Overview of the ImageCLEF 2014 Scalable Concept Image Annotation Task

Mauricio Villegas and Roberto Paredes

mauvilsa@upv.es rparedes@prhlt.upv.es

Slides available at:
http://mvillegas.info/pub/Villegas14_CLEF_Annotation-Overview_presentation.pdf

Outline

Introduction

- **Automatic image annotation** is the process by which a computer assigns to an image, metadata that describes its content.
- In this work the metadata considered is only the presence or absence of concepts in the images, e.g.

→ Dog
→ Table
→ Rural
→ Grass
→ Daytime
→ Tree
→ ...

Introduction – Motivation

- Image annotation research has mostly relied on manually labeled training data. Examples of available datasets are:
  - **ImageNet:** $\approx$1.2M images, 1000 concepts, but only one concept per image.
  - **NUS-WIDE:** $\approx$269k images, multiple concepts per image, but only 81 concepts.
- Even though crowdsourcing has proved to be very useful, it is expensive and difficult to scale to a large amount of concepts.

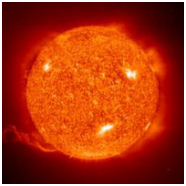
Are there alternatives that do scale concept-wise?

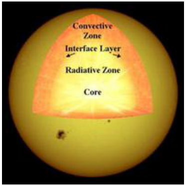
- Millions of images and corresponding related text can be cheaply crawled from the Internet for practically any topic.

Introduction – Motivation

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 search query of “sun”:





Outline

Task description

- **Objective:** To use only automatically gathered data for developing concept scalable image annotation systems.
  - Any data could be used as training, except for hand labeled images, e.g. crawled data, WordNet, dictionaries, stemmers, etc.

- **Participants were provided with:**
  - Crawled dataset (500,000 images and respective webpages).
  - Development set (1,940 samples, labeled for 107 concepts).
  - Implementation of a baseline system and code for computing the performance measures.

Task description

- **Test set:** 7,291 samples, the participants had to label them for 207 concepts, 100 unseen in development (max. 10 runs could be submitted per group).
  Divided into 4 subsets with different concept lists:
  - Previous ImageCLEF (116 concepts).
  - Related to animals (52 concepts).
  - Related to foods (41 concepts).
  - Complete list (207 concepts).

- **Concepts:** Defined as WordNet synsets and for most of them also a Wikipedia article.

Task description – Lines of work

In contrast to traditional image annotation tasks, the proposed one involves more lines of work:

- Which representation to use for the images (visual features).
- How to use unsupervised web data as training.
  - Automatically assign concepts to the images using the textual data?
  - How to pre-process and clean the textual data?
  - Use other resources:
    - Ontologies
    - Language dictionaries
    - Automatic translation
- Which method to use for modelling the concepts.
- What strategy to use for deciding how many and which concepts are assigned to an image.

Task description – Training dataset

- Web training dataset[1] composed of 500,000 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.
- Subset of images selected using only the used concepts to ease data download and handling by participants.

[1] Dataset available at `http://risenet.prhlt.upv.es/webupv-datasets`

Task description – Training dataset

Visual Features:

Feature	Dimensionality	Training data size
Thumbnails	Max. 200 pixels high	30 GB
GIST	480	1.6 GB
Color Hist.	576	330 MB
GETLF	256	60 MB
SIFT	5,000 BoW	1.5 GB
C-SIFT	5,000 BoW	1.3 GB
RGB-SIFT	5,000 BoW	1.5 GB
OPP-SIFT	5,000 BoW	1.4 GB

Task description – Training dataset

Textual Features:

1. Words used to find the images (5MB).
2. Relative URLs of images in webpages (50MB).

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. Image webpages as valid XML (4.7GB).
4. Webpage text (218M): 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 – Training dataset

Textual Features:

1. Words used to find the images (5MB).
2. Relative URLs of images in webpages (50MB).

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. Image webpages as valid XML (4.7GB).
4. Webpage text (218M): 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

