Do We Really Need to Collect Millions of Faces for Effective Face Recognition?

Iacopo Masi*1, Anh Tuan Tran*1, Tal Hassner*2,3, Jatuporn Toy Leksut1 and Gerard Medioni1

1. Institute for Robotics and Intelligent Systems, USC, CA, USA
2. Information Sciences Institute, USC, CA, USA
3. The Open University of Israel, Israel

* Denotes equal authorship

Based on work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA 2014-14071600010.
**Motivation**

### Intra-subject Variations

<table>
<thead>
<tr>
<th>Method</th>
<th># train img.</th>
<th>subj.</th>
<th>img. / subj.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepFaces’14 (FB)</td>
<td>4 m</td>
<td>4,030</td>
<td>1k</td>
</tr>
<tr>
<td>VGG Face’15</td>
<td>2.6 m</td>
<td>2,622</td>
<td>1k</td>
</tr>
<tr>
<td>Face Net’15 (Google)</td>
<td>200 m</td>
<td>8 m</td>
<td>25</td>
</tr>
<tr>
<td>Fusion’15 (FB)</td>
<td>500 m</td>
<td>10 m</td>
<td>50</td>
</tr>
<tr>
<td>MegaFace’16</td>
<td>1.02 m</td>
<td>690.5 k</td>
<td>1.5</td>
</tr>
</tbody>
</table>

### Pose (yaw) Variations

- LFW
- CASIA

Even with lots of resources, it's hard to ensure sufficient intra-subject and pose variations.
Motivation

Intra-subject Variations

<table>
<thead>
<tr>
<th>Method</th>
<th># train img.</th>
<th>subj.</th>
<th>img. / subj.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepFaces’14 (FB)</td>
<td>4 m</td>
<td>4,030</td>
<td>1k</td>
</tr>
<tr>
<td>VGG Face’15</td>
<td>2.6 m</td>
<td>2,622</td>
<td>1k</td>
</tr>
<tr>
<td>Face Net’15 (Google)</td>
<td>200 m</td>
<td>8 m</td>
<td>25</td>
</tr>
<tr>
<td>Fusion’15 (FB)</td>
<td>500 m</td>
<td>10 m</td>
<td>50</td>
</tr>
<tr>
<td>MegaFace’16</td>
<td>1.02 m</td>
<td>690.5 k</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Pose (yaw) Variations

Even with lots of resources, it’s hard to ensure sufficient intra-subject and pose variations.
The two keys to successful face recognition

1. **During training**: Learn the variability of same-subject appearances

   Increase training set intra-subject appearance variations

2. **During testing**: Make same subjects easier to compare

   Reduce test set intra-subject appearance variations
Domain (face) specific data augmentation

Increasing appearance variability in the training set

3D pose
Domain (face) specific data augmentation

Increasing appearance variability in the training set

3D pose

3D shape
Domain (face) specific data augmentation

Increasing appearance variability in the training set

3D pose

3D shape

Expression
Reducing appearance variability in the test set

CNN trained on augmented data

\[ S_{\text{real}} + S_{\text{rend}} = S \]
Reducing appearance variability in the test set

In case we have multiple frames from videos in a template (set of images)
Reducing appearance variability in the test set

- In case we have multiple frames from videos in a template (set of images)
- Each video track is pooled across frames in the feature space with average
- Pair-wise similarity scores are then pooled with Soft-Max operator

\[ s^* = \frac{\sum_{i=1}^{N} s_i \exp(\alpha s_i)}{\sum_{i=1}^{N} \exp(\alpha s_i)} \]
What does this do to performance?

Example: IJB-A

Training better CNNs with less effort using domain (face) specific data augmentation!!!
Do We Really Need to Collect Millions of Faces for Effective Face Recognition?

Iacopo Masi*,1, Anh Tuan Tran*,1, Tal Hassner*,2,3, Jatuporn Toy Leksut1 and Gerard Medioni1

Come see us at the poster (S-4B-09) or visit our webpage for more information, code and results

Thank you!