Transfer learning and embedding

Tambet Matiisen, Ardi Tammpuu, Konstantin Tretjakov, Tanel Pärnamaa

29.04.2015
Labeled Faces in the Wild

13233 images of 5749 people.
Transfer learning experiments

<table>
<thead>
<tr>
<th>Training set size</th>
<th>Test set size</th>
<th>Number of classes</th>
<th>Min images per class</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFW dataset</td>
<td>4913</td>
<td>1072</td>
<td>423</td>
</tr>
<tr>
<td>Fotis dataset</td>
<td>815</td>
<td>179</td>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NN accuracy</th>
<th>kNN accuracy with LFW features</th>
<th>kNN accuracy with Fotis features</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFW dataset</td>
<td>70%</td>
<td>61%</td>
</tr>
<tr>
<td>Fotis dataset</td>
<td>~61%</td>
<td>54%</td>
</tr>
</tbody>
</table>

- **NN training:**
  - Caffe + NVidia Tesla K20
  - CIFAR-10 inspired network
  - ~1 hour training time

- **kNN parameters:**
  - L2 distance
  - 5 nearest neighbors
  - Voting weighted by distance
Interactive portals

To find your celebrity matches, please upload an image of your face, or specify a URL:

[Form for uploading image or specifying URL]

Selleks et leida oma teisiku Eesti Rahvusarhiivist, lae üles oma pilt või kopeeri URL:

[Form for uploading image or specifying URL]
TODO

- NN hyperparameter tuning
  - Possibly with [www.whetlabs.com](http://www.whetlabs.com)
- Try different network architecture
  - Siamese network, Google FaceNet
- Try different pre- and postprocessing
  - Histogram equalization, L2-normalization
- Try stacked autoencoder
  - Learn unsupervised from bigger dataset.
- Try with more datasets
  - YouTubeFaces, Weakly Labeled Faces etc
Siamese network

Uses two images as input and drives same class instances close and different class instances apart in L1-space.

Google FaceNet: network architecture

- Zeiler&Fergus model:
  - 22 layers
  - 140M parameters
  - 1.6B FLOPS

- Inception model:
  - 17 layers
  - (11 inception modules)
  - 7.5M parameters
  - 1.6B FLOPS
Google FaceNet: triple loss

Uses three images as input: anchor, same class instance, other class instance. Drives same class instance close and other class instance away in L2-distance.

\[
L = \sum_{i}^{N} \left[ \| f(x_i^a) - f(x_i^p) \|_2^2 - \| f(x_i^a) - f(x_i^n) \|_2^2 + \alpha \right]
\]

Illumination and pose invariance

Threshold distance < 1.1 would classify all cases correctly.

False accept

Accuracy $98.87\% \pm 0.15$ with center crop and $99.63\% \pm 0.09$ with extra alignment.

False reject

Just one more iteration...

Thanks!

tambet@ut.ee