

# Face Recognition IV

CSE 40537/60537 Biometrics

**Daniel Moreira**  
Spring 2022



# Today you will...

*Get to know*  
Deep-learning-based face recognition.

# Feature Extraction

**RECAP**

## Focus

2D-appearance-based methods.



## Types

Handcrafted features from Computer Vision.

Data-driven learned features from Machine Learning.

# Feature Extraction

**RECAP**

## Focus

2D-appearance-based methods.



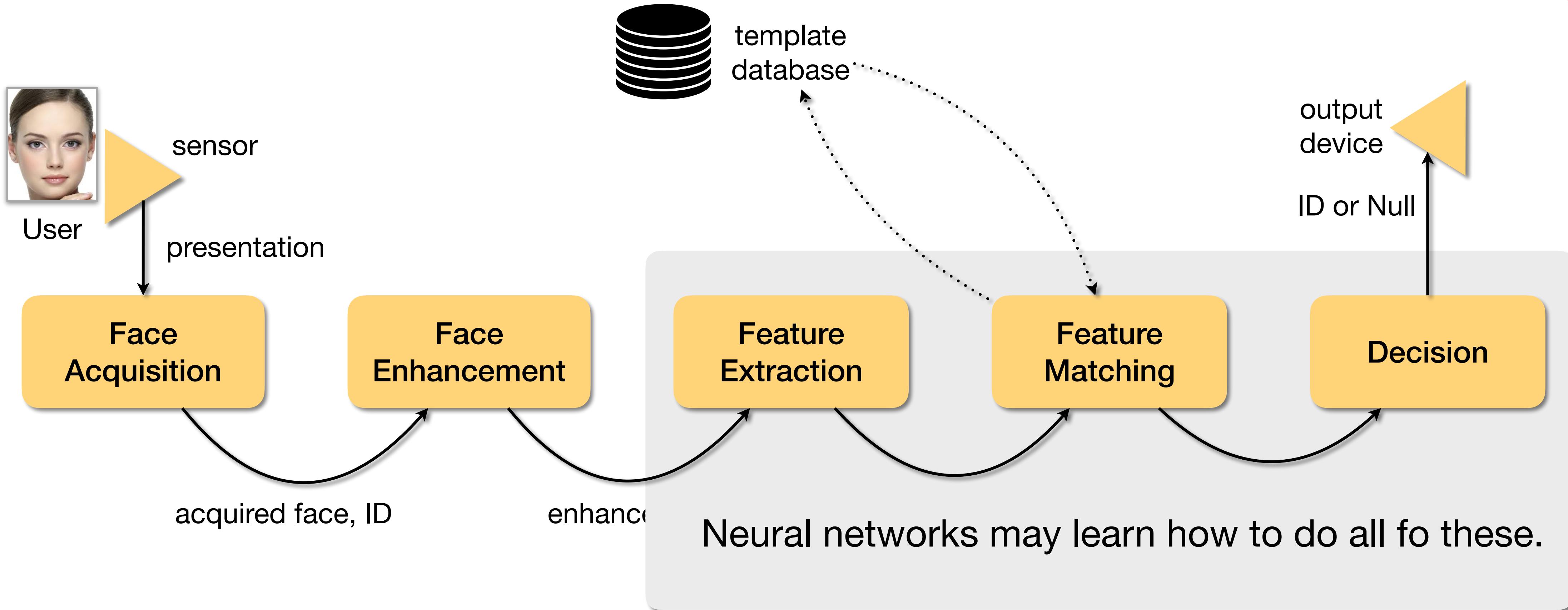
## Types

Handcrafted features from Computer Vision.

**Data-driven learned features from Machine Learning.**

# Face Recognition

**RECAP**



# Data-Driven Face Recognition

Deep Convolutional Neural Networks (CNN)

NON CAP

# Data-Driven Face Recognition

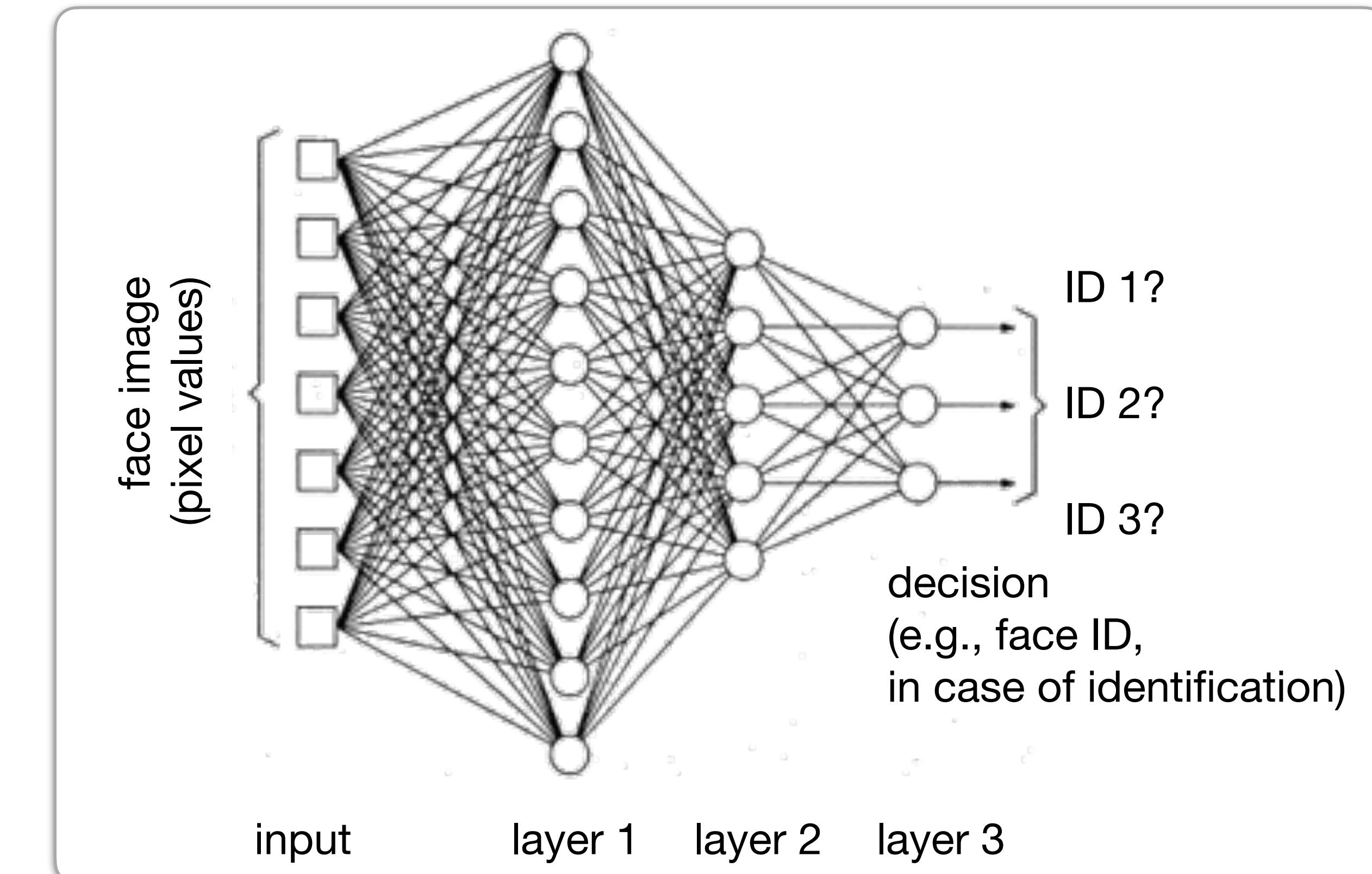
~~Non-CAP~~

## Deep Convolutional Neural Networks (CNN)

From pixels to classification decision.

Hierarchy of feature extractors.

Each layer extracts features from previous layer.



# Data-Driven Face Recognition

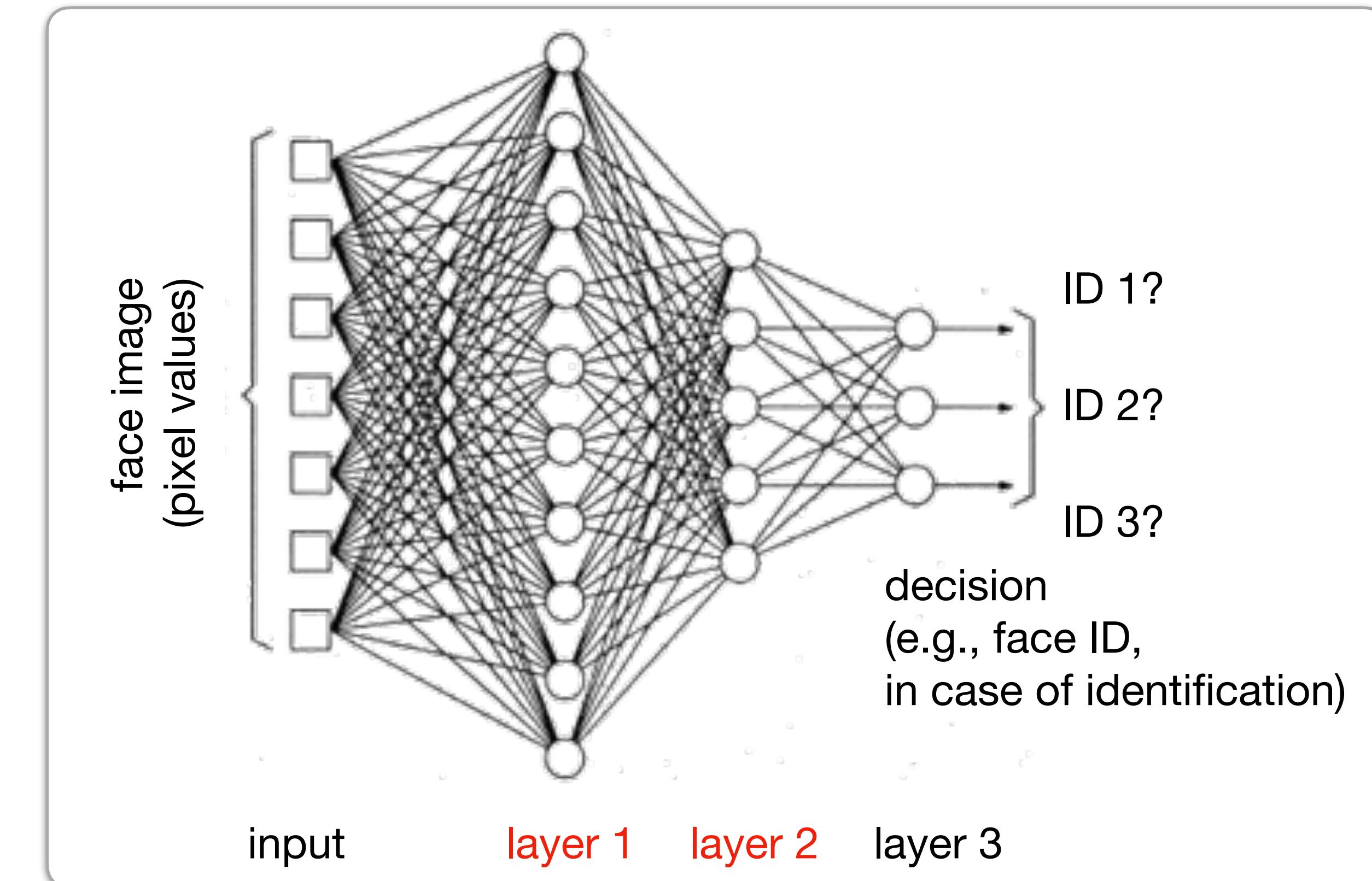
~~Non-CAP~~

## Deep **Convolutional** Neural Networks (CNN)

### Convolutional Layers

E.g., layers 1 and 2.

Feature extractors are convolutional operations which are performed on the output of the previous layer.



# Data-Driven Face Recognition

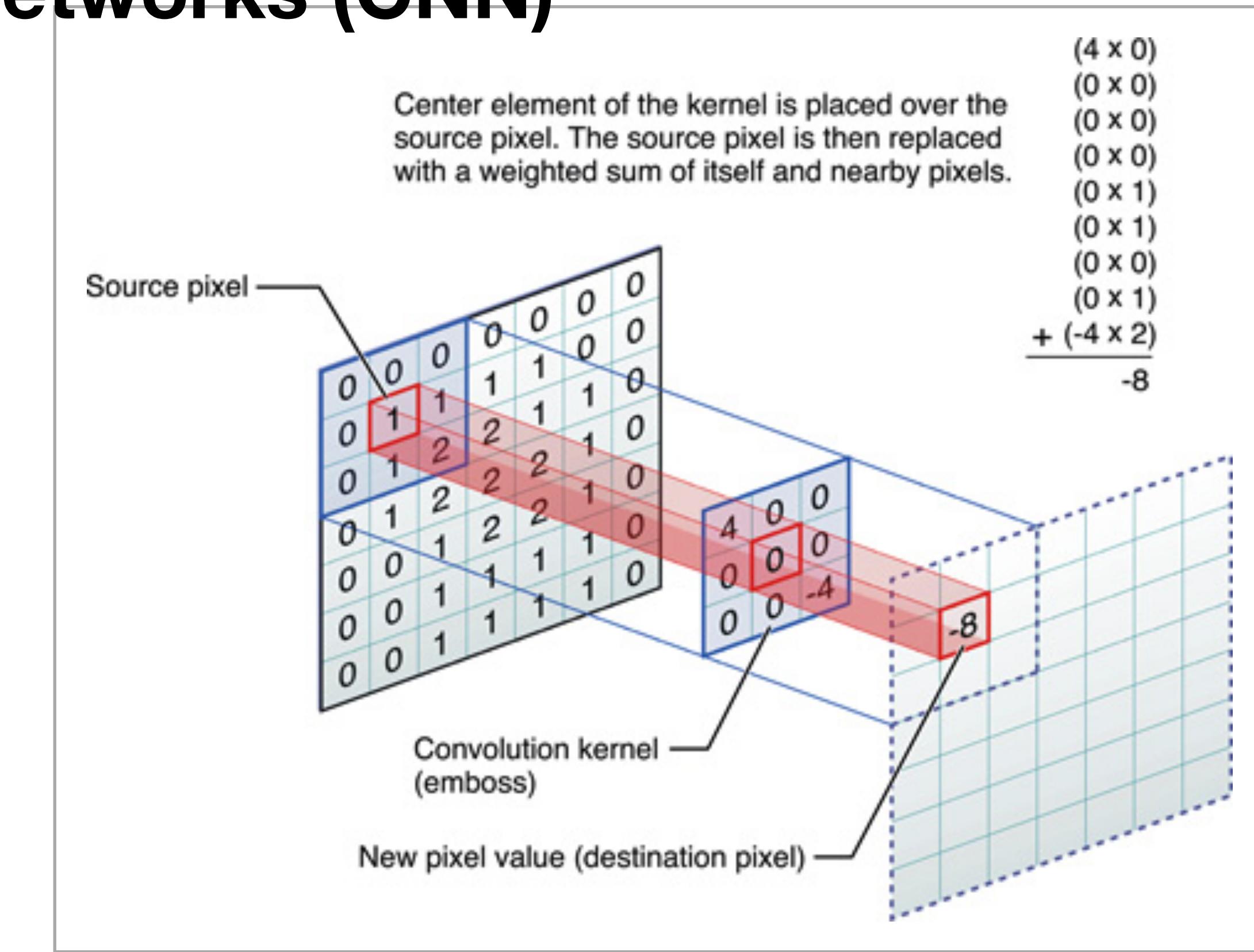
~~Homework CAP~~

## Deep **Convolutional** Neural Networks (CNN)

### Convolutional Layers

E.g., layers 1 and 2.

Feature extractors are convolutional operations which are performed on the output of the previous layer.



Source:<https://developer.apple.com/library/archive/documentation/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html>

# Data-Driven Face Recognition

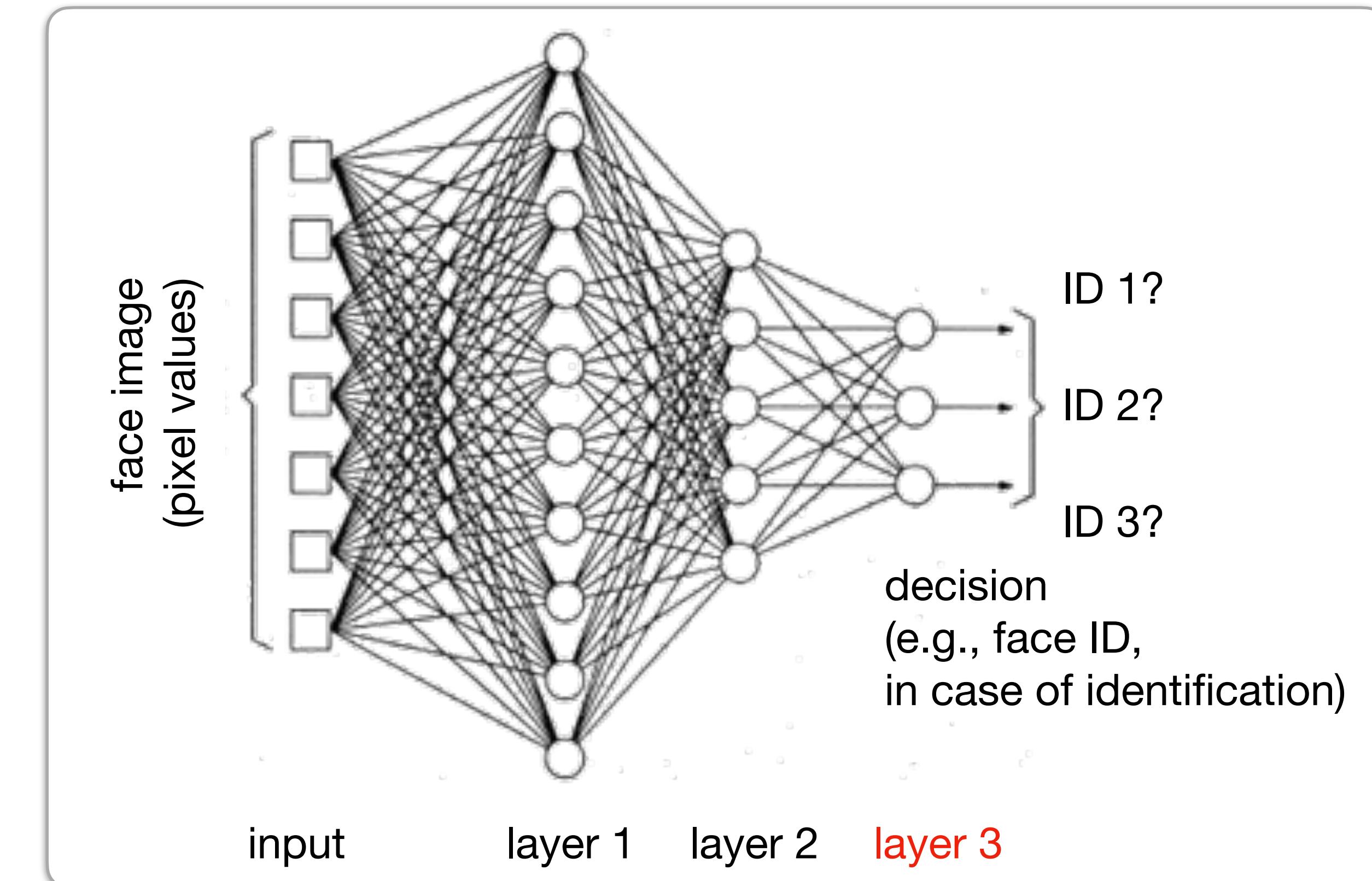
~~Homework~~

## Deep **Convolutional** Neural Networks (CNN)

### Fully Connected Layer

E.g., layer 3.

It performs the classification, presenting one score output for each class (identity, in the case of Biometrics).



# Data-Driven Face Recognition

~~Non-CAP~~

## Deep Convolutional Neural Networks (CNN)

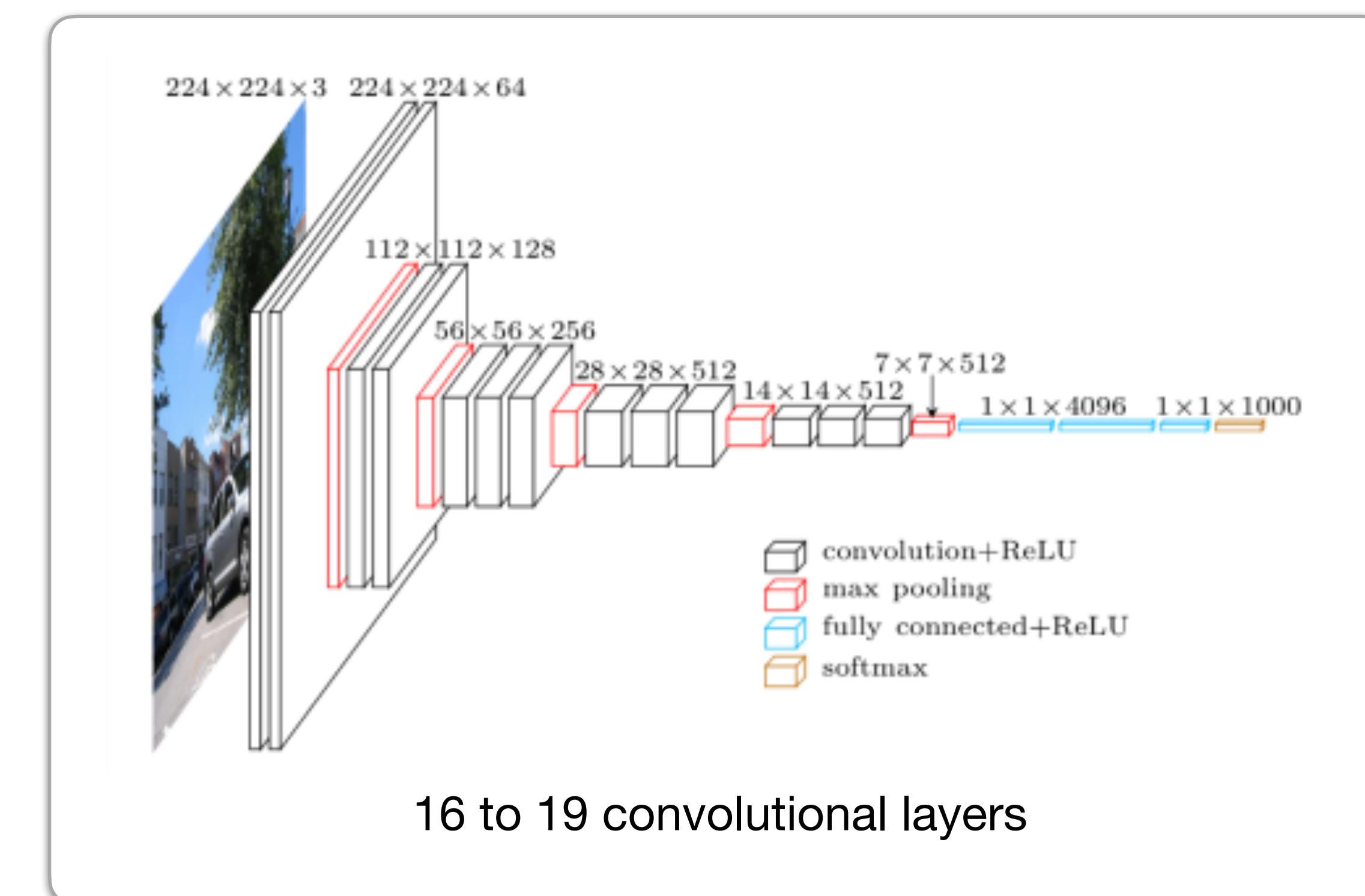
**How deep can they be?**

“Deep” refers to the number of layers.

