Deep Learning Face Attributes in the Wild
Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang
The Chinese University of Hong Kong
{ziwei13, pluo, xwangle}@ie.cuhk.edu.hk, vgwang@ee.cuhk.edu.hk

1. Overview

Problem
Face Attributes Prediction in the Wild

Performance

<table>
<thead>
<tr>
<th></th>
<th>CelebA</th>
<th>LFWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>FaceTracer (RCVM)</td>
<td>81%</td>
<td>74%</td>
</tr>
<tr>
<td>PANDA-W (CVP14)</td>
<td>79%</td>
<td>71%</td>
</tr>
<tr>
<td>PANDA-I (CVP14)</td>
<td>85%</td>
<td>81%</td>
</tr>
<tr>
<td>SC+ANet</td>
<td>83%</td>
<td>76%</td>
</tr>
<tr>
<td>LNets+ANet</td>
<td>83%</td>
<td>79%</td>
</tr>
<tr>
<td>LNets+ANet (w/o)</td>
<td>87%</td>
<td>84%</td>
</tr>
</tbody>
</table>

Running Time
LNets: 35ms, ANet: 14ms

• Project Page: http://personal.ie.cuhk.edu.hk/~lz013/projects/FaceAttributes.html

2. Motivation

- Existing methods: global and local methods
- Global methods: not robust to deformations of objects
- Local methods: rely on face localization and alignment, which would fail under unconstrained face images with complex variations
- Our idea: joint face localization and attribute prediction using only image-level attribute tags

3. Large-scale CelebFaces Attributes Dataset

- 202,599 face images
- 10,177 human identities
- 5 landmarks per image
- 40 attributes per image

- 20x larger than previous

Available at: http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

4. Overall Pipeline

- LNets:
  - (i) pre-trained with massive general objects
  - (ii) face localization with weak supervision

- ANet:
  - (i) pre-trained with massive face identities
  - (ii) attribute prediction by leveraging local features

- LNets and ANet are jointly learned

5.1. Experimental Results (Face Localization)

- 202,599 face images
- 10,177 human identities
- 5 landmarks per image
- 40 attributes per image

5.2. Experimental Results (Attribute Prediction)

- With carefully designed pre-training strategies, our approach is robust to background clutters and face variations.
- We devise a new fast feed-forward algorithm for locally shared filters to save redundant computation.
- We have also revealed multiple important facts about learning face representation, which shed a light on new directions of face localization and representation learning.