# Attributes, Bias and Hashing

D.A. Forsyth, UIUC,
with input from J. Hockenmaier, UIUC, D. Hoiem, UIUC,
A. Farhadi, UWashington,
M. Rastegari, I. Endres, A. Sadeghi, B. Liao, Y. Wang, V. Hedau, K. Karsch, all of UIUC
G. Wang, NTU

# Conclusion

- Recognition is subtle
  - goal uncertain
  - strong basic methods based on classifiers
- Attributes have been helpful
  - the unfamiliar
  - better representations of the familiar
- Could address serious problems
  - intellectual underpinnings of recognition are shaky
    - bias
    - categorization
- Biggest open problem
  - what does recognition do?

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

# 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

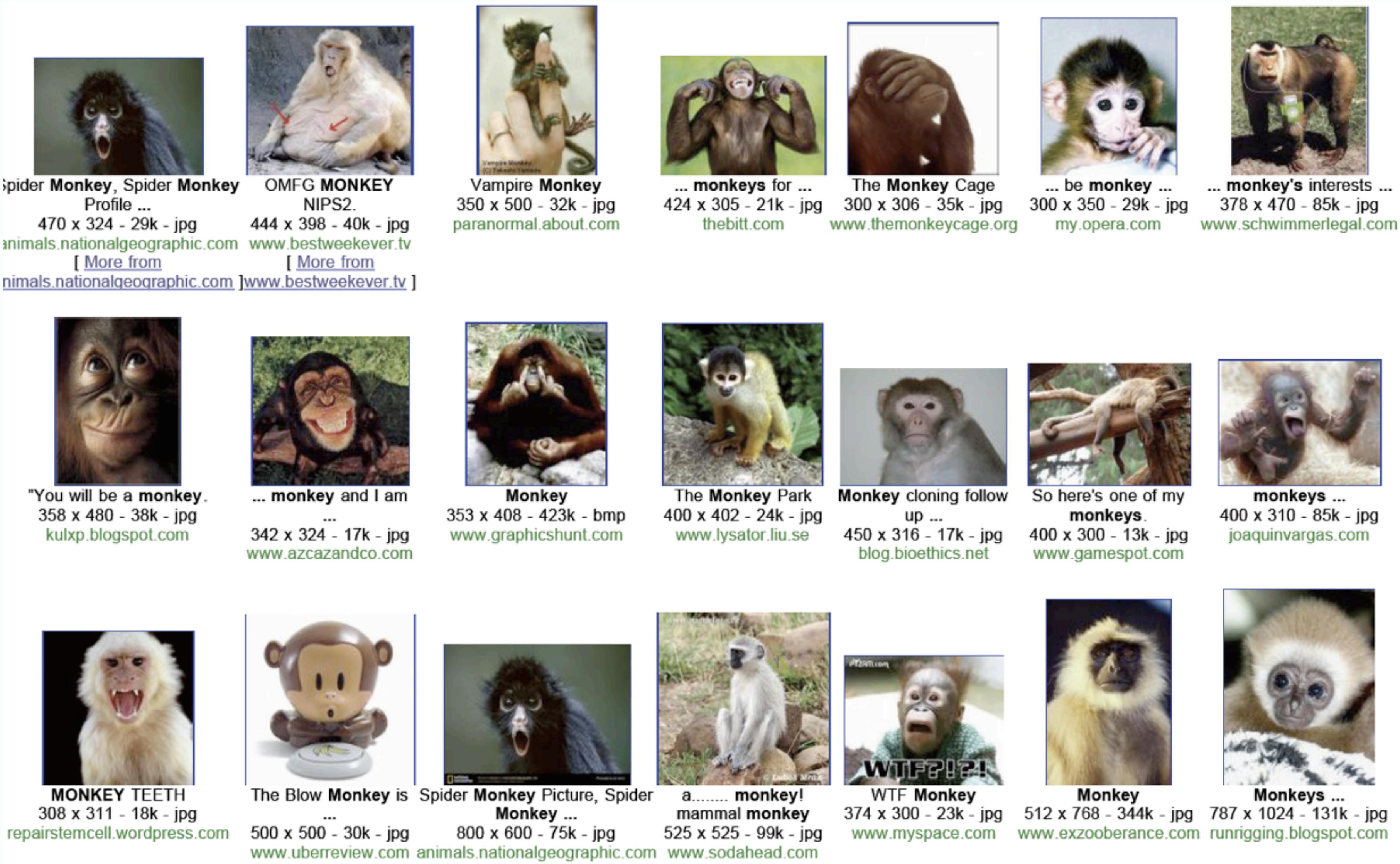
- Subtleties
  - What about the unfamiliar?
  - What kinds of things should we recognize?
  - What environmental knowledge helps?
  - What should we say about pictures?
  - How does utility affect the output?

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

- What should we say about visual data?

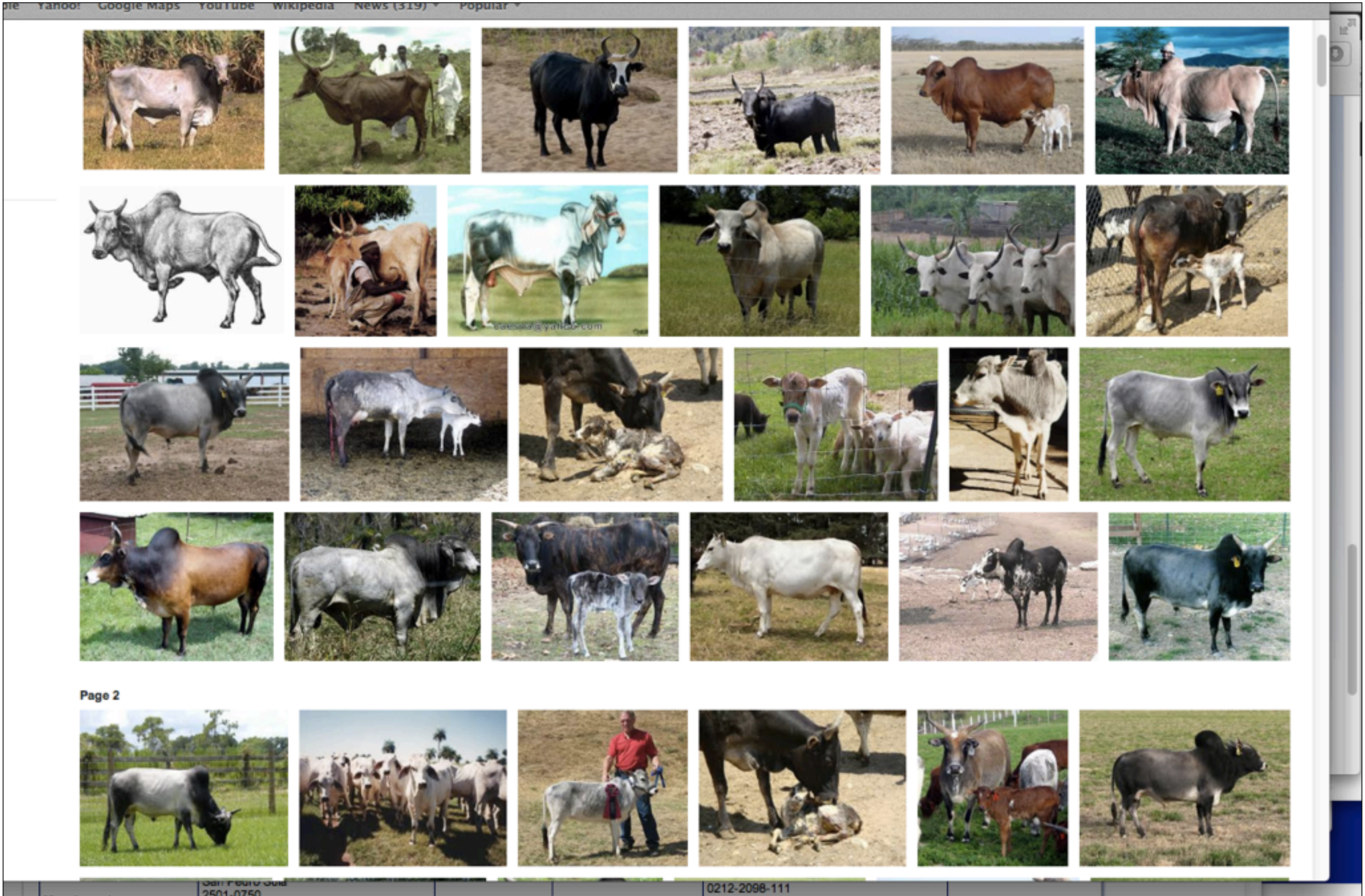
Produce representation

Computer vision

# Conclusion

- Recognition is subtle
    - goal uncertain
    - strong basic methods based on classifiers
- Attributes have been helpful
    - the unfamiliar
    - better representations of the familiar
- Could address serious problems
    - intellectual underpinnings of recognition are shaky
        - bias
        - categorization
- Biggest open problem
    - what does recognition do?

