Unsupervised Discovery of Mid-Level Discriminative Patches

Saurabh Singh, Abhinav Gupta, Alexei A. Efros
Carnegie Mellon University

Presented by Chuan Wang
Which is from Paris?

visual primitives: patches
Discriminative Patches

Frequent + Discriminative

representative discriminative

representative discriminative
Mid-level Discriminative Patches

- Can be detected in a large number of images with high recall and precision
- Do not explicitly aim to discover whole semantic units
- Fully unsupervised: no separate training and test set
Mid-level Discriminative Patches

Highly Ranked Patches
Approach

- **K-Means**

- **Discriminative Clustering**
  - Input: discovery set $\mathcal{D}$ + “natural world” set $\mathcal{N}$
  - Make each cluster a detector
  - Train to find similar patches like those it already owns (linear SVM)
  - Not just against $\mathcal{D}$, but also against $\mathcal{N}$
Algorithm 1 Discover Top n Discriminative Patches

Require: Discovery set $D$, Natural World set $N$
1: $D \Rightarrow \{D_1, D_2\}; \quad N \Rightarrow \{N_1, N_2\}$  \hspace{1cm} \triangleright \text{Divide } D, N \text{ into equal sized disjoint sets}
2: $S \Leftarrow \text{rand.sample}(D_1)$  \hspace{1cm} \triangleright \text{Sample random patches from } D_1
3: $K \Leftarrow \text{kmeans}(S)$  \hspace{1cm} \triangleright \text{Cluster patches using KMeans}
4: \textbf{while not converged()} \textbf{do}
5: \hspace{1cm} \textbf{for all } i \text{ such that } \text{size}(K[i]) \geq 3 \text{ do}  \hspace{1cm} \triangleright \text{Prune out small ones}
6: \hspace{1cm} C_{\text{new}}[i] \Leftarrow \text{svm.train}(K[i], N_1)  \hspace{1cm} \triangleright \text{Train classifier for each cluster}
7: \hspace{1cm} K_{\text{new}}[i] \Leftarrow \text{detect.top}(C[i], D_2, m)  \hspace{1cm} \triangleright \text{Find top m new members in other set}
8: \hspace{1cm} \textbf{end for}
9: \hspace{1cm} K \Leftarrow K_{\text{new}}; \quad C \Leftarrow C_{\text{new}}
10: \hspace{1cm} \text{swap}(D_1, D_2); \quad \text{swap}(N_1, N_2)  \hspace{1cm} \triangleright \text{Swap the two sets}
11: \textbf{end while}
12: A[i] \Leftarrow \text{purity}(K[i]) + \lambda \times \text{discriminativeness}(K[i]) \forall i  \hspace{1cm} \triangleright \text{Compute scores}
13: \textbf{return} \text{select.top}(C, A, n)  \hspace{1cm} \triangleright \text{Sort according to scores and select top n patches}
Approach
Ranking Patches

- **Purity**
  - Summing up the SVM detection scores of top r cluster members (r>m)

- **Discriminativeness**
  - The number of firings on $\mathcal{D}$ / the number of firings on $\mathcal{D} \cup \mathcal{N}$

- **Rank**: linear combination of the two terms
Doublets

- Cluster converges to two or more concepts sometimes (b)
- Discover second-order spatial relationships and cleaning up
Highly Ranked Doublets
Evaluation (PASCAL VOC2007)

what percentage of cluster members correspond to the same visual entity

the number of images in the dataset covered by a given cluster
Works Even Better if Weakly Supervised
Works Even Better if Weakly Supervised
## Quantitative Evaluation (MIT Indoor-67)

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIST</td>
<td>29.7</td>
</tr>
<tr>
<td>Spatial Pyramid HOG <em>(SPHOG)</em> †</td>
<td>29.8</td>
</tr>
<tr>
<td>Spatial Pyramid SIFT <em>(SP)</em> †</td>
<td>34.4</td>
</tr>
<tr>
<td>Scene DPM [4]</td>
<td>30.4</td>
</tr>
<tr>
<td>MM-Scene [37]</td>
<td>28.0</td>
</tr>
<tr>
<td>Object Bank [3]</td>
<td>37.6</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>38.1</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Combination</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours+GIST</td>
<td>44.0</td>
</tr>
<tr>
<td>Ours+SP</td>
<td>46.4</td>
</tr>
<tr>
<td>Ours+GIST + SP</td>
<td>47.5</td>
</tr>
<tr>
<td>Ours+DPM</td>
<td>42.4</td>
</tr>
<tr>
<td>Ours+GIST+DPM</td>
<td>46.9</td>
</tr>
<tr>
<td>Ours+SP+DPM</td>
<td>46.4</td>
</tr>
<tr>
<td>GIST+SP+DPM [4]</td>
<td>43.1*</td>
</tr>
<tr>
<td>Ours+GIST+SP+DPM</td>
<td><strong>49.4</strong></td>
</tr>
</tbody>
</table>
Discussion

● Strong Points
  ○ Unsupervised: train classifier against natural world
  ○ Mid-level patches: semantically meaningful and discriminative
  ○ Cross validation: simple but effective

● Weak Points
  ○ $m=5$: quite a small number and reason?
  ○ Perspective changes / deformation: difficult?

● Discussion
  ○ What if shuffle picking subset during the iterative step?
Thank You!

Question?