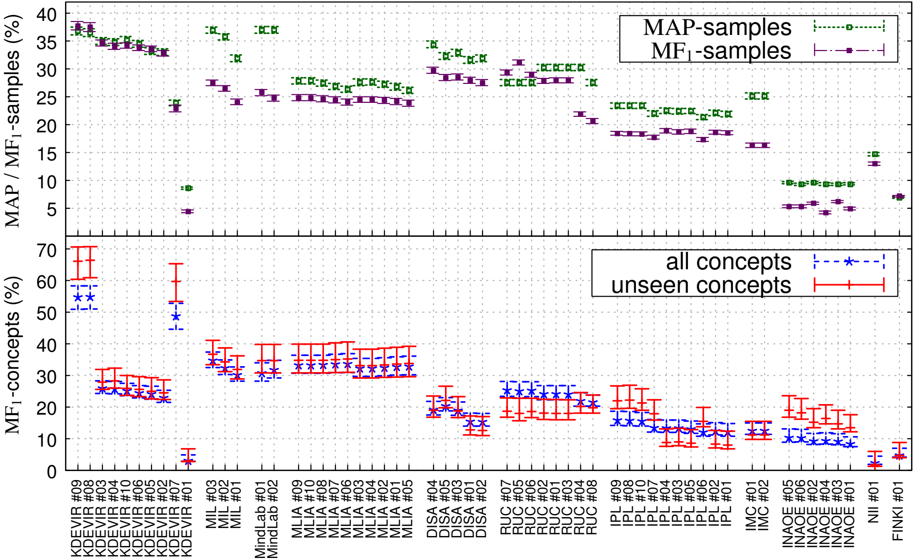
Evaluation – Participation

Groups that registered	43
Total submitted runs	58
Groups that participated	11
Groups that submitted working notes paper	9
Data downloads	> 100

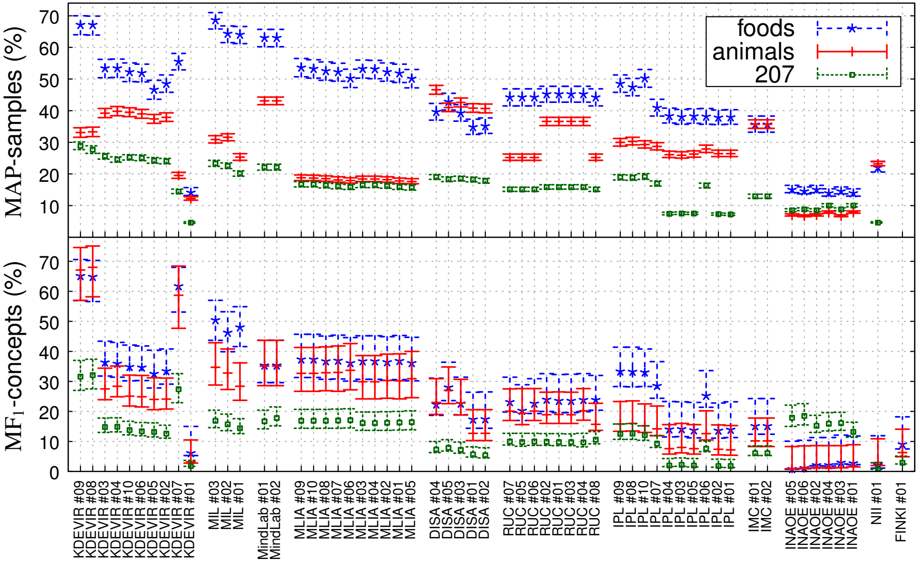
Participants:

- **DISA:** Laboratory of Data Intensive Systems and Applications of the Masaryk University (Brno, Czech Republic).
- **IPL:** Information Processing Laboraroty of the Athens University of Economics and Business (Athens, Greece).
- **KDEVIR:** Computer Science and Engineering department of the Toyohashi University of Technology (Aichi, Japan).
- **MIL:** Machine Intelligence Lab of the University of Tokyo (Tokyo, Japan).
- **MindLab:** Machine learning, perception and discovery Lab from the Universidad Nacional de Colombia (Bogotá, Colombia).
- **MLIA:** Department of Advanced Information Technology of the Kyushu University (Fukuoka, Japan).
- **RUC:** School of Information of the Renmin University of China (Beijing, China).
- **FINKI:** Faculty of Computer Science and Engineering of the Ss. Cyril and Methodius University (Skopje, Republic of Macedonia).
- **IMC:** Institute of Media Computing of the Fudan University (Shanghai, China).
- **INAOE:** Instituto Nacional de Astrofísica, Óptica y Electrónica (Puebla, Mexico).
- **NII:** National Institute of Informatics (Tokyo, Japan).

