Unsupervised Learning of Object Landmarks through Conditional Image Generation

Mamshad Nayeem Rizve
Outline

- The problem
- Solution architecture
- Proposed solution
- Experimental setup
- Experiments
The Problem

Landmark detectors for visual objects.
Solution architecture
Proposed Solution

● A function to learn structure of an image:
  \[ y' = \Phi(x') \]

● Conditional image generation:
  \[ \Psi : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}, \quad (x, y') \mapsto x' \]

● Reconstructed target image:
  \[ \hat{x}' = \Psi(x, \Phi(x')) \]

● Loss:
  \[ \min_{\Psi, \Phi} E_{x, x'} [\mathcal{L}(x', \Psi(x, \Phi(x'))) \]
Proposed Solution

Heatmaps bottleneck

- Heatmap renormalisation:
  \[ u_k^*(x) = \frac{\sum_{u \in \Omega} u e^{S_u(x;k)}}{\sum_{u \in \Omega} e^{S_u(x;k)}} \]

- Separable implementation:
  \[ u_{ik}^*(x) = \frac{\sum_{u_i \in \Omega_i} u_i e^{S_{u_i}(x;k)}}{\sum_{u_i \in \Omega_i} e^{S_{u_i}(x;k)}} \quad \text{and} \quad S_{u_i}(x;k) = \sum_{u_j \in \Omega_j} S_{(u_1,u_2)}(x;k), \]

- Gaussianization of heatmaps:
  \[ \Phi_u(x;k) = \exp \left( -\frac{1}{2\sigma^2} \| u - u_k^*(x) \|^2 \right) \]
Proposed Solution

Perceptual loss
Experimental setup
Encoder network architecture
Experimental setup

Regressor network architecture
Experiments
Experiments
Experiments
Learning facial landmarks (qualitative results)
Experiments
Learning facial landmarks (quantitative results)

<table>
<thead>
<tr>
<th>Method</th>
<th>$K$</th>
<th>MAFL</th>
<th>AFLW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCPR [2]</td>
<td>–</td>
<td>11.60</td>
<td></td>
</tr>
<tr>
<td>CFAN [54]</td>
<td>15.84</td>
<td>10.94</td>
<td></td>
</tr>
<tr>
<td>Cascaded CNN [41]</td>
<td>9.73</td>
<td>8.97</td>
<td></td>
</tr>
<tr>
<td>TCDCN [57]</td>
<td>7.95</td>
<td>7.65</td>
<td></td>
</tr>
<tr>
<td>RAR [41]</td>
<td>–</td>
<td>7.23</td>
<td></td>
</tr>
<tr>
<td>MTCNN [56]</td>
<td>5.39</td>
<td>6.90</td>
<td></td>
</tr>
</tbody>
</table>
Experiments

Ablation study
Experiments
Learning Human Pose
Experiments
Unsupervised Landmarks on Human3.6M
Experiments
Learning 3D object landmarks: pose, shape, and illumination invariance
Experiments
Disentangling appearance and geometry
Thank You
Questions?