TWO-STAGE POOLING OF DEEP CONVOLUTIONAL FEATURES FOR IMAGE RETRIEVAL

Tiancheng Zhi, Ling-Yu Duan, Yitong Wang, Tiejun Huang
Institute of Digital Media, School of EE&CS, Peking University

Impact

Step 2: Mid-level Feature Extraction

- Extract the last convolutional/pooling layer feature maps
- Sort the feature within each feature map to avoid hard coding of location

Step 1: Patch Detection

- Use object proposal algorithm to detect N regions with high objectness
- Handle multiple scales, noisy background and abundant subjects

Step 3: Two-stage Partial Mean Pooling

1) Intra-patch Partial Mean Pooling

- Compute mean value of top $K_1$ responses for each feature map
- Capture the discriminative responses on the feature maps
- Transform the feature to a low-dimensional representation

2) Inter-patch Partial Mean Pooling

- Compute mean value of top $K_2$ patch features along each dimension
- Aggregate patch-level features into compact global representation

Experiment Results

Impact of Parameters

$N$: number of detected patches
$K_1$: parameter in intra-patch partial mean pooling
$K_2$: parameter in inter-patch partial mean pooling

Benchmark Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Dimension</th>
<th>Holidays</th>
<th>UKBench</th>
<th>Oxford5K</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT + TE + DA</td>
<td>1024</td>
<td>72.0</td>
<td>3.51</td>
<td>56.0</td>
</tr>
<tr>
<td>Neural Codes</td>
<td>4096</td>
<td>79.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MOP-CNN + PCA</td>
<td>512</td>
<td>78.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CNN + SfPoc</td>
<td>256</td>
<td>80.2</td>
<td>3.65</td>
<td>58.9</td>
</tr>
<tr>
<td>CNN + DPs + Max Pooling</td>
<td>4096</td>
<td>81.0</td>
<td>3.67</td>
<td>56.0</td>
</tr>
<tr>
<td>Two-stage PMP (VGGNet, with PCA-whitening)</td>
<td>512</td>
<td>86.6</td>
<td>3.80</td>
<td>64.0</td>
</tr>
</tbody>
</table>

Example Retrieval Result

Query

Retrieval Result

Ground Truth

Conclusions & Feature Work

- We have proposed a two-stage partial mean pooling strategy towards an advanced CNN feature extraction framework.
- The proposed compact and discriminative image representation outperforms state-of-the-art methods.
- How to incorporate low-level invariant features into this feature extraction framework will be included in our future work.