E.g., VGG16

Simonyan and Zisserman

*Very Deep Convolutional Networks  
for Large-Scale Image Recognition*



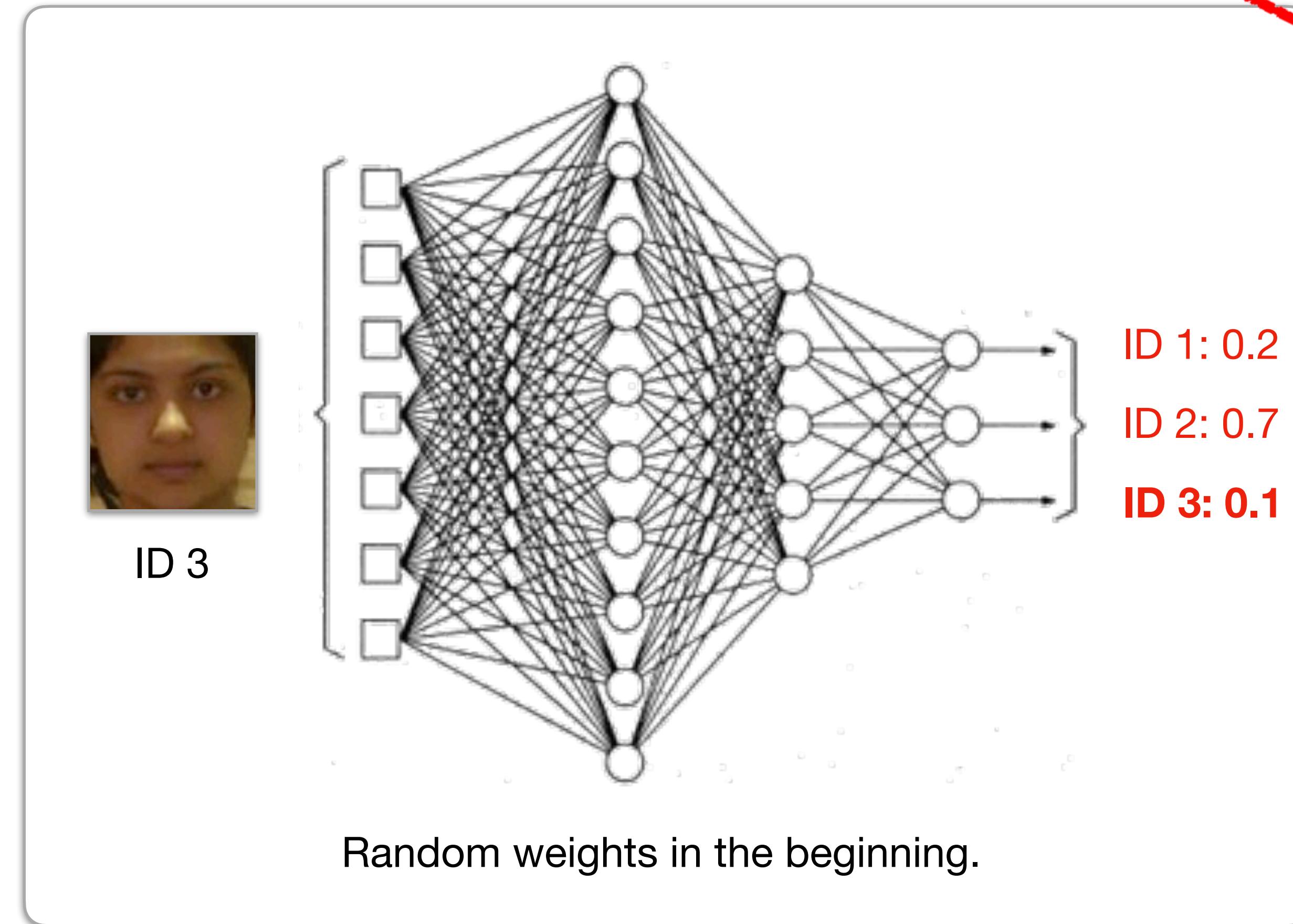
# Data-Driven Face Recognition

~~Homework~~

## Deep Learning

### Training

Labeled examples  
(e.g., faces and expected IDs)  
are used to teach the network  
to classify them correctly.



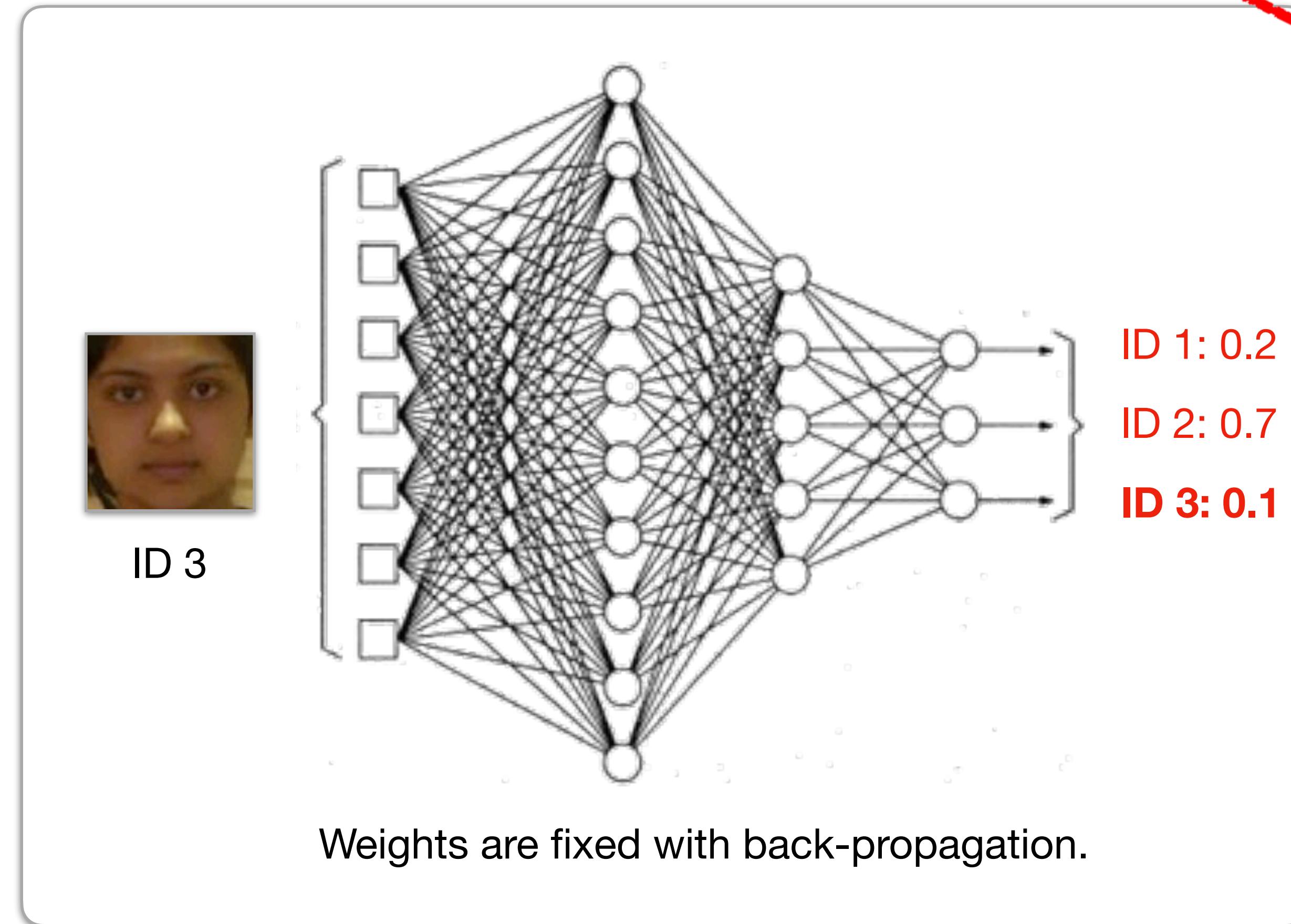
# Data-Driven Face Recognition

~~Face CAP~~

## Deep Learning

### Training

Back-propagation is used to fix the weights of the convolutions within the network.

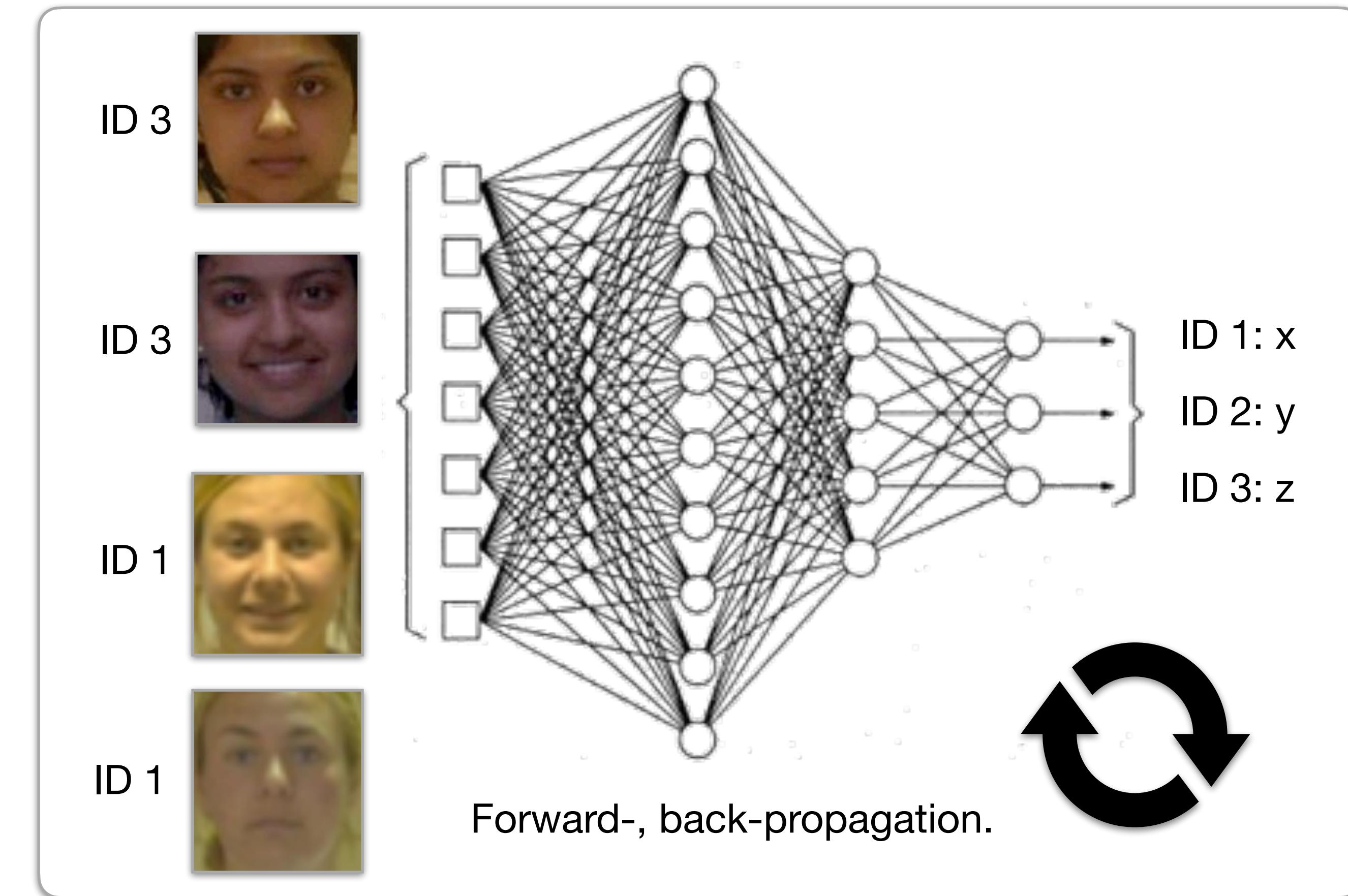


# Data-Driven Face Recognition

~~Homework CAP~~

## Deep Learning

Present various examples of each class and perform forward-, back-propagation.

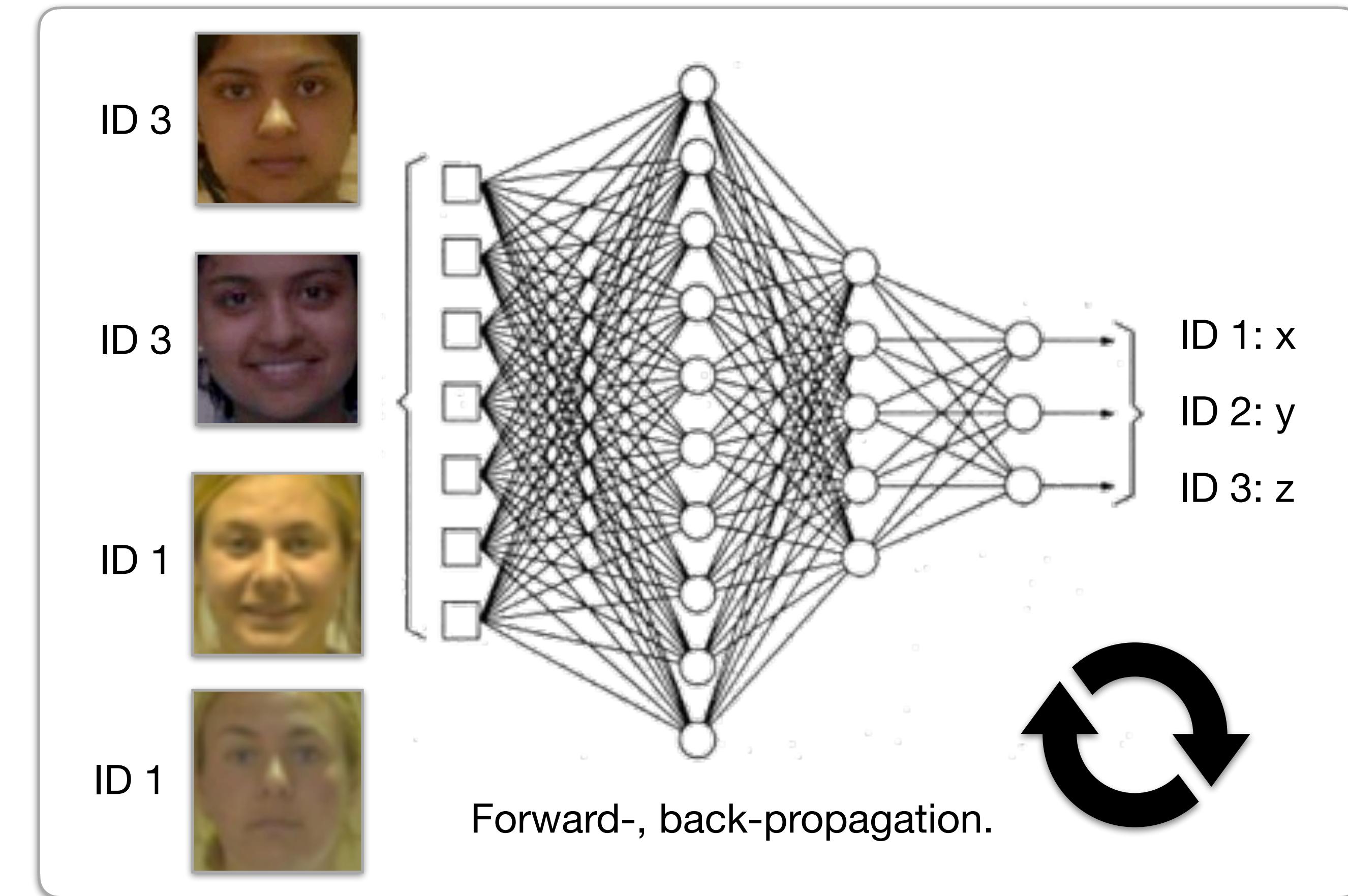


# Data-Driven Face Recognition

## Deep Learning

Optimization target:  
minimize classification  
error through  
**loss function.**

Popular function:  
**cross-entropy loss.**



# Data-Driven Face Recognition

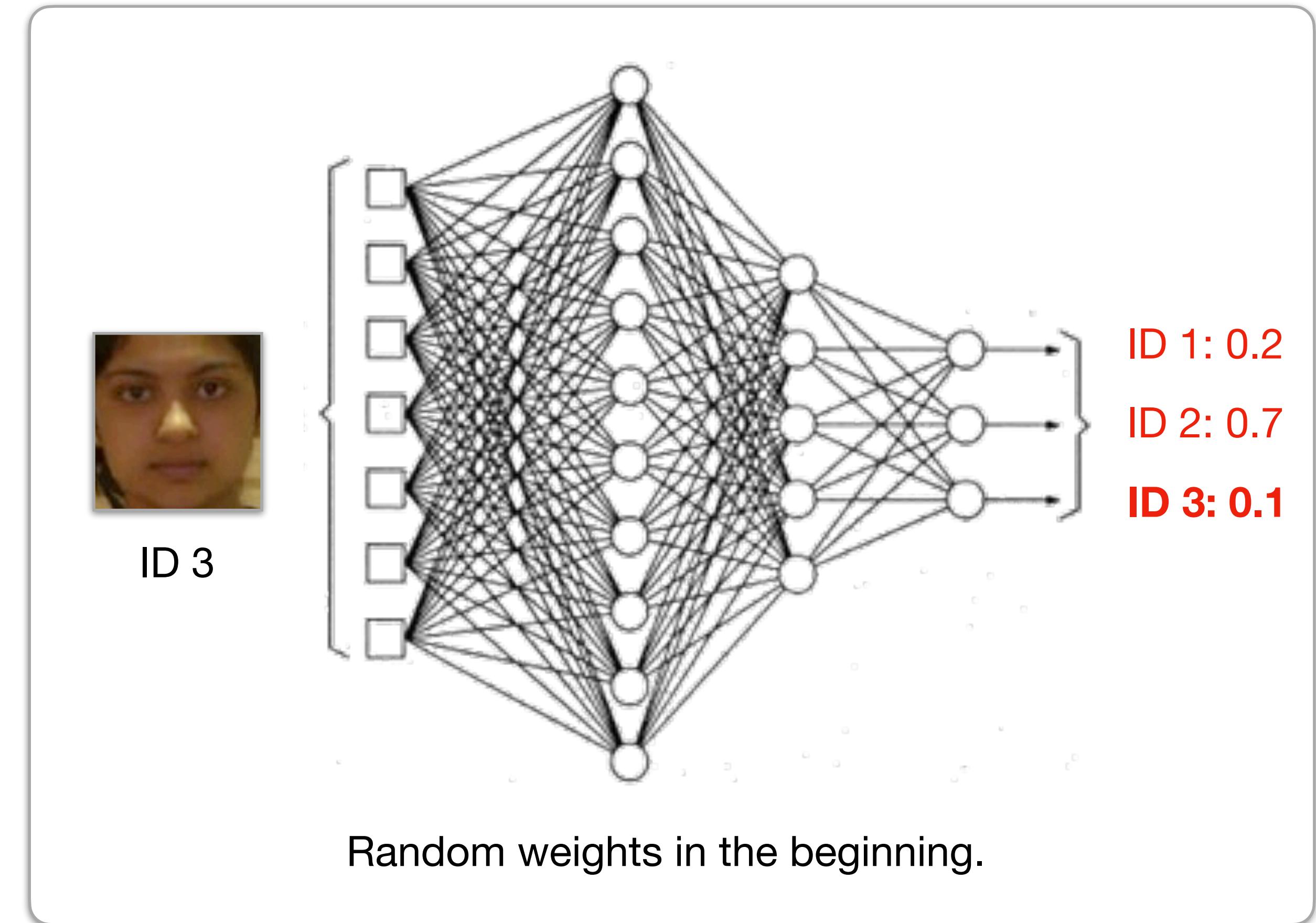
## Deep Learning

### Cross-entropy Loss (CE)

$$CE = \sum_{face=1}^m \sum_{ID=1}^n (-\log(output(ID)))$$

CNN output for ID

#training faces      #people's IDs



# Data-Driven Face Recognition

# Deep Learning

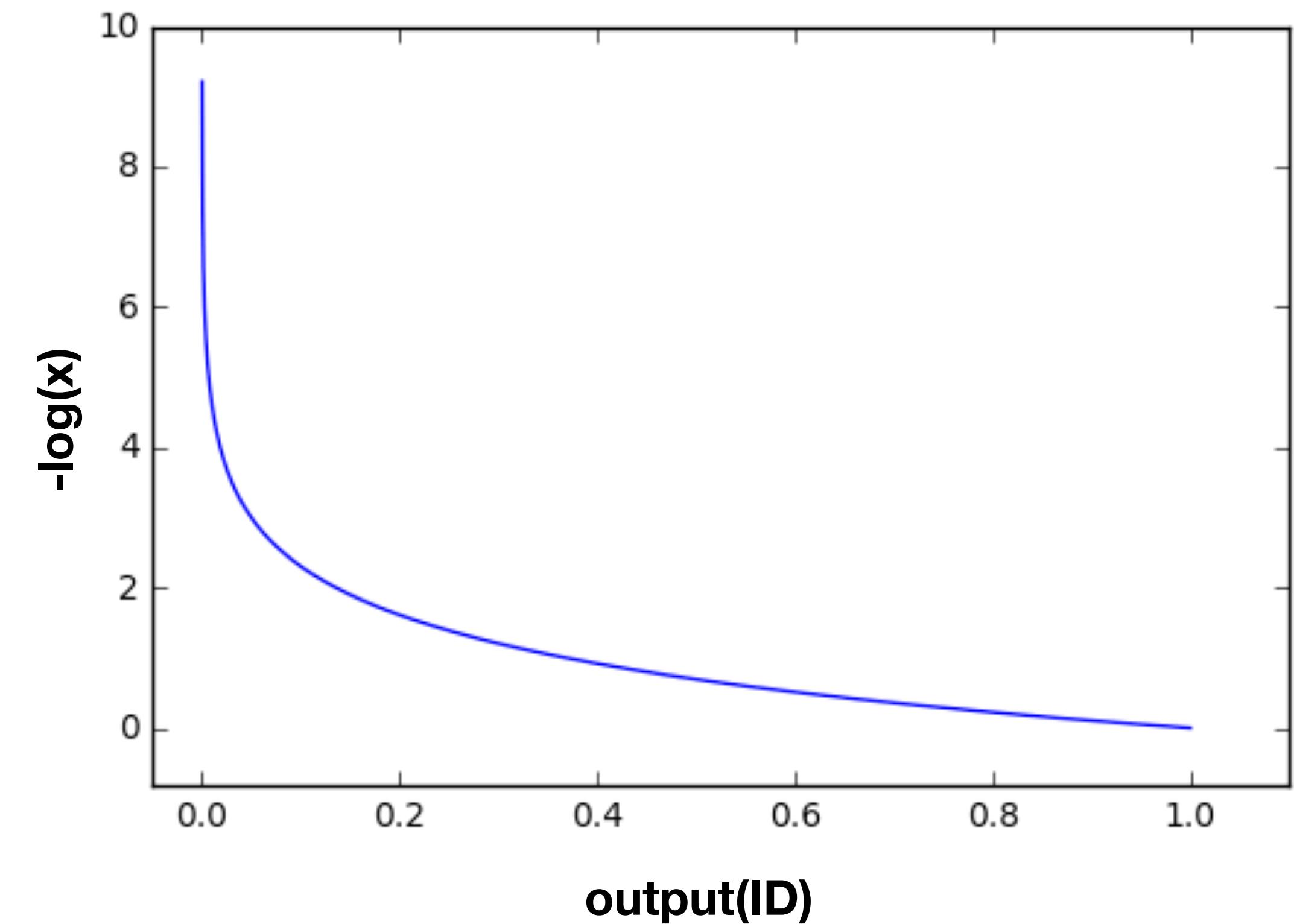
# Cross-entropy Loss (CE)

