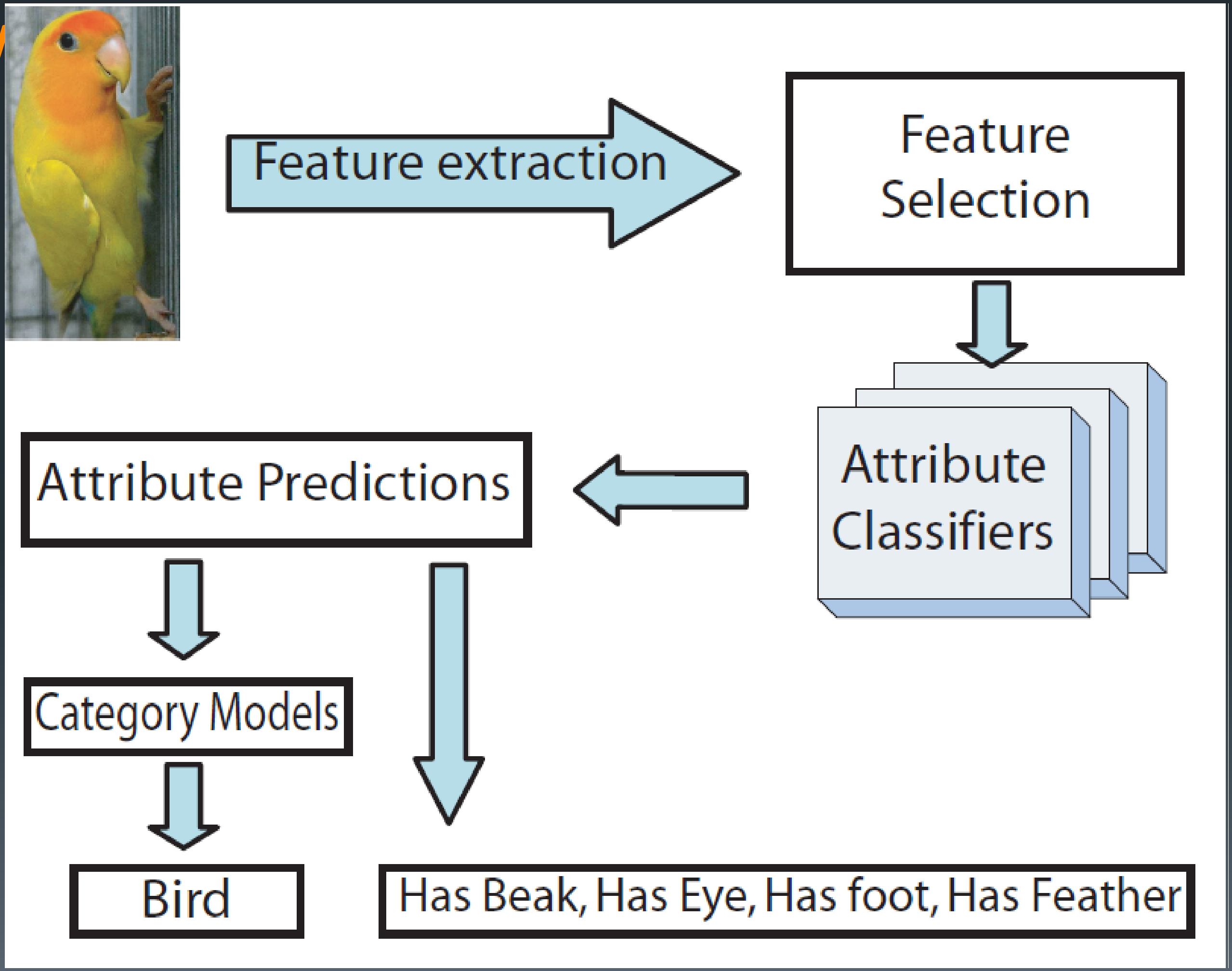


Shape:
Part: Head, Ear, Nose, Mouth, Hair, Face, Torso, Hand, Arm
Material: Skin, Cloth

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

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

Flow


Feature extraction
Feature Selection
Attribute Classifiers
Attribute Predictions
Category Models
Bird
Has Beak, Has Eye, Has foot, Has Feather

Features

Spatial pyramid histograms of quantized

- Color (LAB) and texture (Texton) for materials
- Histograms of gradients (HOG) for parts
- Canny edges for shape
- 9751 Dimensional -> 7 Histograms for each feature type (128 + 256 + 1000 + 9).
- Feature vector reflects distribution only within bounding box.

