Face Recognition IV

COMP 388-002/488-002 Biometrics

Daniel Moreira

Fall 2023

LOYOLA UNIVERSITY CHICAGO

Today you will…

Get to know
Deep learning-based face recognition.

LOYOLA UNIVERSITY CHICAGO

Today's attendance

Please fill out the form

https://forms.gle/fMtd4K4DC3LEhWPMA

Feature Extraction

RECAP

Focus
2D-appearance-based methods.

Types
Handcrafted features from Computer Vision.

Data-driven learned features with Machine Learning.

LOYOLA UNIVERSITY CHICAGO

Feature Extraction

RECAP

Focus
2D-appearance-based methods.

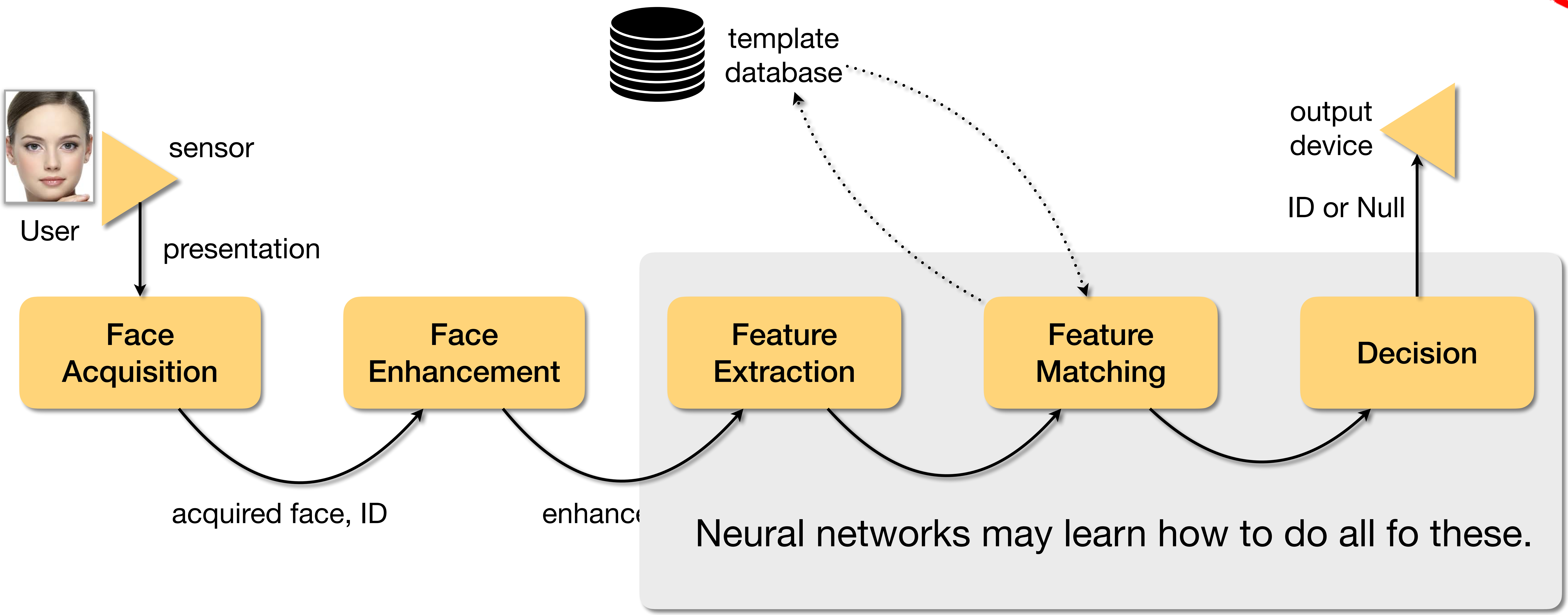
Types
Handcrafted features from Computer Vision.

Data-driven learned features with Machine Learning.

LOYOLA UNIVERSITY CHICAGO

Face Recognition

RECAP



LOYOLA
UNIVERSITY CHICAGO

Data-driven Face Recognition

RECAP

Deep Convolutional Neural Networks (CNN)