# The unfamiliar



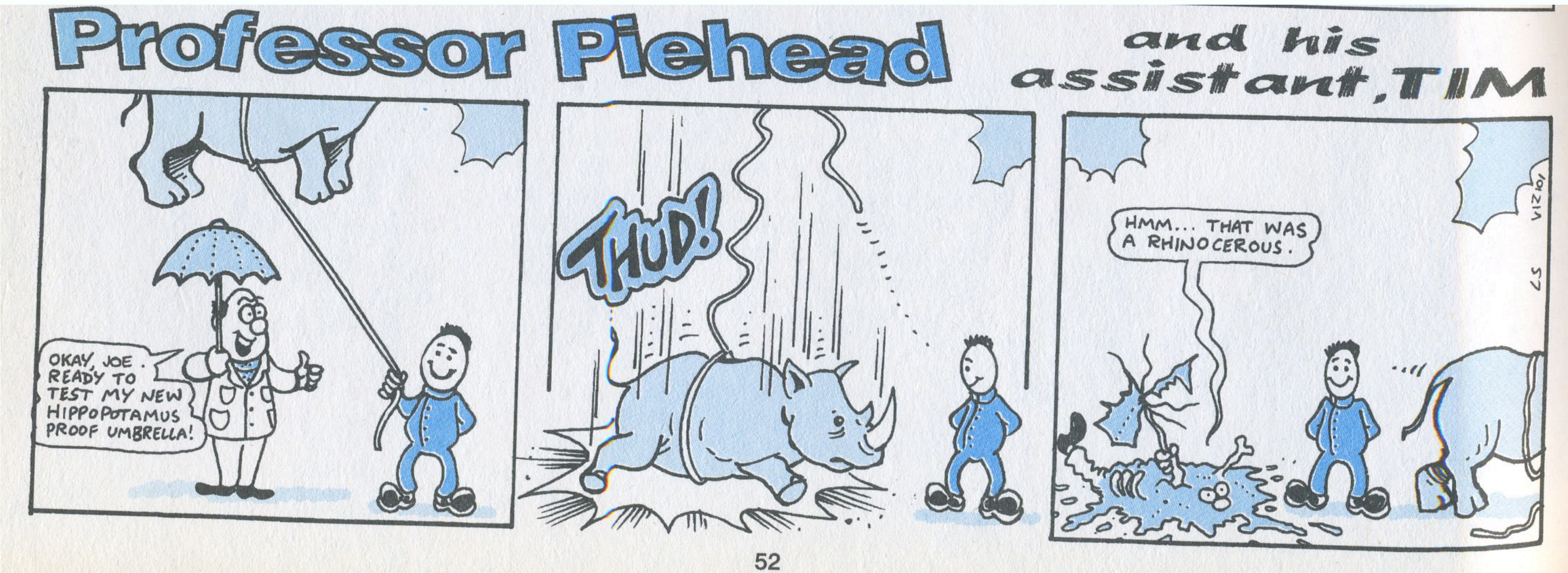

At least two categories for which you probably don't have a name

FIG. 109.—[redacted]

San Pedro Sula 2501-0750 0212-2098-111

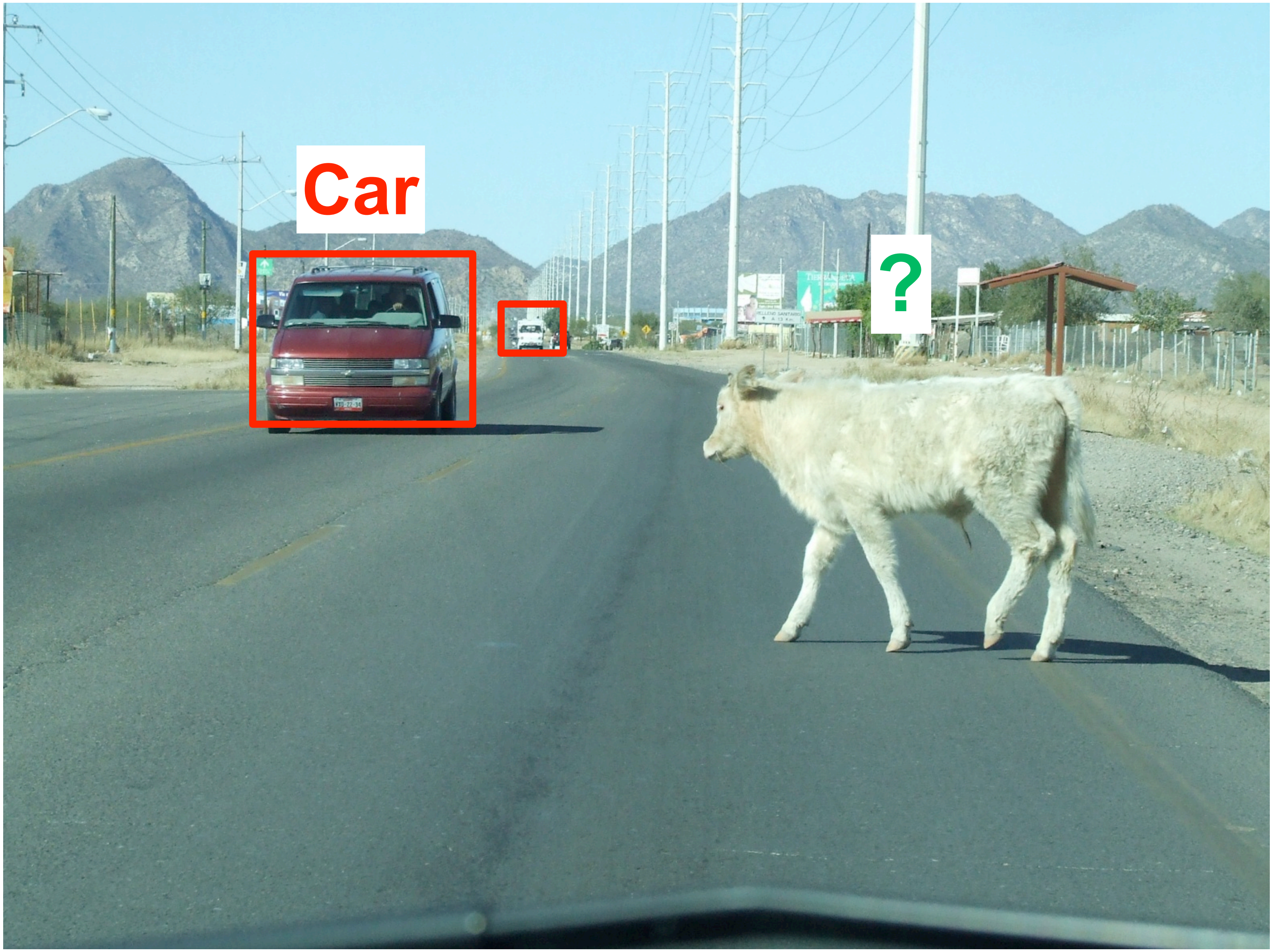
Name in common use among sailors in 19’th century is deeply shocking to modern ears; appears in Aubrey Maturin novels by Patrick O’Brien