Learning Attributes

Simplest approach: Train classifier using all features for each attribute independently

“Has Wheels”

“No Wheels Visible”

Original slides by Derek Hoiem - http://www.cs.illinois.edu/homes/dhoiem/

Dealing with Correlated Attributes



Most things that "have wheels" are "made of metal"

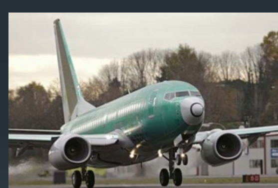
Learning "has wheels", may accidentally learn "made of metal"!

Has Wheels, Made of Metal?

Feat lecti

 vs. 

 vs. 

 vs. 

"Has Wheels" "No Wheels"

Feature selection (L1 logistic regression) for each class separately and pool features

Car Wheel Features

Boat Wheel Features

Plane Wheel Features

All Wheel Features

Experiments

Predicting attributes for unfamiliar objects

Learning new categories

- From limited examples
- From text description alone

Identifying what is unusual about an object

Across category generalization

Datasets

- a-Pascal
  - 20 categories from PASCAL 2008 trainval dataset (10K object images)
    - airplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, sofa, train, tv monitor
  - Ground truth for 64 attributes
  - Annotation via Amazon's Mechanical Turk
- a-Yahoo
  - 12 new categories from Yahoo image search
    - bag, building, carriage, centaur, donkey, goat, jet ski, mug, monkey, statue of person, wolf, zebra
  - Categories chosen to share attributes with those in Pascal, but different correlation statistics!
- Attribute labels are somewhat ambiguous
  - Agreement among "experts" 84.3
  - Between experts and Turk labelers 81.4
  - Among Turk labelers 84.1

a - Pascal

Chair

3D-Boxy Occluded
Furn-Leg Plastic

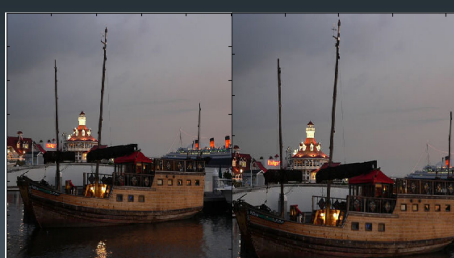
Person

Head Ear Hair Face
Eye Torso Hand
Arm Leg Foot/Shoe
Skin Cloth

Aeroplane

3D-Boxy Round Horiz-Cyl
Occluded Wing Jet-engine
Window Row-Wind Wheel
Door Text Metal Shiny

