The attributes of objects

D.A. Forsyth, UIUC
channelling Derek Hoiem, UIUC,
with Ali Farhadi, Ian Endres, Gang Wang all of UIUC

Obtain dataset

Build features

Mess around with classifiers, probability,

Produce representation

Computer vision

Obtain dataset

Build features

Light entertainment
(the way we do it)

Mess around with classifiers, probability,

Computer vision

Produce representation

Big questions

Computer vision

- What signal representation should we use ?

PLUMBING

Computer vision

- What should we say about visual data?

What should recognition do?

- Report objects present
- Make useful reports about objects
  - which likely involve categories
- Categories
  - allow generalization
    - future behavior; non-visual properties of objects
  - are opportunistic, rather than fixed
    - one person's intra class variation is another's class boundary
  - likely don't form an inclusion hierarchy
  - visual categorization vs. other categorization

Good properties of recognition

- Bias robust
  - biases, sparsity in training data don't affect test behaviour (much)
- Unfamiliarity
  - Make useful statements about objects whose name isn't yet known
- Manage deviant objects
  - Say how a detected object is different from the usual
- Learn by X
  - Single picture
  - Reading
    - Description (0 pictures; zero shot learning)
- Accuracy
  - be good at recognizing known objects

Big questions

Computer vision

- What signal representation should we use ?

PLUMBING

MODELS

Computer vision

- What should we say about visual data?

Taxonomy

Recognition - desirable properties

- Bias robust
  - biases in training data don’t affect test behaviour (much)
- Unfamiliarity
  - Make useful statements about objects whose name isn’t yet known
- Manage deviant objects
  - Say how a detected object is different from the usual
- Learn by X
  - Single picture
  - Reading
    - Description (0 pictures; zero shot learning)
- Accuracy
  - be good at recognizing known objects

Fallacy

If you know your problem well

you can collect an unbiased dataset

Should not be perjorative

Bias

- Frequencies in the data may misrepresent the application
  - Because the labels are often wrong — Label error
  - Because of what gets labelled — Label bias
    - P(labelled|X) is not uniform
    - eg obscure but important objects in complex clutter
    - eg pedestrians in crowds
  - Because of what gets collected — Curation bias
    - eg. pictures from the web are selected - not like a camera on head
    - eg. "Profession" labelling for faces in news pictures

X=data

Labels that are wrong

- Fact of life
- Can fix when there are many instances
  - consistency (Zhao et al 08)
  - smoothing (Berg, 06; Li, 06; Wang 08; Collins 08)
- Might be able to fix with hierarchy+generalization
  - we should never mix up "cat"'s and "truck"'s

Selection for labelling

- P(labelled|X) is not uniform
  - or P(X|labelled) is not the same as P(X|not labelled)

X_j are not like X_i

- There are models
  - problem sometimes called dataset shift, see (Quinonero-Candela 09)
  - can be addressed with, say, large unlabelled datasets
    - build smoothed estimate of p(labelled|X), reweight
- Important effect
  - can make high capacity classifiers generalize better than low capacity
  - (maybe) be very cautious about linear SVM's

Curation bias

Loeff et al, 06

- Collected data is not a fair sample of X
  - labelled AND unlabelled data
- Images on the web are “curated”
- Iconography seems to be a big effect here
  - visual “modes” of representation
    - see Berg+Berg 09
  - we might not see them all
  - cf Google image search with Flickr

X=data

Web Images Videos Maps News Shopping Gmail more ▼ Search settings | Sign in

Google

lion Search SafeSearch off ▼

About 23,100,000 results (0.05 seconds) Advanced search

Everything
Images
Videos
More

Any size
Medium
Large
Icon
Larger than...
Exactly...

Any type
Face
Photo
Clip art
Line drawing

Any color
Full color
Black and white

Related searches: lion roaring lioness lion drawing lion tattoo


Lions Kill Giraffe
479 × 450 - 48k - jpg
abolitionist.com
Find similar images


Lion on Horseback
468 × 393 - 39k - jpg
raincoaster.com
Find similar images


3, Lion
434 × 341 - 41k - jpg
bluepyramid.org
Find similar images


Interestingly, the
470 × 324 - 30k - jpg
bostonherald.com
Find similar images


Description : Aslan
792 × 768 - 99k - jpg
photocase.org
Find similar images

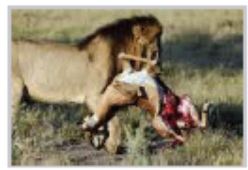

I was doing research on
400 × 300 - 27k - jpg
lowkayhwa.com
Find similar images


Lion Tiger Size
500 × 553 - 65k - jpg
indrajit.wordpress.com
Find similar images


Lion Park, South
450 × 300 - 30k - jpg
africa-nature-photog...
Find similar images


