BUPT-MCPRL at Trecvid2015
Instance Search Task

Wenhui Jiang (jiang1st@bupt.edu.cn)
Zhicheng Zhao, Fei Su, Mei Liu, Shanwei Zhao, Anni Cai

MCPR Lab
Beijing University of Posts and Telecommunications
Brief Overview

• Three local features
  – MSER + RootSIFT
  – Hessian Affine + RootSIFT
  – Deep Conv5

• One global feature
  – Deep FC6

• Feature fusion
  – Manual tuned
  – Query adaptive

• Trial feature
  – Hessian Affine + Deep Conv
**Brief Overview**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MSER + RootSIFT</td>
<td>15.86</td>
<td>13.00</td>
<td></td>
</tr>
<tr>
<td>Hessian Affine + RootSIFT</td>
<td>21.59</td>
<td>17.03</td>
<td></td>
</tr>
<tr>
<td>Deep Conv5</td>
<td>16.58</td>
<td>18.37</td>
<td></td>
</tr>
<tr>
<td>Deep Fc6</td>
<td>4.52</td>
<td>4.03</td>
<td></td>
</tr>
</tbody>
</table>
Deep Conv Feature

- Fully connected layer
- Locally connected layer

Deep FC

Deep Conv
Deep Conv Feature

Receptive field sizes and strides for AlexNet

<table>
<thead>
<tr>
<th>Layer</th>
<th>Rf size</th>
<th>Stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv1</td>
<td>11 X 11</td>
<td>4 X 4</td>
</tr>
<tr>
<td>Conv2</td>
<td>51 X 51</td>
<td>8 X 8</td>
</tr>
<tr>
<td>Conv3</td>
<td>99 X 99</td>
<td>16 X 16</td>
</tr>
<tr>
<td>Conv4</td>
<td>131 X 131</td>
<td>16 X 16</td>
</tr>
<tr>
<td>Conv5</td>
<td>163 X 163</td>
<td>16 X 16</td>
</tr>
<tr>
<td>Pool5</td>
<td>195 X 195</td>
<td>32 X 32</td>
</tr>
</tbody>
</table>

Reference: Exploiting Local Features from Deep Networks for Image Retrieval, CVPR Workshop 2015
Deep Conv Feature

Feature representation workflow for Deep conv features

1. Input image
2. Dense sampling conv5 activations
3. 1M codebook
4. BoW feature (Deep Conv5)
## Deep Conv Feature

<table>
<thead>
<tr>
<th>Features</th>
<th>mAP (2013)</th>
<th>mAP (2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSER + RootSIFT</td>
<td>15.86</td>
<td>13.00</td>
</tr>
<tr>
<td>Hessian Affine + RootSIFT</td>
<td>21.59</td>
<td>17.03</td>
</tr>
<tr>
<td><strong>Deep Conv5</strong></td>
<td><strong>16.58</strong></td>
<td><strong>18.37</strong></td>
</tr>
<tr>
<td>Deep Fc6</td>
<td>4.52</td>
<td>4.03</td>
</tr>
</tbody>
</table>
Multiple Features Fusion

Query → Feature 1 → Rank list 1 → Late Fusion → Final rank list
Feature 2 → Rank list 2
......
Feature 4 → Rank list 4

Weight: W₁, W₂, W₄

Courtesy: Query-Adaptive Late Fusion for Image Search and Person Re-identification, CVPR2015
Multiple Features Fusion

Query

Feature 1

Rank list 1

W₁(q)

Late Fusion

Feature 2

Rank list 2

W₂(q)

Feature 4

Rank list 4

W₄(q)

Final rank list

q

Courtesy: Query-Adaptive Late Fusion for Image Search and Person Re-identification, CVPR2015
Multiple Features Fusion

Good feature: L-shaped score curve

Bad Feature: Flat score curve

Courtesy: Query-Adaptive Late Fusion for Image Search and Person Re-identification, CVPR2015
Multiple Samples Fusion

Take four samples as four features
Dense VS Sparse

Feature representation workflow for SIFT baselines

Feature representation workflow for Deep conv features

Courtesy: Dense Interest Points, CVPR2010
Tentative Experiment

Visual system of human

SIFT Descriptor
Tentative Experiment

Layer 1 Visualization

AlexNet

384-D feature vector
# Tentative Experiment

<table>
<thead>
<tr>
<th>Features</th>
<th>mAP (2013)</th>
<th>mAP (2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSER + RootSIFT</td>
<td>15.86</td>
<td>13.00</td>
</tr>
<tr>
<td>Hessian Affine + RootSIFT</td>
<td>21.59</td>
<td>17.03</td>
</tr>
<tr>
<td>Deep Conv5</td>
<td>16.58</td>
<td>18.37</td>
</tr>
<tr>
<td>Hessian Affine + Deep Conv1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>
Thank you