# What is an object like?

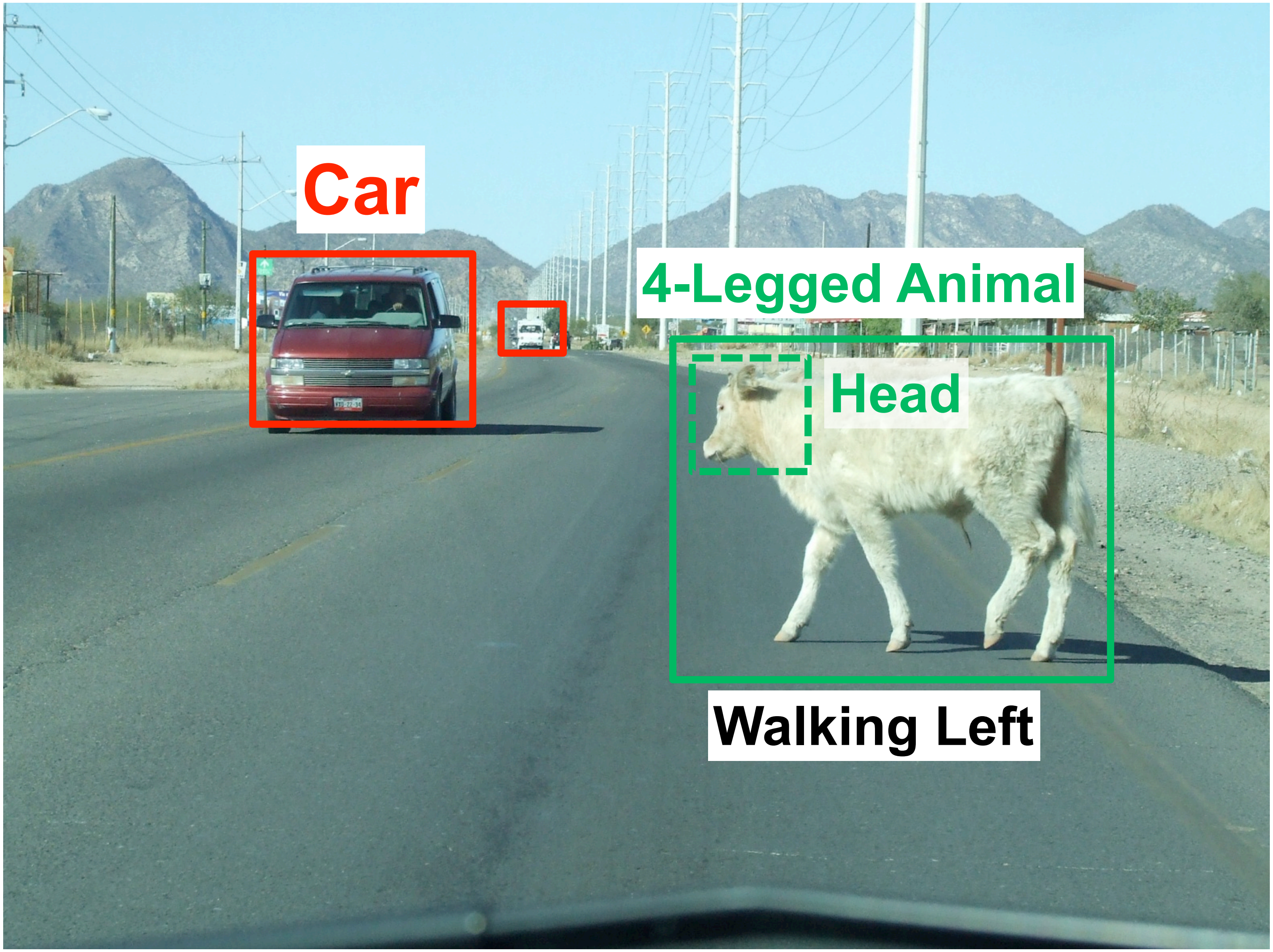



Viz comic, issue 101

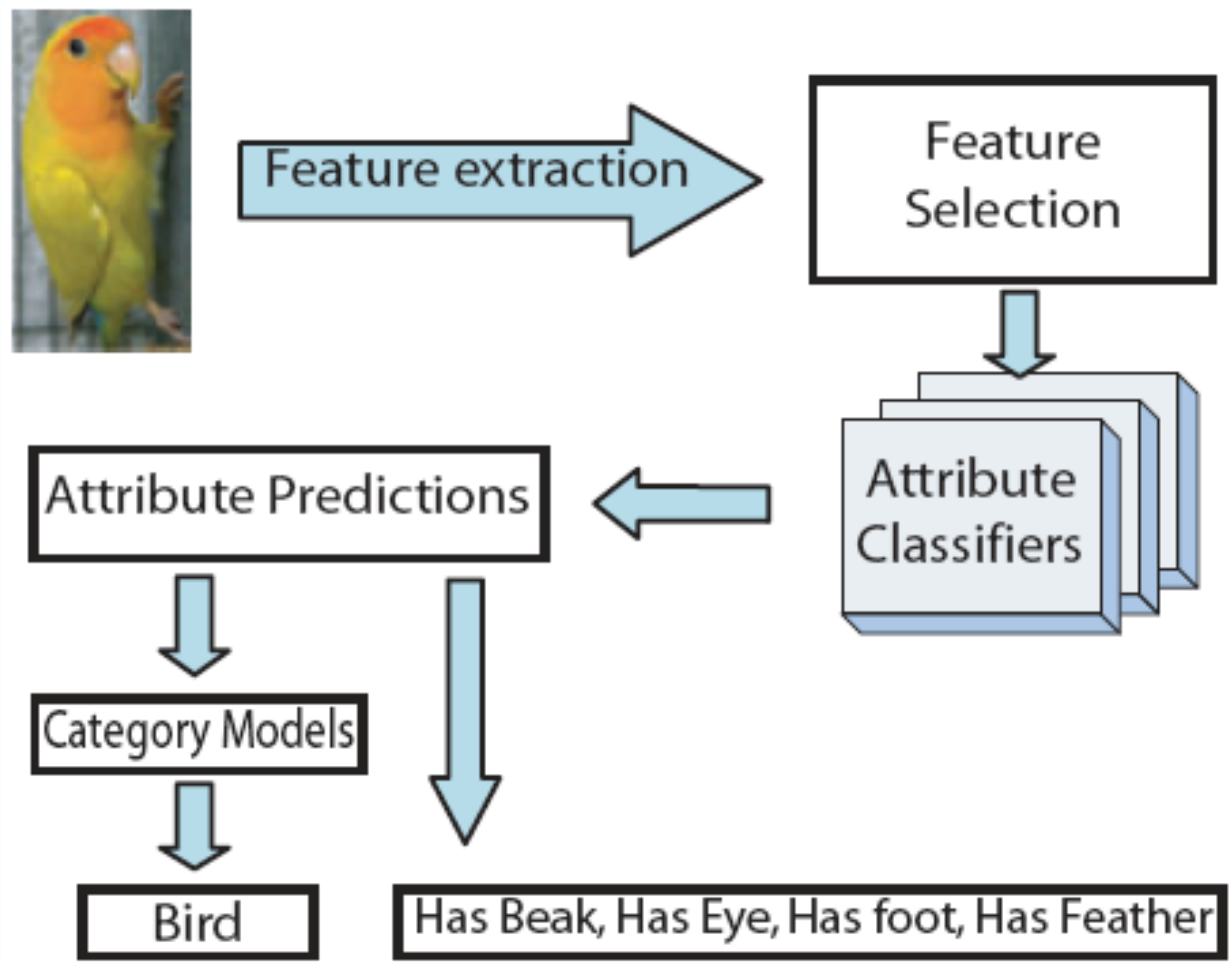
# Vision for driving



# Vision for driving



# General architecture

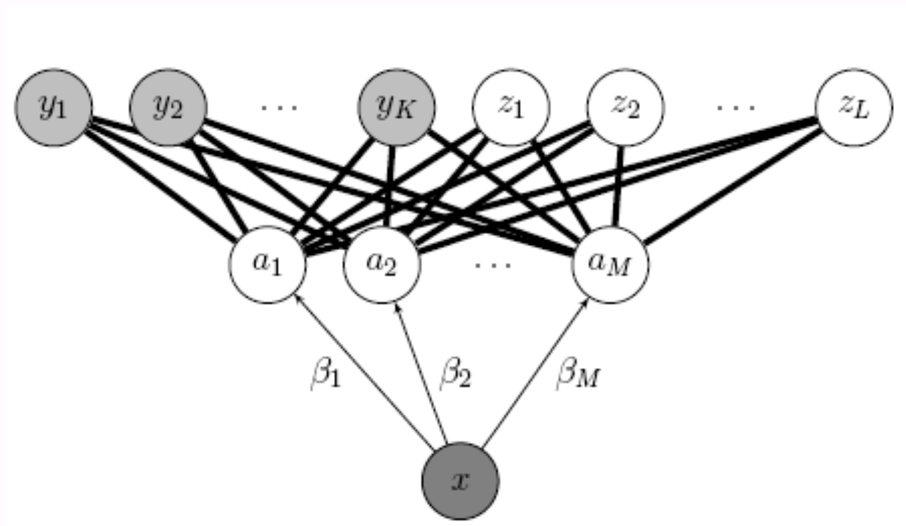



Farhadi et al 09; cf Lampert et al 09

# Direct Attribute Prediction

Known classes

Unknown classes




Attribute layer

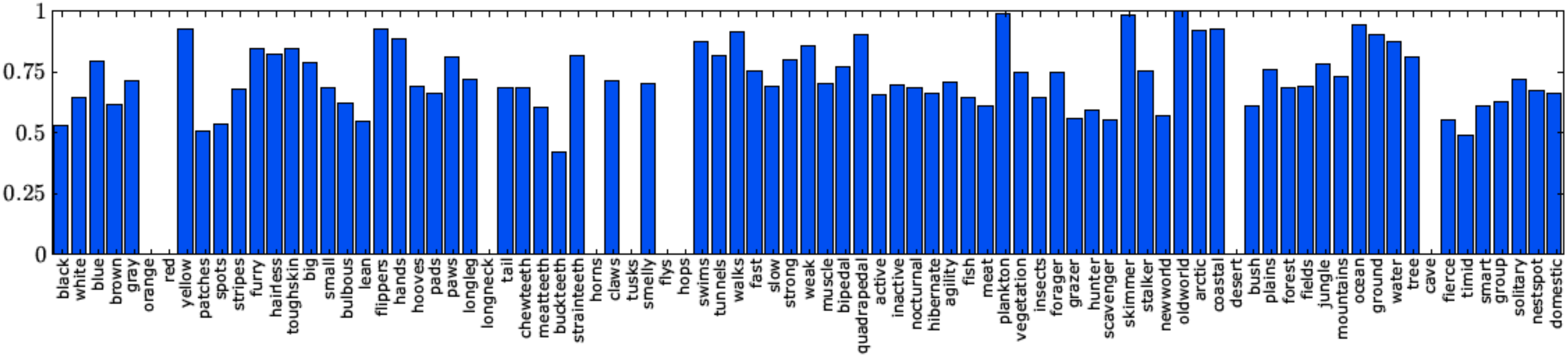
Image features

Lampert ea 09; Farhadi ea 09

Stuff attributes

# Attribute predictions for unknown objects


Farhadi et al 09; cf Lampert et al 09

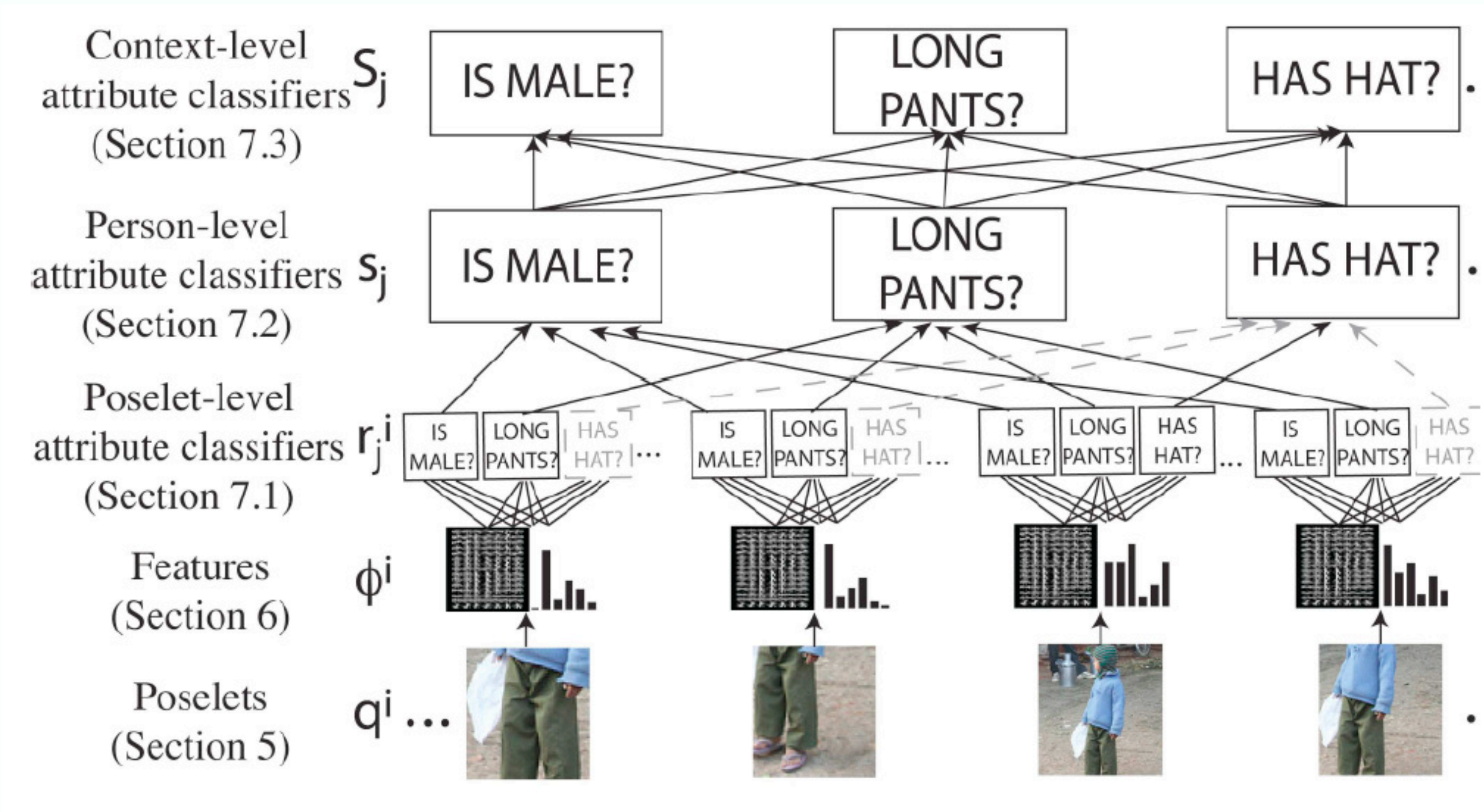



Lampert ea 09

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

# Individual attributes are often wrong, but...

Bourdev et al 11



# Conclusion

- Recognition is subtle
  - goal uncertain
  - strong basic methods based on classifiers
- Attributes have been helpful
  - the unfamiliar
  - better representations of the familiar
