Overview of the ImageCLEF 2012 Scalable Web Image Annotation Task

Mauricio Villegas and Roberto Paredes

{mvillegas,rparedes}@iti.upv.es

CLEF 2012

Rome, 19th of September 2012

Outline

Introduction

Image Annotation: Detecting concepts present in an image.

- Dogs
- Wheelchair
- Breeds
- Table
- Rural
- Grass
- Daytime
- Trees
- ...

Introduction – Motivation

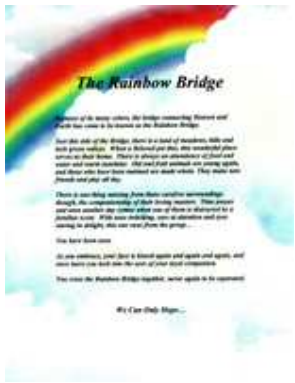
- The research on image annotation has mostly relied on manually labeled training data, for which crowdsourcing has become a common practice.
- Even though crowdsourcing has proved to be very useful, it is expensive and difficult to scale to a large amount of concepts.
- Millions of images and corresponding related text can be cheaply crawled from the Internet for practically any topic.
- With the aim of exploring possible complements or alternatives to the crowdsourcing approach by using Web data, we proposed this new task for ImageCLEF 2012.

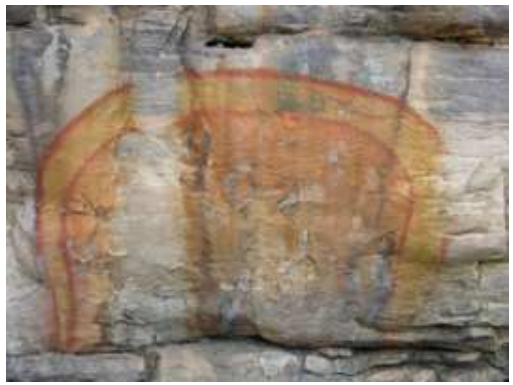
Introduction – Challenge

How to effectively use Web data for image annotation?

- The text in websites is noisy and the degree of relationship to the images varies greatly.
- The types of images also varies. Take for example images from a Web search query of “rainbow”:







Outline

Task Description – Subtasks

Subtask 1: Complementing Manually Labeled Data

- **Objective:** Try to use both automatically gathered Web data and labeled data to enhance the performance in comparison to using only the labeled data.
- **Training set:**
  - Web (250,000 images, unlabeled, textual features).
  - Flickr (15,000 images, labeled for 94 concepts).
- **Test set:** Flickr (10,000 images, labeled for same 94 concepts).
- **Submission:** Concept scores and which were annotated per image (max. 5 runs per group).

Task Description – Subtasks

Subtask 2: Scalable Concept Image Annotation

- **Objective:** Use only automatically gathered Web data and language resources to develop a concept scalable annotation system.
- **Training set:** Web (250,000 images, unlabeled, textual features).
- **Development set:** Web (1,000 images, labeled for 95 concepts).
- **Test set:** Web (2,000 images, labeled for 105 concepts).
- **Submission:** Concept scores and which were annotated per image (max. 5 runs per set per group).

Task Description – Web Training Dataset

- Web training dataset composed of 250,000 images, 7 visual features types and 4 textual feature types.
- Images found by querying Google, Bing and Yahoo using the words from the English dictionary.
- Precautions taken to avoid “message images”, duplicates and near-duplicates.
- To ease data download and handling by participants, the subset of 250,000 images was selected using 158 concepts (including the concepts for the task).