3D-Boxy Occluded
Furn-Leg Plastic



Boat

Tail Beak Head Eye Torso
Leg Foot/Shoe Feather

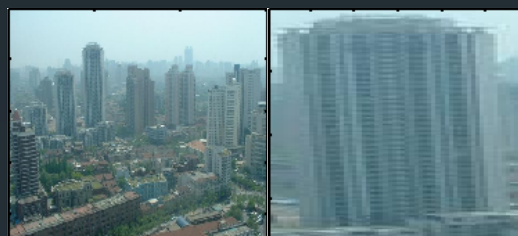
Bird

Vrt-Cyl Leaf Stem/Trunk
Pot Vegetation

Potted Plant

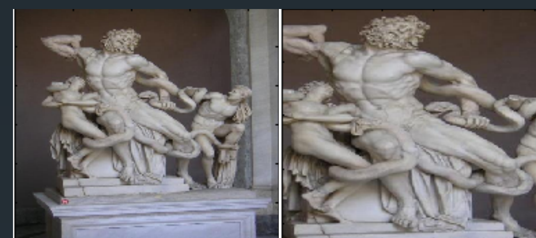
a - Yahoo

Building



2D-Boxy Window Row Wind Metal Glass Shiny

Statue



Head Nose Mouth Face Eye Torso Hand Arm Leg Foot/Shoe

Centaur

Tail Head Ear Hair Face Eye Torso Hand Arm Leg Foot/Shoe Wing Horn Rei

Tail Head Ear Snout Eye Torso Leg Foot/Shoe Horn Furry

Goat

3D-Boxy Vert-Cyl Metal Plastic Shiny

Mug

2D Boxy Horiz-Cyl Metal Shiny Leather

Bag

Predicting attributes

- Train on 20 object classes from a-Pascal train set
  - Feature selection for each attribute
  - Train a linear SVM classifier

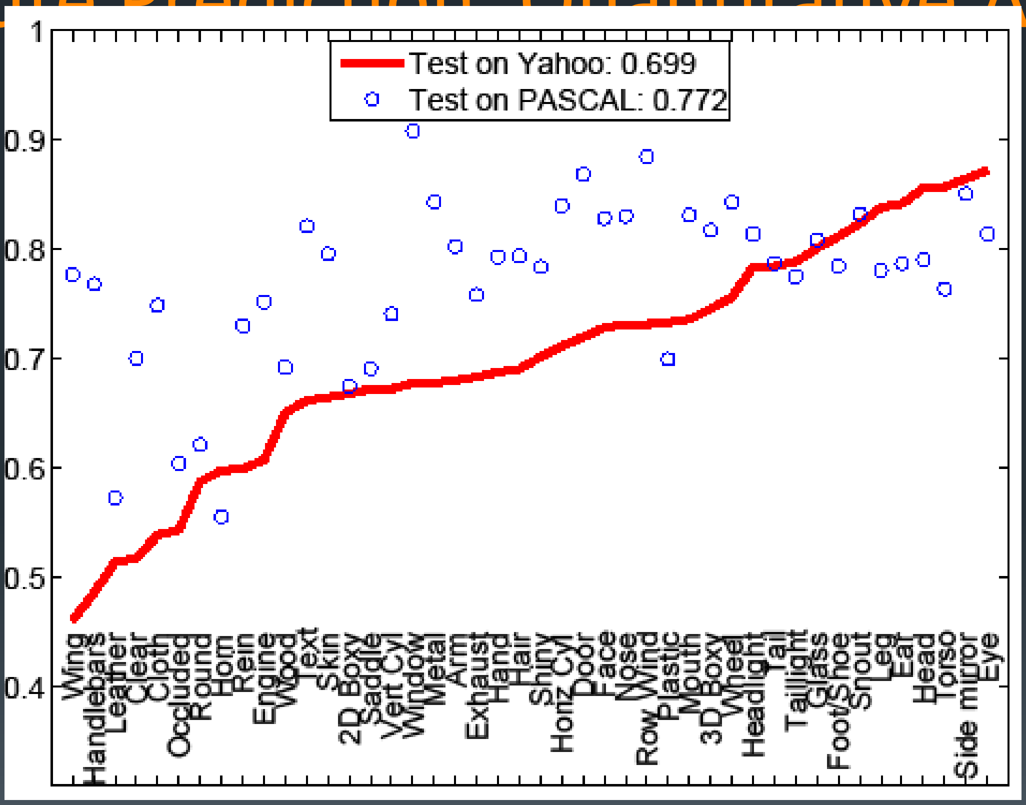
- Test on 12 object classes from Yahoo image search (cross-category) or on a-Pascal test set (within-category)
  - Apply learned classifiers to predict each attribute

Describing Objects by their Attributes

No examples from these object categories were seen during training

Attribute Prediction: Quantitative Analysis

Worst
Wing
Handlebars
Leather
Clear
Cloth



Area Under the ROC for Familiar (PASCAL) vs. Unfamiliar (Yahoo) Object Classes

Best
Eye
Side Mirror
Torso
Head
Ear

Average ROC Area

Test Objects	Parts	Materials	Shape
a-PASCAL	0.794	0.739	0.739
a-Yahoo	0.726	0.645	0.677

