Deep Feature Learning for Face Alignment and Facial Age Estimation

Hao Liu
Tsinghua University, China
**Personal Introduction**

**Hao Liu**
- Visual Intelligence Group, Department of Automation, Tsinghua University
- 3rd Year of Ph.D. candidate
- Supervised by Professor Jie Zhou and Associate Professor Jiwen Lu
- Researching Areas: Deep Learning, Facial Age Estimation and Face Alignment

**Publications**

TIP 2017: Deep Sharable and Structural Detectors for Face Alignment
Deep Sharable and Structural Detectors for Face Alignment[1]

- **Motivation**
  - Facial landmarks are usually spatially correlated.
  - Conventional approaches utilizing hand-crafted features might lose shape-sensitive details.

- **Structural Feature Learning**: model the correlation of neighbouring landmarks to dynamically cover more semantic details.
- **Sharable Detectors**: remove the noises of spatially overlapped patches.
- **Nonlinear Regression**: infer occluded part by non-occluded parts.
# Experimental Results

<table>
<thead>
<tr>
<th>Method</th>
<th>LFPW 68-pts</th>
<th>HELEN 68-pts</th>
<th>HELEN 192-pts</th>
<th>Common Set 68-pts</th>
<th>Challenging Set 68-pts</th>
<th>Full Set 68-pts</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPLL [28]</td>
<td>8.29</td>
<td>8.16</td>
<td>-</td>
<td>8.22</td>
<td>18.33</td>
<td>10.20</td>
</tr>
<tr>
<td>RCP [23]</td>
<td>6.56</td>
<td>5.93</td>
<td>6.50</td>
<td>6.18</td>
<td>17.26</td>
<td>8.35</td>
</tr>
<tr>
<td>GN-DPM [25]</td>
<td>5.92</td>
<td>5.69</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SDM [8]</td>
<td>5.67</td>
<td>5.50</td>
<td>5.85</td>
<td>5.57</td>
<td>15.40</td>
<td>7.50</td>
</tr>
<tr>
<td>CFAN [16]</td>
<td>5.44</td>
<td>5.53</td>
<td>-</td>
<td>5.50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BPCPR [19]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.24</td>
<td>16.56</td>
<td>7.46</td>
</tr>
<tr>
<td>ESR [9]</td>
<td>-</td>
<td>-</td>
<td>5.70</td>
<td>5.28</td>
<td>17.00</td>
<td>7.58</td>
</tr>
<tr>
<td>LBF [10]</td>
<td>-</td>
<td>-</td>
<td>5.41</td>
<td>4.95</td>
<td>11.98</td>
<td>6.32</td>
</tr>
<tr>
<td>LBF fast [10]</td>
<td>-</td>
<td>-</td>
<td>5.80</td>
<td>5.38</td>
<td>15.50</td>
<td>7.37</td>
</tr>
<tr>
<td>Deep Reg [15]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.51</td>
<td>13.80</td>
<td>6.31</td>
</tr>
<tr>
<td>CFSS [12]</td>
<td>4.87</td>
<td>4.63</td>
<td>4.74</td>
<td>4.73</td>
<td>9.98</td>
<td>5.76</td>
</tr>
<tr>
<td>TCDCN [47]</td>
<td>4.57</td>
<td>4.60</td>
<td>4.63</td>
<td>4.80</td>
<td>8.60</td>
<td>5.54</td>
</tr>
<tr>
<td>DCRFA [39]</td>
<td>4.57</td>
<td>4.25</td>
<td>-</td>
<td>4.19</td>
<td>8.42</td>
<td>5.02</td>
</tr>
<tr>
<td>R-DSSD*</td>
<td>4.77</td>
<td>4.31</td>
<td>4.95</td>
<td>4.57</td>
<td>10.86</td>
<td>5.91</td>
</tr>
<tr>
<td>R-DSSD</td>
<td><strong>4.52</strong></td>
<td><strong>4.08</strong></td>
<td><strong>4.62</strong></td>
<td><strong>4.16</strong></td>
<td><strong>9.20</strong></td>
<td><strong>5.59</strong></td>
</tr>
</tbody>
</table>

