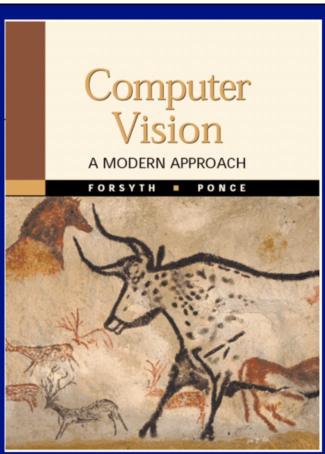
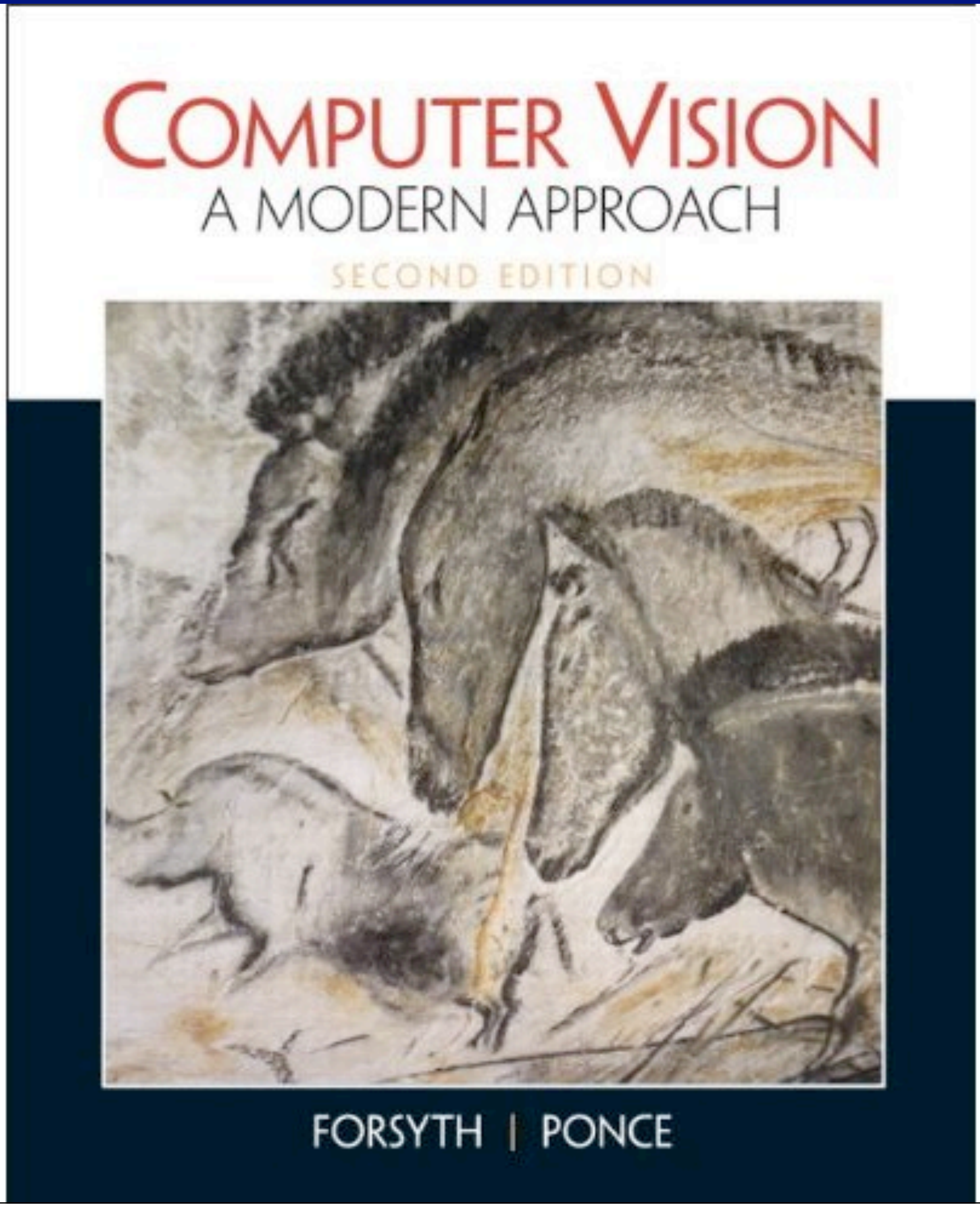
More words and Bigger Pictures

D.A. Forsyth, UIUC,
with input from J. Hockenmaier, UIUC, D. Hoiem, UIUC,
T. Berg, SUNYSB, P. Duygulu, Bilkent,
K. Barnard, U. Arizona,
A. Farhadi, U. Washington
I. Endres, A. Sadeghi, B. Liao, Y. Wang, V. Hedau, K. Karsch,
all of UIUC
G. Wang, NTU,
Nicolas Loeff, RGM Advisors



10 years old

8 months old



Why is visual object recognition useful?

- If you want to act, you must draw distinctions
- For robotics
  - recognition can predict the future
    - is the ground soggy?
    - is that person doing something dangerous?
    - does it matter if I run that over?
    - which end is dangerous?
- For information systems
  - recognition can unlock value in pictures
    - for search, clustering, ordering, inference, ...
- General engineering
  - recognition can tell what people are doing
- If you have vision, you have some recognition system

Observation

Query on

“Rose”

Example from Berkeley
Blobworld system

Annotation results in complementary words and pictures

Annotation results in complementary words and pictures

Query on

Example from Berkeley
Blobworld system

Annotation results in complementary words and pictures

Query on

“Rose”

and

Example from Berkeley
Blobworld system

Example: Predicting word tags

It was there and we didn't

It was there and we predicted it

It wasn't and we did

Substantial literature; this figure from Loeff Farhadi 08; see also Quattoni Darrell 07

Words and pictures affect one another

Marc by Marc Jacobs
Adorable peep-toe pumps, great for any occasion. Available in an array of uppers. Metallic fabric trim and bow detail. Metallic leather lined footbed. Lined printed design. Leather sole. 3 3/4" heel.

Zappos.com

soft and glassy patent calfskin trimmed with natural vachetta cowhide, open top satchel for daytime and weekends, interior double slide pockets and zip pocket, seersucker stripe cotton twill lining, kate spade leather license plate logo, imported
2.8" drop length
14"h x 14.2"w x 6.9"d

Katespade.com

It's the perfect party dress. With distinctly feminine details such as a wide sash bow around an empire waist and a deep scoopneck, this linen dress will keep you comfortable and feeling elegant all evening long. Measures 38" from center back, hits at the knee.
* Scoopneck, full skirt.
* Hidden side zip, fully lined.
* 100% Linen. Dry clean.

bananarepublic.com

E-commerce transactions in 2004, 2005, 2006 of $145 billion, $168 billion, and $198 billion (Forrester Research).

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - serious problems with intellectual underpinnings
- Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - selection
- Crucial open questions
  - dataset bias
  - links to utility

Platonism?

A belief space about recognition

- Object categories are fixed and known
  - Each instance belongs to one category of k
- Good training data for categories is available
- Object recognition=k-way classification
- Detection = lots of classification

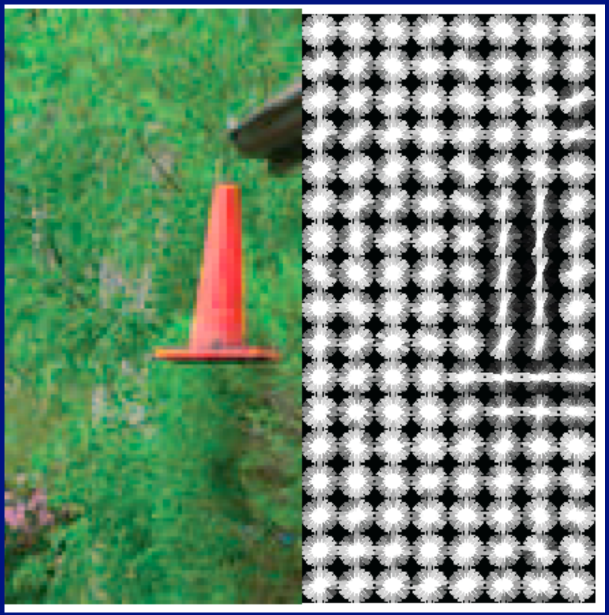
Obtain dataset

Build features

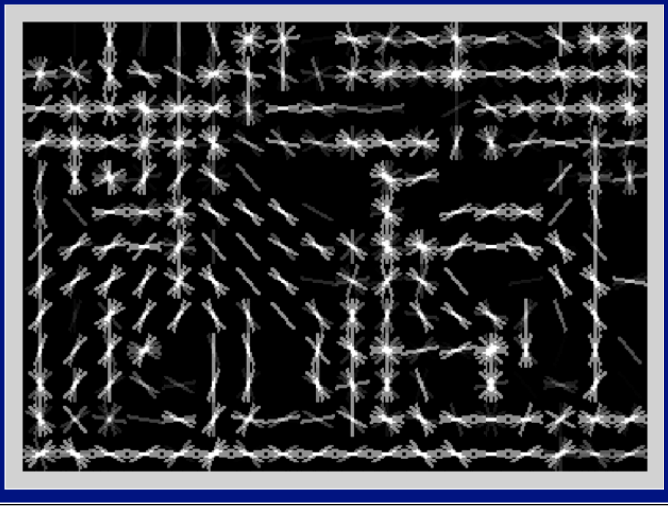
Mess around with classifiers, probability, etc

Produce representation

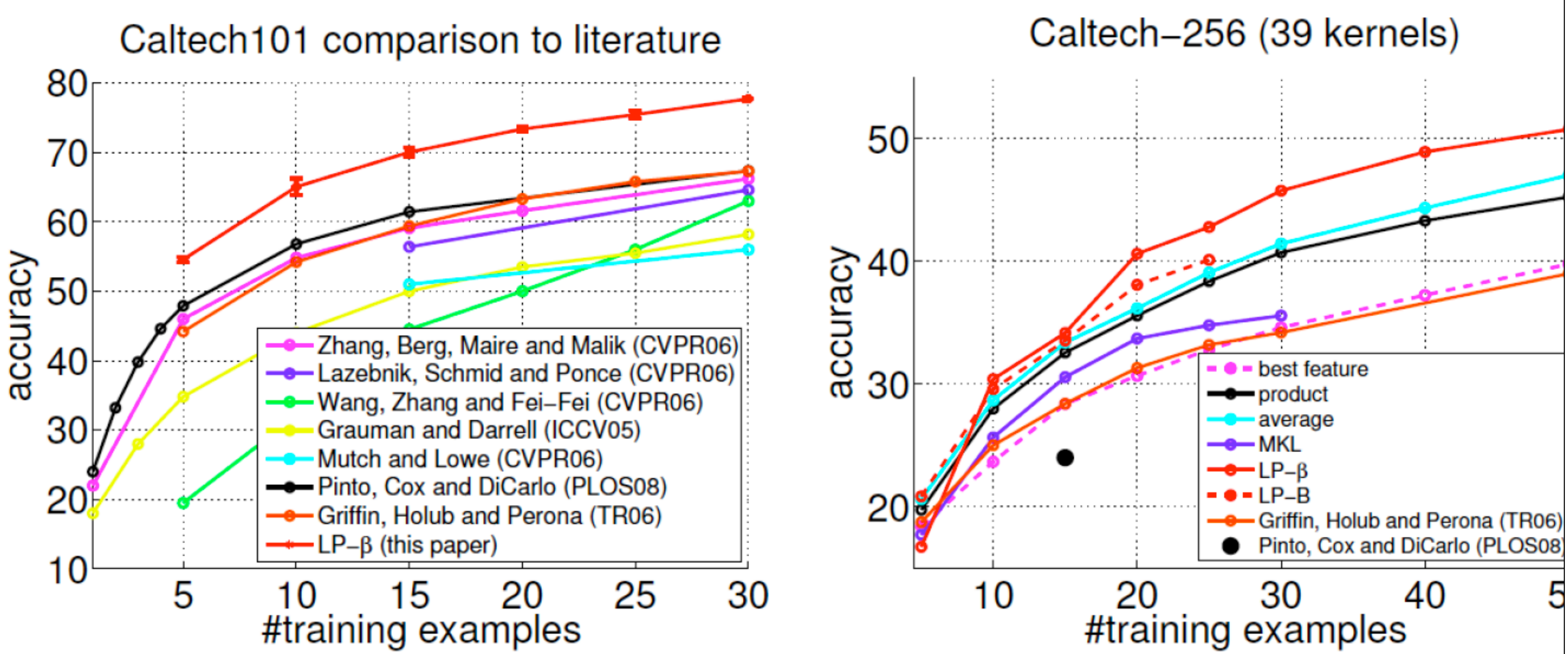
Features

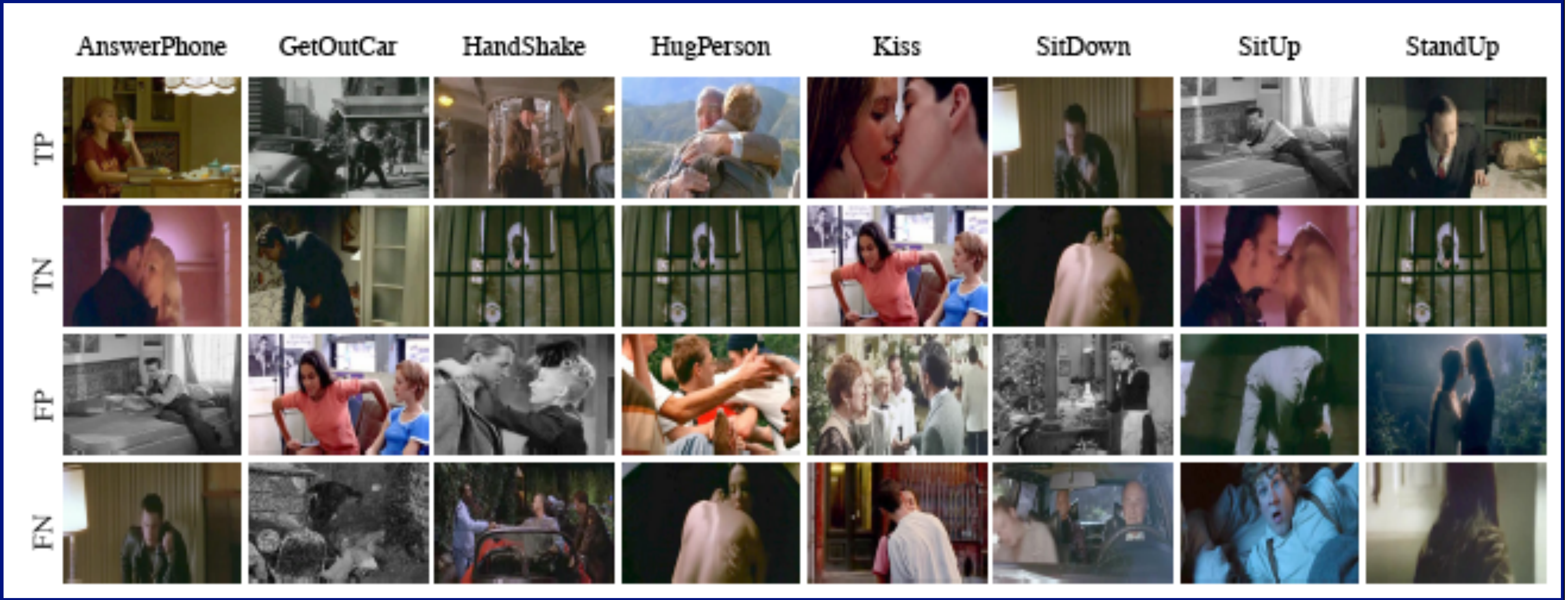


- Principles
  - illumination invariant (robust) -> gradient orientation features
  - windows always slightly misaligned -> local histograms
- HOG, SIFT features (Lowe, 04; Dalal+Triggs 05)



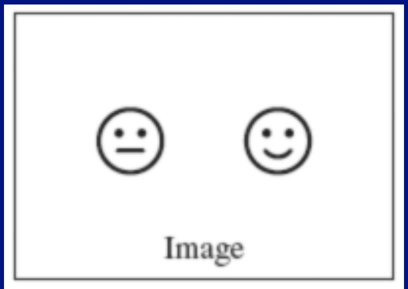
Classification works well

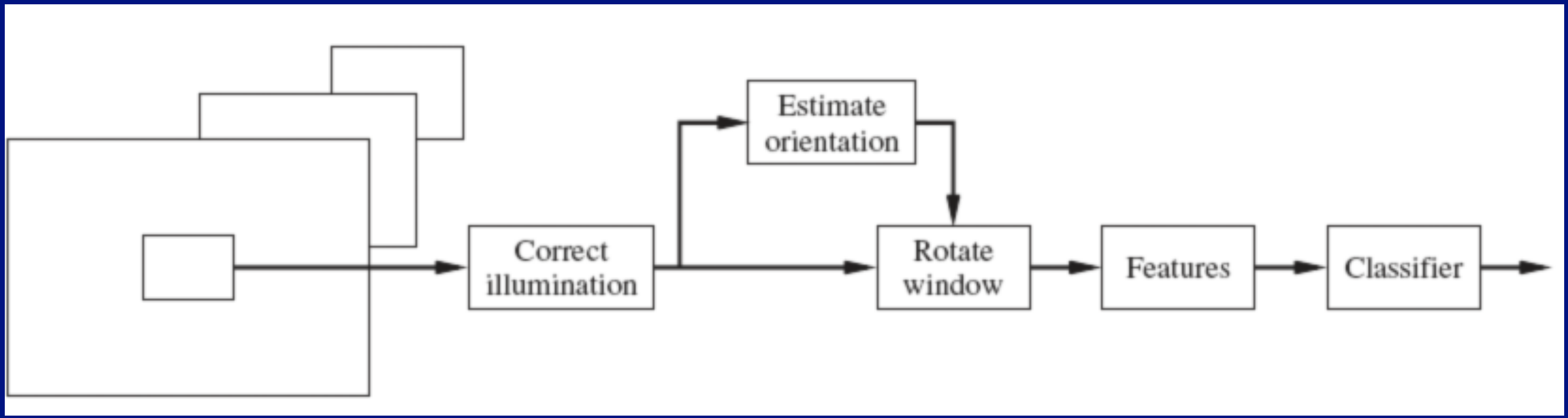




Movies and captions: Laptev et al 08

Detection with a classifier





P. Felzenszwalb, D. McAllester, D. Ramanan. "A Discriminatively Trained, Multiscale, Deformable Par
Model" CVPR 2008.

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - serious problems with intellectual underpinnings
- Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - selection
- Crucial open questions
  - dataset bias
  - links to utility

Platonism?

A belief space about recognition

- Object categories are fixed and known — Obvious nonsense
  - Each instance belongs to one category of k — Obvious nonsense
- Good training data for categories is available — Obvious nonsense
- Object recognition=k-way classification
- Detection = lots of classification

What have we inherited from this view?

- Deep pool of information about feature constructions
- Tremendous skill and experience in building classifiers
- Much practice at empiricism
  - which is valuable, and hard to do right

- Subtleties
  - What about the unfamiliar?
  - What kinds of things should we recognize?
  - What environmental knowledge helps?
  - What should we say about pictures?
  - How does utility affect the output?

Platonism?

A belief space about recognition

- Object categories are fixed and known — Obvious nonsense
  - Each instance belongs to one category of k — Obvious nonsense
- Good training data for categories is available — Obvious nonsense
- Object recognition=k-way classification
- Detection = lots of classification

Are these monkeys?

Big questions

Computer vision

Obtain dataset

- What signal representation should we use ?

Build features

PLUMBING

Classifiers, probability
(Light entertainment)

MODELS

What aspects of the world
should we represent and how?

Mess around
with classifiers,
probability, etc

Computer vision

- What should we say about visual data?