$$CE = \sum_{face=1}^m \sum_{ID=1}^n (-\log(output(ID)))$$

#training faces

#people's IDs

CNN output for ID



# Data-Driven Face Recognition

~~Non-CAP~~

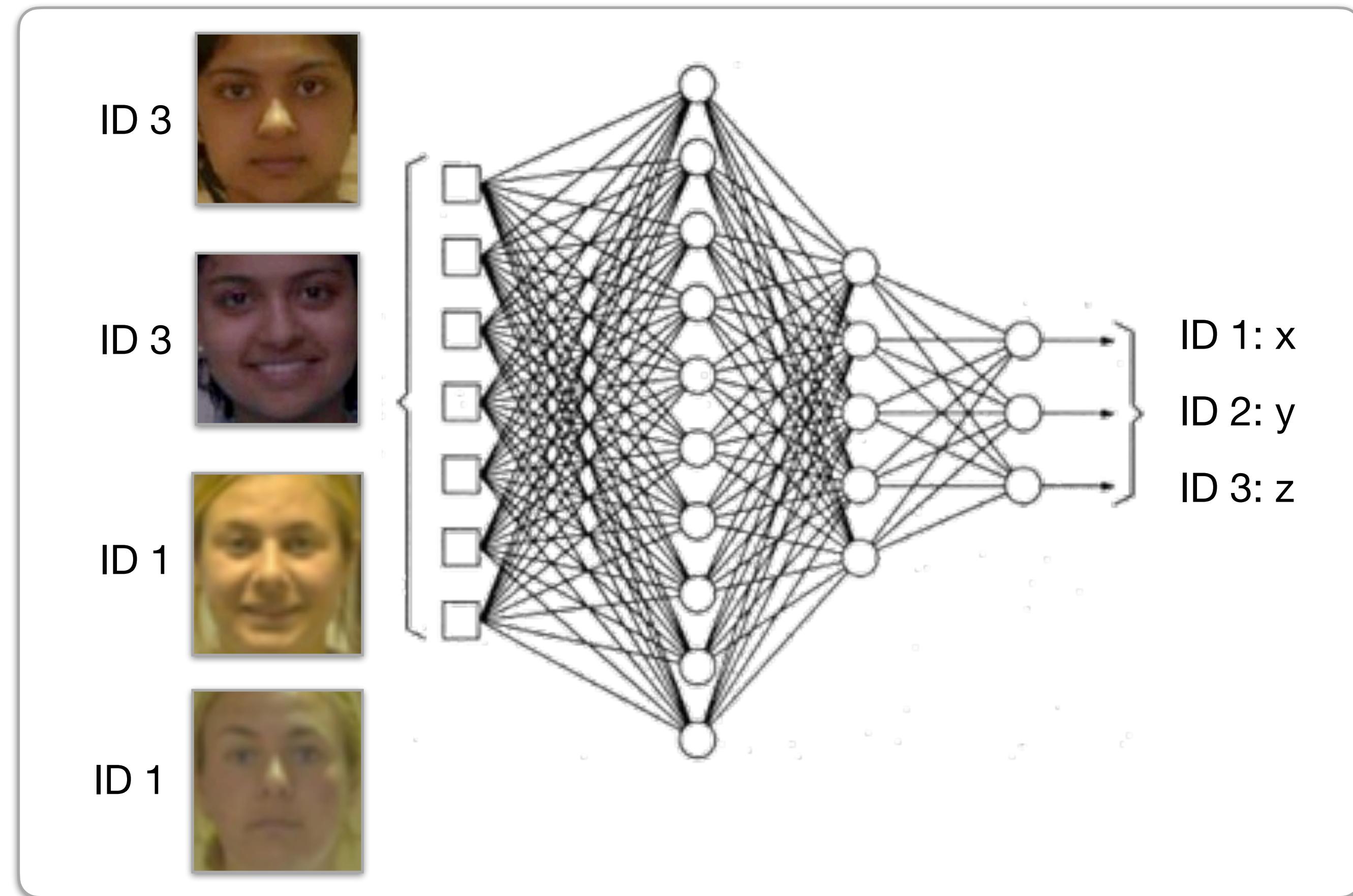
## Deep Learning

**What are the cons here?**

How to enroll a new person?

Fixed number of classes  
(i.e., persons).

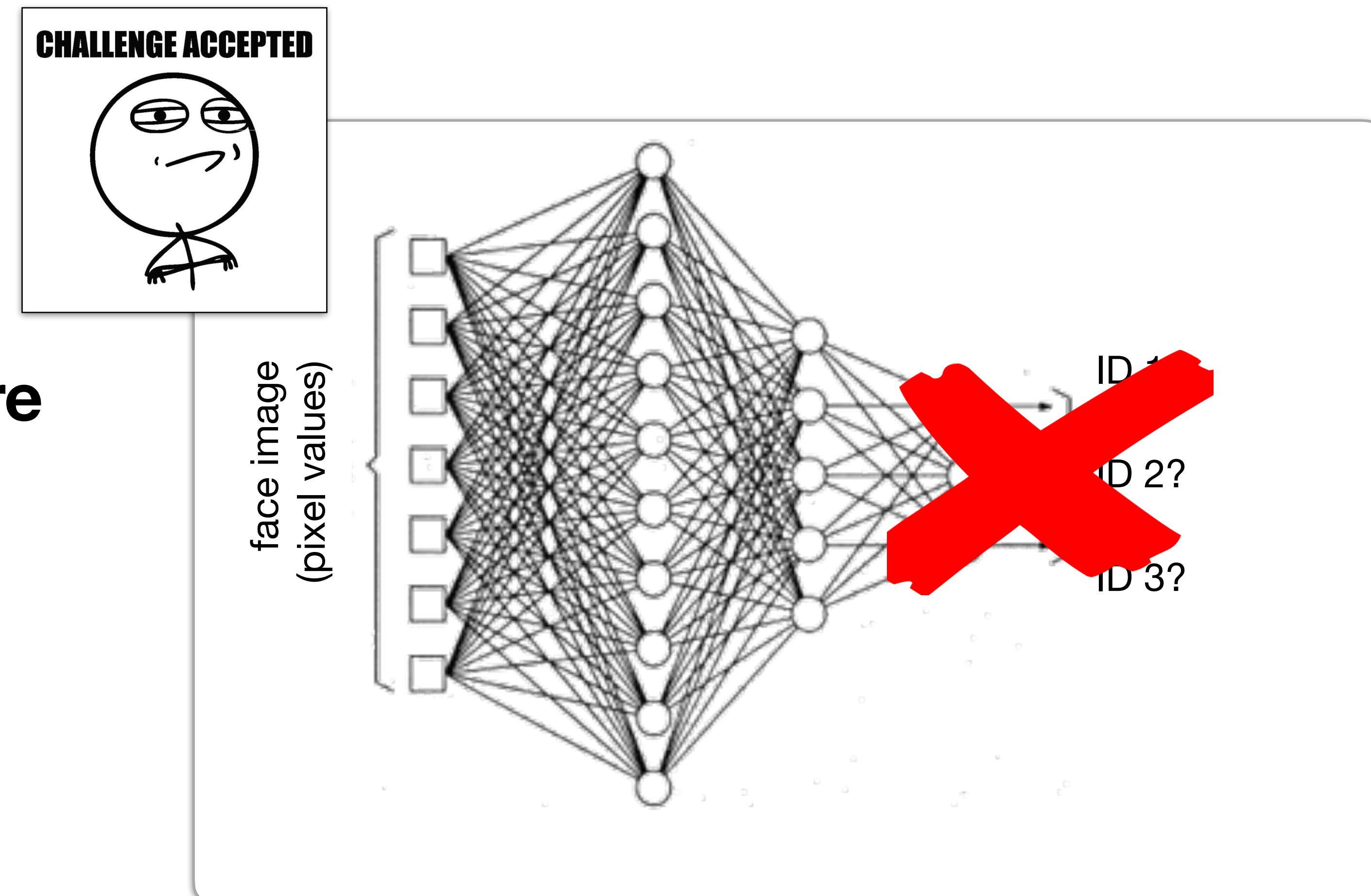
Need for large training dataset (thousands of sample per class).



# Data-Driven Face Recognition

## Deep Learning

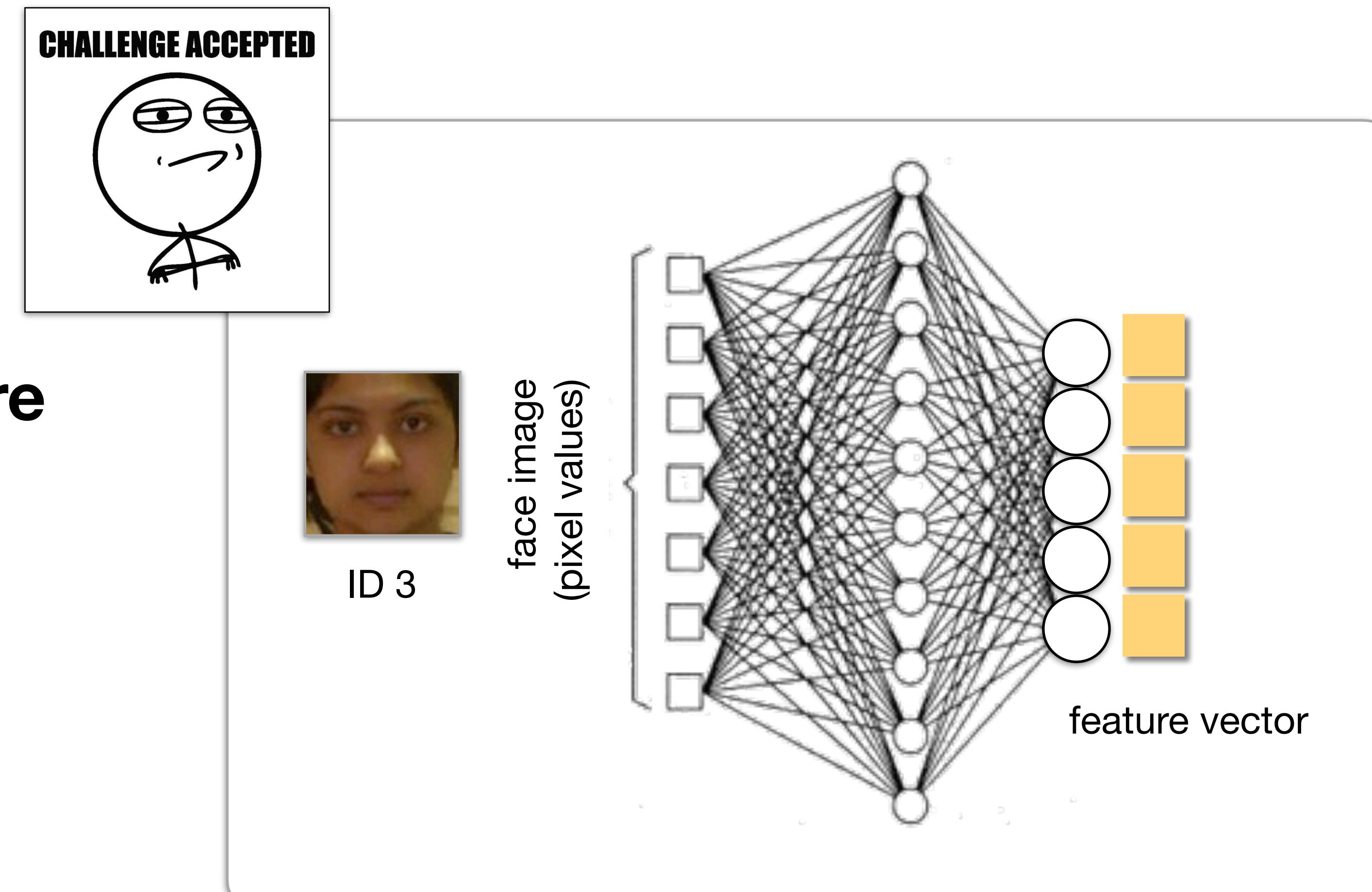
**How to make CNN more flexible?**  
Remove fully connected layer and use last convolutional layers as a feature descriptor.



# Data-Driven Face Recognition

## Deep Learning

**How to make CNN more flexible?**  
Remove fully connected layer and use last convolutional layers as a feature descriptor.

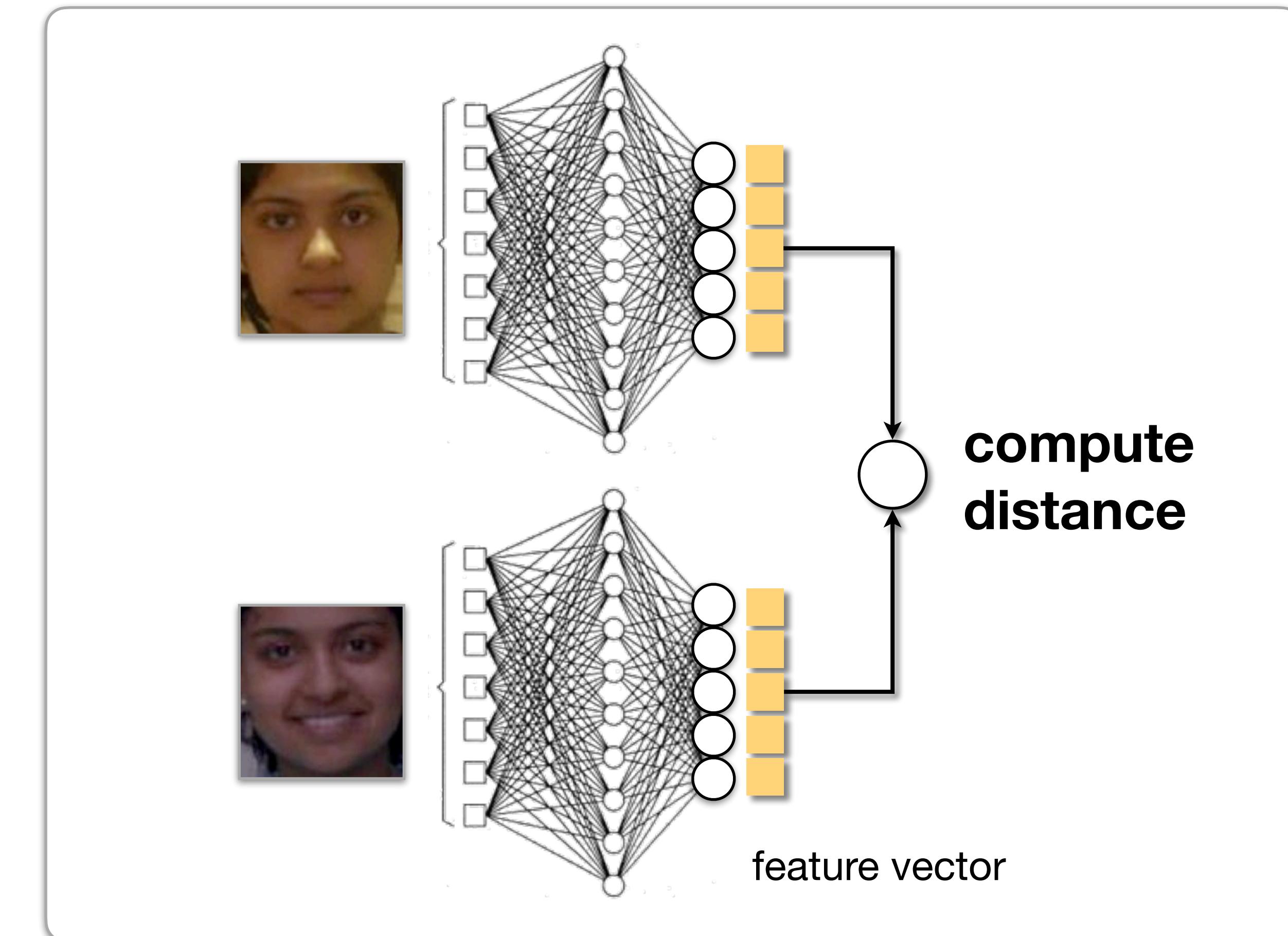


# Data-Driven Face Recognition

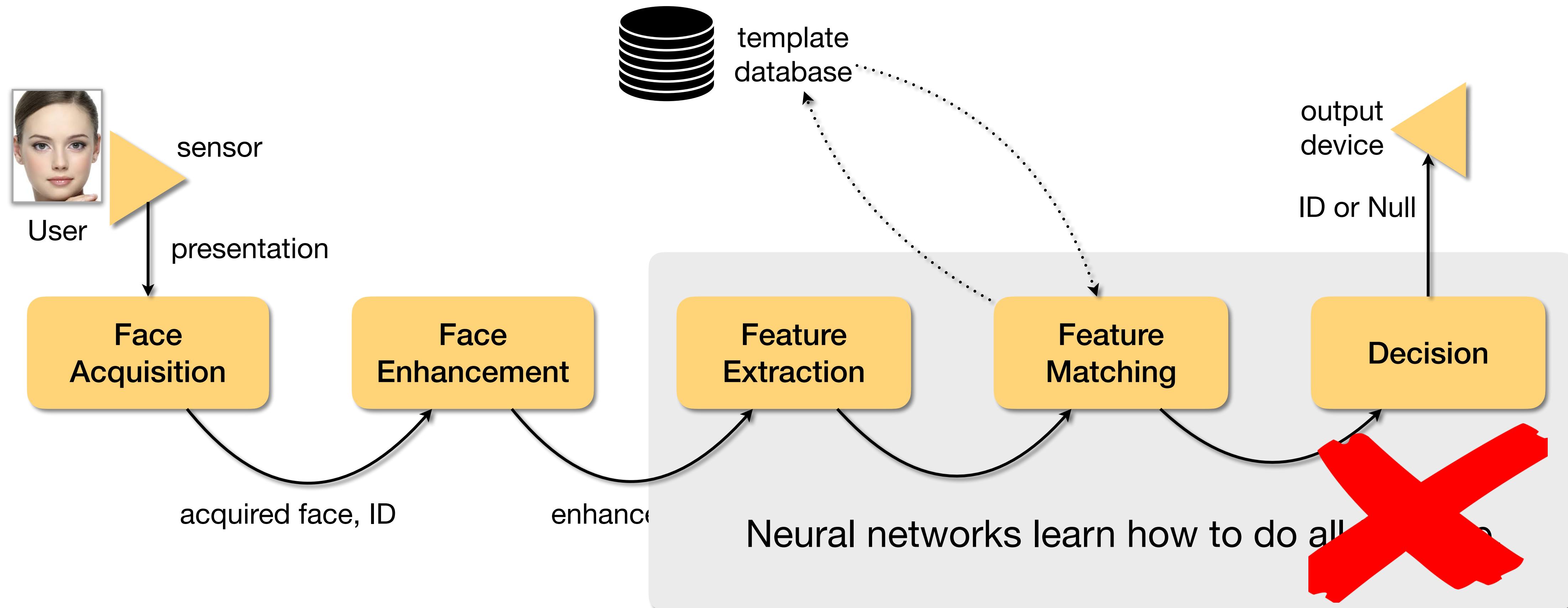
## Deep Learning

**How to make CNN more flexible?**

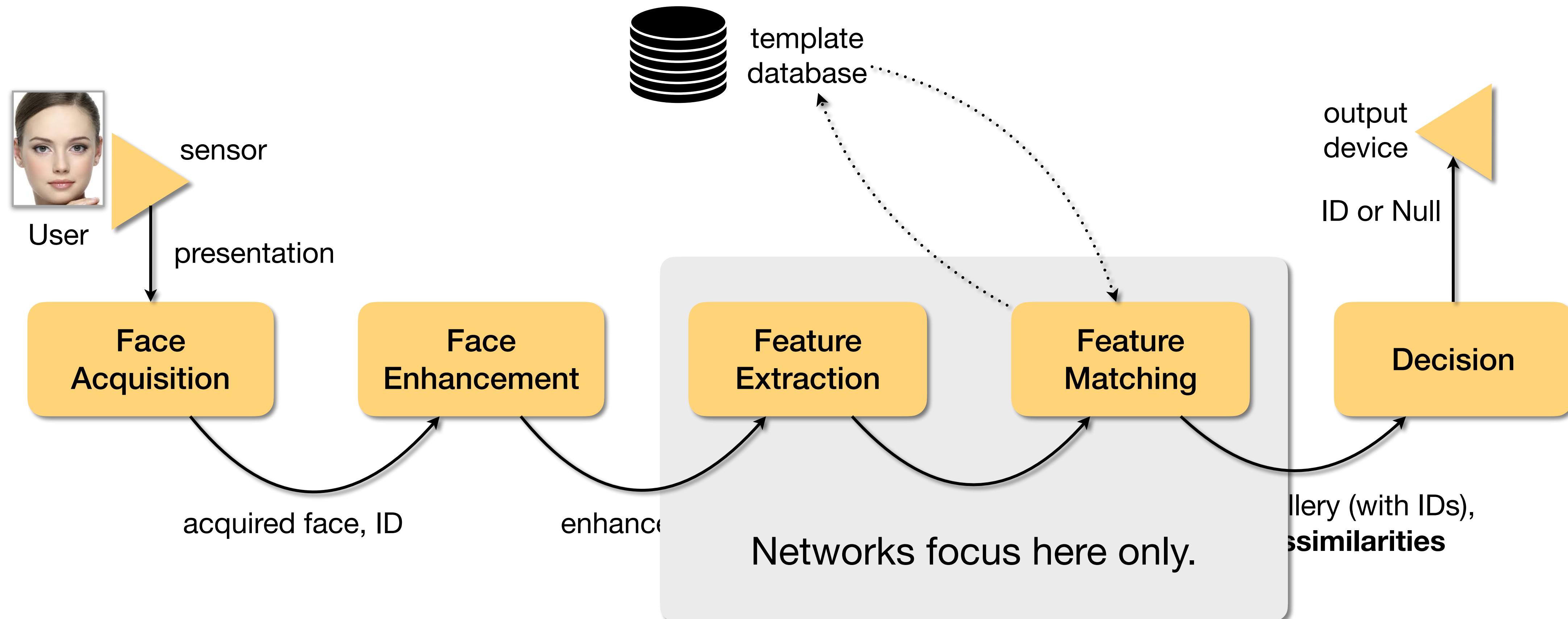
To speed up training, use **siamese networks** (same architecture, same weights).



