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, I. Endres, A. Sadeghi, B. Liao, Y. Wang, V. Hedau,
K. Karsch, all of UIUC
G. Wang, NTU,
Nicolas Loeff, RGM Advisors

Why is visual object recognition useful?

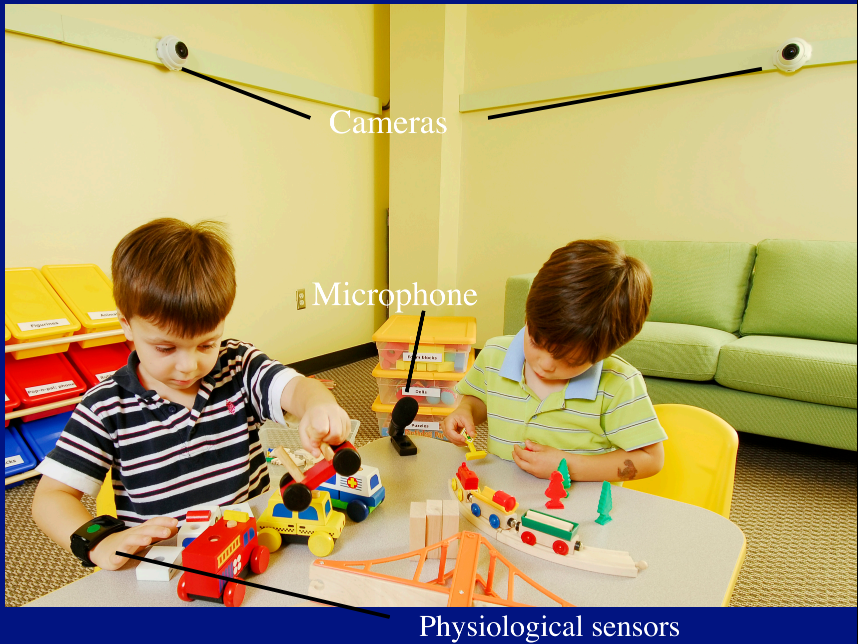
- If you want to act, you must draw distinctions
- For robotics
  - recognition can predict the future
    - is the ground soggy?
    - is that person doing something dangerous?
    - does it matter if I run that over?
    - which end is dangerous?
- For information systems
  - recognition can unlock 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

Example: Humans

- Surveillance
  - prosecution; intelligence gathering; crime prevention
  - HCI; architecture;
- Synthesis
  - games; movies;
- Safety applications
  - pedestrian detection
- People are interesting
  - movies; news

Computational Behavioural Science

- Observe people
  - Using vision, physiological markers
- – Interacting, behaving naturally
  - In the wild
- drive feedback for therapy
  - Eg reward speech
- Applications
  - Model: screen for ASD
  - Other:
    - Anywhere large scale observations help
      - Support in home care
      - Support care for demented patients
      - Support stroke recovery
      - Support design of efficient buildings
- 10M$, 5yr NSF award under Expeditions program
  - GaTech, UIUC(DAF, Karahalios), MIT, CMU, Pittsburgh, USC, Boston U



Words near pictures are informative

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.

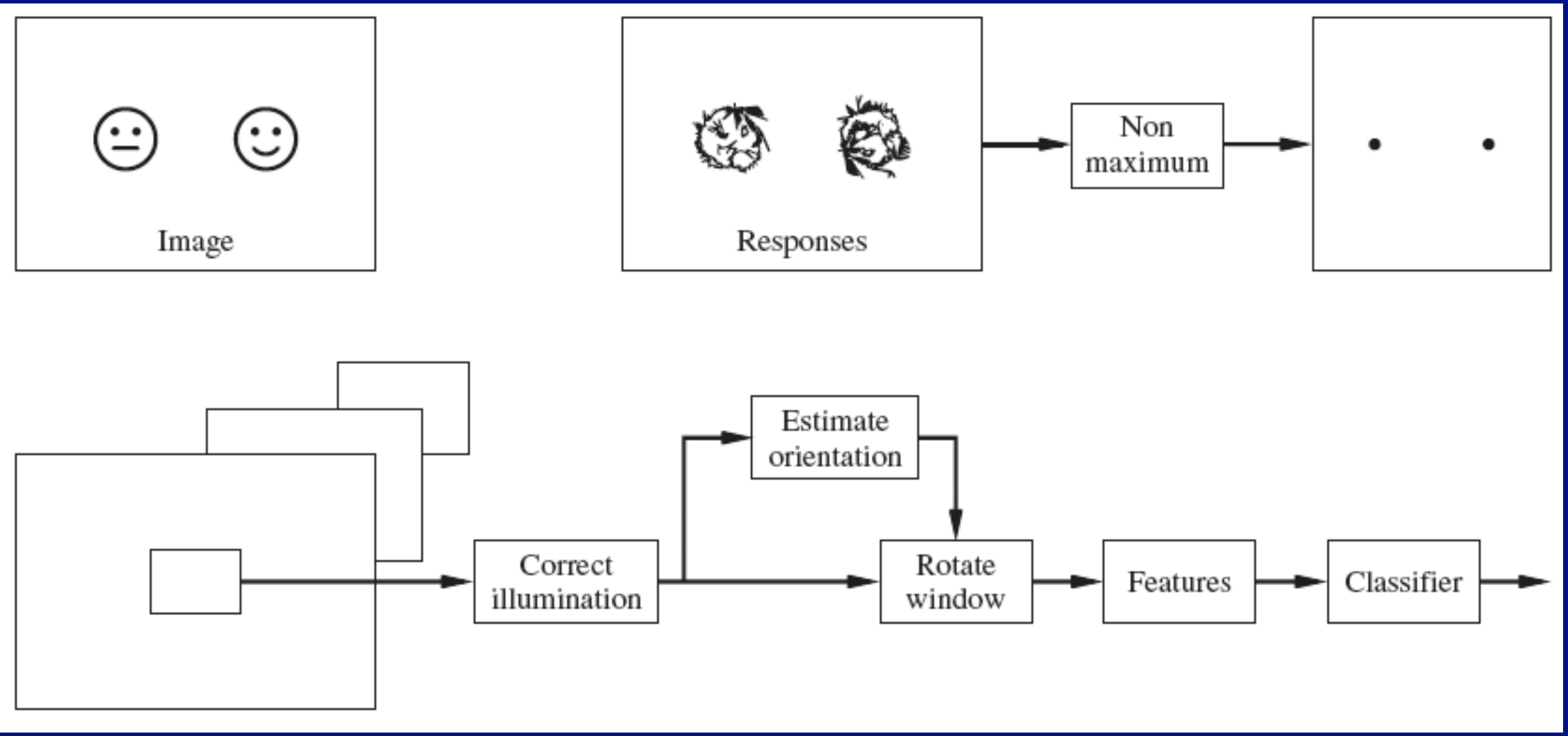
bananarepublic.com

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
  - many meanings, useful in different contexts
- Important recognition technologies coming
  - attributes
  - phrases
  - geometry
  - sentences
- Crucial open questions
  - dataset bias
  - links to utility

Detection with a classifier



Obtain dataset

Build features

Mess around with classifiers, probability, etc

Produce representation

Computer vision

Big questions

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?

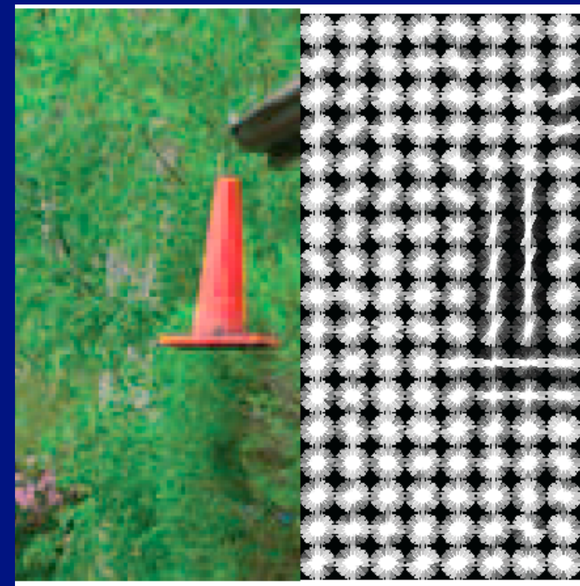
Mess around
with classifiers,
probability, etc

Computer vision

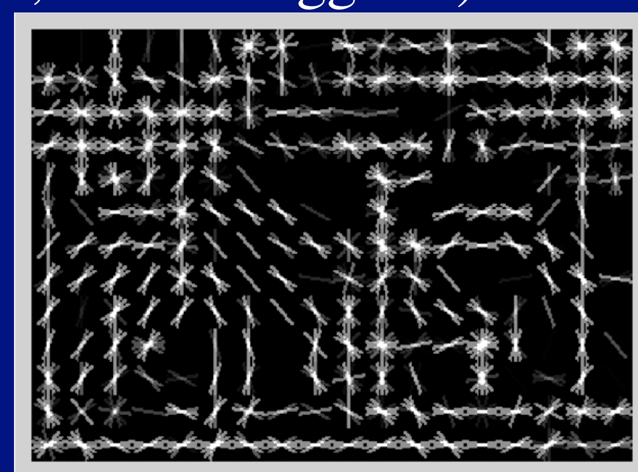
- What should we say about visual data?

