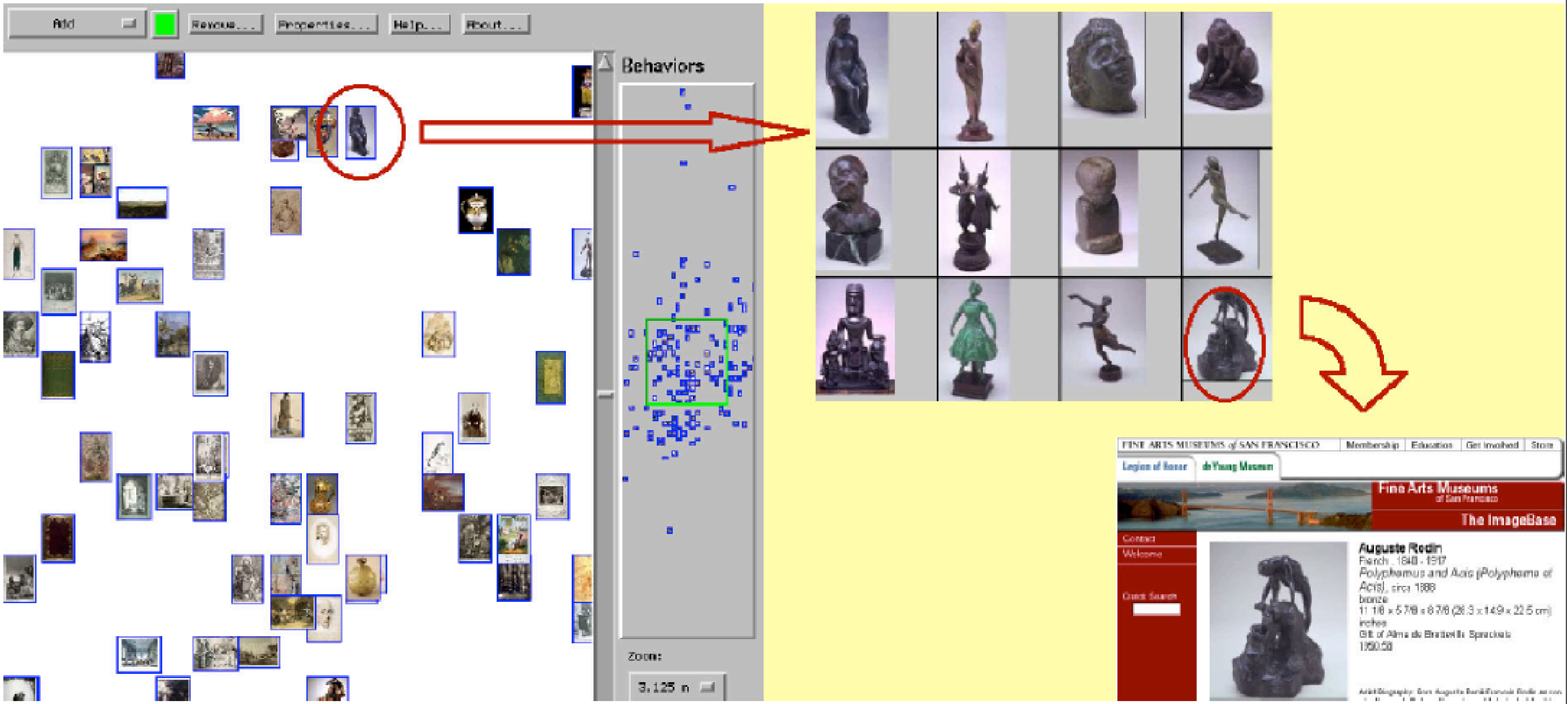
Sentences and pictures:
not just “more words” and pictures

D.A. Forsyth, UIUC
with
Julia Hockenmaier, Derek Hoiem, Ian Endres, Ali Farhadi,
Nicolas Loeff, Cyrus Rashtchian, Gang Wang
all of UIUC
with echoes from
Kobus Barnard (U. Arizona), Pinar Duygulu (Bilkent U.),
Nando de Freitas (UBC