Lion Limited
500 × 500 - 76k - jpg
onlineartdemos.co.uk
Find similar images


Lion
395 × 480 - 47k - jpg
ibexinc.wordpress.com
Find similar images


lions
1200 × 800 - 243k - jpg
lifeasastudentnurse...
Find similar images


African Lion
500 × 333 - 57k - jpg
itsnature.org
Find similar images


LIONS:
604 × 800 - 225k - jpg
edge.org
Find similar images


Lion. Panthera leo
459 × 480 - 35k - jpg
shoarns.com
Find similar images


lions, cuddle
620 × 400 - 70k - jpg
telegraph.co.uk
Find similar images


lion
350 × 504 - 28k - jpg
sodahead.com
Find similar images


LION!
500 × 385 - 74k - jpg
firemice.wordpress.com
Find similar images

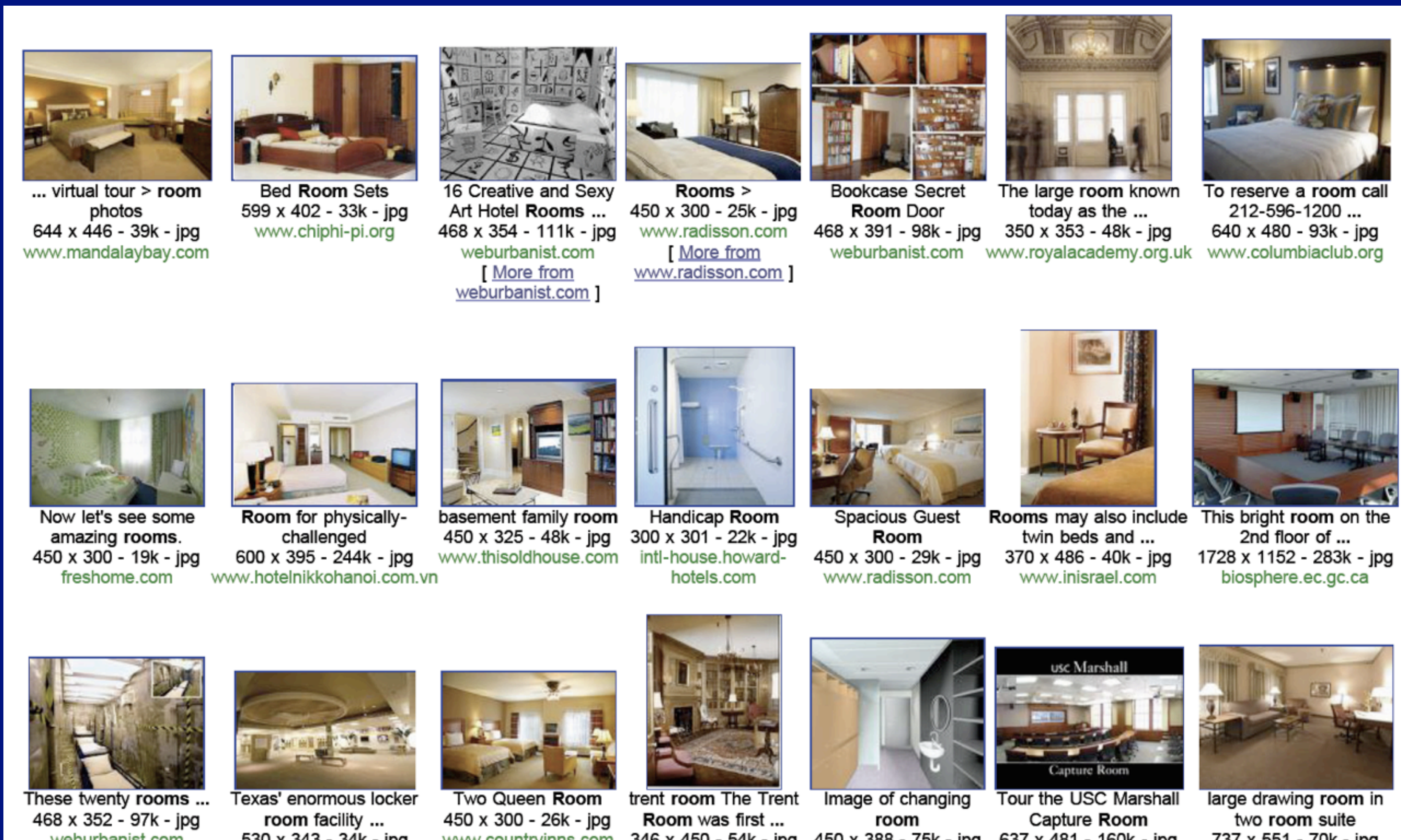

Starring horse-riding
800 × 626 - 53k - jpg
dailymail.co.uk
Find similar images


Picture: 17 stone
468 × 602 - 93k - jpg
dailymail.co.uk
Find similar images


human-lion
470 × 324 - 31k - jpg
seesdifferent...
Find similar images

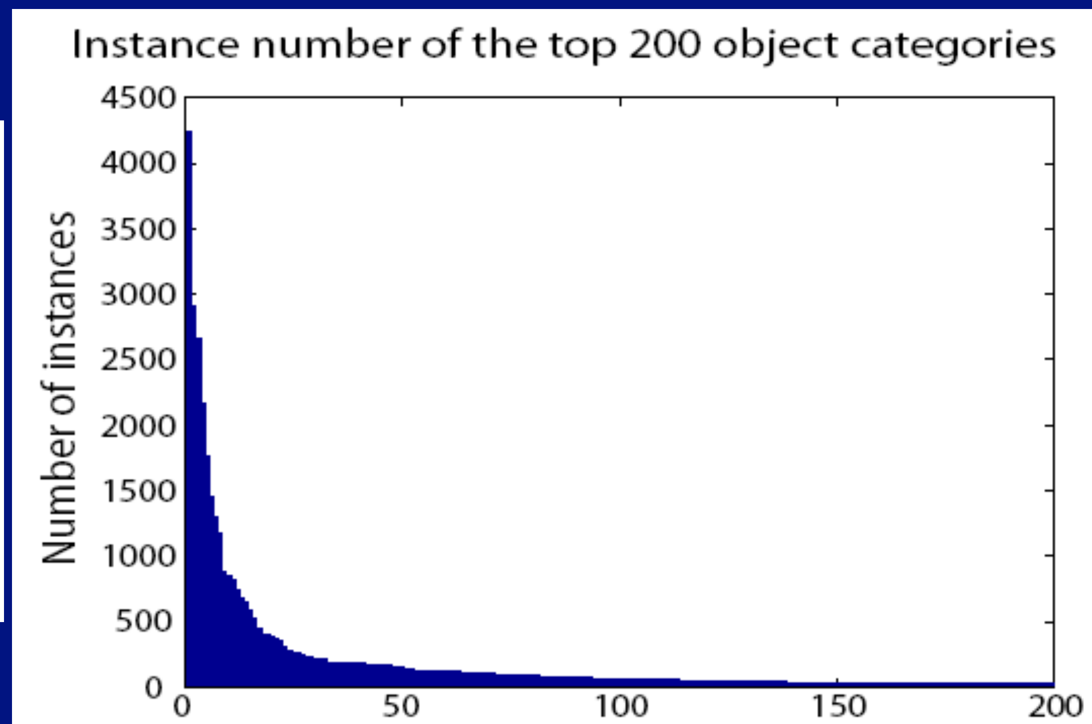

Lion at Sunset
400 × 318 - 25k - jpg
art.com
Find similar images

Google “rooms”



Flickr “rooms”

Many things are rare



Wang et al, 10
frequencies of labels in labelme dataset

Sparsity and within class variation

- Variation within classes has some meaningful structure
  - big cars vs small cars; big dogs vs small dogs
  - blue cars vs yellow cars; blue dogs vs yellow dogs
- Cannot be treated as pure variance with few examples
- Perverse to treat as pure variance

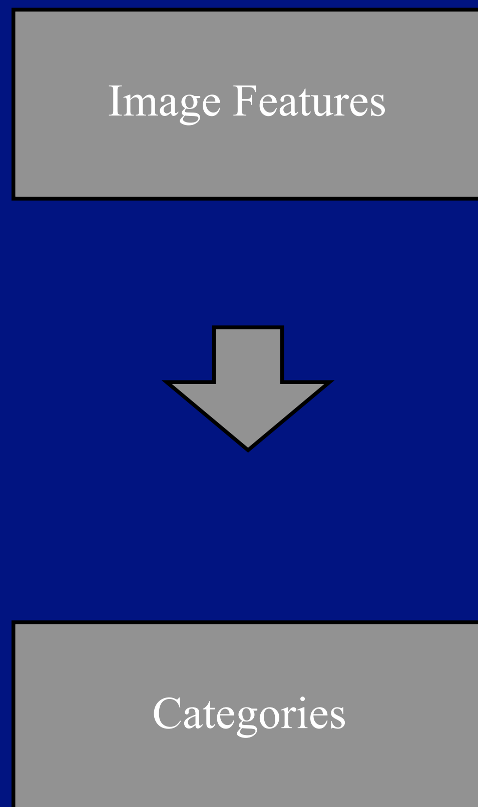
Induction

- Fundamental principle of machine learning
  - if the world is like the dataset, then future performance will be like training
    - Chernoff bounds, VC dimension, etc., etc.
- But what if the world can't be like the dataset?

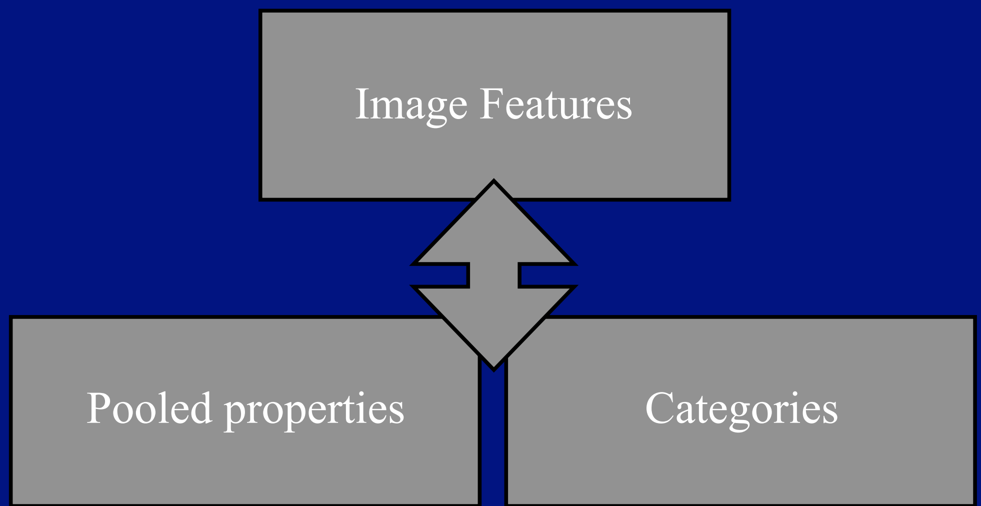
Object recognition

- The world can’t be like the training data because
  - many things are rare in plausible datasets
  - within class variation can’t be properly represented for each class
- Strategies
  - Ensure training data fairly represents the future
  - train by comparison to similar objects
  - Try only to learn things that are fairly represented
  - represent in terms of pooled properties

Old way



Bias suggests



Attributes

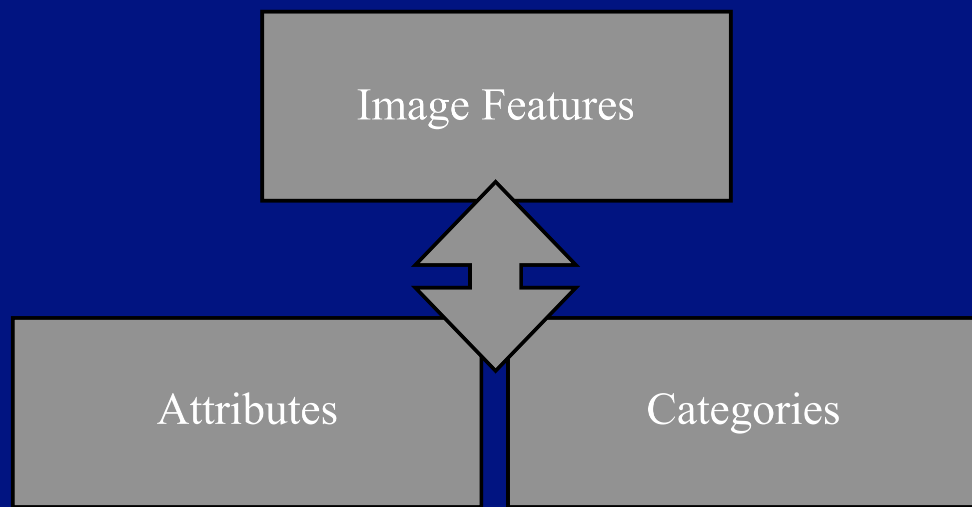
- Properties shared by many object categories
  - with explicit, exposed semantics

- Material (like)
  - glass, wood, furry, red, etc.
- Part (like)
  - has wheel, has head, has tail, etc.
- Shape (like)
  - is 2D Boxy, is cylindrical, etc

cf Ferrari Zisserman 07

NOT Latent Variables - Semantics is explicit, exposed

Architectural consequence



Recognition - desirable properties

Represent in terms of pooled properties

Pooled properties are attributes

- Bias robust
  - biases in training data don’t affect test behaviour (much)
- Unfamiliarity
  - Make useful statements about objects whose name isn’t yet known
- Manage deviant objects
  - Say how a detected object is different from the usual
- Learn by X
  - Single picture
  - Reading
    - Description (0 pictures; zero shot learning)
- Accuracy
  - be good at recognizing known objects

Attribute phenomena

- Some are easily predicted from pictures
  - eg “red”, “wooden”
- Some are properly inherited from category
  - eg “mammal”
- They are heavily correlated
  - easy binary variable argument
- Some are “stuff”-like
  - eg “red”, “wooden”
- Others “thing”-like
  - eg “wheel”, “leg”
- Within class variation
  - Different instances of the same category could have different attributes