Produce representation

Bias

Should not be perjorative

- 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

Size doesn't make bias go away

- And could make it worse...
  - eg your dataset collector really likes red cars
- cf next slide

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

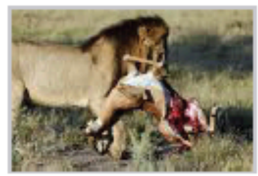

I was doing research on
400 × 300 - 27k - jpg
lowkayhwa.com
Find similar images


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

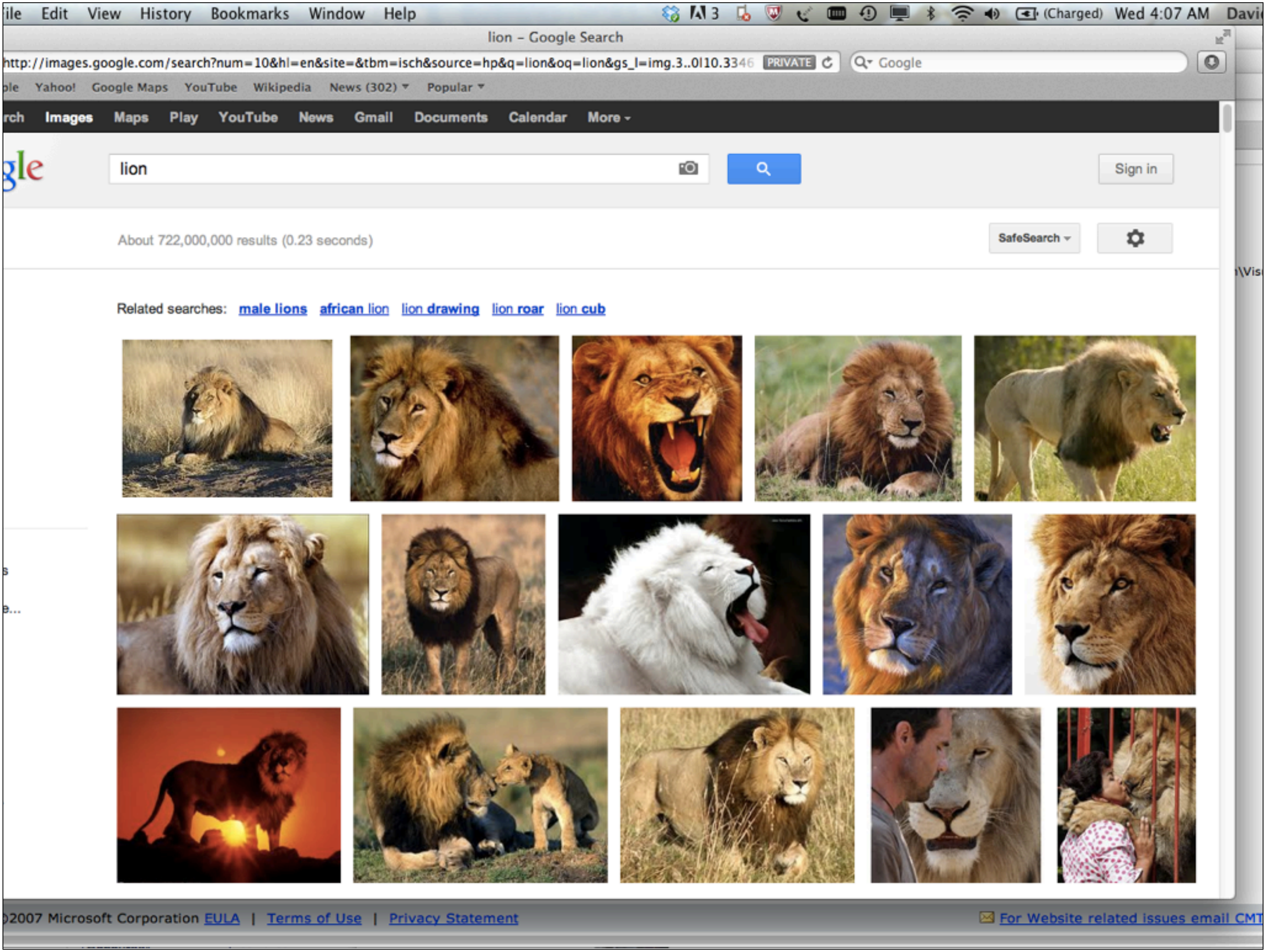

LION!
500 × 385 - 74k - jpg
firemice.wordpress.com
Find similar images


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

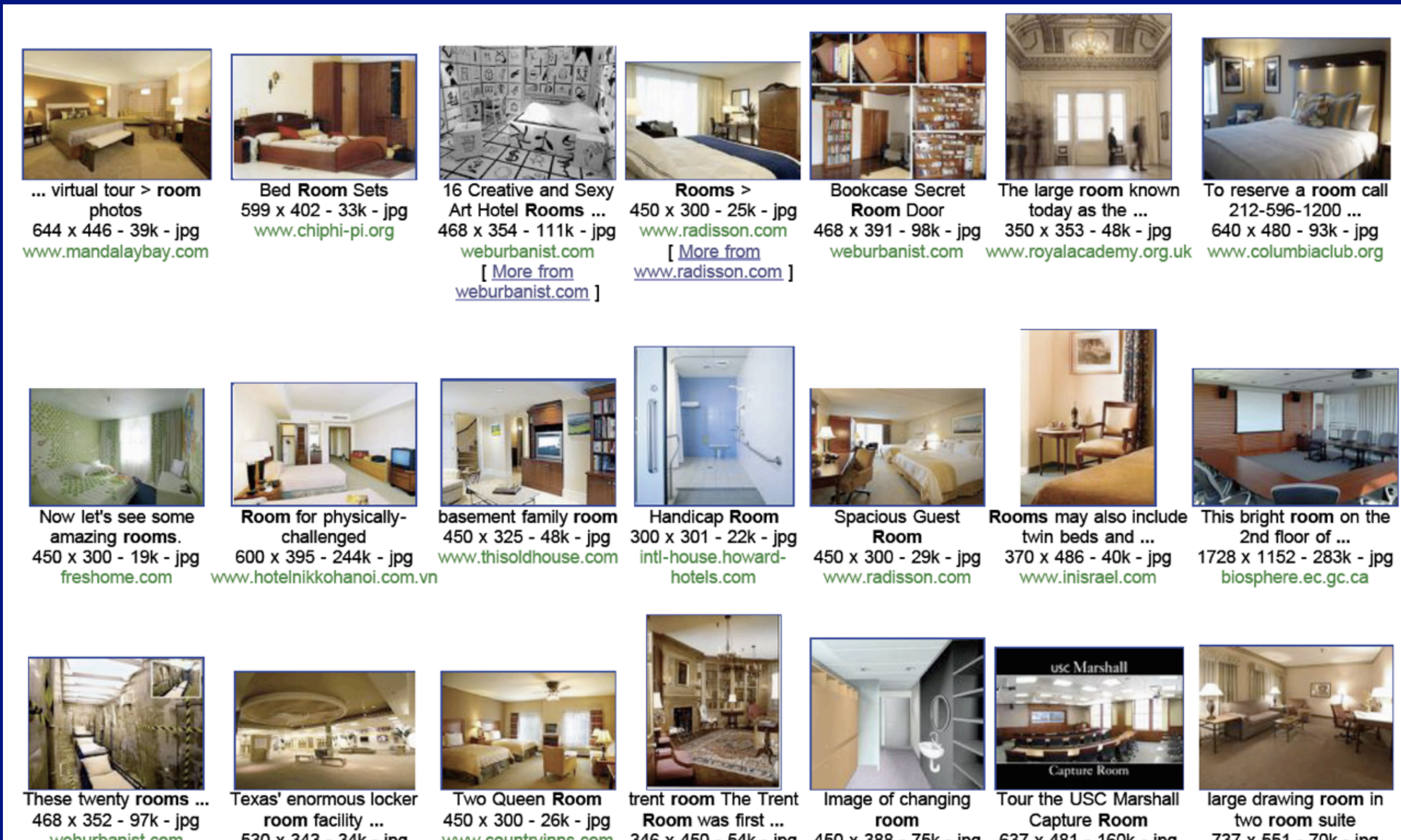

Picture: 17 stone
468 × 602 - 93k - jpg
dailymail.co.uk
Find similar images


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


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


ile Edit View History Bookmarks Window Help
(Charged) Wed 4:07 AM Davi
lion – Google Search
http://images.google.com/search?num=10&hl=en&site=&tbm=isch&source=hp&q=lion&oq=lion&gs_l=img.3..0l10.3346
PRIVATE
Google
ple Yahoo! Google Maps YouTube Wikipedia News (302) Popular
rch Images Maps Play YouTube News Gmail Documents Calendar More
gle
lion
Sign in
About 722,000,000 results (0.23 seconds)
SafeSearch
Related searches: male lions african lion lion drawing lion roar lion cub
©2007 Microsoft Corporation EULA | Terms of Use | Privacy Statement
For Website related issues email CMT

Google “rooms”



Flickr “rooms”

Bias isn't always bad

- If all the faces on the web are politicians
  - one needs only to be good at politicians to be good at the web
- If no users can tell an ape from a monkey
  - you might not have to either
- If people really only want to search videos for "kissing"
  - then you don't need a general activity recognition strategy

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?

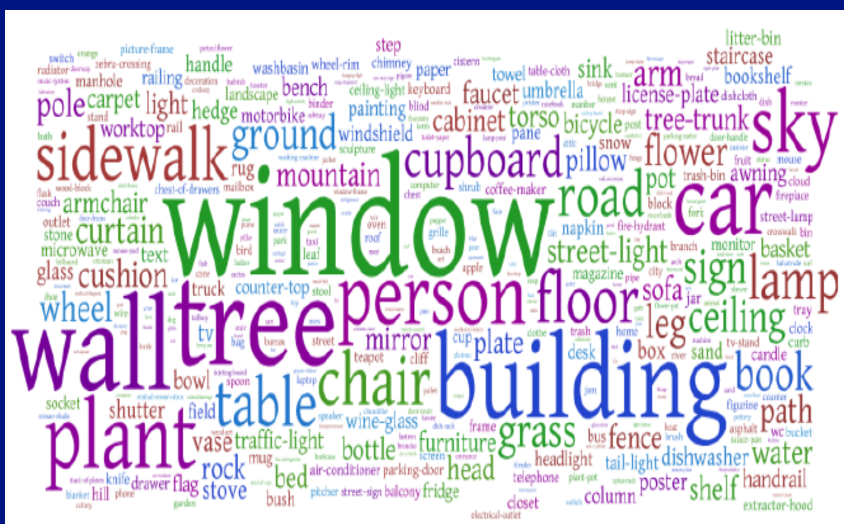
Pedestrian Detection

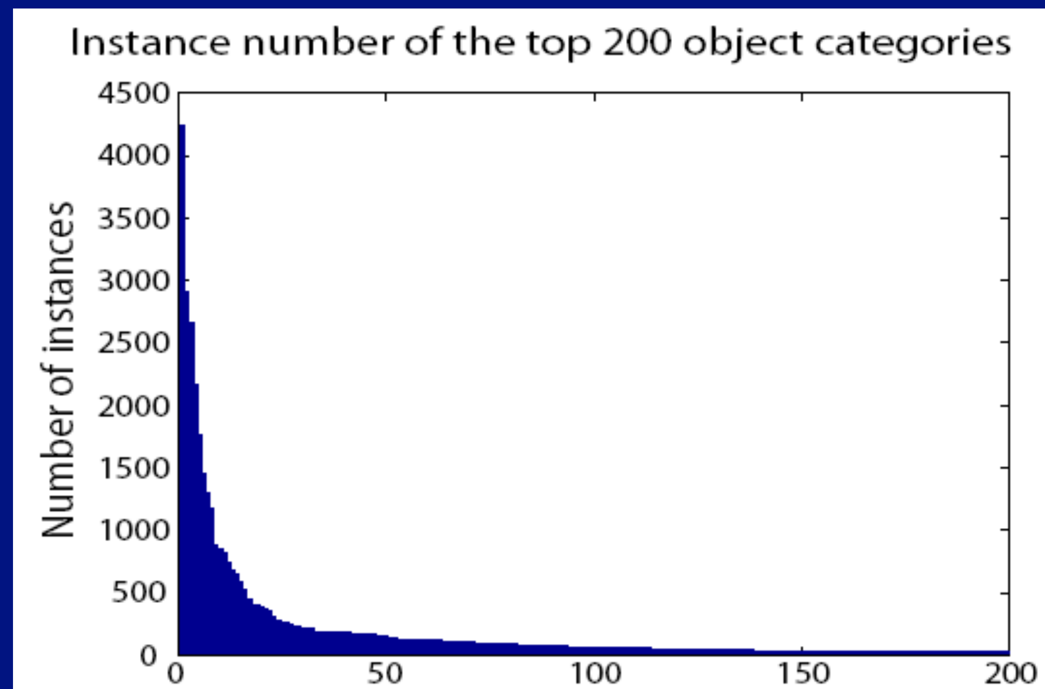
- Pedestrian detection:
  - We may not run down people who behave strangely
    - want "will fail to detect with frequency ..."
    - can do "..." IF test set is like training set
  - There is a large weight of easy cases which may conceal hard cases
- Resolution (frankly implausible)
  - ensure that training set is like test set
- Resolution (perhaps)
  - try only to learn things that are "fairly represented" in datasets
  - i.e. build models

Object recognition

- The world can't be like the dataset because
  - many things are rare in plausible datasets
    - but not in the world
  - this exaggerates bias
- Strategies
  - train by comparison to similar objects
  - represent in terms of pooled properties

Many things are rare





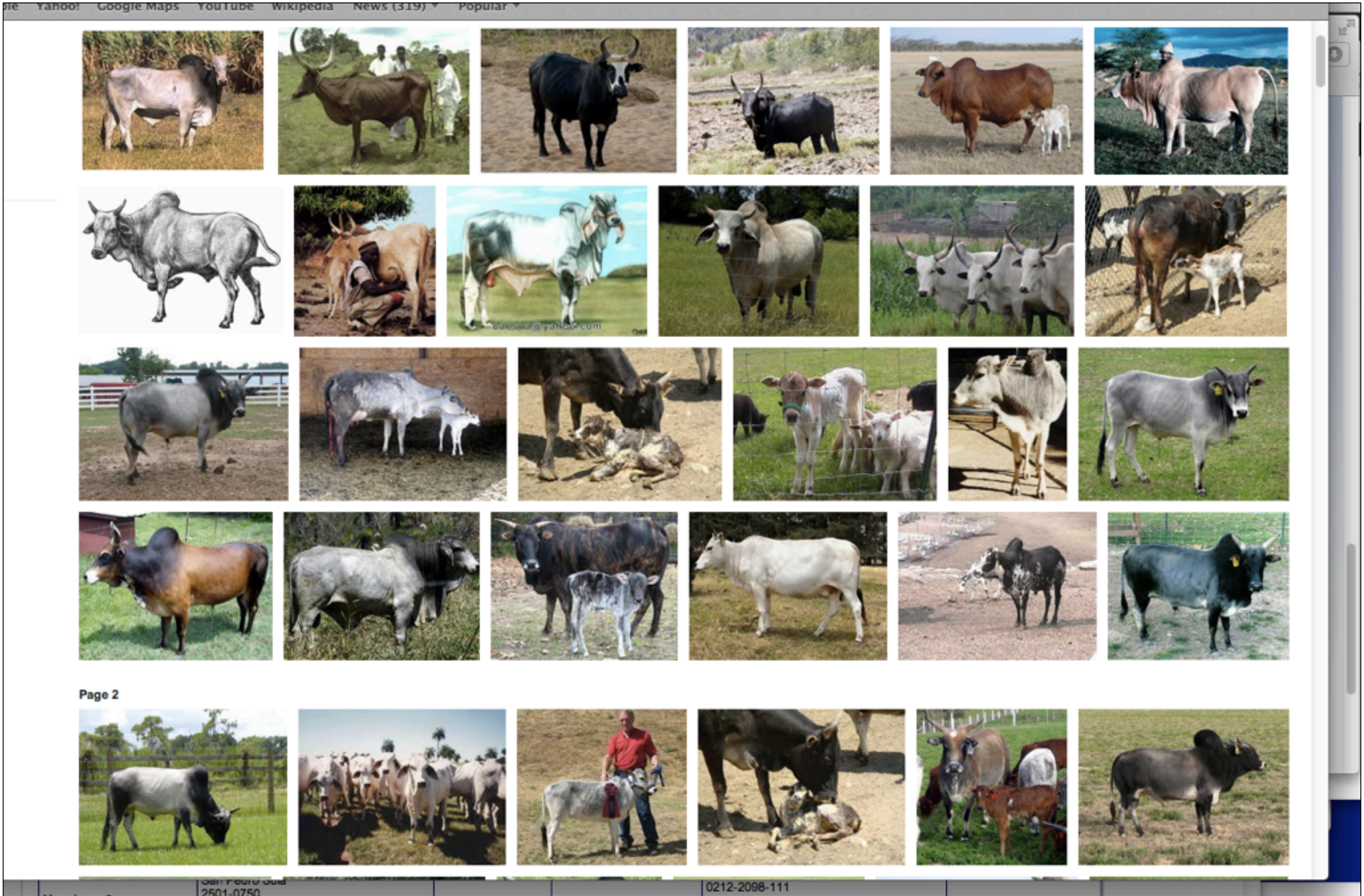
Wang et al, 10, LabelMe data cf word frequencies, which also tend to be like this

Defenses against Bias

- Appropriate feature representations
  - eg illumination invariance
- Appropriate intermediate representations
  - which could have less biased behavior
  - perhaps attributes? scenes? visual phrases?
- Appropriate representations of knowledge
  - eg geometry --- pedestrian example

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - serious problems with intellectual underpinnings
- Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - sentences
- Crucial open questions
  - dataset bias
  - links to utility


Yahoo! Google Maps YouTube Wikipedia News (319) Popular
Page 2
San Pedro Sula
2501-0750
Honduras
0212-2098-111

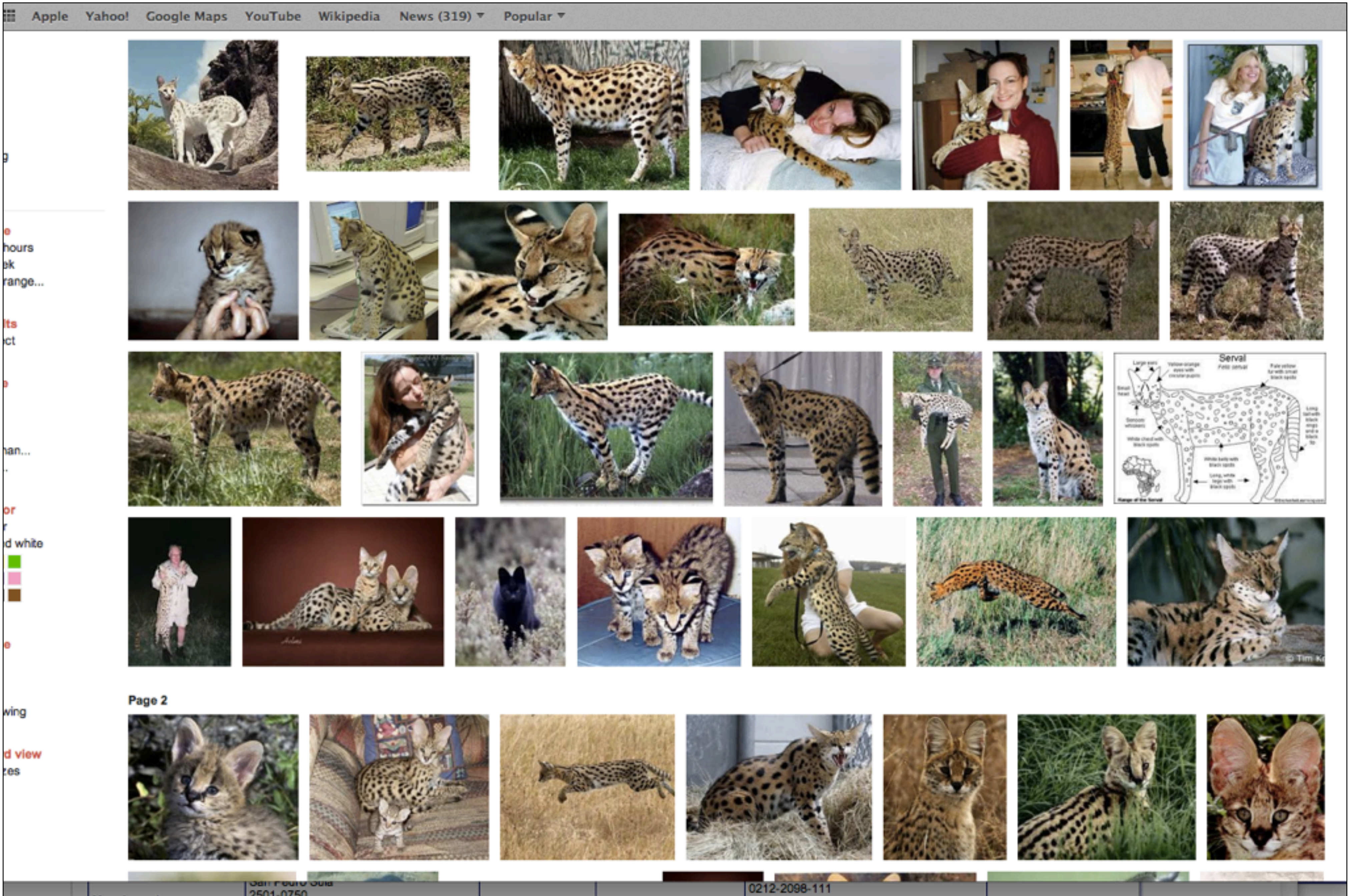

Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News (319)
Popular
hours
range...
d white
d view
Serval
Page 2
2501-0750
0212-2098-111