Task Description – Web Training Dataset

Visual Features:

Feature	BoW	Raw desc.
SIFT	5k dim. (780M) and 50k dim. (1.3G)	128 dim. (7G)
C-SIFT	5k dim. (680M) and 50k dim. (1.2G)	384 dim. (20G)
RGB-SIFT	5k dim. (760M) and 50k dim. (1.3G)	384 dim. (18G)
OPP-SIFT	5k dim. (630M) and 50k dim. (1.2G)	384 dim. (19G)
SURF	–	64 dim. (11G)
GIST	480 dim. (570M)	
Color Hist.	576 dim. (170M)	
Thumbnails	Max. 200 pixels high	

Task Description – Web Training Dataset

Textual Features:

1. Words used to find the images (3M).
2. Relative URLs of images in webpages (25M).

Dogs can tell size of another dog by listening to its growls

Washington, Dec 21 : A new study has shown that dogs can tell the size of another dog by listening to its growls.
Peter Pongracz and his team recruited 96 dogs of various breeds ...

```
<html>
<head>
 <title> Dogs can tell size of another dog by listen-
  ing to its growls | Science / Technology </title>
</head>
<body>
 <h2> Dogs can tell size of another dog by listening
  to its growls </h2>
 <img src="img/dogs.jpg" alt="dogs in the park" />
 <p> Washington, Dec 21 : A new study has shown that
  dogs can tell the size of another dog by listening
  to its growls. </p>
 <p> Peter Pongracz and his team recruited 96 dogs of
  various breeds ... </p>
</body>
</html>
```

3. Website text 1 (300M): ... to its growls . dogs in the park {X} . Washington . Dec 21 . A new study has ...
4. Website text 2 (110M): dogs 0.09 of 0.0422 by 0.0336 growls 0.33 to 0.0326 dog 0.0321 can 0.0309 size 0.0307 ...

Task Description – Web Training Dataset

Textual Features:

1. Words used to find the images (3M).
2. Relative URLs of images in webpages (25M).

Dogs can tell size of another dog by listening to its growls

Washington, Dec 21 : A new study has shown that dogs can tell the size of another dog by listening to its growls.
Peter Pongracz and his team recruited 96 dogs of various breeds ...

```
<html>
<head>
 <title> Dogs can tell size of another dog by listen-
  ing to its growls | Science / Technology </title>
</head>
<body>
 <h2> Dogs can tell size of another dog by listening
  to its growls </h2>
 <img src="img/dogs.jpg" alt="dogs in the park" />
 <p> Washington, Dec 21 : A new study has shown that
  dogs can tell the size of another dog by listening
  to its growls. </p>
 <p> Peter Pongracz and his team recruited 96 dogs of
  various breeds ... </p>
</body>
</html>
```

3. Website text 1 (300M): ... to its growls . dogs in the park {X} . Washington . Dec 21 . A new study has ...
4. Website text 2 (110M): dogs 0.09 of 0.0422 by 0.0336 growls 0.33 to 0.0326 dog 0.0321 can 0.0309 size 0.0307 ...

Outline

Subtask 1 – Participation

Received 16 runs from 3 groups (useful from only 2 groups).

- **KIDS-NUTN (National University of Tainan):**
  - Proposed a fusion of several visual features and textual.
  - For annotation, they tried Random Forests and Multiple Bernoulli Relevance Models.
  - Unclear how they handled the Web data.
- **ISI (University of Tokyo):**
  - Method focused on scalability.
  - Used a combination of several *SIFT features.
  - For annotation, they used their online learning method Passive-Aggressive with Averaged Pairwise Loss.
  - Tackled the Web data by labeling it using the appearance of concept words in the textual features.

Subtask 1 – Results

Best results[1] using only Flickr data:

	MAP (%)	MF_1 (%)
Random Baseline	10.3	10.0
ISI 1424	70.8	55.3
KIDS-NUTN 1451	57.9	45.4

Results[1] using both Web and Flickr data:

	MAP (%)	MF_1 (%)
ISI 1393	25.0	18.2
ISI 1398	24.7	18.1
ISI 1399	24.5	17.8
ISI 1400	24.1	17.5
KIDS-NUTN 1369	52.1	39.9
KIDS-NUTN 1370	53.8	39.7
KIDS-NUTN 1371	49.3	33.1
KIDS-NUTN 1372	52.8	40.0

[1] MAP (mean average precision) and MF_1 (mean F-measure) computed per image.

