Gender classification by Deep Learning on mobile device

Xiaofei Fu (stevenfu@stanford.edu)

Department of Electrical Engineering, Stanford University

Motivation

Powerful mobile devices with accessibility features had already helped visually impaired people a lot.

By building a face detection and gender classification app running completely offline on Mobile device, we hope to use latest computer vision techniques to further aid these people to ease social interactions.

For second part we also build a function to represent picture by synthesise sound to represent the RGB value of touch point. By scrubbing finger across screen and listen to sound changes, one can roughly tell the shape of object presented on screen.

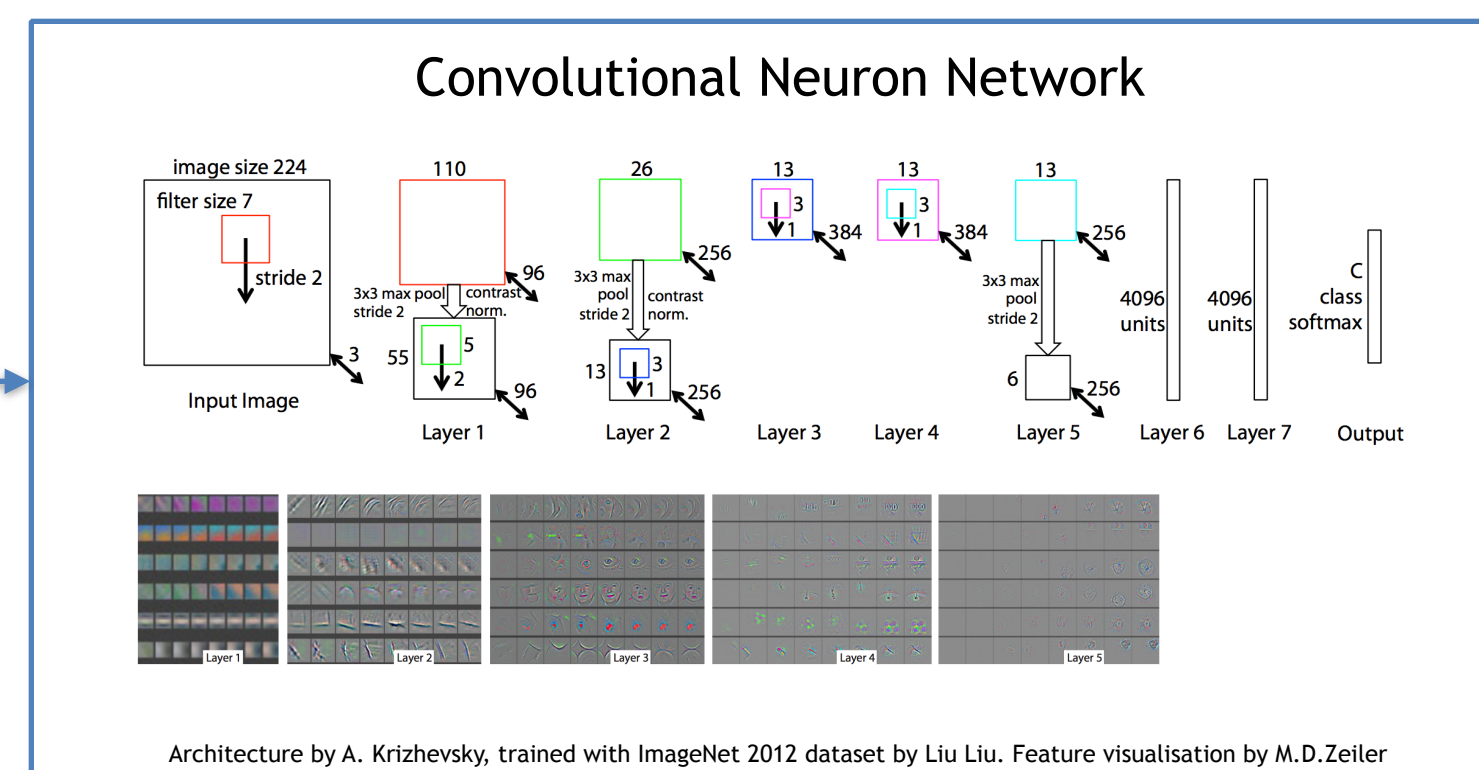## Gender classification process

Camera capture

Crop head and shoulder

Feature extraction by CNN

Face detection

Convolutional Neuron Network

Architecture by A. Krizhevsky, trained with ImageNet 2012 dataset by Liu Liu. Feature visualisation by M.D.Zeiler.