Trained on a-PASCAL objects

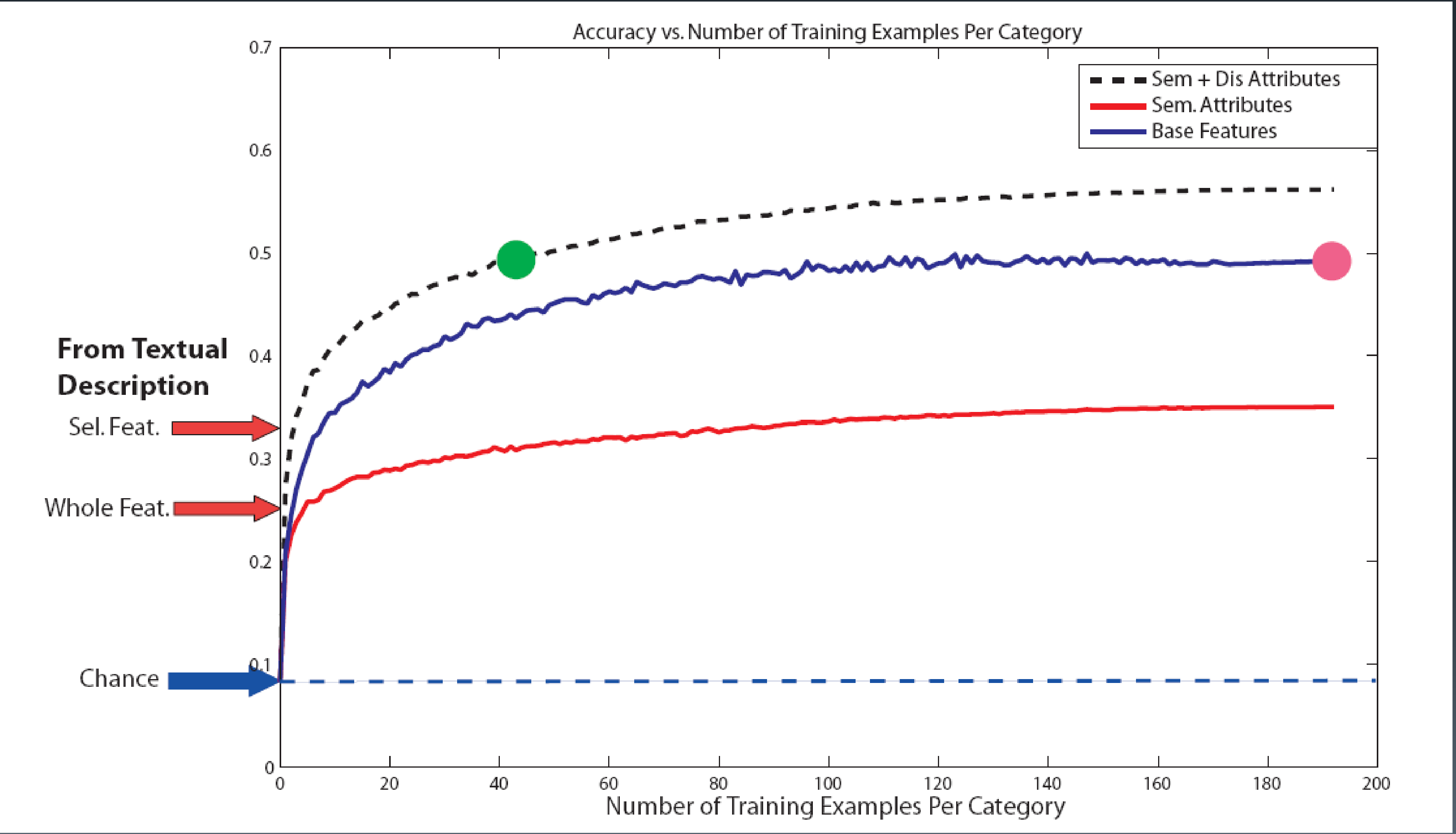
Category Recognition

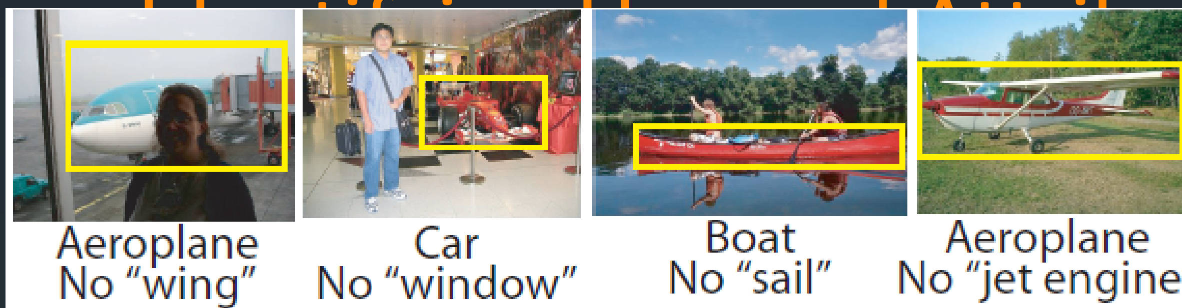
- Attribute predictions as features
- Linear SVM trained to categorize object each object
- Discriminative attributes
  - Train 10,000 and select 1,000 most reliable, according to a validation set

PASCAL 2008	Base Features	Semantic Attributes	All Attributes
Classification Accuracy	58.5%	54.6%	**59.4%**
Class-normalized Accuracy	35.5%	28.4%	**37.7%**

Learning New Categories

- Limited examples
  - Nearest neighbor of attribute predictions
- From textual description
  - nearest neighbor to verbally specified attributes
    - Goat: “has legs, horns, head, torso, feet”, “is furry”
    - Building: “has windows, rows of windows”, “made of glass, metal”, “is 3D boxy”