Produce representation

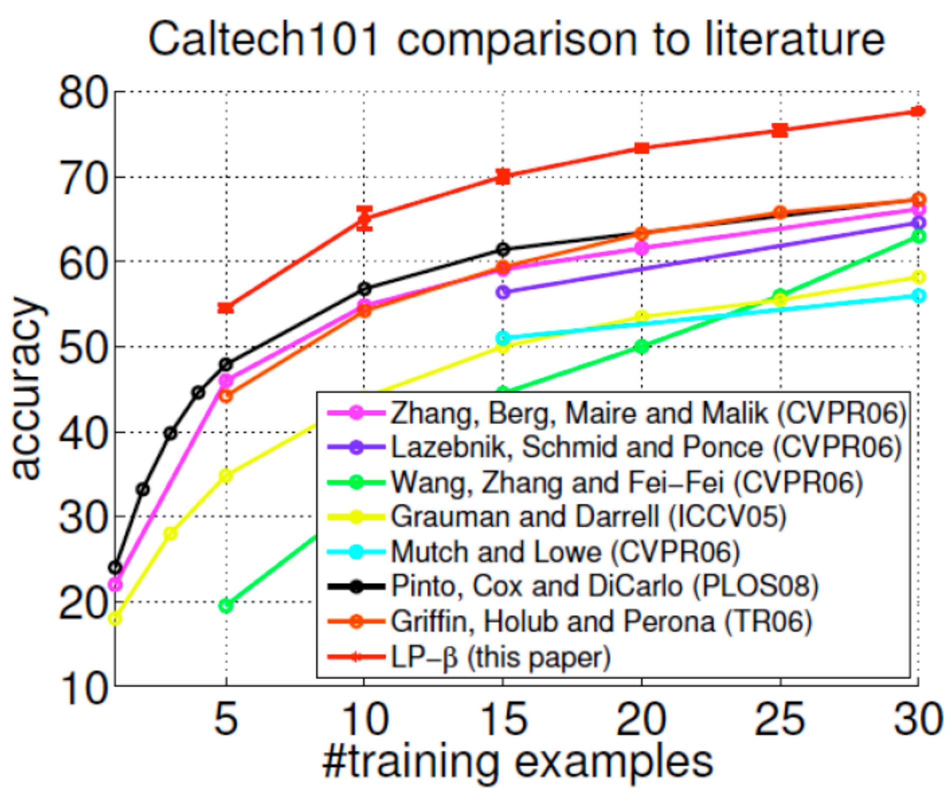
Features

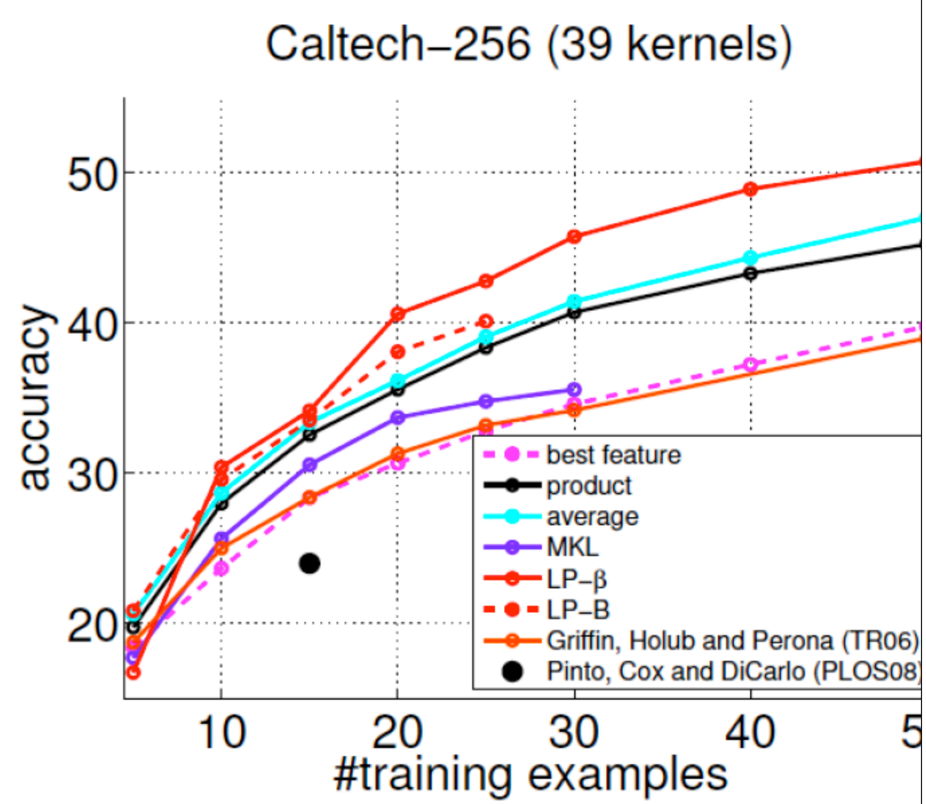


- Principles
  - illumination invariant (robust) -> gradient orientation features
  - windows always slightly misaligned -> local histograms
- HOG, SIFT features (Lowe, 04; Dalal+Triggs 05)



Classification works well





Movies and captions: Laptev et al 08

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

Platonism?

A belief space about recognition

- Categories are fixed and known
  - Each instance belongs to one category of k

Obvious nonsense
Obvious nonsense

- Object recognition=k-way classification
- research agenda:
  - more features, better classifiers:
  - perhaps category hierarchies for statistical leverage (tying)

Are these monkeys?

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?

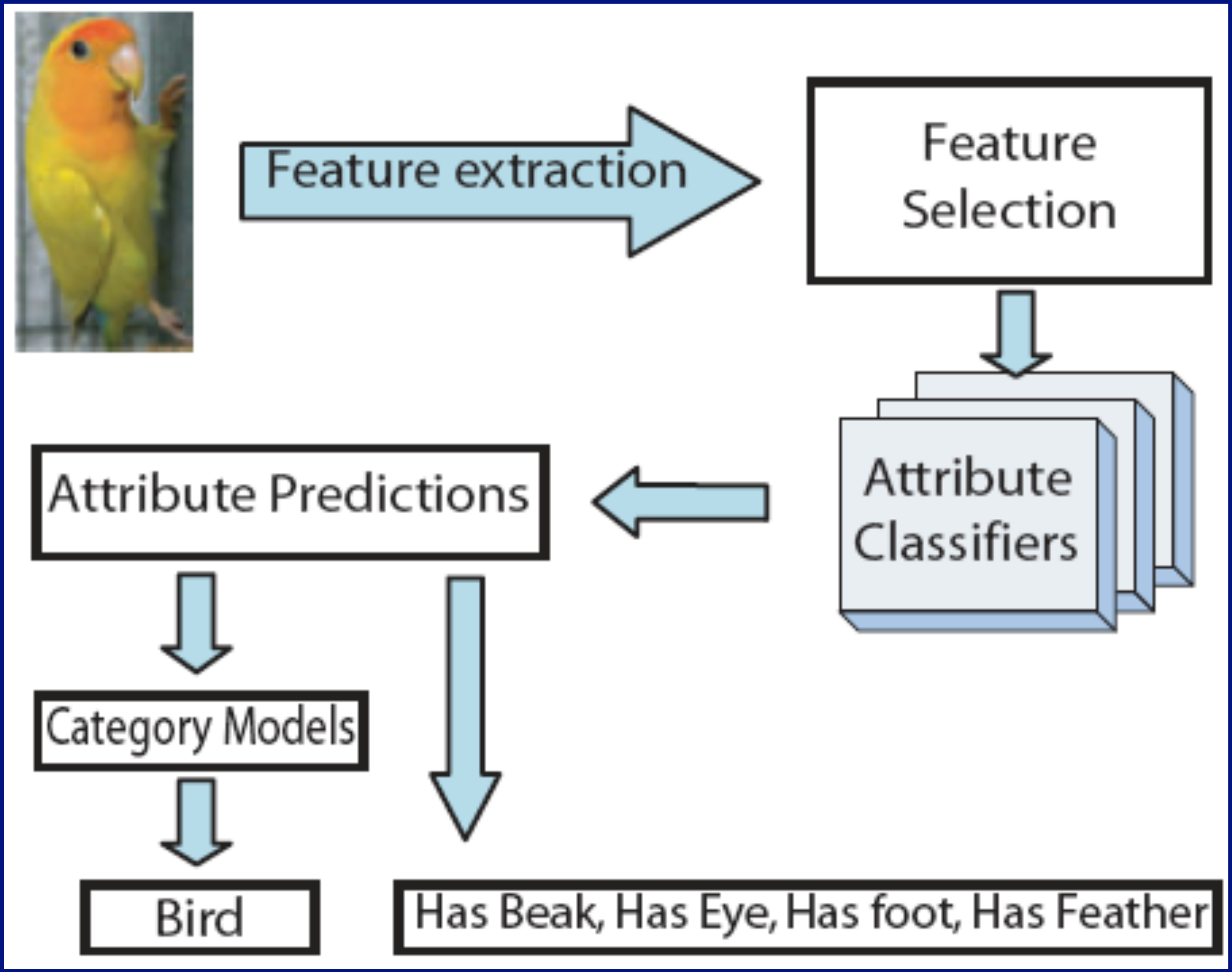
Conclusion

- 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

Subtleties: What about the unfamiliar?

Subtleties: What about the unfamiliar?