![Graphs showing performance comparison for different methods](image-url)
Annotated Results

Evaluation on 300-W, where 68 landmarks were employed
Annotated Results

Evaluation regarding with occlusion (68 landmarks):

Evaluation regarding with denser landmarks (192 landmarks):

Running Speed: 30FPS@Intel-i5 CPU
TPAMI 2017: Two-Stream Transformer Networks for Video-based Face Alignment
Framework

- Two-Stream Transformer Networks
  - Spatial appearance information
  - Temporal consistency
  - Weighted Fusion
Results

Fig. 3. CED curves of our TSTN compared to the state-of-the-arts on three categories in 300-VW [34] separately. In contrast to the state-of-the-art methods, our TSTN achieves comparable results in category two and superior performance in category one and the most difficult category three.

Fig. 4. Resulting examples of our TSTN on the 557th video clip in 300-VW [34] Category Three, where the selected tracked subject undergoes severe poses over time. The bottom subfigure shows that our TSTN exhibits robustness to difficult cases like large variations of facial aspect ratios.
Facial Age Estimation in the wild

挫折

- Large variances of facial expressions and occlusions
- Appearance changes with facial makeup
- Label correlation for human age labels

我们的聚焦

- Exploiting Label Correlation for Ages
- Learning Robust Features
State-of-the-arts

 Facial Age Estimation Framework

 Conventional Methods for Facial Age Estimation:

  - Feature Extraction (Requires strong prior knowledge)
    - Hand-crafted Features: BIF, LBP, SIFT
    - Shallow Feature Learning: CS-LBFL [Lu et al, TIP 2015]
  - Age Estimator (Should explicitly explore the ordinal relation for ages)
    - LDL [Geng et al, PAMI 2013]
    - OHRANK [Chen et al, CVPR 2012]
Facial Age Estimation x 2
Deep Feature Learning for Facial Age Estimation

- **Why Deep Feature Learning?**
  - **Learning Features** directly from raw pixels
  - **Modeling Nonlinear Relationship** between Pixels and Labels
  - Transfer Learning (Fine-tuning)

- **Facial Age Estimation**
  - Label Correlation (FG 2017)
  - Missing Labels (PR 2017)

**Jointly learning feature descriptors for face representation and exploiting the relationship of human age labels**
Motivation

- Deep Convolutional neural networks (CNN) works very well for face recognition.
- Human age labels are chronologically correlated and age estimation is an ordinal learning computer vision problem.


Given a quadruplet of batched data, we construct a label ordinal graph based on the ordinal embeddings [1].

The dissimilarity of face pairs in the learned feature space should be isotonic to that of the ordinal relations within the label ordinal graph.

The distance of the face pair with a larger age gap should be smoothly bigger than that of the face pair with a smaller age gap.

Weighting Function:

\[
\omega_{y_{p1}, y_{p2}} = \begin{cases} 
(\left| y_{p1} - y_{p2} \right| + 1)^\eta, & \text{if } y_{p1} \neq y_{p2} \\
1, & \text{otherwise}
\end{cases}
\]
Evaluation on Challenge Data

Face resulting samples where the errors are less than one year old.
Group-Aware Deep Feature Learning for Facial Age Estimation[1]

- Motivation
  - Densely collecting face samples across a large range of age labels is difficult (unbalanced training samples across age classes)

Basic Idea

- **Group-Aware Relationship**: 1) inter-group variances are maximized; 2) intra-group variances are minimized.

- **Smoothness of Overlaps**: face samples within overlaps should be smoothly weighted according to the age differences.
Evaluation on MORPH and FG-NET

- **MORPH:**
- **FG-NET:**
THANK YOU!