Deep Learning Face Representation by Joint Identification-Verification

Tongzhou Mu
Feb 02, 2018
Face Recognition
Basic Concepts

- **Face detection**
  - Give a bounding box for face in image

- **Face alignment**
  - Find landmarks in face

- **Face identification**
  - Image classification
  - Category number will be very large

- **Face verification**
  - Check whether two images are from one person
Several Methods in History
Pipeline

- Face alignment: SDM
- Feature extraction: DeepID2, Dimension Reduction
- Face verification: Joint Bayesian model
Deep Feature Representation: DeepID2

PART 2
Deep Feature Representation

- Input
  - Face patch

- Output
  - Feature vector

- Previous work
  - DeepFace from Facebook AI Research
ConvNets

Weights shared in 2x2 local regions

Weights unshared

Multi-scale
Property of Good Feature

Intra-personal variation

Inter-personal variation
Supervisory Signals

- **Face identification signal**
  - Classifies each face image into one of $n$ (e.g., $n = 8192$) different identities
  - Increases the *inter*-personal variations

- **Face verification signal**
  - Encourages DeepID2 extracted from faces of the same identity to be similar
  - Reduces the *intra*-personal variations
Connect DeepID2 to a n-way softmax layer

\[
\text{Ident}(f, t, \theta_{id}) = - \sum_{i=1}^{n} -p_i \log \hat{p}_i = - \log \hat{p}_t
\]

Must form discriminative identity-related features

Large inter-personal variations
Verification Signal

- Loss function based on the $L_2$ norm

\[
\text{Verif}(f_i, f_j, y_{ij}, \theta_{ve}) = \begin{cases} 
\frac{1}{2} ||f_i - f_j||_2^2 & \text{if } y_{ij} = 1 \\
\frac{1}{2} \max(0, m - ||f_i - f_j||_2)^2 & \text{if } y_{ij} = -1 
\end{cases}
\]

- Encourage the distance larger than a margin $m$

- Loss function based on the $L_1$ norm is similar
Verification Signal

- Loss function based on cosine similarity

\[
\text{Verif}(f_i, f_j, y_{ij}, \theta_{ve}) = \frac{1}{2} (y_{ij} - \sigma(wd + b))^2
\]

\[
d = \frac{f_i \cdot f_j}{\|f_i\|_2 \|f_j\|_2}
\]
The DeepID2 learning algorithm

**input:** training set $\chi = \{(x_i, l_i)\}$, initialized parameters $\theta_c$, $\theta_id$, and $\theta_ve$, hyperparameter $\lambda$, learning rate $\eta(t)$, $t \leftarrow 0$

**while** not converge **do**

$t \leftarrow t + 1$  
**sample two training samples** $(x_i, l_i)$ and $(x_j, l_j)$ from $\chi$

$f_i = \text{Conv}(x_i, \theta_c)$ and $f_j = \text{Conv}(x_j, \theta_c)$

$\nabla \theta_id = \frac{\partial \text{Ident}(f_i, l_i, \theta_id)}{\partial \theta_id} + \frac{\partial \text{Ident}(f_j, l_j, \theta_id)}{\partial \theta_id}$

$\nabla \theta_ve = \lambda \cdot \frac{\partial \text{Verif}(f_i, f_j, y_{ij}, \theta_ve)}{\partial \theta_ve}$, where $y_{ij} = 1$ if $l_i = l_j$, and $y_{ij} = -1$ otherwise.

$\nabla f_i = \frac{\partial \text{Ident}(f_i, l_i, \theta_id)}{\partial f_i} + \lambda \cdot \frac{\partial \text{Verif}(f_i, f_j, y_{ij}, \theta_ve)}{\partial f_i}$

$\nabla f_j = \frac{\partial \text{Ident}(f_j, l_j, \theta_id)}{\partial f_j} + \lambda \cdot \frac{\partial \text{Verif}(f_i, f_j, y_{ij}, \theta_ve)}{\partial f_j}$