Data-driven Face Recognition

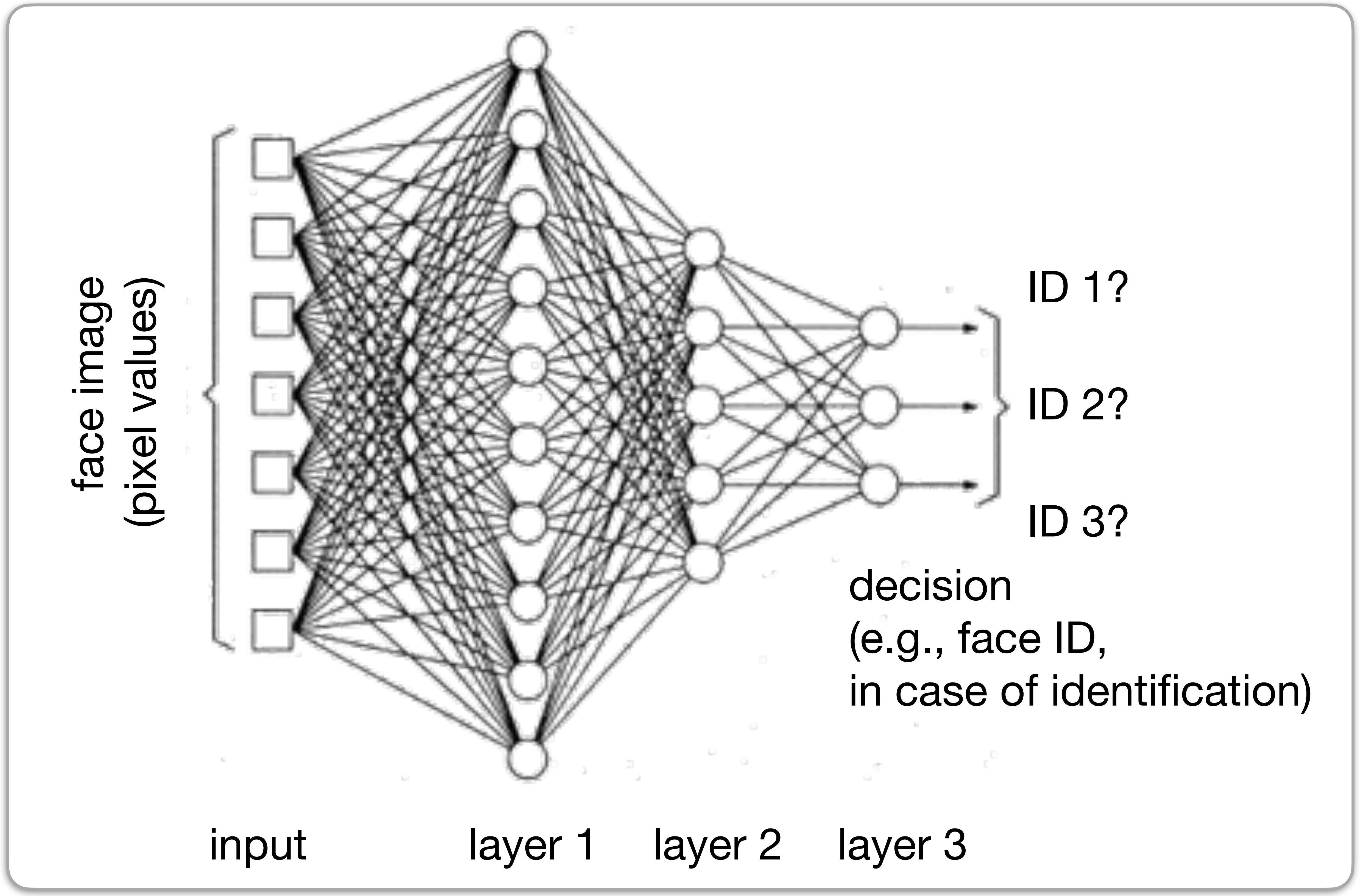
RECAP

Deep Convolutional Neural Networks (CNN)

From pixels to classification decision.

Hierarchy of feature extractors.

Each layer extracts features from previous layer.

