BUPT-MCPRL at Trecvid2014
Instance Search Task

Wenhui Jiang (jiang1st@bupt.edu.cn)
Zhicheng Zhao, Qi Chen, Jinlong Zhao, Yuhui Huang,
Xiang Zhao, Lanbo Li, Yanyun Zhao, Fei Su, Anni Cai

MCPR Lab
Beijing University of Posts and Telecommunications
Our submission

- BOW baseline + CNN as global feature: 22.7%
  CNN as global feature boosts the performance by about 3% (estimated in INS2013).

- BOW baseline + Query expansion + CNN as global feature: 22.1%
  That’s not normal. We are investigating on it.

- BOW baseline + Localized CNN search: 21.6%
  Localized CNN search boosts the performance by about 0.5%.

- Interactive Run: BOW baseline + Query expansion (Interactive): 23.8%
Brief introduction

- **Reference Dataset**
  - 470K shots
  - 2 key frames per second
  - Max pooling for shot score

- **Query Images**
  - Average pooling for query score

- **Feature Model**
  - Bag-of-words
  - Convolutional neural networks
System Overview
BOW Highlights

• Three kinds of local features + BOW framework
  + ≈9% mAP

• Contextual weighting
  + ≈3% mAP

• Burstiness
  + ≈2% mAP
Three kinds of local features

- Hessian detector + RootSIFT (128D)
- MSER detector + RootSIFT (128D)
- Harris Laplace + HsvSIFT (384D)
- AKM for training codebook of size 1M

<table>
<thead>
<tr>
<th>local features</th>
<th>points per image</th>
<th>mAP(INS2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSER + RootSIFT</td>
<td>around 150</td>
<td>16.308</td>
</tr>
<tr>
<td>Hessian + RootSIFT</td>
<td>around 500</td>
<td>12.739</td>
</tr>
<tr>
<td>Harris + HsvSIFT</td>
<td>around 250</td>
<td>12.967</td>
</tr>
<tr>
<td>Total</td>
<td>around 900</td>
<td><strong>21.731</strong></td>
</tr>
</tbody>
</table>

Rich features are important, because they are complementary.
Contextual weighting

- Set different weights on ROI and backgrounds:

  In the aspect of metric

  Typical scheme:

  \[
  \text{sim}(q, d) = \sum_{i=1}^{D} \alpha_i q_i d_i , \text{where } \alpha_i = \begin{cases} 
  \beta & (\in \text{ROI}) \\
  1 & (\notin \text{ROI}) 
  \end{cases}
  \]

  (1)

  Similarity (take inner product and L2-normalization as an example, and set \( \beta=3 \)):

  \[
  \text{sim}(Q, I_1) = 1.47 \\
  \text{sim}(Q, I_2) = 1.33
  \]
Contextual weighting

• A good similarity measurement include of consistent:
  — Similarity kernel.
  — Normalization scheme.

• Good similarity measurement satisfies:
  — Self-similarity equals to one;
  — Self-similarity is the largest.

• L2-norm + inner product ✓
  L1-norm + inner product ✗

• Advise:
  — When you want to set larger weights on ROI descriptors, you may also need to modify the normalization scheme.

Boost the mAP by 3%
Burstiness

**Definition:** A visual word is more likely to appear in an image if it already appeared once in that image.

[Jegou. CVPR 2009]

- If we first normalize the feature vector, then calculate the similarity: image with very few descriptors equals to the image contains several dominant descriptors. This also leads to burstiness.
- Advise: L1-based similarity kernel rather than L2-based.

**Boost the mAP by 2%**
What’s next?

• Local features are **unable** to solve
  — Smooth objects or objects are more suitable to describe using shape etc.
  — Small objects which could extract few local features

• What’s next?
  — Introduce better similarity measurement?
  — Keep ensembling more features?
What’s next?

• How well would Deep Learning work for instance search?

[Razavian et al. CVPRw 2014]
Convolutional neural network

- Decaf has shown that CNN trained on ImageNet2012 1000CLS has good generalization.

[Krizhevsky et al. NIPS 2012]
Convolutional neural network

- **Two schemes**
  - As global features
    - $+ \approx 3\% \text{ mAP}$
  - Generic object detection + CNN
    - $+ \approx 1\% \text{ mAP}$
Convolutional neural network

• **Scheme 1: As global features**
  — Activations from a certain layer as global features.
  — CNN takes the entire image as the input, therefore it is unable to emphasize the ROI.
  — Relatively strict geometric information

<table>
<thead>
<tr>
<th>Layer</th>
<th>Dim</th>
<th>Metric</th>
<th>mAP (using CNN only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fc6</td>
<td>4096</td>
<td>L2</td>
<td>3.84</td>
</tr>
<tr>
<td>Fc6 + Relu</td>
<td>4096</td>
<td>SSR</td>
<td>3.43</td>
</tr>
<tr>
<td>Fc7 + Relu</td>
<td>4096</td>
<td>L2</td>
<td>3.07</td>
</tr>
<tr>
<td>Fc7 + Relu</td>
<td>4096</td>
<td>SSR</td>
<td>2.67</td>
</tr>
<tr>
<td>Fc8</td>
<td>1000</td>
<td>SSR</td>
<td>1.34</td>
</tr>
</tbody>
</table>

*Boost the mAP by 3% (combined with BOW)*
Convolutional neural network

• **Scheme 2: Localized search**
  — Instance search is inherently asymmetric.
  — CNN is not like BOW, it has fewer geometric correspondences, especially for the output of fully connected layer.

• How to deal with the asymmetric problem of CNN?
  — Train a specific CNN
    But where is the training set come from?

  — Generic object detection (derived from RCNN) + CNN feature comparison

**Problem:** Designing an efficient indexing system is important. As a trial run, we only use it for reranking the top 100 results.

**Boost the mAP by 1%**
Topic 9113, result from BOW baseline. Images in red box are false results.
Topic 9113, result after reranking.
Failure examples
Failure examples: After reranking
Problems

- The input region is limited to a rectangle, not arbitrary shape.
### Problems

<table>
<thead>
<tr>
<th>Instance Search</th>
<th>Object Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. No suitable training data;</td>
<td>1. Enough training data;</td>
</tr>
<tr>
<td>2. Focus on <strong>both intra-class and inter-class</strong> analysis;</td>
<td>2. Mainly focus on <strong>inter-class</strong> analysis;</td>
</tr>
<tr>
<td>3. Objects to be retrieved could be anything;</td>
<td>3. Object class to be detected is specified ahead of time;</td>
</tr>
<tr>
<td>4. Require real-time response.</td>
<td>4. Could be performed off-line.</td>
</tr>
<tr>
<td>5. Focus on finding relevant image from a large dataset.</td>
<td>5. Focus on detecting relevant object in a given image.</td>
</tr>
</tbody>
</table>
Thanks!

jiang1st@bupt.edu.cn
https://sites.google.com/site/whjiangpage/
http://www.bupt-mcprl.net