- Could address serious problems
  - intellectual underpinnings of recognition are shaky
    - bias
    - categorization
- Biggest open problem
  - what does recognition do?

“Man with a dog on a leash.”



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

# Missing attributes

Aeroplane No “wing” | Car No “window” | Boat No “sail” | Aeroplane No “jet engine” | Motorbike No “side mirror” | Car No “door”

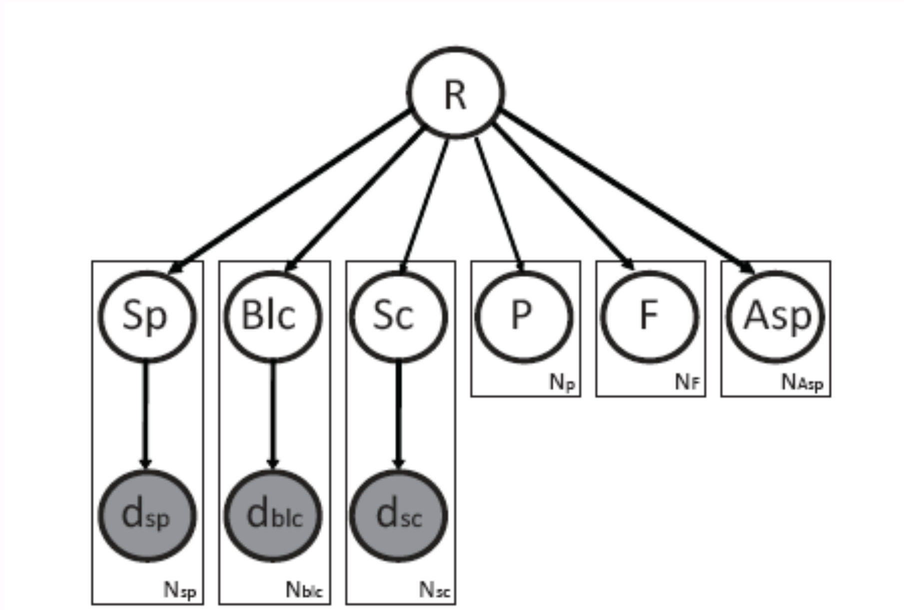
Bicycle No “wheel” | Sheep No “wool” | Train No “window” | Sofa No “wood” | Bird No “tail” | Bird No “leg” | Bus No “door”

# Extra attributes

# Conclusion

- Recognition is subtle
  - goal uncertain
  - strong basic methods based on classifiers
- Attributes have been helpful
  - the unfamiliar
  - better representations of the familiar
- Could address serious problems
  - intellectual underpinnings of recognition are shaky
    - bias
    - categorization
- Biggest open problem
  - what does recognition do?