General architecture

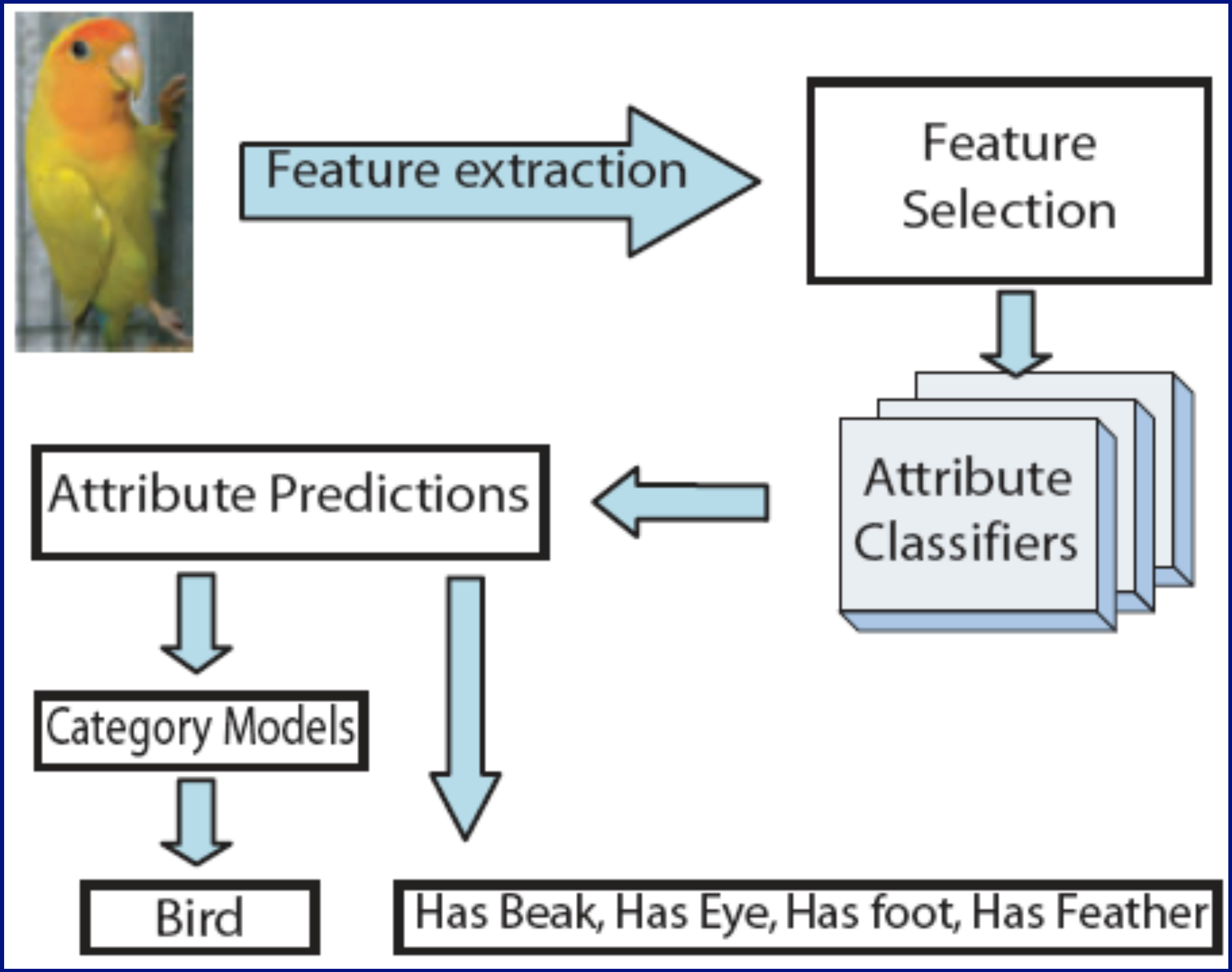


Farhadi et al 09; cf Lampert et al 09

Attribute predictions for unknown objects

Farhadi et al 09; cf Lampert et al 09

General architecture

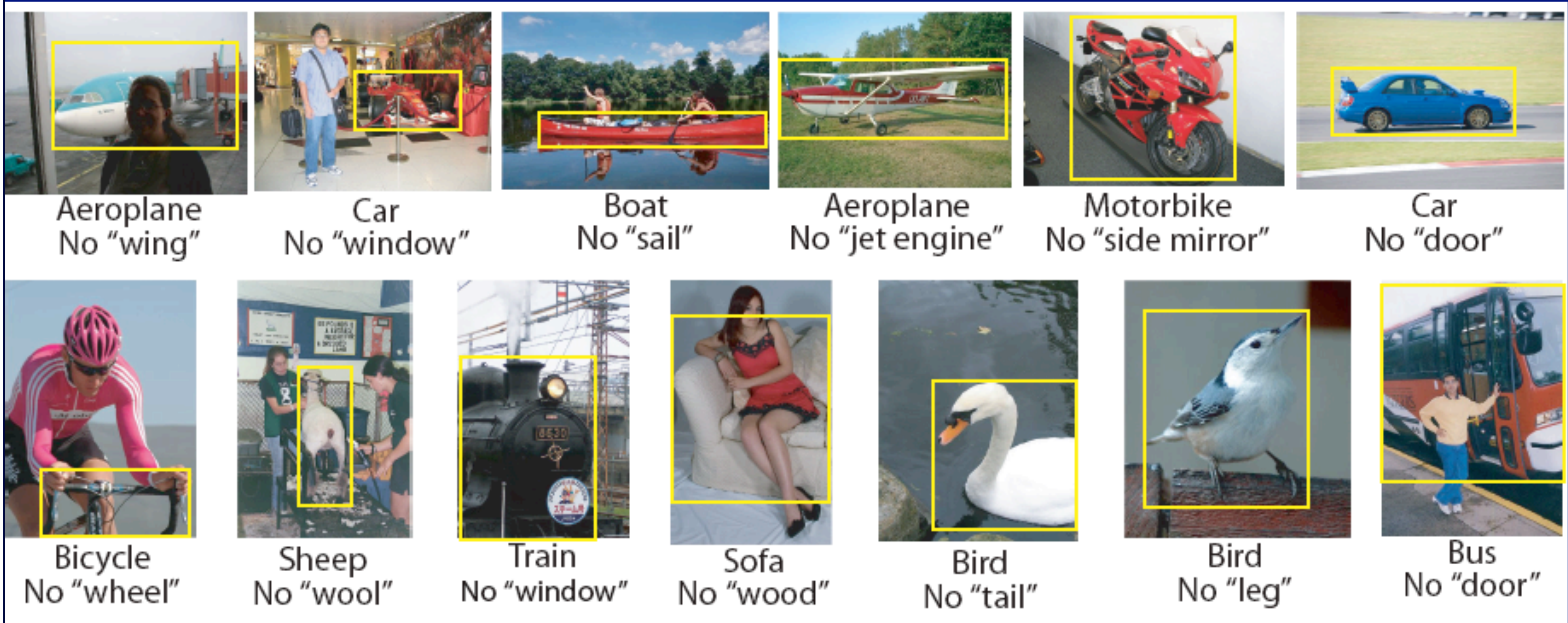


Farhadi et al 09; cf Lampert et al 09

Known objects could be unfamiliar

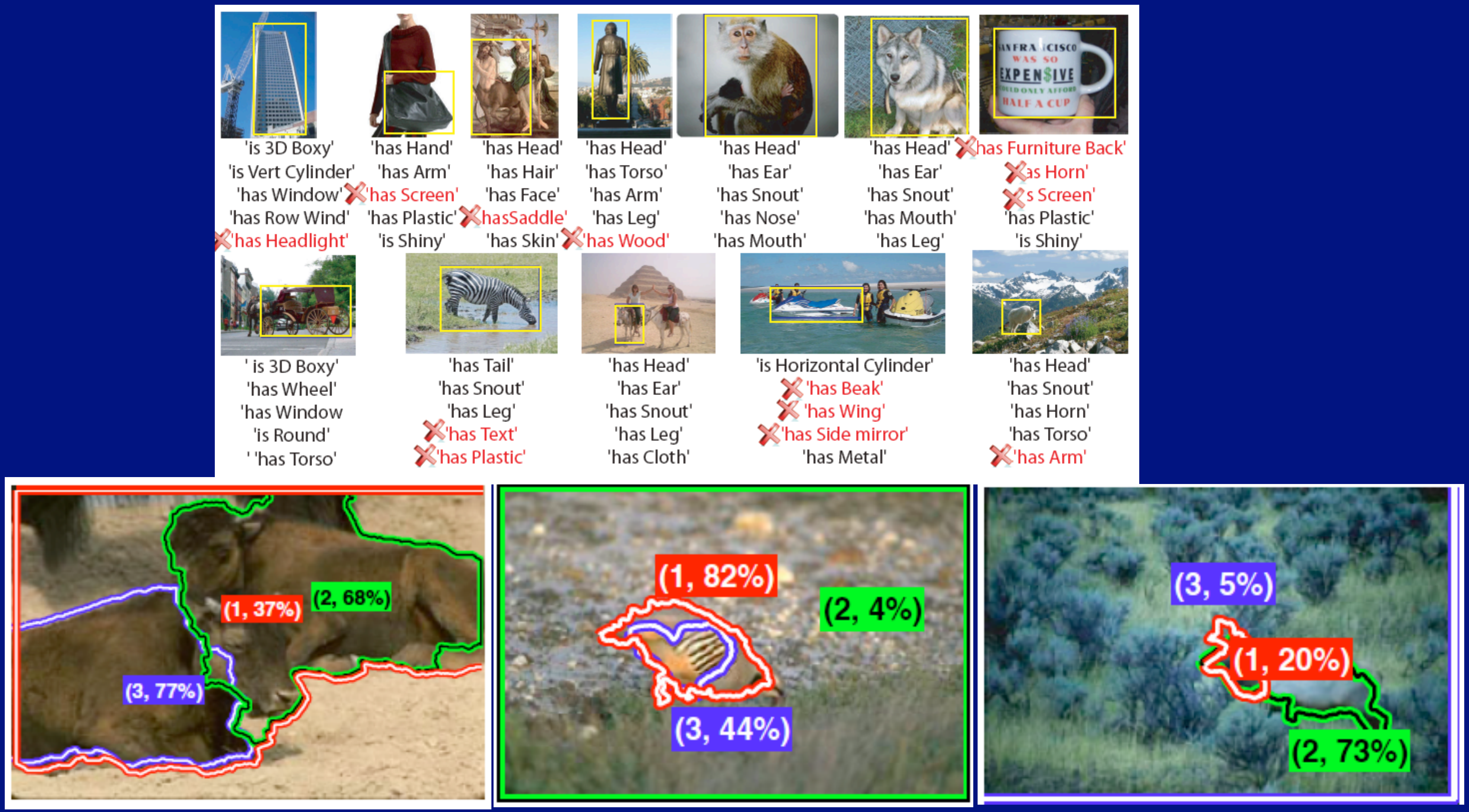
- By being different from the typical
- Pragmatics suggests this is how adjectives are chosen
  - If we are sure it's a cat, and we know that
    - an attribute is different from normal
    - the detector is usually reliable
  - we should report the missing/extra attribute