# Data-Driven Face Recognition



# Data-Driven Face Recognition



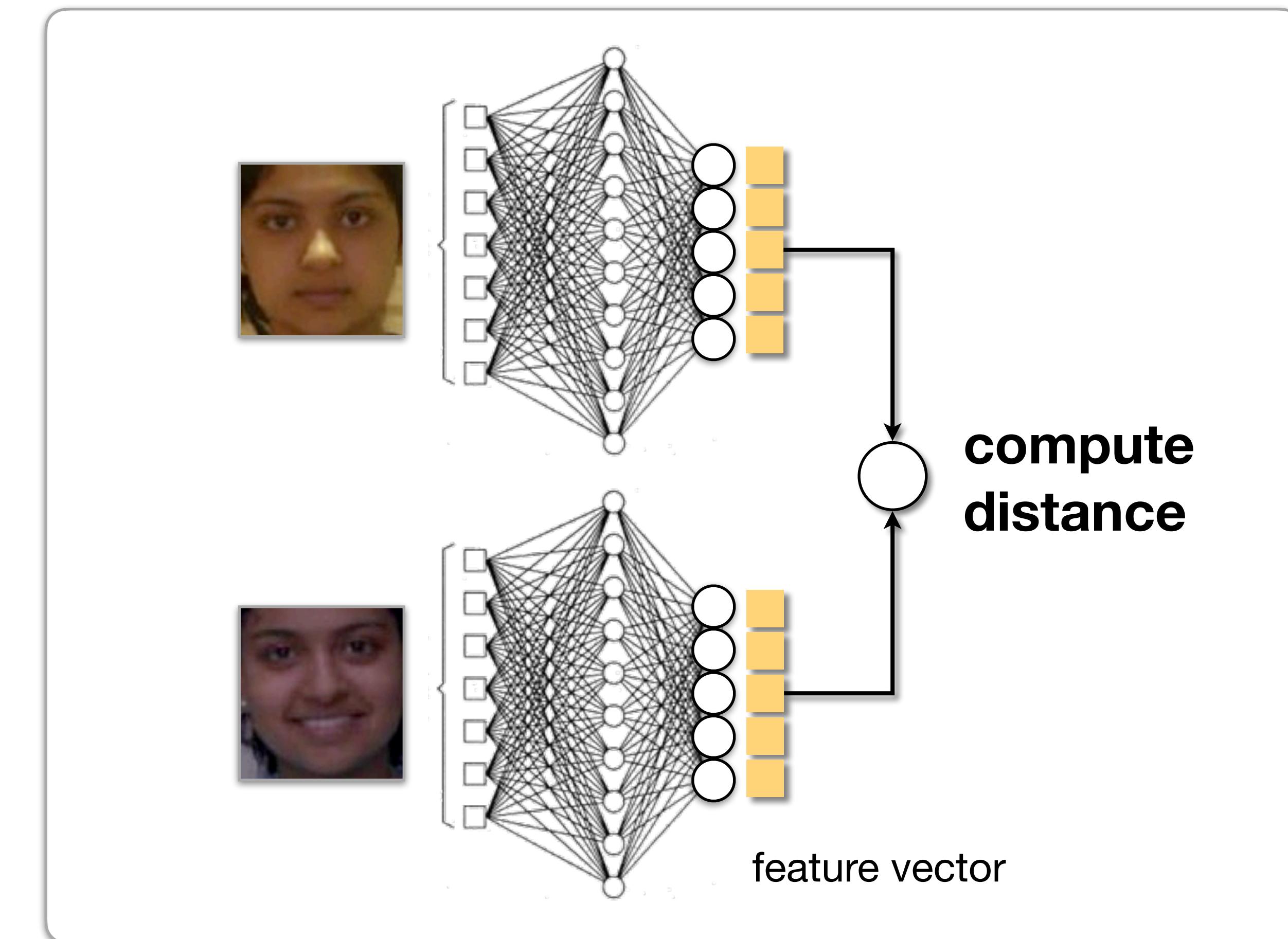
# Data-Driven Face Recognition

## Deep Learning

### Training Approaches

Pairwise-loss-based

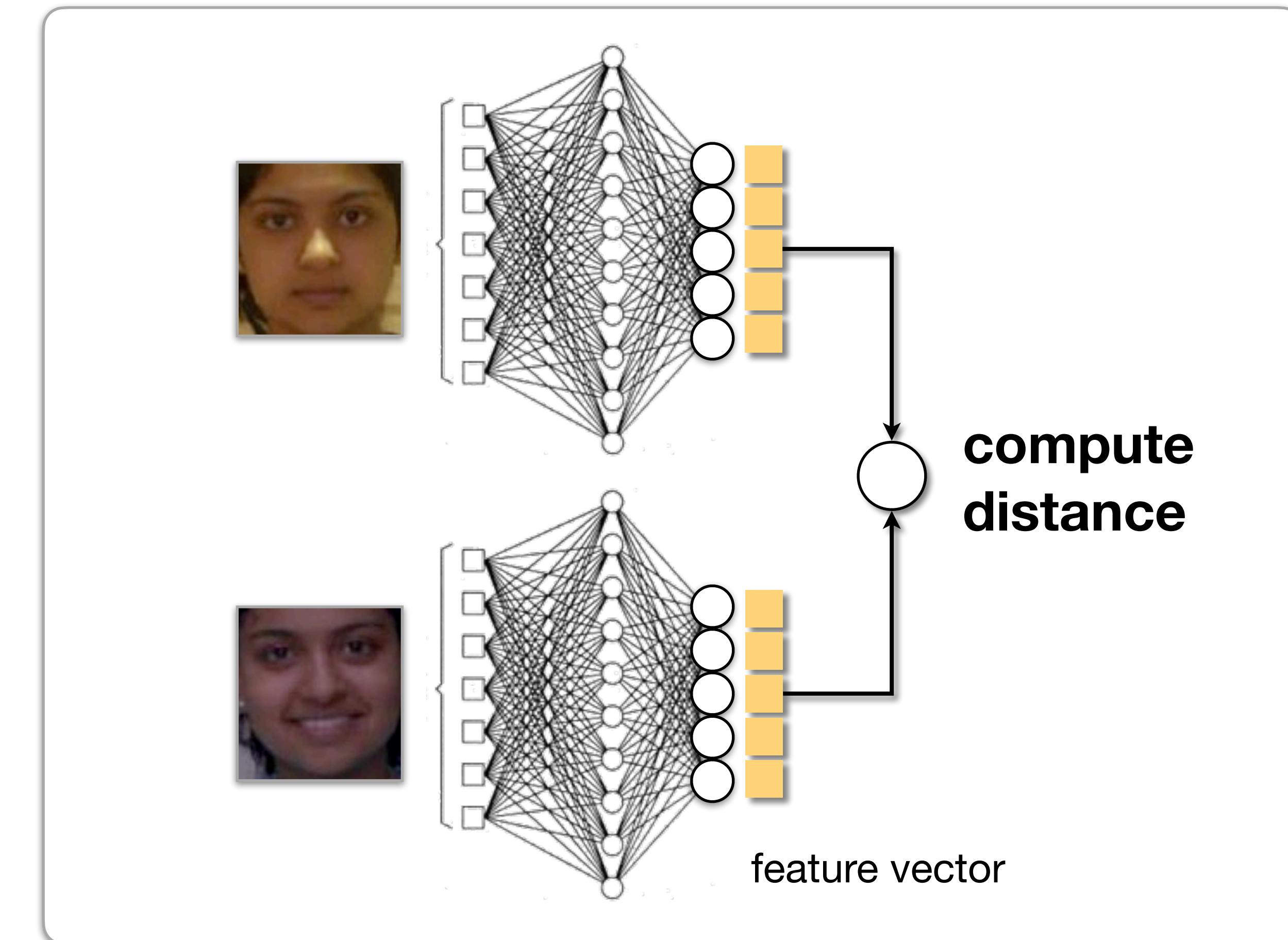
Triplet-loss-based



# Data-Driven Face Recognition

## Deep Learning

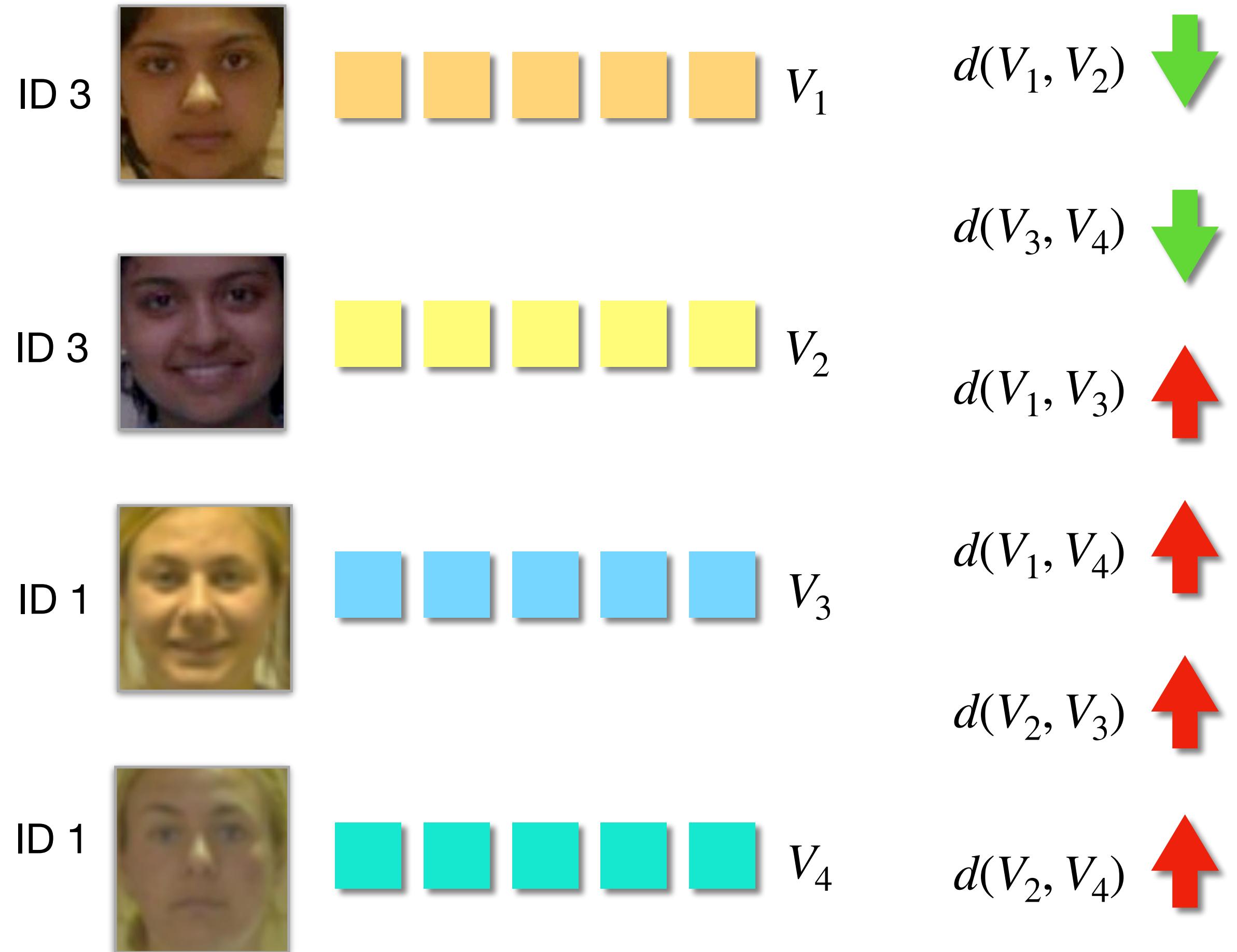
**Training Approaches**  
**Pairwise-loss-based**  
Triplet-loss-based



# Pairwise Face Recognition

## Pairwise Loss (PL)

Train the network in a way that feature vectors of the same class have small distance, while feature vectors from different classes have large distance.



# Pairwise Face Recognition

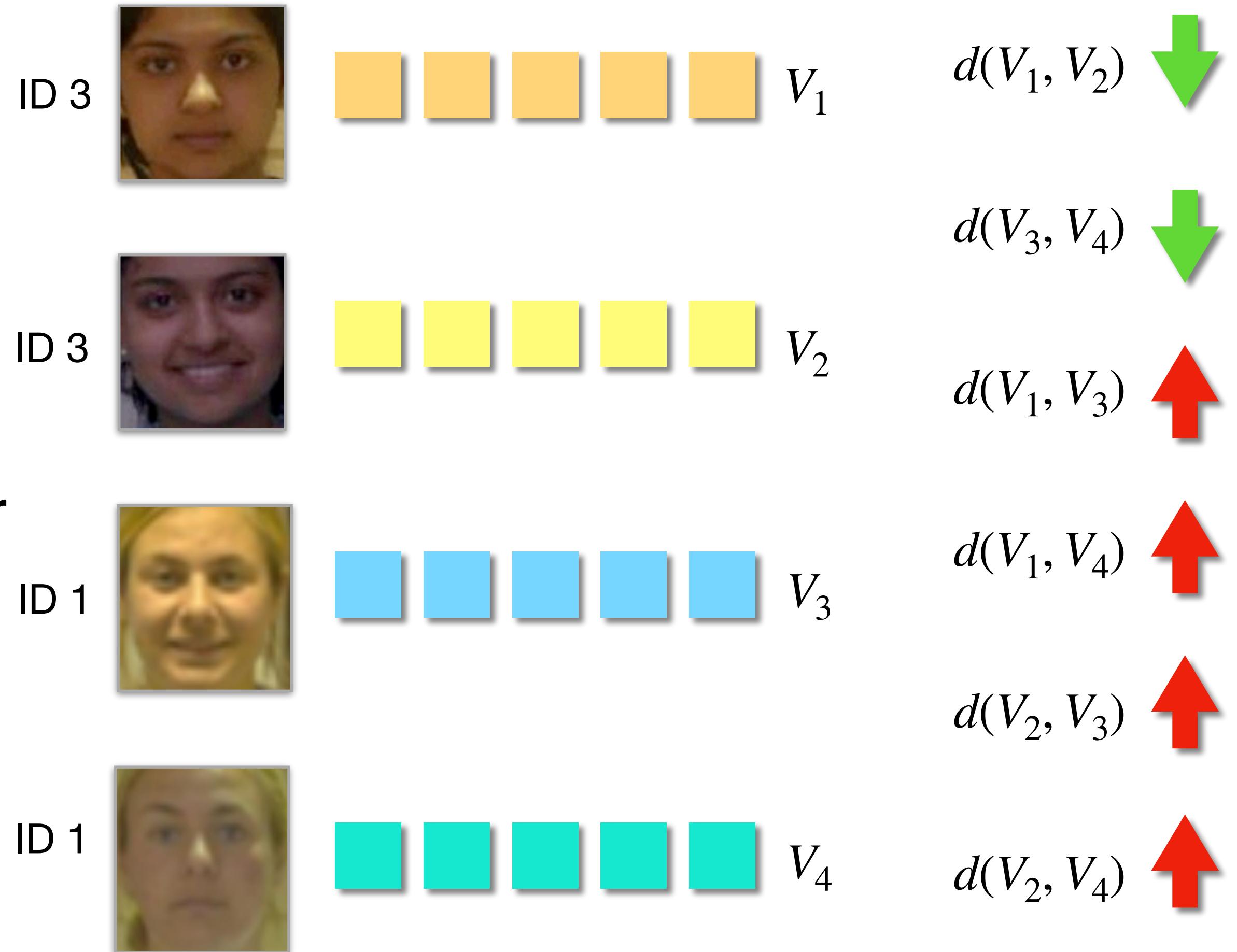
## Pairwise Loss (PL)

the smaller, the better

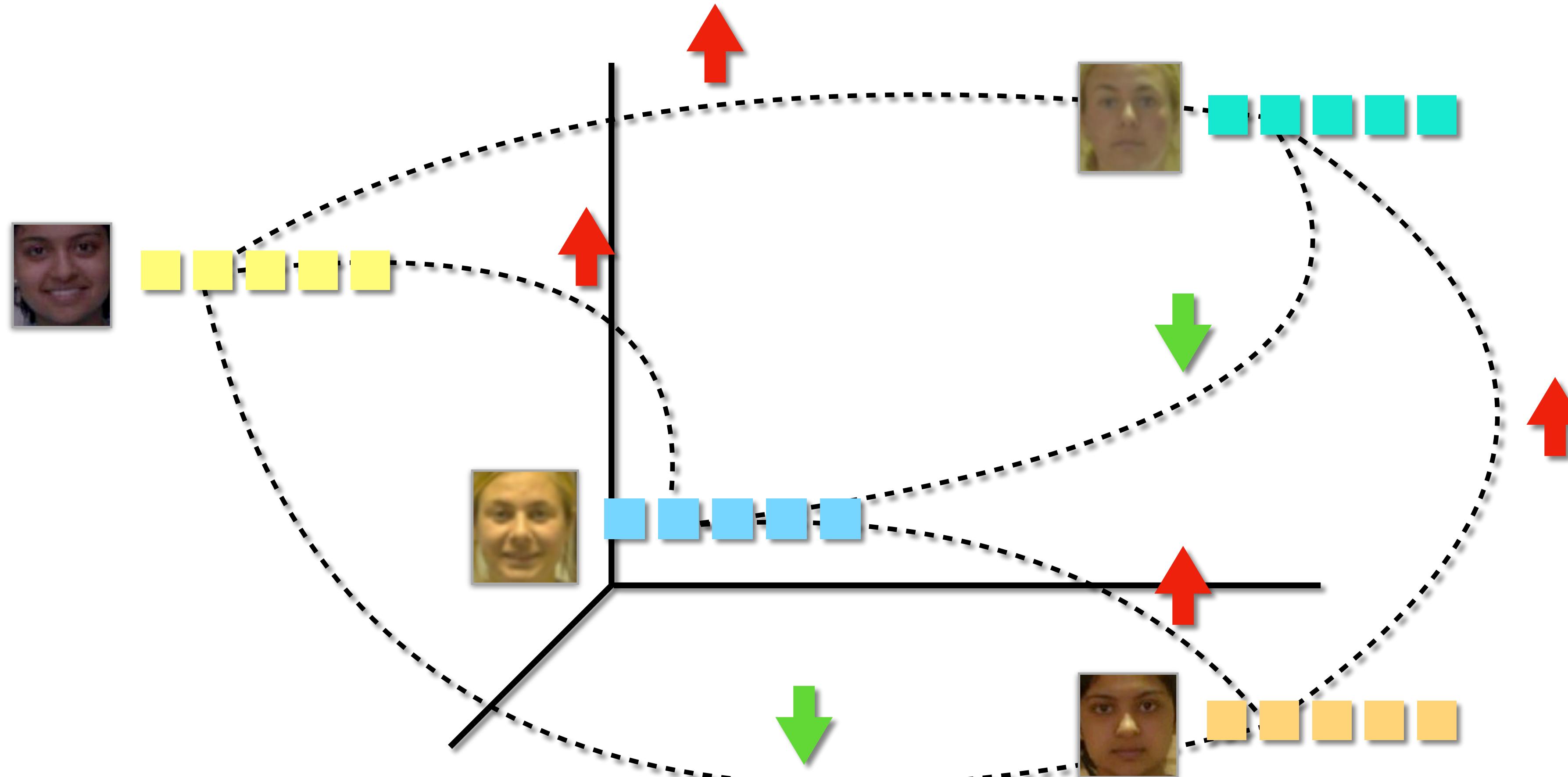
$$PL = \begin{cases} d(V_x, V_y) & \text{if genuine pair} \\ \max(0, m - d(V_x, V_y)) & \text{if impostor pair} \end{cases}$$