# Latent Root




Root

Visual attributes

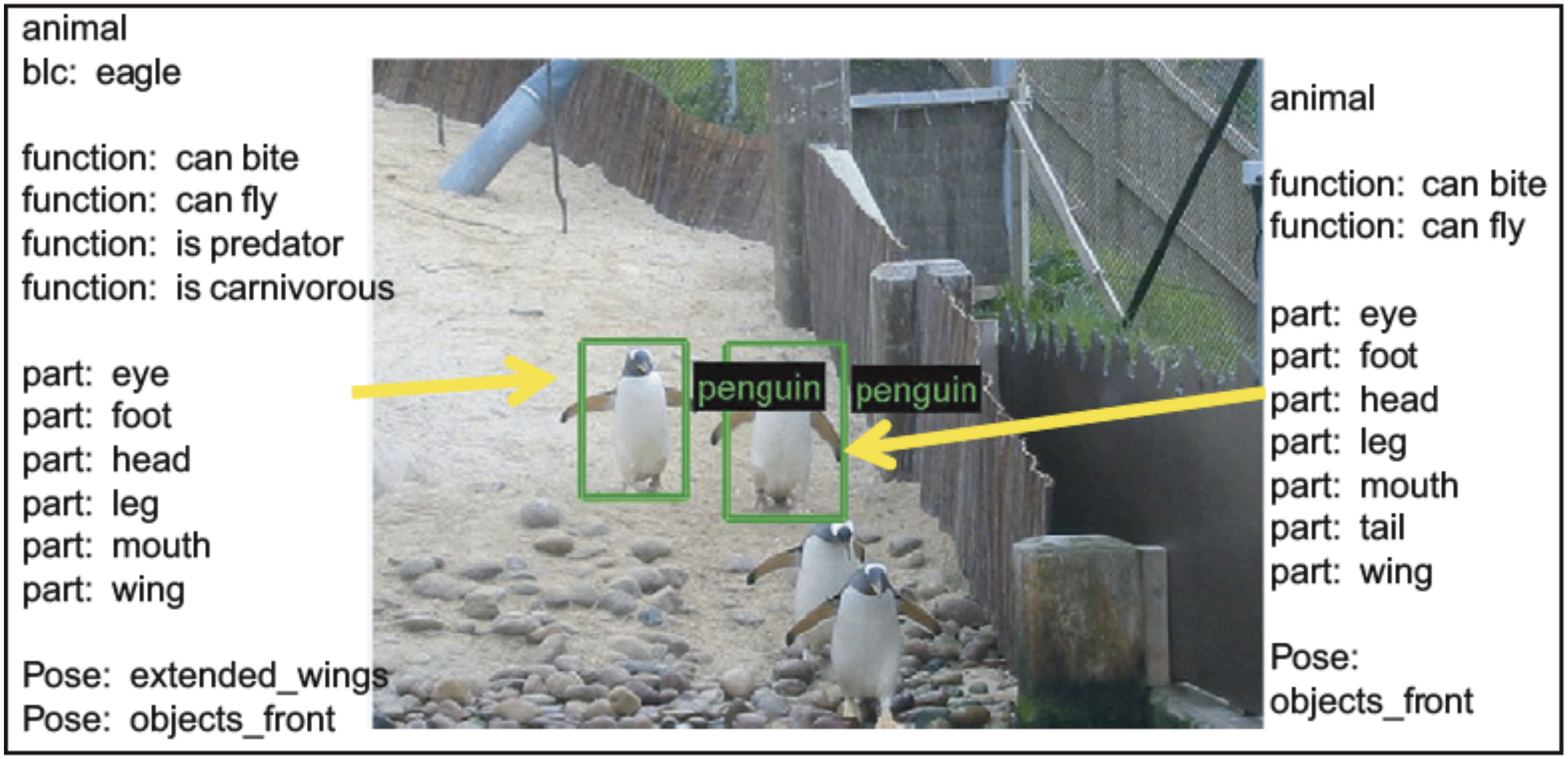
Other attributes

Detector Responses

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

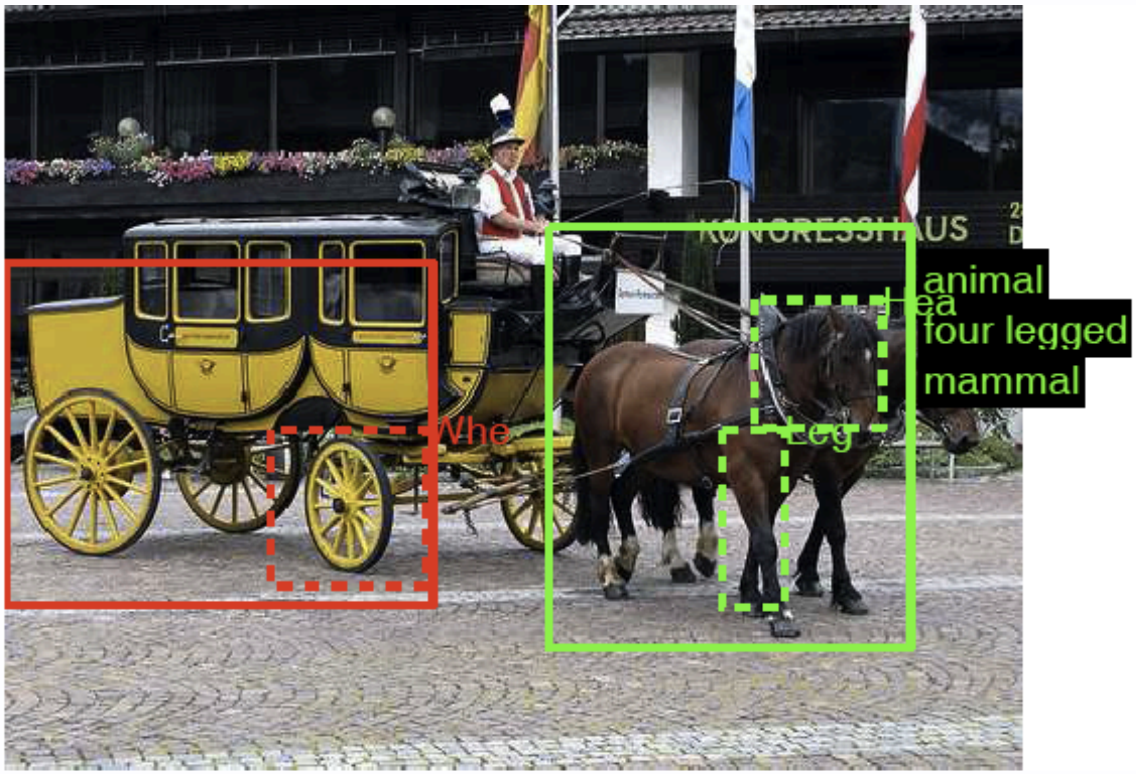
Farhadi ea 10

P: predicate
F: functional attribute
Asp: aspect




Farhadi ea 10

No horses or carriages in training set




Farhadi ea 10

# Discovering attributes


**Fig. 4.** Automatically discovered handbag attributes, sorted by visualness.

Berg et al 12

# Eliciting attribute information

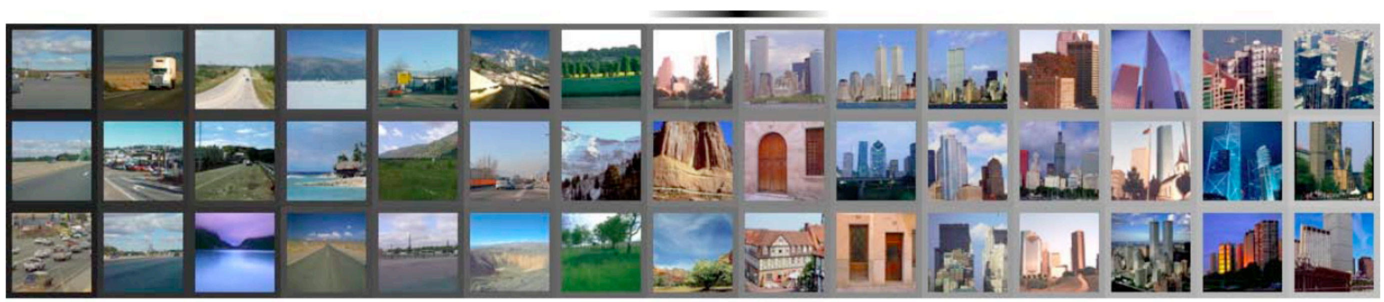

Parkash+Parikh 2012

(a) "This image is not perspective enough to be a street scene."

(b) "Zac Effon is too young to be Hugh Laurie (bottom right)."

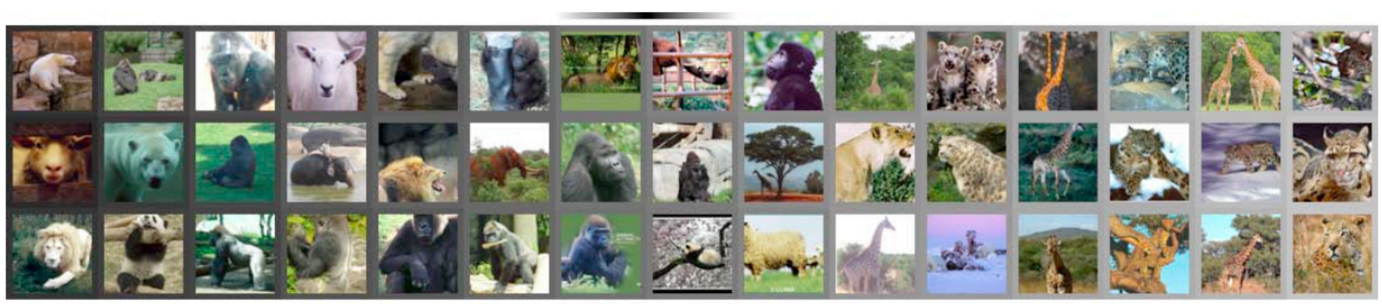
# Eliciting attributes


Figure 1. Interactively building a vocabulary of nameable attributes.



Parikh Grauman 11

(a) OSR gist: nameable, "congested"

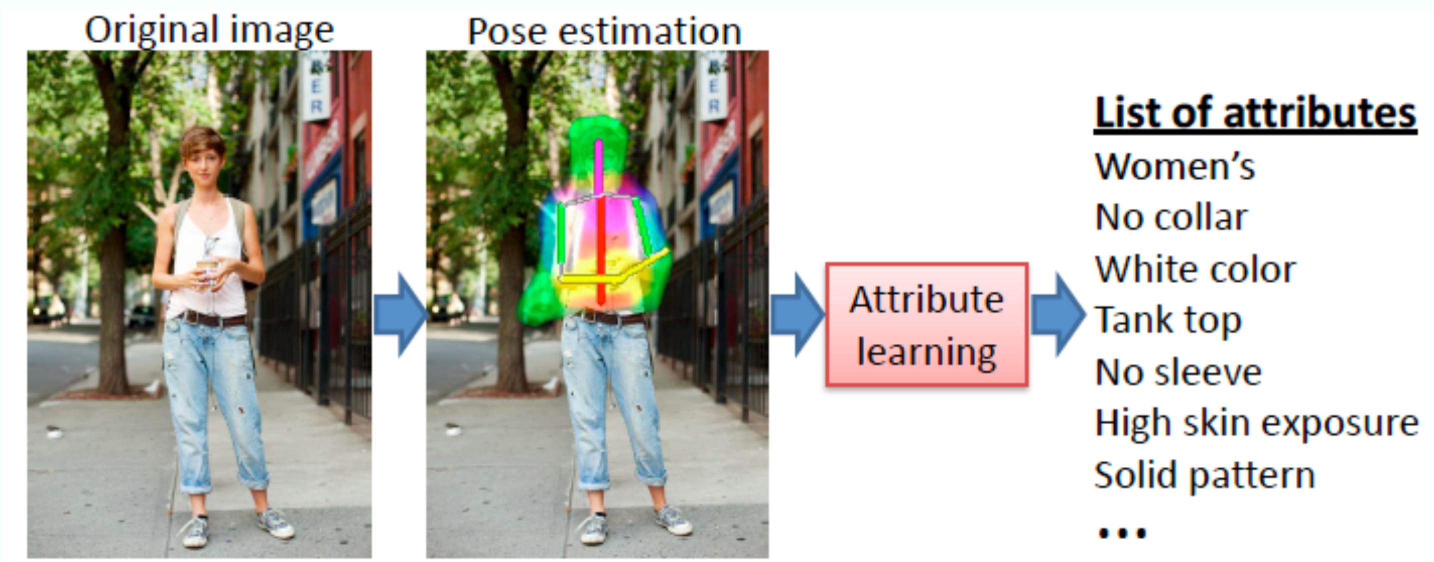


(c) AWA gist: nameable, "spotted"

# Eliciting Attributes

- Listen to Subhransu
- Keep in mind that
  - data labelled “red hat”, “blue car”, “red bicycle” is easier to learn from
  - because “red” and “bicycle” detectors have to agree (Wang, 10)