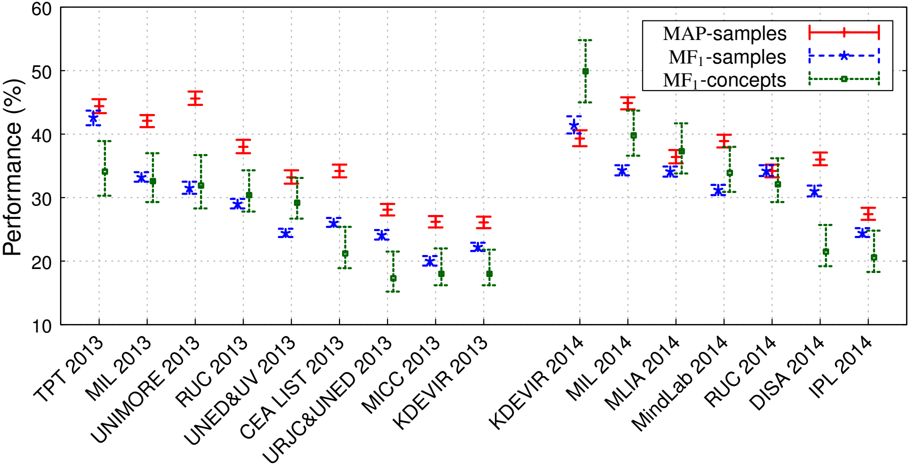
Evaluation – Results (complete test set)



Evaluation – Results (subsets)



Evaluation – Results (comparison with 2013)



Evaluation – Details of some submitted systems

System	Visual Feat.	Training Data Processing	Annotation Technique
KDEVIR run #9	Provided by organizers	- Ontology built per concept using WordNet and Wikipedia - Training positive and negative samples selected by exploiting ontologies	- Multiple SVMs per concept with context dependent kernel - Annotation of top-k concepts exploiting ontologies
MIL run #3	Fisher Vectors & ImageNet CNN	- Extract webpage title, image tag attributes and singularize nouns - Label training images by appearance of concept (synonyms and hyponyms)	- Linear multilabel classifier learned by PAAPL - Annotation of the 4% top scored concepts
MindLab run #1	ImageNet CNN	- Extract webpage words, stopword removal and stemming	- A logistic regression (soft-max) model - Annotation based on threshold
MLIA run #9	Provided by organizers	- Provided webpage features, stopword removal and stemming - Concepts assigned to training images by appearance of synonyms, filtered using Overfeat	- One SVM per concept, F-measure cross-validation - Annotation based on SVM classification decision

Evaluation – Details of some submitted systems

System	Visual Feat.	Training Data Processing	Annotation Technique
KDEVIR run #9	**Provided by organizers**	- Ontology built per concept using WordNet and Wikipedia - Training positive and negative samples selected by exploiting ontologies	- Multiple SVMs per concept with **context dependent kernel** - Annotation of top-k concepts exploiting ontologies
MIL run #3	Fisher Vectors & ImageNet CNN	- Extract webpage title, image tag attributes and singularize nouns - Label training images by appearance of concept (synonyms and hyponyms)	- Linear multilabel classifier learned by PAAPL - Annotation of the 4% top scored concepts
MindLab run #1	ImageNet CNN	- Extract webpage words, stopword removal and stemming	- A logistic regression (soft-max) model - Annotation based on threshold
MLIA run #9	Provided by organizers	- Provided webpage features, stopword removal and stemming - Concepts assigned to training images by appearance of synonyms, filtered using Overfeat	- One SVM per concept, F-measure cross-validation - Annotation based on SVM classification decision