Missing attributes



Extra attributes

Some regions “want” to be objects



Endres Hoiem 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

Subtleties: Clumps of meaning

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

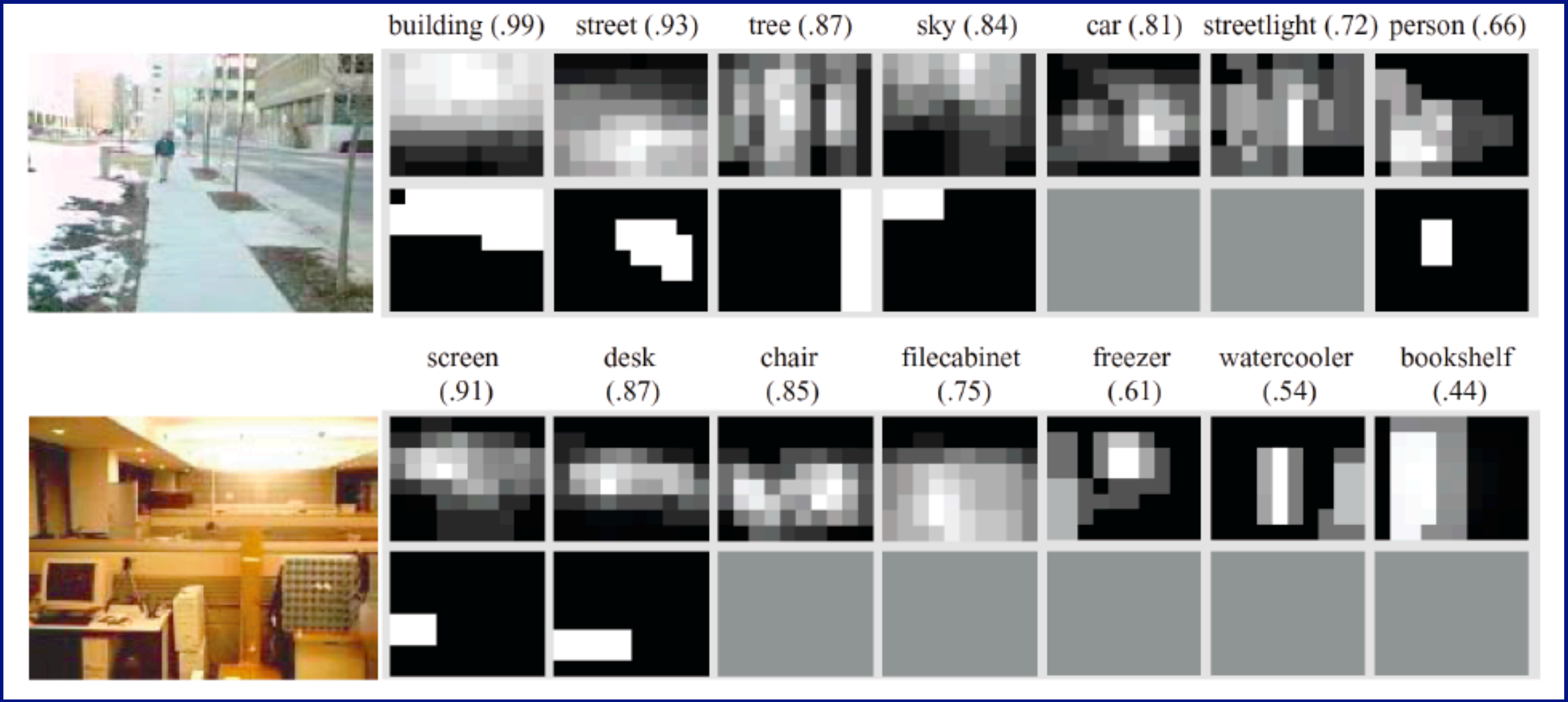
Scenes

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

Xiao et al 10

Scenes

Torralba et al '93



Correlated words

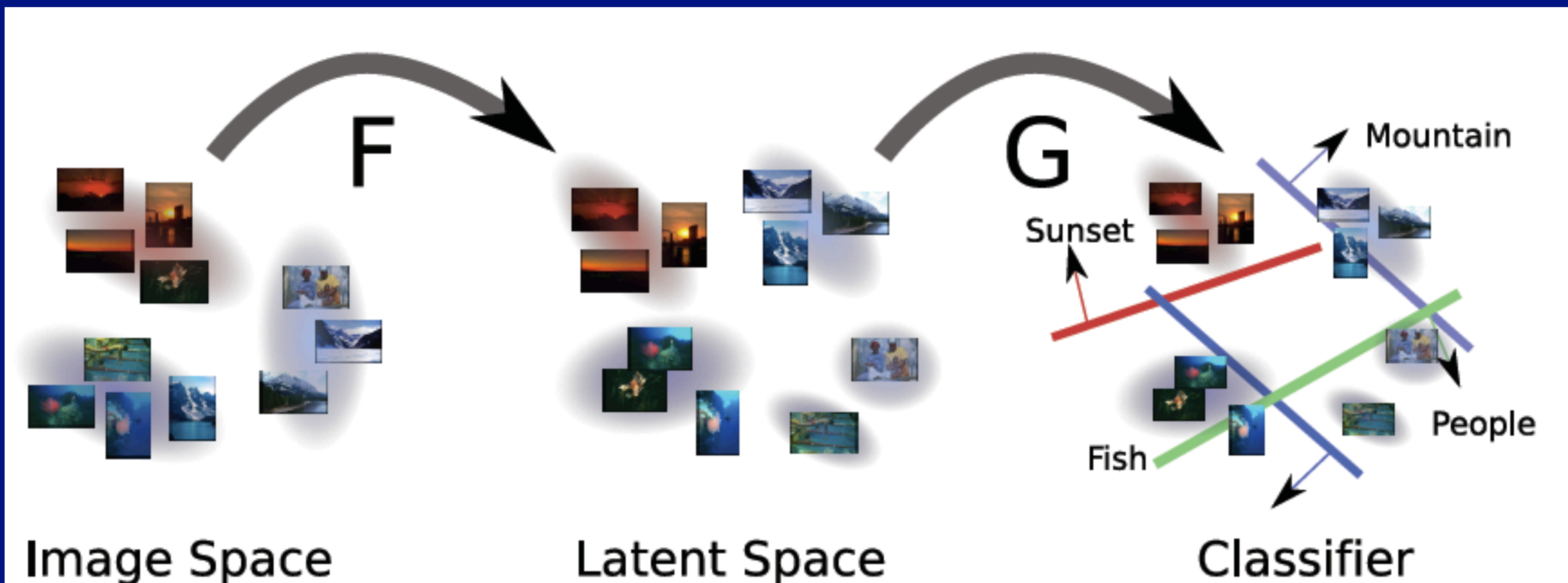
- Idea
  - some features are not helpful
  - a low dimensional subspace is good at predicting most things (Ando +Zhang,)
  - We can find this space by penalizing rank in the matrix of linear classifiers

Learn this

$$\mathcal{D} \approx \mathcal{G}\mathcal{F}\mathcal{X}$$

Word data (observed)

Image representation (observed)