$\nabla \theta_c = \nabla f_i \cdot \frac{\partial \text{Conv}(x_i, \theta_c)}{\partial \theta_c} + \nabla f_j \cdot \frac{\partial \text{Conv}(x_j, \theta_c)}{\partial \theta_c}$

update $\theta_id = \theta_id - \eta(t) \cdot \nabla \theta_id$, $\theta_ve = \theta_ve - \eta(t) \cdot \nabla \theta_ve$, and $\theta_c = \theta_c - \eta(t) \cdot \nabla \theta_c$.

end while

output $\theta_c$
Other Parts of The Pipeline

PART 3
Pipeline

Face alignment
• SDM

Feature extraction
• DeepID2
• Dimension Reduction

Face verification
• Joint Bayesian model

Crop 400 face patches

200 ConvNets

25 DeepID2 Features (4000-d)

21 landmarks

180-d vector
Selected Patches
Joint Bayesian Model

- Feature representation $f$ as the sum of inter- and intra-personal variations, or $f = \mu + \varepsilon$
- $\mu$ and $\varepsilon$ are Gaussian distributions
- log-likelihood ratio test

$$\log \frac{P(f_1, f_2 | H_{inter})}{P(f_1, f_2 | H_{intra})}$$
Datasets

- **Training**
  - CelebFaces+ (202,599 face images of 10,177 identities)
  - CelebFaces+A (8192 identities) for DeepID2
  - CelebFaces+B (1985 identities) for feature selection and Joint Bayesian

- **Testing**
  - LFW (13,233 face images of 5749 identities)
Balance identification & verification signals
Balance identification & verification signals

First 2 PCA dimensions
Richer identity information helps to form DeepID2 with diverse inter-personal variations.
Investigating the verification signals

Table 2: Comparison of different verification signals.

<table>
<thead>
<tr>
<th>verification signal</th>
<th>L2</th>
<th>L2+</th>
<th>L2-</th>
<th>L1</th>
<th>cosine</th>
<th>none</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 norm (%)</td>
<td>94.95</td>
<td>94.43</td>
<td>86.23</td>
<td>92.92</td>
<td>87.07</td>
<td>86.43</td>
</tr>
<tr>
<td>Joint Bayesian (%)</td>
<td>95.12</td>
<td>94.87</td>
<td>92.98</td>
<td>94.13</td>
<td>93.38</td>
<td>92.73</td>
</tr>
</tbody>
</table>
Table 3: Face verification accuracy with DeepID2 extracted from an increasing number of face patches.

<table>
<thead>
<tr>
<th># patches</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy (%)</td>
<td>95.43</td>
<td>97.28</td>
<td>97.75</td>
<td>98.55</td>
<td>98.93</td>
<td>98.97</td>
</tr>
<tr>
<td>time (ms)</td>
<td>1.7</td>
<td>3.4</td>
<td>6.1</td>
<td>11</td>
<td>23</td>
<td>35</td>
</tr>
</tbody>
</table>
Comparison with other methods

Table 4: Accuracy comparison with the previous best results on LFW.

<table>
<thead>
<tr>
<th>method</th>
<th>accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>high-dim LBP [4]</td>
<td>95.17 ± 1.13</td>
</tr>
<tr>
<td>TL Joint Bayesian [2]</td>
<td>96.33 ± 1.08</td>
</tr>
<tr>
<td>DeepFace [22]</td>
<td>97.35 ± 0.25</td>
</tr>
<tr>
<td>DeepID [21]</td>
<td>97.45 ± 0.26</td>
</tr>
<tr>
<td>GaussianFace [14]</td>
<td>98.52 ± 0.66</td>
</tr>
<tr>
<td>DeepID2</td>
<td>99.15 ± 0.13</td>
</tr>
</tbody>
</table>
Conclusion & Extension

PART 5
Main Contributions

- Combine identification and verification supervisory signals
- Reducing intra-personal variations while enlarging inter-personal differences
Extension

- **DeepID3**: Face Recognition with Very Deep Neural Networks
- Borrow ideas from VGG net and GoogLeNet
- Achieve 99.53% on LFW
Thanks