The Video-Based Age and Gender Recognition with Convolution Neural Networks

Kharchevnikova Angelina
angelina.kharchevnikova@gmail.com
Outline

• Age and gender recognition problem
• Literature survey
• Proposed algorithm
• Experimental results and discussion
• Concluding comments and future plans
Age and gender characteristics can be applied:

- Retail for contextual advertising
- Video surveillance systems
- Photo processing applications
- Analysis of multimedia data group

The task of video classification is to assign the newly arriving (to the input) sequence of $T$ frames $\{X(t)\}, t=1,2,\ldots,T$, with the face of an individual to one of the $L$ classes

The reliability of the existing solutions remains insufficient for practical application
Age prediction:

- Choi, S.E.: Age estimation using a hierarchical classifier based on global and local facial features (2011)

Gender prediction:


**VGG-16** (500MB) (Rasmus Rothe, Radu Timofte, Luc Van Gool DEX: Deep EXpectation of apparent age from a single image 2015)
Proposed algorithm

Video camera

Frame selection → Face detection

Application of classifier committees → Recognition using CNN

Probabilities

Final decision

Reference database
Aggregation

The output of the CNN - the Softmax layer

\[
P(l|X(t)) = \text{softmax } z_l(t) = \frac{\exp z_l(t)}{\sum_{j=1}^{L} \exp z_j(t)}, \quad l = 1, 2, \ldots, L
\]

Simple voting

\[
l^* = \arg \max_{l=1,L} T \sum_{t=1}^{T} \delta(l^*(t) - l)
\]

Arithmetical mean

\[
l^* = \arg \max_{l-1,L} \frac{1}{T} \sum_{t=1}^{T} P(l|X(t))
\]

The geometric mean

\[
l^* = \arg \max_{l-1,L} \prod_{t=1}^{T} P(l|X(t)) = \arg \max_{l=1,L} \sum_{t=1}^{T} \log P(l|X(t))
\]

An expected value (mathematical expectation)

\[
l^* = \sum_{l=1}^{L} l \cdot P(l|X(t))
\]
Experimental results and discussion

[Image of gender/age recognition software with two images: one showing a male with a gender classification of 'MALE' and age '35', and another showing a female with a gender classification of 'FEMALE' and age '18'.]
Intel Core i5-2400 CPU, 64-bit machine with NVIDIA GeForce GT 440
Inference time

- OpenCV dnn: 4,805 seconds
- OpenCV dnn+mvn: 8,734 seconds
- Caffe+mean img deduction: 34,984 seconds
- Caffe+mvn: 9,012 seconds

Categories:
- Gender_net/Age_net
- VGG-16
Datasets

IARPA Janus Benchmark A (IJB-A): 2043 video, 13900 frames + gender information

Kinect: 104 video, 936 frames + gender and age information
Datasets

Indian Movie: 332 video, 28312 frames + gender and age information
Gender recognition accuracy with classifier fusion

**Gender_net**

- IJB-a: 60, 59, 59
- Indian Movie: 71, 72, 75
- Kinect: 73, 75, 77

**VGG-16**

- IJB-a: 81, 81, 82
- Indian Movie: 81, 87, 88
- Kinect: 84, 84, 84

Accuracy in %

- Simple voting
- Sum rule
- Product rule
Age recognition accuracy with classifier fusion

**Age_net**

<table>
<thead>
<tr>
<th></th>
<th>Simple voting</th>
<th>Sum rule</th>
<th>Product rule</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>23</td>
<td>23</td>
<td>31</td>
<td>41</td>
</tr>
<tr>
<td>31</td>
<td>41</td>
<td>43</td>
<td>45</td>
<td>50</td>
</tr>
</tbody>
</table>

**VGG-16**

<table>
<thead>
<tr>
<th></th>
<th>Simple voting</th>
<th>Sum rule</th>
<th>Product rule</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>62</td>
<td>65</td>
<td>65</td>
<td>63</td>
<td>73</td>
</tr>
<tr>
<td>73</td>
<td>79</td>
<td>79</td>
<td>78</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

• The geometric mean (product rule) with normalization of the input video images is the most accurate in gender classification task
• The most accurate age prediction is achieved with the computation of the expected value
• The accuracy of the VGG-16 architecture is about 15% and 20% higher for the gender recognition and age prediction than Age and Gender net models
• The inference time of the VGG-16 is 4-9 times lower

Future work

• Implementation the gender/age recognition in Android mobile offline application
• Applying the modern techniques for fast classification and optimization of deep CNNs
Thank you for your attention!