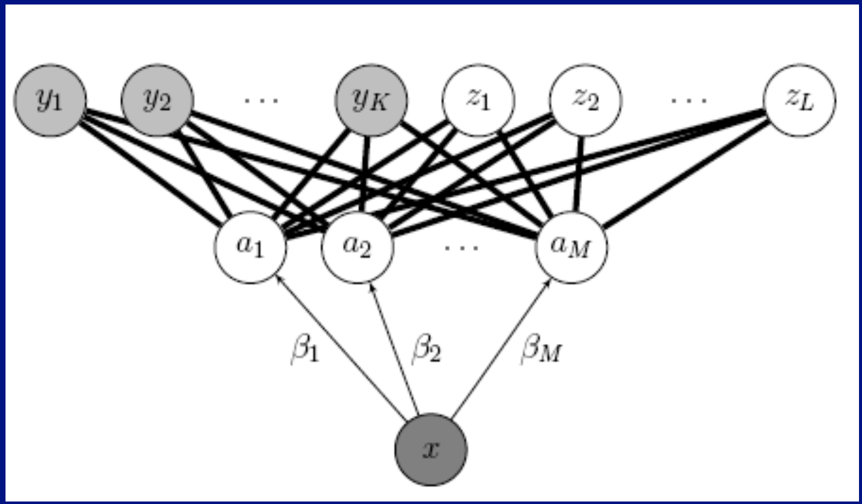
“Stuff” -- shape doesn’t matter (sky, grass, bush)
cf mass noun

“Thing” -- shape matters (cow, cat, car)
cf count noun

Direct Attribute Prediction

Known classes

Unknown classes



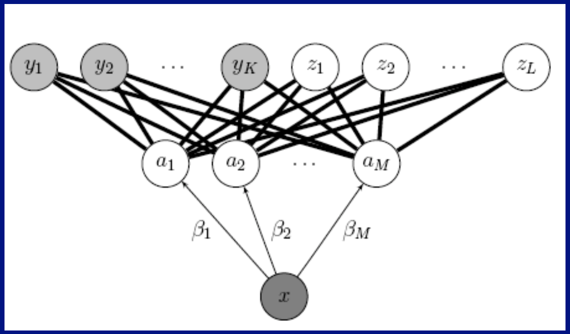
Attribute layer

Image features

Lampert ea 09; Farhadi ea 09

Stuff attributes

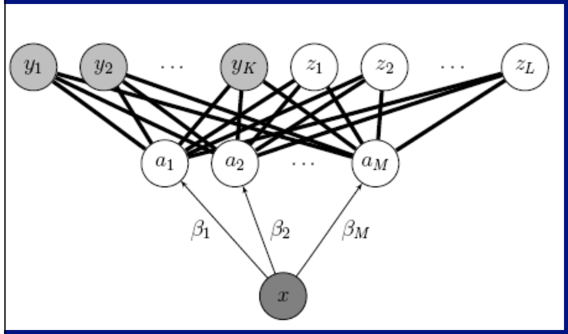
Direct Attribute Prediction



- Training
  - Lampert ea
    - objects labelled, attributes inherited from object labels
  - Farhadi ea
    - attributes labelled in images
- Architecture
  - Lampert ea
    - undirected object attribute links
    - deterministic links
  - Farhadi ea
    - directed attribute -> object links

Lampert ea 09; Farhadi ea 09

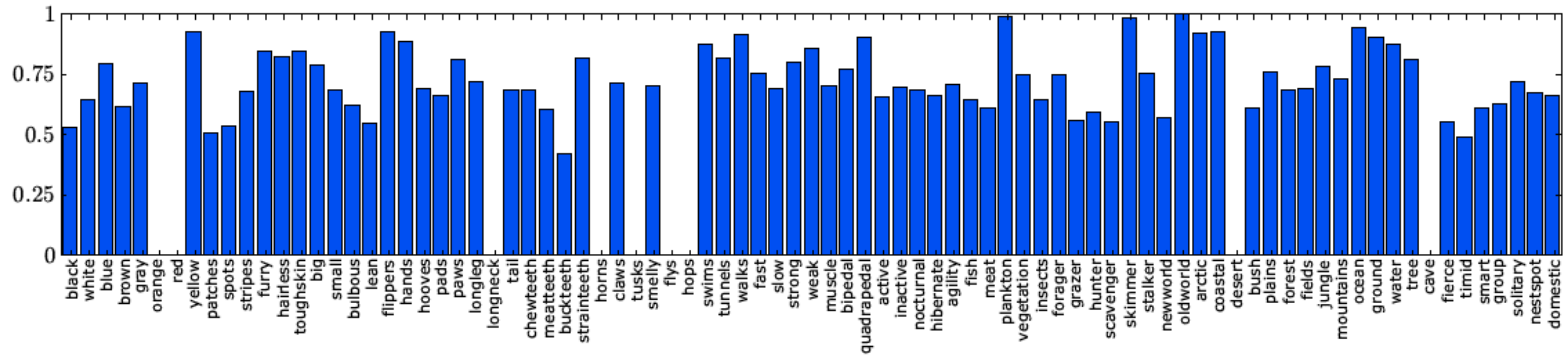
Direct Attribute Prediction



- Attractions
  - Pooling allows improved generalization of attributes
  - learning by X (few examples; description)
  - sensible statements about the unfamiliar
  - accuracy (evidence complex, but supportive)
- Inherited vs observed training
  - inherited: easier labelling
  - observed: cleaner semantics
- Disadvantage
  - only for directly visual attributes

Lampert ea 09; Farhadi ea 09

Farhadi ea 09



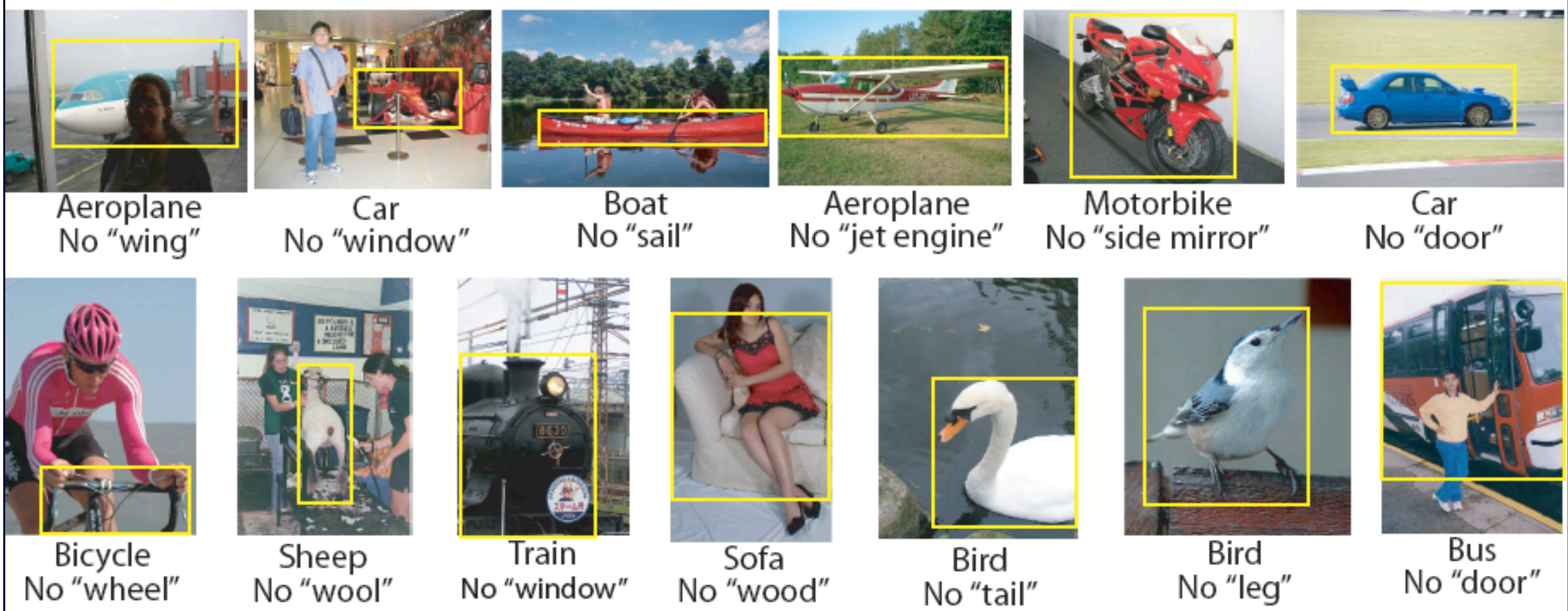
Lampert ea 09

Object categories in test set are not same categories as in training set

How is an object different from typical?