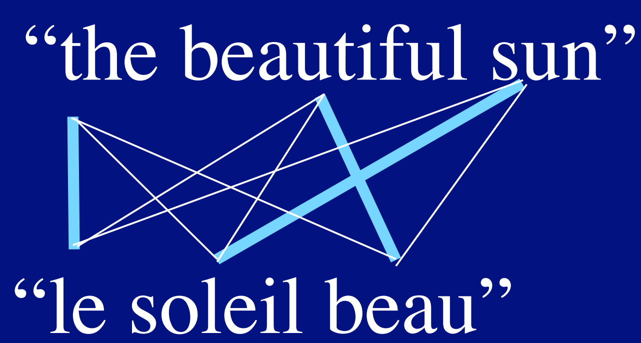
Words and pictures: Implicit csp



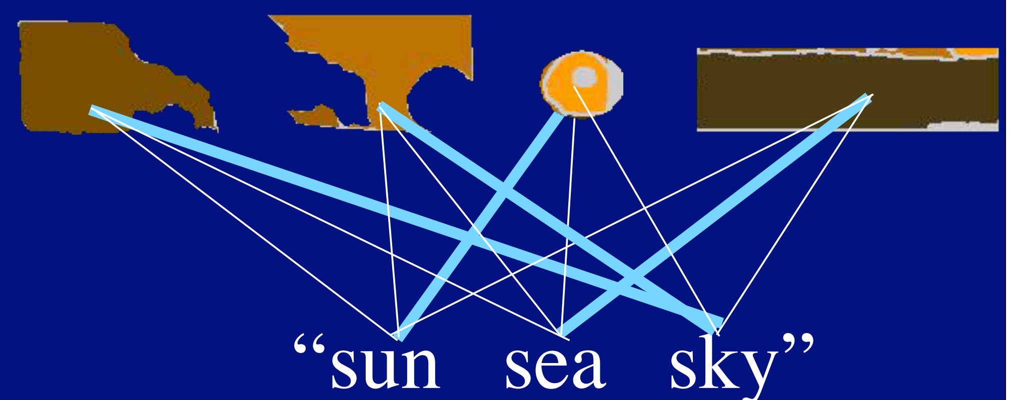
Barnard et al 01, 01

Words and pictures: Explicit csp

- In its simplest form, missing variable problem
- Pile in with EM
  - given correspondences, conditional probability table is easy (count)
  - given cpt, expected correspondences could be easy
- Caveats
  - might take a lot of data; symmetries, biases in data create issues



Brown, Della Pietra, Della Pietra & Mercer 93;
Melamed 01



See Duygulu et al 02

How to generalize words and pictures?

- More accuracy
- More words
- Predict more structure

Accuracy

Method	P	R	F1	Ref	
Co-occ	0.03	0.02	0.02	[53]	Y. Mori et al 99
Trans	0.06	0.04	0.05	[27]	Duygulu et al, 02
CMRM	0.10	0.09	0.10	[37]	Jeon et al 03
TSIS	0.10	0.09	0.10	[19]	Celebi et al 05
MaxEnt	0.09	0.12	0.10	[39]	Jeon et al 04
CRM	0.16	0.19	0.17	[44]	Lavrenko et al 03
CT-3×3	0.18	0.21	0.19	[82]	Yavlinsky et al, 05
CRM-rect	0.22	0.23	0.23	[31]	Feng et al 04
InfNet	0.17	0.24	0.23	[50]	Metzler et al 04
MBRM	0.24	0.25	0.25	[31]	Feng et al 04
MixHier	0.23	0.29	0.26	[17]	Carneiro et al, 05
PicSOM	0.35*	0.35*	0.35*	[73]	Viitaniemi et al 07

More words

- Easy case
  - learn with larger vocabularies
  - tricky bits, but...
- Hard case
  - what do we do about out-of-example words?
  - one simple answer doesn’t work (later)

Structure

- Correlated words
  - waves go with beaches not cats
- Attributes
  - has nose
- Adjectives
  - green hat
- Relations
  - cat on mat
- Sentences
  - A dolphin holds a basketball as it swims on its back

A **dolphin** holds a **basketball** as it swims on its **back** .

A **dolphin** swimming upside down while holding a **basketball**

A **dolphin** swims upside down holding a **basketball** between it 's **flippers** .

A **seal** floats on it 's back in the **water** , holding a **basketball** .