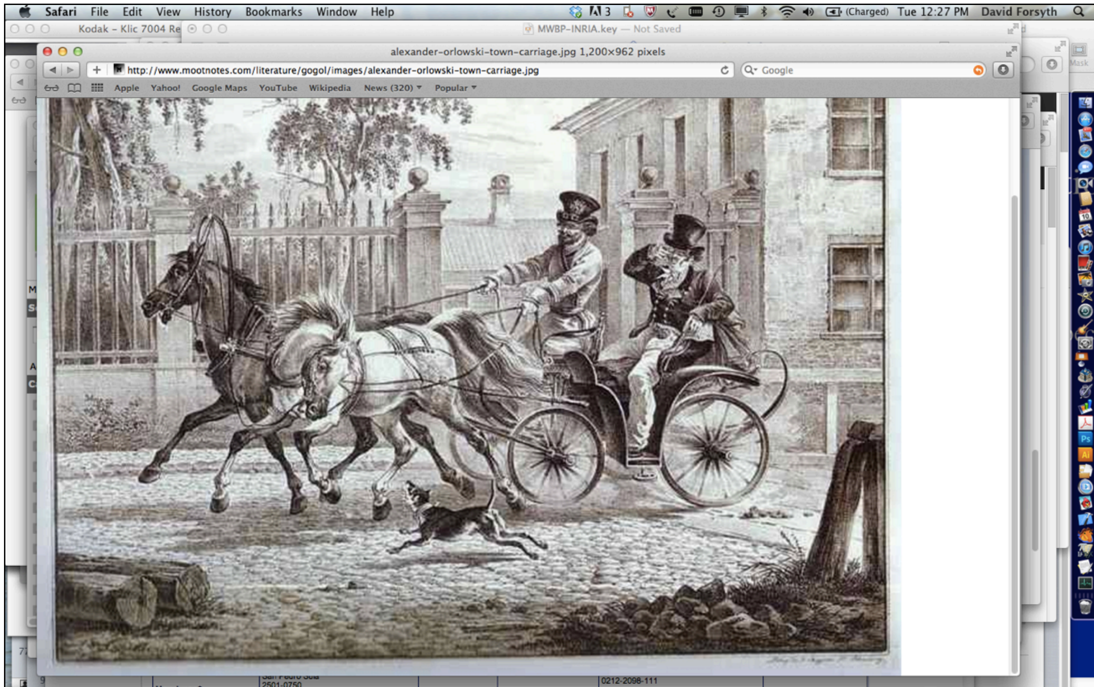
FIG. 109.—[redacted]

2501-0750 0212-2098-111

Name in common use among sailors in 19'th century is deeply shocking to modern ears; appears in Aubrey Maturin novels by Patrick O'Brien

Honduras*	San Pedro Sula 2501-0750		0212-2098-111

9	2501-0750			0212-2098-111

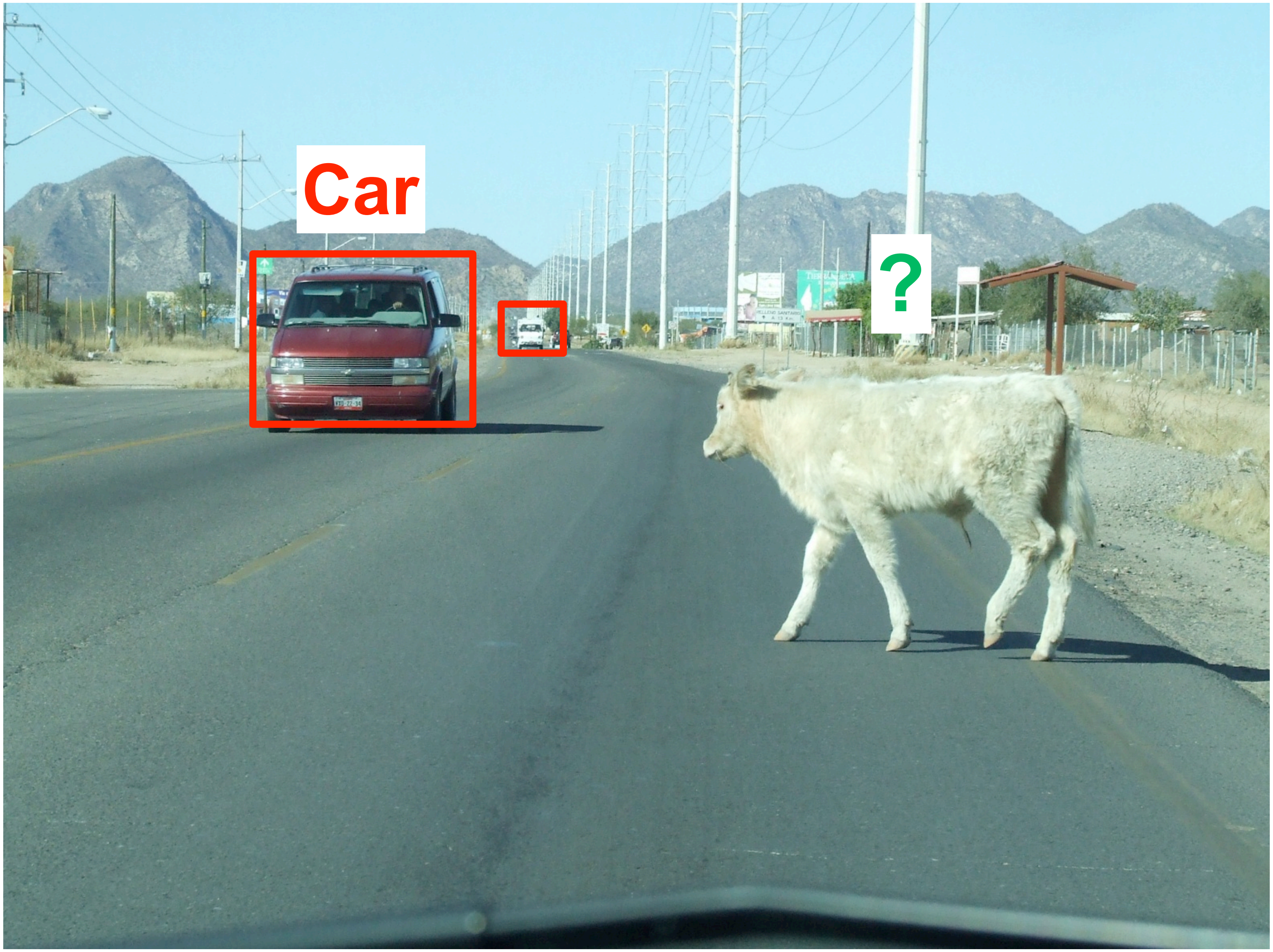


britchka

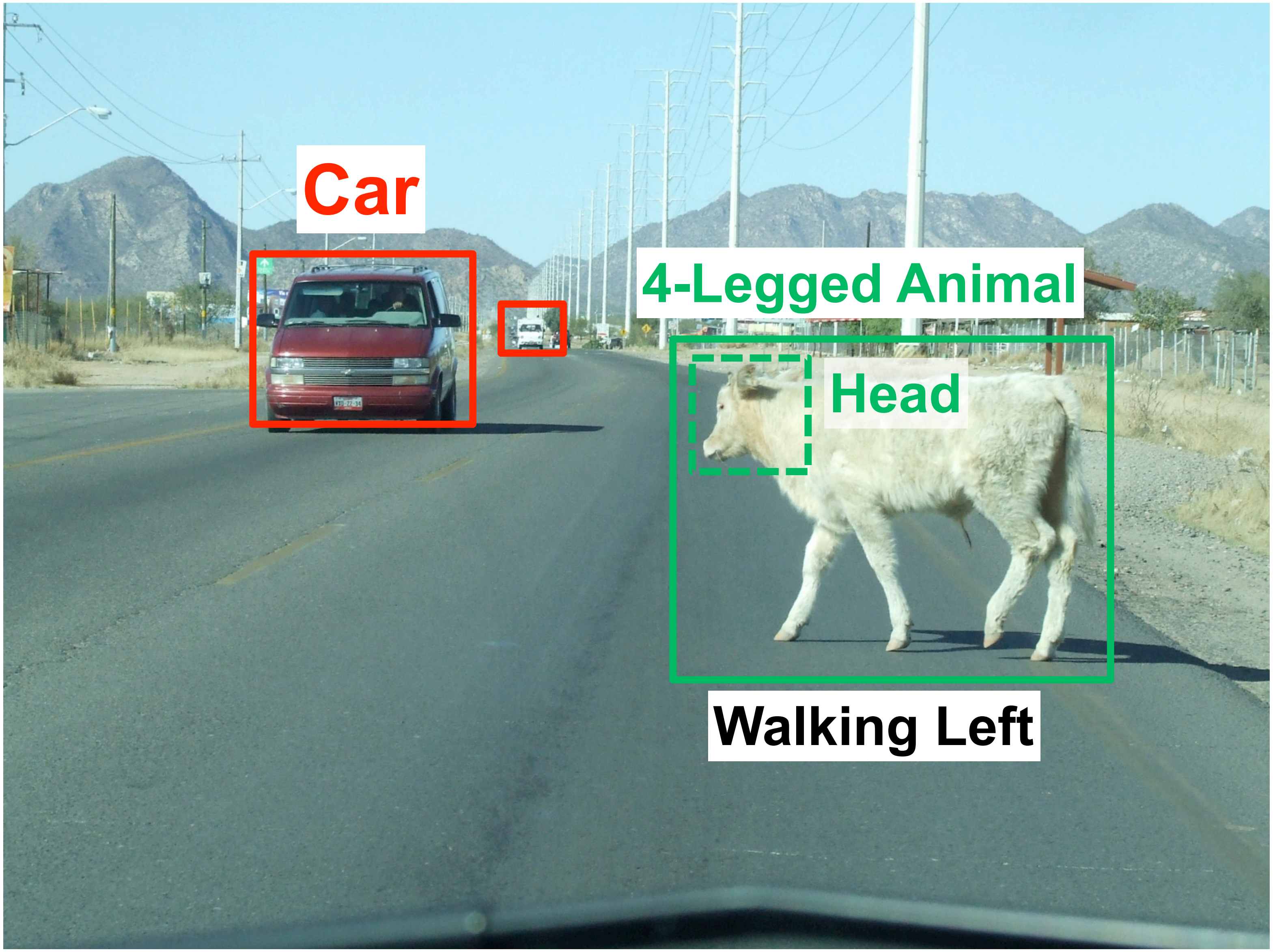
brougham

The Unfamiliar

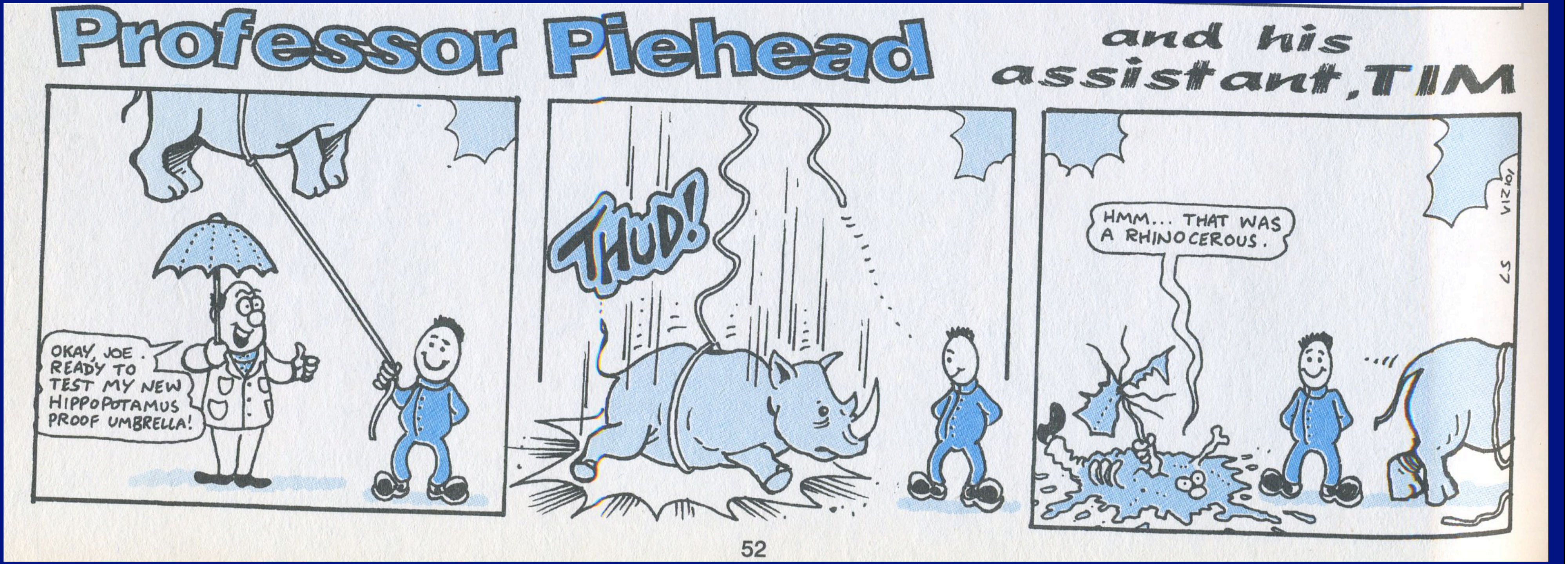
Vision for driving



Vision for driving

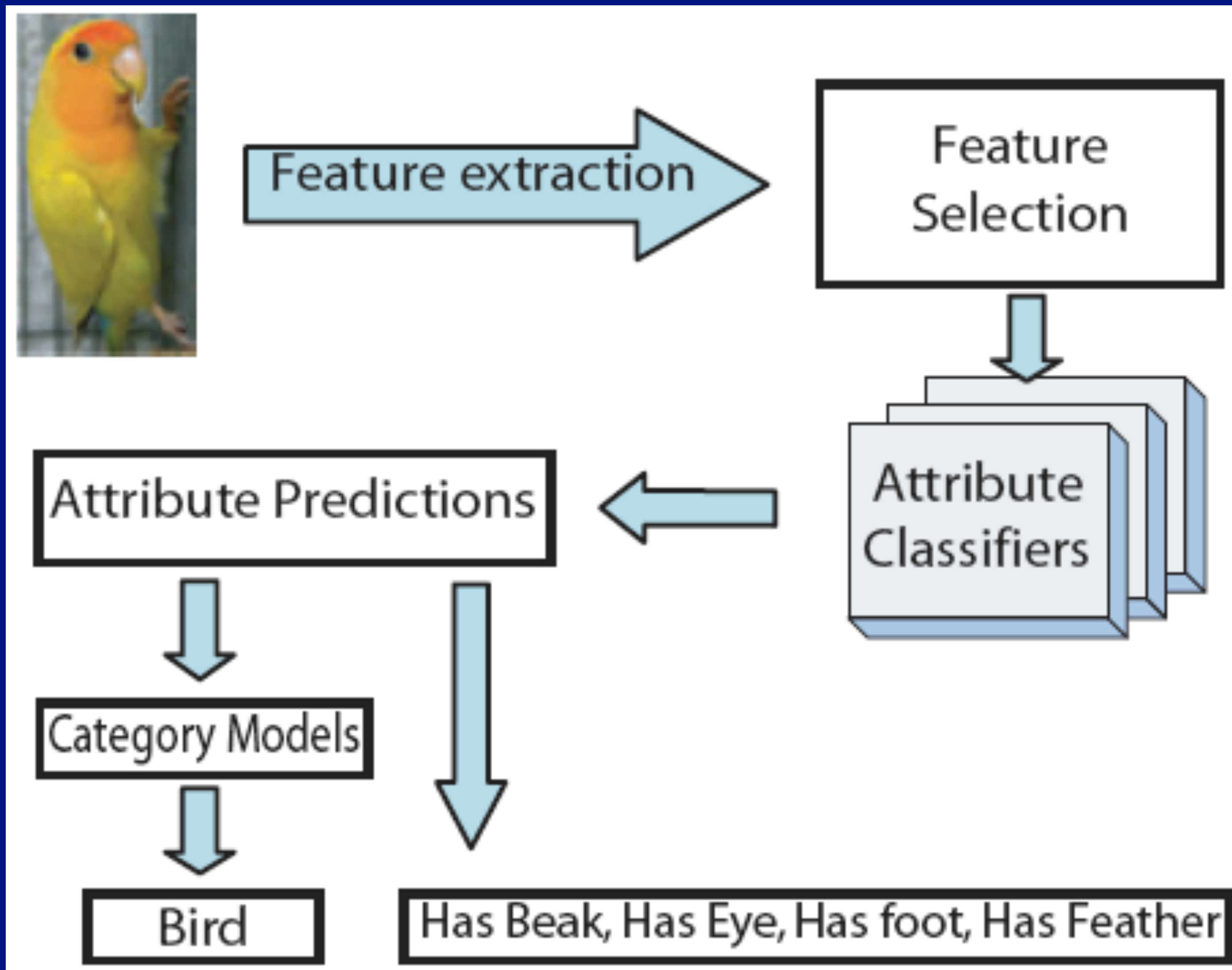


What is an object like?