- Pragmatics suggests this is how adjectives are chosen
  - If we are sure it’s a cat, and we know that
    - an attribute is different from normal
    - the detector is usually reliable
  - we should report the missing/extra attribute

Missing attributes



Farhadi ea 09

Extra attributes

Farhadi ea 09

Indirect Direct Attribute Prediction

Unknown classes

Attribute layer

Known classes

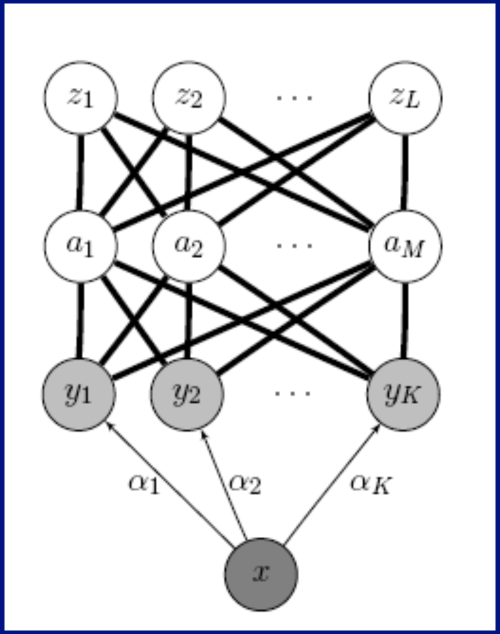
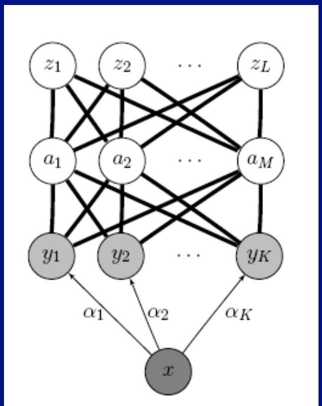


Image features

Lampert ea 09

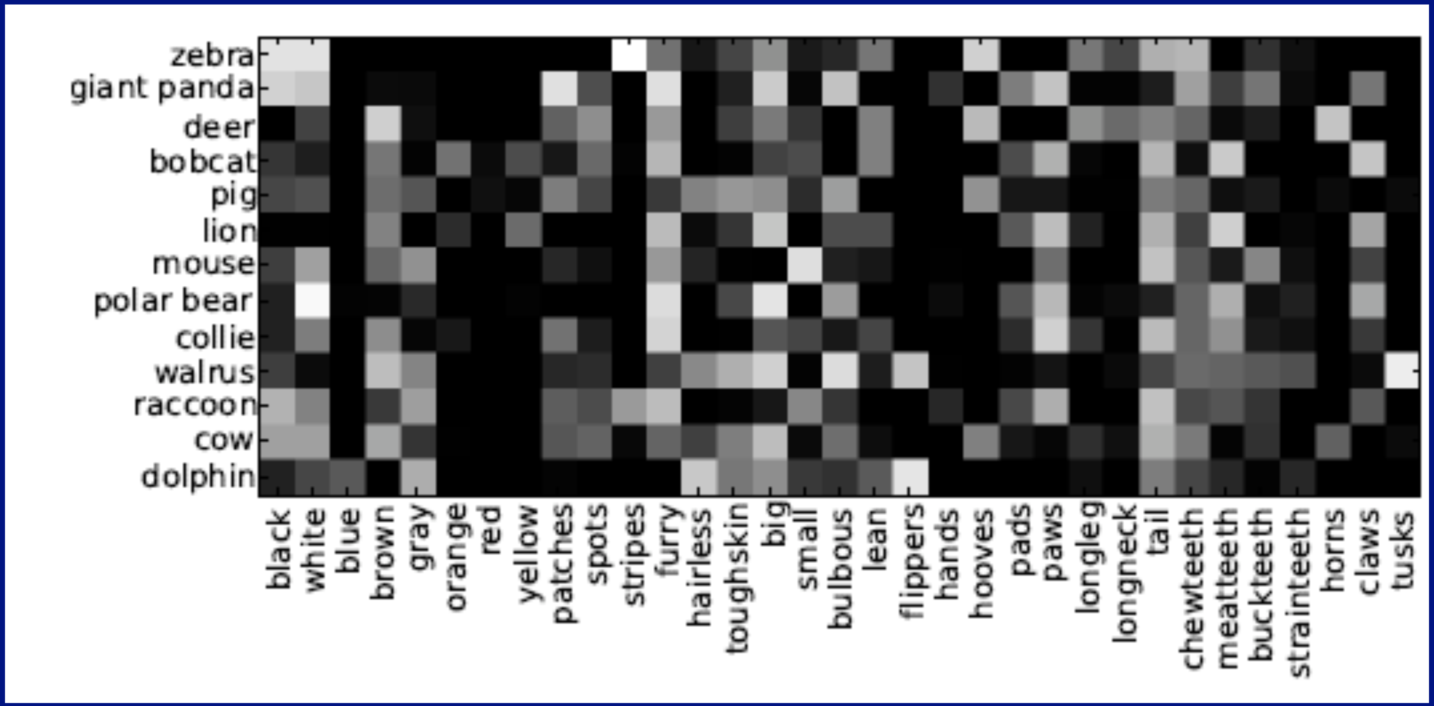
Stuff attributes

Indirect Attribute Prediction



- Training
  - learn predictors for known classes, usual procedure
  - y-a, a-z links from object semantics
    - all instances of a class have the same attribute vector
- Test
  - inference
- Property:
  - attributes from class predictions
    - so non-visual prediction should be OK
  - attribute predictions are “like” natural attribute vectors

Attribute Correlations



Lampert ea 09 after Osherson ea 91; Kemp ea 06

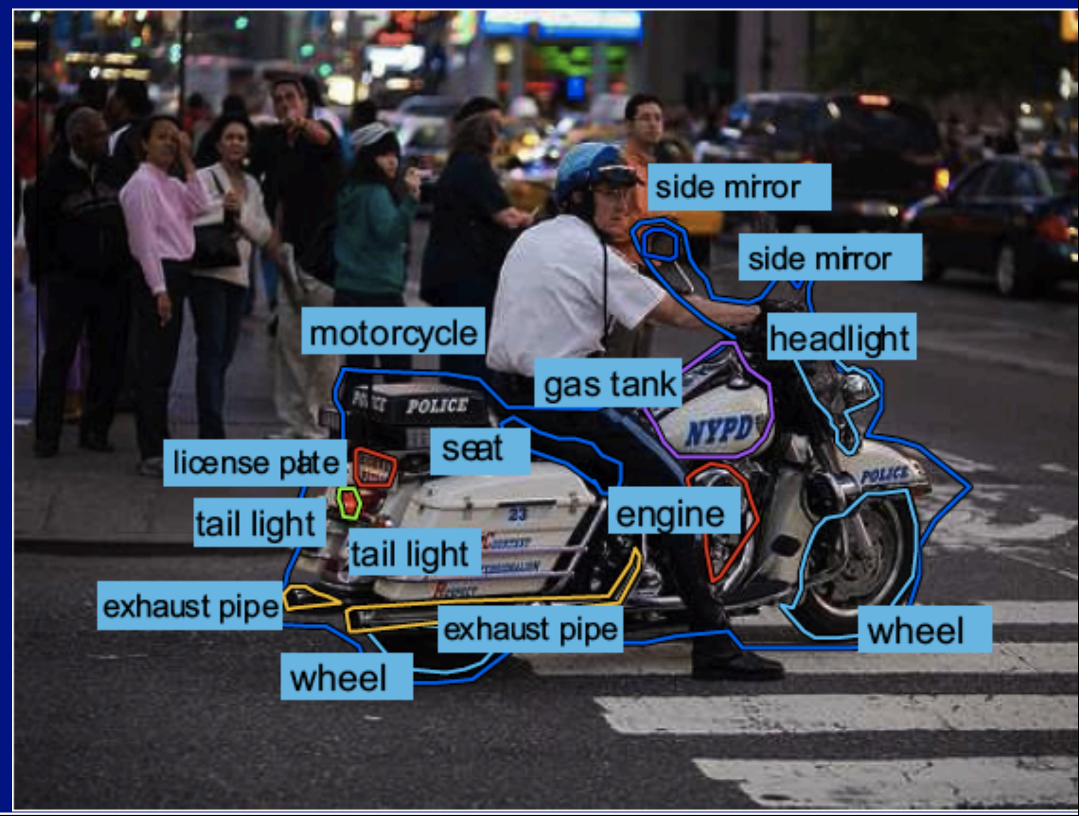
Attribute vectors as bit vectors

- N binary attributes => 2^N attribute vectors
- N should be large
  - so attributes must be heavily correlated
  - how to model?
    - indirect attribute prediction
    - latent variable models

Thing attributes

- Parts
  - in the old fashioned sense, as having semantics
    - “leg”, “wheel”, etc.
- Improved representation of localized objects
- Detection

Farhadi ea 10



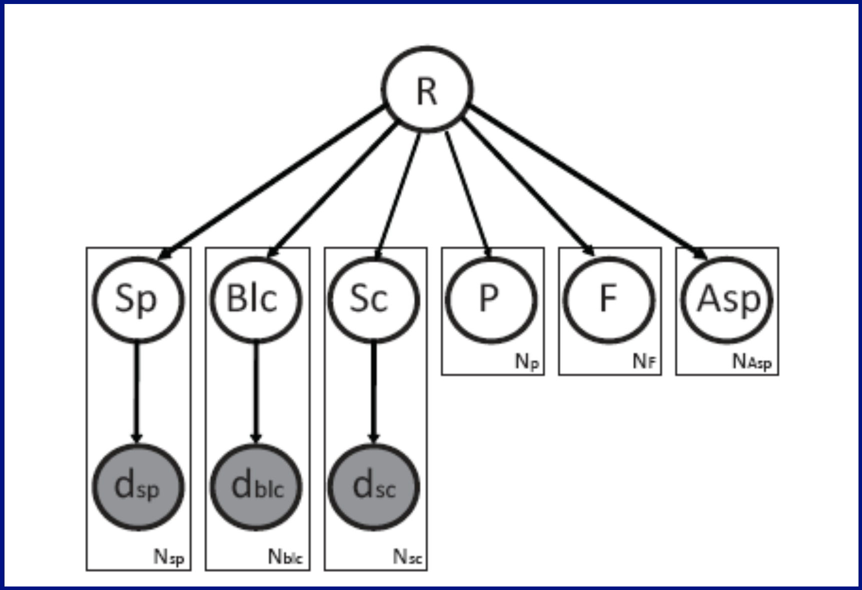
Latent Root

Root

Visual attributes

Other attributes

Detector Responses



Sp: spatial part (gridded location)
Blc: basic level category
Sc: superordinate category