enforced margin

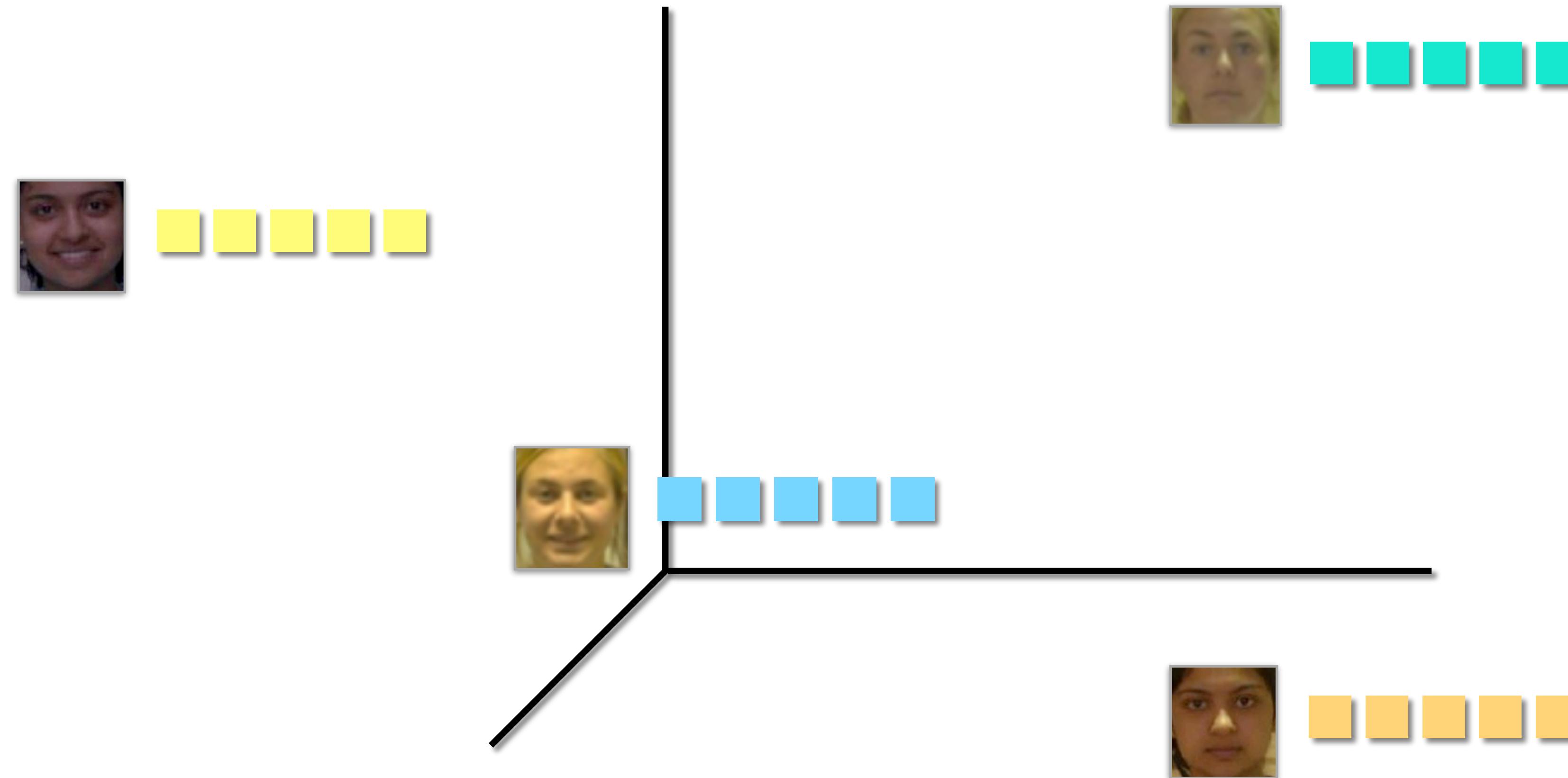
it must be larger than  $m$



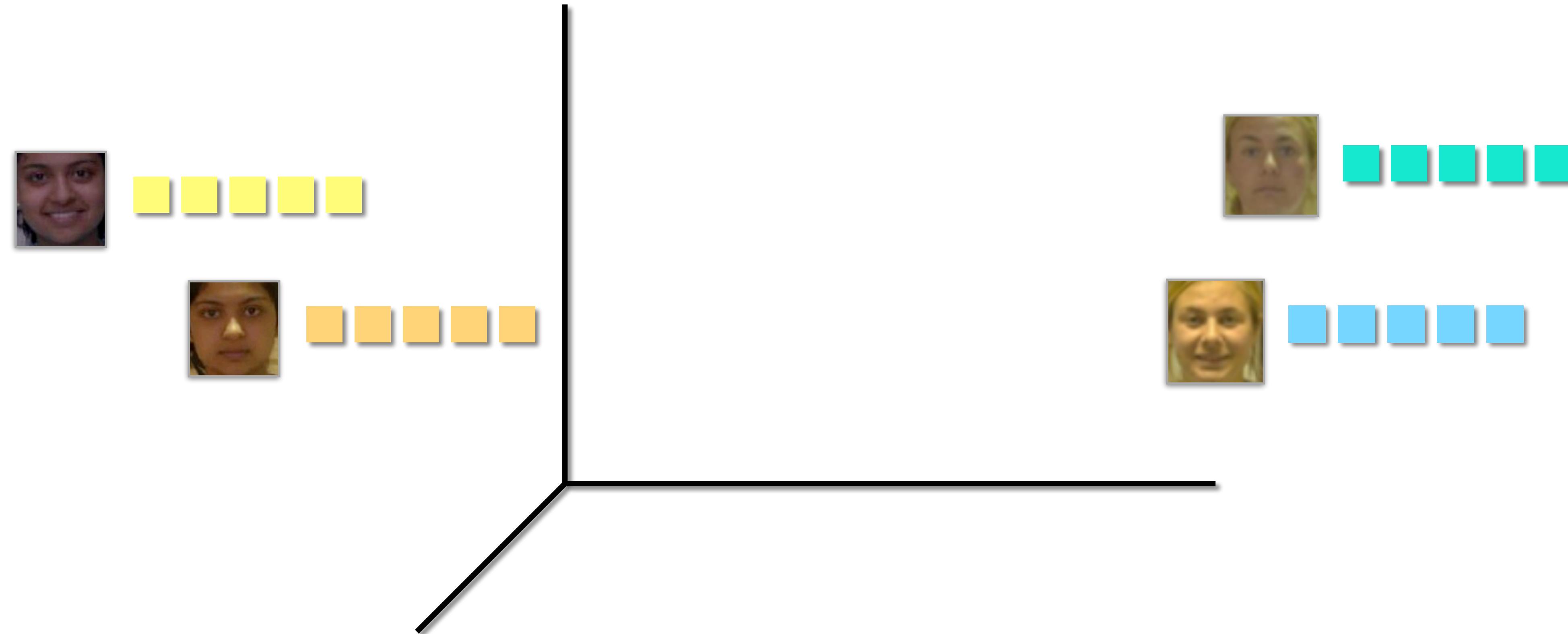
# Pairwise Face Recognition



# Pairwise Face Recognition



# Pairwise Face Recognition



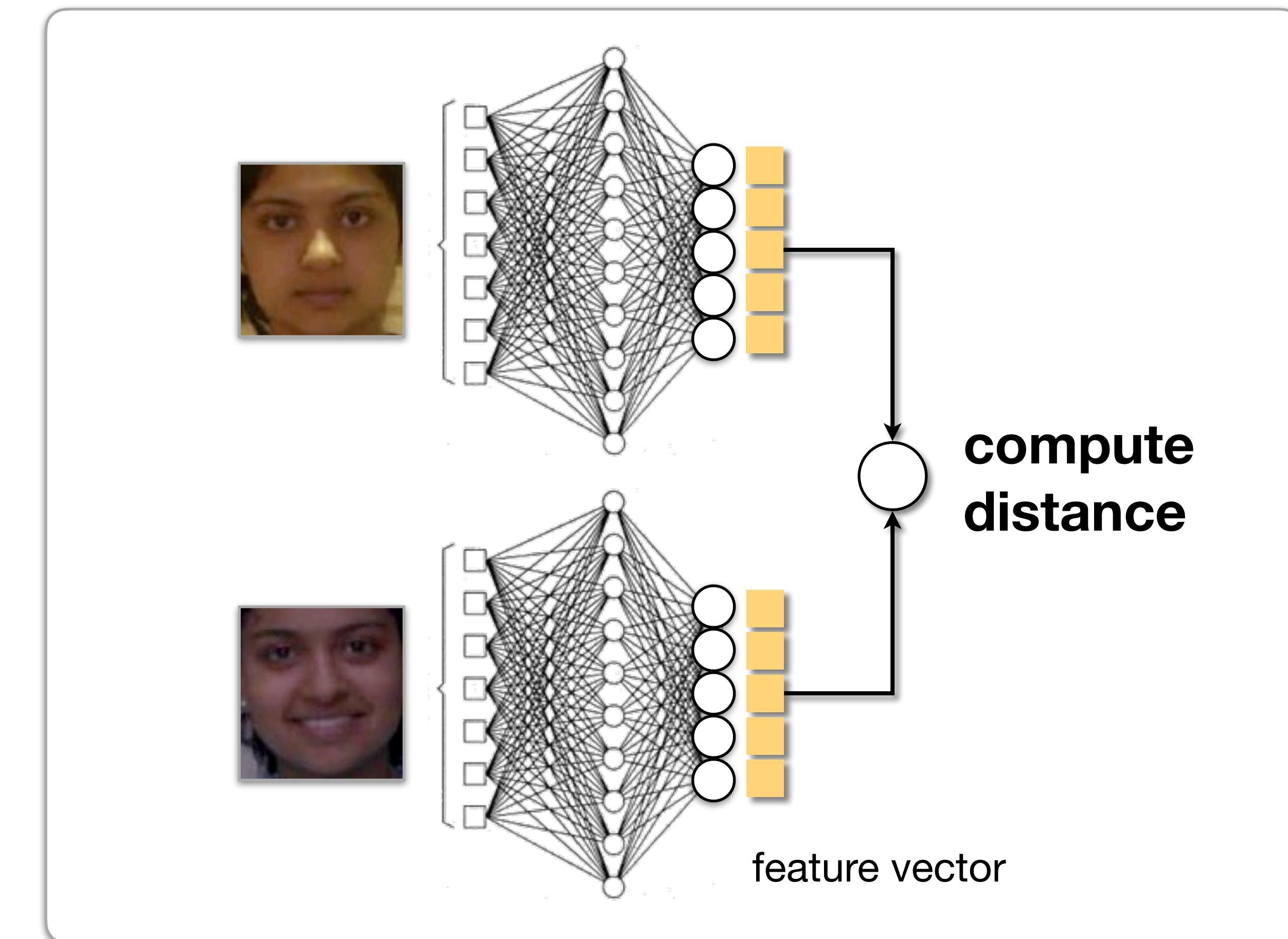
# Data-Driven Face Recognition

## Deep Learning

### Training Approaches

Pairwise-loss-based

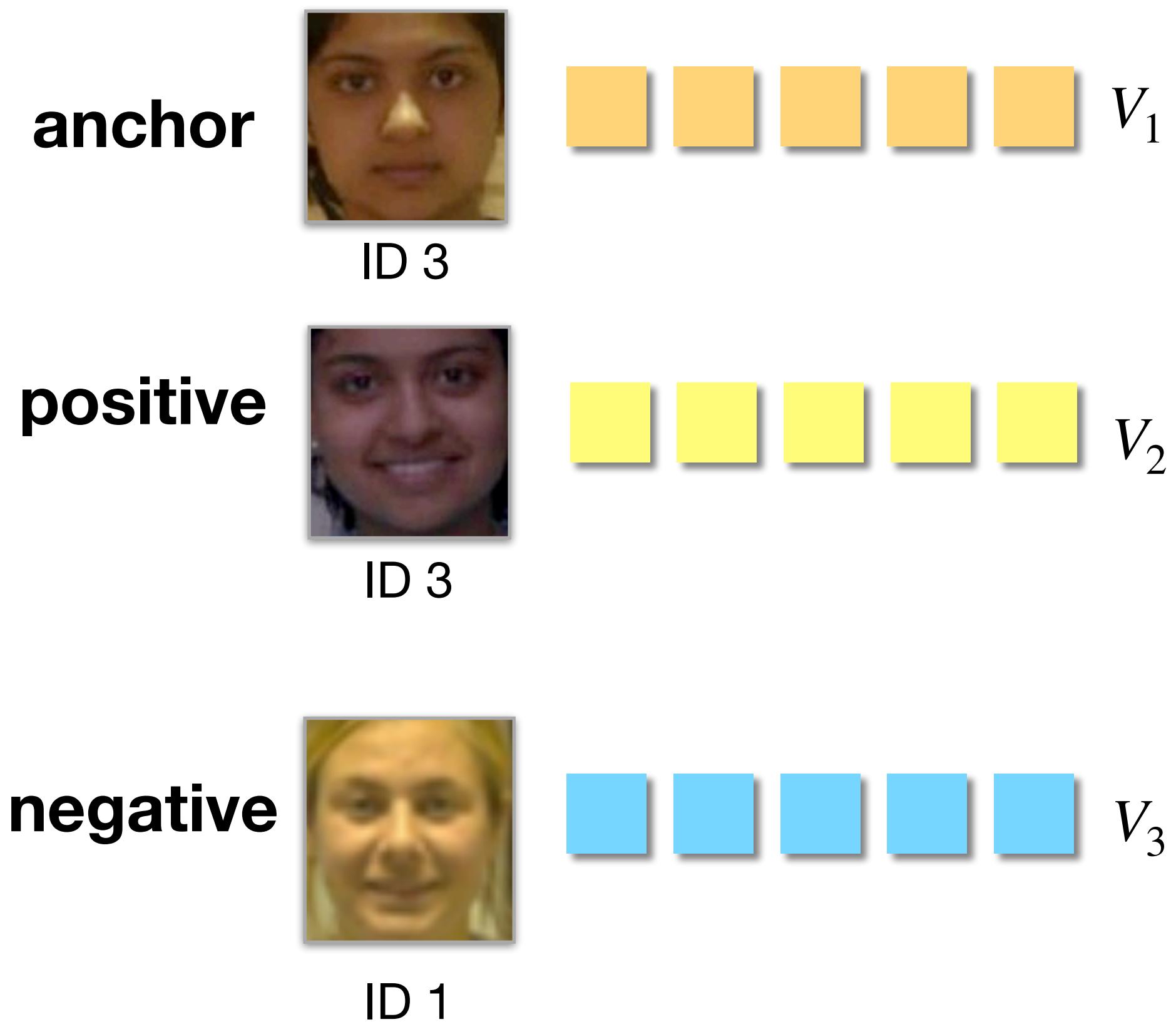
**Triplet-loss-based**



# Triplet Face Recognition

## Triplet Loss (TL)

Choose a reference data sample (the **anchor**) and a **positive** and a **negative** data samples to optimize their distances.



# Triplet Face Recognition

## Triplet Loss (TL)

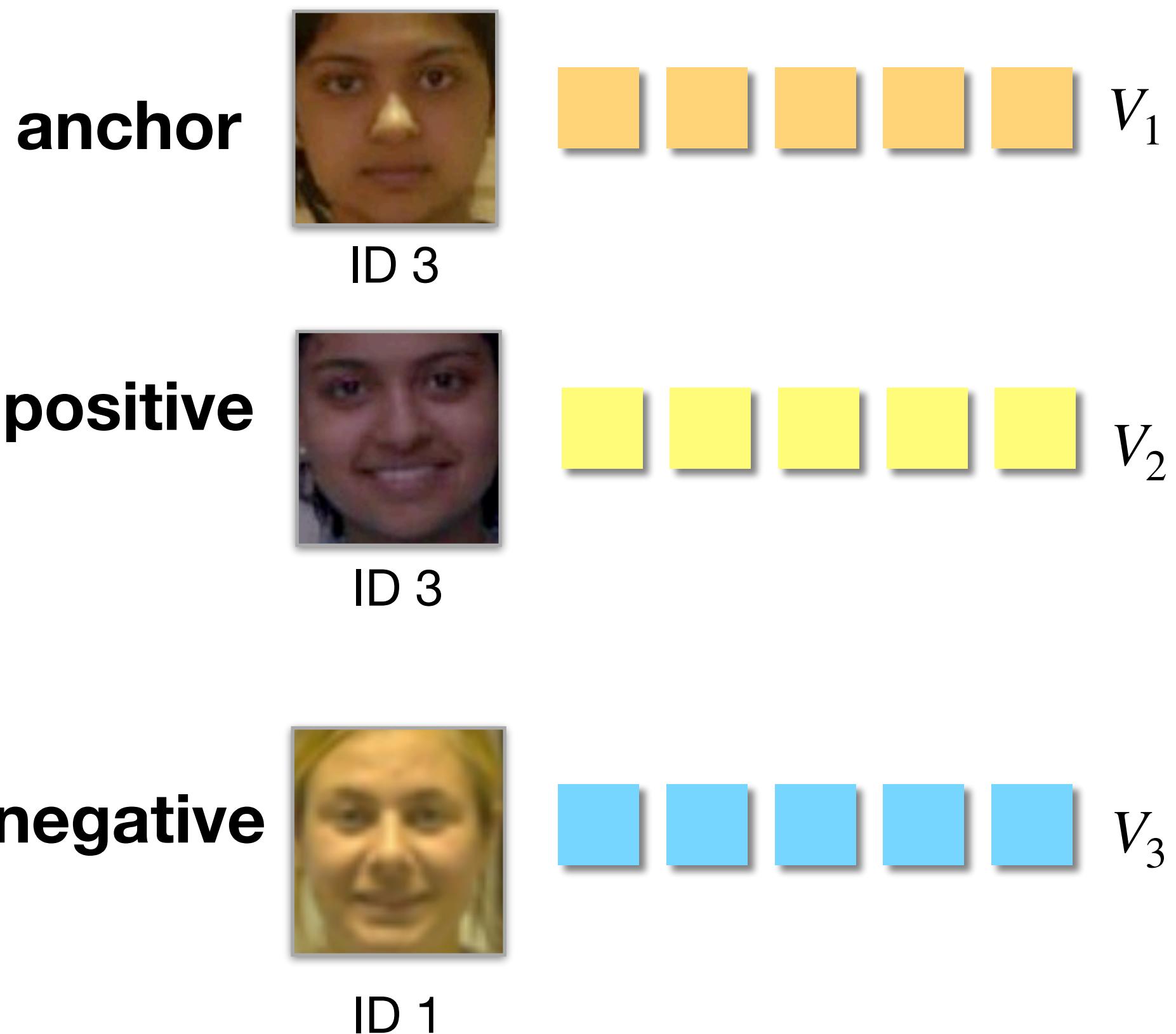
Choose a reference data sample (the **anchor**) and a **positive** and a **negative** data samples to optimize their distances.

Minimize  $d(V_1, V_2)$  and maximize  $d(V_1, V_3)$ .

Schroff et al.

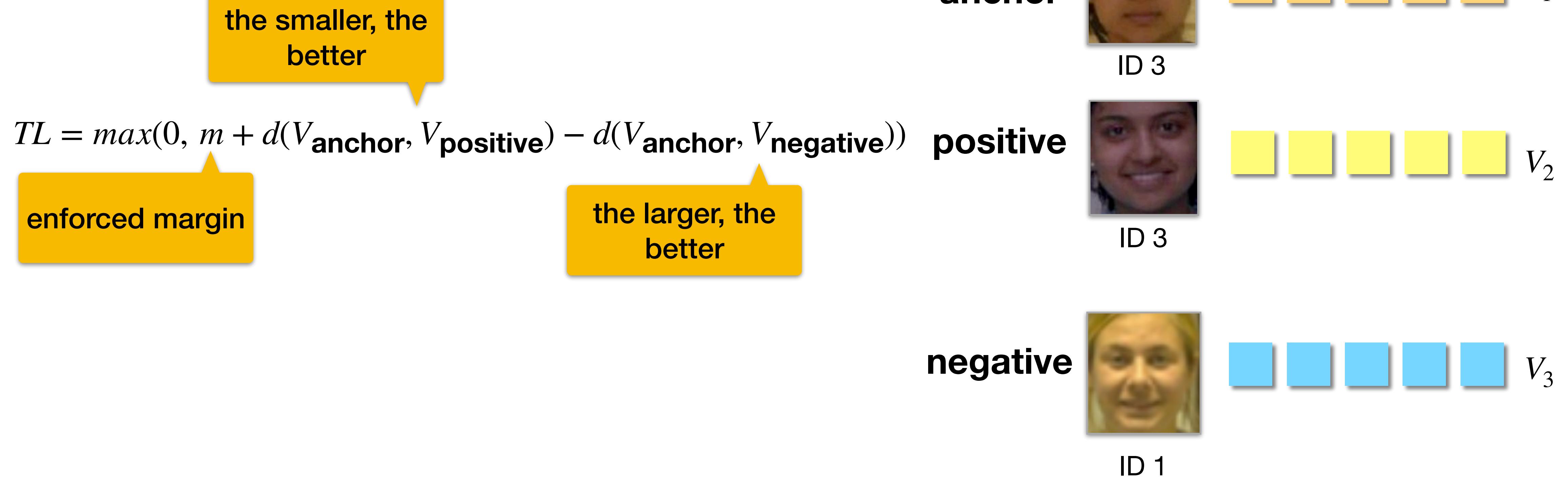
*Facenet: A unified embedding for face recognition and clustering.*

CVPR 2015



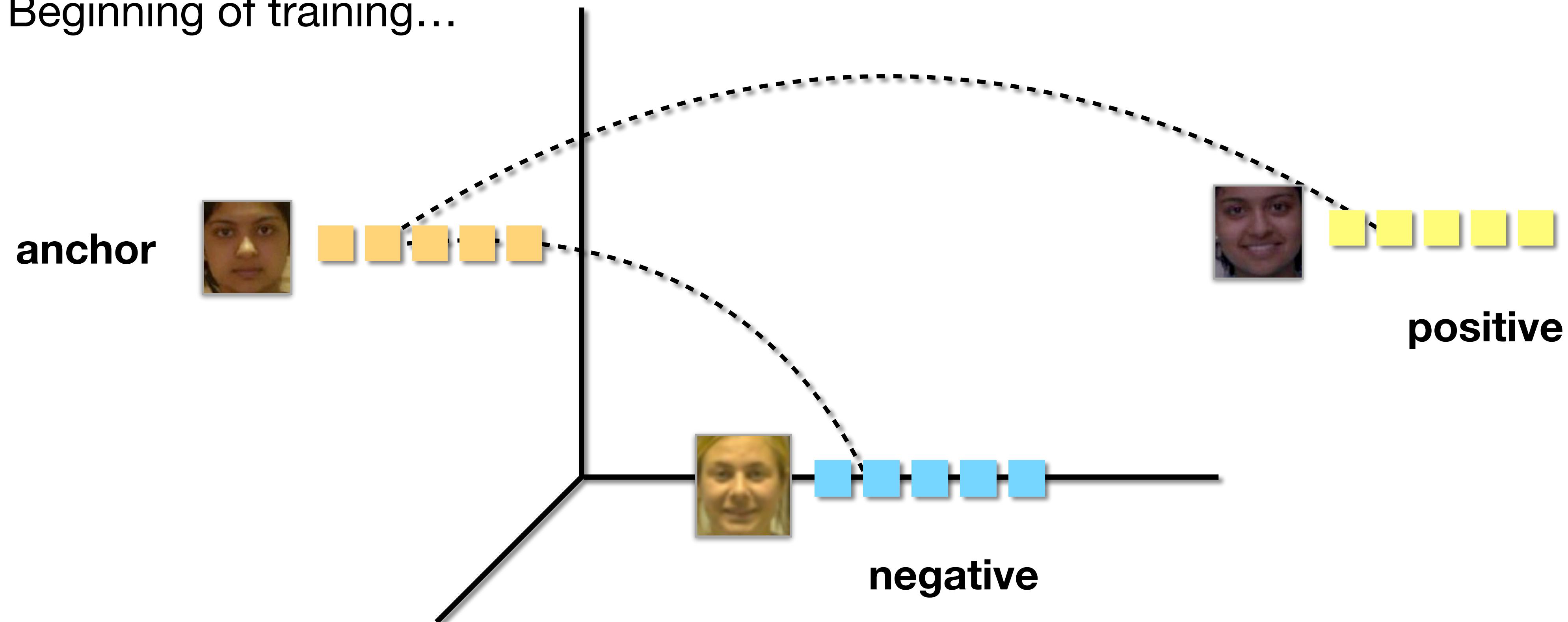
# Triplet Face Recognition

## Triplet Loss (TL)

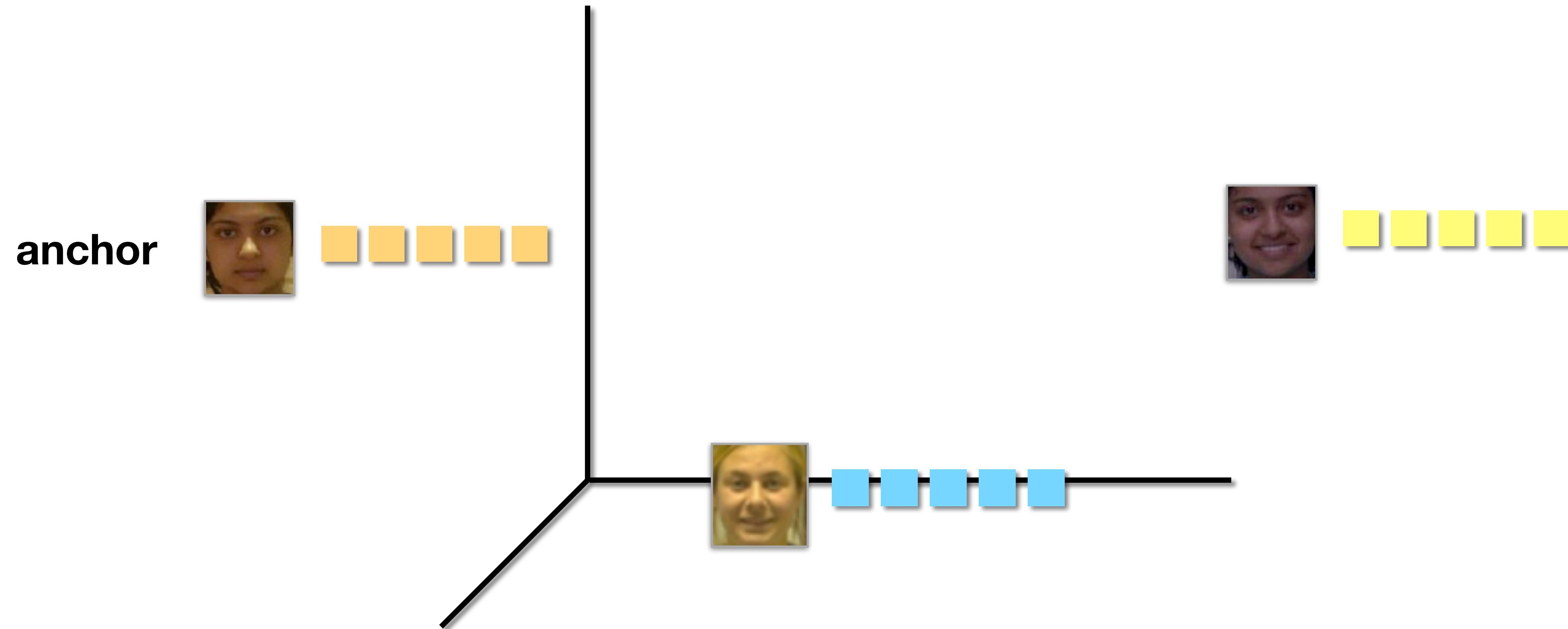


# Triplet Face Recognition

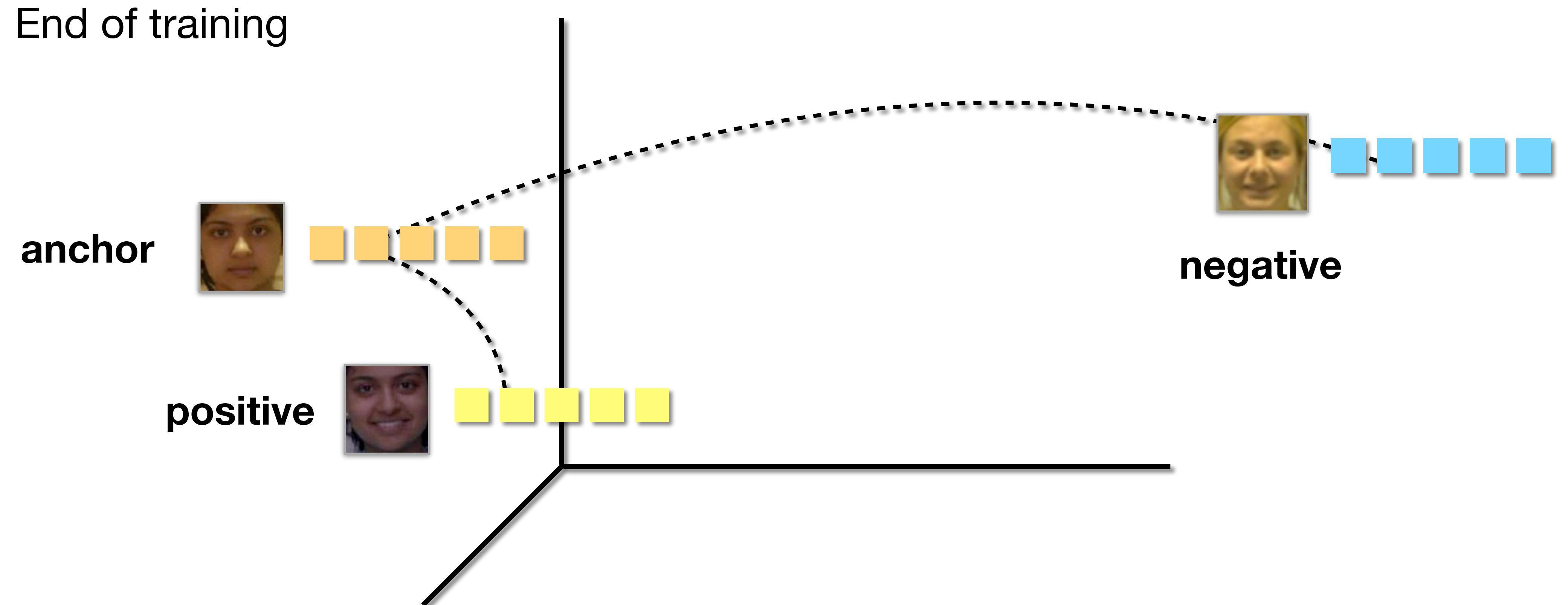
Beginning of training...