Viz comic, issue 101

General architecture

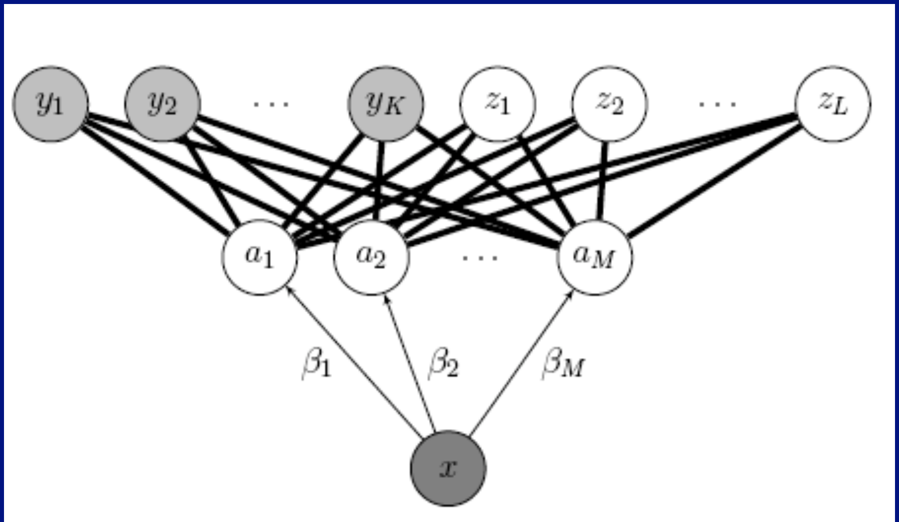


Farhadi et al 09; cf Lampert et al 09

Direct Attribute Prediction

Known classes

Unknown classes



Attribute layer

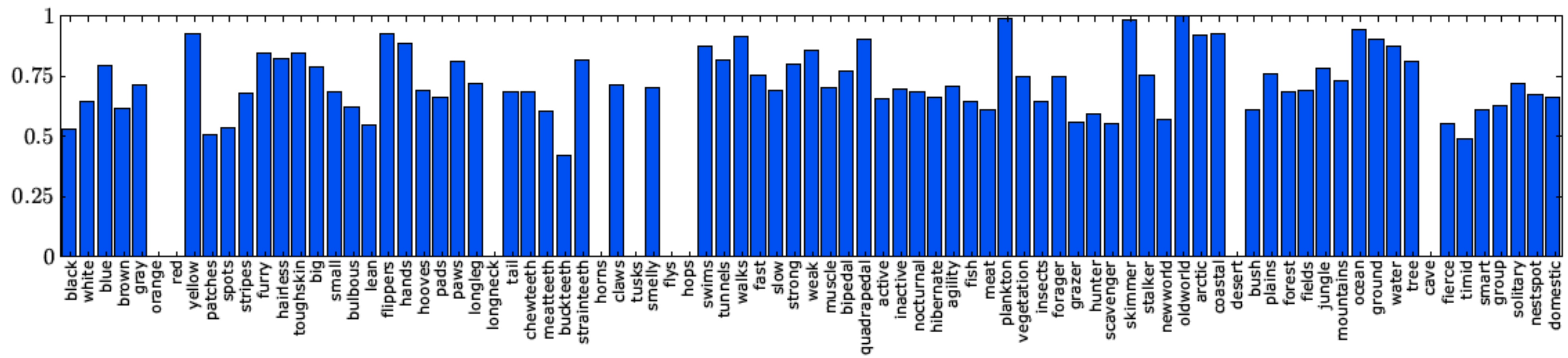
Image features

Lampert ea 09; Farhadi ea 09

Stuff attributes

Attribute predictions for unknown objects

Farhadi et al 09; cf Lampert et al 09



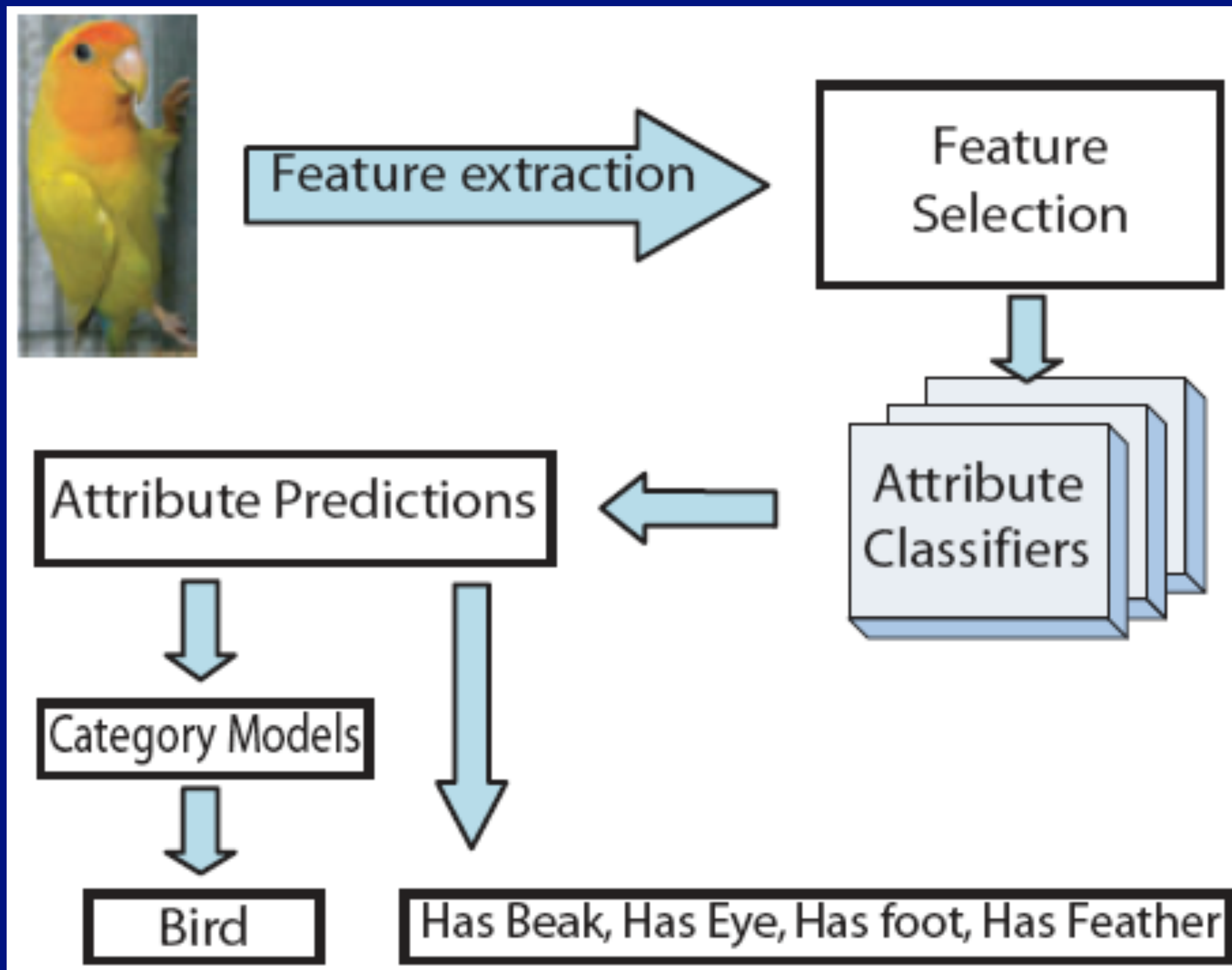
Lampert ea 09

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

Known objects could be unfamiliar

- By being different from the 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

General architecture

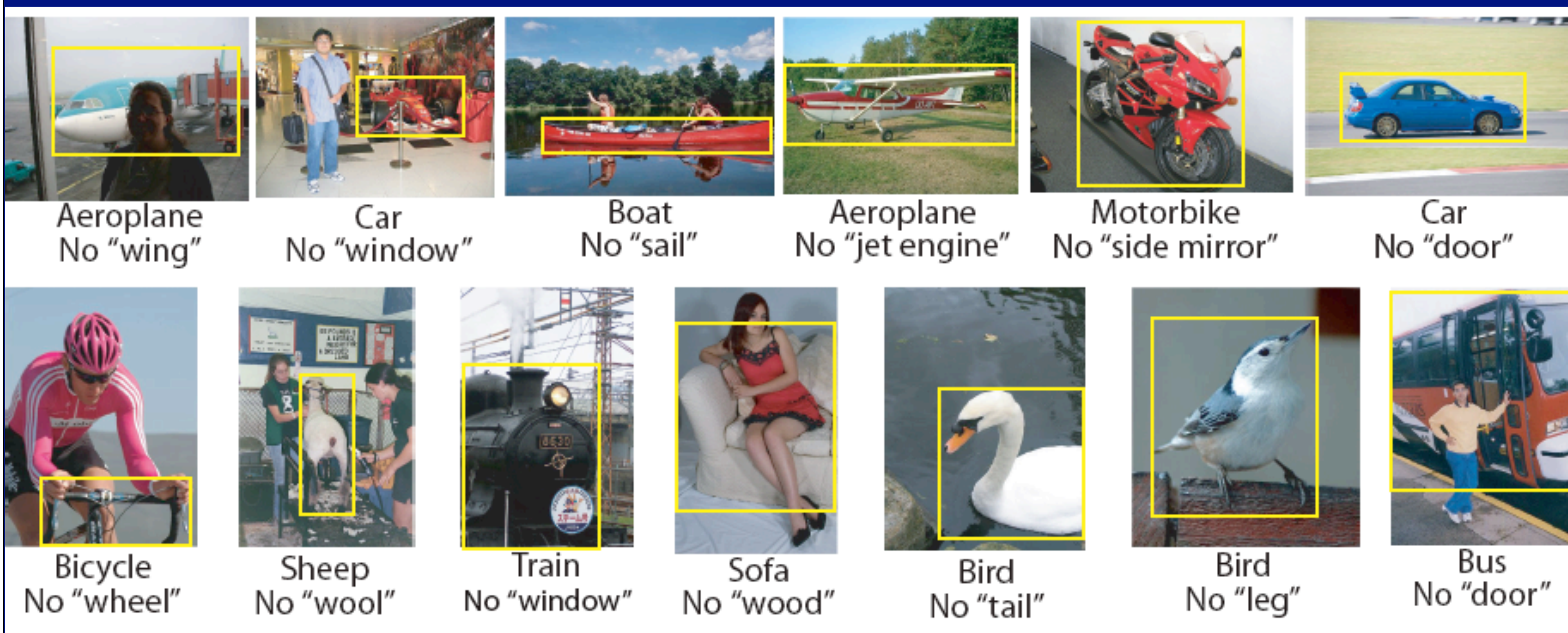


Farhadi et al 09; cf Lampert et al 09

"Man with a dog on a leash."

“Man in camouflage clothes restraining a vicious attack dog with a leash.”

Missing attributes



Extra attributes

Indirect Direct Attribute Prediction

Unknown classes

Attribute layer

Known classes

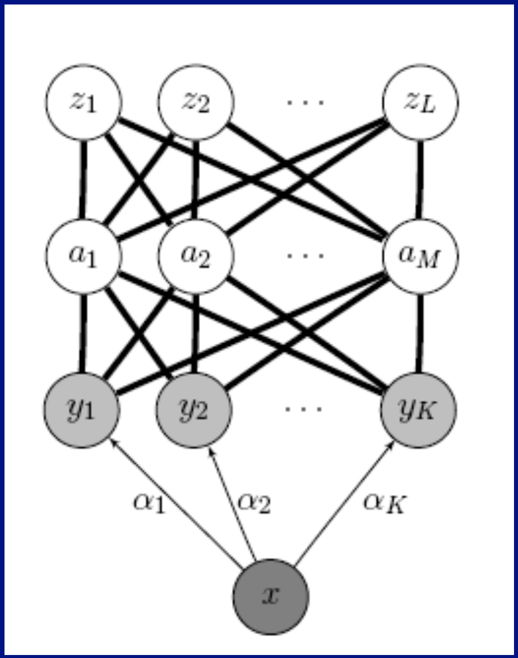
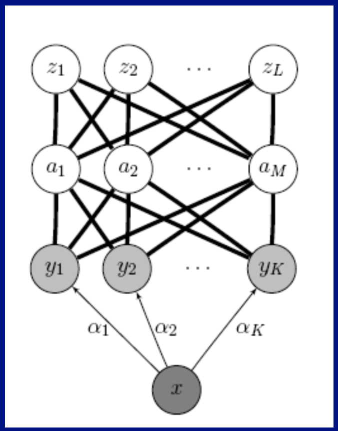


Image features

Lampert ea 09

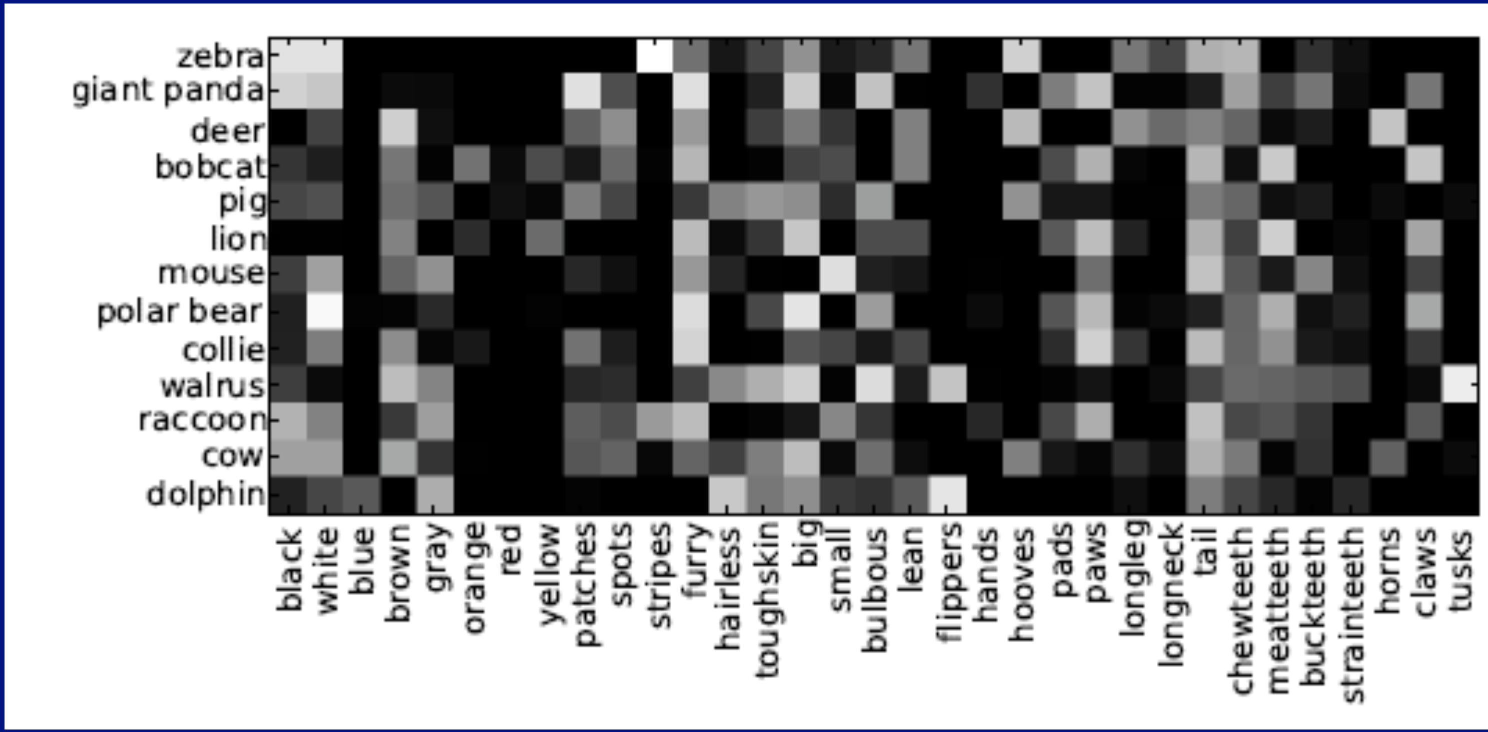
Stuff attributes

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

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

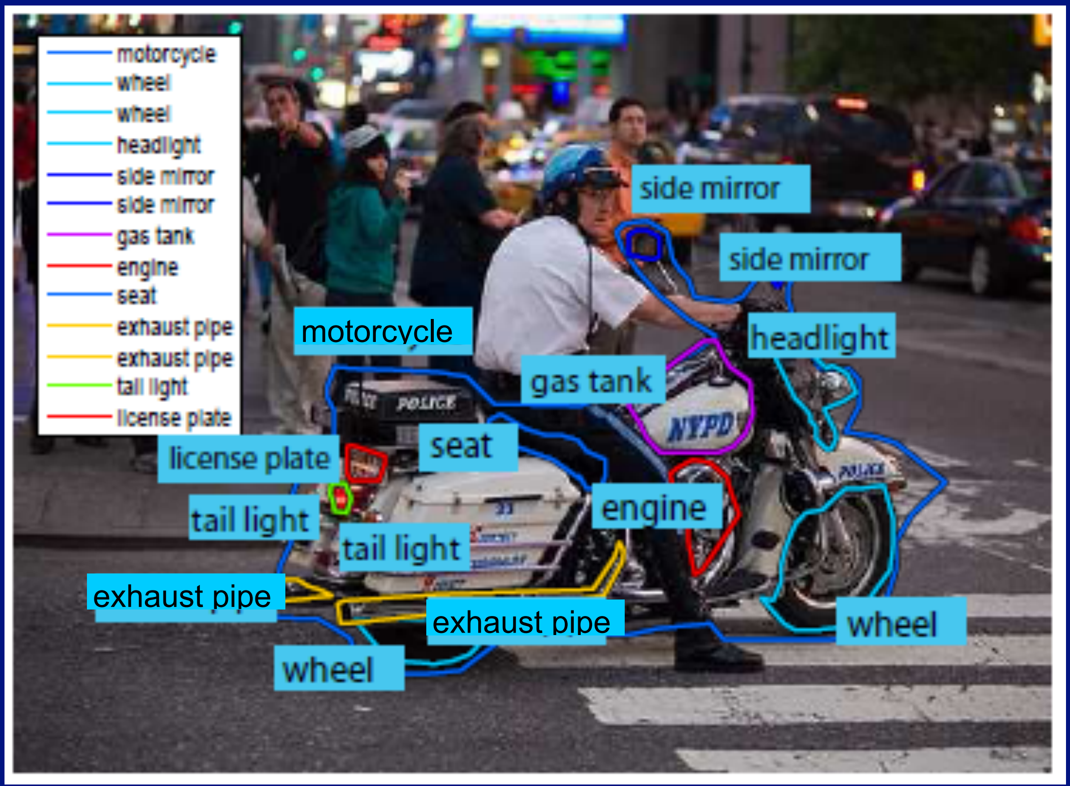
Datasets - II

- 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

UIUC PASCAL Sentence Dataset

- 5 Sentences from AMT: "Please describe the image in one complete but simple sentence."
- Quality control: qualification test + AMT grading task
- 8000 images for ~$1000

A large sheep standing between large trees in a rural area.

A ram stands in the middle of a group of trees.

The sheep is standing under the trees.

A sheep standing in a forest.

a sheep under pine trees

http://vision.cs.uiuc.edu/pascal-sentences/

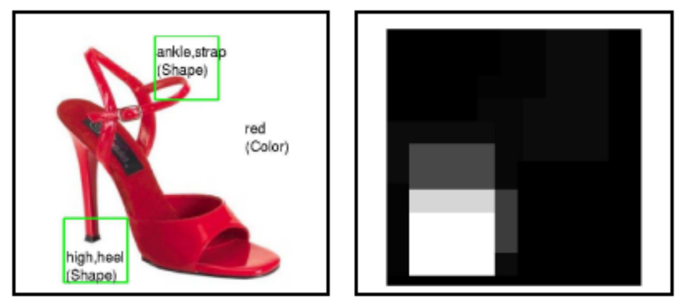
Rashtchian et al. HLT NAACL 2010

Attribute Discovery Data

- Gather pictures/captions of shoes, handbags, ties, earings, handbags
- Parse text into attributes
- Automatically learn which are visual
  - Visual attributes are more accurately classified
  - Human-Computer agreement on which attributes are visual: 70-90%
- Produces 37705 annotated examples
- Automatically characterize attribute localizability and type

The 12K pink and green gold leaves gently cascade down on these delicate beaded 10K gold earrings.

pink, green, gold, leaves, delicate, beaded



http://tamaraberg.com/attributesDataset/index.html

Berg et al. ECCV 2010

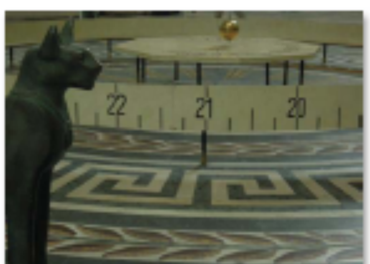
SBU Captioned Photo Dataset

- Query images with captions from Flickr
- Filter: minimum length, at least two words from keyword list, at least one spatial preposition
- Dataset contains 1,000,000 captioned images

Man sits in a rusted car buried in the sand on Waitarere beach.

Little girl and her dog in northern Thailand. They both seemed interested in what we were doing.

Interior design of modern white and brown living room furniture against white wall with a lamp hanging.



The Egyptian cat statue by the floor clock and perpetual motion machine in the pantheon.

Our dog Zoe in her bed.

