

# DataFace

# A Facial Recognition Joint

State of the art facial recognition has surpassed the accuracy of human beings:

- Google's FaceNet Accuracy Score = .9964
- Facebook's DeepFace Accuracy Score = .9735
- Baidu (Minwa) Accuracy Score = .9977
- Humans -- AUC Score = .995

Benchmarked using the LFW Dataset (labeled faces in the wild)

Deep learning convolutional neural networks were used and trained with:

- Facebook used ~4.4 million images
- Google used 260 million images of over 8 million unique individuals

# Questions:

- What kind of results are achievable using publicly available resources?
- Are these results valuable?
- What's next?

# Data:

- 4,324 images (only 3,218 made it out of neural net)
- 12 different people
- > 2.5 minutes of total video

## Summary:

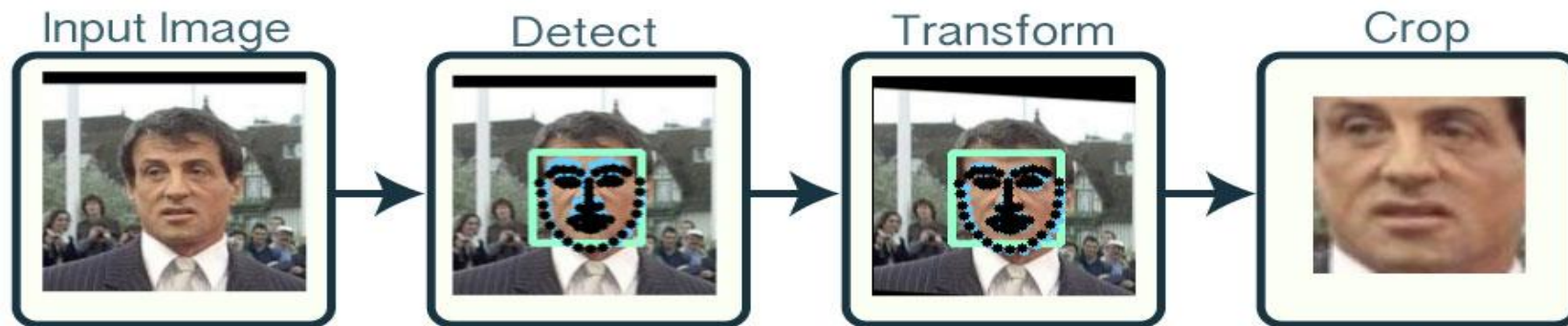
--Colin Clemence - 169 images/~5 sec, --John Marin - 476 images/~11 sec, --Leslie Pham - 129 images/~2 sec,

--Lucy Smoot - 311 images/~10 sec, --Michael Gat - 38 images/~3 sec, --Mike Frantz - 270 images/~6 sec,

- Mike Ludwig - 244 images/~4 sec, --Paul Trichon - 571 images/~17 sec, --Pauline Chow - 567 images/~9 sec,

- Robbie Smoot - 1,203 images/~22 sec, --Roshanak Omrani - 178 images/~3 sec, --Ryan Gin - 296 images/~5 sec