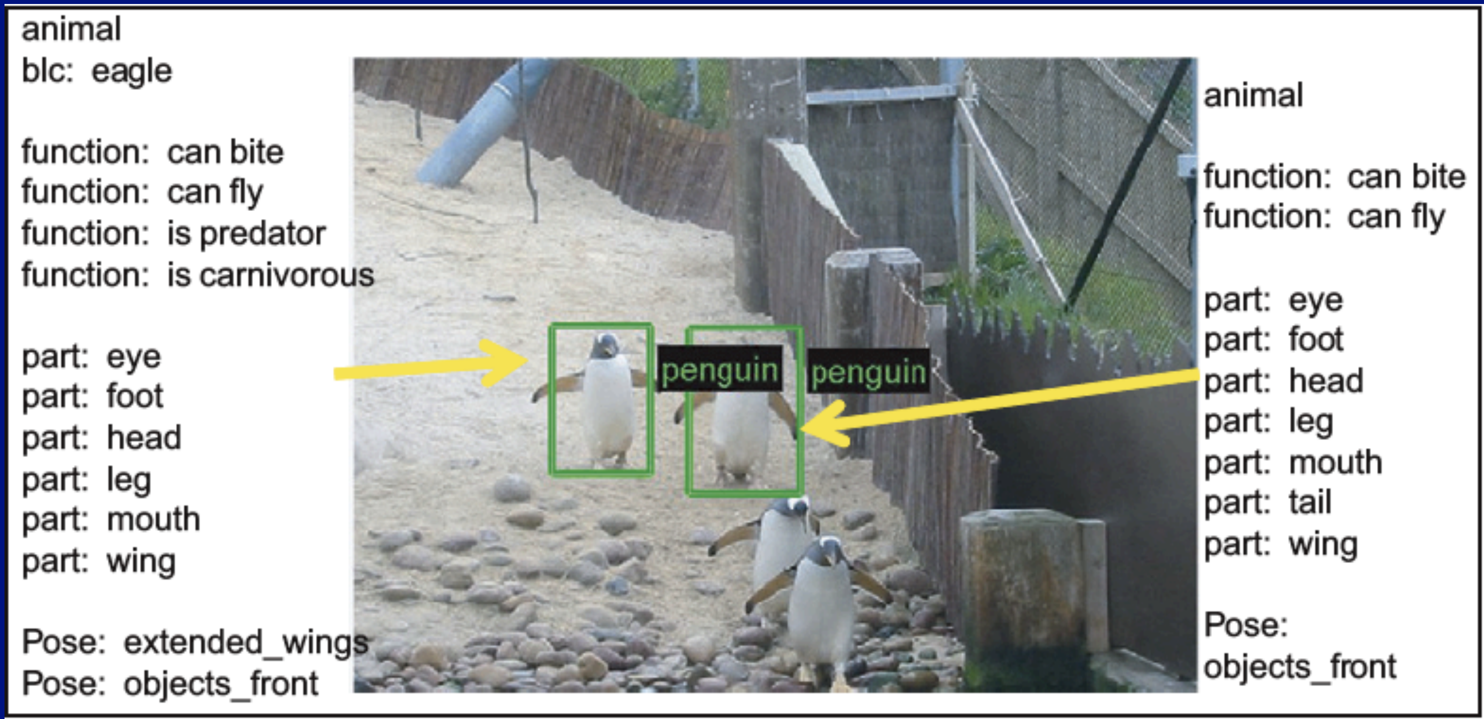
Farhadi ea 10

P: predicate
F: functional attribute
Asp: aspect

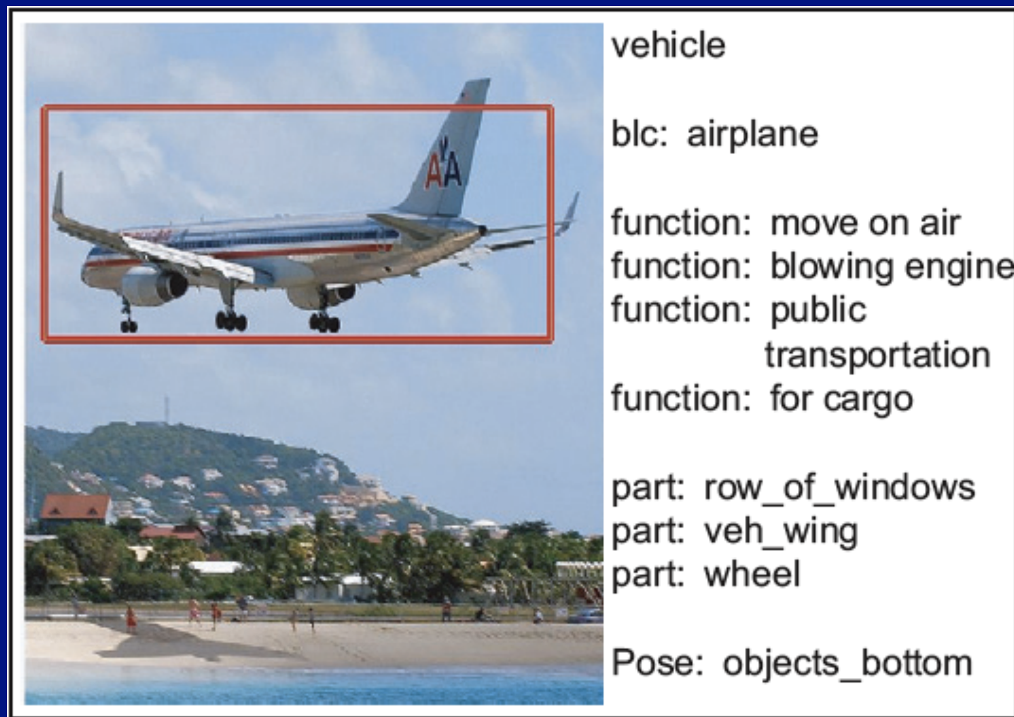
Roots improve prediction

Domain	Method	Average		Has Part		Basic-Cat		Super-Cat		Function		Pose	
		F	UnF	F	UnF	F	UnF	F	UnF	F	UnF	F	UnF
Animal	Root Model	0.757	0.646	0.798	0.747	0.755	NA	0.761	0.591	0.807	0.602	0.665	0.649
	Baseline	0.701	0.591	0.770	0.648	0.721	NA	0.710	0.618	0.732	0.567	0.571	0.532
Vehicle	Root Model	0.854	0.700	0.929	0.752	0.885	NA	0.891	0.778	0.922	0.691	0.643	0.578
	Baseline	0.781	0.652	0.870	0.723	0.841	NA	0.849	0.717	0.801	0.637	0.544	0.533

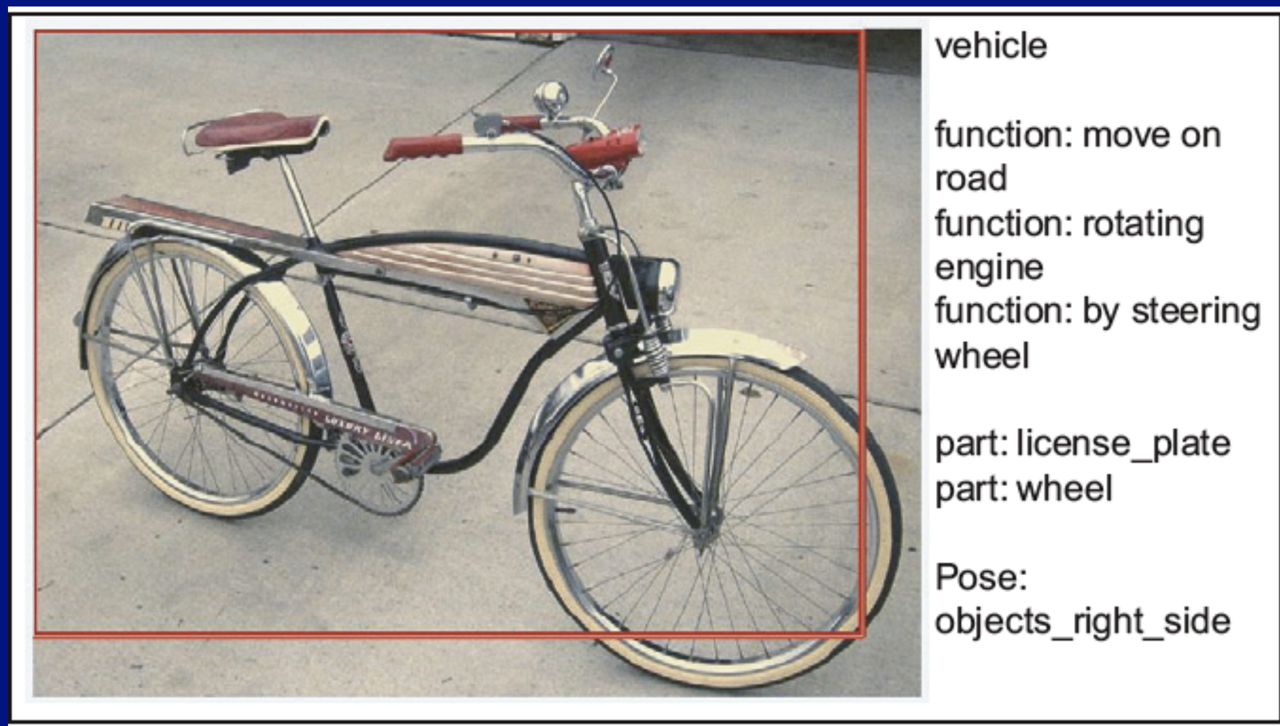
AUC for root/baseline for various types of attribute
baseline: inherit from blc prediction
F: familiar test
UnF: unfamiliar test



