Diverse Concept-Level Features for Multi-Object Classification

Youssef Tamaazousti, Hervé Le Borgne and Céline Hudelot

youssef.tamaazousti@cea.fr
Image Classification

- **OFFLINE**
  - Labels
  - Training database
  - Image Description
  - Model training

- **ONLINE**
  - Test image
  - Image Description
  - Comparator
    - Bear
    - Wood
    - Grass
Image Classification

OFFLINE

Labels

Training database

Image Description

Model training

ONLINE

Test image

Image Description

comparator

• Bear
• Wood
• Grass
Image description

- Low/Mid-Level Features
  - Image described in terms of contours and shapes

- Semantic Features
  - Image described in terms of semantic concepts
Semantic Features

- Torresani et al., 2010 – Li et al., 2010
  - Describe images in terms of outputs of concept-detectors
  - Each value is associated to a humanly-understandable word
Sparsification

- Wang et al., 2010 – Ginsca et al., 2015
  - Keep only the K highest values of the vector and set all others to zero
# Positioning

## Sparsification

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mid-level</strong></td>
<td>Csurska et al., 2004 (Bag of Visual Words)</td>
<td>Wang et al., CVPR 2010</td>
</tr>
<tr>
<td></td>
<td>Perronnin et al., 2007 (Fischer Kernels)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Krizhevsky et al., NIPS 2012 (Fully-connected layers of CNNs)</td>
<td></td>
</tr>
<tr>
<td><strong>Semantic</strong></td>
<td>Torresani et al., CVPR 2010</td>
<td>Ginsca et al., MMM 2015</td>
</tr>
<tr>
<td></td>
<td>Li et al., NIPS 2010</td>
<td>Tamaazousti et al., ICMR 2016 (Ours)</td>
</tr>
<tr>
<td></td>
<td>Bergamo et al., CVPR 2012</td>
<td></td>
</tr>
</tbody>
</table>
Classifcaticin with Semantic Features

- Object classification
  - Without sparsification
    - No missing information but noisy values (not good)
  - With sparsification
    - No missing information (good)
Classificatin with Semantic Features

- Multi-Object classification
  - Without sparsification
    - No missing information but noisy values (not good)
  - With sparsification
    - Missing information (not good)
Problem

• Typical problematic case
  • Image with multiple objects

• Observation
  • When the concept of the largest object is activated, a set of its annex concepts is also activated

• Why are we losing information?
  • Naive sparsification
    • Would select one principal concept and its annex concepts
    • Other principal concepts could be set to zero
Usual formalism

- **Sparsification** [Wang et al., 2010, Ginsca et al., 2015]
  - Principle
    - Set to zero « some » values of the vector
  - Objective
    - Keep the **good concepts** and delete the **bad ones**

- **Usual definition**
  - Good concepts = **highest** values
  - Bad concepts = all others (**lowest** values)
Proposed formalism

• Proposed definition
  • Good concepts = principal concepts and their annex concepts (not necessarily the highest values)

  • Bad concepts = all others (not necessarily the lowest values)

• Questions
  1. How to get the good concepts?
  2. What are the good concepts?
1. How to get the good concepts?

• Get the good concepts is a hard problem!

• Bergamo et al., 2012 (Bottom-up)
  • Get generic concepts (good concepts) using unsupervised clustering (hard)
  • Bottom-up: Low-level errors are propagated to upper concepts → limited performances

• Our proposal (Top-Down)
  • Get the good concepts using largely available Human Knowledge databases (hierarchies, human-categorization rules, databases, etc.)
2. What are the good concepts?

- Inspired by Psychological studies

- Rosch, 1978 - Jolicoeur et al., 1984
  - Different levels of good concept in Human minds
  - The concepts mostly known and used by Humans are
    - Superordinate: vehicle
    - Basic-level: car
    - Subordinate: ford mustang
### Observations

<table>
<thead>
<tr>
<th></th>
<th>Number of concepts-detectors</th>
<th>Range of values of concept-detectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superordinate</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Basic-level</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Subordinate</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>
Proposed approach

- Concept-detectors
  - Superordinate
    - Semantic process ➔ High range of values
  - Basic-level
    - Visual process
  - Subordinate
    - Visual process + reduction of number of concepts ➔ Low number of concepts

<table>
<thead>
<tr>
<th></th>
<th>Number of concepts</th>
<th>Range of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superordinate</td>
<td>Low</td>
<td>Low ➔ High</td>
</tr>
<tr>
<td>Basic-level</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Subordinate</td>
<td>High ➔ Low</td>
<td>High</td>
</tr>
</tbody>
</table>
Proposed approach

- S.O.T.A semantic feature

\[ \mathcal{F} = \begin{bmatrix} * & \cdots & * & 0 & \cdots & 0 \end{bmatrix} \]

- Our final semantic feature (D-CL)

\[ \mathcal{F} = \begin{bmatrix} * & \cdots & * & * & \cdots & * & \cdots & 0 \end{bmatrix} \]

**Superordinate**  | **Basic-level**  | **Subordinate**

Semantic process: semantic classifiers  |  Visual process: binary classifiers
In practice

- Hard to set the list of *superordinate*, *basic-level*, and *subordinate* concepts.
Experimental Protocol

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Pascal VOC 07</th>
<th>Pascal VOC 12</th>
<th>Nus-Wide Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of multi-label</td>
<td>45%</td>
<td>30%</td>
<td>20%</td>
</tr>
</tbody>
</table>

- **Evaluation metric**
  - mean Average Precision (mAP)

- **Pascal VOC 07**
  - Train/val: 5k images - Test: 5k images

- **Pascal VOC 12**
  - Train/val: 10k images - Test: 10k images

- **Nus-Wide Object**
  - Train/val: 20k images - Test: 15k images
## Multi-Object Classification Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Nus-Wide Object (20%)</th>
<th>Pascal VOC 2007 (45%)</th>
<th>Pascal VOC 2012 (30%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li <em>et al.</em>, 2010</td>
<td>n.a</td>
<td>45.2</td>
<td>n.a</td>
</tr>
<tr>
<td>Torresani <em>et al.</em>, 2010</td>
<td>n.a</td>
<td>43.8</td>
<td>n.a</td>
</tr>
<tr>
<td>Torresani <em>et al.</em> (reimpl.)</td>
<td>70.3</td>
<td>82.4</td>
<td>81.7</td>
</tr>
<tr>
<td>Bergamo <em>et al.</em>, 2011</td>
<td>n.a</td>
<td>43.7</td>
<td>n.a</td>
</tr>
<tr>
<td>Bergamo <em>et al.</em>, 2012</td>
<td>36.5</td>
<td>53.2</td>
<td>49.3</td>
</tr>
<tr>
<td>Simonyan <em>et al.</em>, 2015</td>
<td>67.3</td>
<td>77.4</td>
<td>77.2</td>
</tr>
<tr>
<td>Ginsca <em>et al.</em>, 2015</td>
<td>74.7</td>
<td>82.8</td>
<td>81.7</td>
</tr>
<tr>
<td>D-CL (ours)</td>
<td>76.0</td>
<td>85.1</td>
<td>83.0</td>
</tr>
</tbody>
</table>

**Naive sparsification**

**Without sparsification**
Conclusions

• **Novelty:**
  - New semantic image-representation
  - New formalism of sparsification
  - New sparsification process based on Human-cognition

• **Results:**
  - Multi-object classification
    - 3 publicly available benchmarks
    - +2 points of mAP compared to the best state-of-the-art semantic features
Thank you (questions ?)