BİL722

Understanding and Predicting Inportance in Images

Alexander C. Berg, et al.

Pınar Küllü

N12247681

- A raft with 3 adults and two children in a river.
- Four people in a canoe paddling in a river lined with cliffs.
- Several people in a canoe in the river.

Problem Overview

- the problem of understanding and predicting perceived importance of image content.
- What factors do people inherently use to determine importance?

What's in this image?

man	chair
baby	boxes
sling	cups
ladder	water bottle
fridge	wall
table	pacifier
glasses	beard
shirt	watermelon

. . .

What do people describe?

"A **bearded man** stands while holding a **small child** in a **green sheet**."

"A **bearded man** with a **baby** in a **sling** poses."

"**Man** standing in **kitchen** with **little girl** in **green sack**."

"**Man** with **beard** and **baby**"

"A **bearded man** is holding a **child** in a **sling**."

Important content:

man, **beard**, **baby**, **sling**, **kitchen**

- There are some underlying consistent factors influencing people's perception of importance in pictures.
  - factors related to image composition such as size and location
  - factors related to content semantics such as category of object and category of scene
  - factors related to context, including object-scene or attribute-object context

Approach

1. Gathering data, content labels and descriptions

2. Mapping from content to description

3. Exploring importance factors

4. Building and evaluating models to predict importance

Data

- **ImageCLEF Dataset**
  - Collection of 20K images covering various aspects of contemporary life, such as sports, cities, animals, people, and landscapes.
  - IAPR TC-12 Benchmark includes a free-text description for each image.
  - Each image is also segmented into constituent objects and labeled according to a set of (275) labels.

Data

- **UIUC Pascal Sentence Dataset**
  - Consists of 1K images sub-sampled from the Pascal Challenge.
  - 5 descriptions written by humans for each image.
  - Annotated with bounding box localizations for 20 object categories.

Collecting Content Labels

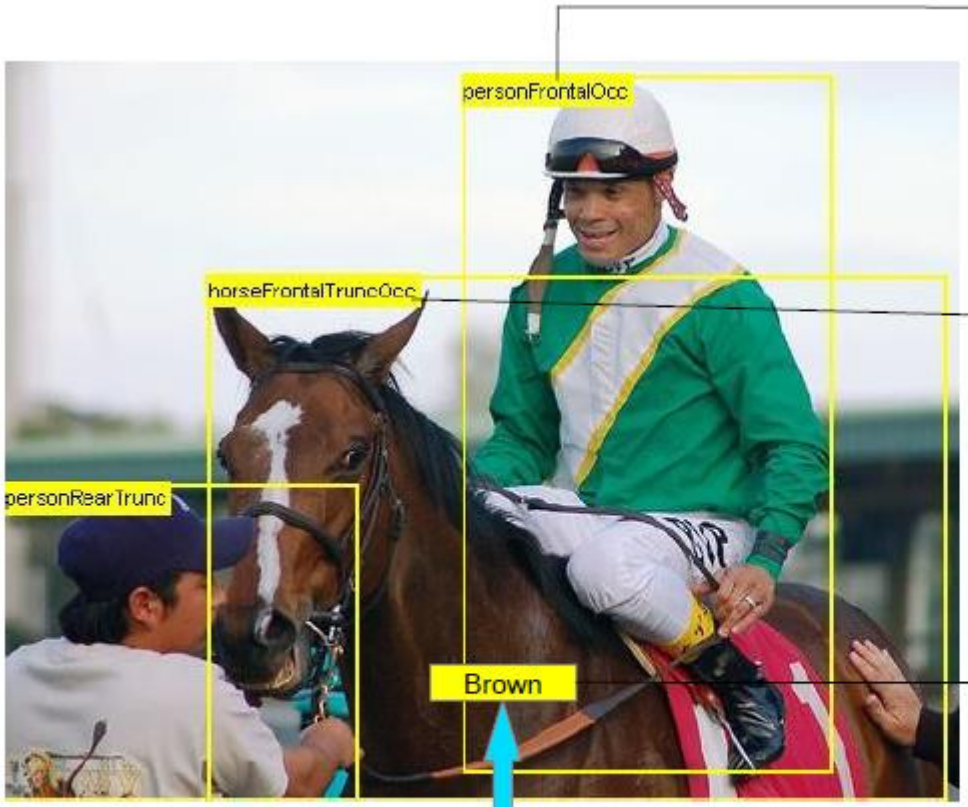
- Object labels
  - present in each of data collections
- Scene labels
- Attribute labels

Mechanical Turk is used for the UIUC dataset.