The **dolphin** on his **back** holds the **orange** **basketball**

Correlated Words

- Simple method:
  - rack up some features, build a bunch of linear classifiers one per word
  - works poorly
    - few examples per word
    - many features, only some are stable

Learn this

$$\mathcal{D} \approx \mathcal{M}\mathcal{X}$$

Word data (observed)

Image representation (observed)

Correlated words

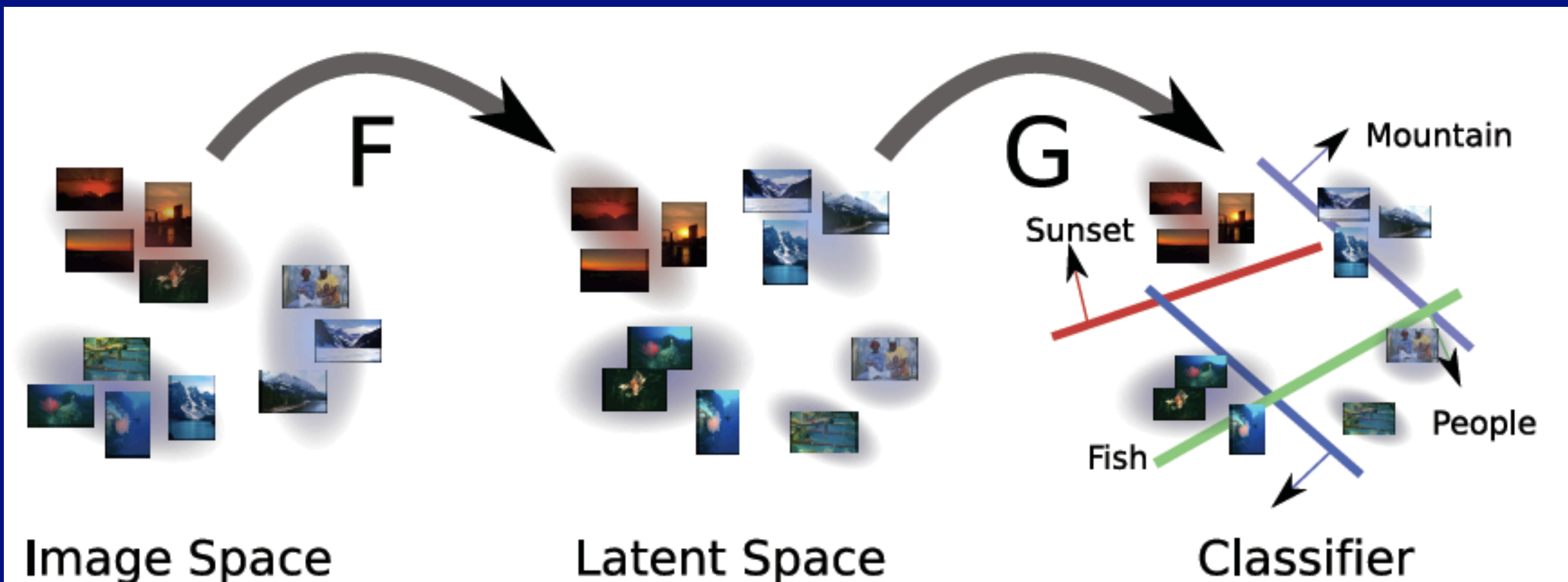
- Idea
  - some features are not helpful
  - a low dimensional subspace is good at predicting most things (Ando +Zhang,)
  - We can find this space by penalizing rank in the matrix of linear classifiers

Learn this

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

Word data (observed)

Image representation (observed)



Loeff Farhadi 08

It was there and we didn't

It was there and we predicted it

It wasn't and we did

Loeff Farhadi 08

Correlated word predictors are quite good

Method	P	R	F1	Ref
Co-occ	0.03	0.02	0.02	[53]
Trans	0.06	0.04	0.05	[27]
CMRM	0.10	0.09	0.10	[37]
TSIS	0.10	0.09	0.10	[19]
MaxEnt	0.09	0.12	0.10	[39]
CRM	0.16	0.19	0.17	[44]
CT-3×3	0.18	0.21	0.19	[82]
CRM-rect	0.22	0.23	0.23	[31]
InfNet	0.17	0.24	0.23	[50]
MBRM	0.24	0.25	0.25	[31]
MixHier	0.23	0.29	0.26	[17]
(section 2.2)	0.27	0.27	0.27	
(section 2.2, kernel)	0.29	0.29	0.29	
PicSOM	0.35*	0.35*	0.35*	[73]

Loeff Farhadi 08

Structure

- Correlated words
  - waves go with beaches not cats
- Attributes
  - has nose
- Adjectives
  - green hat
- Relations
  - cat on mat
- Sentences
  - A dolphin holds a basketball as it swims on its back

A **dolphin** holds a **basketball** as it swims on its **back** .

