Deep Convolutional Neural Network with Independent Softmax for Large Scale Face Recognition

Yue Wu, Jun Li, Yu Kong, Yun Fu

Northeastern University, Boston, USA
Outline

• Task and Challenges
• Our solution
• Final Evaluation
• Conclusion
Task and Challenges

• Recognize **1M** celebrities

• A classification problem
  – Large number of classes
    • **100K** celebrities
  – Large number of images
    • **10M** images

Outline

• Task and Challenges
• Our solution
• Final Evaluation
• Conclusion
First Try

- ResNet 18-layer* model on full dataset
  - Convergence?
  - Training very slow: 4 months (estimated)

- ResNet-18 on 10K classes out of 100K classes
  - Training fast: 5 days
  - Top-5 error: 32%

*Deep Residual Learning for Image Recognition, CVPR 2016
What if we test faces not in 10K?

• 1000 images are tested
  – ‘scores’ is top-1 scores
  – ’in’ means faces in 10K
  – ’out’ means faces not in 10K
    • Almost a uniform output

• We can train several non-overlapped, independent models distributed in different machines.
Independent Softmax Models (ISM)

- **5 ResNet-18 models**
  - Training separately
  - **One week** to finished
  - Distributed
    - A single GeForce Titan X
    - Two GeForce 970

- **Classify a face**
  - Get scores for each class from 5 models
  - Concentrate all scores
  - Get the top-1 result for 100K classes

### 100K Partition Information

<table>
<thead>
<tr>
<th>Fold</th>
<th>Classes</th>
<th>Train Images</th>
<th>Val Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10,001</td>
<td>760,656</td>
<td>89,461</td>
</tr>
<tr>
<td>2</td>
<td>23,000</td>
<td>1,743,649</td>
<td>205,377</td>
</tr>
<tr>
<td>3</td>
<td>23,000</td>
<td>1,740,138</td>
<td>204,921</td>
</tr>
<tr>
<td>4</td>
<td>23,000</td>
<td>1,742,208</td>
<td>205,092</td>
</tr>
<tr>
<td>5</td>
<td>20,890</td>
<td>1,578,945</td>
<td>185,807</td>
</tr>
</tbody>
</table>
Result

• Development set
  – Dev1: hard set
  – Dev2: random set

• Measurement
  – For $N$ images, $M$ images are recognized, among which $C$ images are correct
    • precision = $C/M$
    • coverage = $M/N$
  – Coverage@Precision = 95% (C@P=95)
    • The higher, the better
• Multi-crop testing
  – Multi-crop testing usually give better result but is time-consuming

• Model ensemble
  – Fine-tuned from the 10K model for 3 epochs (SM)
  – Fusion
Outline

• Task and Challenges
• Our solution
• Final Evaluation
• Conclusion
Final Evaluation

Random set C@P=95

- Our: 0.734
- ITRC-SARI: 0.707
- CGIT_NLPR: 0.684
- ms37: 0.646
- 1510: 0.570
- FaceAll: 0.554
- faceman: 0.461

Hard set C@P=95

- CGIT_NLPR: 0.534
- Our: 0.486
- faceman: 0.330
- ms37: 0.260
- FaceAll: 0.254
- BUPT_PRIS: 0.210
Outline

• Task and Challenges
• Our solution
• Final Evaluation
• Conclusion
Conclusion and Thank You!

• Key components:
  – independent Softmax
  – Multi-crop testing
  – Model ensemble

• Public available codes and models
  – https://github.com/wuyuebupt/msceleb2016acmmm