Loeff Farhadi 08; see also Quattoni Darrell 07

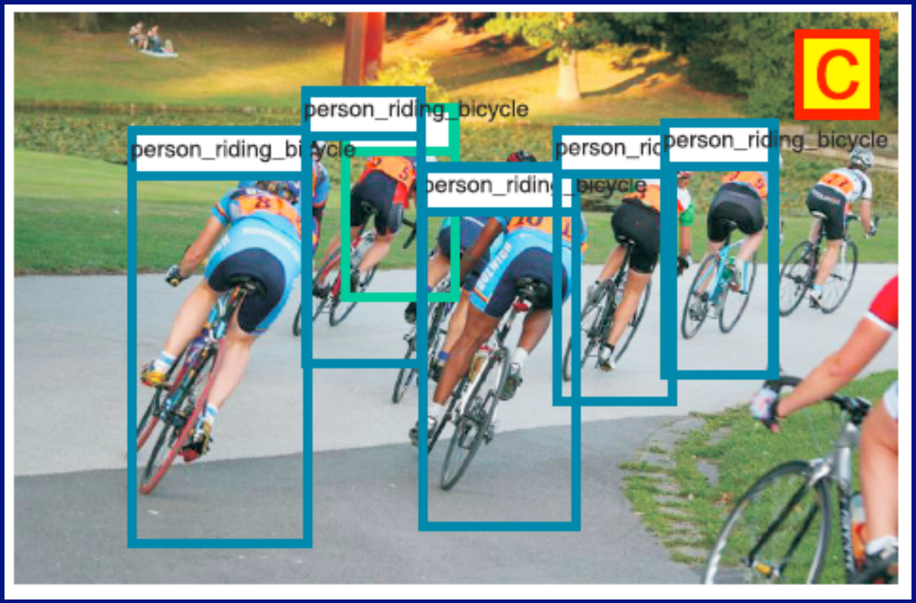
It was there and we didn’t

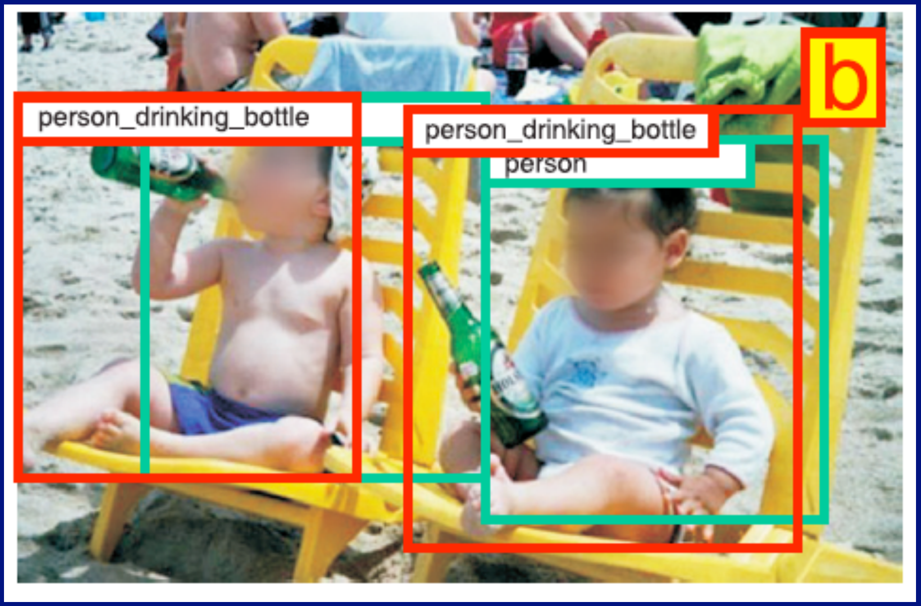
It was there and we predicted it

It wasn’t and we did

Loeff Farhadi 08; see also Quattoni Darrell 07

Scenes > Visual phrases > Objects

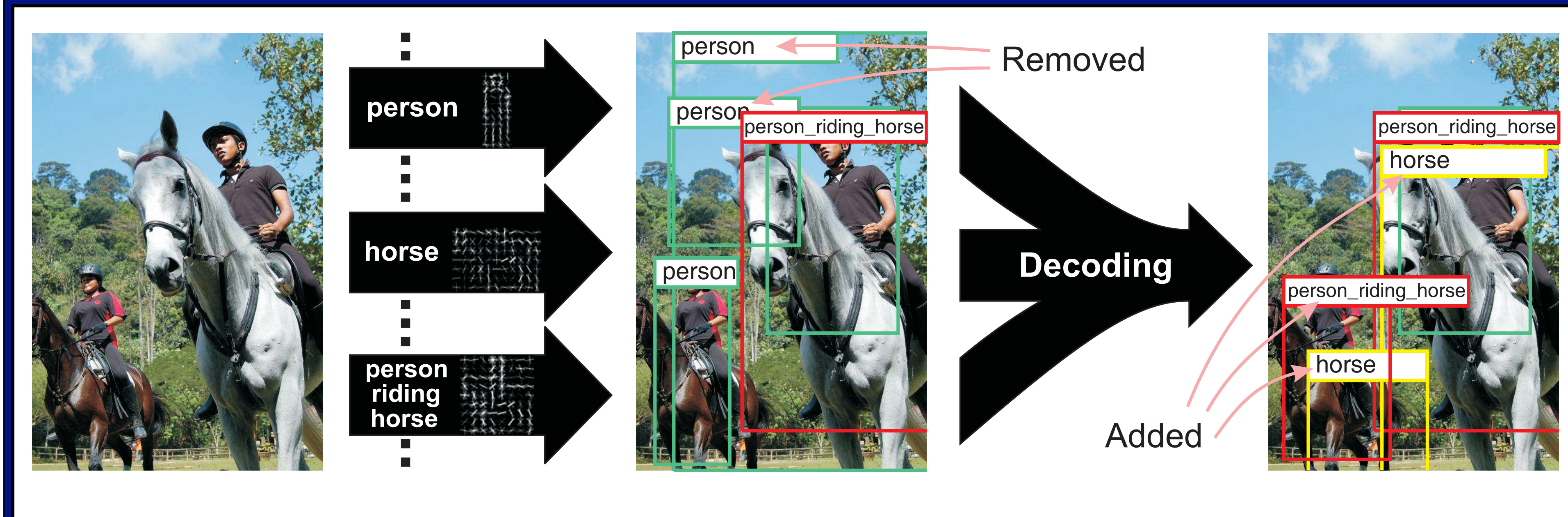




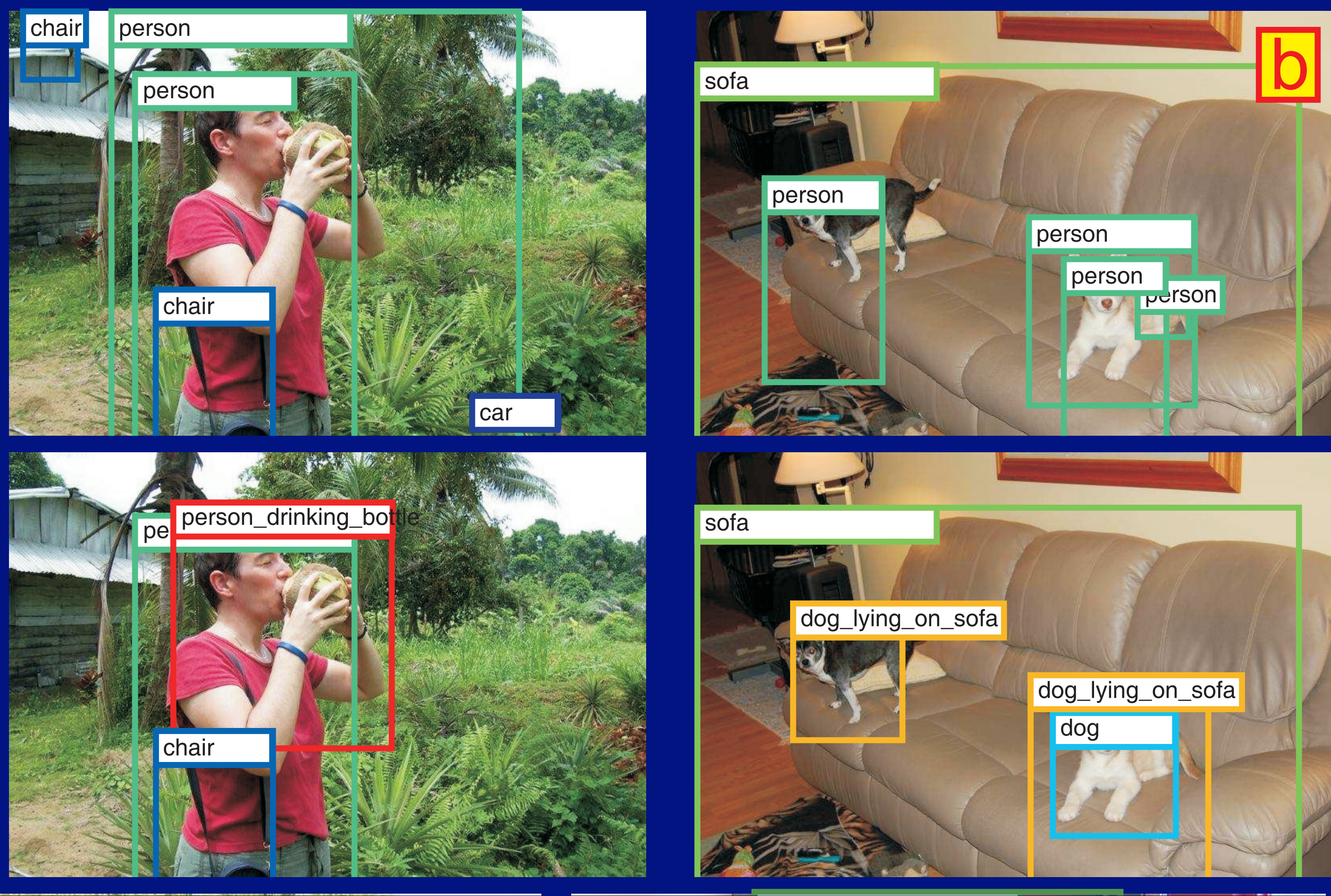
- Composites
  - easier to recognize than their components
  - because appearance is simpler

