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*

\[ P_S(x, y) \neq P_T(x, y) \]
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
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

Not realistic settings!!

Saenko et al., ECCV 2010

Source

many samples

Target

few

Copyright © 2014 Idiap Research Institute
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

<table>
<thead>
<tr>
<th>Correctly classified image</th>
<th>Misclassified image</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1 points</td>
<td>+0 points</td>
</tr>
</tbody>
</table>

Source: IMAGE\textsc{net}, CALTECH 256, PASCAL\textsc{2}

Target: SUN

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

Train: 50 images/class

Copyright © 2014 Idiap Research Institute
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

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>XRCE</td>
<td>228</td>
</tr>
<tr>
<td>2</td>
<td>Hubert Curien Lab Group</td>
<td>158</td>
</tr>
<tr>
<td>3</td>
<td>Idiap</td>
<td>45</td>
</tr>
</tbody>
</table>

### Score per class

<table>
<thead>
<tr>
<th>class</th>
<th>Score XRCE</th>
<th>Score Hubert Curien</th>
<th>Score Idiap</th>
</tr>
</thead>
<tbody>
<tr>
<td>aeroplane</td>
<td>41</td>
<td>36</td>
<td>3</td>
</tr>
<tr>
<td>bike</td>
<td>12</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>bird</td>
<td>15</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>boat</td>
<td>18</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>bottle</td>
<td>20</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>bus</td>
<td>23</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>car</td>
<td>17</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>dog</td>
<td>8</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>horse</td>
<td>17</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>monitor</td>
<td>28</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>motorbike</td>
<td>12</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>people</td>
<td>17</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>
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