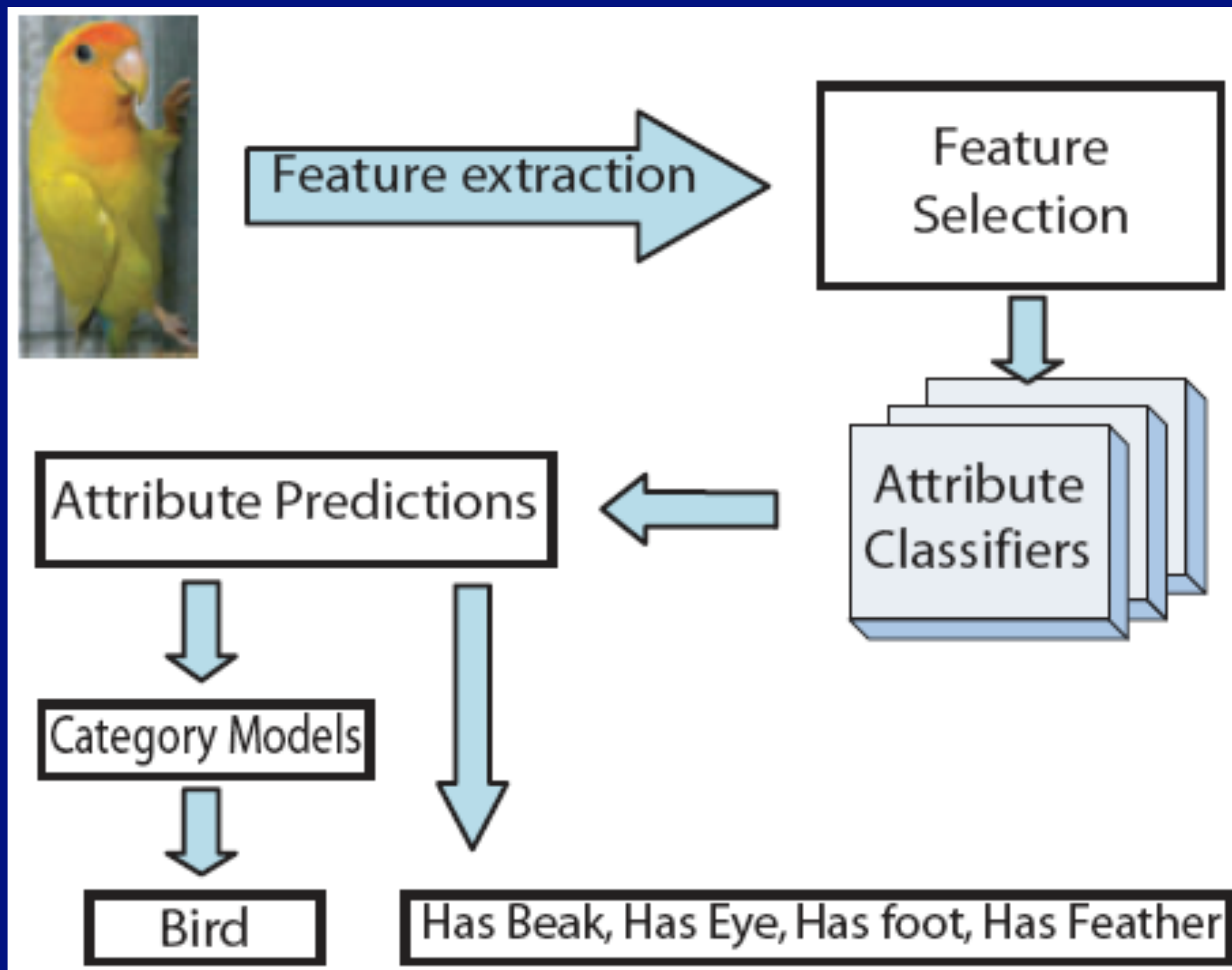
A **dolphin** swimming upside down while holding a **basketball**

A **dolphin** swims upside down holding a **basketball** between it 's **flippers** .