# Describing clothing




Chen et al 12

# Local attributes

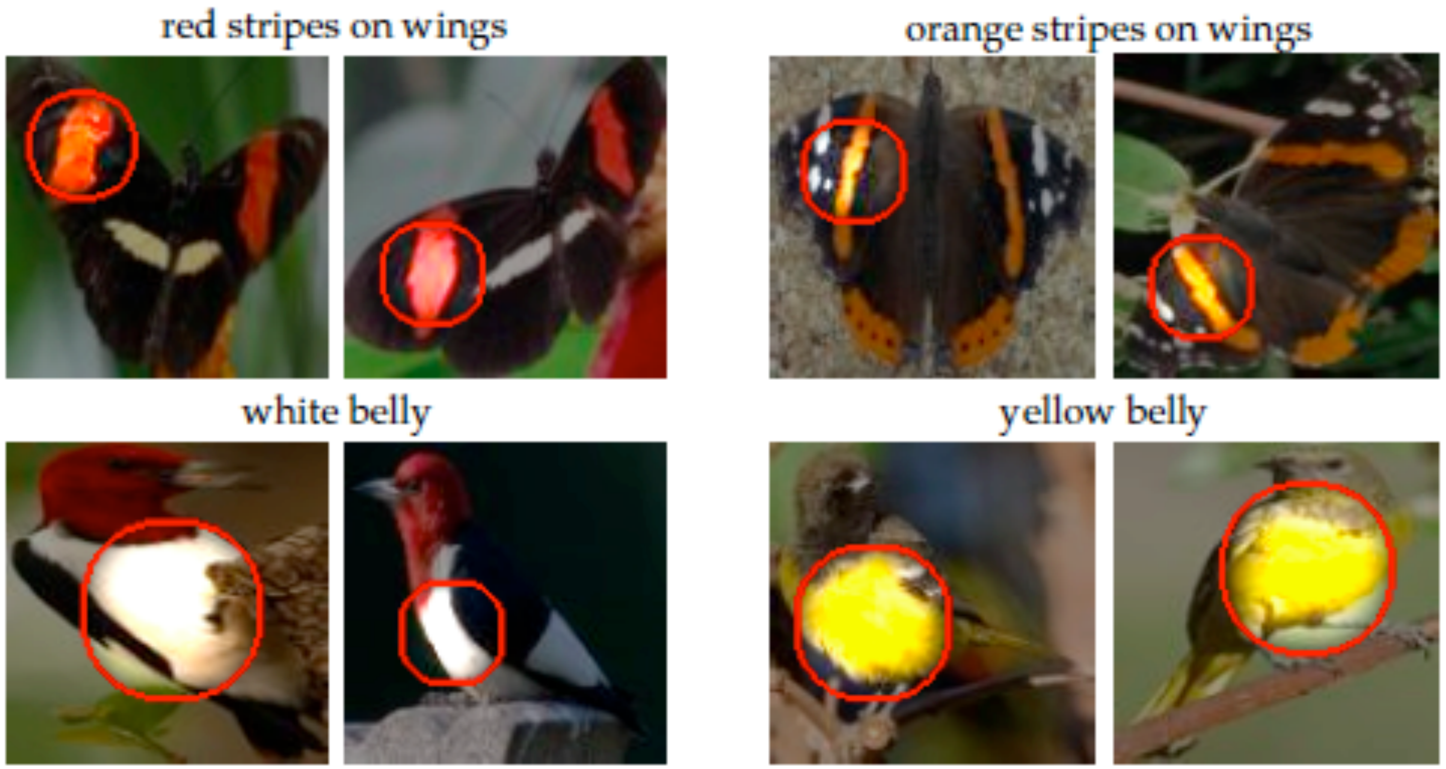



Figure 1. Sample local and semantically meaningful attributes automatically discovered by our approach. The names of the attributes are provided by the user-in-the-loop.

Duan et al 2012

# Describing People with Attributes

"A man with short hair and long sleeves" "A person with long pants" "A woman with long hair, glasses and long pants"

Figure 14. Given a picture of a person our method can generate a natural language description.

Bourdev et al 11

# Conclusion

- Recognition is subtle
  - goal uncertain
  - strong basic methods based on classifiers
- Attributes have been helpful
  - the unfamiliar
  - better representations of the familiar
- Could address serious problems
  - intellectual underpinnings of recognition are shaky
    - bias
    - categorization
- Biggest open problem
  - what does recognition do?