# Video & Image Pre-processing:



Example of 1 frame



Affine transformation on 4 consecutive frames

# OpenFace's Convolutional Neural Net:

Spatial convolutions

ReLU's

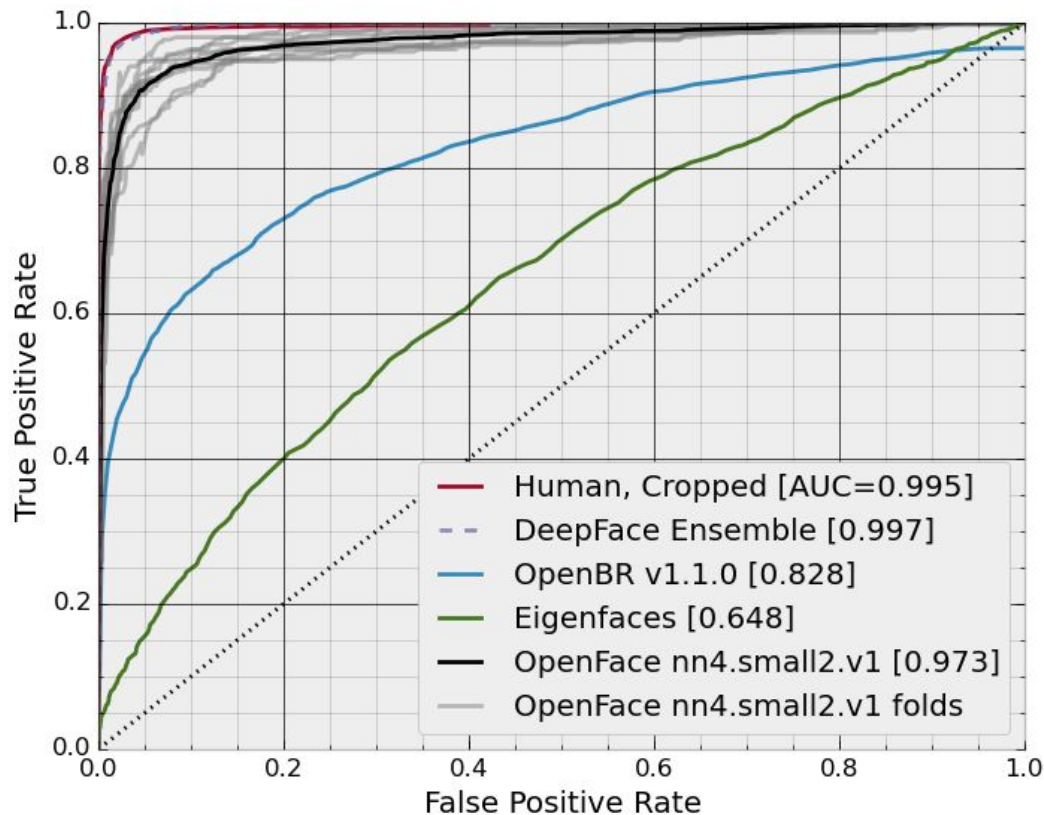
Pooling

Triplet loss function

FaceNets Inception model

Trained with 500K images

Implemented using Torch



# Embeddings → PCA & SVM:

Generated embeddings (3,218 X 128) of Low-dimensional face representations

12 different classes (people to predict)