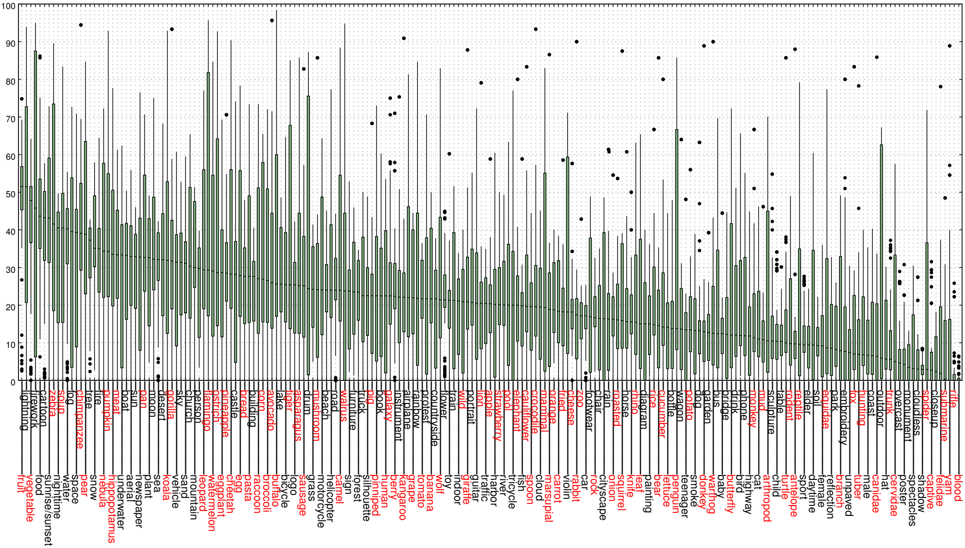
Evaluation – Details of some submitted systems

System	Visual Feat.	Training Data Processing	Annotation Technique
KDEVIR run #9	Provided by organizers	- **Ontology built per concept** using WordNet and Wikipedia - **Training** positive and negative samples selected by **exploiting ontologies**	- Multiple SVMs per concept with context dependent kernel - **Annotation** of top-k concepts **exploiting ontologies**
MIL run #3	Fisher Vectors & ImageNet CNN	- Extract webpage title, image tag attributes and singularize nouns - Label training images by appearance of concept (synonyms and hyponyms)	- Linear multilabel classifier learned by PAAPL - Annotation of the 4% top scored concepts
MindLab run #1	ImageNet CNN	- Extract webpage words, stopword removal and stemming	- A logistic regression (soft-max) model - Annotation based on threshold
MLIA run #9	Provided by organizers	- Provided webpage features, stopword removal and stemming - Concepts assigned to training images by appearance of synonyms, filtered using Overfeat	- One SVM per concept, F-measure cross-validation - Annotation based on SVM classification decision

Evaluation – Details of some submitted systems

System	Visual Feat.	Training Data Processing	Annotation Technique
KDEVIR run #9	Provided by organizers	- Ontology built per concept using WordNet and Wikipedia - Training positive and negative samples selected by exploiting ontologies	- Multiple SVMs per concept with context dependent kernel - Annotation of top-k concepts exploiting ontologies
MIL run #3	Fisher Vectors & **ImageNet CNN**	- Extract webpage title, image tag attributes and singularize nouns - Label training images by appearance of concept (synonyms and hyponyms)	- Linear multilabel classifier learned by PAAPL - Annotation of the 4% top scored concepts
MindLab run #1	**ImageNet CNN**	- Extract webpage words, stopword removal and stemming	- A logistic regression (soft-max) model - Annotation based on threshold
MLIA run #9	Provided by organizers	- Provided webpage features, stopword removal and stemming - Concepts assigned to training images by appearance of synonyms, filtered using Overfeat	- One SVM per concept, F-measure cross-validation - Annotation based on SVM classification decision

Evaluation – Concept F_1 boxplots for all runs



Outline

Conclusions

- Participation was excellent, and the teams presented diverse approaches to address the proposed challenge.
- The results indicate that the web data can be effectively used for training practical and scalable annotation systems.
- The performance of the systems improved with respect to last year.
- Due to the larger number of unseen concepts, results had narrower confidence intervals, so it made the comparison the systems more conclusive.
- The winning team was KDEVIR. Its success is possibly due to classifier that considers contextual information and usage of concept ontologies both in training and test.

Future work

- The task will hopefully continue for CLEF 2015, pending the notification of acceptance of ImageCLEF 2015 lab due the 19th of September.
- Several modifications to the task:
  - Localisation within the images.
  - Description sentence generation.
- New organisers:
  - Andrew Gilbert
  - Luca Piras
  - The ViSen consortium

Thank you for your attention!

Questions? Comments?