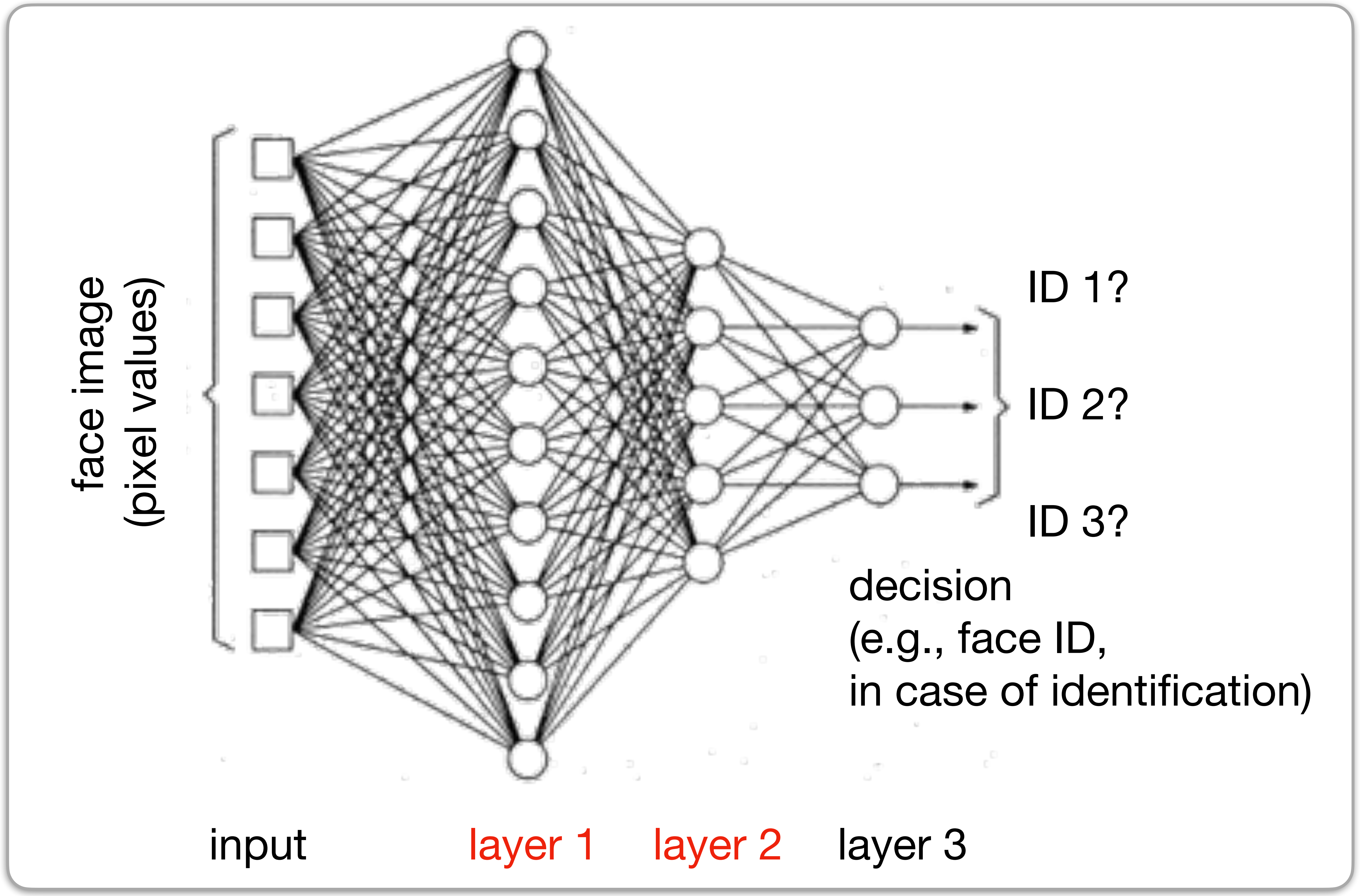


Data-driven Face Recognition

RECAP

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.