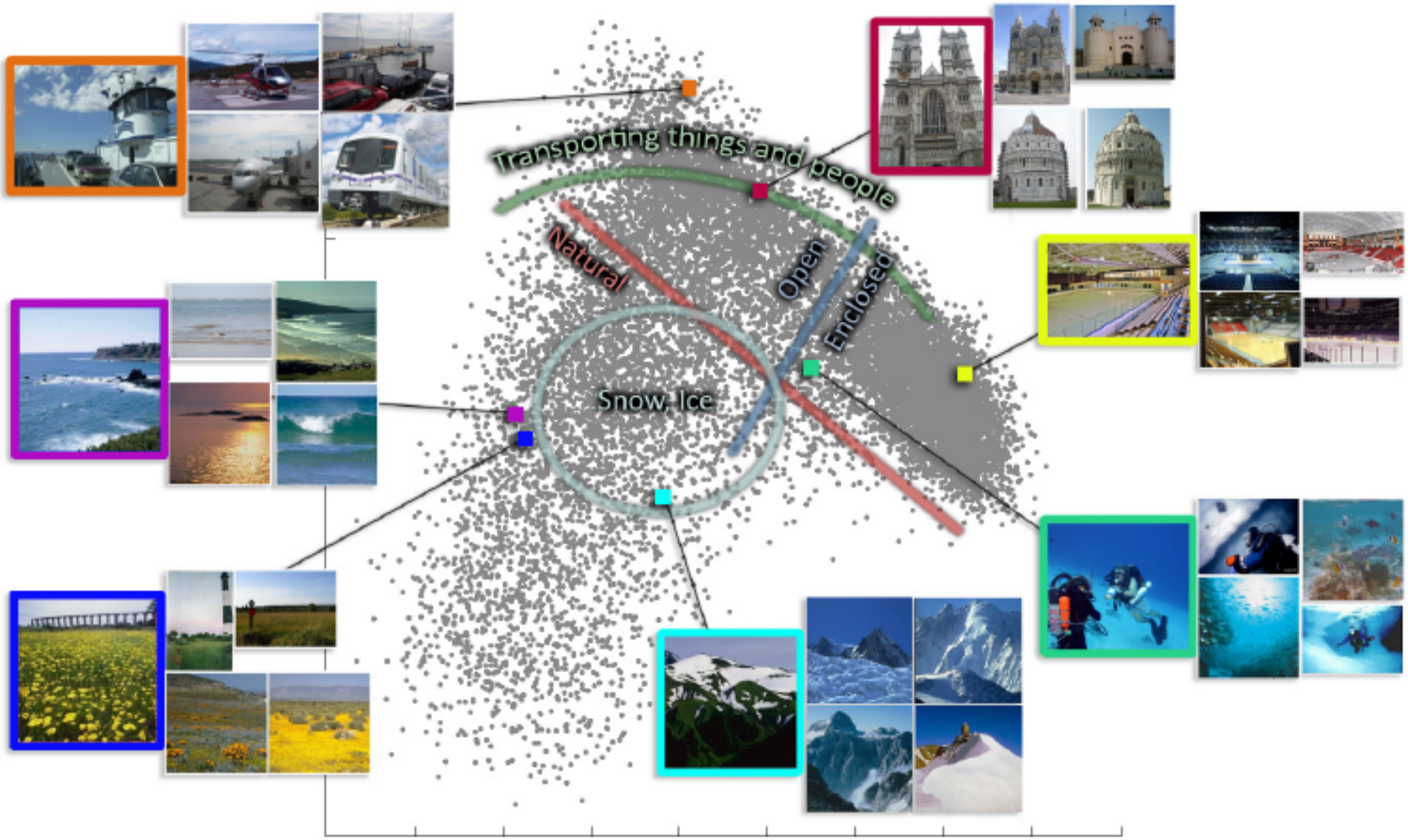
Emma in her hat looking super cute.

http://dsl1.cewit.stonybrook.edu/~vicente/sbucaptions/

Ordonez et al. NIPS 2011

Other Attribute Datasets

SUN Attributes Dataset

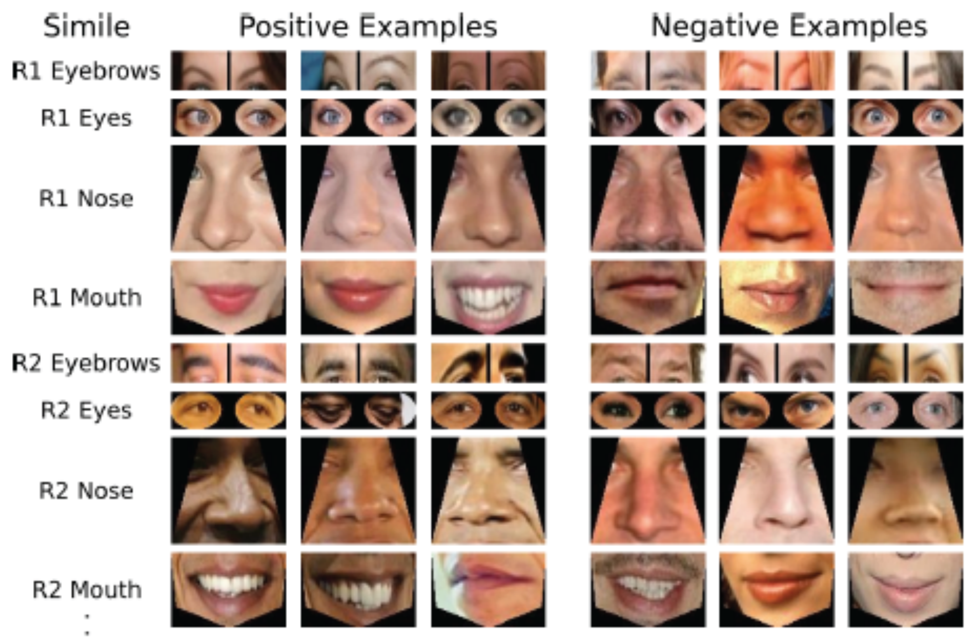


http://cs.brown.edu/~gen/sunattributes.html

Patterson Hays CVPR 2012

Other Attribute Datasets

PubFig



http://www.cs.columbia.edu/CAVE/databases/pubfig/ Kumar et al. ICCV 2009

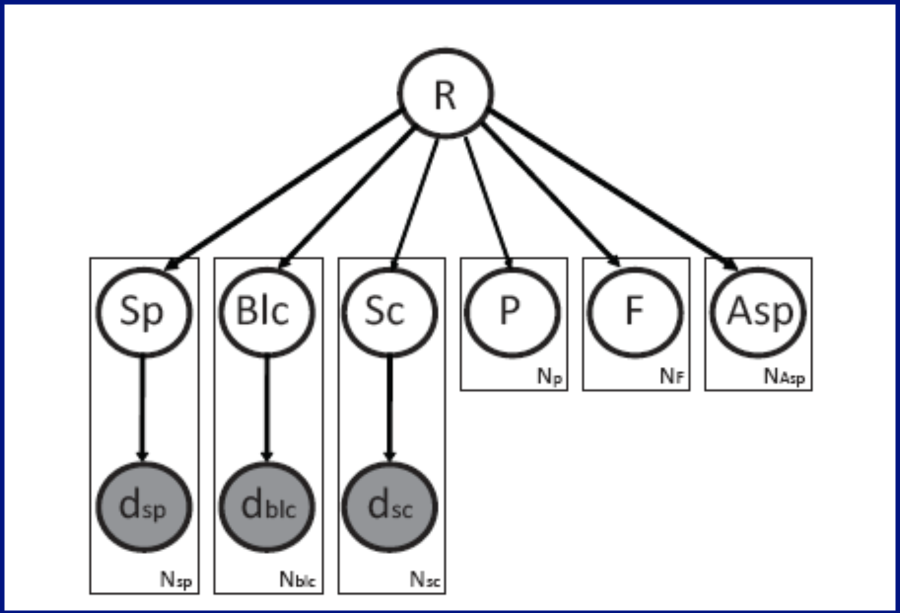
Latent Root

Root

Visual attributes

Other attributes

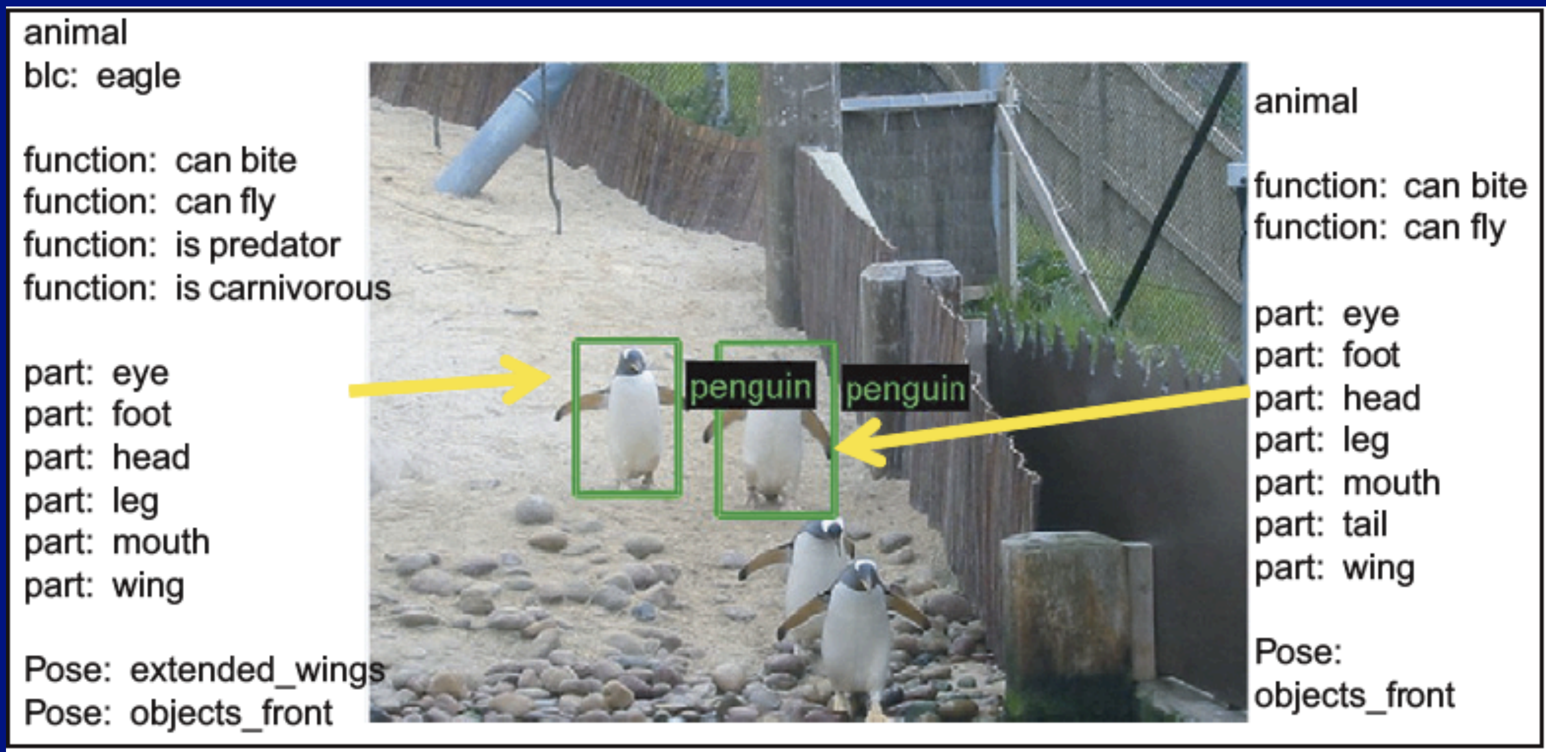
Detector Responses



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

Farhadi ea 10

P: predicate
F: functional attribute
Asp: aspect

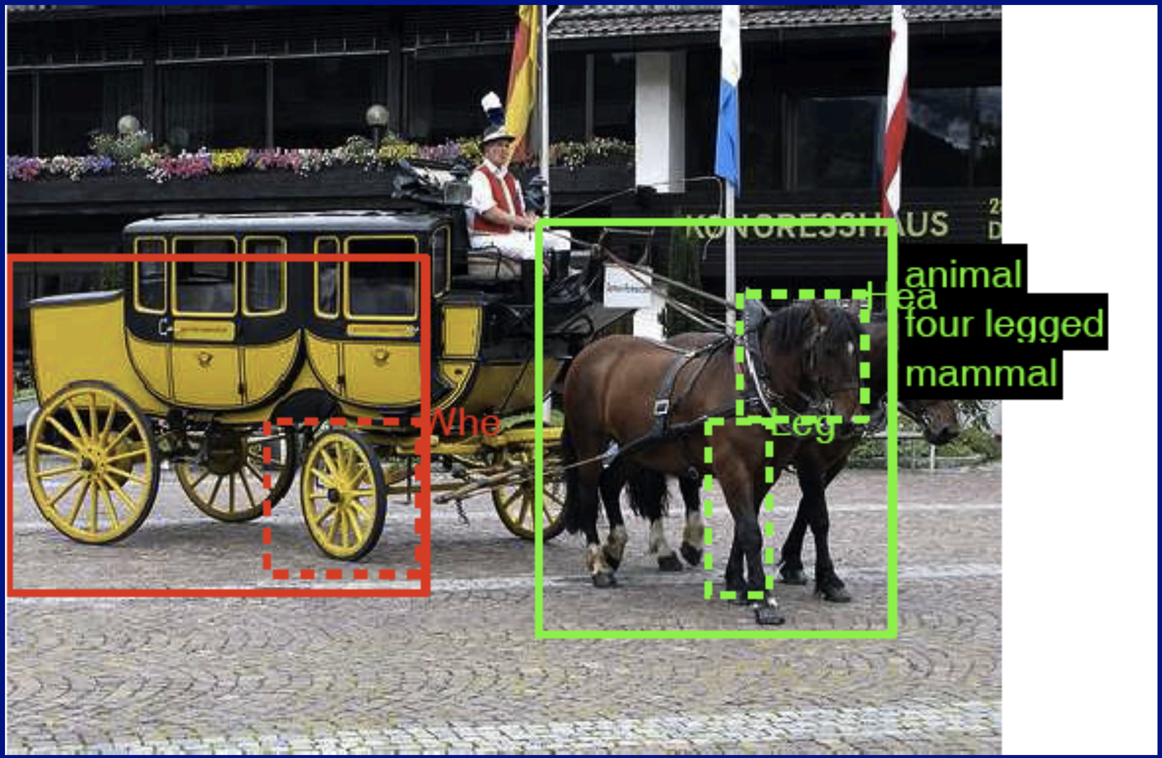


Farhadi ea 10

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

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - many meanings, useful in different contexts
- Important recognition technologies coming
  - attributes
  - phrases
  - geometry
  - sentences
- Crucial open questions
  - dataset bias
  - links to utility

Meaning comes in clumps

“Sledder”
Is this one thing?
Should we cut her off her sled?

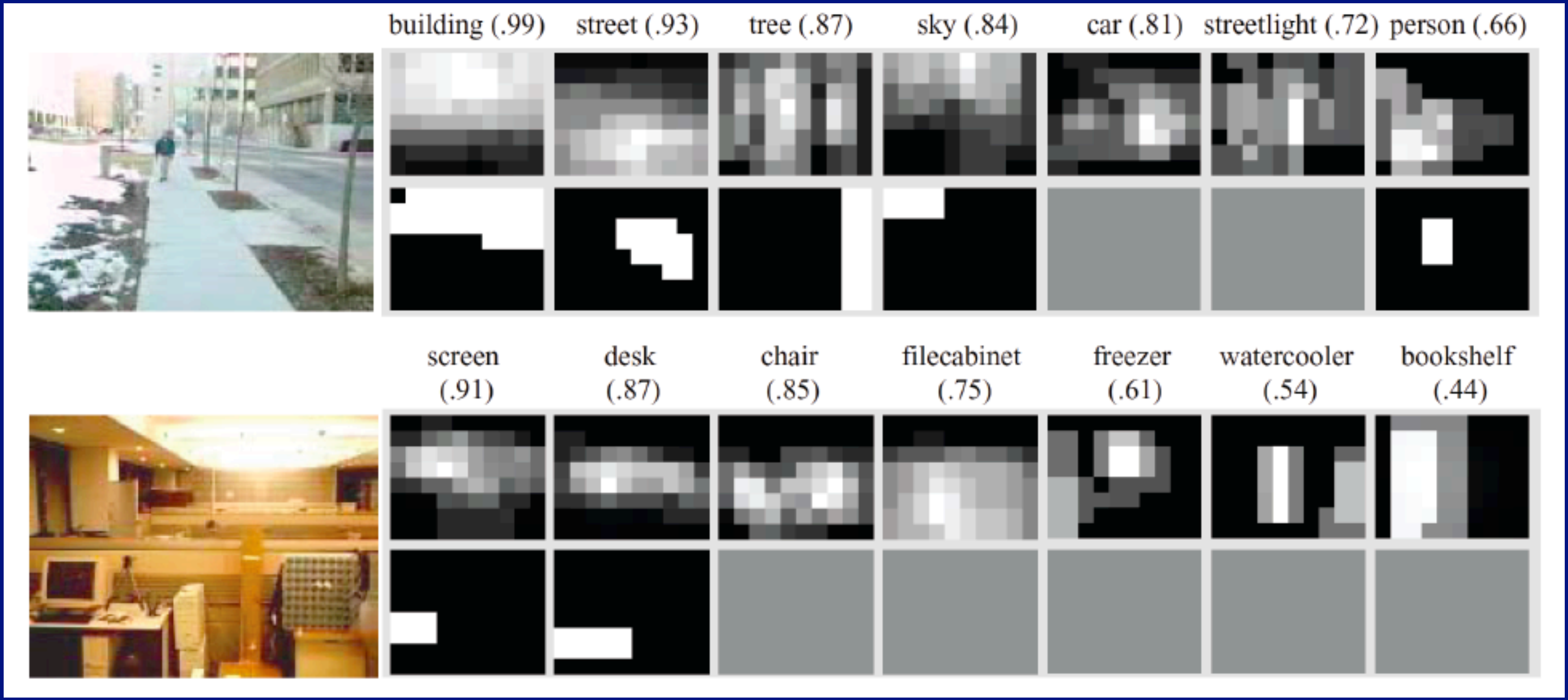
Scenes

- Likely stages for
  - Particular types of object
  - Particular types of activity

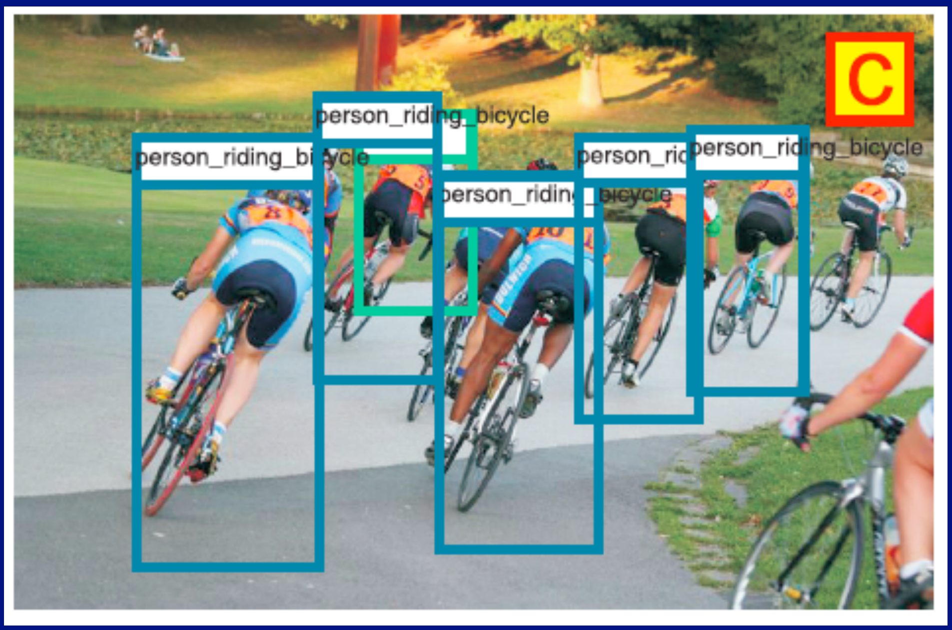
Xiao et al 10

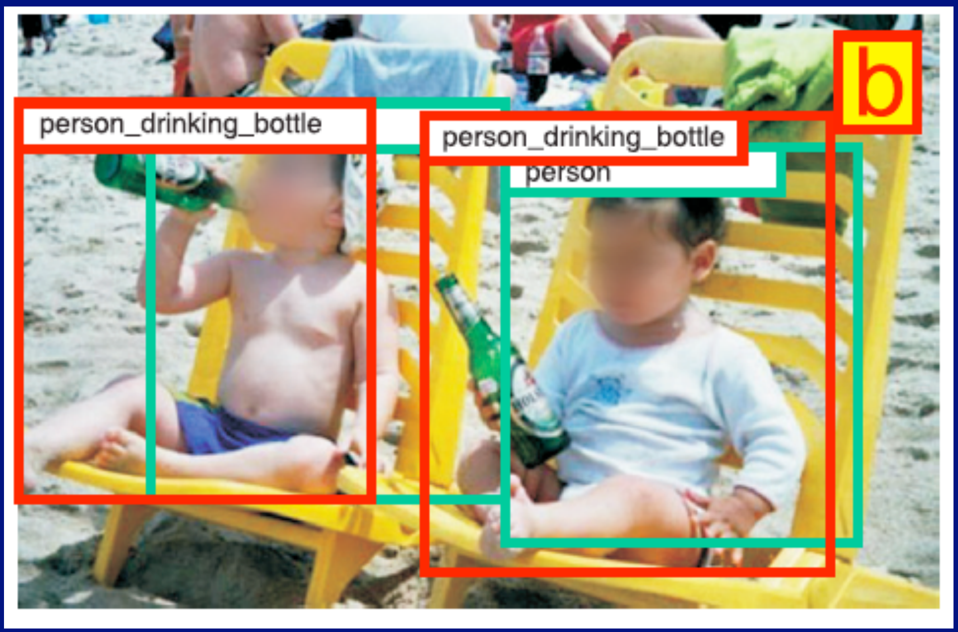
Scenes

Torralba et al ’93



Scenes > Visual phrases > Objects





- Composites
  - easier to recognize than their components
  - because appearance is simpler

Farhadi + Sadeghi 11

Issue: what should one recognize?

