Image Data Research Notes

04/18/2017

**Best Dataset so far**:

[IMDB-WIKI 500K face images with age and gender labels](https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/)

The team crawled 0.5 million images of celebrities from IMDB and Wikipedia and make the data available on the website.

This blog also provide their DEX(Deep Expectation) model to download, the model is build with Caffe, and it is a CNN uses the VGG-16 architecture and pre-trained on Image-net.

The raw images size are quite big, about 270 gbs, a better version to download is their cropped faces version with 40% margin, this version is only 7 gbs for IMDB images, WIKI is 1 gb.

The age estimation the output layer ha 101 neurons(0-100 years) one to each year, and gender estimation layer has 2 neurons(0 for female, 1 for male).

**Second best dataset:**

<http://faculty.chicagobooth.edu/bernd.wittenbrink/cfd/download/download.html>

1.5 GBs with very clear, no background noise, id photo style front shoot, labels of age and gender is provided, con is only 600 unique people images in total, each person is 4 photos; as a general rule of thumb is for DL, better has at least a few thousands examples per class.

**Other datasets:**

<http://www.wilmabainbridge.com/facememorability2.html>

10k US adult faces db, download request has been sent to the owner, waiting for approval

<http://www.face-rec.org/databases/>

<http://web.mit.edu/emeyers/www/face_databases.html>

<https://www.kairos.com/blog/60-facial-recognition-databases>

<http://vis-www.cs.umass.edu/lfw/>

this is a wild image set which has no label at all, thus not recommended.

<http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html>

CMU face db, 750k images of 337 people, but need to pay to get delivered in a hard drive,

Since the data size is huge: 308 gbs.

<http://www.jdl.ac.cn/peal/index.html>

China Science Academy Face DB: close to 1 million pictures with good label