# 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

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

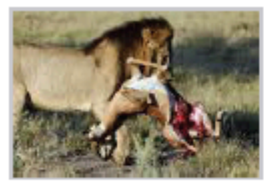

I was doing research on
400 × 300 - 27k - jpg
lowkayhwa.com
Find similar images


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

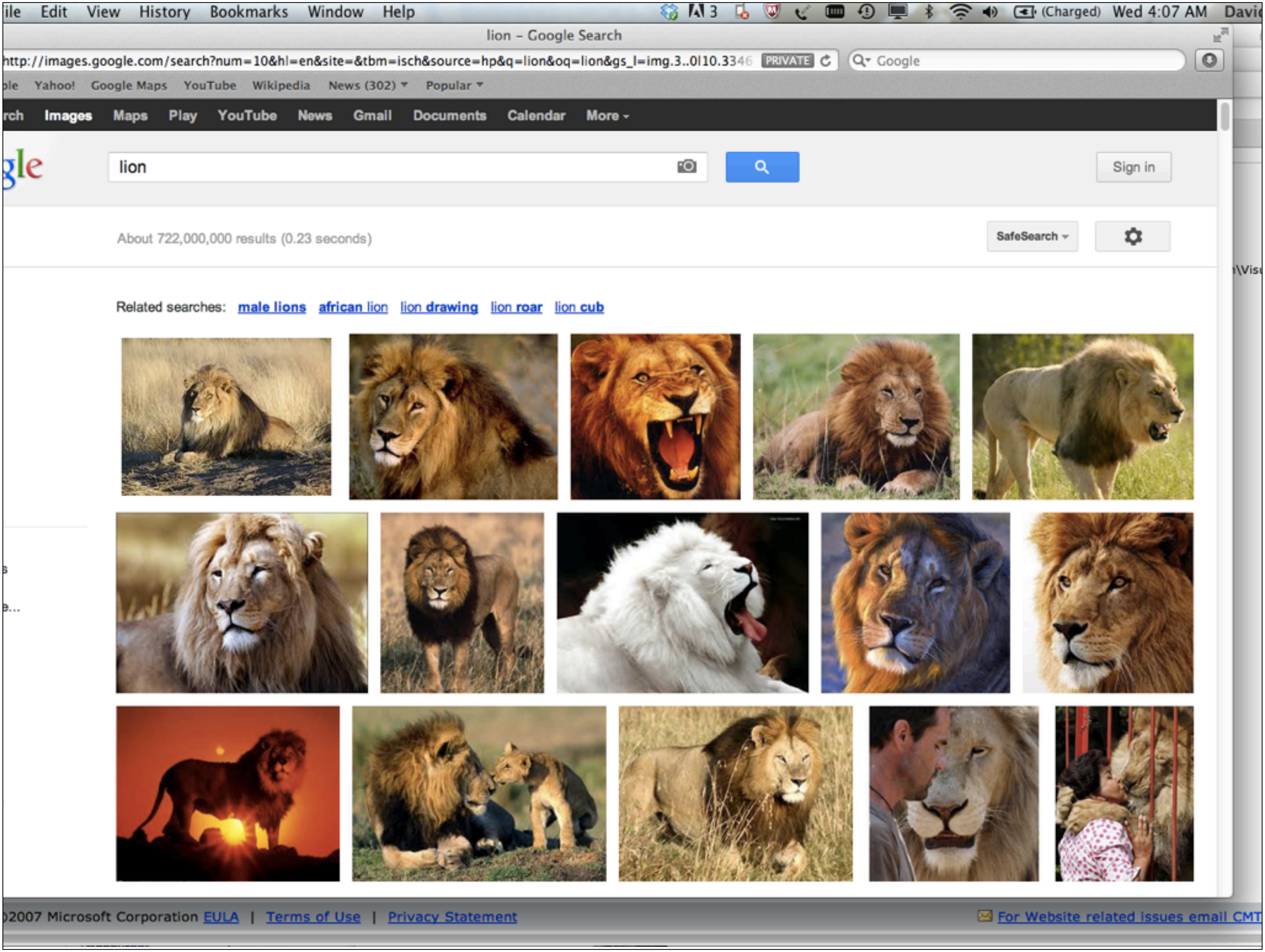

LION!
500 × 385 - 74k - jpg
firemice.wordpress.com
Find similar images


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

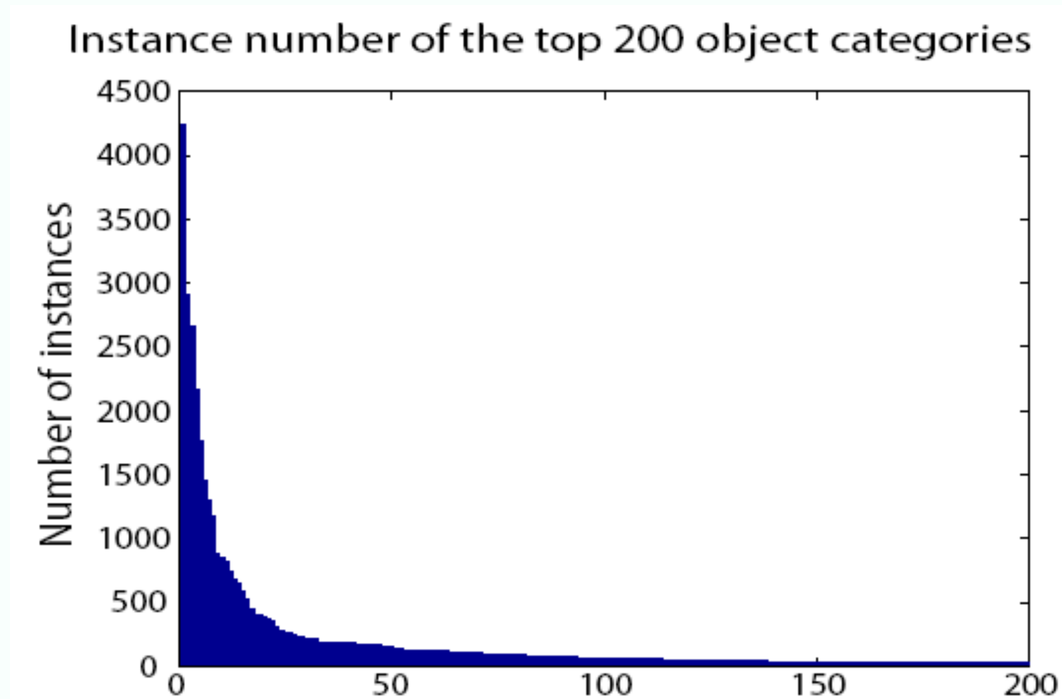
# Induction

- 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.
- Learning = Poison Kool-Aid
  - learning is sweet
  - but the label has some very scary fine print
  - drink carefully, and not too much
- But what if the world can't be like the dataset?

# The world will never be like the dataset

- Because
  - Bias is persistent
  - many things are rare in plausible datasets
    - but not in the world
  - this exaggerates bias
- Strategies
  - Fix datasets  (don’t believe it)
  - Use representations that are well-behaved in the presence of bias

# 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?

Surprising/disturbing absence of results here

- Appropriate representations of knowledge
  - eg geometry --- pedestrian example

Second surprising absence - attributes and aspect/shape

# Conclusion

- Recognition is subtle
  - goal uncertain
  - strong basic methods based on classifiers
- Attributes have been helpful
  - the unfamiliar
  - better representations of the familiar
- Could address serious problems
  - intellectual underpinnings of recognition are shaky
    - bias
    - categorization
- Biggest open problem
  - what does recognition do?

# What should we say about visual data?

- Most important question in vision
  - What does the output of a recognition system consist of?
- 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...

# Nouns <> Categories

- People are very good at managing
  - Sometimes we don’t have words for things
  - Sometimes we don’t have things for words
- Nouns are sometimes too extensive
  - the visual complexity inside the category is too high
- or too prescriptive
  - “buckler” vs “shield”
- or both
  - “man” vs “man on a horse”

# Category sensitive binarization

- Approach:
  - Each image gets a binary code
    - and each can have a distinct code
  - Each bit is predicted by a classifier
  - Choose codes so that
    - they can be predicted accurately
    - within a category, codes tend to cluster
    - across categories, codes tend to be different

Rastegari ea 12

# Category sensitive binarization

- Strategy:
  - allocate codes to images so each in a category has the same code
  - Iterate
    - learn SVMs to predict codes
    - adjust codes so they're
      - closer to predictions
      - cluster within categories
      - are separated across categories

Rastegari ea 12

# Codes = discovered attributes

**Fig. 8.** Discovering attributes: Each bit corresponds to a hyperplane that group the data according to unknown notions of similarity. It is interesting to show what our bits have discovered. On two sides of the black bar we show 8 most confident images for 5 different hyperplanes/bits (Each row). Note that one can easily provide names for these attributes. For example, the bottom row corresponds to all round objects versus objects with straight vertical lines. The top row has silver, metalic and boxy objects on one side and natural images on the other side, the second row has water animals versus objects with checkerboard patterns. Discovered attributes are in the form of contrast: both sides have its own meaning. These attributes are compact representations of standard attributes that only explain one property. For more examples of discovered attributes please see supplementary material.

Rastegari ea 12

# Conclusion

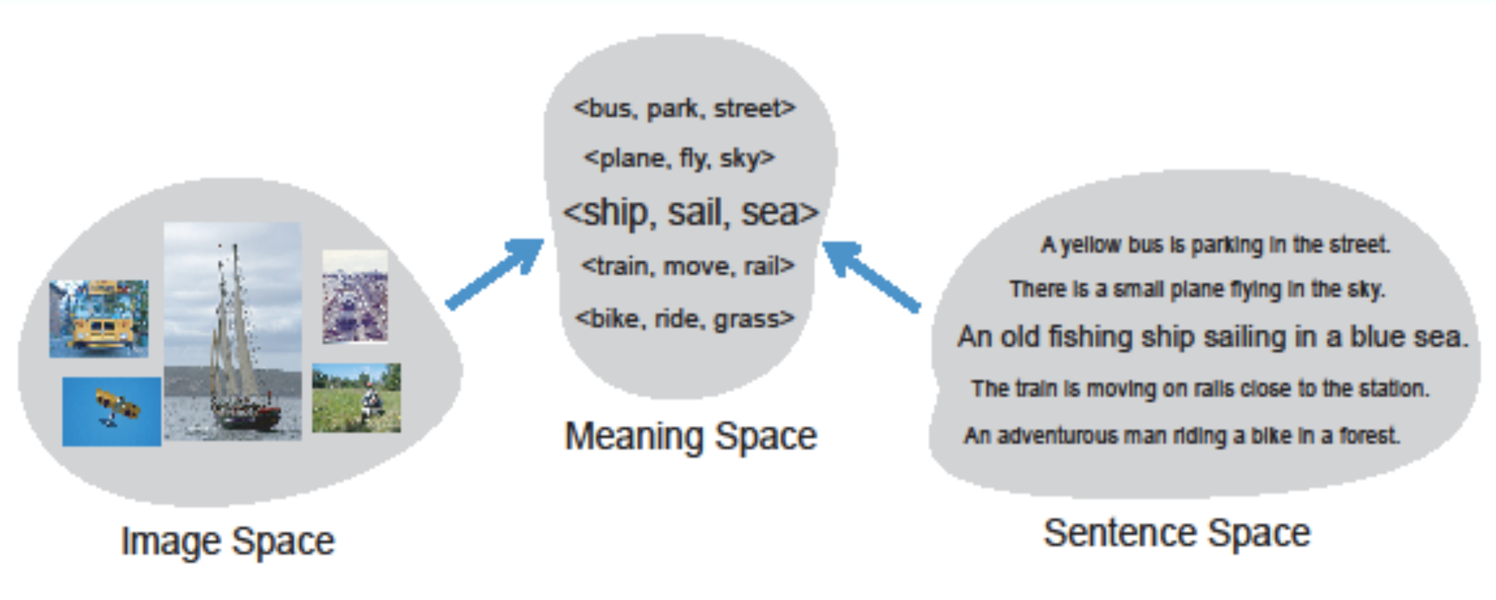
- Recognition is subtle
  - goal uncertain
  - strong basic methods based on classifiers
- Attributes have been helpful
  - the unfamiliar
  - better representations of the familiar
- Could address serious problems
  - intellectual underpinnings of recognition are shaky
    - bias
    - categorization
- Biggest open problem
  - what does recognition do?

# Why I like sentence level descriptions

- Conceptual
  - Many good sentences for an image
    - this is probably what recognition is like
  - You need to select to write sentences
    - you can’t just write down everything
  - Selection could depend on utility
    - what you say about an image depends on what you want to achieve
  - Sentences can be arranged wrt generality
    - “a picture” > “a person”> “a girl in a blue dress playing an ocarina”
- Practical
  - Sentences are compact representations of meaning
    - it’s really hard to get details right
  - People can spot errors quickly
    - and also generate sentences from a quick look

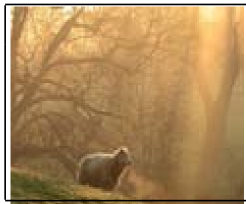
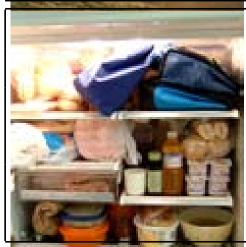
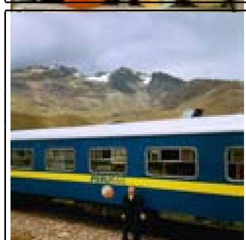
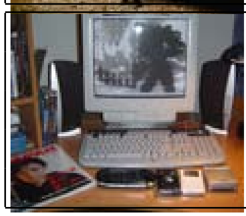
# Rich(ish) sentences from simple intermediates

Object, action, scene




Farhadi ea 10

# Examples

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

Farhadi ea 10

# 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

# Another 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

# Conclusion

- Recognition is subtle
  - goal uncertain
  - strong basic methods based on classifiers
- Attributes have been helpful
  - the unfamiliar
  - better representations of the familiar
- Could address serious problems
  - intellectual underpinnings of recognition are shaky
    - bias
    - categorization
- Biggest open problem
  - what does recognition do?

