1 Motivation

Randomized Max-Margin Compositions vs. part based approaches

RM²C Parsing vs. Standard Object Detection

2 Detection Procedure

recognition phase:

some \( h_i(x_i) \) of part classifier \( i \) at sites \( x_i \) of local image patch

responses of part classifiers at sites \( x \)

randomized compositions \( \gamma \) of query image

randomized compositional classifier \( f_\gamma \)

combined compositional classifier \( f(\gamma_1, \ldots, \gamma_k) \)


3 Randomized Parts & Compositions

Learning Parts without Part Annotations

- without extra annotation as difficult as finding the object
- clustering parts based on features (e.g. HOG) is not reliable
- to avoid incorrect groups we train parts using single positive patches which are randomly sampled from the training data

\[
m_i(x) = \max_{\gamma} \left( \sum_{\gamma} \max(0, 1 - h_i(x)) \right)
\]

\[
decision function \ h_i(x) = \alpha_i^T x + \beta_i
\]

Randomized Max-Margin Compositions RM²C

\[
\begin{align*}
\text{pars} \times \text{patches} & > 30000 D \\
\text{curse of dimensionality}
\end{align*}
\]

parts are highly uncorrelated

- subspaces / part grouping are not suitable for dim reduction

compositional classifier (SVM training)

\[
\text{evaluation of different grouping strategies on the validation set}
\]

randomized compositions clearly outperform other grouping strategies

final non-linear classifier with

\[
x(F(\gamma), F(\gamma')) = \exp(-||F(\gamma) - F(\gamma')||^2)\]

optimize max

\[
\begin{align*}
\alpha_i & = \frac{1}{2} \sum_{\gamma} \sum_{\gamma'} \alpha_i \alpha_j \exp(-||\gamma - \gamma'||^2) \\
\text{decision function} \ g(F(\gamma)) & = \sum_{\gamma \in T} \alpha_i h_i(x(\gamma))
\end{align*}
\]

Part Evaluation

- without extra annotation as difficult as finding the object
- clustering parts based on features (e.g HOG) is not reliable
- to avoid incorrect groups we train parts using single positive patches which are randomly sampled from the training data

measure individual part performance using keypoint annotations

comparison of the 48 DPM parts with a subset of 48 of our parts

4 Experimental Results

Object Detection Results PASCAL VOC 2007 (comp3)

<table>
<thead>
<tr>
<th>Method</th>
<th>AP</th>
<th>Rm²C</th>
<th>Rm²C</th>
<th>Rm²C</th>
<th>Rm²C</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPM [65]</td>
<td>52.6</td>
<td>50.9</td>
<td>49.9</td>
<td>50.9</td>
<td>51.5</td>
</tr>
<tr>
<td>PASCAL [15]</td>
<td>52.5</td>
<td>51.1</td>
<td>50.3</td>
<td>50.9</td>
<td>50.9</td>
</tr>
<tr>
<td>BC [30]</td>
<td>44.3</td>
<td>35.2</td>
<td>34.7</td>
<td>35.2</td>
<td>35.0</td>
</tr>
<tr>
<td>AOI [30]</td>
<td>44.6</td>
<td>40.5</td>
<td>39.8</td>
<td>40.5</td>
<td>40.3</td>
</tr>
<tr>
<td>RM²C</td>
<td>48.6</td>
<td>46.8</td>
<td>46.5</td>
<td>46.8</td>
<td>46.6</td>
</tr>
</tbody>
</table>
| Results on the MIT Indoor Dataset

Object Detection Results PASCAL VOC 2010 (comp3)

<table>
<thead>
<tr>
<th>Method</th>
<th>AP</th>
<th>Rm²C</th>
<th>Rm²C</th>
<th>Rm²C</th>
<th>Rm²C</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPM [65]</td>
<td>54.5</td>
<td>52.8</td>
<td>52.3</td>
<td>52.8</td>
<td>52.8</td>
</tr>
<tr>
<td>PASCAL [15]</td>
<td>54.3</td>
<td>52.8</td>
<td>52.3</td>
<td>52.8</td>
<td>52.8</td>
</tr>
<tr>
<td>BC [30]</td>
<td>44.3</td>
<td>40.5</td>
<td>40.3</td>
<td>40.5</td>
<td>40.3</td>
</tr>
<tr>
<td>AOI [30]</td>
<td>44.6</td>
<td>40.5</td>
<td>40.3</td>
<td>40.5</td>
<td>40.3</td>
</tr>
<tr>
<td>RM²C</td>
<td>48.6</td>
<td>46.8</td>
<td>46.5</td>
<td>46.8</td>
<td>46.6</td>
</tr>
</tbody>
</table>

Project Website

- parts for all datasets are available for download here:

- performance evaluation with increasing number of parts on the validation set
- mean average precision saturates around 1000 parts

Are all parts needed?

5 Object Parsing Results

- performance evaluation with increasing number of parts on the validation set
- mean average precision saturates around 1000 parts