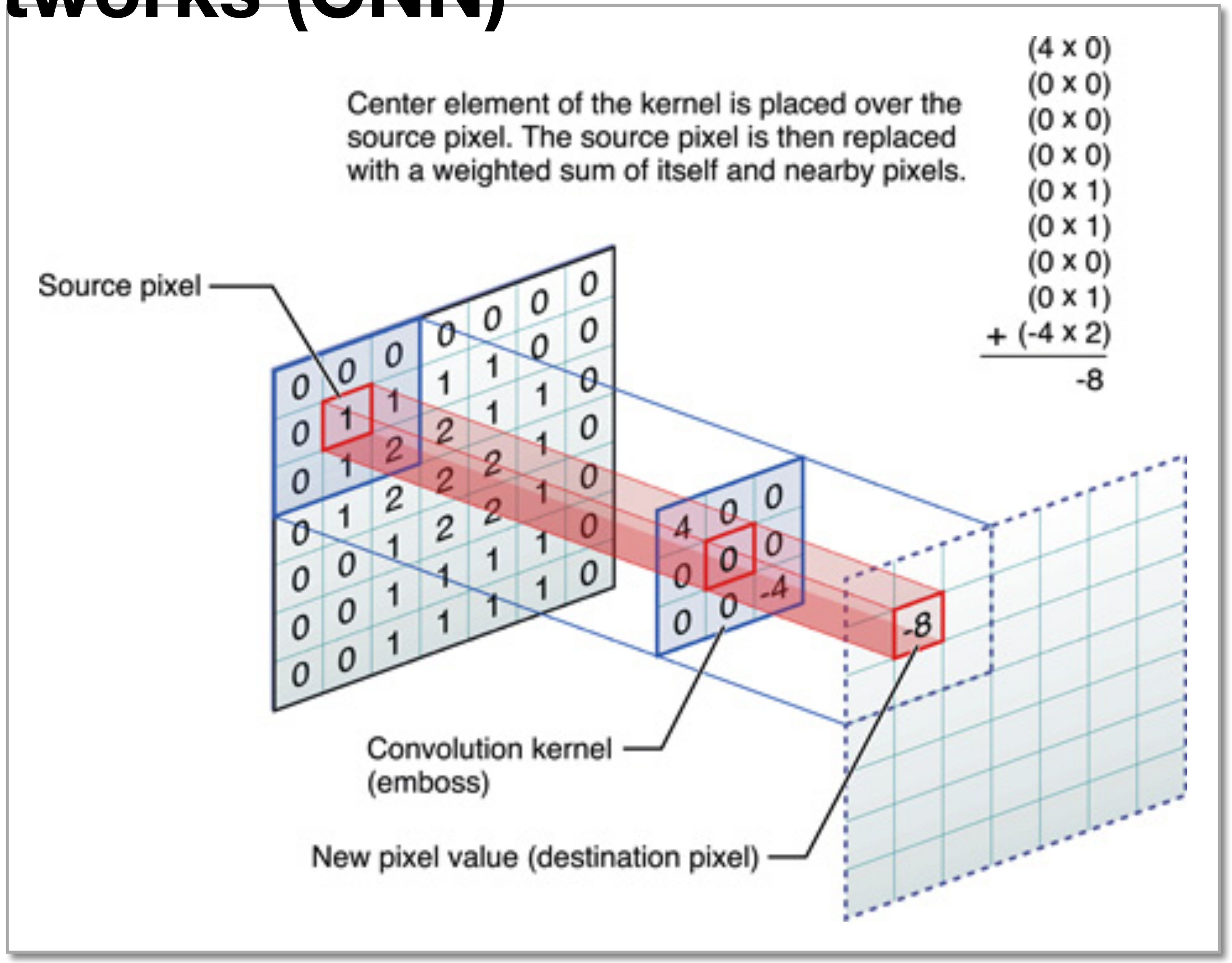
LOYOLA
UNIVERSITY CHICAGO

Data-driven Face Recognition

RECAP

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

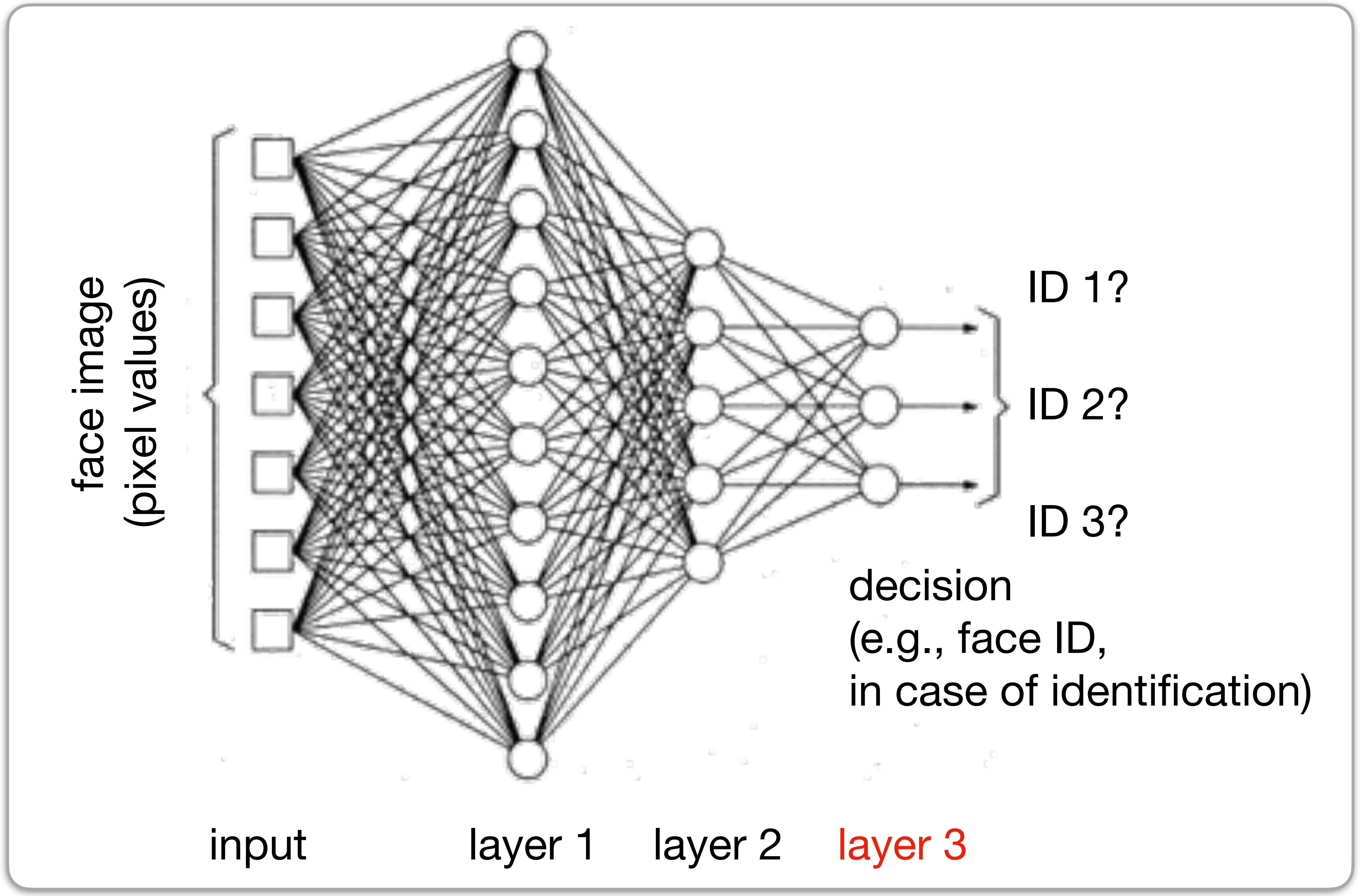
LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

RECAP

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).