Approach

1. Gathering data, content labels and descriptions
2. **Mapping from content to description**
3. Exploring importance factors
4. Building and evaluating models to predict importance

Mapping from content to description



Human description:
A man is riding a brown horse

Amazon Mechanical Turk

Approach

1. Gathering data, content labels and descriptions
2. Mapping from content to description
3. **Exploring importance factors**
4. Building and evaluating models to predict importance

Exploring importance factors
Compositional Factors

Size

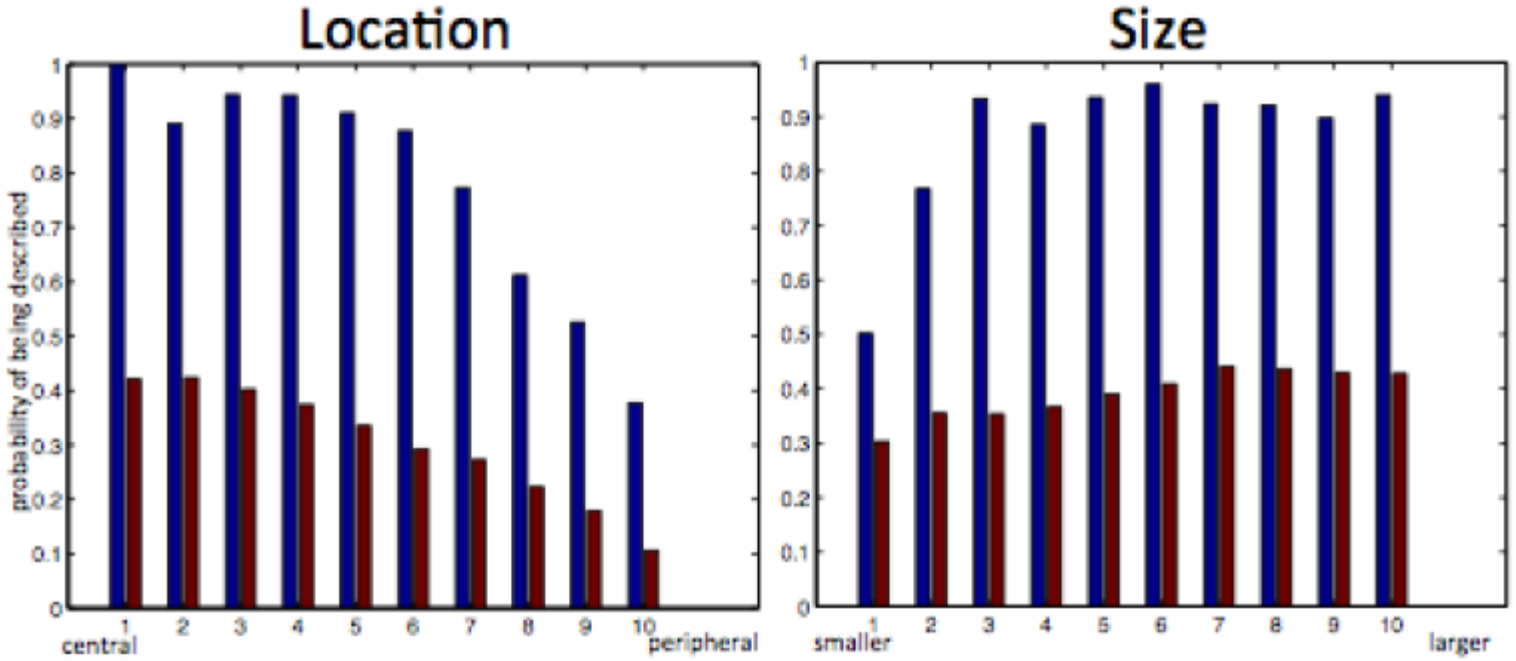
"A sail boat on the ocean."

Location

"Two men standing on beach."

- Size is measured as object size, normalized by image size.
- Location is measured as the distance from the image center to the object center of mass, normalized by image size.

Exploring importance factors
Compositional Factors



Objects further away from the image center are less likely to be mentioned (left). The bigger an object is, the more likely it is mentioned (right), unless it is very large. In general objects from ImageCLEF (red) are less likely to be described than objects from the UIUC (blue) dataset.