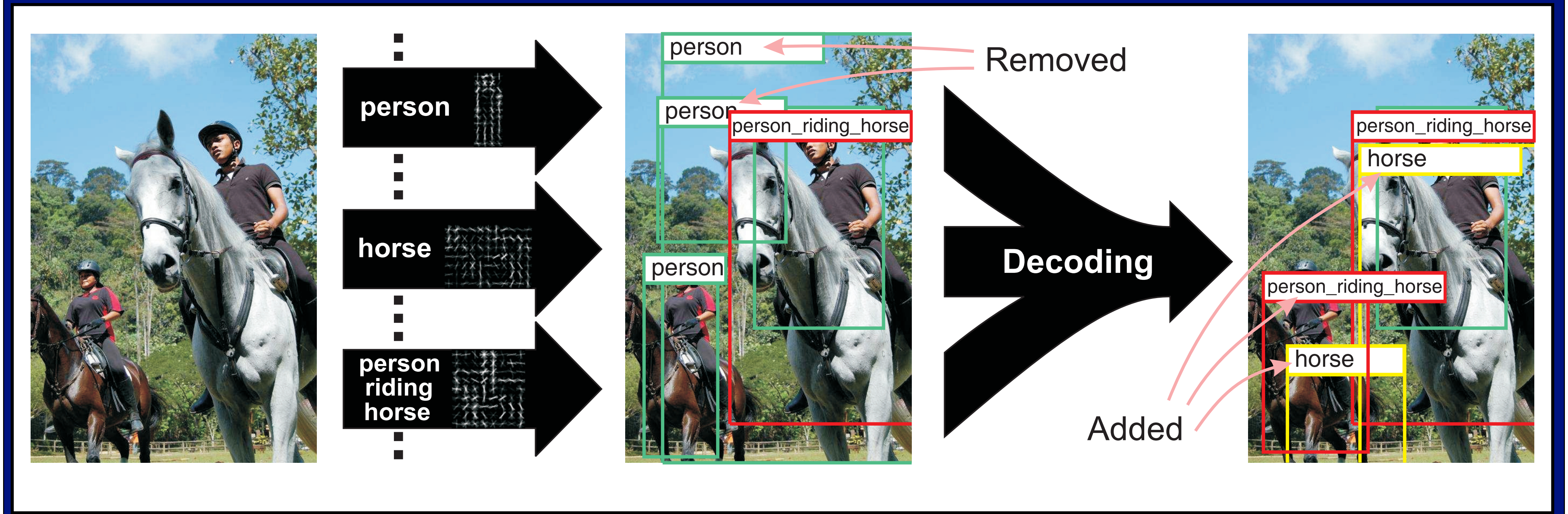
- Single objects
  - potentially inaccurate
  - which ones?
  - crosstalk between detectors
    - eg bottles and humans

- Visual phrases
  - chosen opportunistically for accuracy
  - potentially far too many
  - which ones?
  - crosstalk between detectors
    - eg bottle, person, person drinking from bottle, etc.

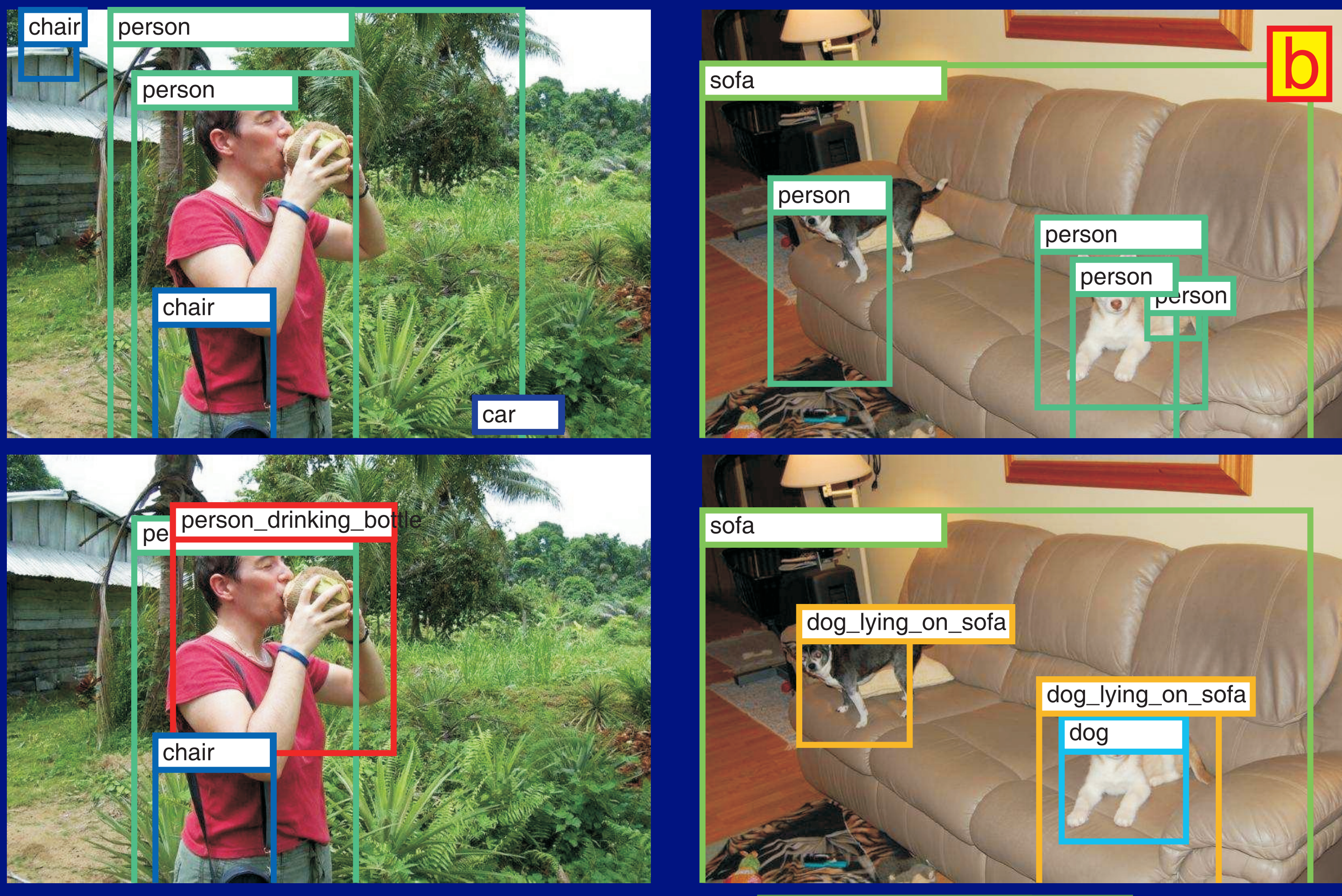
Decoding

- Take pool of detector responses, decide what to believe
- Standard pastime, usually unremarked
    - eg test against threshold (pretty much everyone, all the time)
    - eg greedy algorithm (Desai et al 09; Kang et al 06)
    - vote on location (Bourdev+Malik 09)
    - single classifier looks at all best detector responses (Maji et al 11)
    - each detector response retested, using others as features (Farhadi Sadeghi 11)
- Probably much richer topic than currently allowed

Decoding


person
horse
person riding horse
person
person
person_riding_horse
person
Removed
Decoding
Added
person_riding_horse
horse
person_riding_horse
horse

Decoding helps

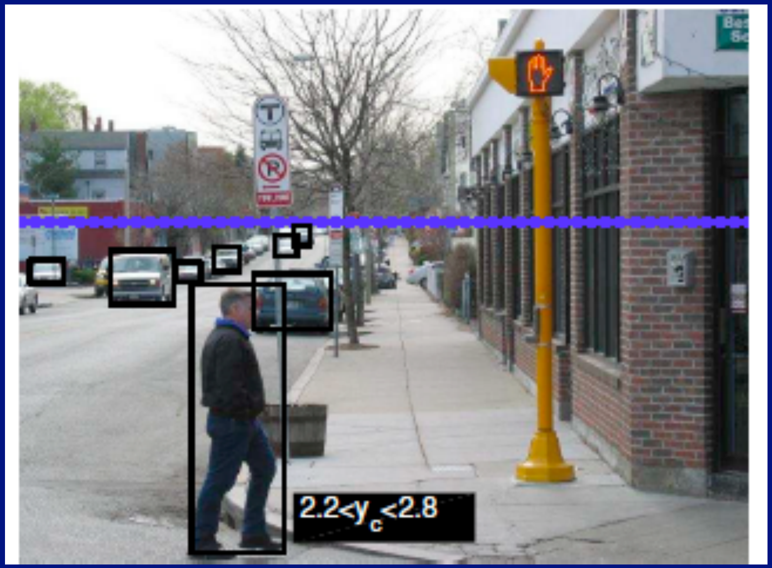


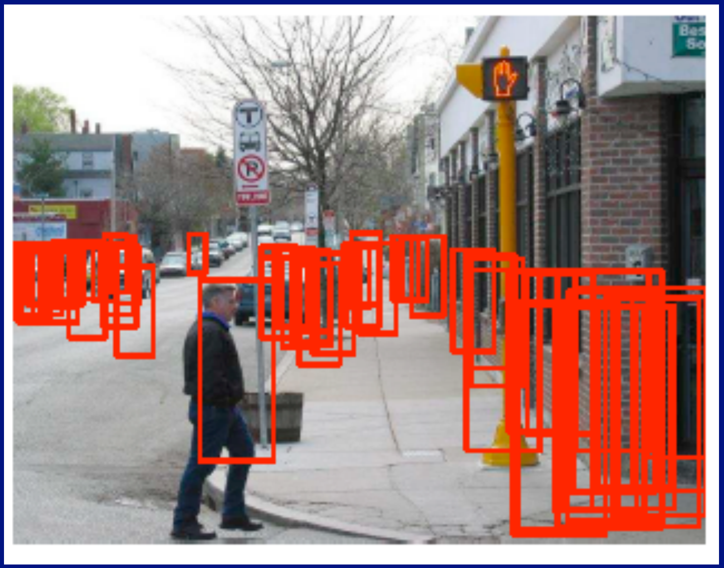
Farhadi Sadeghi 11

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
- Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - selection
- Crucial open questions
  - dataset bias
  - links to utility

Environmental knowledge





Hoiem et al 06

Environmental knowledge is powerful

(b) Local Detection (b) Full Model Detection

(a) Local Detection (a) Full Model Detection

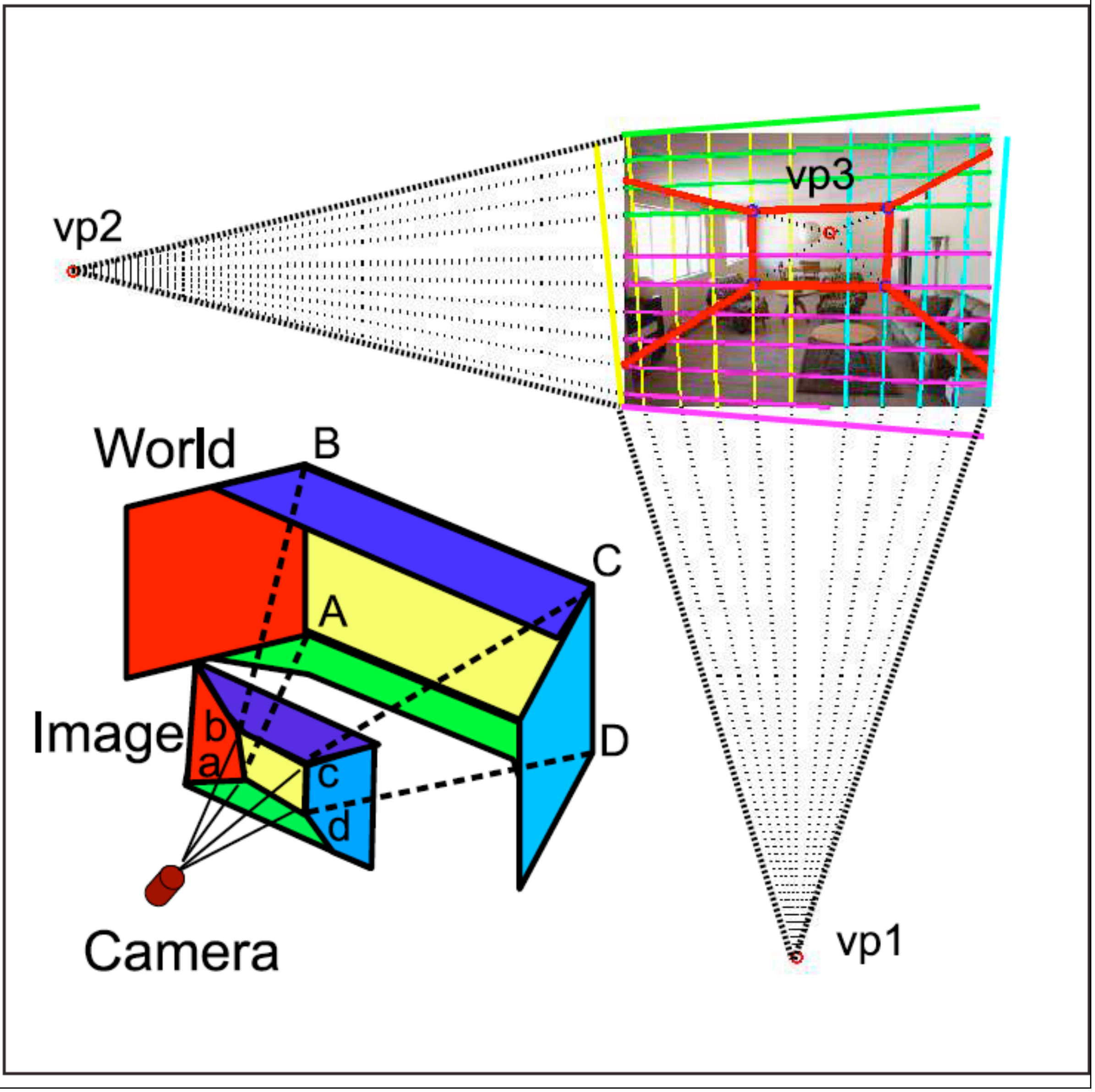
Hoiem et al 06

Vanishing points

- Cluster long straight edges into three clusters (after Rother, 02)

Estimating layout

- Choice of layout= 4DOF in image
- Search cost function
  - learned from examples