LOYOLA
UNIVERSITY CHICAGO

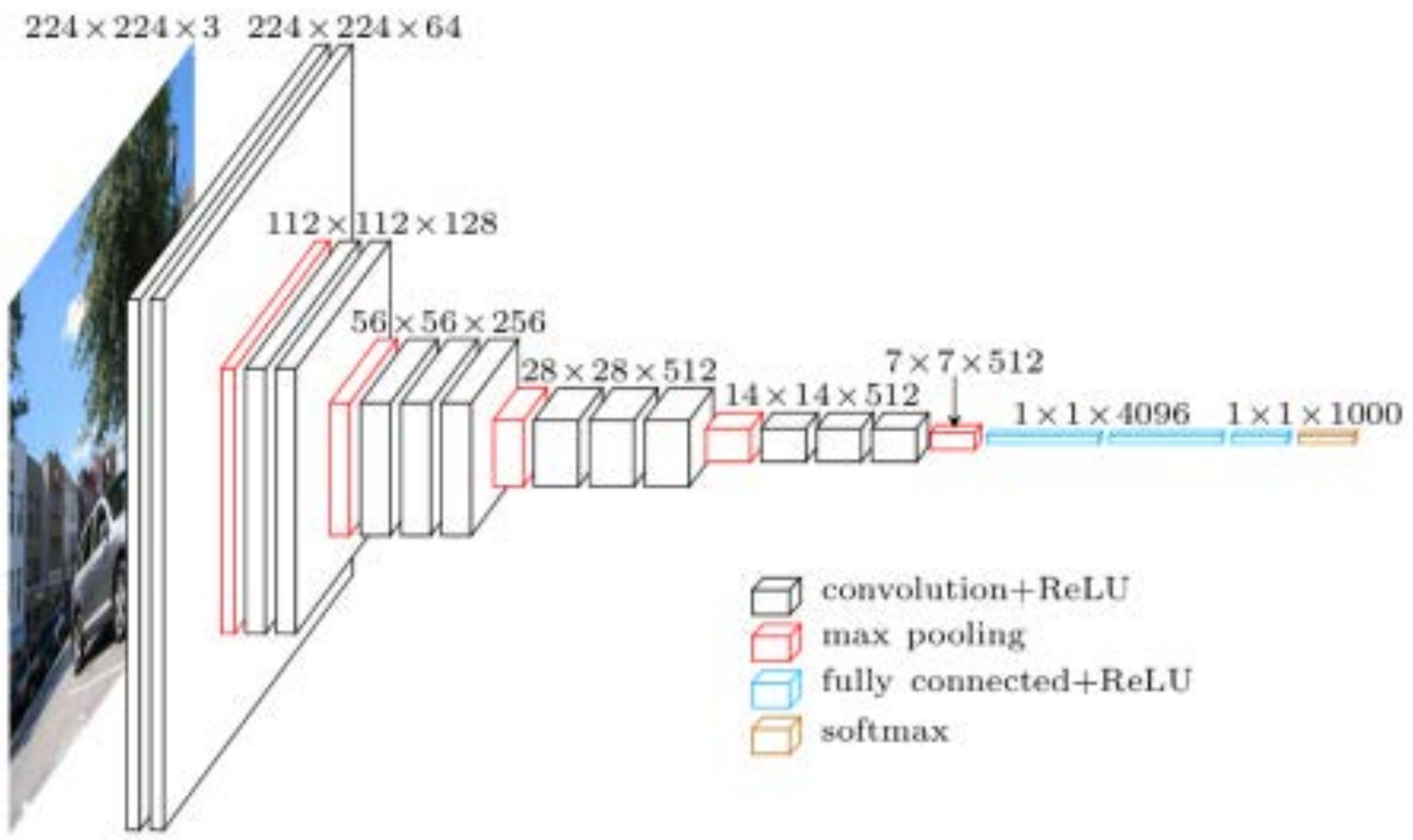
Data-driven Face Recognition

RECAP

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



16 to 19 convolutional layers