Exploring importance factors
Semantic Factors

- 2 kinds of semantic information
  - How the category of an object influences the probability that the object will be described
  - How the scene category of an image and strength of its depiction influences the probability that the scene type will be described

Exploring importance factors
Semantic Factors

Object Type

"Girl in the street"

Scene Type & Depiction Strength

"kitchen in house"

Exploring importance factors Semantic Factors-Object Type

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

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

Object	Prob	Object	Prob
horse	0.99	bus	0.80
sheep	0.99	motorbike	0.75
train	0.99	bicycle	0.69
cat	0.98	sofa	0.59
dog	0.96	dining table	0.56
aeroplane	0.97	tv/monitor	0.54
cow	0.95	car	0.43
bird	0.93	potted plant	0.26
boat	0.90	bottle	0.26
person	0.81	chair	0.26

Probability of being mentioned when present for various object categories (UIUC).

- Very unusual objects tend to be mentioned.
- Human pays attention to people inherently.
- Animate objects are much more likely to be mentioned.

Exploring importance factors Semantic Factors-Scene Type

office	airport	kitchen	dining room	field	living room	street	river	restaurant	sky	forest	mountain
0.29	0.13	0.36	0.21	0.16	0.13	0.18	0.1	0.28	0.18	0.0	0.07

Probability of description for each Scene Type.

- The scene category is much more likely to be mentioned for indoor scenes (avg. 0.25) than outdoor scenes (avg 0.12).

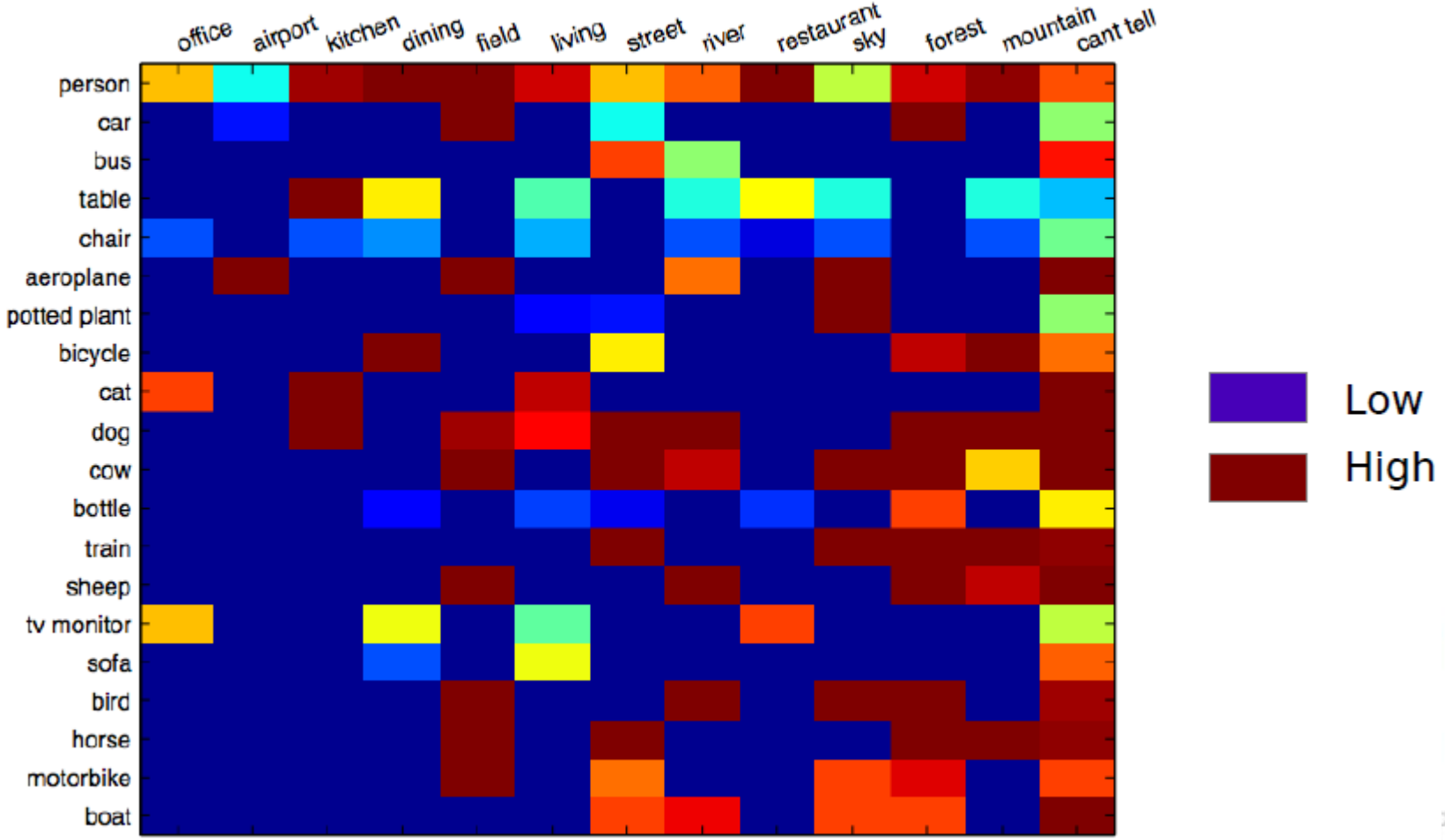
Rating	1	2	3	4	5
Prob	0.15	0.21	0.21	0.22	0.26

