SAPIENZA
UNIVERSITÀ DI ROMA

Overview of the ImageCLEF 2014 Domain Adaptation Task

Barbara Caputo[1], Novi Patricia[2]

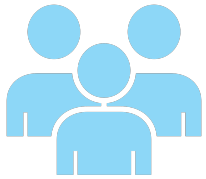
[1]University of Rome La Sapienza, Italy

[2]Idiap Research Institute, Switzerland

DomainAdaptation@ImageCLEF2014: Outline

- Introduction & Motivation
- DomainAdaptation@ImageCLEF2014: The Task
  - ❑ Challenges
  - ❑ Data & Features
- Participants and Results
- Analysis
- Conclusions

Introduction & Motivation

- The amount of annotated image collections is increased over the last years **Amazon Mechanical Turk** 
- Leveraging over and across large data sources is a challenging research
- For a given task, training on a dataset (e.g. PASCAL VOC) and testing on another (e.g. ImageNet) produces a poor result, although learning same categories

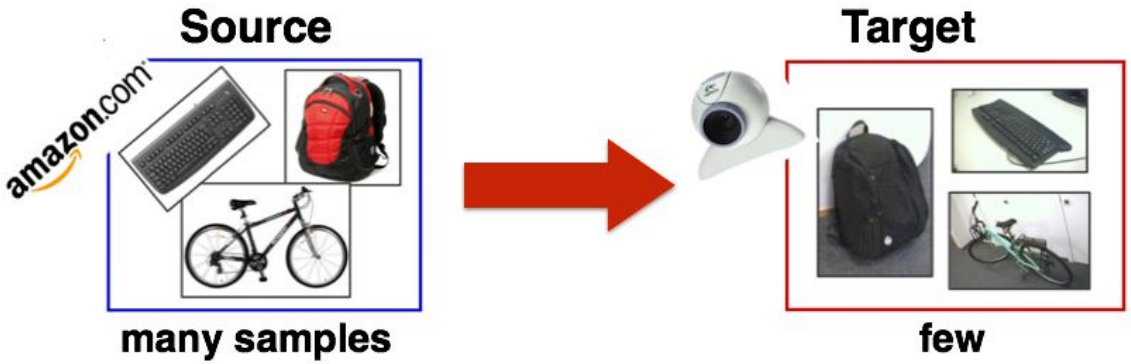
Introduction & Motivation

- The problem to generalize object categorization across databases is known as the ***domain adaptation*** challenge
- A source domain (S) has a large amount of labeled images
- A target domain (T) has different image set, and few or no labeled samples
- Formally:

  two domains ***differ*** ➔ probability distributions are ***different***

$$P_S(x, y) \neq P_T(x, y)$$

Domain Adaptation Task

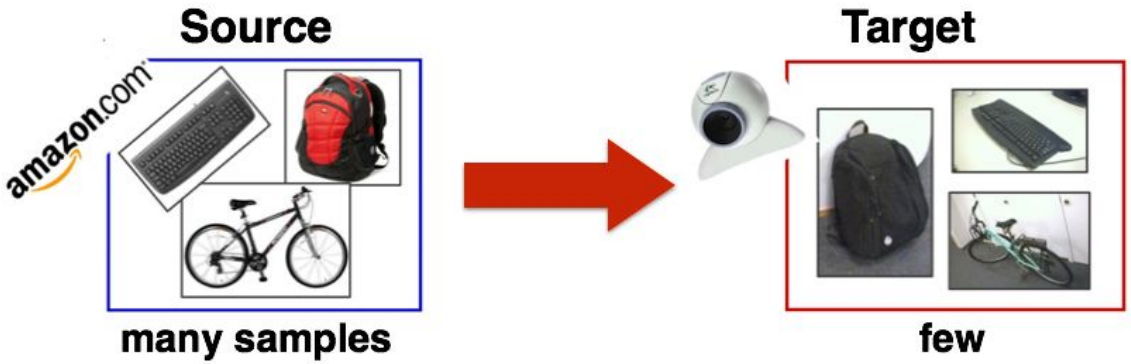


Saenko et al, ECCV 2010

Current research focuses on:

- ❑ The source consists of one or maximum two databases
- ❑ The labels on both domain are the same
- ❑ The number of annotated training data for target domain are limited

Domain Adaptation Task



Saenko et al., ECCV 2010

Current research focuses on:

Not realistic settings !!