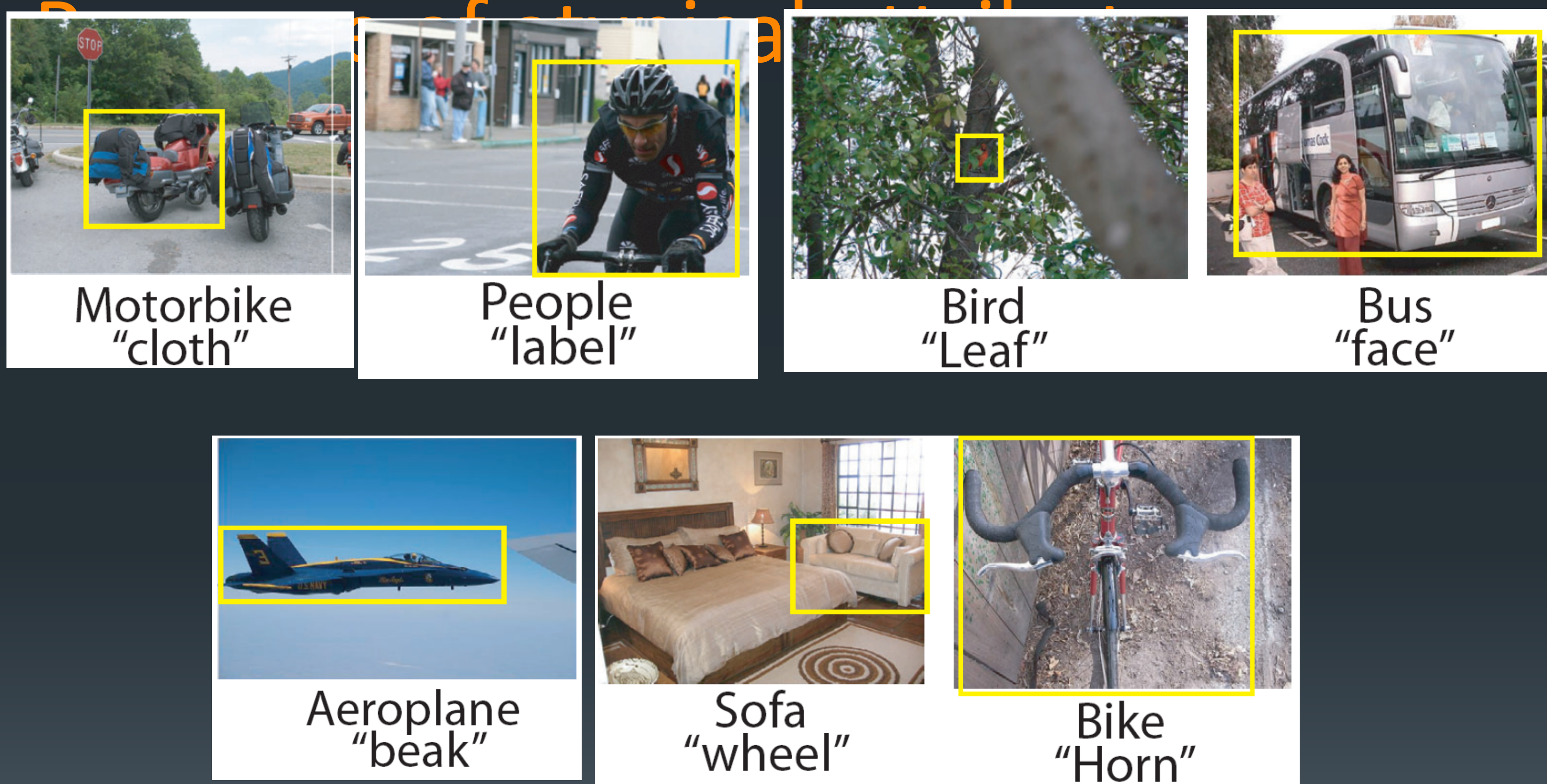

Accuracy vs. Number of Training Examples Per Category
Sem + Dis Attributes
Sem. Attributes
Base Features
From Textual Description
Sel. Feat.
Whole Feat.
Chance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
20
40
60
80
100
120
140
160
180
200
Number of Training Examples Per Category