A **seal** floats on it 's back in the **water** , holding a **basketball** .

The **dolphin** on his **back** holds the **orange** **basketball**

General architecture

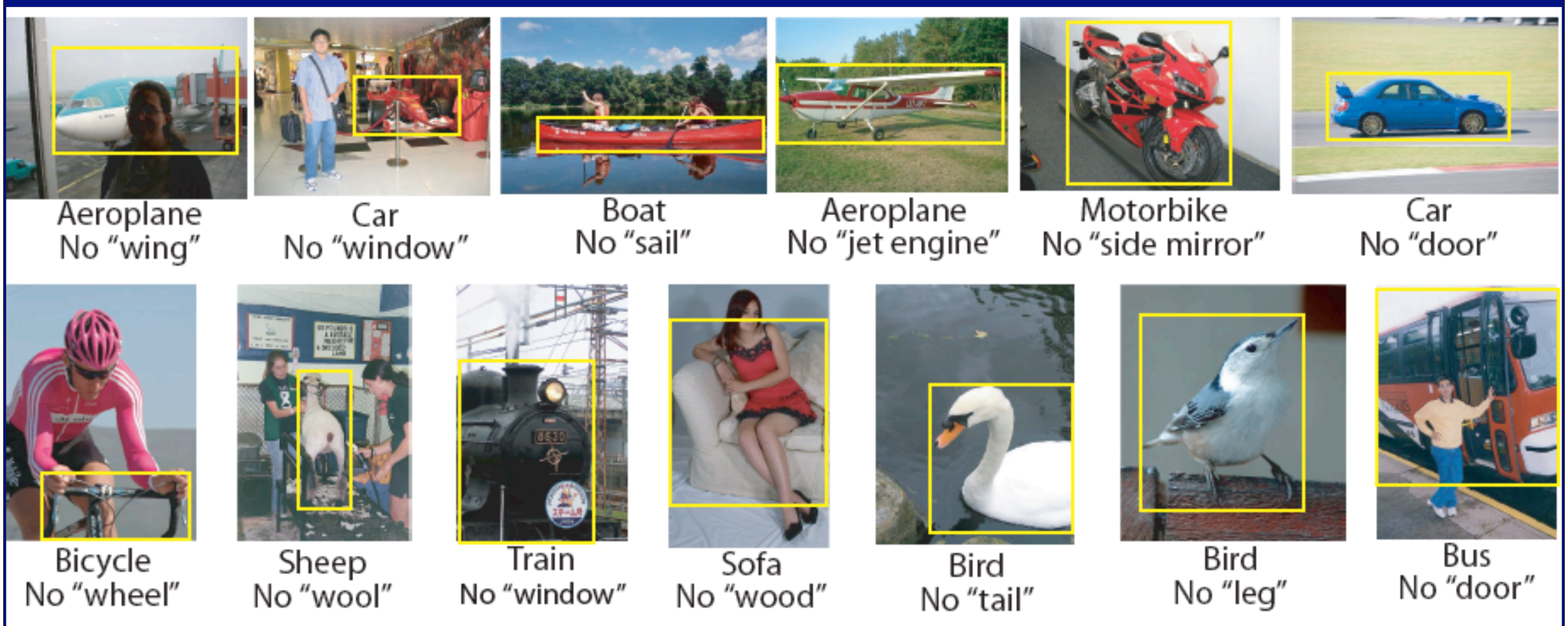


Farhadi et al 09; cf Lampert et al 09

How is an object different from typical?

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

Missing attributes



Extra attributes

Structure

- Correlated words
  - waves go with beaches not cats
- Attributes
  - has nose
- Adjectives
  - green hat
- Relations
  - cat on mat
- Sentences
  - A dolphin holds a basketball as it swims on its back

A **dolphin** holds a **basketball** as it swims on its **back** .