1000 Dimension Feature Vector (actually ImageNet object IDs)

2070x Feature Vectors, manually labeled gender

Logistic Regression

Training (on PC)

coefficiency vector + intercept
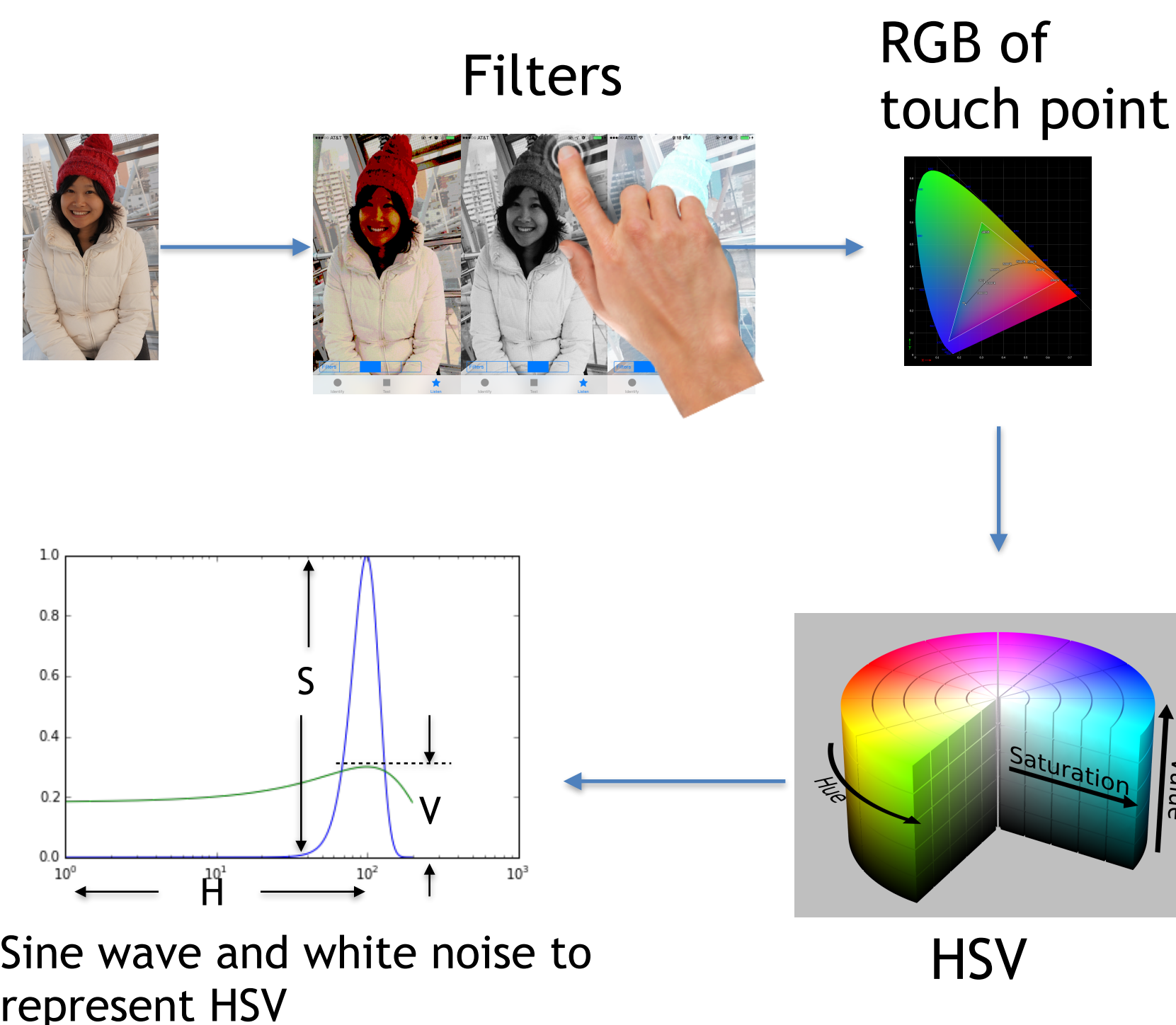
Prediction (Mobile)

coefficiency vector

.X   +   intercept   =   Score

>0 Female

<0 Male

Above process can also support multi-class classification, though not tested yet for this time.

Sound representation of picture

Filters

RGB of touch point

HSV

Sine wave and white noise to represent HSV

Results

We trained the Logistical Regression with CNN feature vectors of 1662 pictures(1123 male, 579 female) that was downloaded from ImageNet dataset and a stock photo website.

From these pictures we cropped 2070x(1111 male, 959 female) head and shoulder regions by face detection and rotation with face angle. We also cropped all face regions from them.

We experimented with different combination of training vector count, vector length, also experimented training with full picture, head and shoulder corp, and face crop, and found "head and shoulder" crop gives best results. We have ~**94%** accuracy when **validating** with training data.

Then we tested our result with a new dataset(1090 pictures in total) that were not used in the training. Due to the randomness inherited from CNN, different test runs show different results.