Farhadi + Sadeghi 11

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

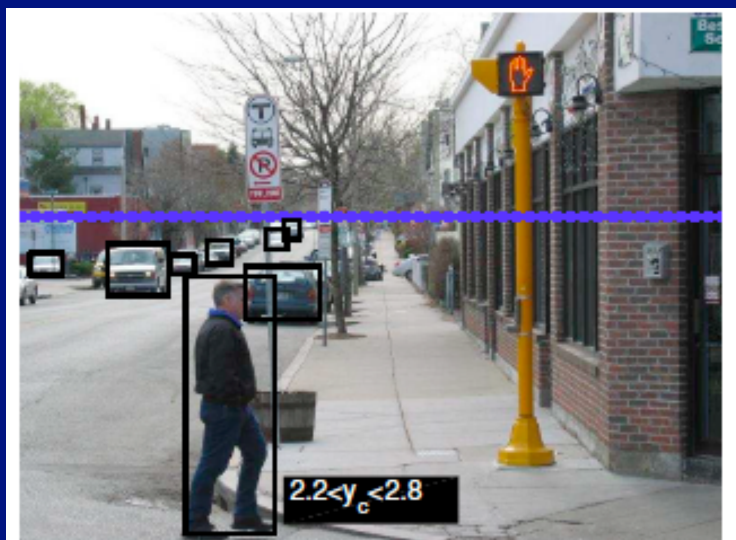


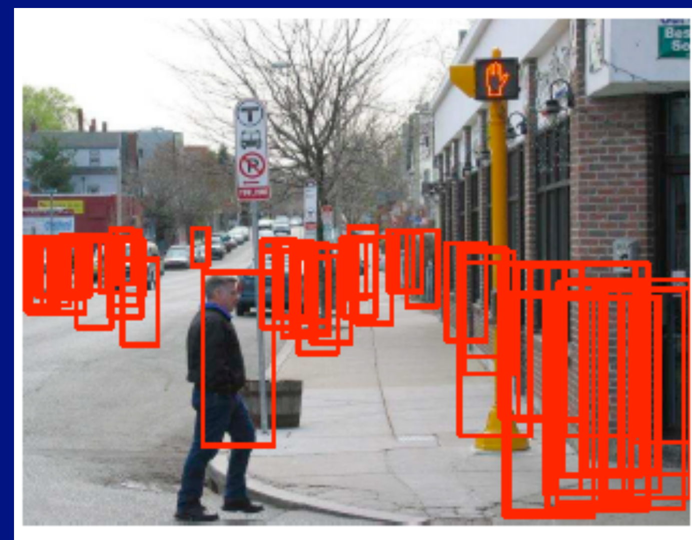
Farhadi Sadeghi 11

Conclusion

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

Subtleties: Environmental knowledge



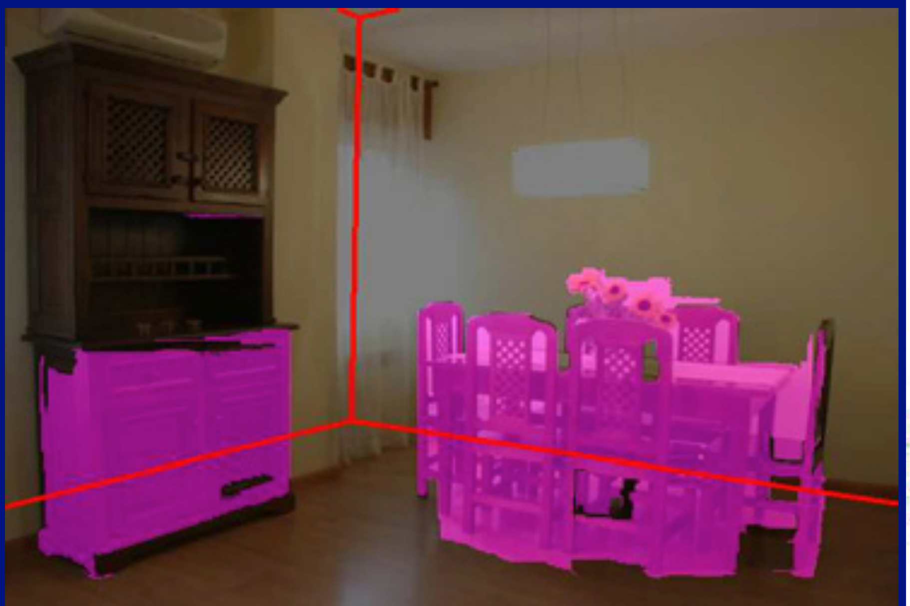


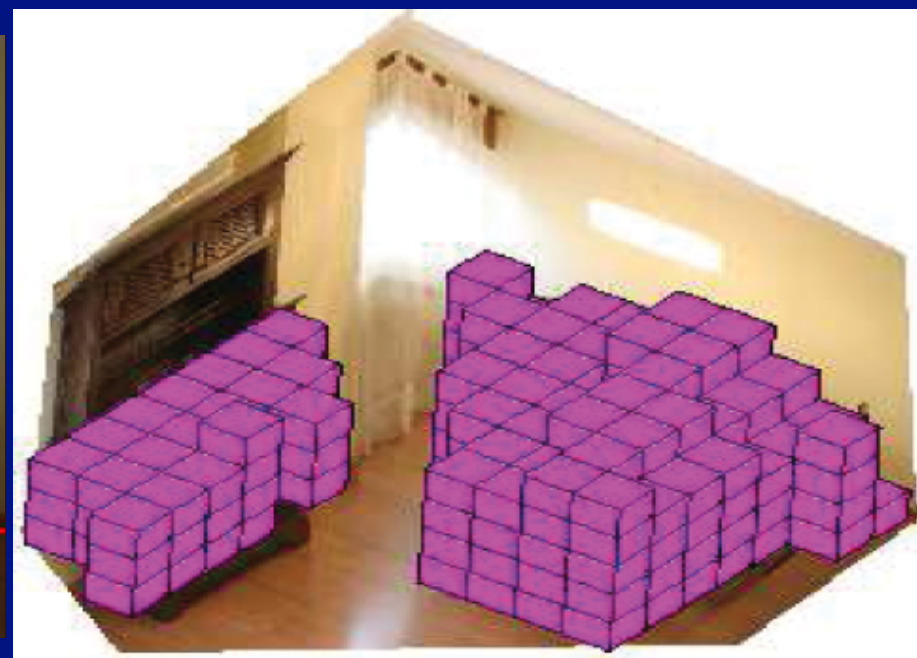
Hoiem et al 06

Environmental knowledge is powerful

Hoiem et al 06

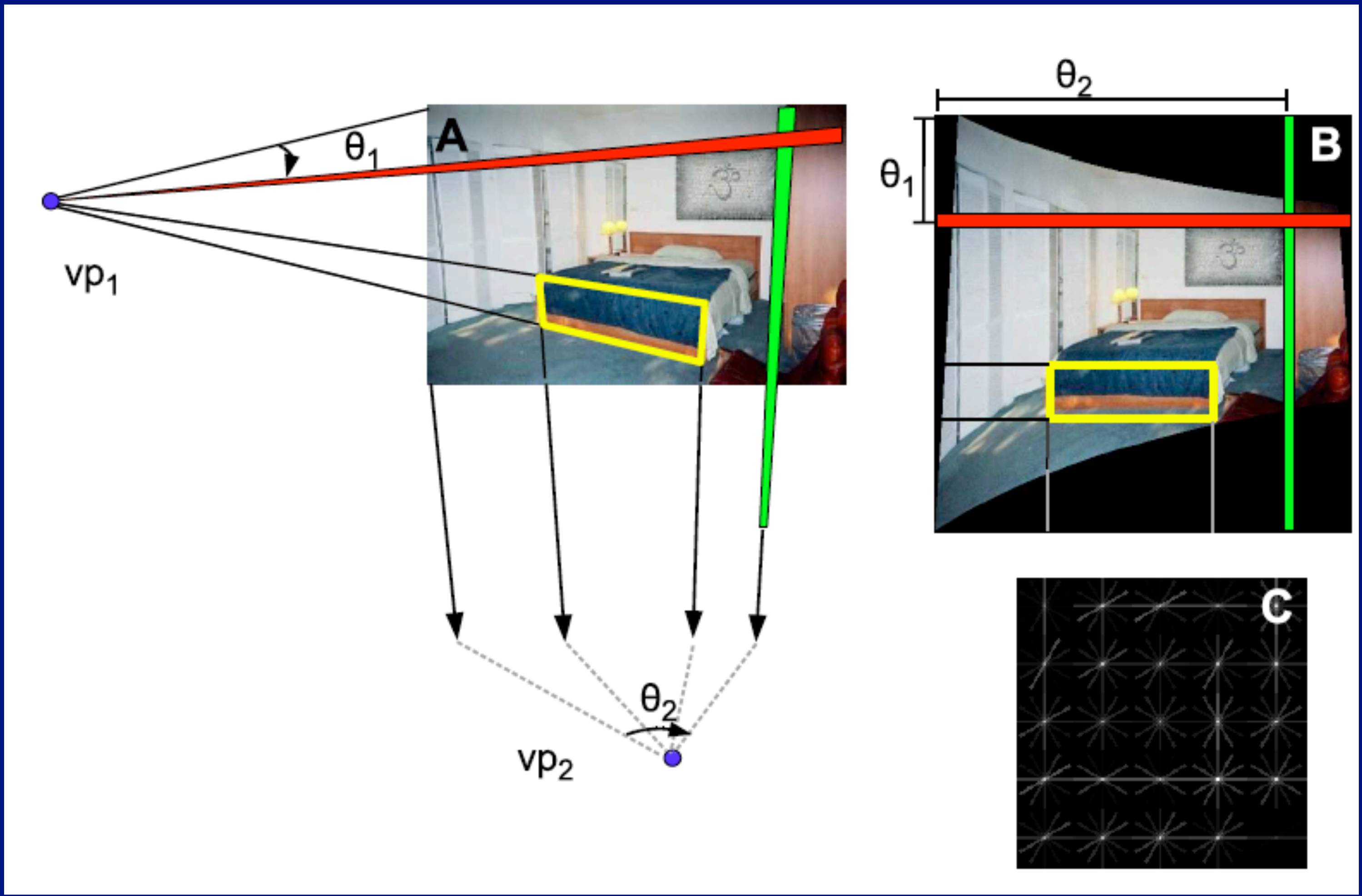
V. Hedau et al ‘09





V. Hedau et al '09

Environmental knowledge is powerful



Hedau et al 2010

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

Subtleties: 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 .

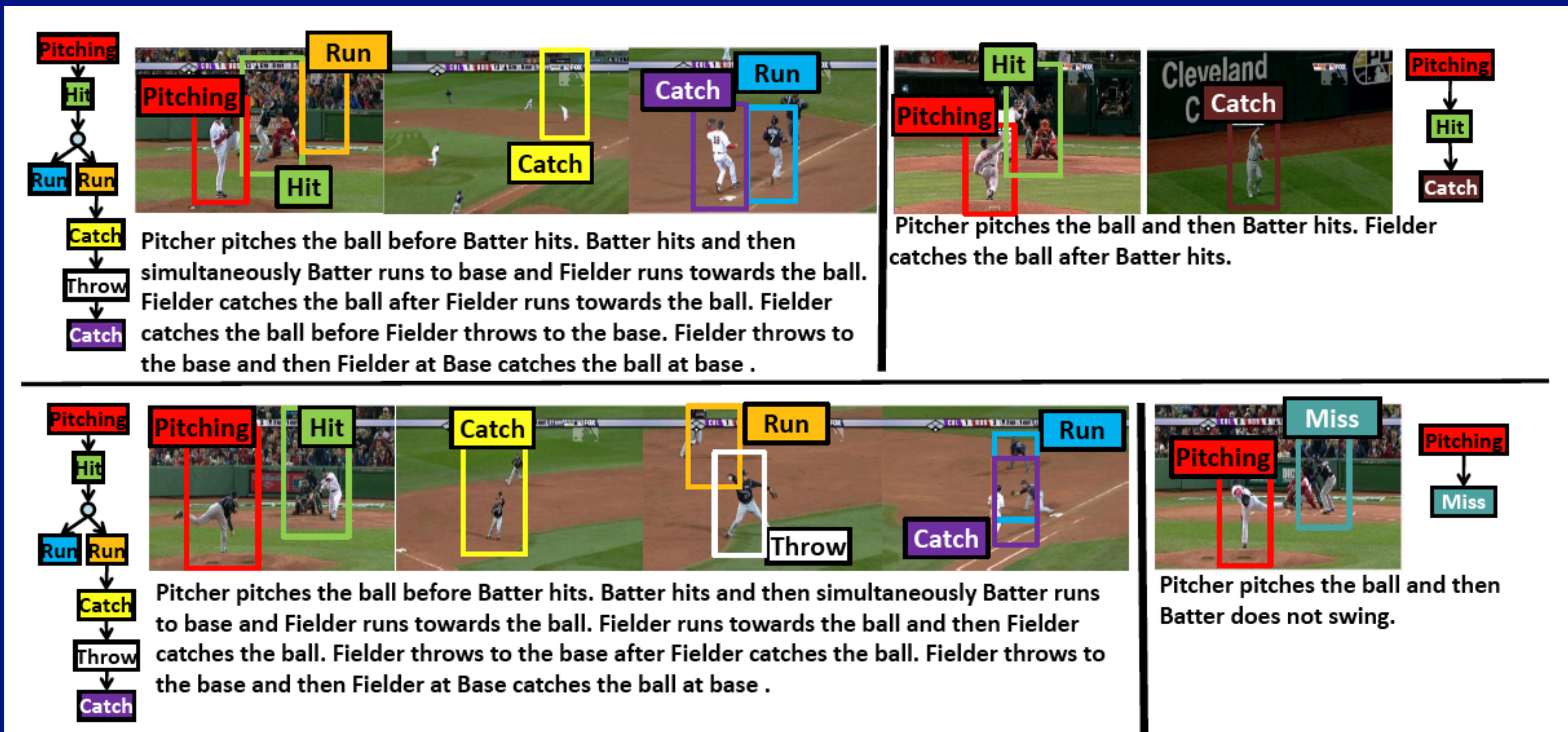
Two women wearing jeans , one with a blue scarf around her head , sit and talk .

Sentences from Julia Hockenmaier’s work

Rashtchian ea 10

For language people: Pragmatics - what is worth saying?

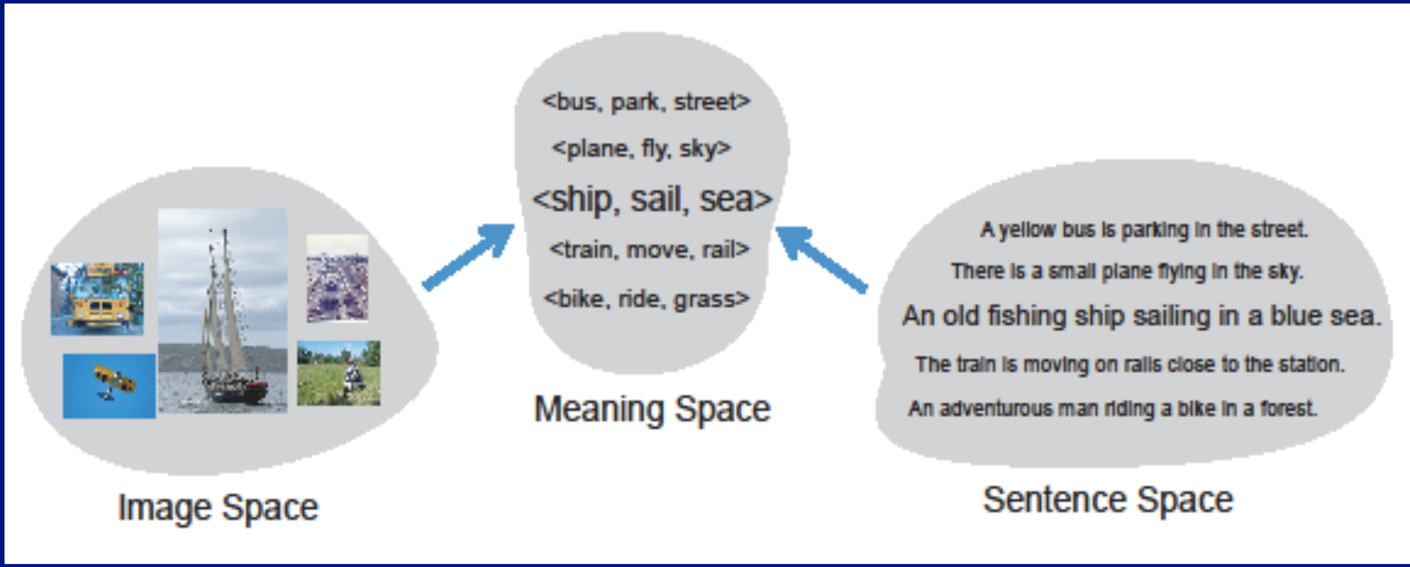
Predicting stylized narrations



Gupta ea 09

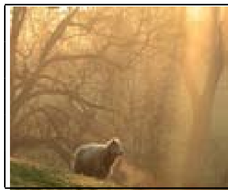
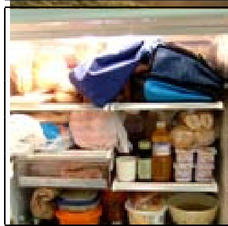
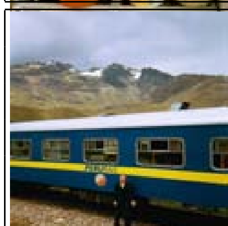
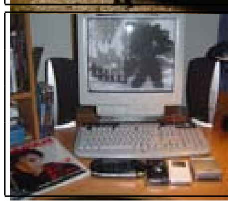
Rich(ish) sentences from simple intermediates

Object, action, scene



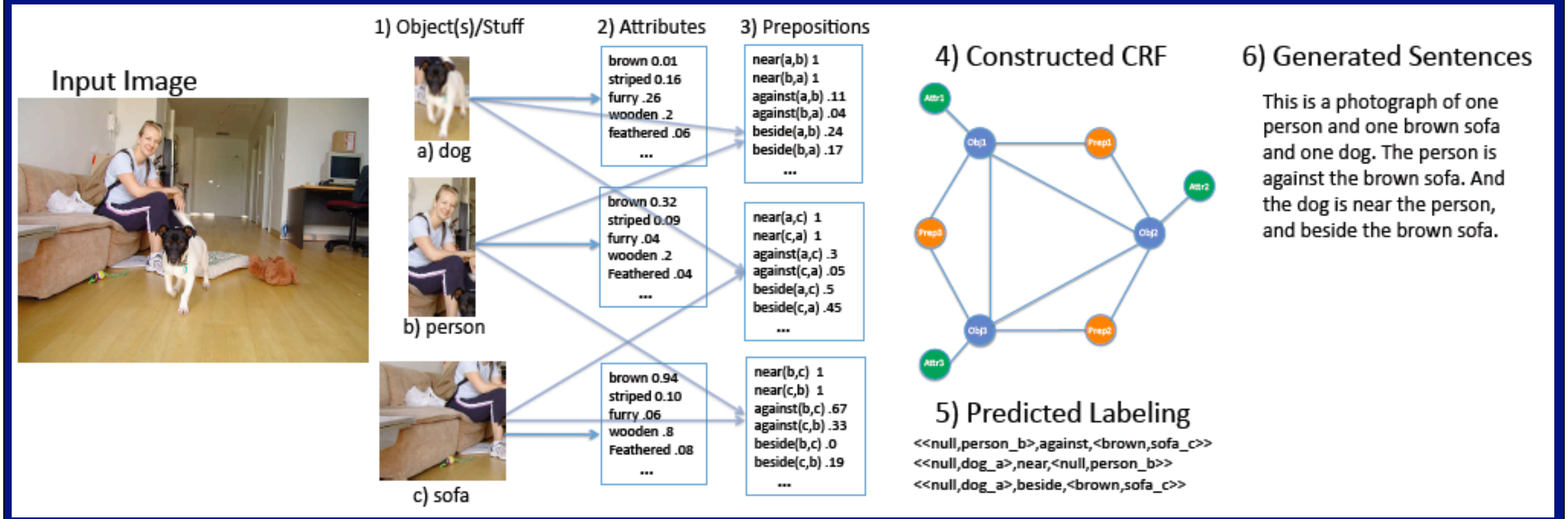
Farhadi ea 10

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

Nobody was hurt in the coming movie


/10/2009
PM

How many adults were on the platform and what were they doing?