- ❑ The source consists of one or maximum two databases
- ❑ The labels on both domain are the same
- ❑ The number of annotated training data for target domain are limited

DomainAdaptation@ImageCLEF2014: Challenges

- General settings existed in the community:
  - ❑ **One** source, one target: Gong et al., CVPR 2012
  - ❑ At most **two** sources, one target: Saenko et al., ECCV 2010
- DomainAdaptation@ImageCLEF2014 (1st edition): focus on **the number of sources**
  - ❑ 4 sources
  - ❑ existing available resources
  - ❑ semi-supervised setting (limited samples)

Data

- Publicly available databases:
  - ❑ Caltech-256: 256 categories, 30607 images
  - ❑ ImageNet (ILSVRC2012): WordNet hierarchy, 500 images/node
  - ❑ PASCAL VOC2012: 20 classes
  - ❑ Bing: 256 categories, collected using Bing search engine
  - ❑ SUN: scene understanding, 899 categories, 130.519 images
- 12 common classes: aeroplane, bike, bird, boat, bottle, bus, car, dog, horse, monitor, motorbike, and people

Evaluation Metrics

Correctly classified image	Misclassified image
+1 points	+0 points

Source

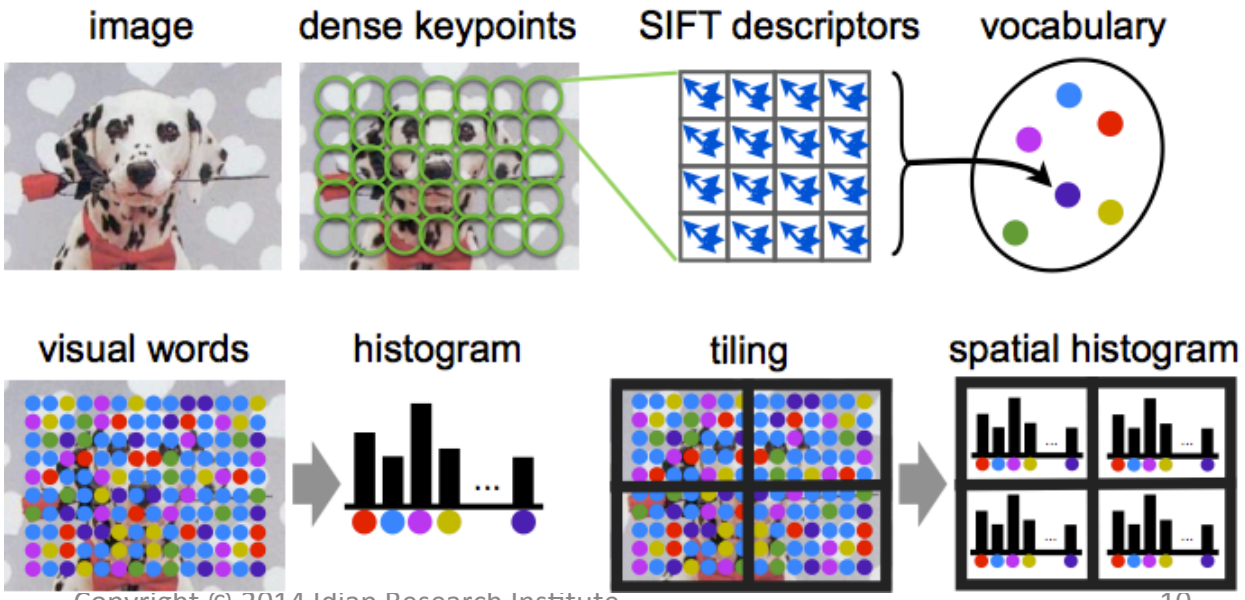
Train: 50 images/class

Target

Train: 5 images/class
Test: 50 images/class

Features Extraction

- Dense SIFT descriptors
- Grid of 128 pixels
- Quantized into 256 visual words ➔1024 dimension
- www.vlfeat.org



Participants

- 19 groups registered
- 3 groups submitted runs
  - ❑ XRCE: combine several heterogeneous domain adaptation methods, use majority voting to improve overall accuracy
  - ❑ Hubert Curien Lab: no working notes
  - ❑ Idiap Research Institute: treat source classifier as experts, then combine the output confidence with high-level cue integration

Results

Total score

Rank	Group	Score
1	XRCE	228
2	Hubert Curien Lab Group	158
3	Idiap	45

Score per class

class	Score XRCE	Score Hubert Curien	Score Idiap
aereoplane	41	36	3
bike	12	7	1
bird	15	15	0
boat	18	5	4
bottle	20	25	3
bus	23	10	6
car	17	13	7
dog	8	8	3
horse	17	6	2
monitor	28	15	3
motorbike	12	7	3
people	17	11	10

Analysis

- XRCE shows that the current methods are not able to address effectively the problem of leveraging over multiple sources
- Ensemble methods appear instead to be a viable option in the realistic condition, whether:
  - combine the output of various DA algorithms, or
  - combine several sources output confidence
- Enough interest from the participants (19 groups registered)
- Pre-computed features did not allow flexibility

Wrapping up

- The 1st edition of DA Task focused on the problem of building a classifier in a **target domain** while leveraging over 4 sources
- Ensemble learning based method is able to tackle the problem

Next year competition:

- Provide raw images ➔ a wider generality of approaches
- Multiple sources, possibly by augmenting the number of classes
- Partial overlap of classes between domains