# Triplet Face Recognition

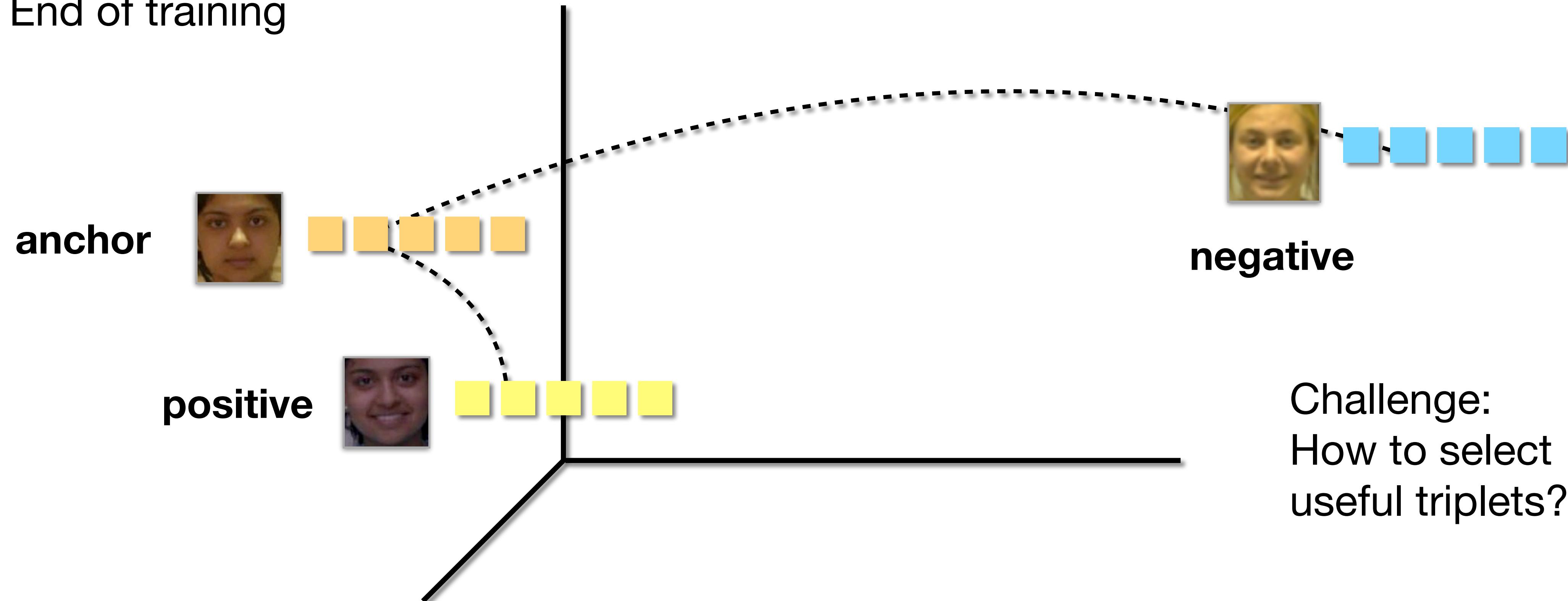


# Triplet Face Recognition



# Triplet Face Recognition

End of training



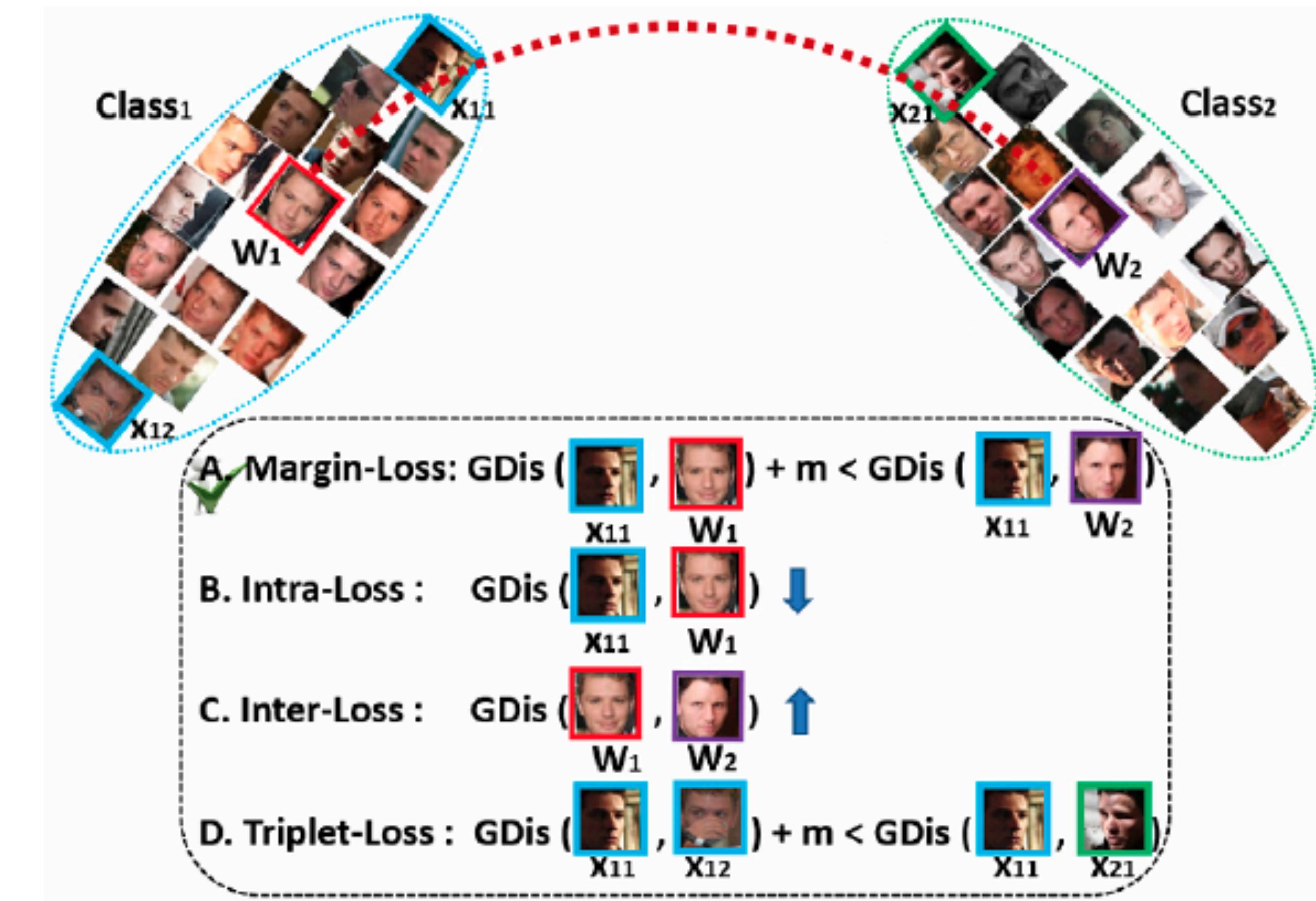
# Improvements

Source: Deng et al.  
*Additive Angular Margin Loss  
for Deep Face Recognition.*  
CVPR 2019

## Centre Loss

Use class clusters' centers to improve the convergence of the learning process.

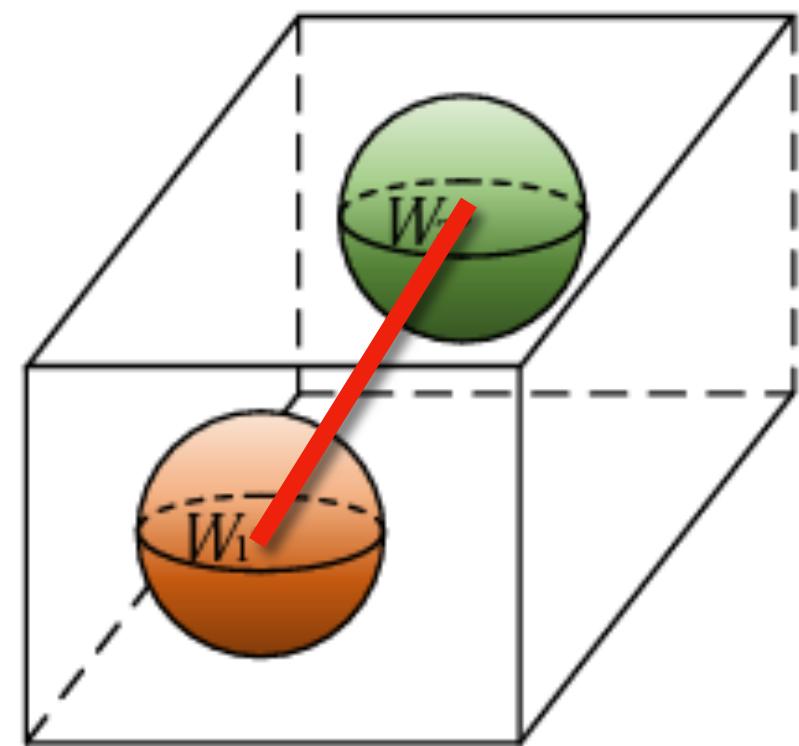
Liu et al.  
*Sphereface: Deep hypersphere embedding for face recognition.*  
CVPR 2017



# Improvements

## SphereFace

Transform feature space into hypersphere and compute the distances as the **angles** between the feature vectors.



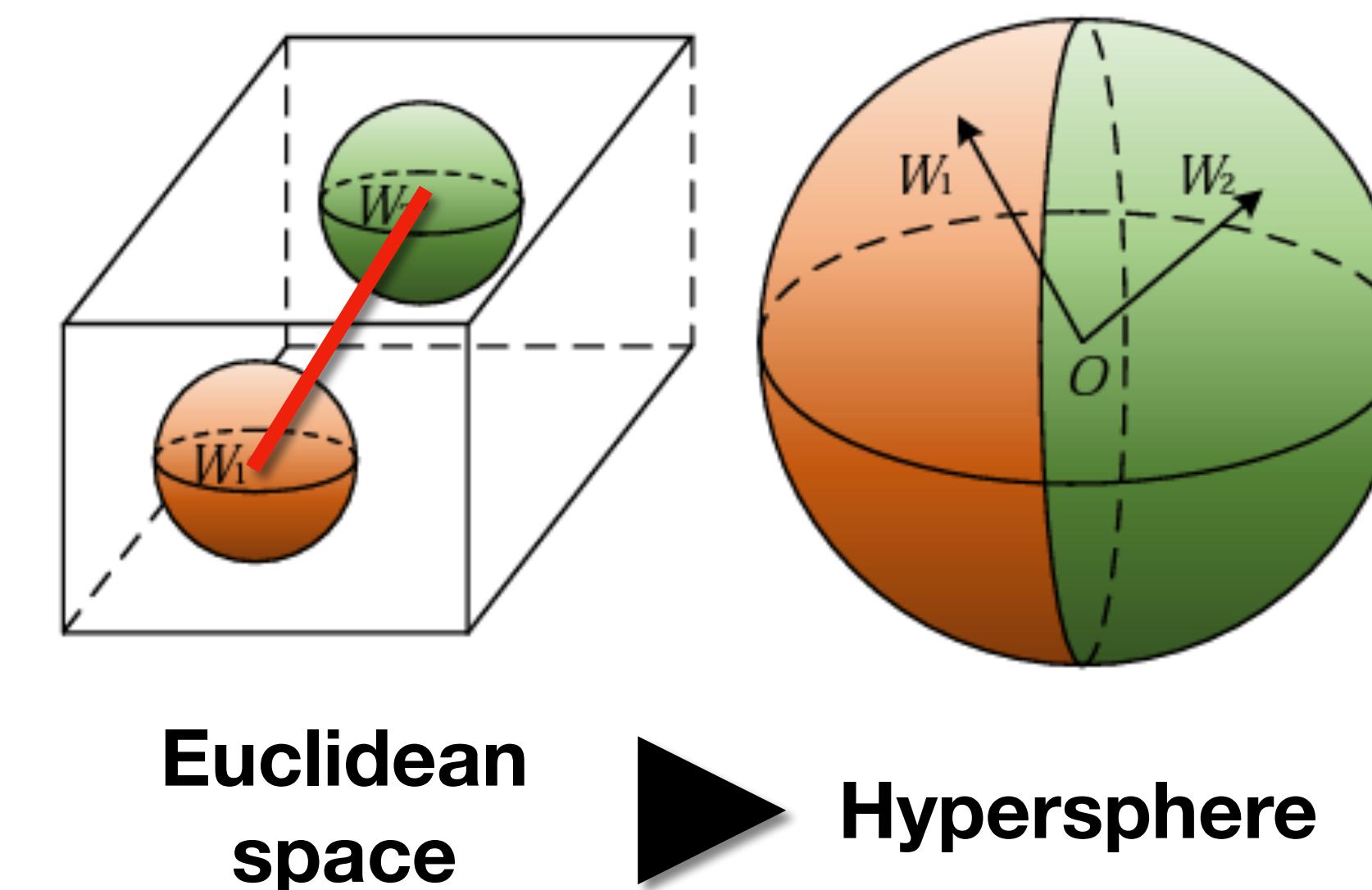
Euclidean  
space

Liu et al.  
*Sphereface: Deep hypersphere embedding for face recognition.*  
CVPR 2017

# Improvements

## SphereFace

Transform feature space into hypersphere and compute the distances as the **angles** between the feature vectors.



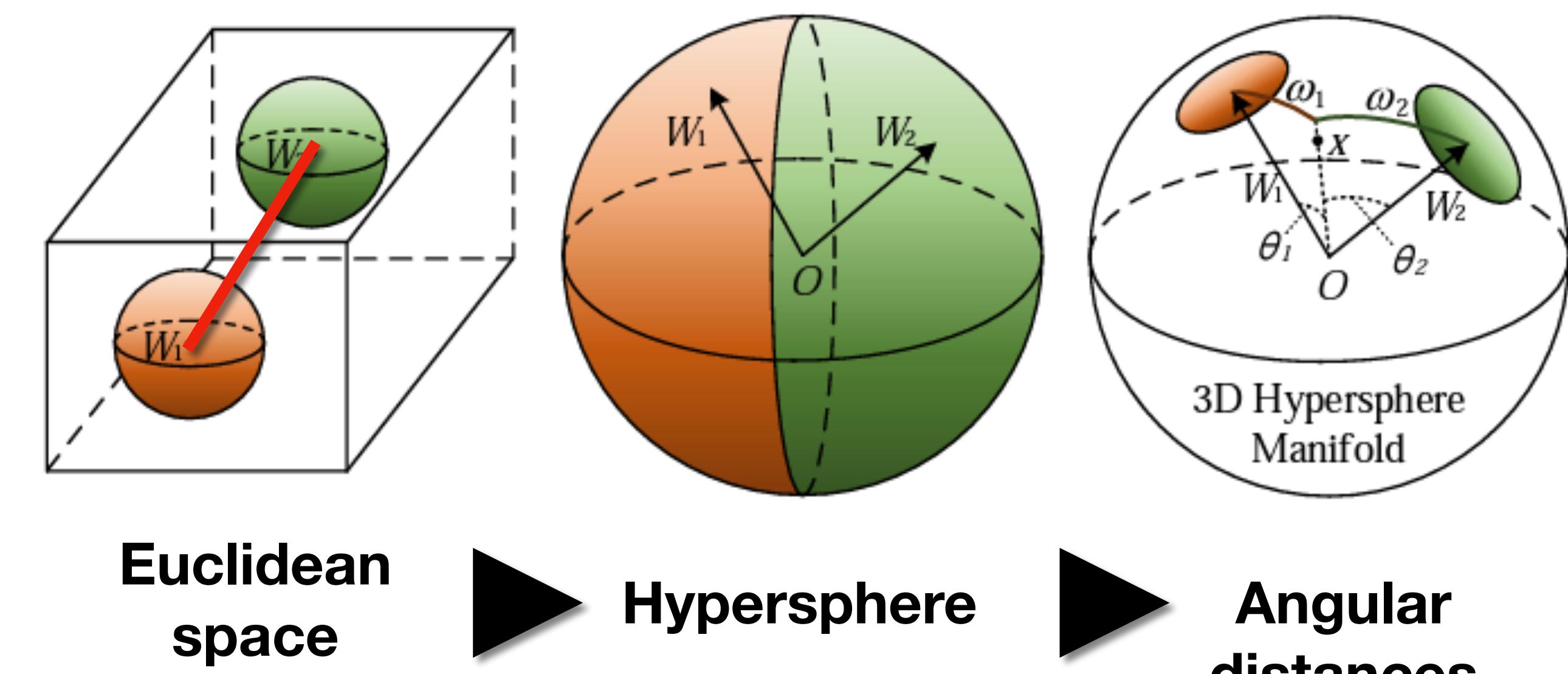
Liu et al.

*Sphereface: Deep hypersphere embedding for face recognition.*  
CVPR 2017

# Improvements

## SphereFace

Transform feature space into hypersphere and compute the distances as the **angles** between the feature vectors.



Liu et al.  
*Sphereface: Deep hypersphere embedding for face recognition.*  
CVPR 2017

# Improvements

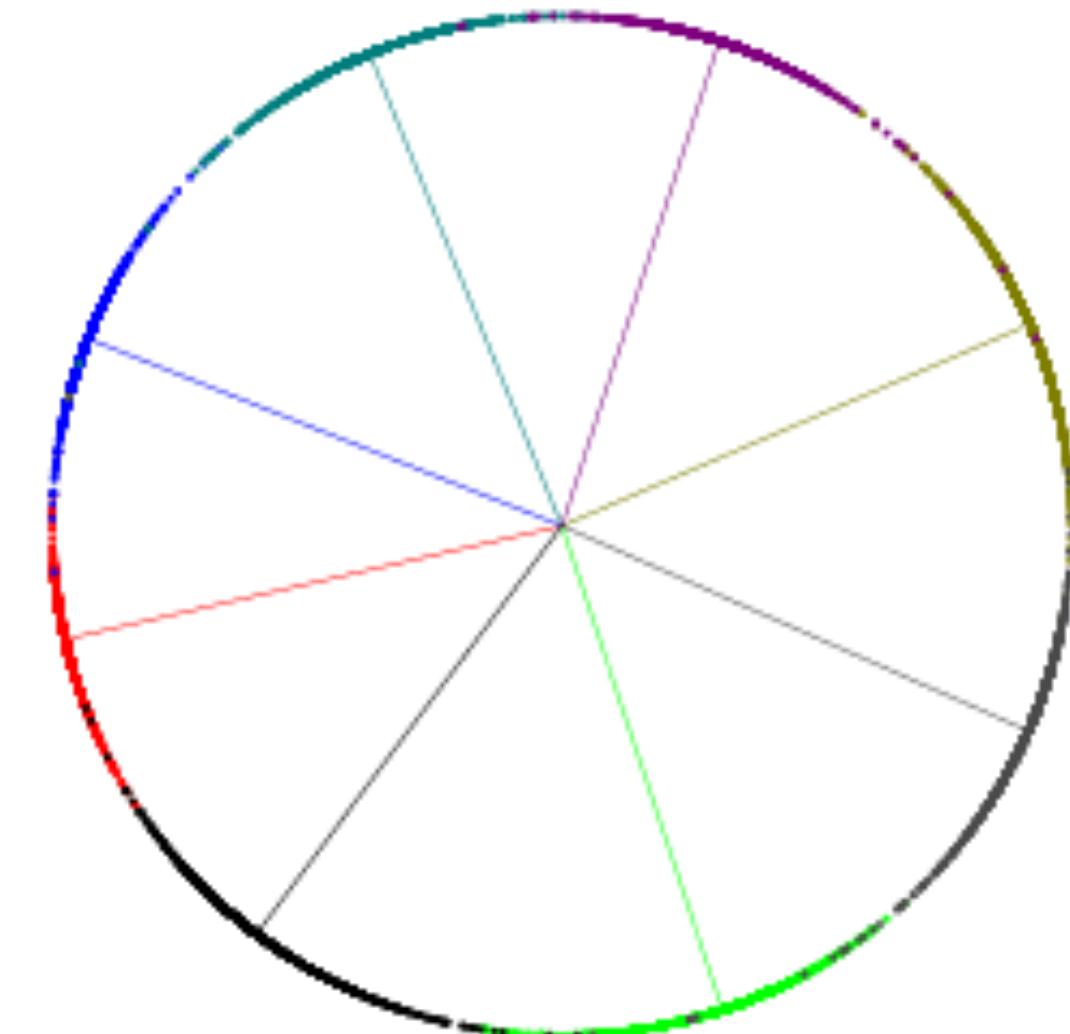
## ArcFace

Current state of the art.

Deng et al. proposed the **additive angular margin loss** to the problem of face recognition.

Deng et al.