Subtask 2 – Participation

Received 10 runs from 1 group.

- **ISI (University of Tokyo):**
  - Method focused on scalability.
  - Used a combination of several *SIFT features.
  - For annotation, they used their online learning method Passive-Aggressive with Averaged Pairwise Loss.
  - Tackled the Web data by labeling it using the appearance of concept words in the textual features.

Subtask 2 – Results

Results[1] for test set:

	MAP (%)	MF_1 (%)
Random Baseline	6.7	5.5
Co-occurrence Baseline	22.1	17.1
ISI 1407	31.5	24.6
ISI 1408	32.2	25.1
ISI 1411	32.4	25.2
ISI 1412	32.3	25.4
ISI 1415	32.1	24.9

[1] MAP (mean average precision) and MF_1 (mean F-measure) computed per image.

Subtask 2 – Results

ISI results[2] for Flickr ann. subtask:

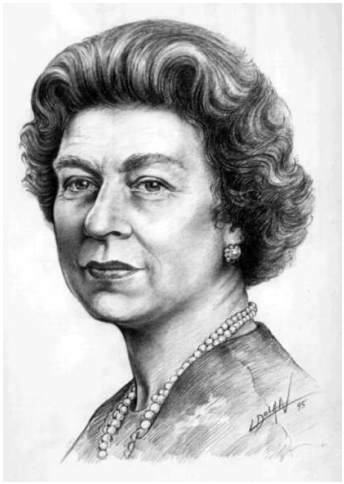
Concept	F_1 (%)
none	87.2
noblur	82.7
dog	72.2
fireworks	66.7
flower	66.2
partialblur	64.8
fooddrink	62.3
adult	61.2
one	59.3
female	58.9
outdoor	58.8
tree	58.1

ISI results[2] for Web subtask 2:

Concept	F_1 (%)
fireworks	70.3
pencil	69.2
stars	64.6
sunrise/sunset	56.5
drawing/diagram	56.4
galaxy	50.0
space	48.3
newspaper	46.2
lightning	45.2
forest	43.0
pool	42.6
fire	42.4

[2] F_1 (F-measure) computed per concept.

Subtask 2 – Example annotations

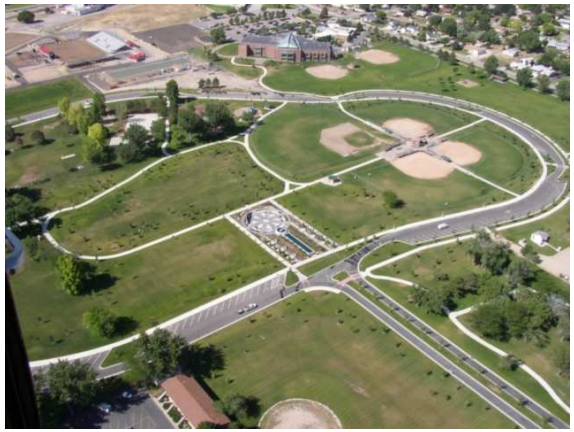


annotations:
person/people
drawing/diagram
child
pencil
baby

textbfmissed:
–

annotations:
person/people
garden/park
horse
sign
sports

missed:
–



annotations:
water
aerial
garden/park
road
grass

missed:
building
tree

Outline

Conclusions and Future Work

- Participation was disappointingly low, making it hard to draw good conclusions.
- Subtask 1:
  - None of the participants were able to take advantage of the Web data.
- Subtask 2:
  - The ISI system obtained a considerable better performance than the provided baselines.
  - The processing of the textual data of ISI is rather simple, so possibly there is much room for improvement.
  - For some concepts, the performance is relatively good, indicating that the Web data can be quite useful.

Conclusions and Future Work

- Results in subtask 2 were interesting, thus repeating the task with more participants would be desirable.

- However, we would like to know:
  - Why so few groups participated?
  - Right now, is there enough interest to repeat the task?
  - Suggestions?

Thank you for your attention !

Questions ?