A **dolphin** swimming upside down while holding a **basketball**

A **dolphin** swims upside down holding a **basketball** between it 's **flippers** .

A **seal** floats on it 's back in the **water** , holding a **basketball** .

The **dolphin** on his **back** holds the **orange** **basketball**

“Pink” from Google

Yanai Barnard 05

“Pink” after 10 EM iterations

Yanai Barnard 05

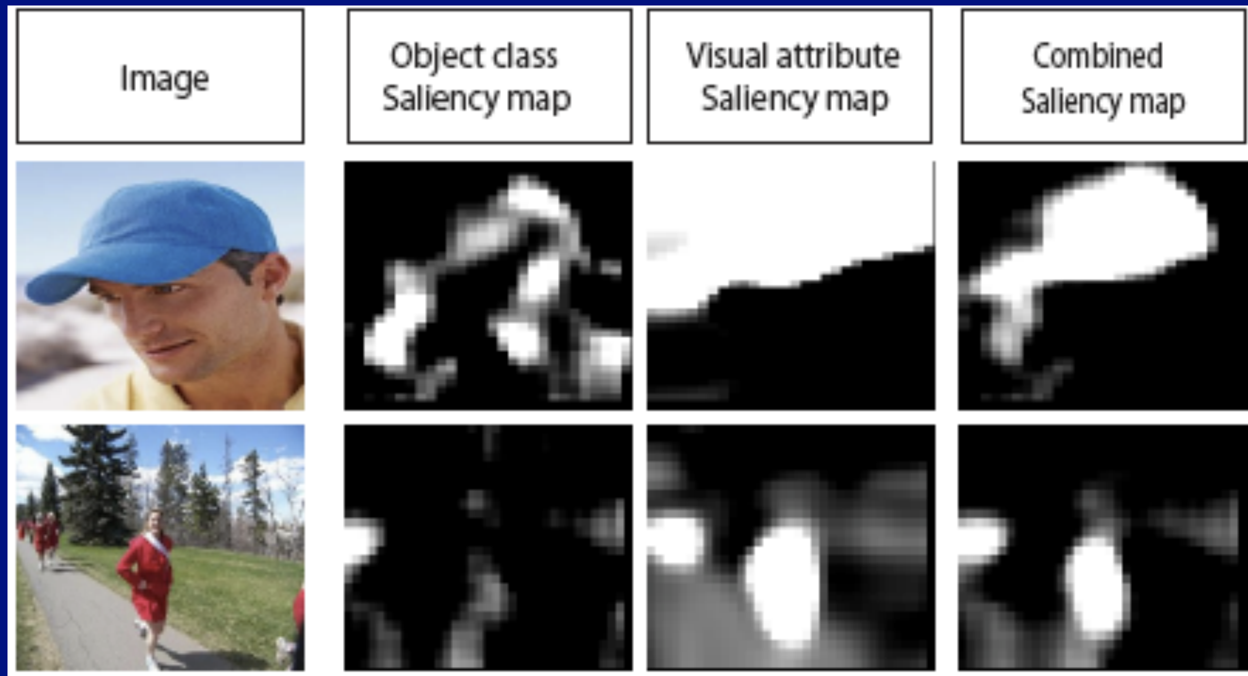


Image
Object class Saliency map
Visual attribute Saliency map
Combined Saliency map

Wang et al 09


Ground truth
Round 1
Round 2
Round 3

Structure

- Correlated words
  - waves go with beaches not cats
- Modifiers
  - pink cadillac
- Attributes
  - has nose
- Modifier-noun pairs
  - green hat
- Relations
  - cat on mat

Gupta and Davis 08, but there is still a lot here

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

Structure

- Correlated words
  - waves go with beaches not cats
- Attributes
  - has nose
- Adjectives
  - green hat
- Relations
  - cat on mat

Gupta and Davis 08,
but there is still a lot here

- Sentences
  - A dolphin holds a basketball as it swims on its back

A **dolphin** holds a **basketball** as it swims on its **back** .