*Additive Angular Margin Loss for Deep Face Recognition.*  
CVPR 2019



**Margin-less  
class separation**

# Improvements

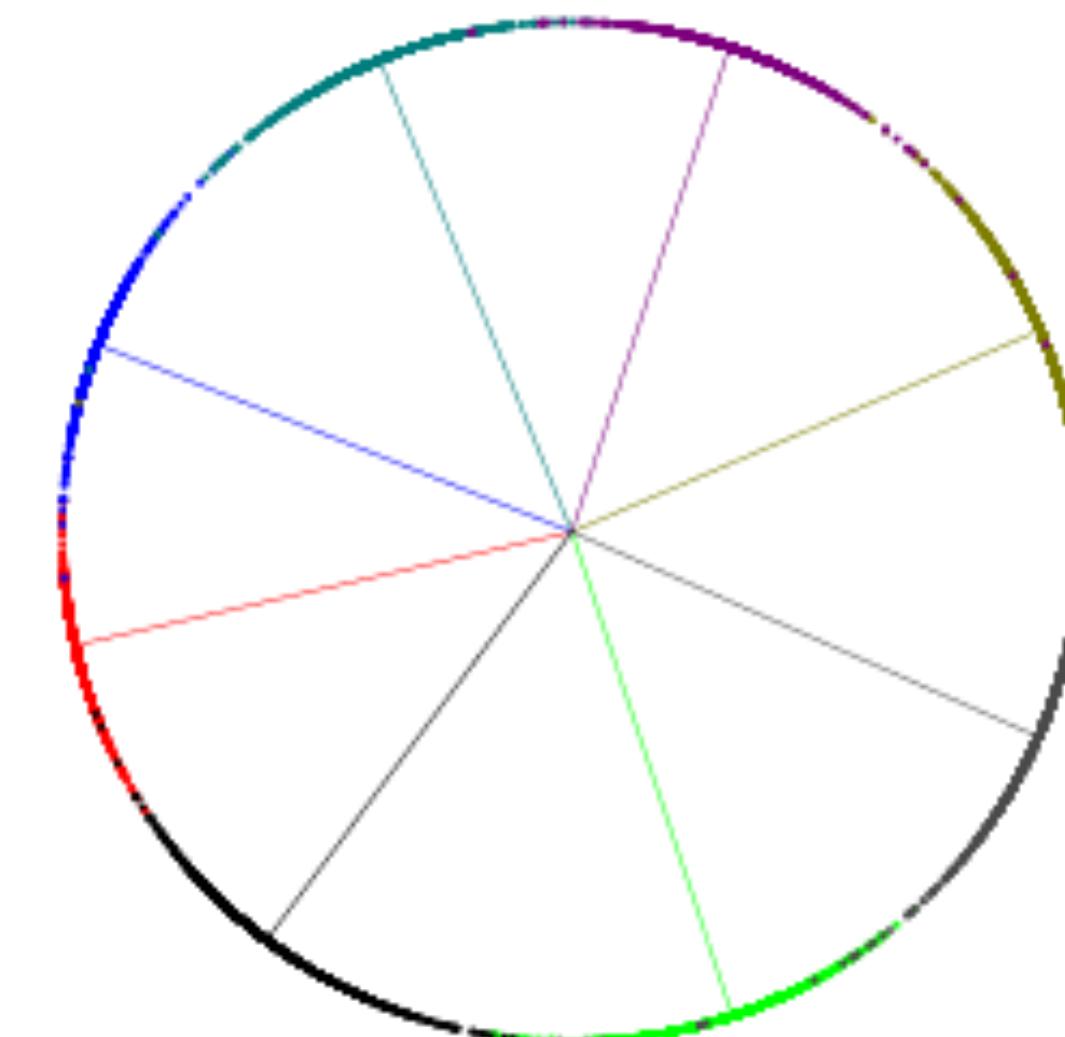
## ArcFace

Current state of the art.

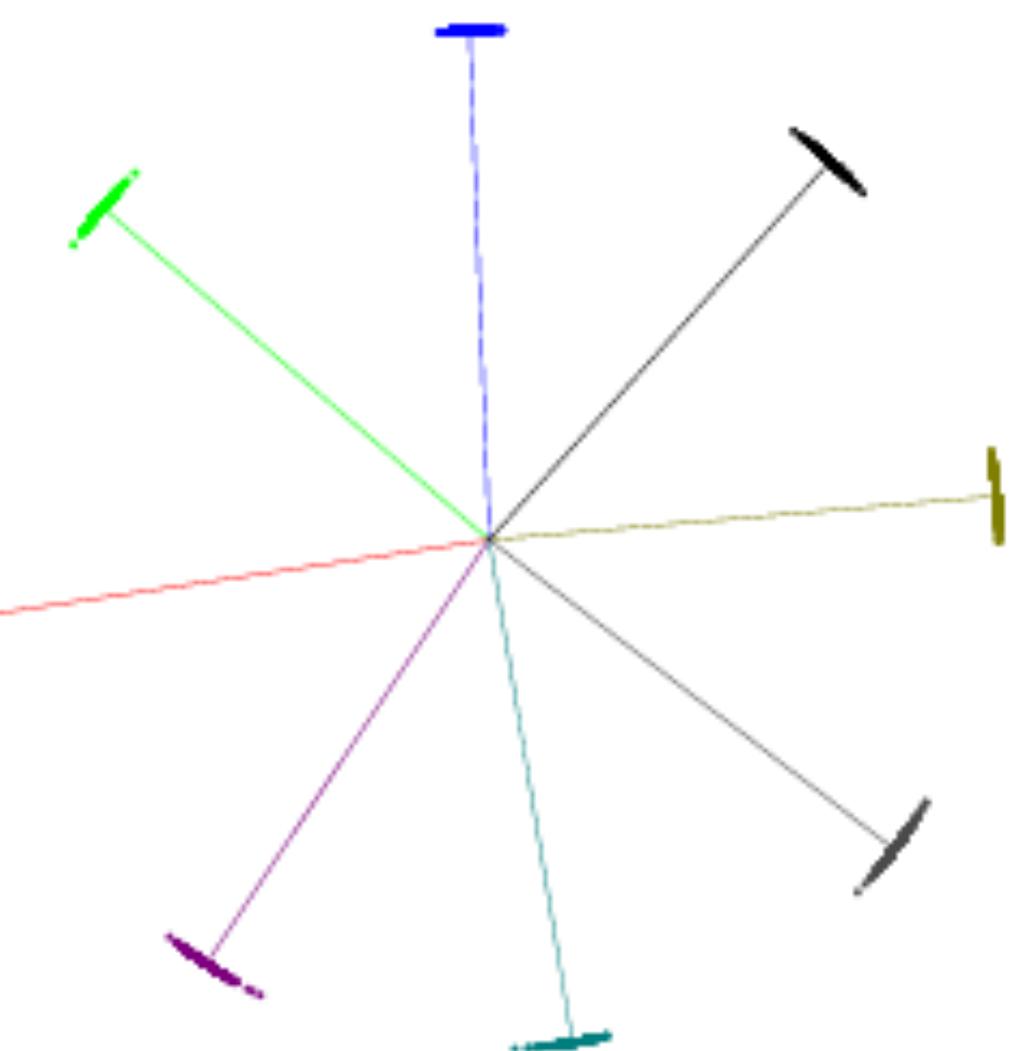
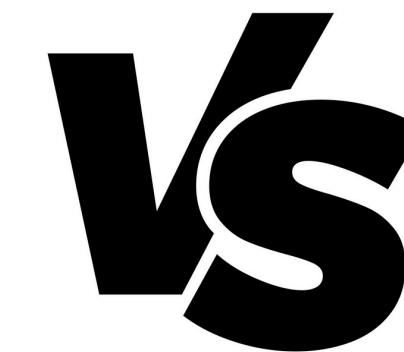
Deng et al. proposed the **additive angular margin loss** to the problem of face recognition.

Deng et al.

*Additive Angular Margin Loss for Deep Face Recognition.*  
CVPR 2019



Margin-less  
class separation



Additive angular  
margin loss

# Improvements

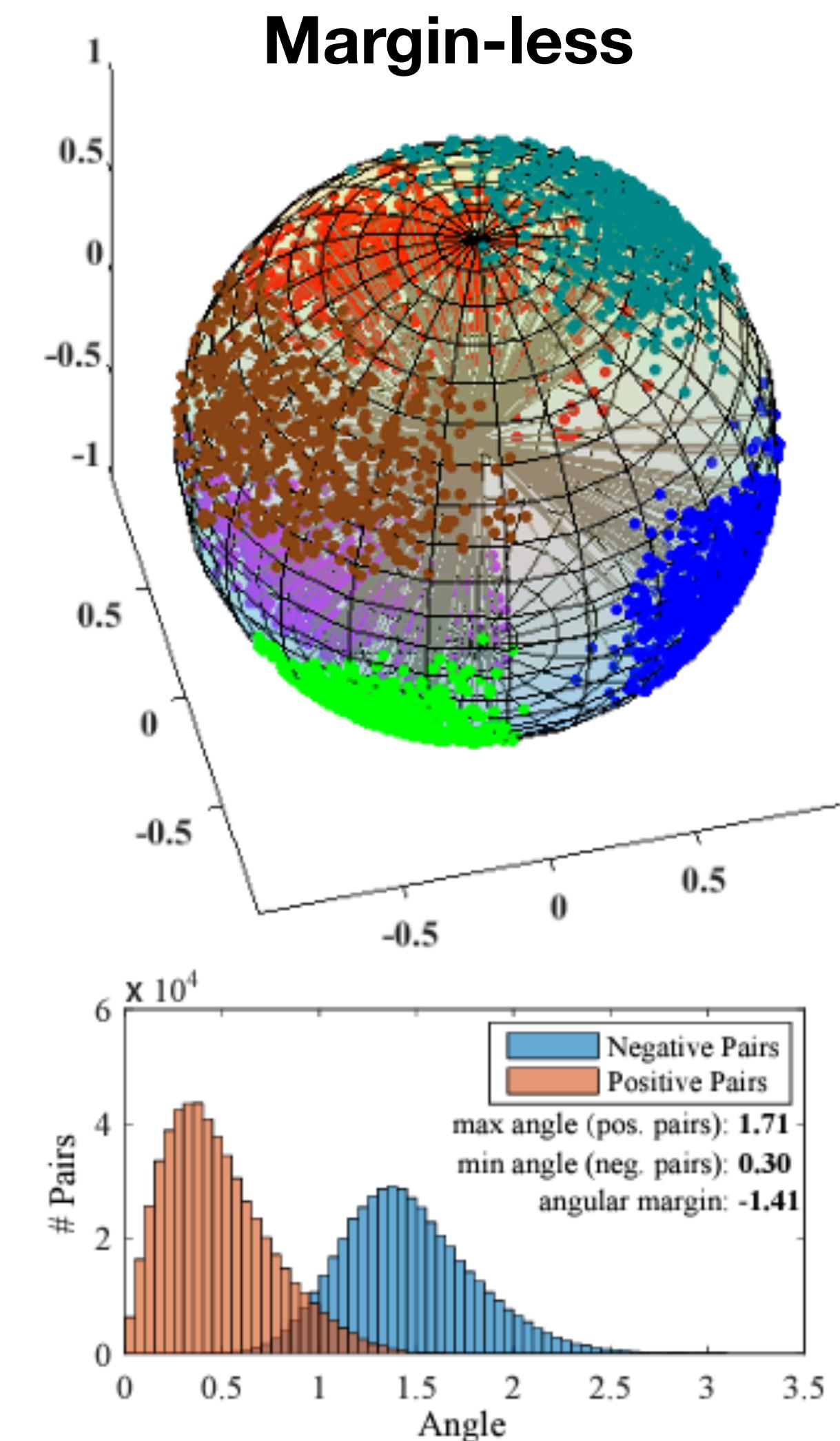
## ArcFace

Current state of the art.

Deng et al. proposed the **additive angular margin loss** to the problem of face recognition.

Deng et al.

*Additive Angular Margin Loss for Deep Face Recognition.*  
CVPR 2019



# Improvements

Source: Liu et al.  
*Sphereface: Deep hypersphere embedding for face recognition.*  
CVPR 2017

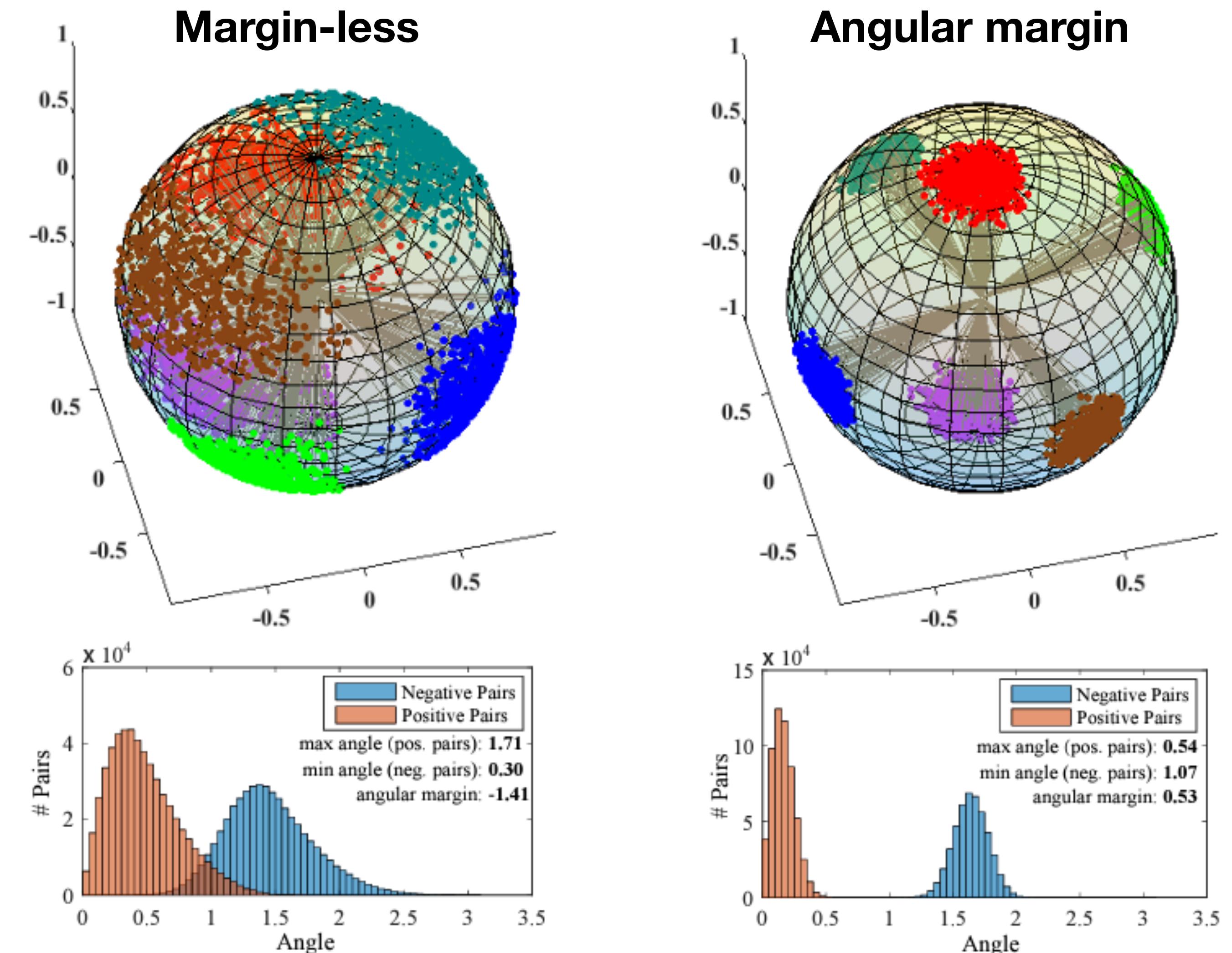
## ArcFace

Current state of the art.

Deng et al. proposed the **additive angular margin loss** to the problem of face recognition.

Deng et al.

*Additive Angular Margin Loss for Deep Face Recognition.*  
CVPR 2019



# Data-Driven Face Recognition

## Problems

### Accountability

You must understand what the network is using to classify samples.

You must avoid this  
in the case of  
Face Recognition!

<https://twitter.com/EricTopol/status/1161657580675985409>

The screenshot shows a tweet from Eric Topol (@EricTopol) with a blue checkmark. The text reads: "How surgical skin markings faked out a deep learning #AI neural net-- a commercially approved product for algorithm-aided melanoma diagnosis. Highly instructive. Machines can be dumb." Below the tweet is a link: [jamanetwork.com/journals/jamaderm](http://jamanetwork.com/journals/jamaderm) by @JAMADerm by @UniHeidelberg.

Below the tweet is a thumbnail image from a medical journal article titled "Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition". The image shows two side-by-side dermoscopic images of a skin lesion. The left image is labeled "Unmarked" and the right image is labeled "E Unmarked".

The journal article abstract includes the following key points:

- IMPORTANCE:** Deep learning convolutional neural networks (CNNs) have shown a performance at the level of dermatologists in the diagnosis of melanoma. Accordingly, further exploring the potential limitations of CNN technology before broadly applying it is of special interest.
- OBJECTIVE:** To investigate the association between gentian violet surgical skin markings in dermoscopic images and the diagnostic performance of a CNN approved for use as a medical device in the European market.
- DESIGN AND SETTING:** A cross-sectional analysis was conducted from August 1, 2018, to November 30, 2018, using a CNN architecture trained with more than 120 000 dermoscopic images of skin neoplasms and corresponding diagnoses. The association of gentian violet skin markings in dermoscopic images with the performance of the CNN was investigated in 3 image sets of 180 melanocytic lesions each (0.07 benign, 0.25 melanocytic).
- EXPOSURES:** The same lesions were sequentially imaged with and without the application of gentian violet surgical skin marker and then evaluated by the CNNs for their probability of being a melanoma. In addition, the markings were removed by manually cropping the dermoscopic images to focus on the melanocytic lesion.
- MAIN OUTCOMES AND MEASURES:** Sensitivity, specificity, and area under the curve (AUC) of the receiver operating characteristic (ROC) curve for the CNNs diagnostic classification in unmarked, marked, and cropped images.
- RESULTS:** In all, 180 melanocytic lesions (107 benign nevi and 23 melanomas) were imaged. In unmarked lesions, the CNN achieved a sensitivity of 95.7% (95% CI, 79.6–99.2%) and a specificity of 84.7% (95% CI, 76.0%–89.8%). The ROC AUC was 0.969. In marked lesions, an increase in melanoma probability scores was observed that resulted in a sensitivity of 100% (95% CI, 85.7%–100%) and a significantly reduced specificity of 45.8% (95% CI, 36.7%–55.2%;  $P < .001$ ). The ROC AUC was 0.932. Cropping images led to the highest sensitivity of 100% (95% CI, 85.7%–100%) and specificity of 97.2% (95% CI, 92.1%–99.0%), and ROC AUC of 0.993. Heat maps created by vanilla gradient descent backpropagation indicated that the blue markings were associated with the increased false positive rate.

# Data-Driven Face Recognition

## Problems

### Accountability

You must understand what the network is using to classify samples.

Comments on:

<https://www.youtube.com/watch?v=rga2-d1oi30>

### Automated Inference on Criminality using Face Images

Xiaolin Wu  
Shanghai Jiao Tong University  
xwu510@gmail.com

Xi Zhang  
Shanghai Jiao Tong University  
zhangxi.19930818@sjtu.edu.cn

#### Abstract

*We study, for the first time, automated inference on criminality based solely on still face images. Via supervised machine learning, we build four classifiers (logistic regression, KNN, SVM, CNN) using facial images of 1856 real persons controlled for race, gender, age and facial expressions, nearly half of whom were convicted criminals, for discriminating between criminals and non-criminals. All four classifiers perform consistently well and produce evidence for the validity of automated face-induced inference on criminality based on facial images.*

people share the belief that the face alone suffices to reveal innate traits of a person. Aristotle in his famous work Prior Analytics asserted, "It is possible to infer character from features, if it is granted that the body and the soul are changed together by the natural affections". Psychologists have known, for as long as a millennium, the human tendency of inferring innate traits and social attributes (e.g., the trustworthiness, dominance) of a person from his/her facial appearance, and a robust consensus of individuals' inferences . These are the facts found through numerous studies [2, 32, 4, 5, 9, 20, 21, 27, 25].

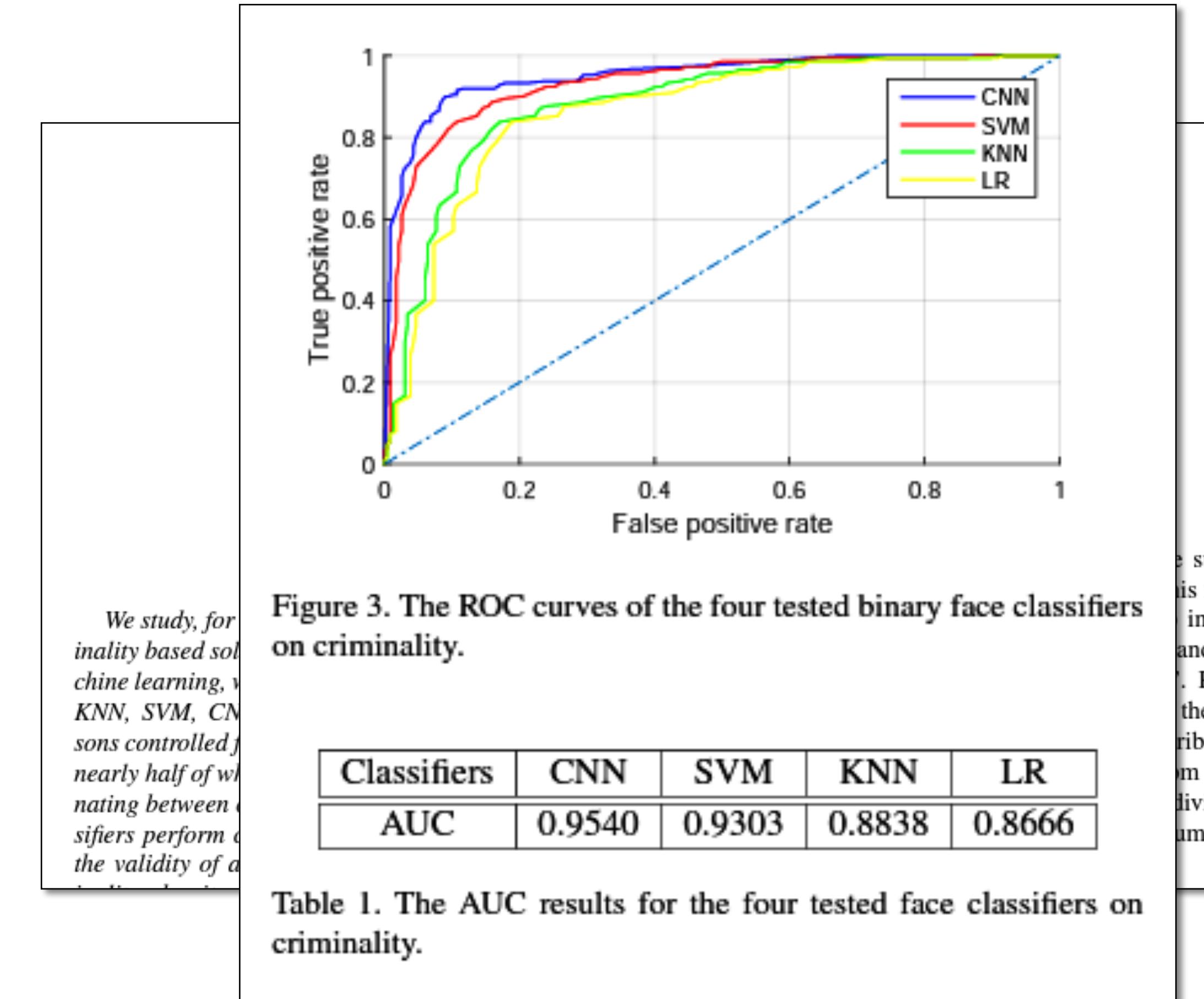
# Data-Driven Face Recognition

## Problems

### Accountability

You must understand what the network is using to classify samples.