Probability of Scene term mentioned given Scene depiction strength (1 implies the scene type is very uncertain, and 5 implies a very obvious example of the scene type, as rated by human evaluators). Scenes are somewhat more likely to be described when users provide higher ratings.

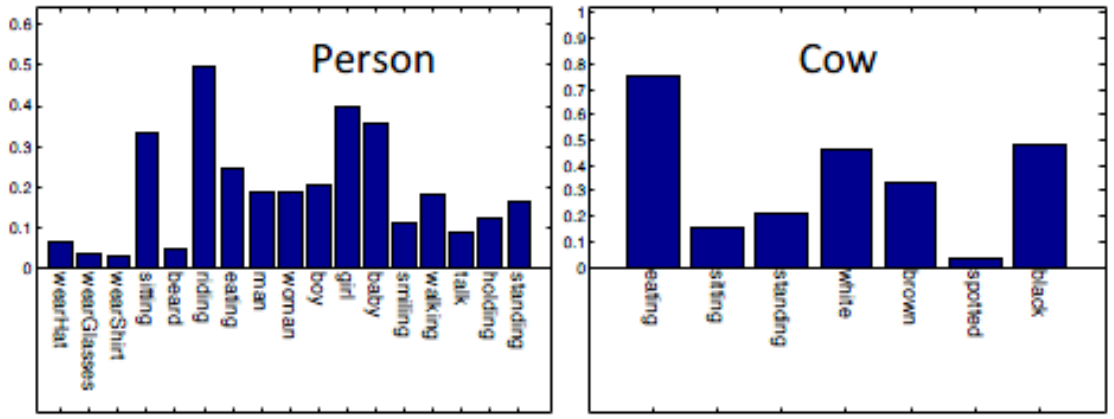
Exploring importance factors
Context Factors

- 2 kinds of contextual factors
  - Object-scene context
    - The probability of an object being described given that it occurs in a particular scene.
  - Attribute-object context
    - the probability of an attribute being described given that it occurs as a modifier for a particular object category.