Farhadi ea 10



Farhadi ea 10

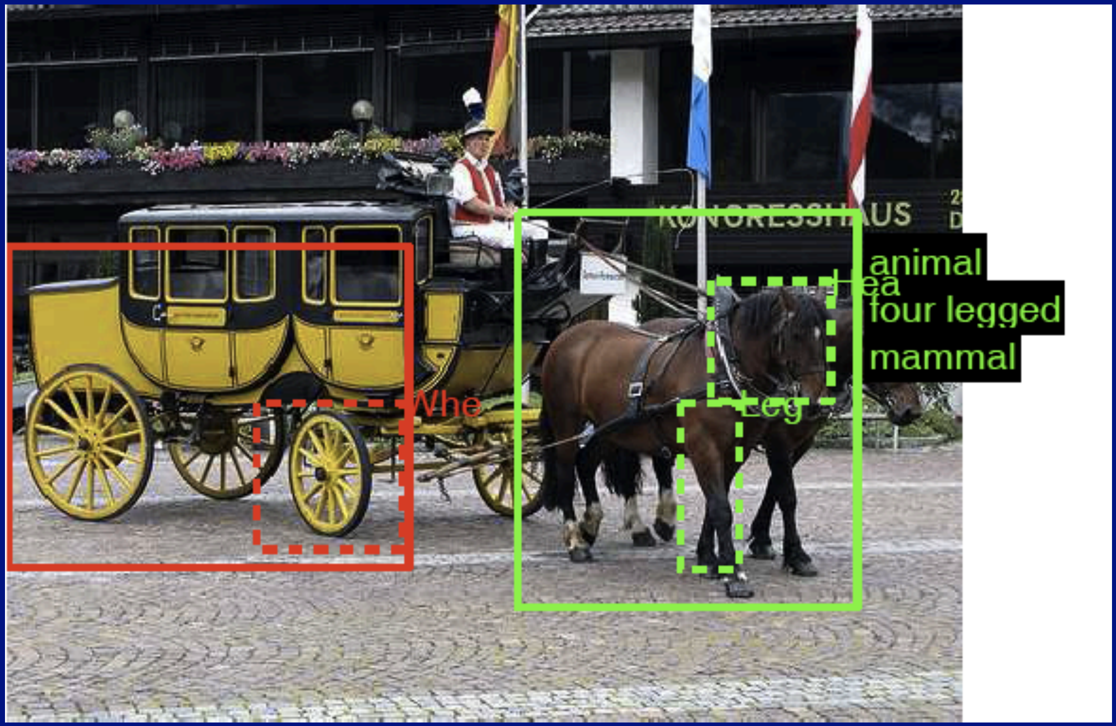


Farhadi ea 10

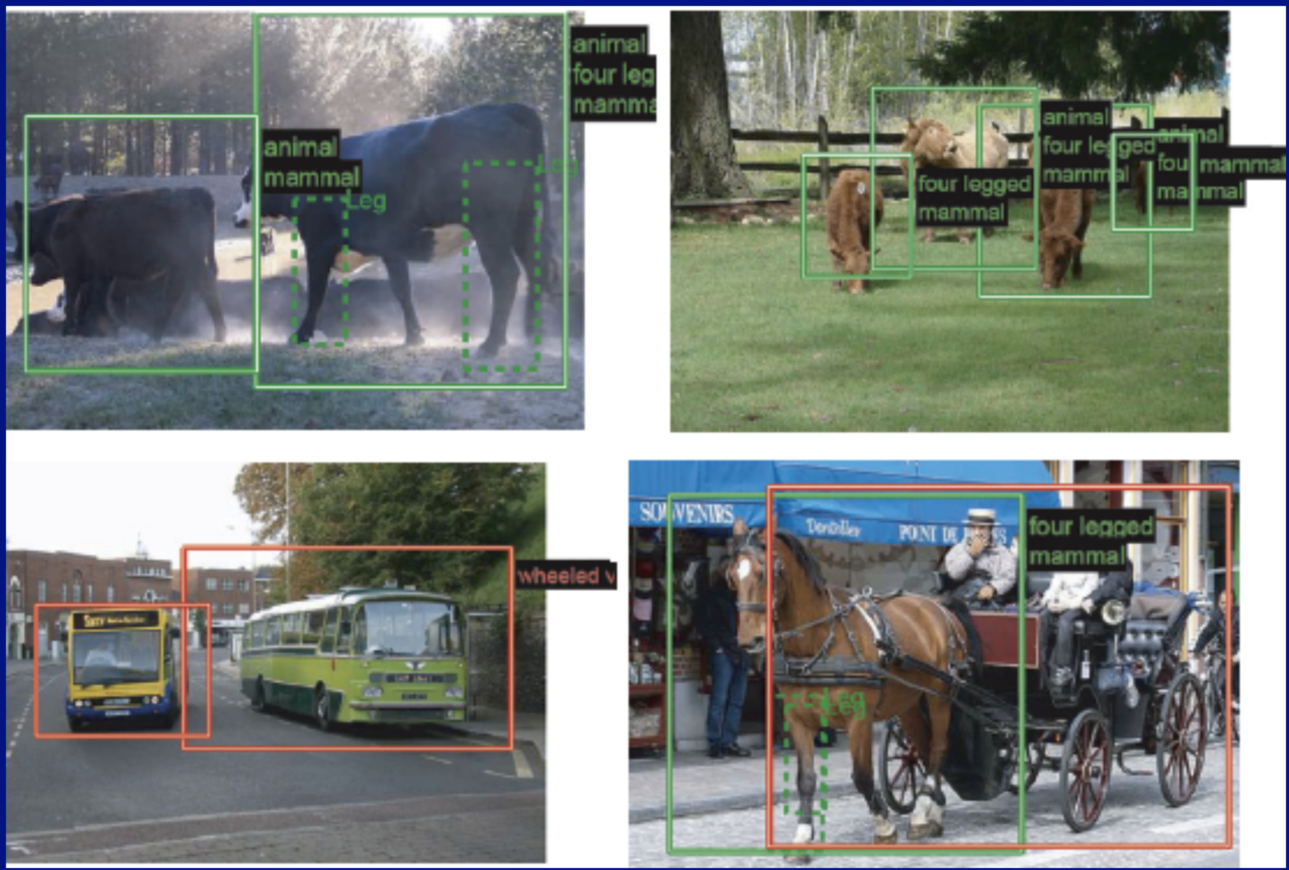
Localizing unfamiliar categories

- Detect by:
  - Part detectors (eg leg - over several example categories)
  - BLC detectors (eg animal - ditto)
  - vote on location
- Train on familiar animals/vehicles, test on unfamiliar