What outcome do we expect?

How are other people feeling?

What will they do?

What's going to happen to the baby?


/10/2009 3:55
PM

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

Bias

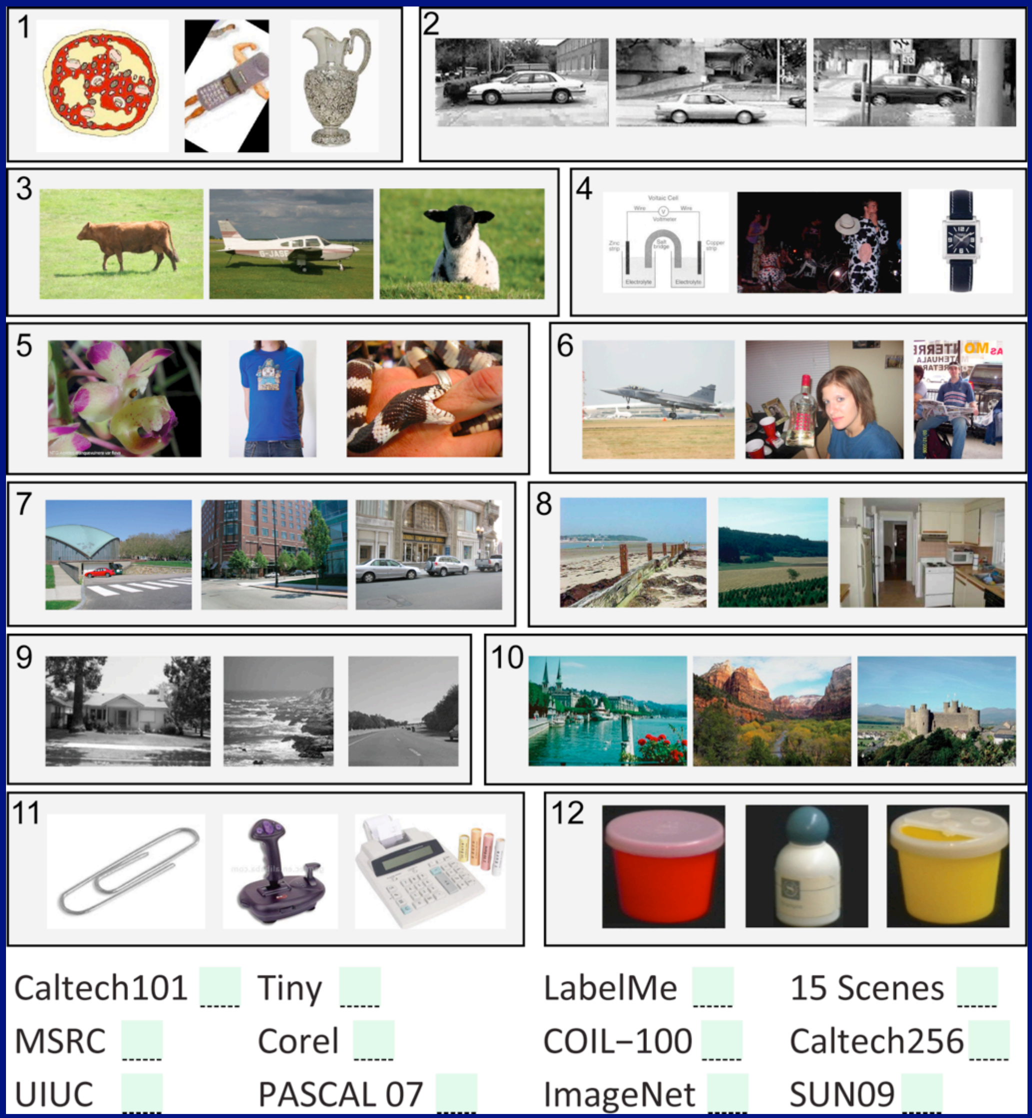
Should not be perjorative

- 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

Bias is pervasive

Torralba+Efros 11



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

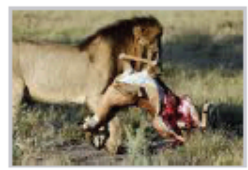

I was doing research on
400 × 300 - 27k - jpg
lowkayhwa.com
Find similar images

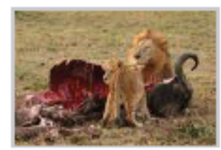

Lion Tiger Size
500 × 553 - 65k - jpg
indrajit.wordpress.com
Find similar images


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

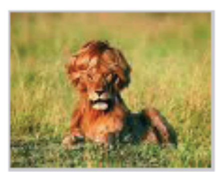

African Lion
500 × 333 - 57k - jpg
itsnature.org
Find similar images


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

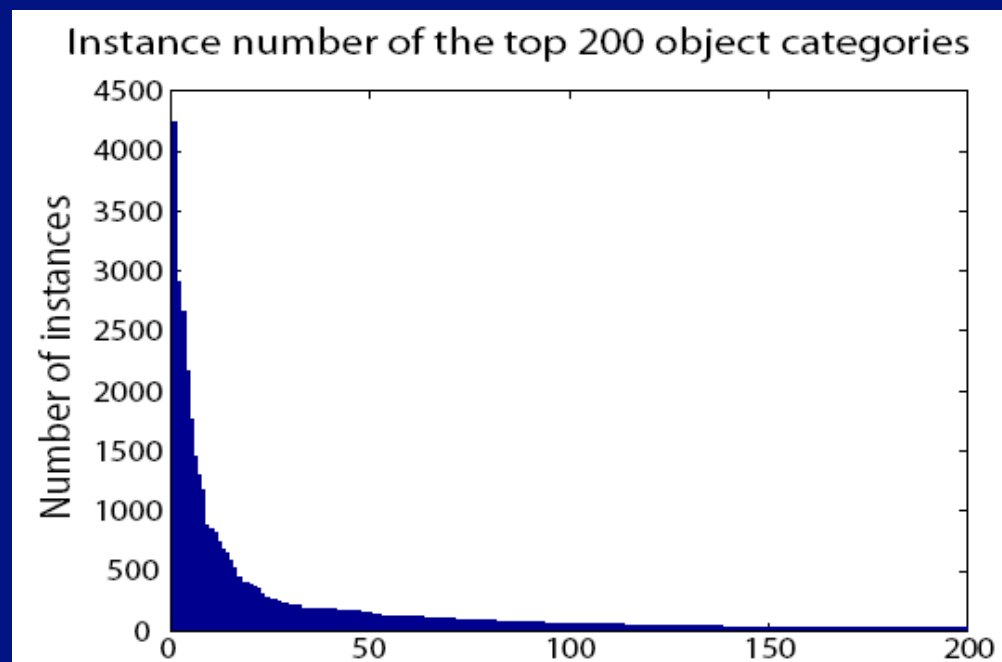

Lion at Sunset
400 × 318 - 25k - jpg
art.com
Find similar images

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 dataset because
  - many things are rare
  - this exaggerates bias



Wang et al, 10

Defenses against Bias

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

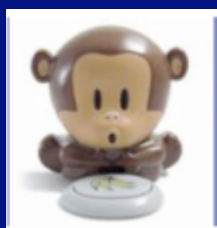
Conclusion

- 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

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

Co-existing category systems



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...

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

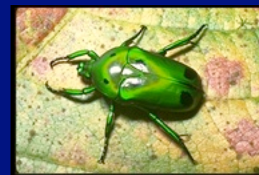
The end

- Thanks to
  - ONR, NSF, Google

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

Roots

- Observation:
  - Pictures affect nearby words (Barnard ea 01a, 01b; Duygulu ea 02; probably many others)
  - Even if they're not really annotations (Berg, 06; quite likely Google, too?)
  - now a really useful commonplace
- Analogy:
  - Object recognition seems somewhat like machine translation, etc.
  - Fertile
    - attributes ?=? adjectives
    - visual phrases
- These are correlations - what's the latent variable?

Meaning

What we can do

Vision has first rate intellectual tools for attacking recognition; we're in amazingly good shape.

- Primary machine is the classifier
  - features in, decision out
- Immensely powerful feature constructions
- Decision is typically label
  - "cat", "dog", "motorcycle", etc.
  - drawn from vocabularies of 20-1000
    - (or so, depending on paper)

Aspects of Meaning

- What is the output of an object recognition system like?
  - not just a list of what's there
- What fragments should be recognized?
  - objects
  - ?
  - scenes
- What are the things we recognize like?
  - what can they do?
  - how do we learn about it?
- How can we deal with the unfamiliar?

Aspects of Meaning

- What is the output of an object recognition system like?
  - not just a list of what's there
- What fragments should be recognized?
  - objects
  - ?
  - scenes
- What are the things we recognize like?
  - what can they do?
  - how do we learn about it?
- How can we deal with the unfamiliar?