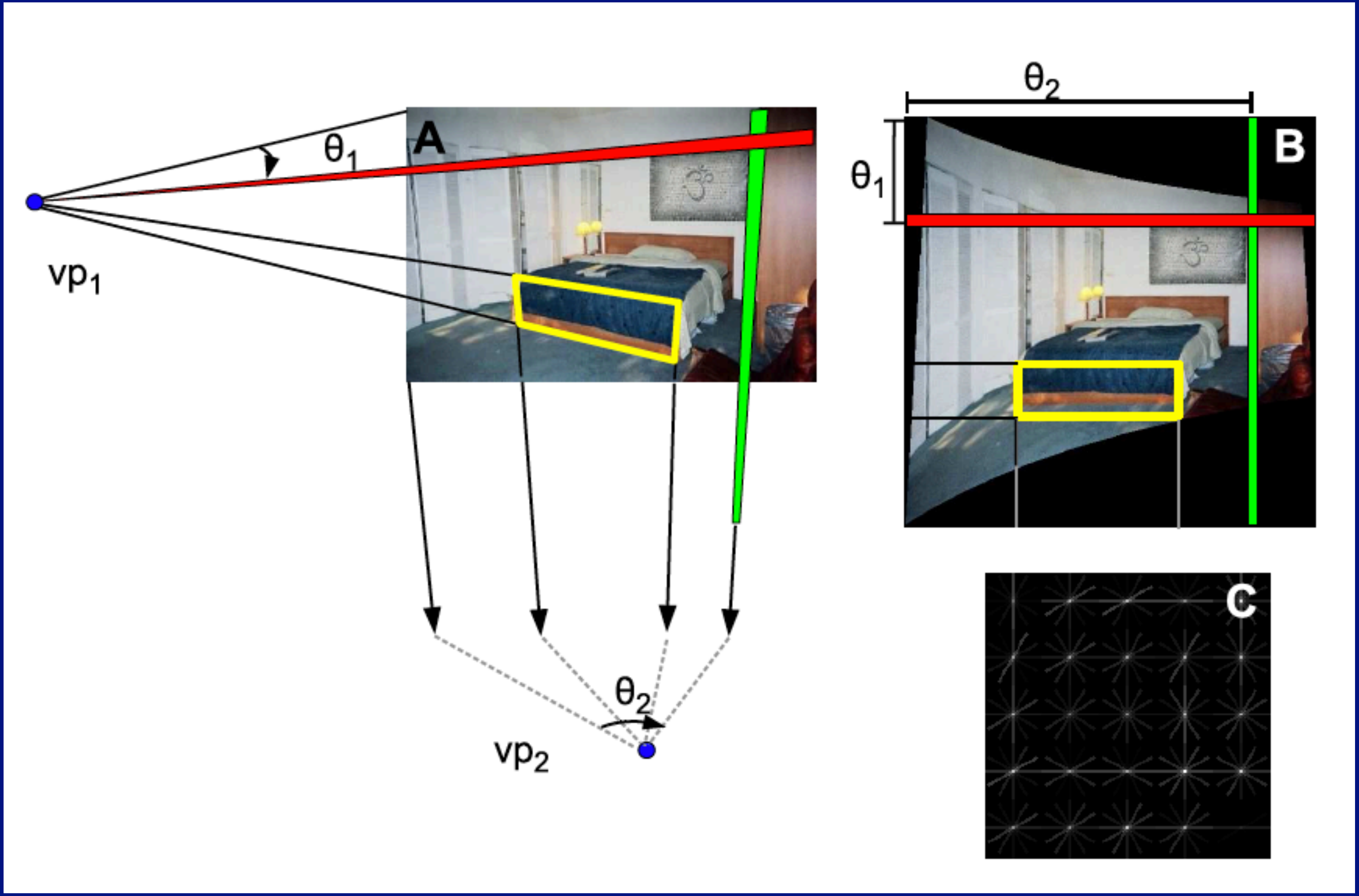


Clutter maps

Detecting beds

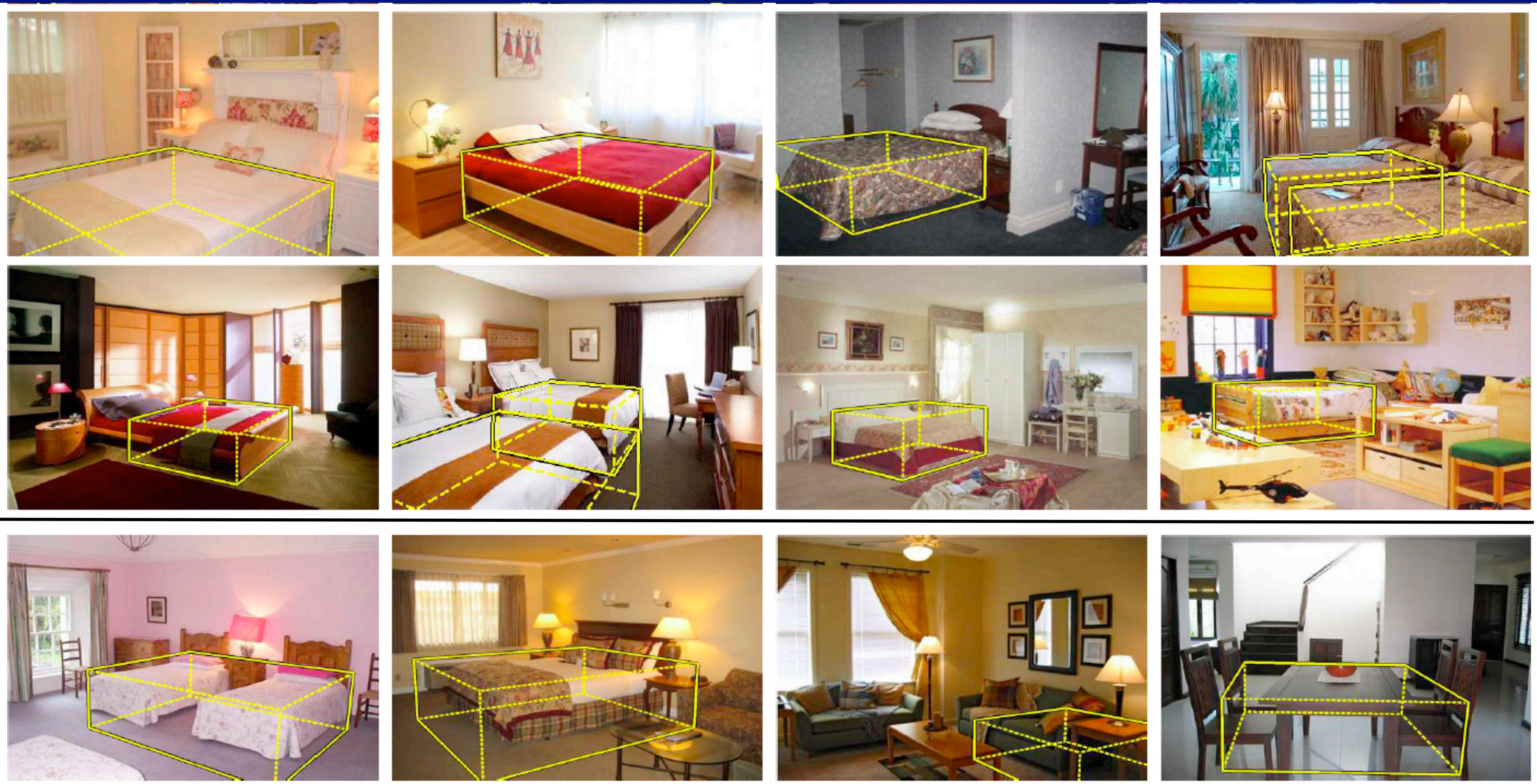
- Natural strategy
  - mark up data, apply Felzenswalb et al, '08 (FMRG)
- Problem
  - changes in viewpoint lead to changes in appearance
    - FMRG doesn't know this - must be less efficient
- Using a room box
  - rectify the image to each face of the room box
  - look for FACES of beds in each rectified image using FMRG
  - find three that share a corner

Detecting beds - I



Detecting beds - II

True positives

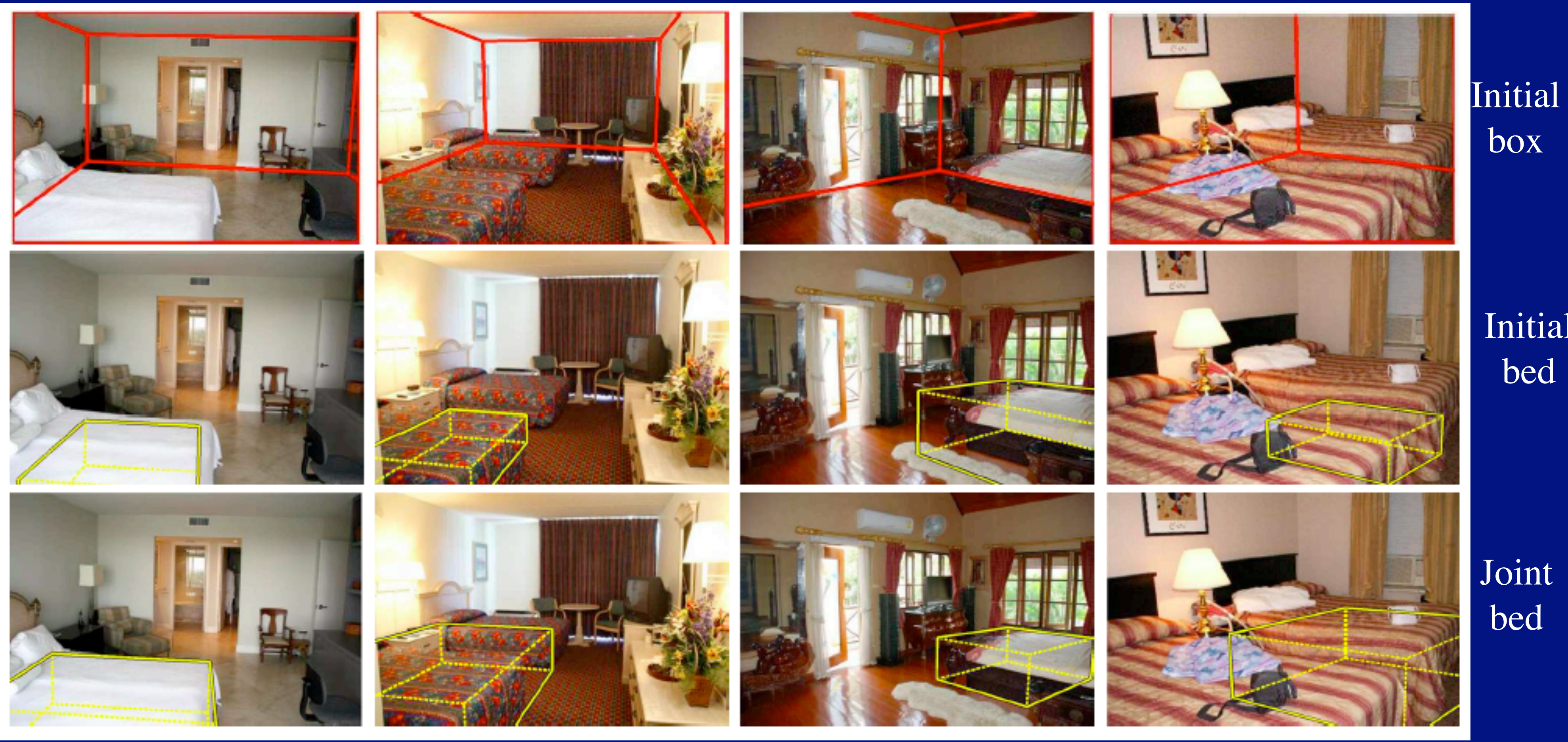


False positives

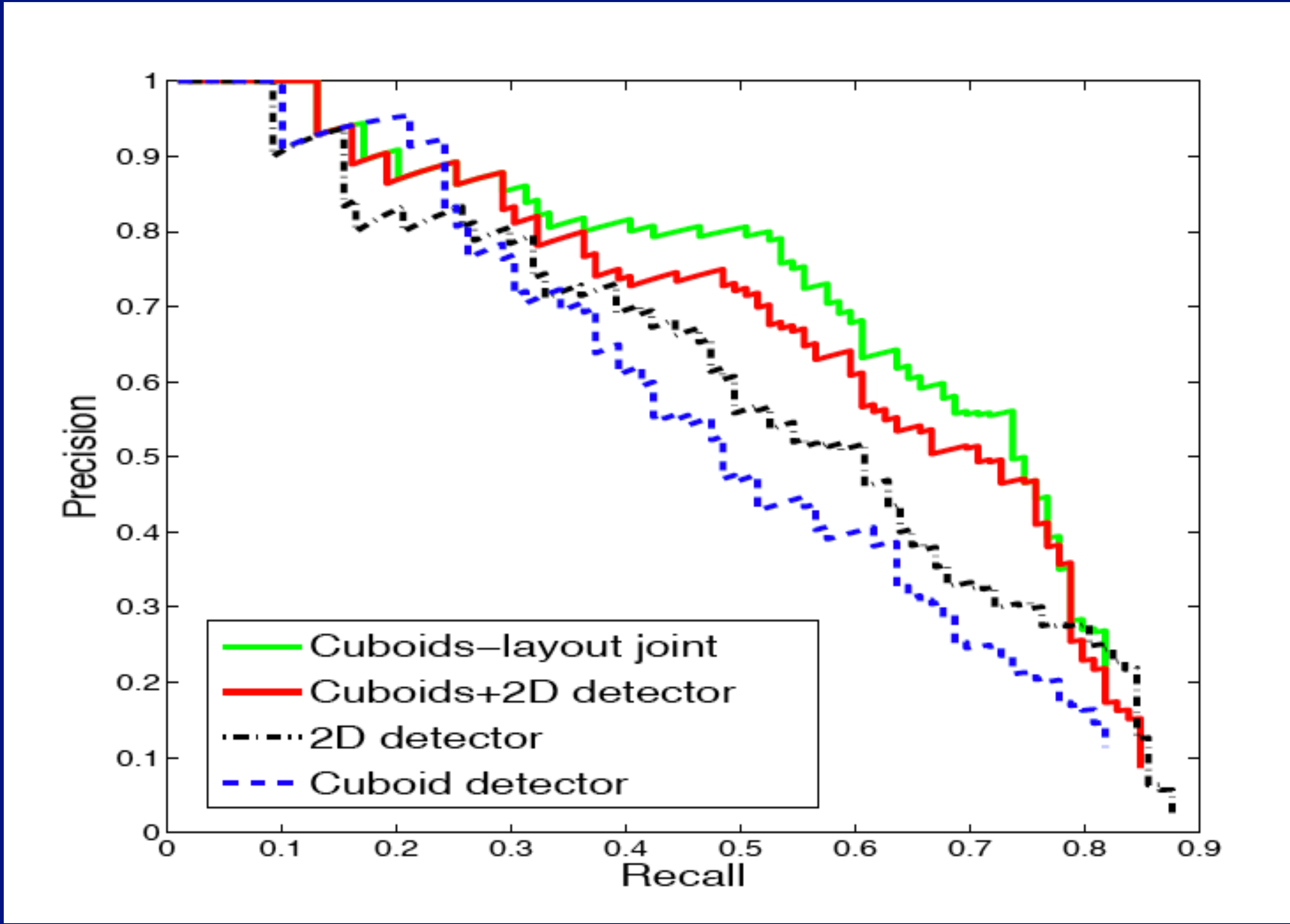
Detecting beds - III

- Beds constrain rooms
  - are axis-aligned
  - can’t pierce walls
- Variants
  - Box only (OK)
  - Box + 2D (better)
  - Jointly estimate room box, bed box(es) (best)

Joint estimation helps



Joint estimation helps



Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - many meanings, useful in different contexts
- Important recognition technologies coming
  - attributes
  - phrases
  - geometry
  - selection
- Crucial open questions
  - dataset bias
  - links to utility

Selection: What is worth saying?

Two girls take a break to sit and talk .

Two women are sitting , and one of them is holding something .

Two women chatting while sitting outside

Two women sitting on a bench talking .

Two women wearing jeans , one with a blue scarf around her head , sit and talk .

Sentences from Julia Hockenmaier’s work

Rashtchian ea 10

For language people: Pragmatics - what is worth saying?

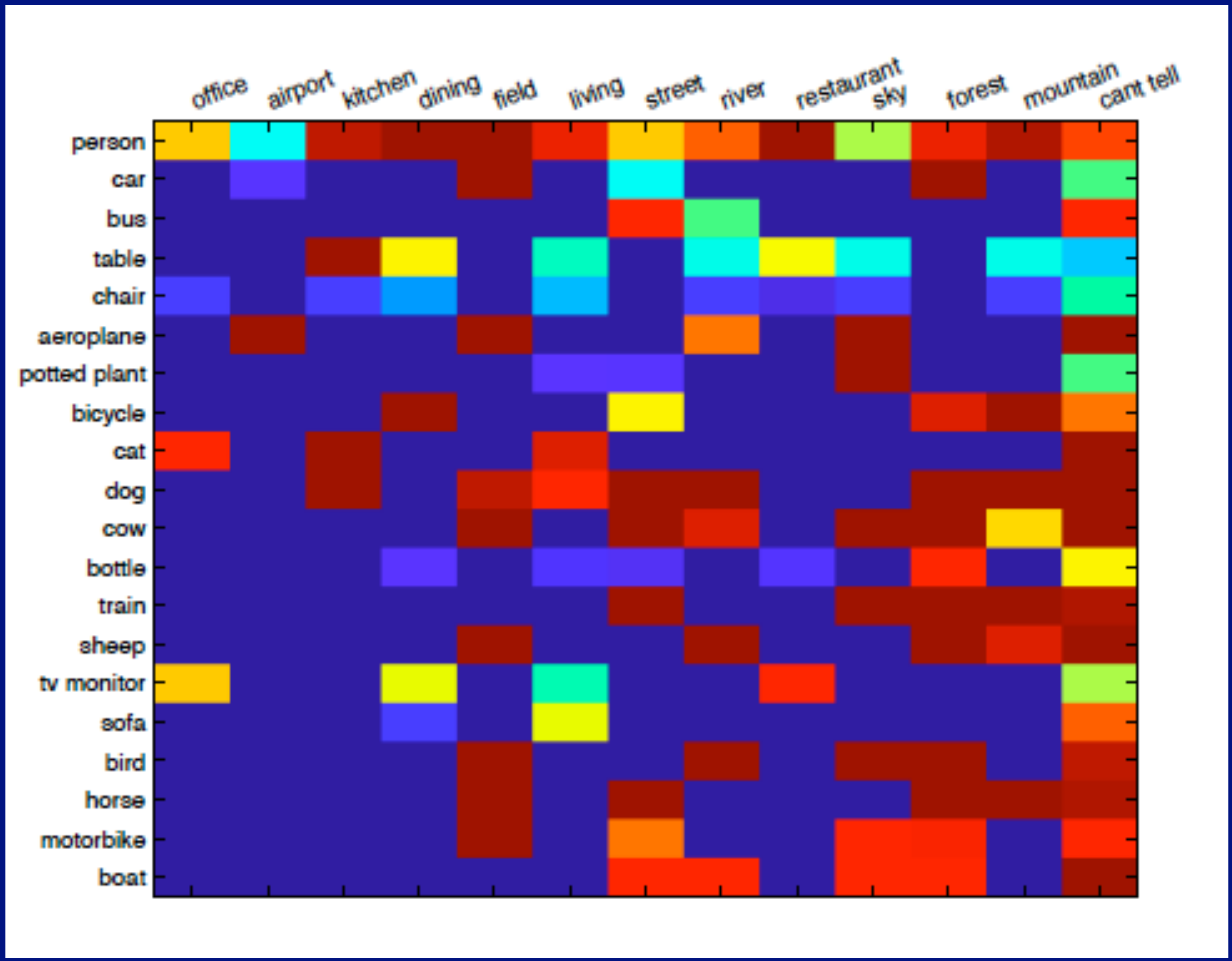
- Some factors conducive to being mentioned (Berg ea 12)

Top10	Prob	Last10	Prob
firework	1.00	hand	0.15
turtle	0.97	cloth	0.15
horse	0.97	paper	0.13
pool	0.94	umbrella	0.13
airplane	0.94	grass	0.13
bed	0.92	sidewalk	0.11
person	0.92	tire	0.11
whale	0.91	smoke	0.09
fountain	0.89	instrument	0.07
flag	0.88	fabric	0.07

Table 1. Probability of being mentioned when present for various object categories (ImageCLEF).

Context

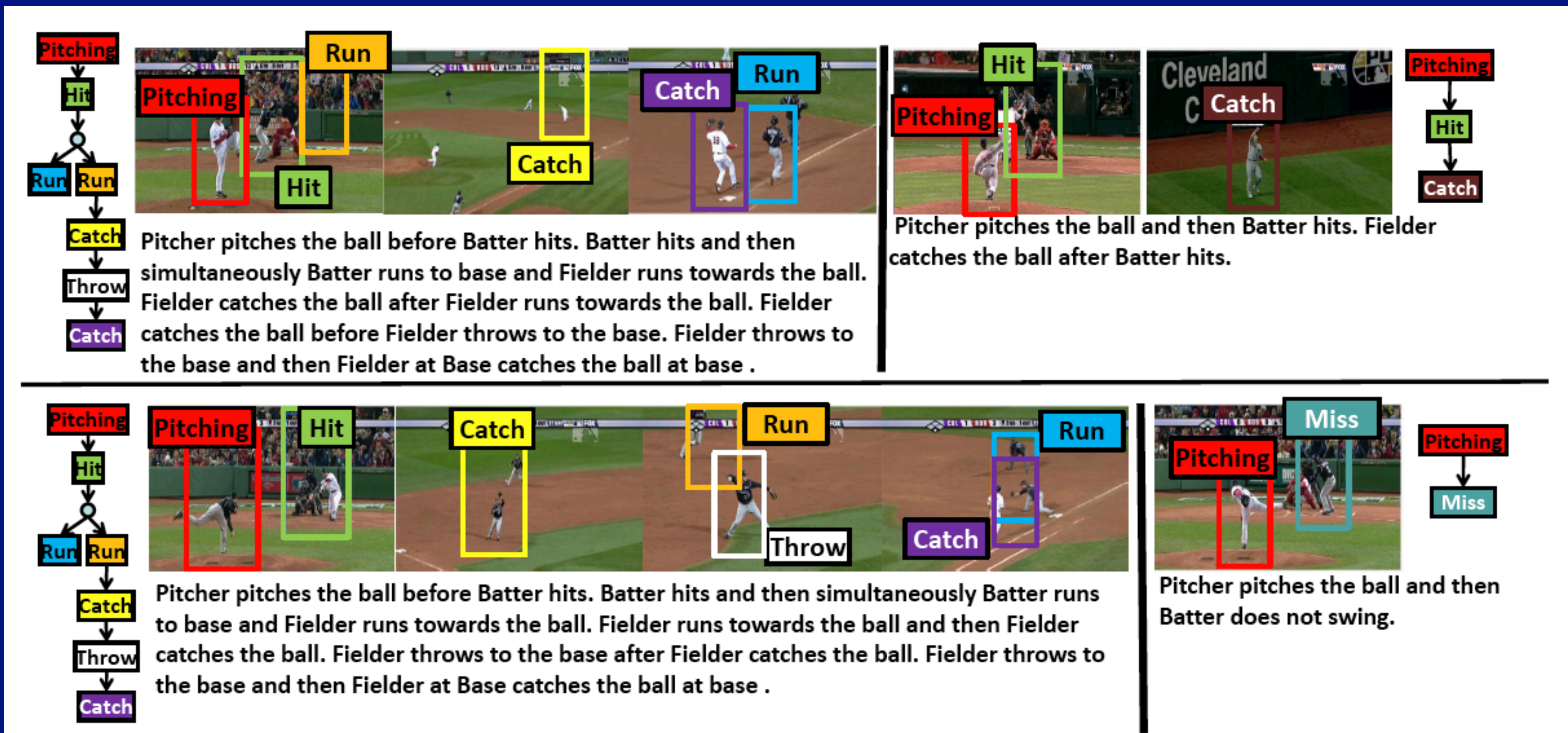
Object



Red - high
Blue - low

Berg et al 12: Objects more likely to be mentioned in uncommon context, figure shows probability of being mentioned conditioned on appearing