# 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

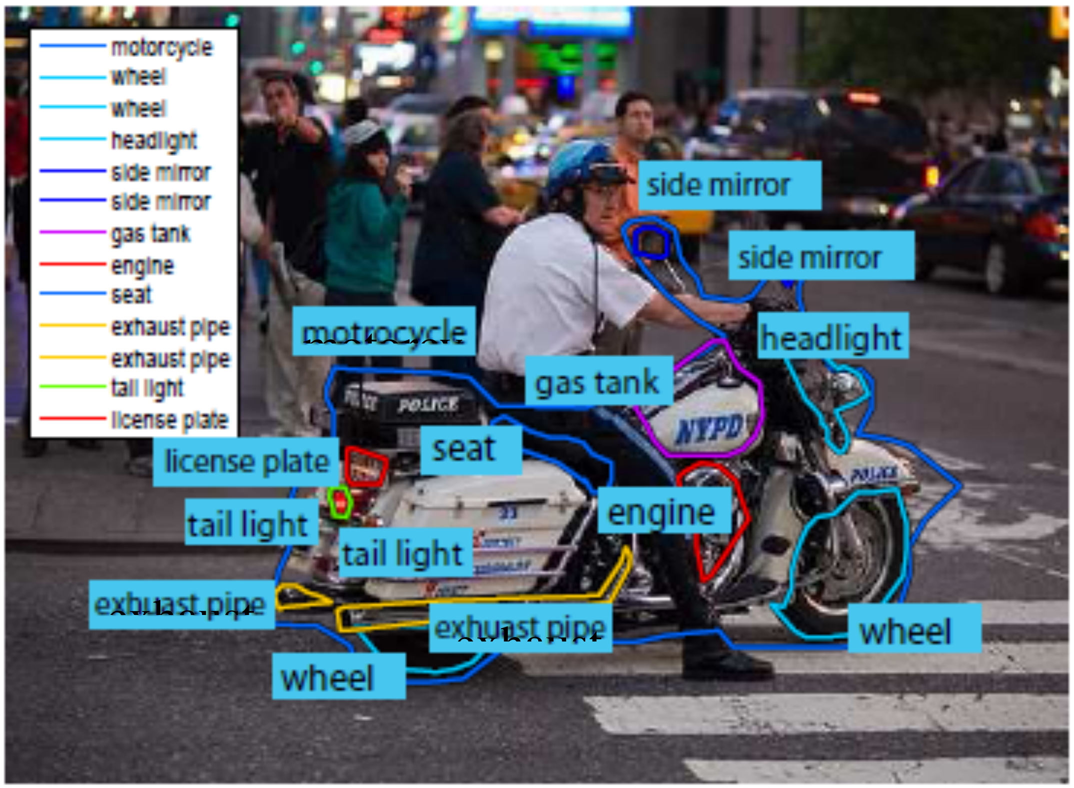
# Datasets - II

- 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 ImageI
3192 Objects – 28 Categor
26695 Parts – 71 types
30046 Attributes – 34 typ
1052 Material Images – 10 t

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/

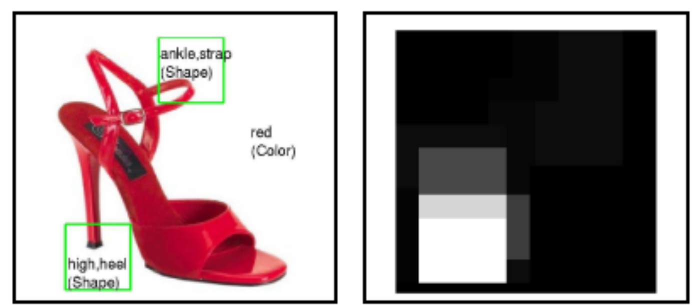
Rashtchian et al. HLT NAACL 2010

# Attribute Discovery Dataset

- 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](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.

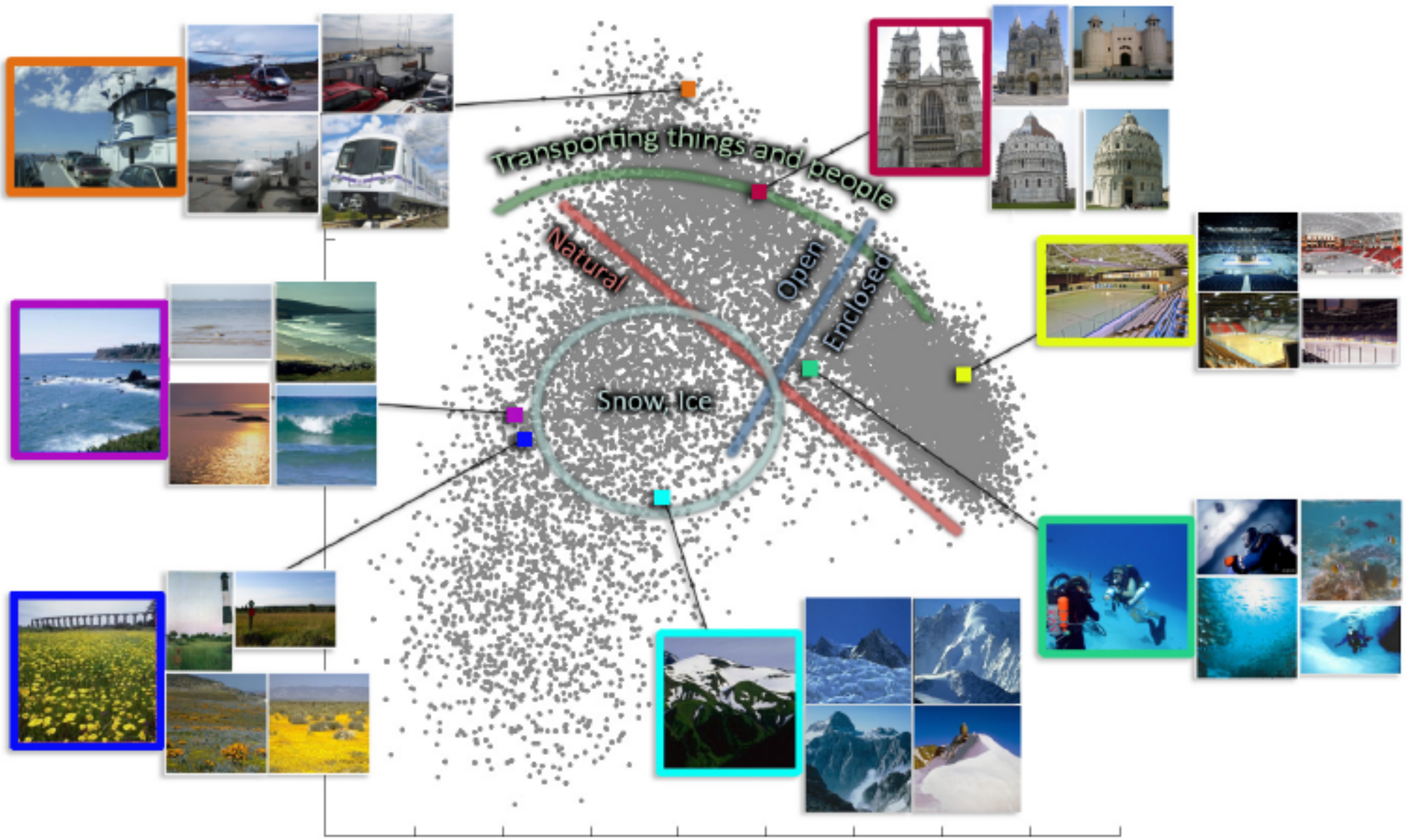
Emma in her hat looking super cute.

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

Ordonez et al. NIPS 2011

# Other Attribute Datasets

## SUN Attributes Dataset

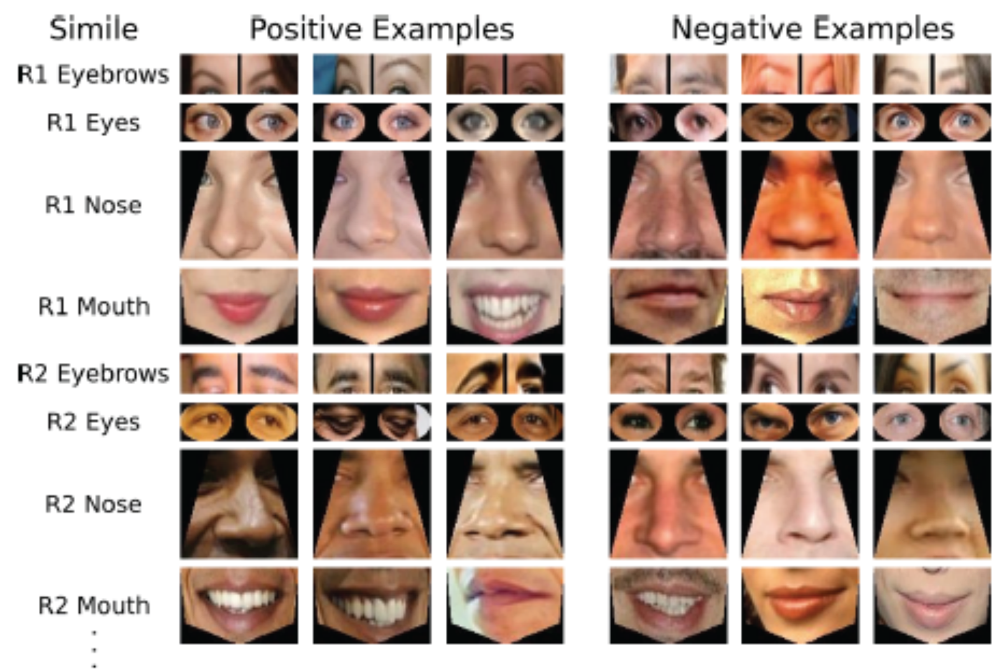



http://cs.brown.edu/~gen/sunattributes.html

Patterson Hays CVPR 2012

# Other Attribute Datasets

## PubFig


http://www.cs.columbia.edu/CAVE/databases/pubfig/

Kumar et al. ICCV 2009