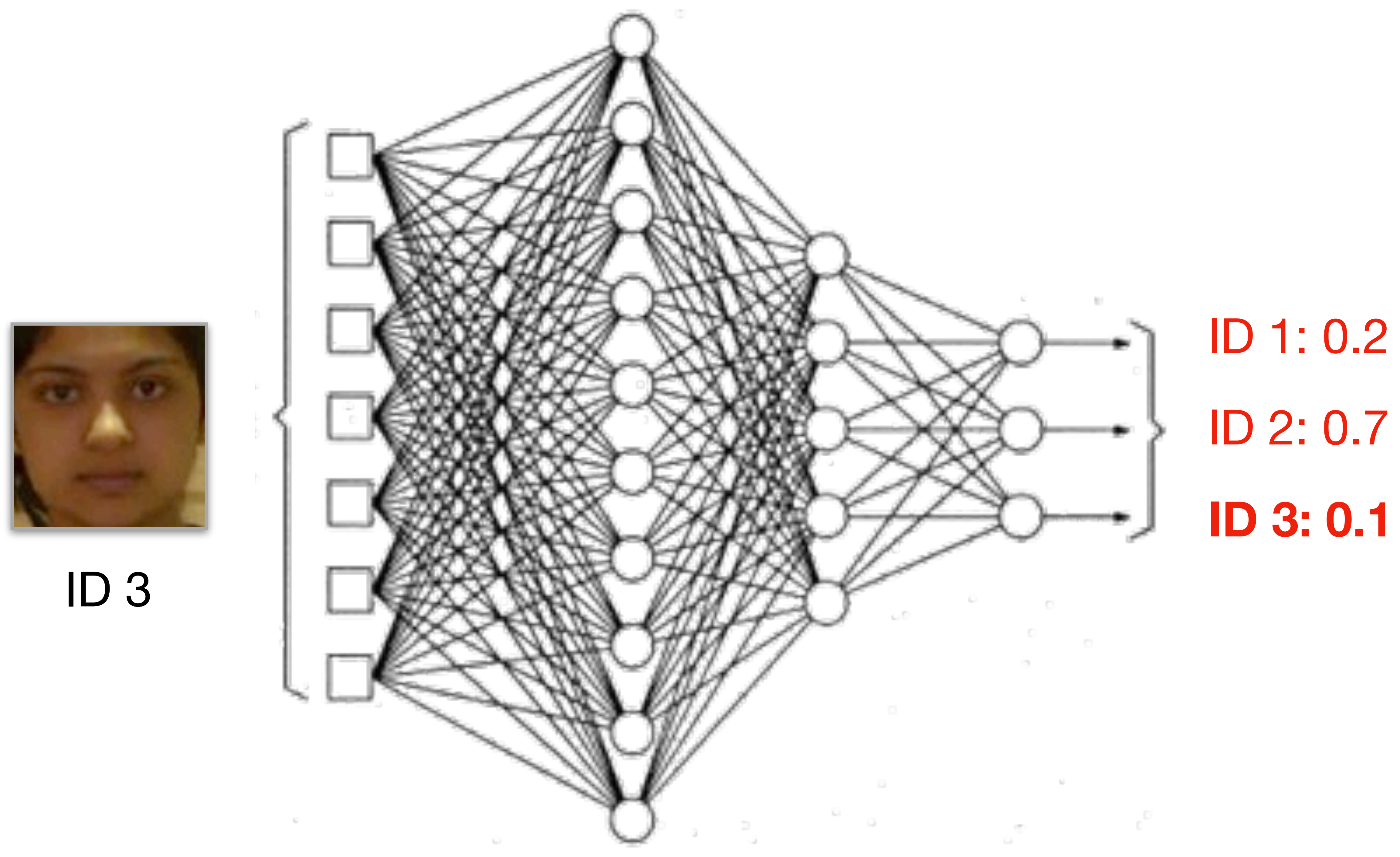
LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

RECAP

Deep Learning

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



Random weights in the beginning.