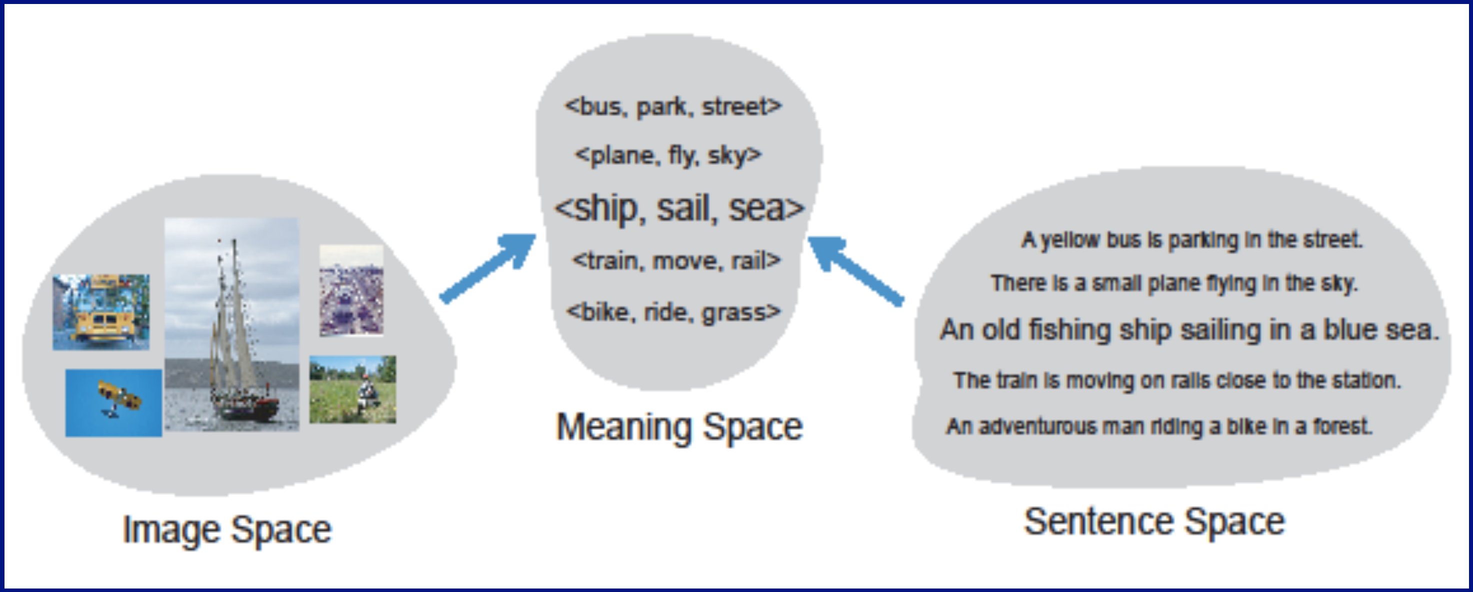
Predicting stylized narrations



Gupta ea 09

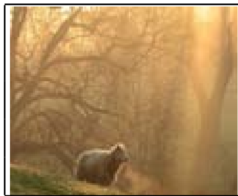
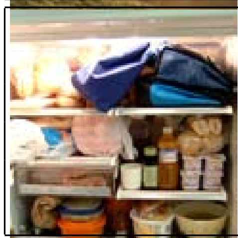
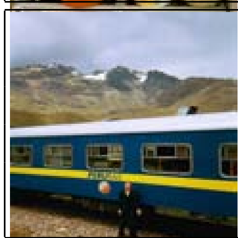
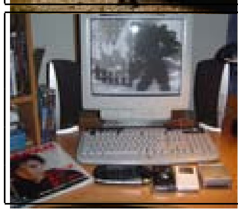
Rich(ish) sentences from simple intermediates

Object, action, scene



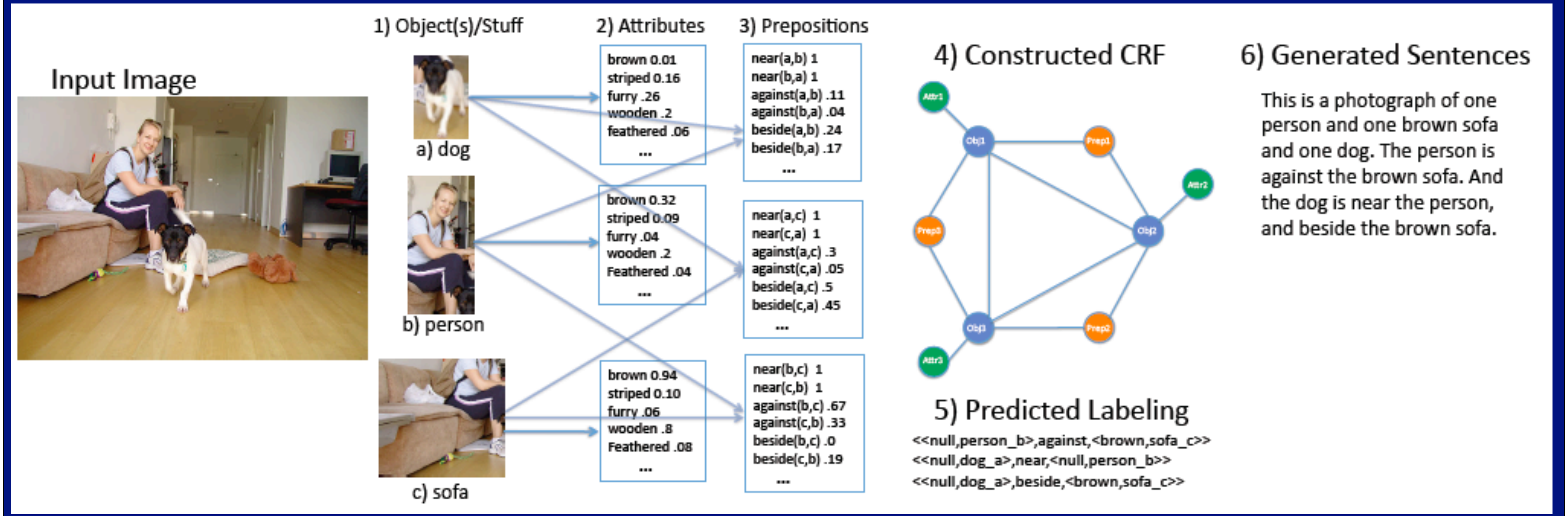
Farhadi ea 10

Examples

	(pet, sleep, ground) (dog, sleep, ground) (animal, sleep, ground) (animal, stand, ground) (goat, stand, ground)	see something unexpected. Cow in the grassfield. Beautiful scenery surrounds a fluffly sheep. Dog hearding sheep in open terrain. Cattle feeding at a trough.
	(furniture, place, furniture) (furniture, place, room) (furniture, place, home) (bottle, place, table) (display, place, table)	Refrigerator almost empty. Foods and utensils. Eatables in the refrigerator. The inside of a refrigerator apples, cottage cheese, tupperwares and lunch bags. Squash apenny white store with a hand statue, picnic tables in front of the building.
	(transportation, move, track) (bike, ride, track) (transportation, move, road) (pet, sleep, ground) (bike, ride, road)	A man stands next to a train on a cloudy day A backpacker stands beside a green train This is a picture of a man standing next to a green train There are two men standing on a rocky beach, smiling at the camera. This is a person laying down in the grass next to their bike in front of a strange white building.
	(display, place, table) (furniture, place, furniture) (furniture, place, furniture) (bottle, place, table) (furniture, place, home)	This is a lot of technology. Somebody's screensaver of a pumpkin A black laptop is connected to a black Dell monitor This is a dual monitor setup Old school Computer monitor with way to many stickers on it

Farhadi ea 10

Adding Attributes and Prepositions



Kulkarni et al 11

Adding Attributes and Prepositions

This is a photograph of one sky, one road and one bus. The blue sky is above the gray road. The gray road is near the shiny bus. The shiny bus is near the blue sky.

There are two aeroplanes. The first shiny aeroplane is near the second shiny aeroplane.

There are one cow and one sky. The golden cow is by the blue sky.

There are one dining table, one chair and two windows. The wooden dining table is by the wooden chair, and against the first window, and against the second white window. The wooden chair is by the first window, and by the second white window. The first window is by the second white window.

Here we see one person and one train. The black person is by the train.

This is a picture of one sky, one road and one sheep. The gray sky is over the gray road. The gray sheep is by the gray road.

Here we see one road, one sky and one bicycle. The road is near the blue sky, and near the colorful bicycle. The colorful bicycle is within the blue sky.

Here we see two persons, one sky and one aeroplane. The first black person is by the blue sky. The blue sky is near the shiny aeroplane. The second black person is by the blue sky. The shiny aeroplane is by the first black person, and by the second black person.

This is a picture of two dogs. The first dog is near the second furry dog.

This is a photograph of two buses. The first rectangular bus is near the second rectangular bus.

Kulkarni et al 11

Integer Program

ILP: Found ***trucks*** parked on first avenue in the east village.

HMM: This is the first ***cellar door*** left back bedroom in center and clothes dryer to the right to the building in the house. This ***HUGE screen*** hanging on the wall outside a burned down building in the house. ***My truck*** parked on first avenue in the east village by the glass buildings in the house.

Human: Flat bed Chisholms truck on display at the vintage vehicle rall y at Astley Green Colliery near Leigh Lancs

ILP:
This is a photo of this little flower sprouted up in defiance against grass. Bright yellow flowers growing in a rock garden at Volcan Mombacho.

HMM: These was taken on the flowers growing in a rock garden in the field in two sorts. This little flower sprouted up in defiance in the field in two sorts. A full open flower sprouted up in defiance in the field in gardens. Bright yellow flowers growing in a rock garden in the field.

Human: Yellow flower in my field

ILP: I think this is a boy's bike lied in saltwater for quite a while.

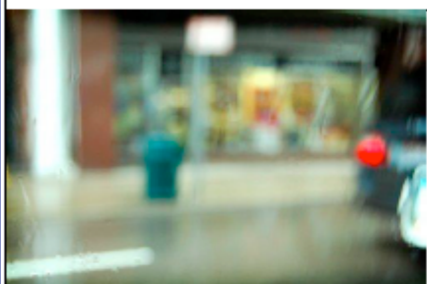
HMM: I liked the way ***bicycles*** leaning against a wall in Copenhagen Denmark in a windy sky in a Singapore bathroom. ***Boy's bike*** lied in saltwater for quite a while in a windy sky in a Singapore bathroom. ***Fruit*** rubbing his face in the encrusted snow in a windy sky in a Singapore bathroom.

Human: You re nobody in Oxford, unless you have a old bike with a basket

Use an integer program to enforce discourse, etc constraints (objects should not be mentioned repeatedly)
ILP: Method (Berg ea 12, ACL paper)
HMM: Yang et al 11 (cf Kulkarni ea 11)
Human: Human annotator

Detecting visual text

Another dream **car** to add to the list, this one spotted in Hanbury St.



Shot out my **car** window while stuck in traffic because people in Cincinnati can't drive in the rain.

"Car" is Visual

Not Visual

Discriminative, using largely lexical features

	Category	Position	AUC
	Words	Phrase	74.7
+	Image	-	74.4
+	Bootstrap	Phrase	74.3
+	Spell	Phrase	75.3
+	Length	-	74.7
+	Words	Before	*76.2*
+	Wordnet	Phrase	*76.1*
+	Spell	After	*76.0*
+	Spell	Before	*76.8*
+	Wordnet	Before	**77.0**
+	Wordnet	After	75.6

Table 6: Results of feature ablation on LARGE data set.

Dodge et al 12

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - many meanings, useful in different contexts
- Important recognition technologies coming
  - attributes
  - phrases
  - geometry
  - sentences
- Crucial open questions
  - dataset bias
  - links to utility

Defenses against Bias

- Appropriate feature representations
  - eg illumination invariance
- Appropriate intermediate representations
  - which could have less biased behavior
  - perhaps attributes? scenes? visual phrases?
- Appropriate representations of knowledge
  - eg geometry --- pedestrian example

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - many meanings, useful in different contexts
- Important recognition technologies coming
  - attributes
  - phrases
  - geometry
  - sentences
- Crucial open questions
  - dataset bias
  - links to utility

What should we say about visual data?

- Most important question in vision
  - What does the output of a recognition system consist of?
- A list of all objects present in scene, and locations
  - obvious nonsense - too big, too sensitive to defn of “object”
- A useful representation of reasonable size
  - dubious answer
    - Useful in what way?
    - How do we make the size reasonable?

Object categories depend on utility

Monkey or Plastic toy or both or irrelevant

Some of this depends on what you're trying to do, in ways we don't understand

Person or child or beer drinker or beer-drinking child or tourist or holidaymaker or obstacle or potential arrest or irrelevant or...

Plausible belief space about recognition

- Categories are highly fluid
  - opportunistic devices to aid generalization
    - affected by current problem, utility
  - instances can belong to many categories
    - simultaneously
  - at different times, the same instance may belong to different categories
  - categories are shaded
    - much "within class variation" is principled
  - Most categories are rare
  - Many might be personal, many are negotiated

Notice that some of these issues have resonant ideas when one thinks about the "meaning" of language

- Understanding (recognition)
  - constant coping with the (somewhat) unfamiliar
  - bias is pervasive, affects representation

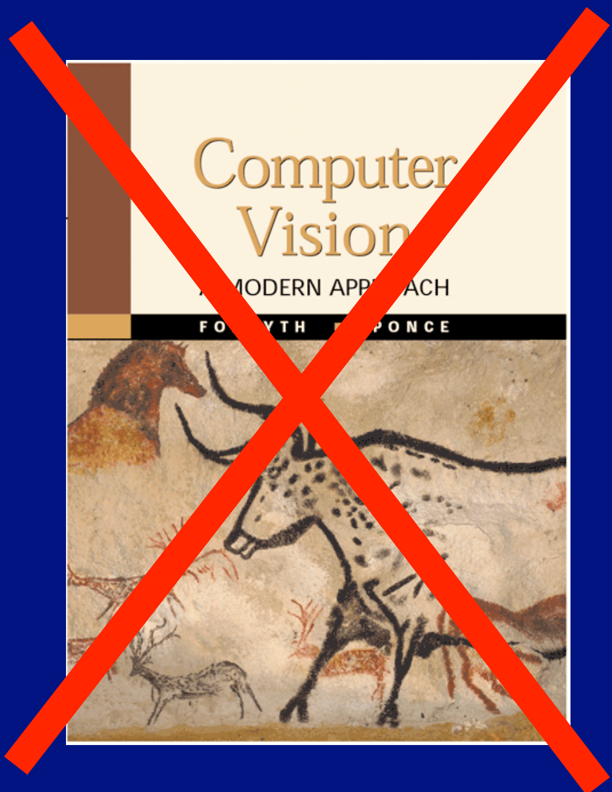
The big question

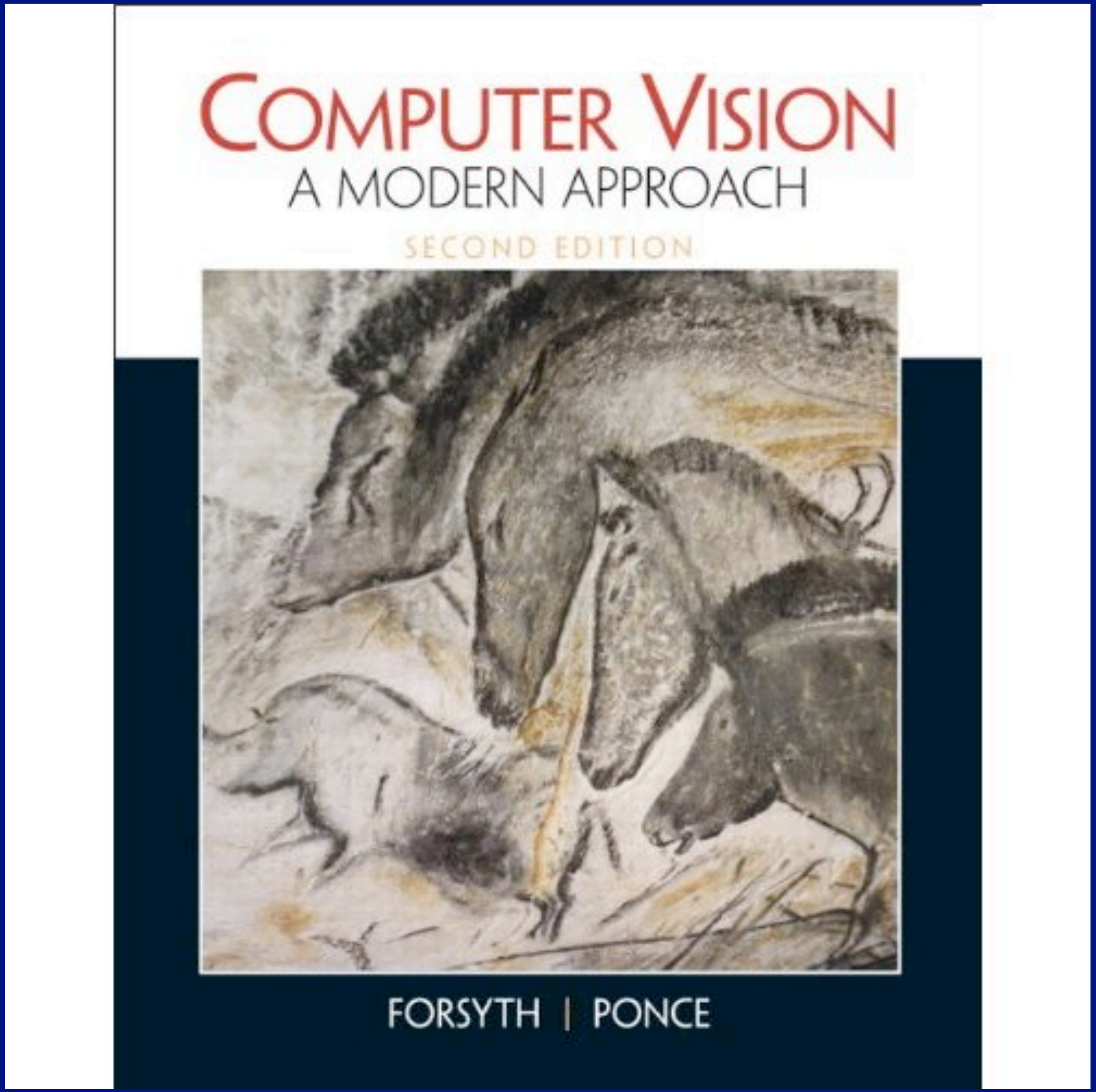
- How to insert object semantics into object recognition?
  - without being silly
  - what is useful knowledge?
  - where does it come from?
  - what is worth saying about objects?
  - what objects are worth saying things about?
  - how should categories be created and destroyed to meet pragmatic needs?

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - serious problems with intellectual underpinnings
- Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - selection
- Crucial open questions
  - dataset bias
  - links to utility

More information





The end

- Thanks to
  - ONR, NSF, Google

What is to be done?

- Cross border raiding by vision, NLP communities is fertile
  - long may it continue
  - even if the details of the analogy are sometimes shaky
- Build a body of knowledge about everyday objects
  - “mundane” knowledge, hard to harvest from the web
- Build a theory of what it means to be “like” something
  - in what respect are things similar? how can we use this idea?
- Build a theory of knowing and reasoning about objects
  - as applied to the concrete world
  - linked to visual observations