Difficult to say what the embeddings mean

Embeddings are a value between -1 & 1

# High dimensional visual representation of clusters of embeddings using t-SNE in 2D

20

15

10

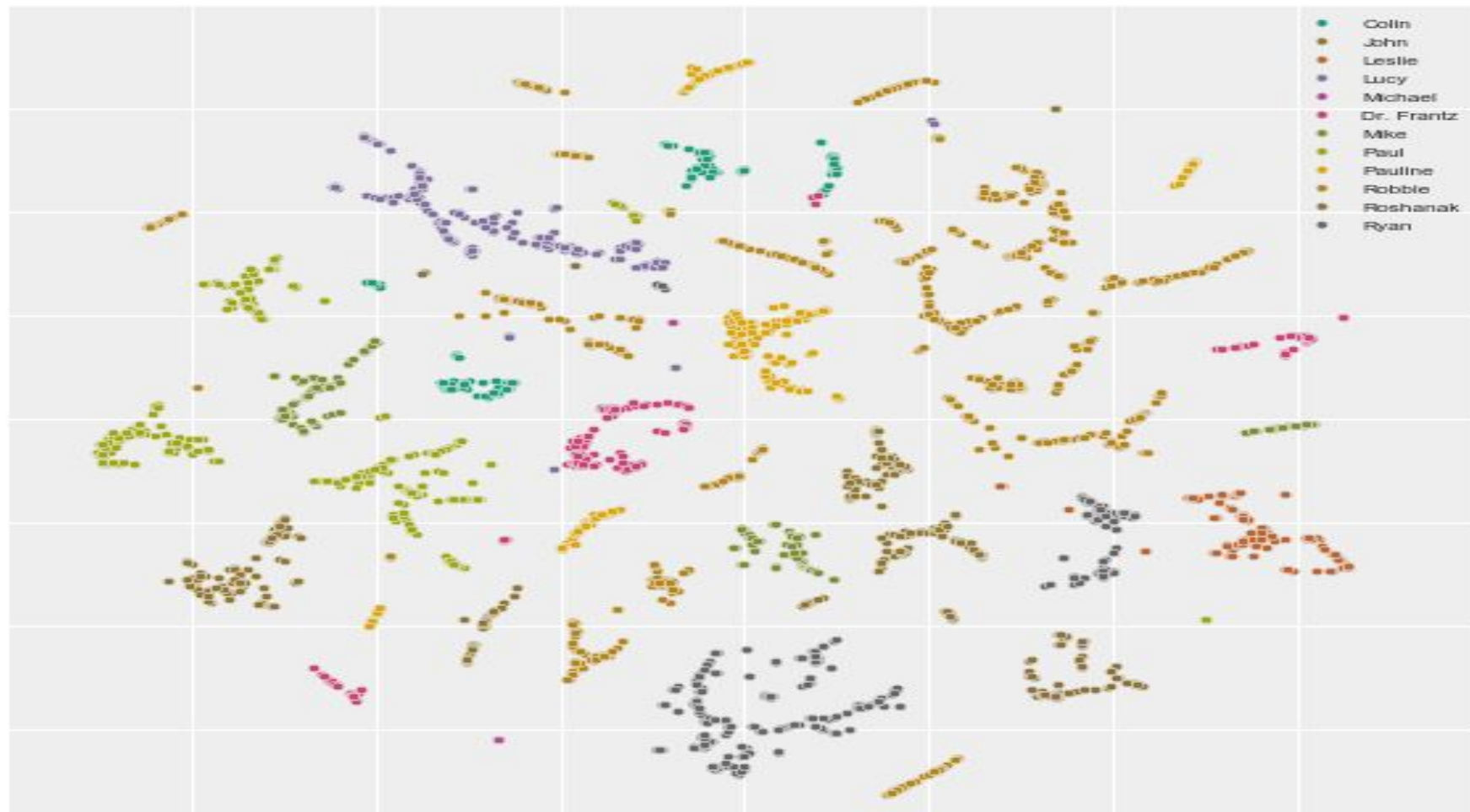
5

0

-5

-10

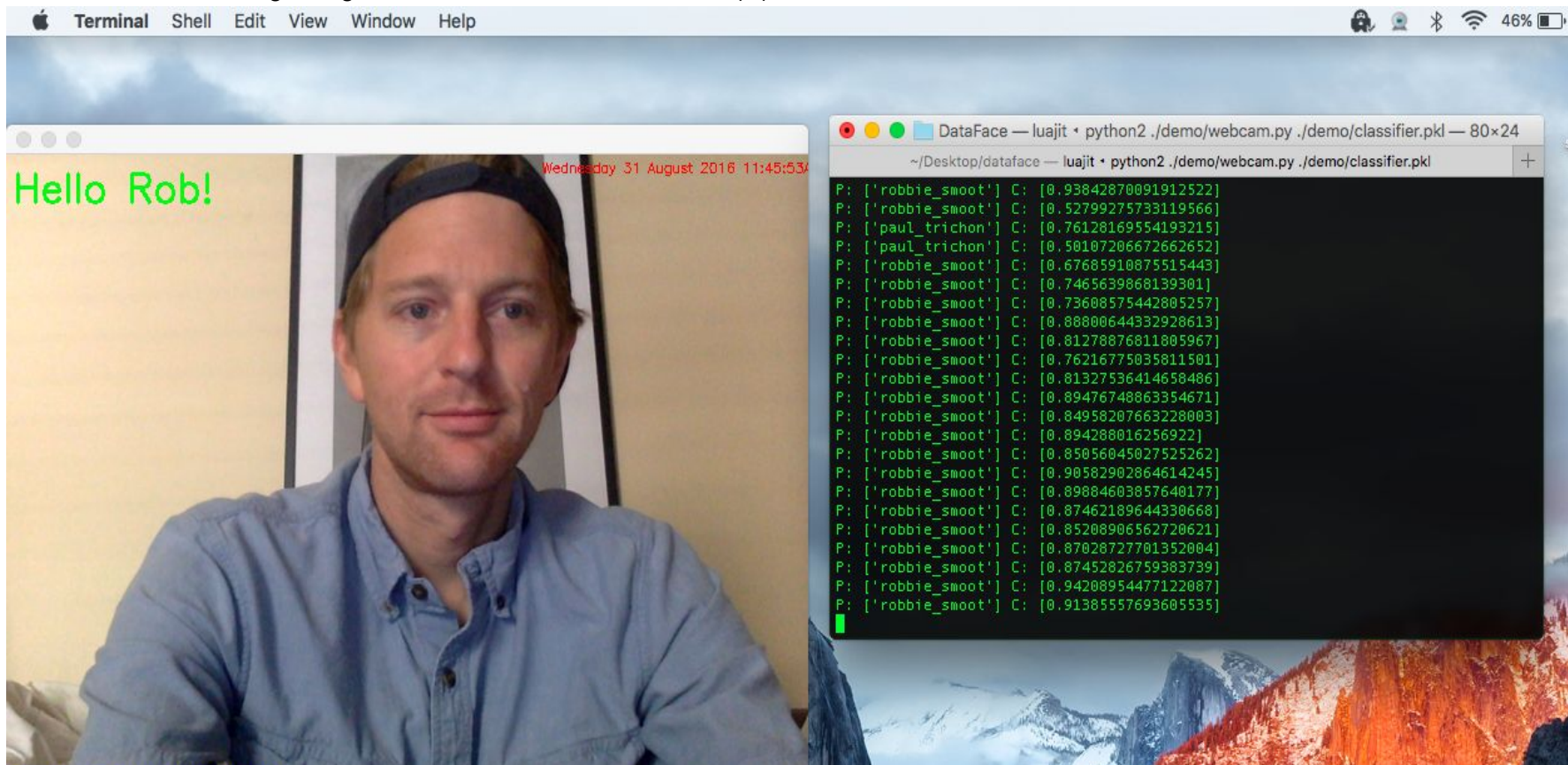
-15





# Demonstration:

Not really able to demonstrate this in Google slides so here is a screenshot. It recognizes me and says hello (L). There is a confidence interval regarding the classification for each frame (R).



# VALUE ?

These results are pretty good, right? It's difficult to judge since we don't really have a test and the model was trained on particular data. Performance is highly dependent on training input. When using less than 3 seconds of video I was able to get good results (unable to really quantify at this point). A lot more testing and analysis needs to go into this to find out how to best maximize results.

# What's next?

- Imagine if you walked into your local coffee shop and the barista knew your name and knew your usual order--and this barista was able to that for EVERY customer. This is the direction I plan to take DataFace. There are a lot of obstacles to getting there but there is much joy in the journey.

# Supervised vs. Unsupervised Learning:

- Facial recognition is still a supervised classification problem requiring labels
- How can we minimize the need for labels?

- pseudo labels

- temporary labels

- users input

# Acknowledgments:

★ Dr. John Marin, Pauline Chow, Mike Ludwig and everyone @ GeneralAssembly

★ OpenFace @ <https://github.com/cmusatyalab/openface>

★ Dlib @ <https://github.com/davisking/dlib>