A Deep Cascade Network for Unaligned Face Attribute Classification

Hui Ding, Hao Zhou, Shaohua Kevin Zhou and Rama Chellappa

Introduction

Face Attribute Classification

- Predict different face properties
- Useful for face verification

Limitations

- Depends heavily on face alignment
- Doesn’t consider the spatial relationship

Proposed Model

- The localization network is responsible to detect the attribute relevant face regions. It is trained in a weakly-supervised manner with attributes labels.
- The classification network is consisted of one global and several parts subnets.

Multi-Net Learning

- The localization network is responsible to detect the attribute relevant face regions. It is trained in a weakly-supervised manner with attributes labels.
- The classification network is consisted of one global and several parts subnets.

Hint-based Model Compression

- The localization network is responsible to detect the attribute relevant face regions. It is trained in a weakly-supervised manner with attributes labels.
- The classification network is consisted of one global and several parts subnets.

Multi-Net Learning Results

<table>
<thead>
<tr>
<th>Methods</th>
<th>Classif. Brancl</th>
<th>Loc. Branch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without MNL</td>
<td>91.05</td>
<td>91.06</td>
</tr>
<tr>
<td>MNL</td>
<td>91.07</td>
<td>91.06</td>
</tr>
</tbody>
</table>

Table 2: Fine-grained classification accuracy on CUB-200 dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>Classif. Branch</th>
<th>Loc. Branch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without MNL on full image</td>
<td>72.10</td>
<td>71.66</td>
</tr>
<tr>
<td>MNL on full image</td>
<td>76.03</td>
<td></td>
</tr>
<tr>
<td>Without MNL on crop</td>
<td>75.76</td>
<td>71.90</td>
</tr>
<tr>
<td>MNL on full image</td>
<td>76.03</td>
<td></td>
</tr>
</tbody>
</table>

Highlight-based Model Compression Results

<table>
<thead>
<tr>
<th>Layer</th>
<th>Core1</th>
<th>Core2</th>
<th>Core3</th>
<th>Core4</th>
<th>Core5</th>
<th>Core6</th>
<th>Classifiers</th>
<th>Accracy</th>
<th>Params</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3x3x32</td>
<td>3x3x32</td>
<td>3x3x32</td>
<td>3x3x32</td>
<td>3x3x32</td>
<td>3x3x32</td>
<td>GAP</td>
<td>90.89</td>
<td>2.75M</td>
</tr>
<tr>
<td></td>
<td>3x3x64</td>
<td>3x3x64</td>
<td>3x3x64</td>
<td>3x3x64</td>
<td>3x3x64</td>
<td>3x3x64</td>
<td>GAP</td>
<td>89.96</td>
<td>3.27M</td>
</tr>
<tr>
<td></td>
<td>3x3x128</td>
<td>3x3x128</td>
<td>3x3x128</td>
<td>3x3x128</td>
<td>3x3x128</td>
<td>3x3x128</td>
<td>GAP</td>
<td>90.74</td>
<td>4.94M</td>
</tr>
<tr>
<td></td>
<td>3x3x256</td>
<td>3x3x256</td>
<td>3x3x256</td>
<td>3x3x256</td>
<td>3x3x256</td>
<td>3x3x256</td>
<td>GAP</td>
<td>88.16</td>
<td>5.52M</td>
</tr>
<tr>
<td></td>
<td>3x3x512</td>
<td>3x3x512</td>
<td>3x3x512</td>
<td>3x3x512</td>
<td>3x3x512</td>
<td>3x3x512</td>
<td>GAP</td>
<td>86.88</td>
<td>6.67M</td>
</tr>
<tr>
<td></td>
<td>3x3x1024</td>
<td>3x3x1024</td>
<td>3x3x1024</td>
<td>3x3x1024</td>
<td>3x3x1024</td>
<td>3x3x1024</td>
<td>GAP</td>
<td>85.62</td>
<td>7.82M</td>
</tr>
</tbody>
</table>

Region Selection Layer Visualization

Classification Results on Unaligned CelebA

Acknowledgement

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.