Absence of typical attributes

752 reports

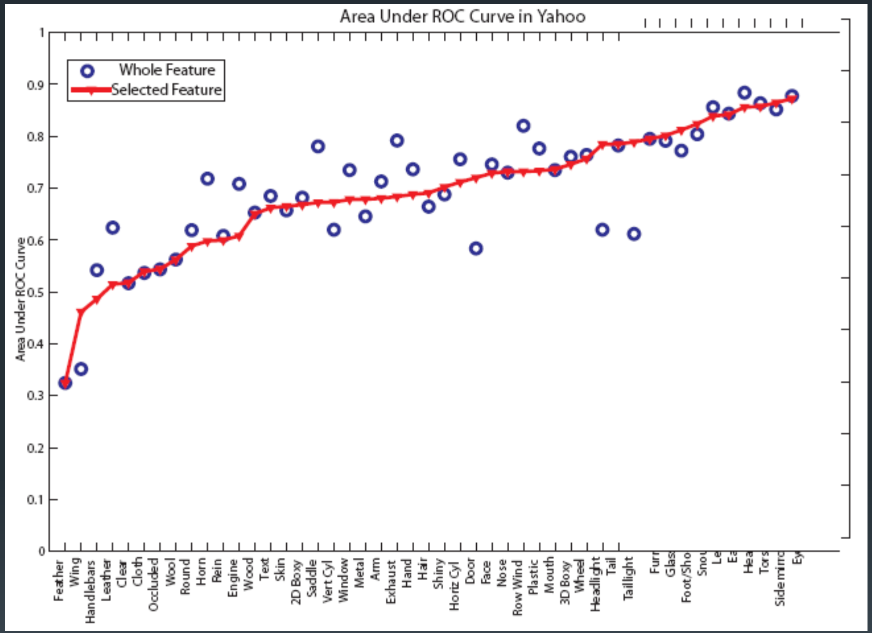
68% are correct

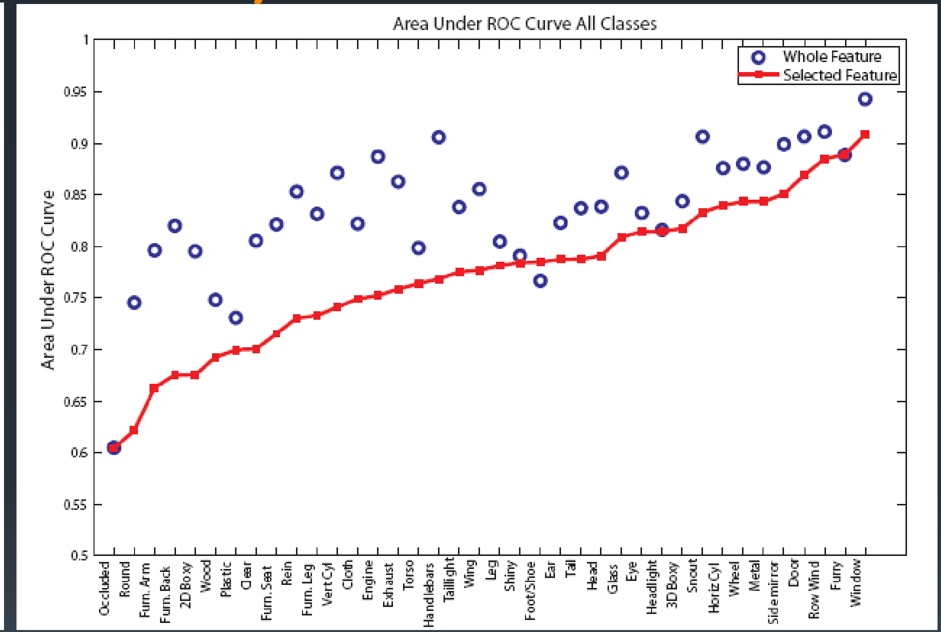


951 reports

47% are correct

Better Semantics vs Accuracy





Train on 20 PASCAL classes
Test on 12 different Yahoo classes

Train and Test on
Same Classes from PASCAL

Extensions

- Comprehensive set of attributes
- Multiple strategies for predicting attributes
- Probabilistic inference to use a subset of attribute classifiers
- Use of context to enable descriptive attributes and priming
- Infer object relationships and use through attributes
- Relative attributes!
- Where is it ? What is it like? What is it?
  - Answering - What is it doing here ? What can I do with it? Can this be important ?

Discussion

Feature Selection - Do we need it if the scene is segmented and annotated ?

A better way to learn attributes ? Using a bounding box seems unfair.

Material, texture is sensitive to lighting - same attribute might not be true for all instances

“Discriminative Attributes” seems similar to learning without attributes!

Comparison with classification results using a Linear SVM seems unfair.

Use of attributes should complement traditional object class recognition.