Exploring importance factors
Context Factors
Object-Scene Context



Exploring importance factors
Context Factors
Attribute-Object Context

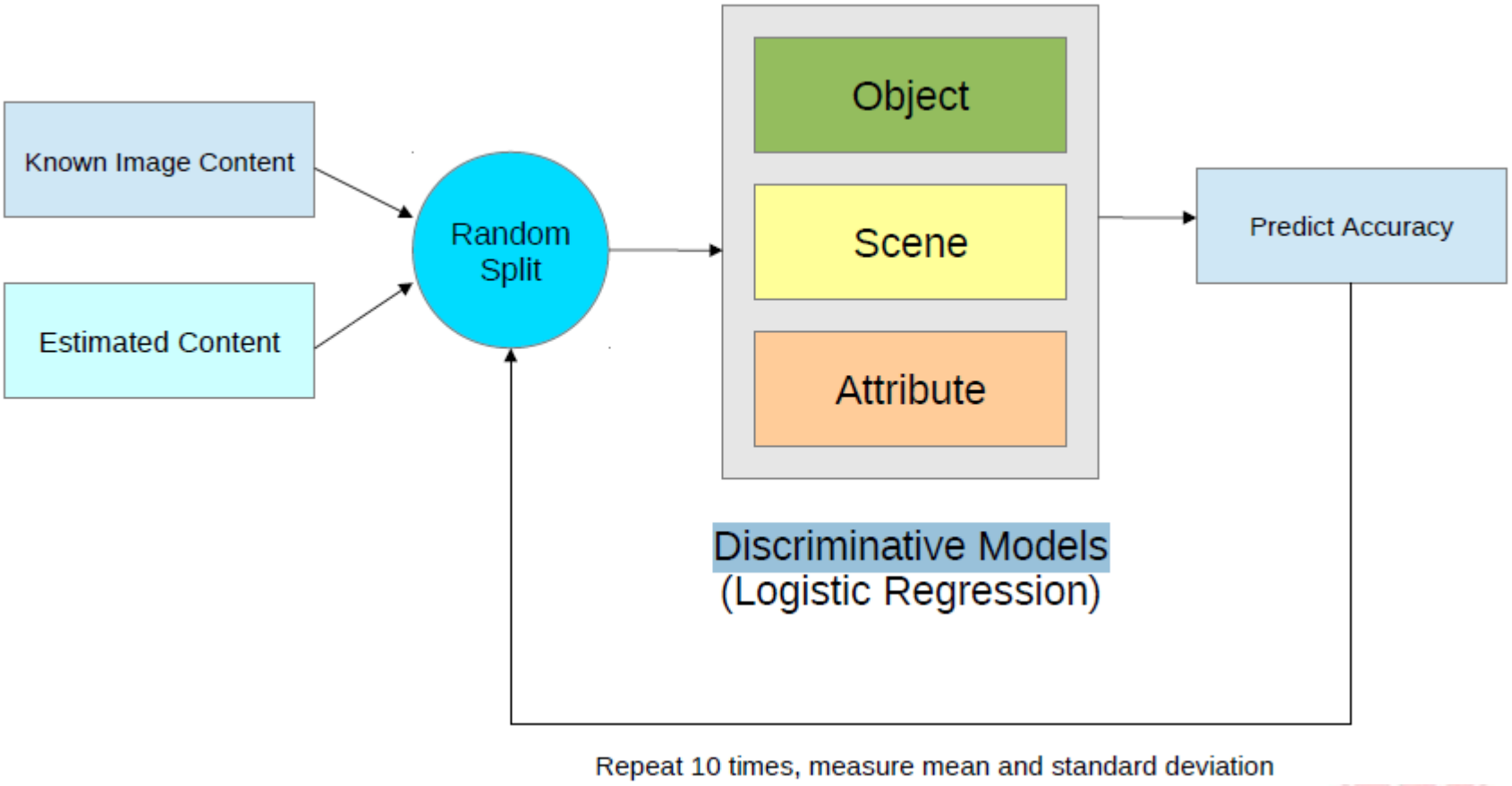


The impact of Attribute-Object context, showing the probability of an attribute being mentioned given that it occurs with a particular object.

Approach

1. Gathering data, content labels and descriptions
2. Mapping from content to description
3. Exploring importance factors
4. **Building and evaluating models to predict importance**

Predicting Importance



Predicting objects in sentences (ImageCLEF+UIUC)

Model	Features	Accuracy% (std)
Baseline (ImageCLEF)		57.5 (0.2)
Log Reg (ImageCLEF)	$K_o^s + K_o^l$	60.0 (0.1)
Log Reg (ImageCLEF)	K_o^c	68.0 (0.1)
Log Reg (ImageCLEF)	$K_o^c + K_o^s + K_o^l$	69.2 (1.4)
Baseline (UIUC-Kn)		69.7 (1.3)
Log Reg (UIUC-Kn)	$K_o^s + K_o^l$	69.9 (0.6)
Log Reg (UIUC-Kn)	K_o^c	79.8 (1.4)
Log Reg (UIUC-Kn)	$K_o^c + K_o^s + K_o^l$	82.0 (0.9)
Baseline (UIUC-Est)		76.5 (1.0)
Log Reg (UIUC-Est)	$E_o^s + E_o^l$	76.9 (1.1)
Log Reg (UIUC-Est)	E_o^c	78.9 (1.4)
Log Reg (UIUC-Est)	$E_o^c + E_o^s + E_o^l$	79.52 (1.2)

Predicting scenes in sentences (UIUC)

Model	Features	Accuracy% (std)
Baseline (UIUC-Kn)		86.0 (0.2)
Log Reg (UIUC-Kn)	K_s^c+K_s^r	96.6 (0.2)
Log Reg (UIUC-Est)	E_s^d	87.4 (1.3)

Predicting attributes in sentences (UIUC)

Model	Features	Accuracy% (std)
Baseline (UIUC-Kn)		96.3 (.01)
Log Reg (UIUC-Kn)	$K_a^c + K_o^c$	97.0 (.01)
Log Reg (UIUC-Est)	$E_a^d + E_o^c$	96.7 (.01)

Conclusion

- They have proposed several factors related to human perceived importance, including factors related to image composition, semantics, and context.
- They explore the impact of these factors individually on two large labeled datasets.
- Finally, they demonstrate discriminative methods to predict object, scene and attribute terms in descriptions given either known image content, or content estimated by state of the art visual recognition methods.

QUESTIONS?