Comments on:  
<https://www.youtube.com/watch?v=rga2-d1oi30>



e suffices to re-  
is famous work  
infer character  
and the soul are  
. Psychologists  
the human ten-  
ributes (e.g., the  
in his/her facial  
individuals' infer-  
umerous studies

# Data-Driven Face Recognition

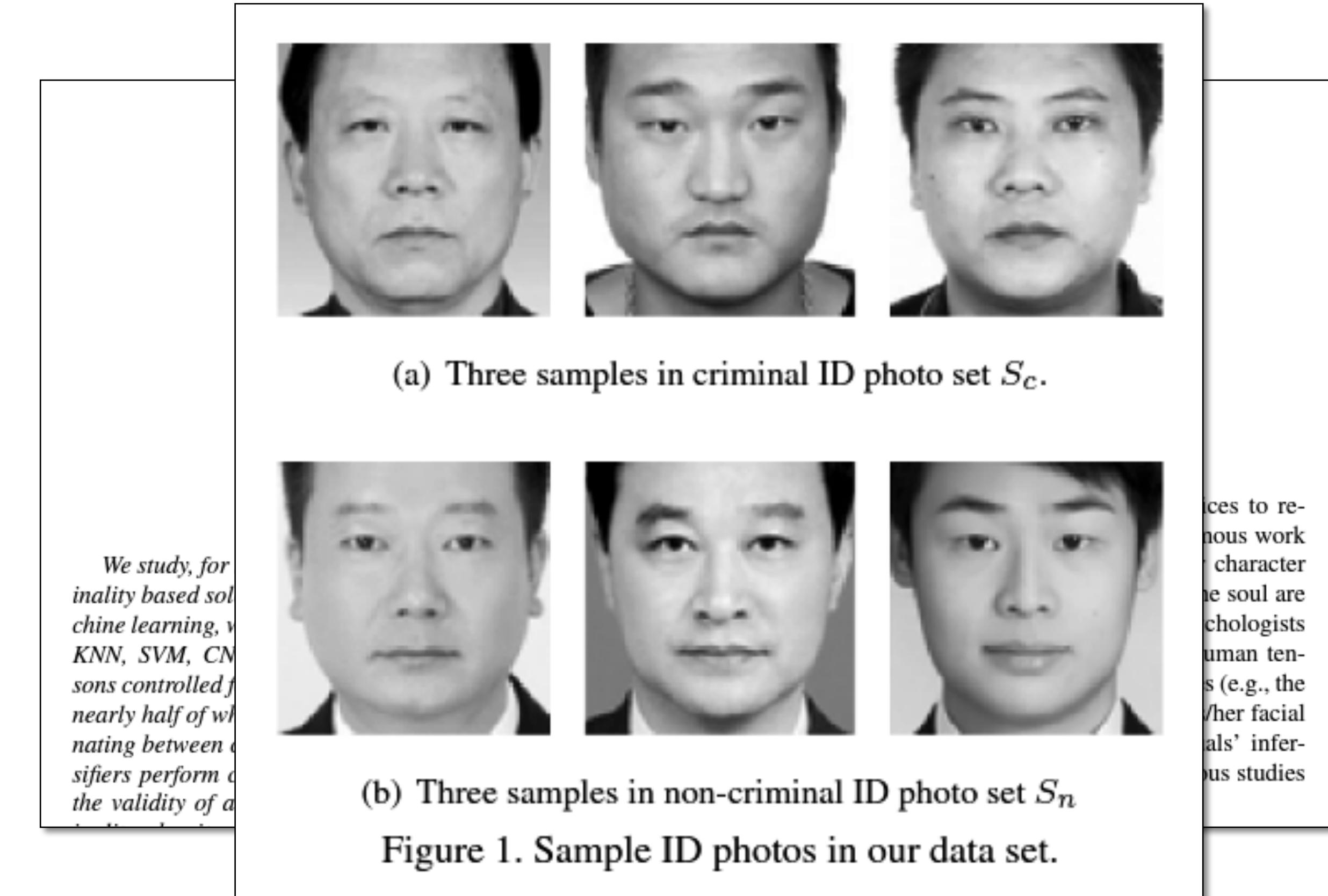
## Problems

### Accountability

You must understand what the network is using to classify samples.

Comments on:

<https://www.youtube.com/watch?v=rga2-d1oi30>



# Data-Driven Face Recognition

## Problems

### Accountability

You must understand what the network is using to classify samples.

#### Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images

By Michal Kosinski, Yilun Wang

*Journal of Personality and Social Psychology*. February 2018, Vol. 114, Issue 2, Pages 246-257.

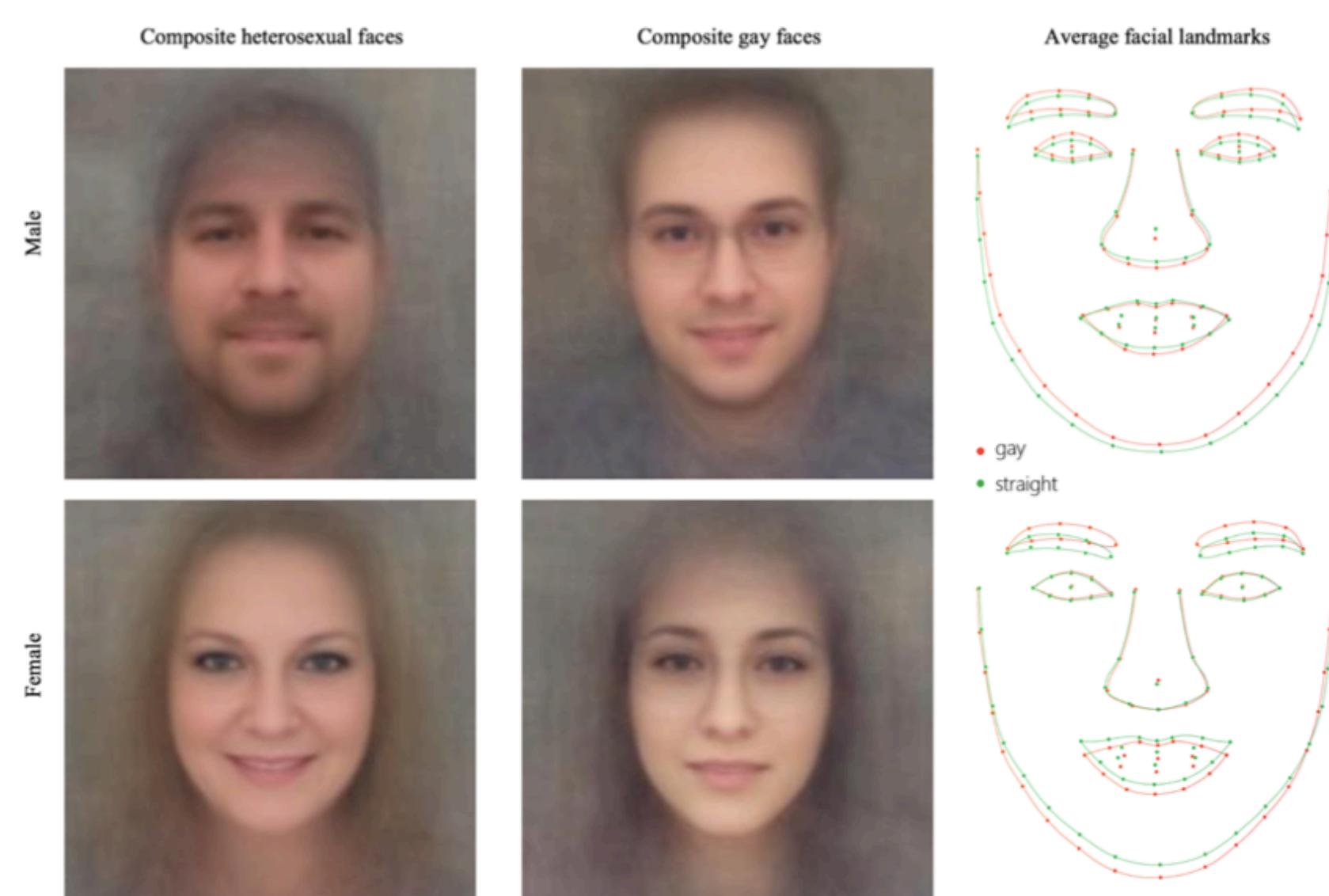
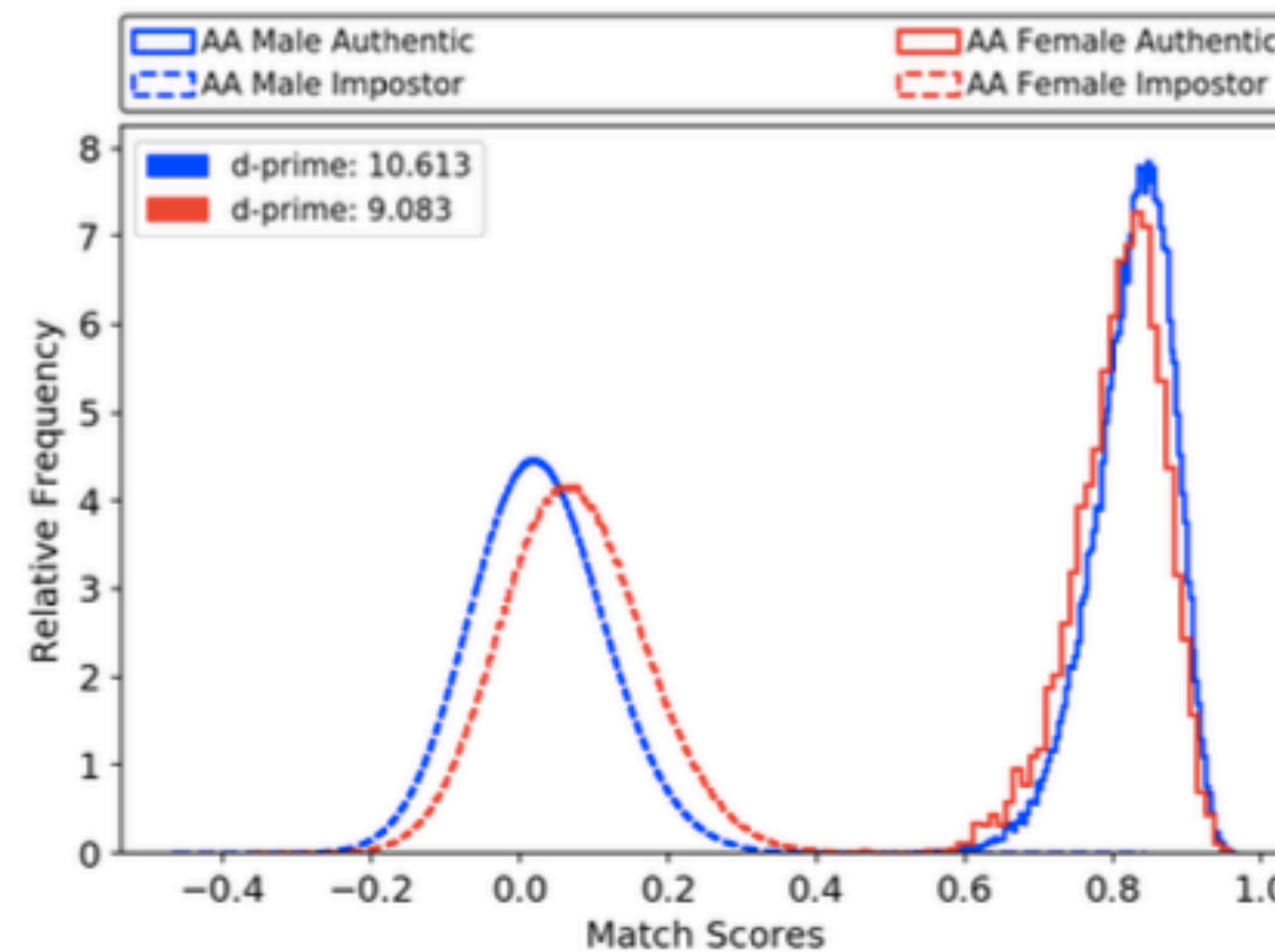


Figure 4. Composite faces and the average facial landmarks built by averaging faces classified as most and least likely to be gay.

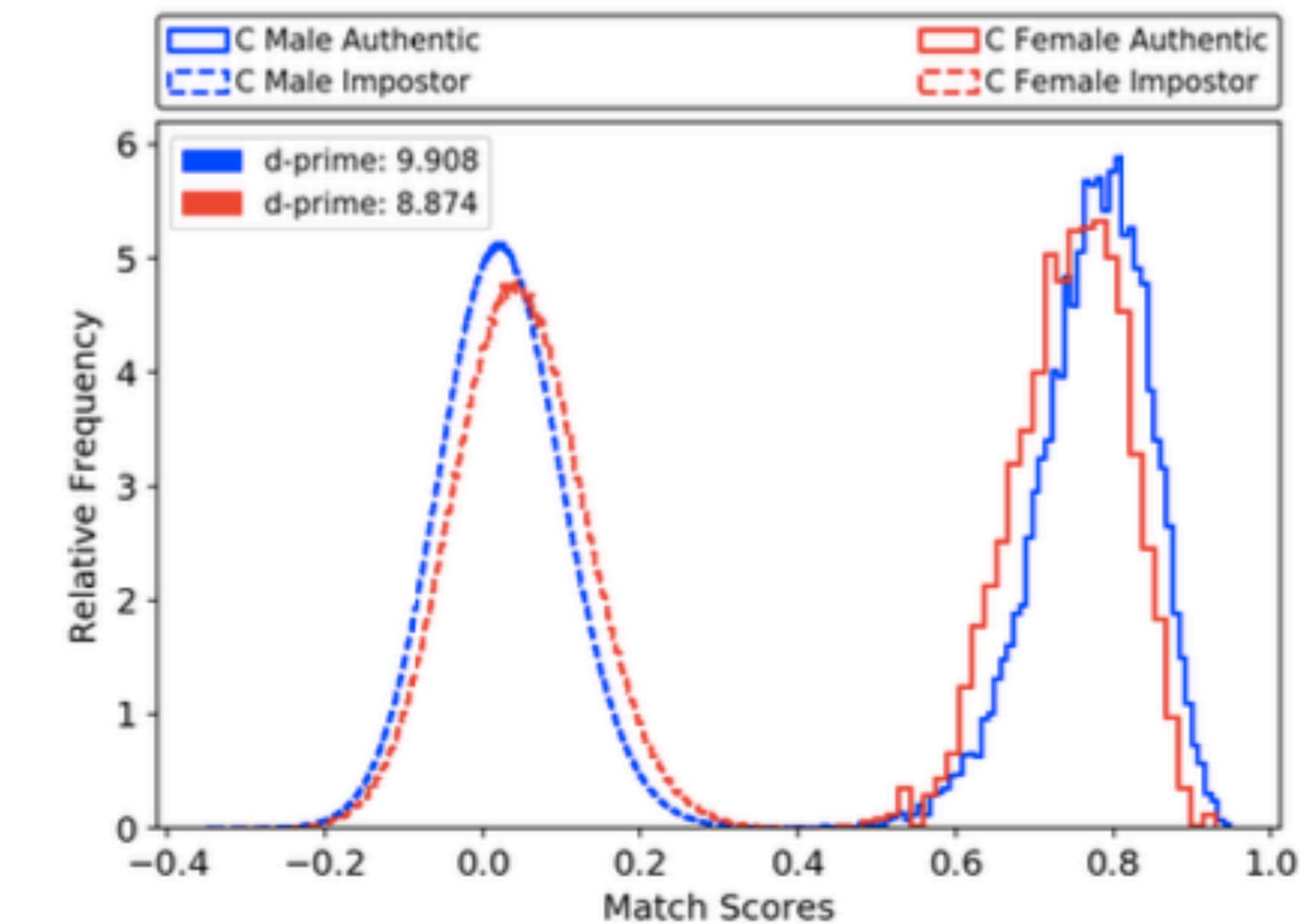
# Data-Driven Face Recognition

## Notre Dame Preliminary Studies Dr. Bowyer at CVRL

ArcFace performance  
trained on MORPH  
dataset.



(a) MORPH African American



(b) MORPH Caucasian

# Data-Driven Face Recognition

## Notre Dame Preliminary Studies

Dr. Bowyer at CVRL

ArcFace performance  
trained on MORPH  
dataset.

### MORPH: A Longitudinal Image Database of Normal Adult Age-Progression

Karl Ricanek Jr., IEEE Senior Member  
Department of Computer Science  
University of North Carolina Wilmington  
Wilmington, North Carolina, USA  
RICANEKK@UNCW.EDU

Tamirat Tesafaye  
Department of Computer Science  
Addis Ababa University  
Addis Ababa, Ethiopia  
TAMIRAT@PROGRAMMER.NET

#### 3.2. Statistics

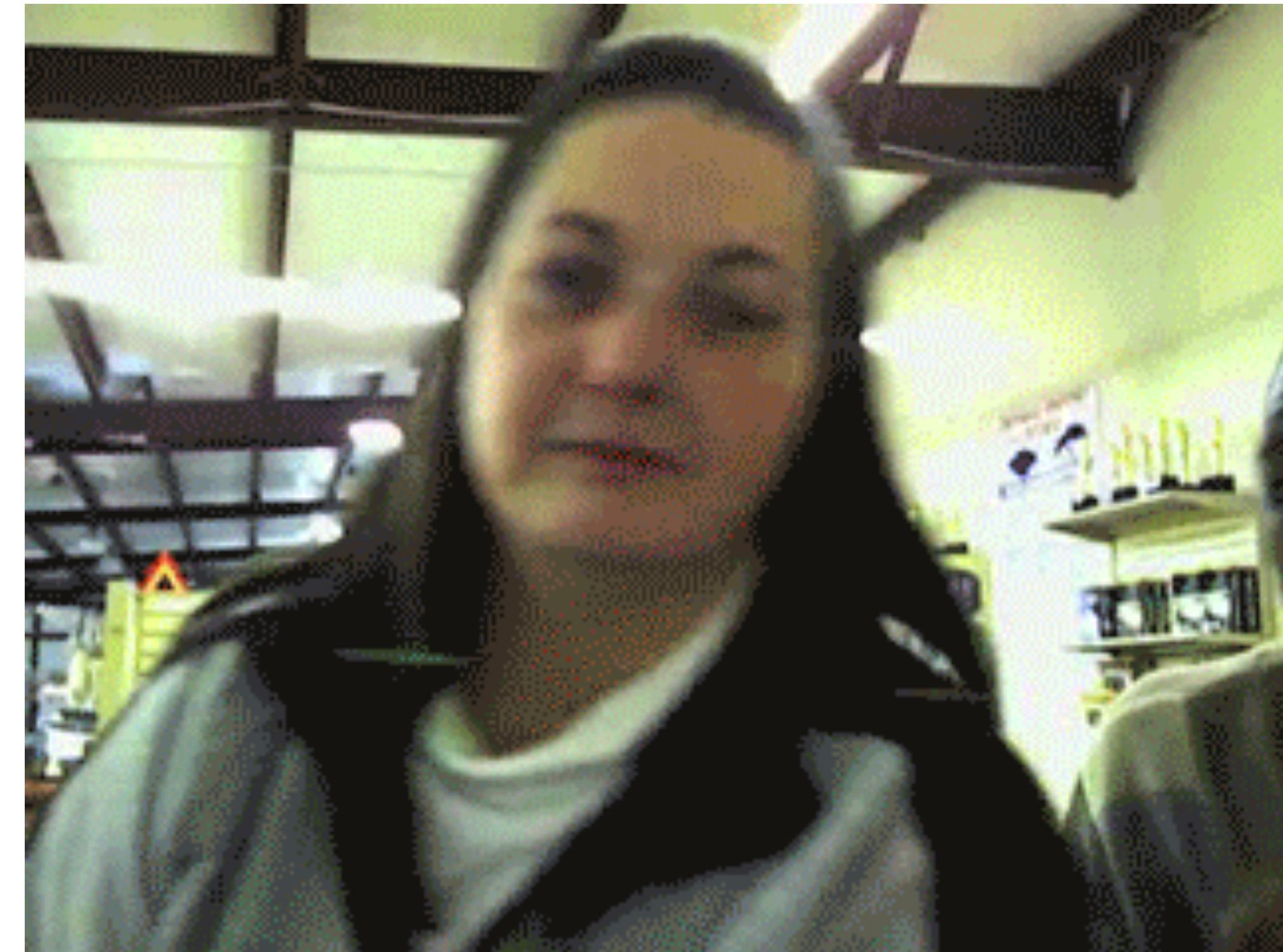
As of this writing, the database contains 1,724 face images of 515 individuals. These images represent a diverse population with respect to age, gender, and ethnicity. There are 1,278 images of individuals of African-American decent, 433 images of individuals of Caucasian decent and 3 images classified as other. There are 294 images of females and 1,430 images of males. For the male images, seventy-six percent have some form of facial hair, usually a mustache.

# Data-Driven Face Recognition

## Problems

### Bias

What happens if you train the network only with one type of faces (e.g., with only young caucasians)?



# Data-Driven Face Recognition

## Problems

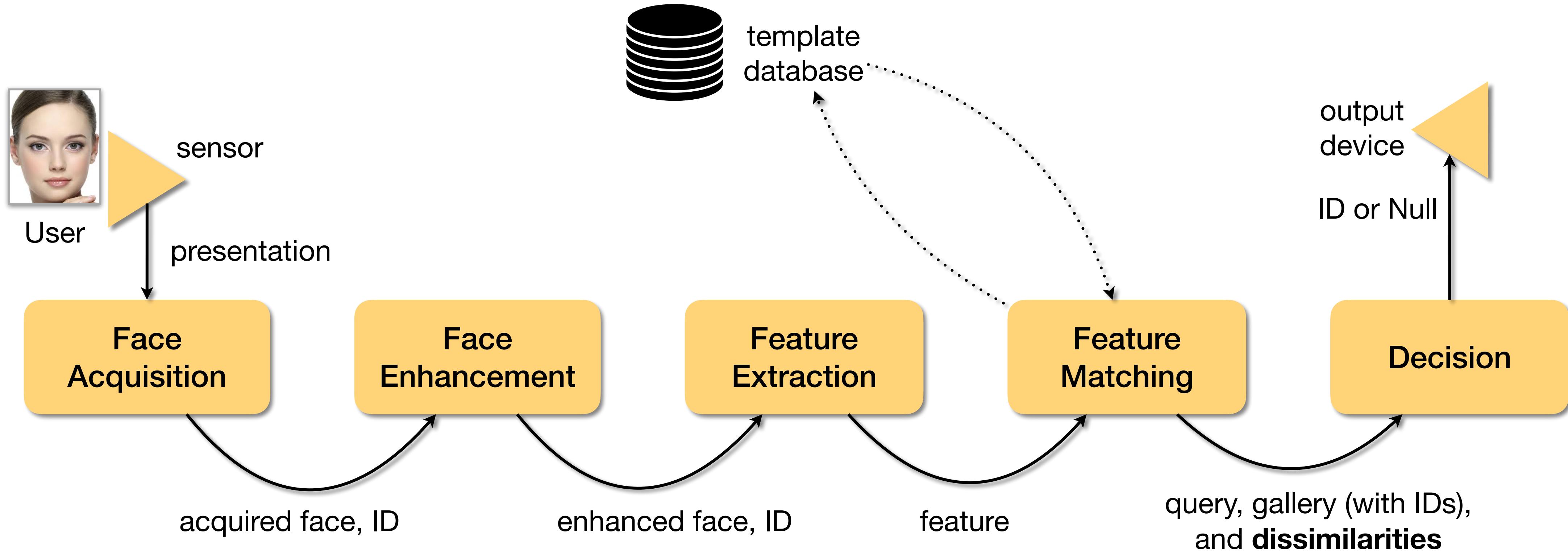
### Avoid Bias

Diversify the training dataset.

There are synthetic ways to do it...  
(FaceGen demonstration)



# S'up Next?



# S'up Next?

**Face Recognition Coding Class**

Please bring your computers.