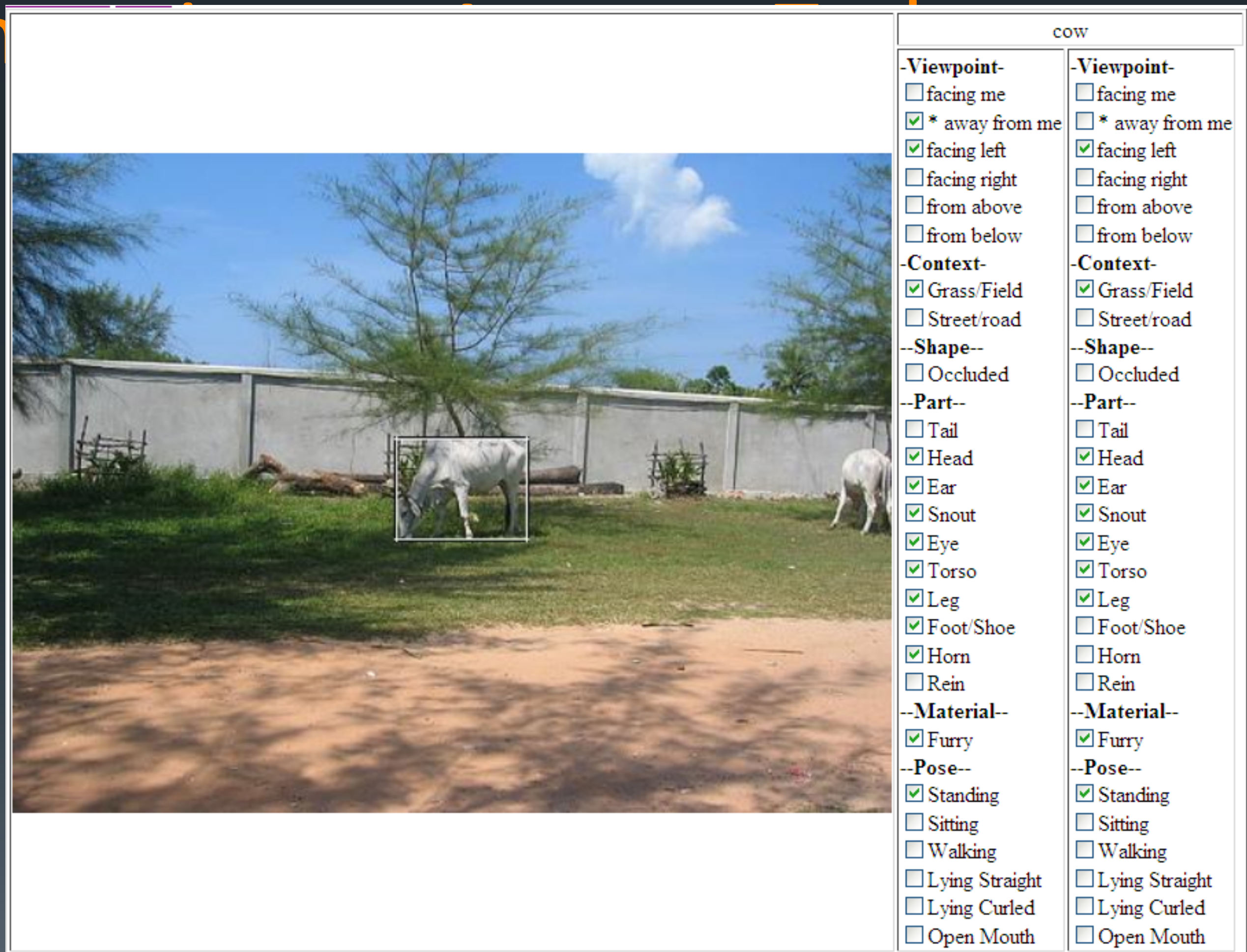
Conclusion

- Inferring object properties should be an important goal of object recognition
- Learning attributes enables several new abilities
  - Predicting properties of new types of objects
  - Identifying unusual about a familiar object
  - Learning from verbal description
- Raises an important issue concerning dataset biases while learning

Thank You!

Additional Slides

Ann


cow
-Viewpoint-
facing me
* away from me
facing left
facing right
from above
from below
-Context-
Grass/Field
Street/road
--Shape--
Occluded
--Part--
Tail
Head
Ear
Snout
Eye
Torso
Leg
Foot/Shoe
Horn
Rein
--Material--
Furry
--Pose--
Standing
Sitting
Walking
Lying Straight
Lying Curled
Open Mouth
-Viewpoint-
facing me
* away from me
facing left
facing right
from above
from below
-Context-
Grass/Field
Street/road
--Shape--
Occluded
--Part--
Tail
Head
Ear
Snout
Eye
Torso
Leg
Foot/Shoe
Horn
Rein
--Material--
Furry
--Pose--
Standing
Sitting
Walking
Lying Straight
Lying Curled
Open Mouth

Describing Objects by their Attributes

No examples from these object categories were seen during training