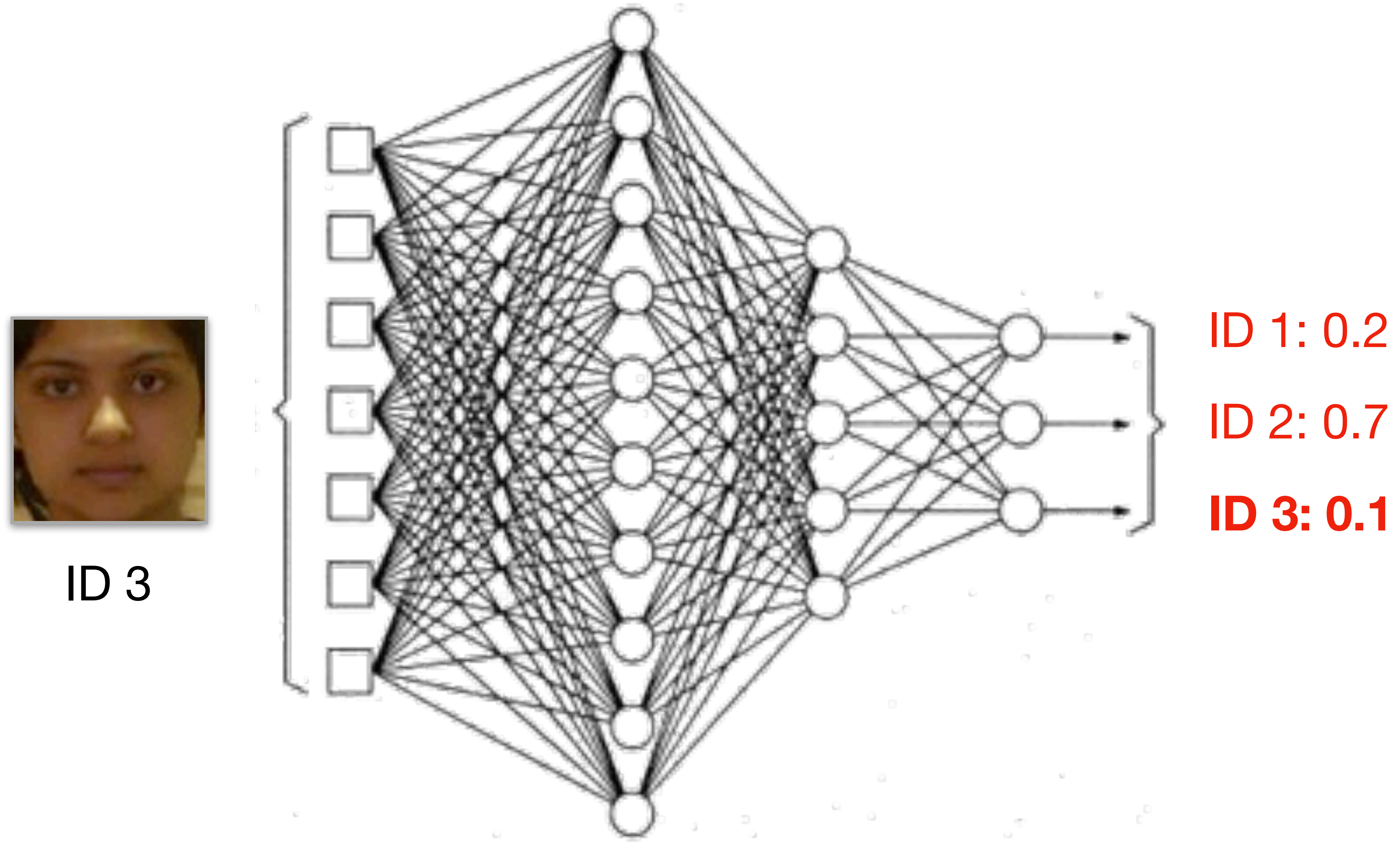
Data-driven Face Recognition

RECAP

Deep Learning

Training

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

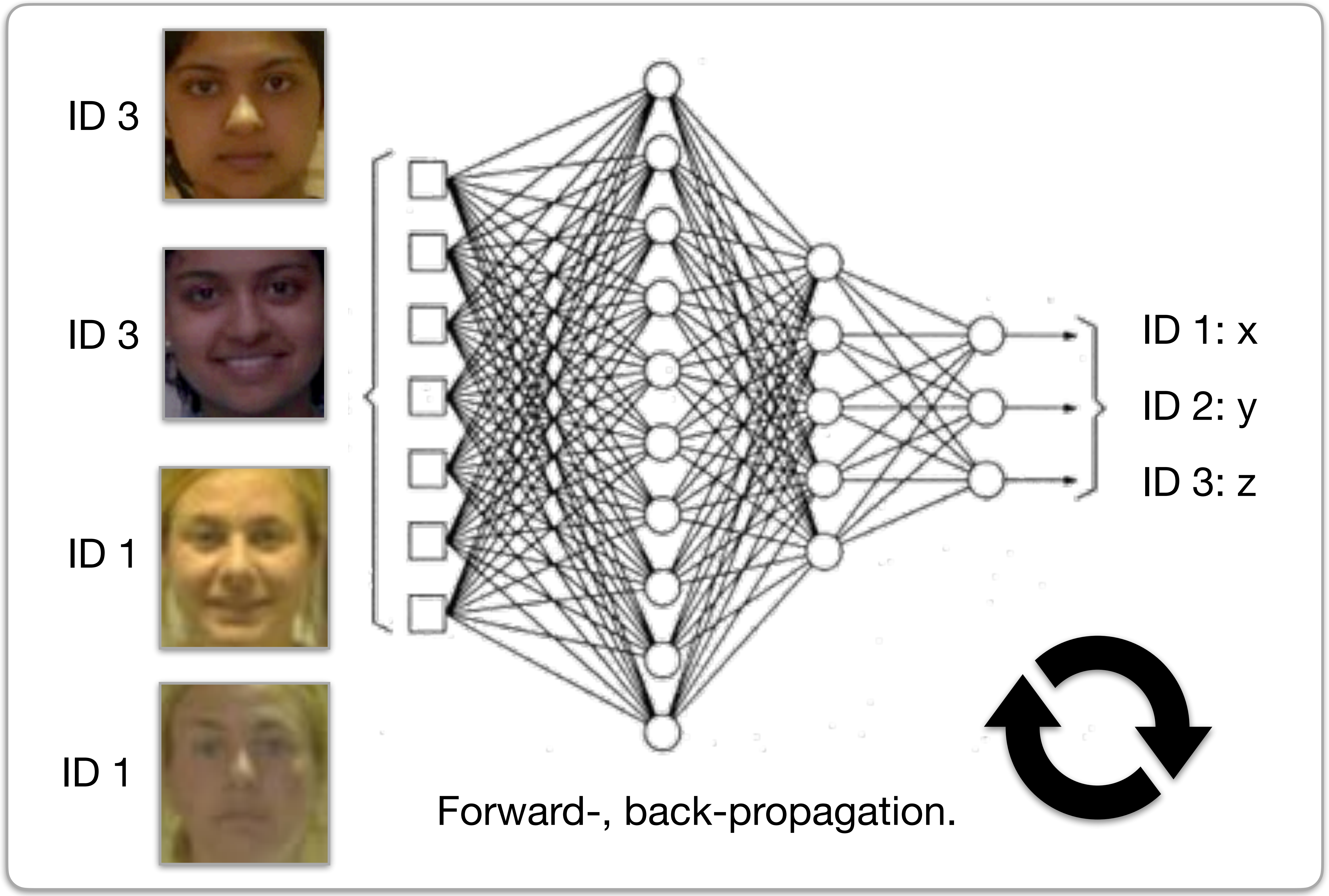


Weights are fixed with back-propagation.

Data-driven Face Recognition

Deep Learning

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