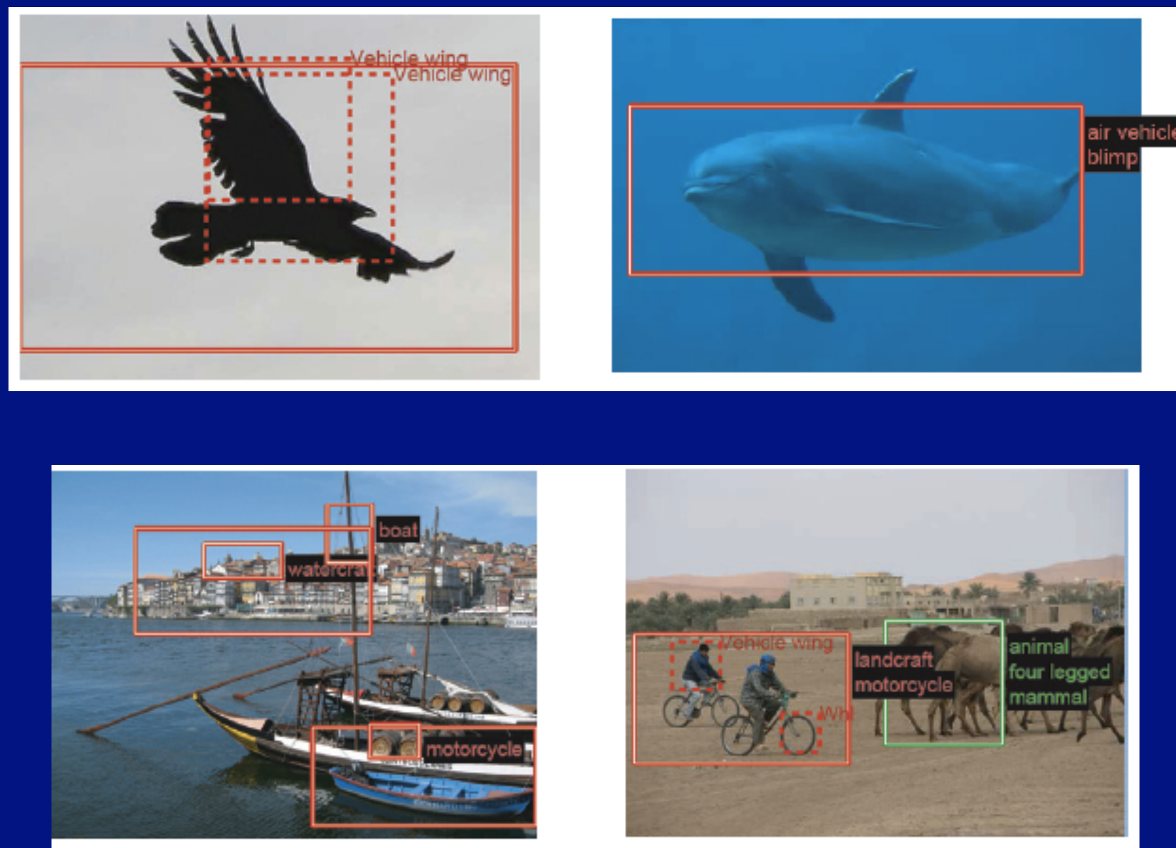
No horses or carriages in training set



Farhadi ea 10



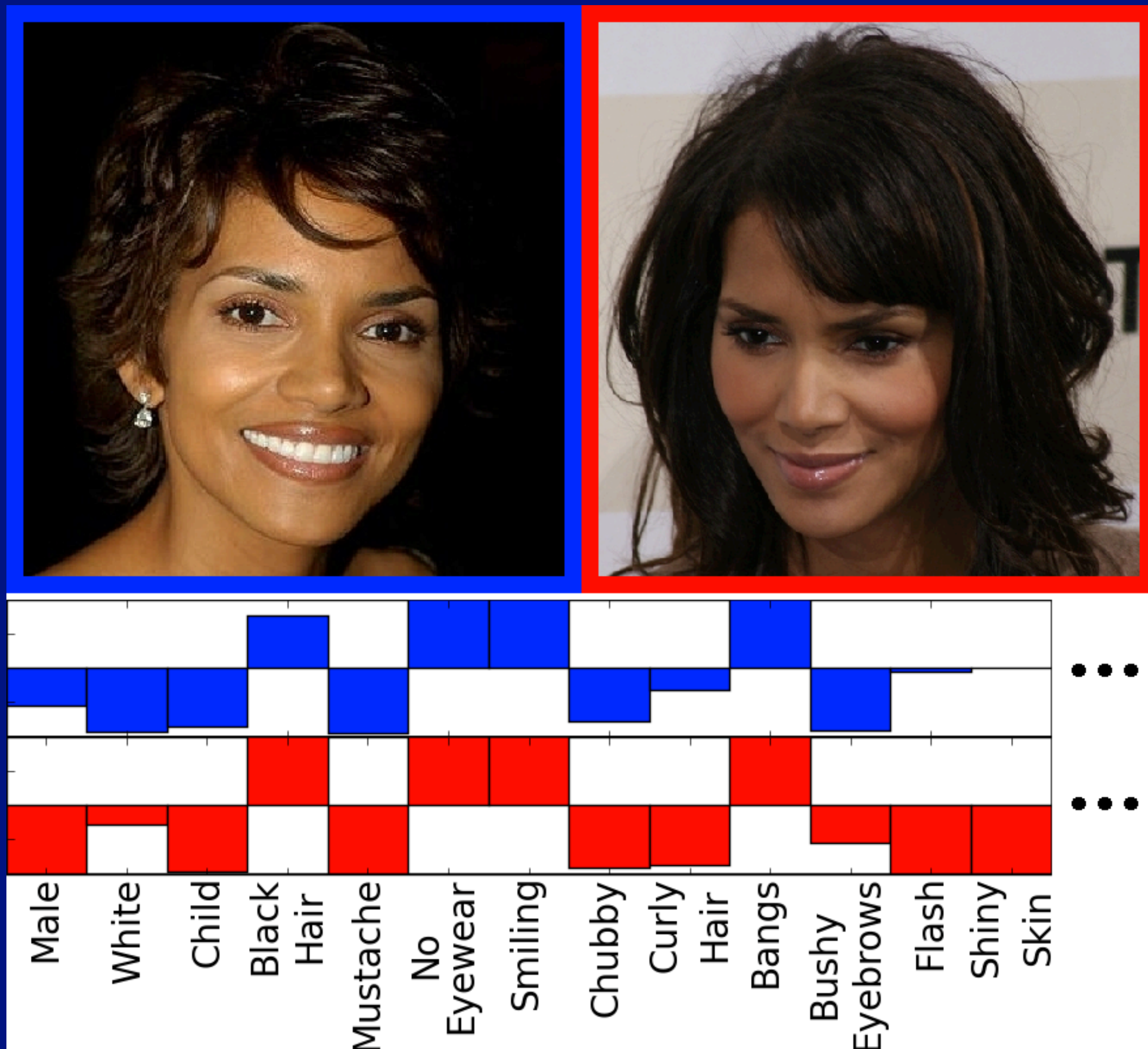
Farhadi ea 10



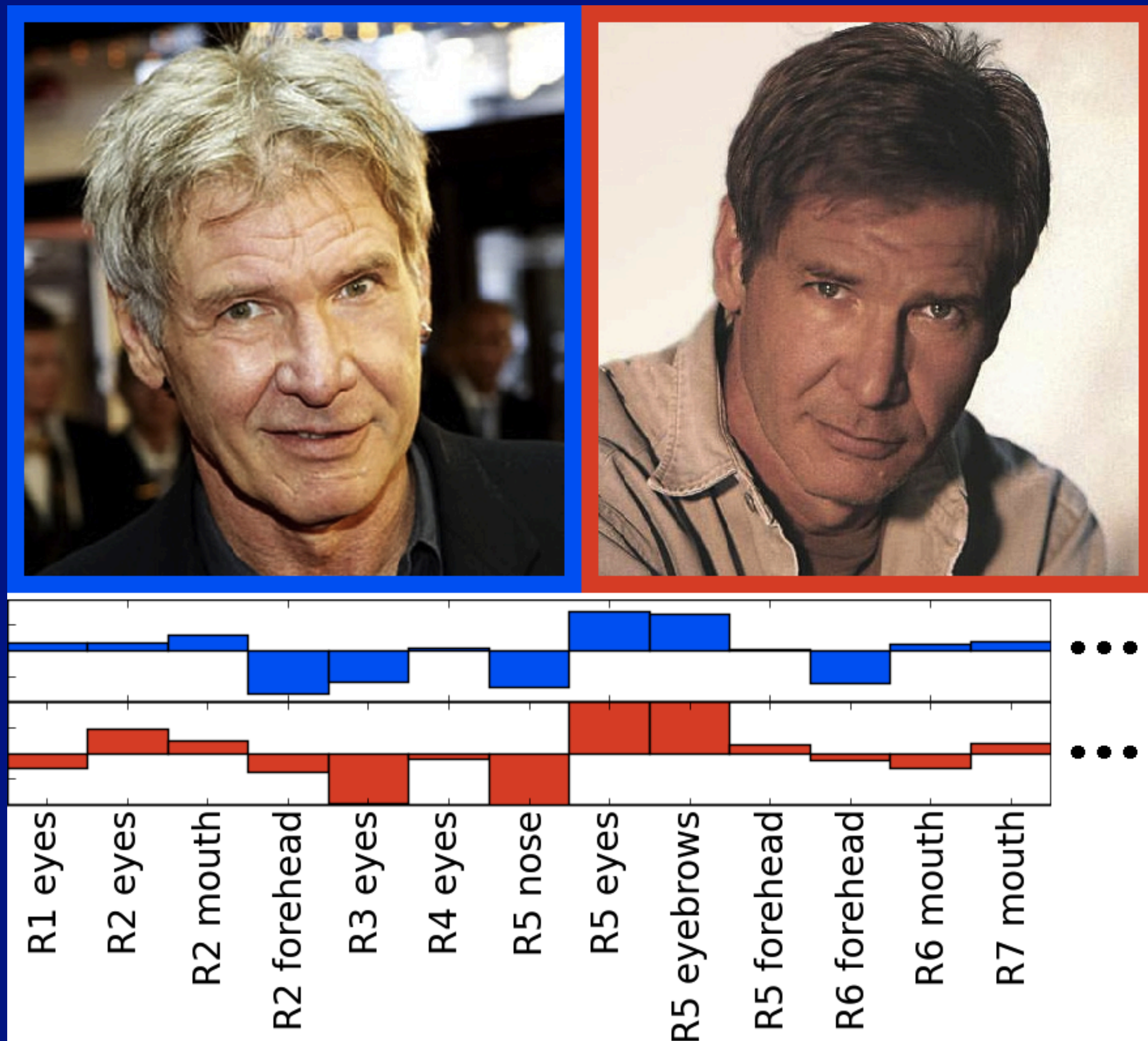
Farhadi ea 10

Accuracy

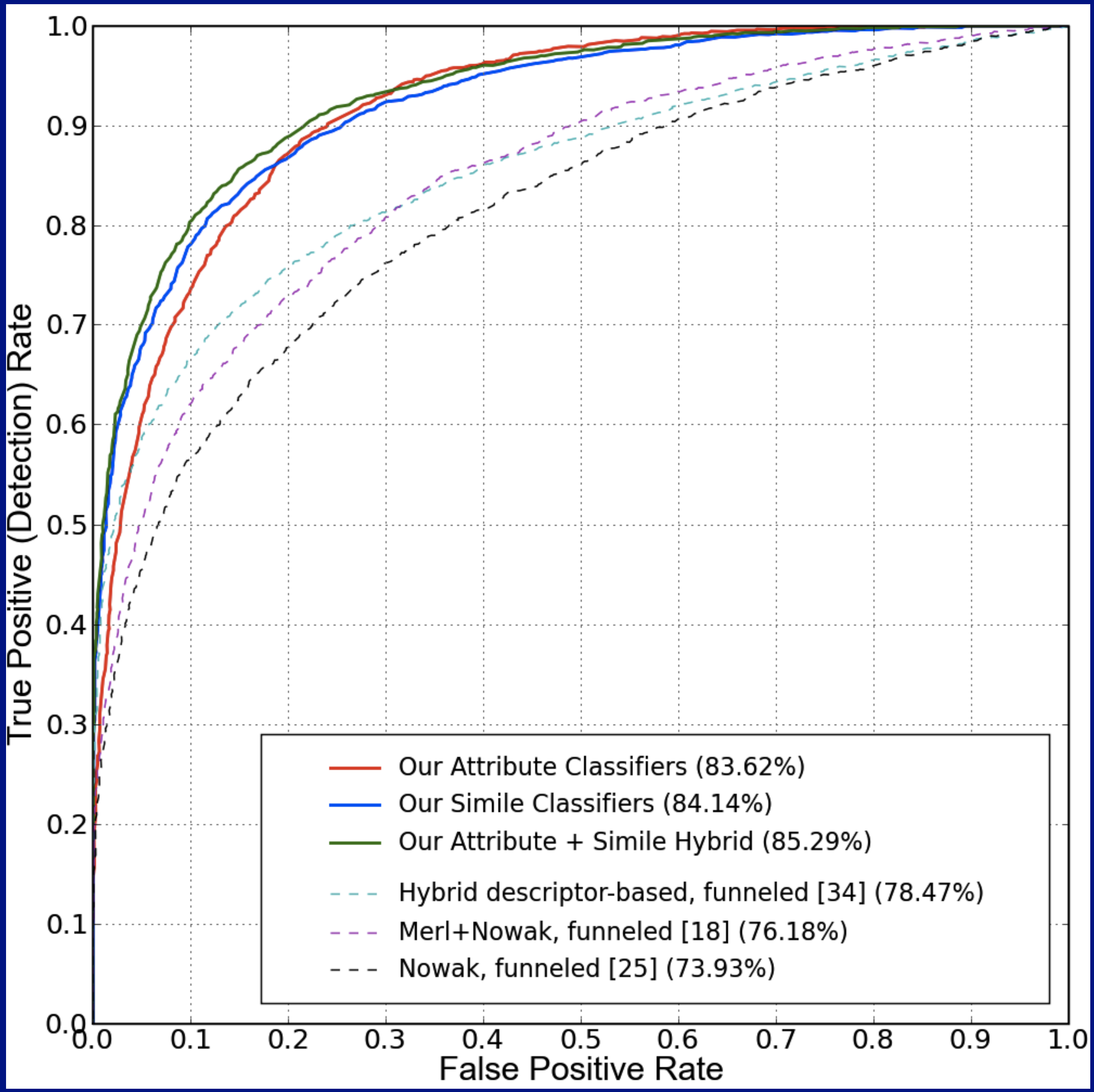
- Papers described are promising, but...
- Standard task: Face Verification
  - is face A the same person as face B?
- Significant improvements using an attribute representation



"Attribute and Simile Classifiers for Face Verification," ICCV 2009. (N. Kumar, A. Berg, P. Belhumeur, S. K. Nayar)



"Attribute and Simile Classifiers for Face Verification," ICCV 2009. (N. Kumar, A. Berg, P. Belhumeur, S. K. Nayar)



"Attribute and Simile Classifiers for Face Verification," ICCV 2009. (N. Kumar, A. Berg, P. Belhumeur, S. K. Nayar)

Datasets - I

- a-Pascal
  - mark up Pascal VOC 2008 with 64 attributes (using Amazon Turk)
  - all of it!
- a-Yahoo
  - 12 additional classes, from Yahoo, with attributes (Amazon Turk)
  - chosen to “mask” Pascal classes
    - Wolf (dog); Centaur (people, horses); goat (sheep); etc.
- Approx 1M annotations! ($600)
- Accuracy
  - Turk inter-annotator agreement 84.1%
  - UIUC inter-annotator agreement 84.3%
  - Turk UIUC agreement 81.4%

Farhadi ea 09

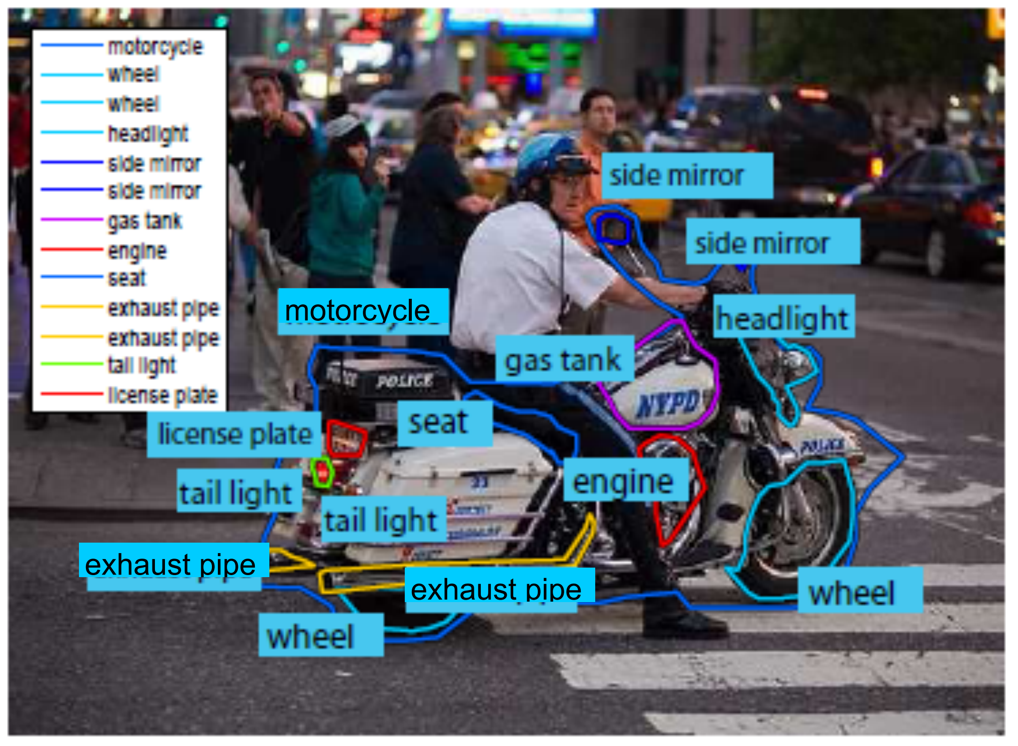
Datasets - II

- Animals with attributes
  - 30475 images
  - animals in 50 classes, min 92 per class
  - classes have attributes from Osherson, 91
  - 85 attributes in total
  - attribute markup inherited from class

Lampert ea 09

Datasets - III

Cross Category Object REcognition Dataset



Endres et al 10; Farhadi ea 10

2780 Images – from ImageNet

3192 Objects – 28 Categories

26695 Parts – 71 types

30046 Attributes – 34 types

1052 Material Images – 10 types

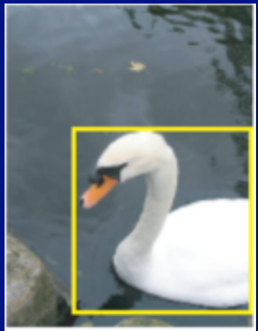
http://vision.cs.uiuc.edu/CORE

Future Directions

- Richer semantics of attributes
- Spatial support and spatial models
- Materials
- Similarity
- Discriminative attributes
- Attribute correlation
- Learning from X

Richer semantics

- Distinguish between:
  - it has one
  - it should have one, but I can't see it
  - it doesn't have one
  - the one it has belongs to something else



Bird
No "tail"

Aeroplane
No "jet engine"

Bird
"Leaf"

Spatial support and spatial models

- simple modifiers can be learned w/o spatial markup
  - e.g. “pink” Yanai + Barnard, 05.

- complex texture modifiers can, too
  - e.g. “spots”, Ferrari+Zisserman, 07