A **dolphin** swimming upside down while holding a **basketball**

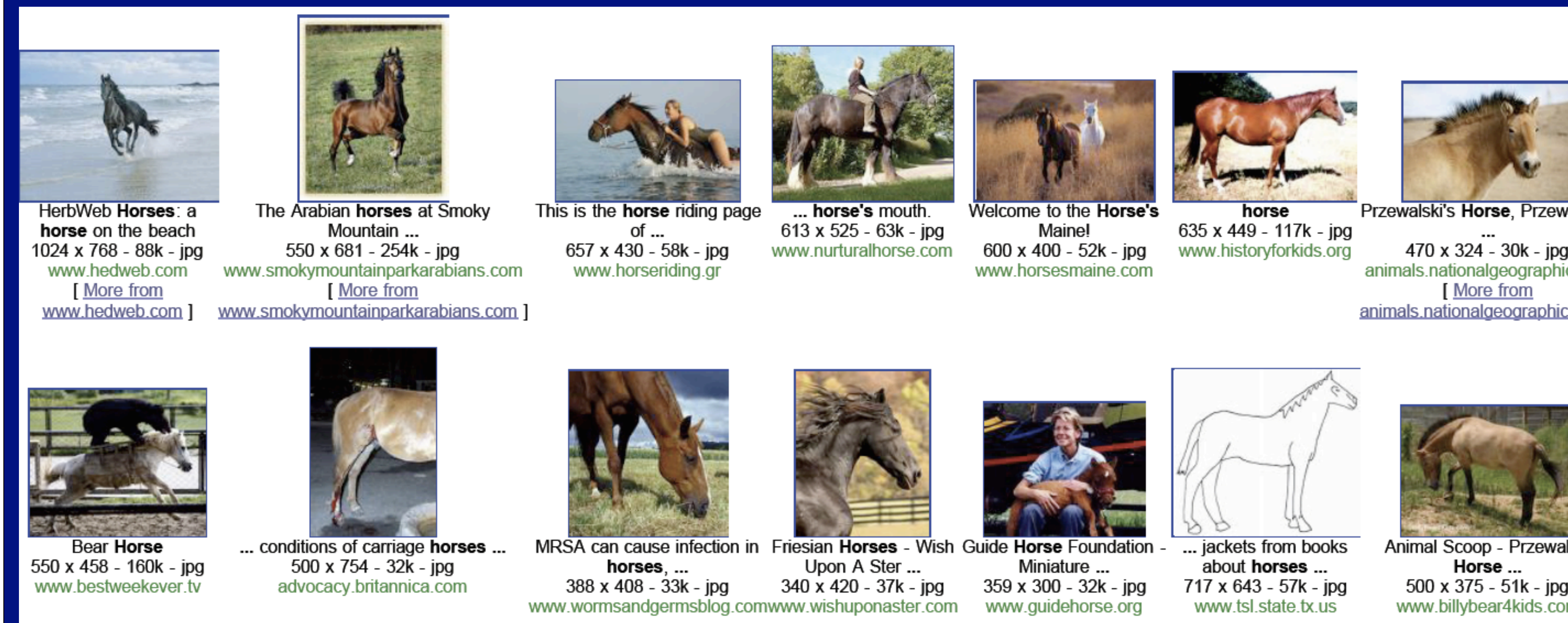
A **dolphin** swims upside down holding a **basketball** between it 's **flippers** .

A **seal** floats on it 's back in the **water** , holding a **basketball** .

The **dolphin** on his **back** holds the **orange** **basketball**

Relations distort participants

Relations distort participants



Relations distort participants

Structure

- Correlated words
  - waves go with beaches not cats
- Attributes
  - has nose
- Adjectives
  - green hat
- Relations
  - cat on mat
- Sentences
  - A dolphin holds a basketball as it swims on its back

A **dolphin** holds a **basketball** as it swims on its **back** .

A **dolphin** swimming upside down while holding a **basketball**

A **dolphin** swims upside down holding a **basketball** between it 's **flippers** .

A **seal** floats on it 's back in the **water** , holding a **basketball** .

The **dolphin** on his **back** holds the **orange** **basketball**

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 .

A crowd of young adults in a dark room .

A girl in a brown shirt and a
blue jean skirt is dancing with a
young man dressed in a blue shirt wearing
a black backpack .

A group of people standing in a
dark building .

A large group of people dancing in a bar

Dancing at club and two guys bucking up

Conclusions

- Real progress in accuracy
- Structure is still hard, but rewarding
- Big problem: predicting sentences