A contextual rescoring method is proposed for improving the detection of body joints of a pictorial structure model for human pose estimation. A set of mid-level parts is incorporated in the model, and their detections are used to extract spatial and score-related features relative to other body joint hypotheses. A technique is proposed for the automatic discovery of a compact subset of poselets that covers a set of validation images while maximizing precision. A rescoring mechanism is defined as a set-based boosting classifier that computes a new score for body joint detections, given its relationship to detections of other body joints and mid-level parts in the image. This new score complements the unary potential of a discriminatively trained pictorial structure model. Experiments on two benchmarks show performance improvements when considering the proposed mid-level image representation and rescoring approach in comparison with other pictorial structure-based approaches.

Abstract

Human Pose Estimation: Pipeline

Mid-level part representation

   - Generate seed random windows.
   - Procrustes alignment to gather training samples.
   - Estimate Gaussian distribution of keypoints.
   - Compute keypoint estimation precision

2. Poselet selection: weighted set cover
   \[
   \min \sum_{j \in \text{val}} (1 - \text{Prec}(j)) x_j \\
   \text{subject to} \sum_{j \in \text{val}} x_j \geq 1 \forall n, x_j \in [0,1].
   \]

Contextual Rescoring & Pictorial structure formulation

1. Mid-level contextual detections

2. Contextual features


4. Pictorial structure formulation


Quantitative results

<table>
<thead>
<tr>
<th></th>
<th>Mean PCP (UIUC Sports)</th>
<th>Mean PCP (LSP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours poselets</td>
<td></td>
<td>57.8, 57.8</td>
</tr>
<tr>
<td>Ours predefined</td>
<td></td>
<td>58.2, 58.6</td>
</tr>
</tbody>
</table>

Qualitative results

References