- joint modifier/noun data make learning easier
  - e.g. blue hat, Wang+Forsyth, 09

Materials



- Material not inherited from object in humans
  - Sharan ea 09
- Material classification hard
  - Liu ea 10; nice dataset
  - Hayman ea 04; nice dataset

Material	Recognition rate (%) SVM	VZ
sandpaper	0.00	1.23
aluminium foil	11.35	12.35
styrofoam	34.72	38.27
sponge	50.62	54.32
corduroy	46.91	59.26
linen	30.41	25.93
cotton	11.11	20.99
brown bread	5.11	7.41
orange peel	11.11	11.11
cracker B	3.70	7.41
AVERAGE	**20.50**	**23.83**

(a) Training only on CUReT

Material	Recognition rate (%) SVM	VZ
sandpaper	77.78	66.67
aluminium foil	91.67	88.89
styrofoam	100.00	91.67
sponge	100.00	100.00
corduroy	80.56	80.56
linen	61.11	41.67
cotton	61.11	47.22
brown bread	77.78	80.56
orange peel	100.00	63.89
cracker B	91.67	80.56
AVERAGE	**84.17**	**74.17**

(b) Training on *both* CUReT and KTH-TIPS



Similarity

- Learn from “A is like B”, “C is not like D”
- Useful representation
  - Kumar 09
- Some improvements in classification with few examples
  - Wang ea 09
- Links to attributes/learning from text give improvements
  - Rohrbach ea 10

Discriminative attributes

- Haven’t written down all attributes
  - Naive bayes does objects from attributes at 74% given ground truth
- Select random splits of objects that are well predicted
  - obtain by random search
    - assign objects to +, -, x randomly
    - learn a classifier
    - keep those that are accurately predicted on held out set
- Use these as attributes, too
- Q: Do they have semantics?
  - “cows and horses have it, cars and buses don’t”

Attribute correlation

- Through latent objects is probably not right
  - some attributes are correlated through objects
  - others through their semantics (eg furry, hairy, fuzzy, soft)
- Error correction (?)
  - natural result of massive correlations and competent modelling
- Fundamental coding limits (?)
  - can we error correct arbitrarily with visual features?

Learning from X

- Descriptions from text can produce OK visual models
  - Farhadi ea 09; Lampert ea 09
- Pragmatics is a major obstacle
  - (dead silence on this issue)

The big question

- How to insert object semantics into object recognition?
  - without being silly
  - what is useful knowledge?
  - where does it come from?
  - what is worth saying about objects?
  - what objects are worth saying things about?
  - how should categories be created and destroyed to meet pragmatic needs?