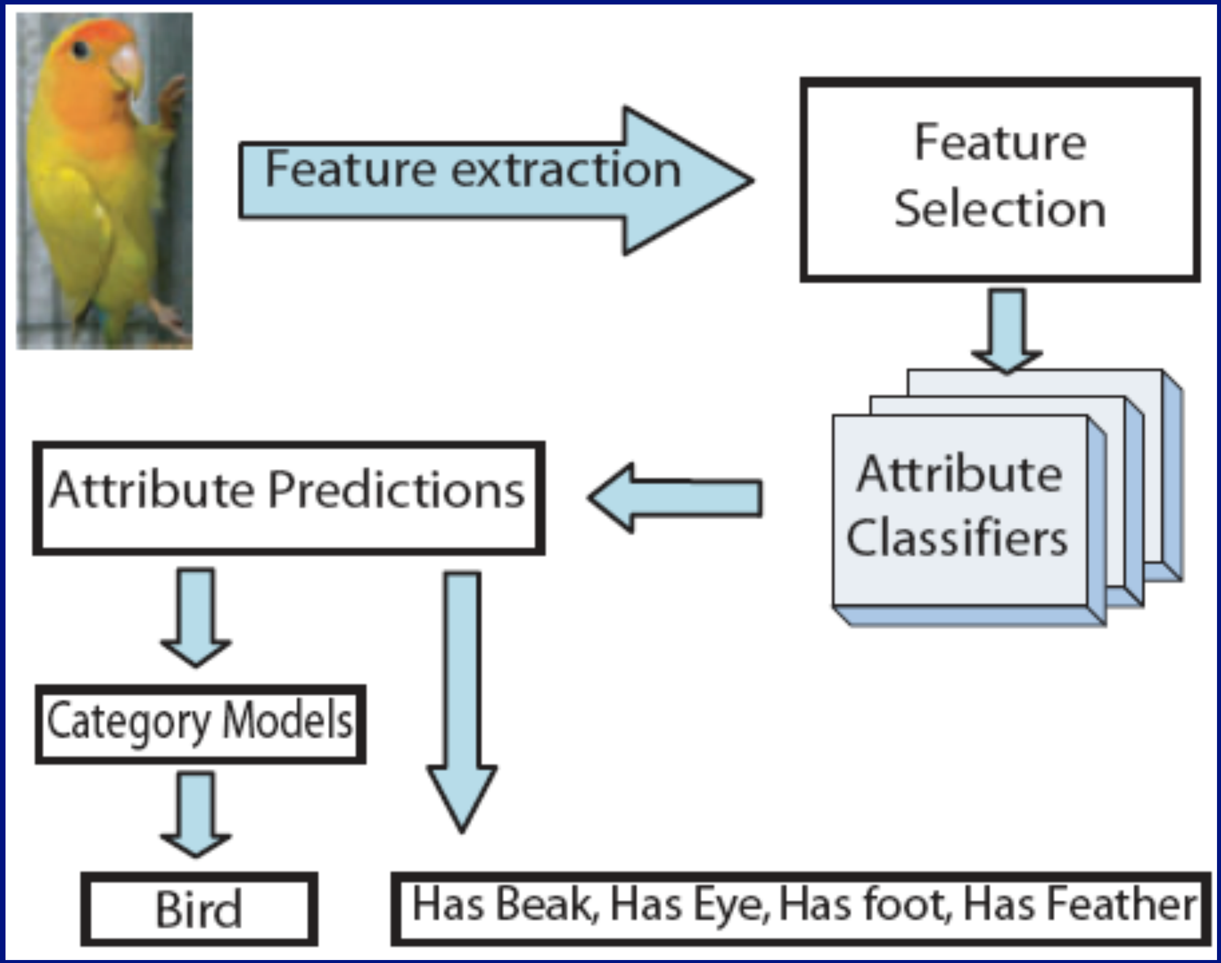
Scenes > Visual phrases > Objects

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

Aspects of Meaning

- What is the output of an object recognition system like?
  - not just a list of what's there
- What fragments should be recognized?
  - objects
  - ?
  - scenes
- What are the things we recognize like?
  - what can they do?
  - how do we learn about it?
- How can we deal with the unfamiliar?

Describing objects



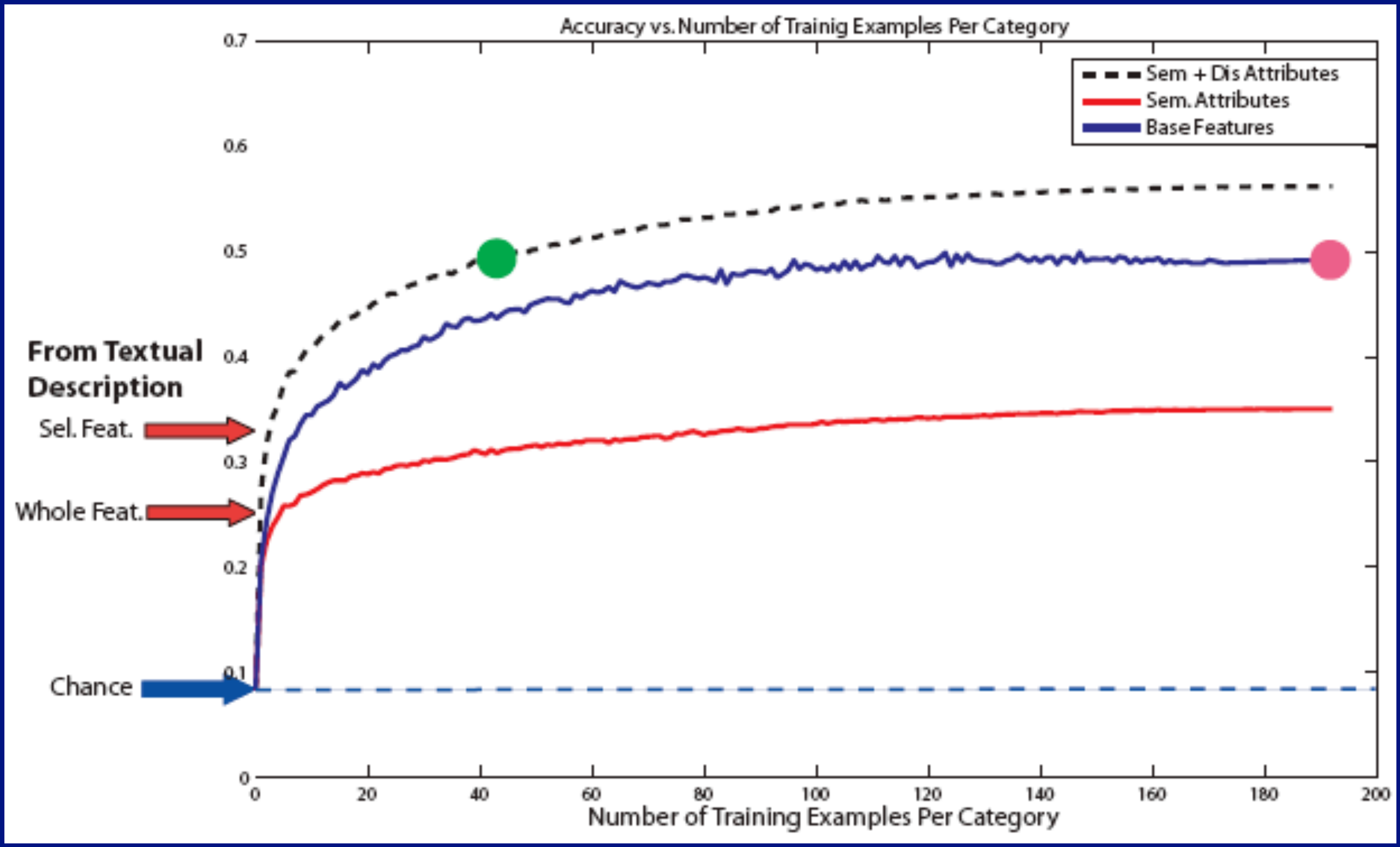
Farhadi ea 09

For language people: a theory of adjectives?

Extra attributes

Farhadi ea 09

Learn by reading

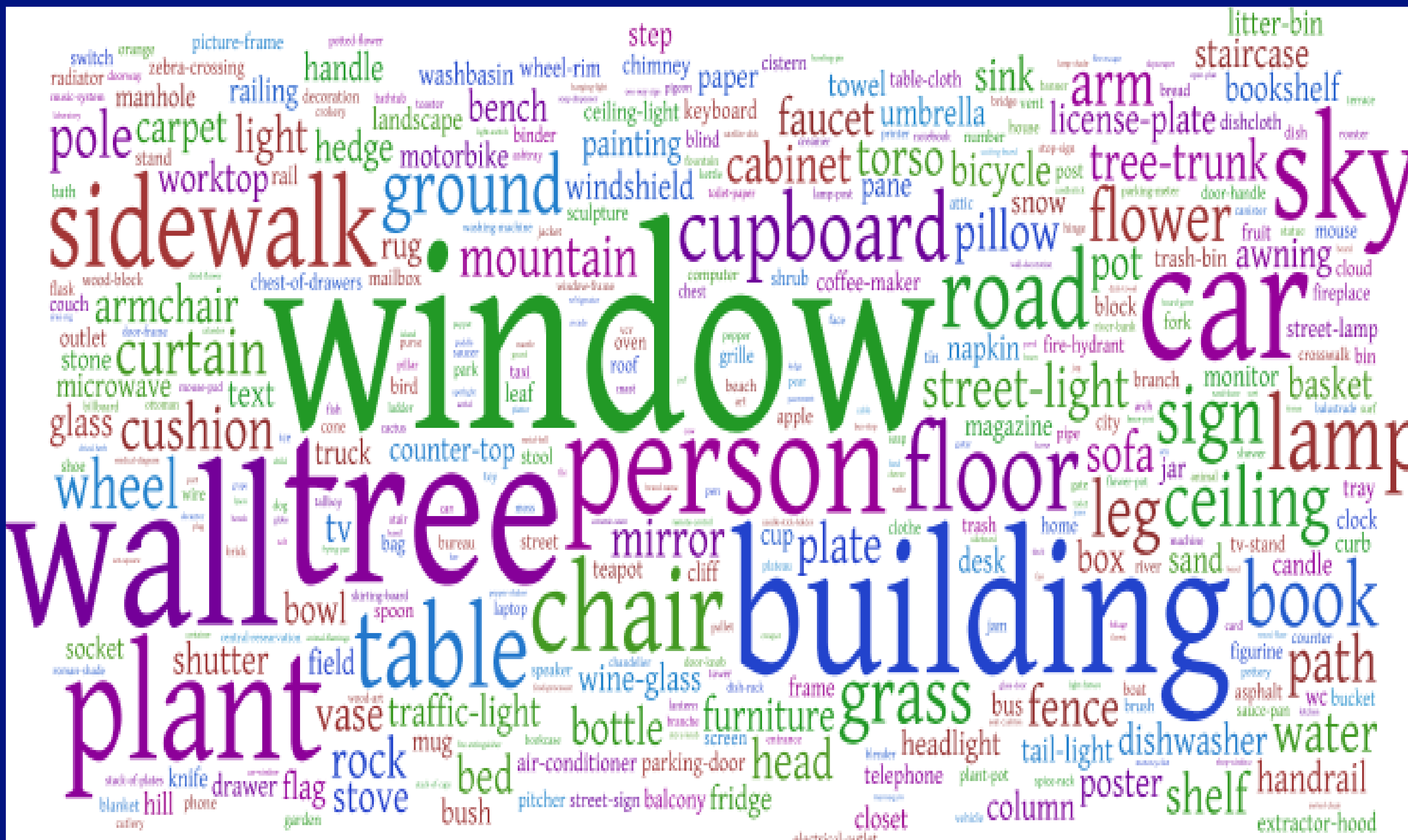


Farhadi ea 09

Aspects of Meaning

- What is the output of an object recognition system like?
  - not just a list of what's there
- What fragments should be recognized?
  - objects
  - ?
  - scenes
- What are the things we recognize like?
  - what can they do?
  - how do we learn about it?
- How can we deal with the unfamiliar?

Most things are unfamiliar



Wang ea 10; labelme data

For language people: distributional semantics?

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