On average we have **68-78%** accuracy of **predicting** gender correctly, interestingly we have around **68%** accuracy predicting back figure **without seeing face at all**.

(Heading to App Stores in your palm, soon!)
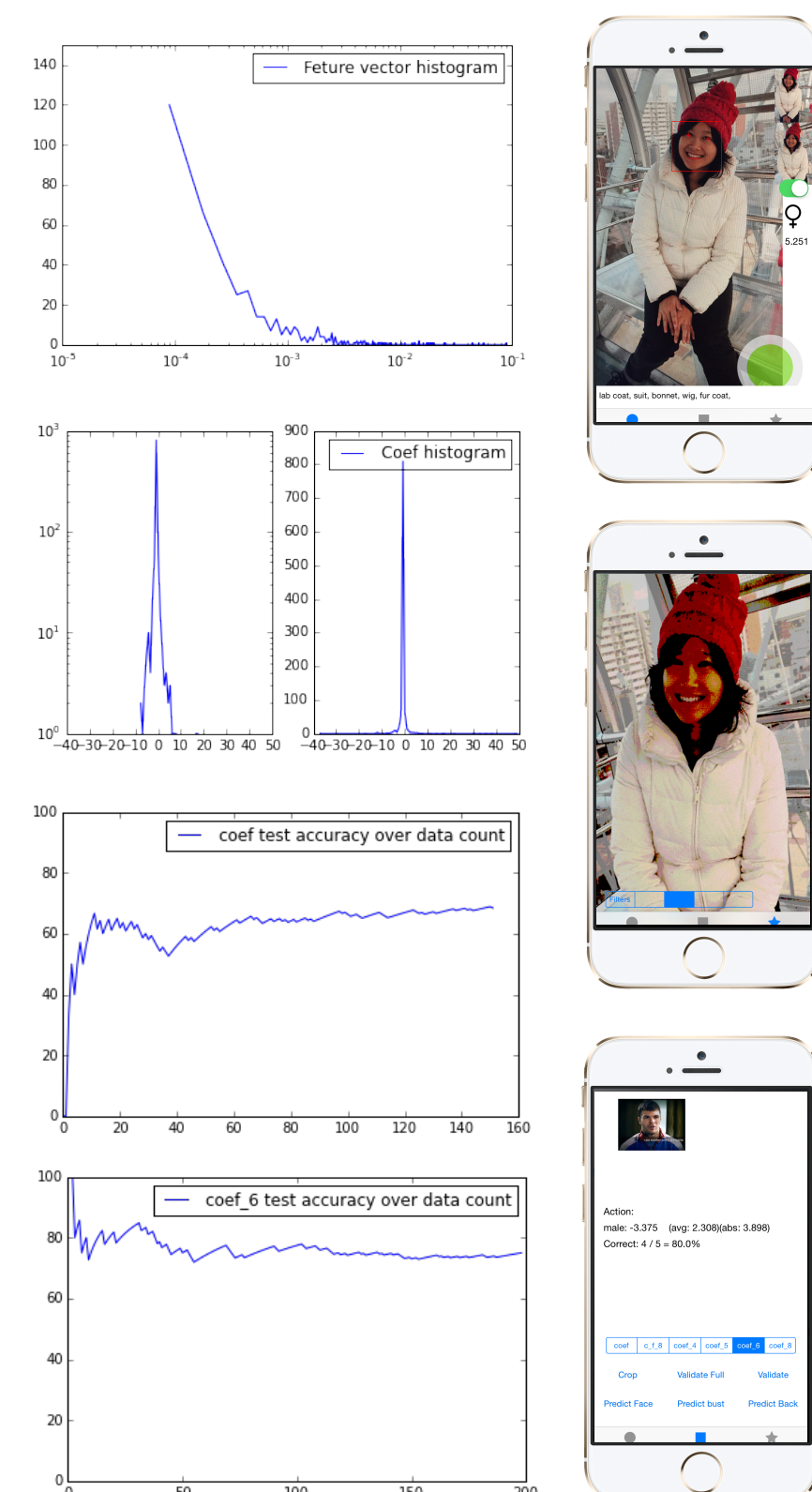
Thanks!

Professor Girod and Wetzstein for a wonderful course.
TA Chen and Boin for helps on everything.

References

[1]Yunchao Gong, Liu Liu, Ming Yang, Lubomir Bourdev. Compressing Deep Convolutional Networks Using Vector Quantization.
[2]Zeiler, Matthew D and Fergus, Rob. Visualizing and understanding convolutional neural networks. *arXiv preprint arXiv:1311.2901*, 2013.
[3]Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 2012
[4]LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., and Jackel, L.D. Handwritten digit recognition with a back-propagation network. *NIPS*, 1990.
[5]libccv.org
[6]developer.apple.com/