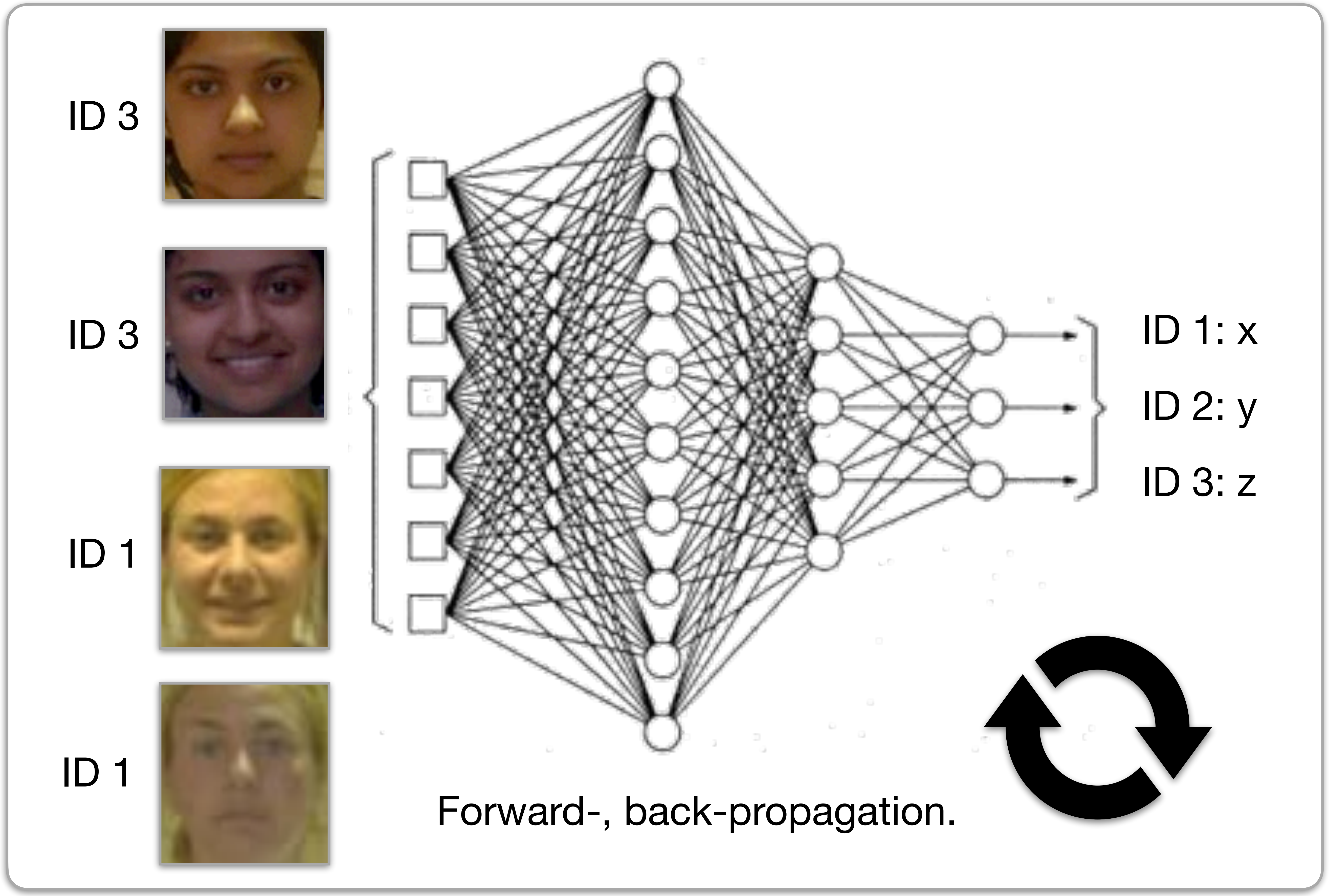
LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

Deep Learning

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

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



LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

Deep Learning

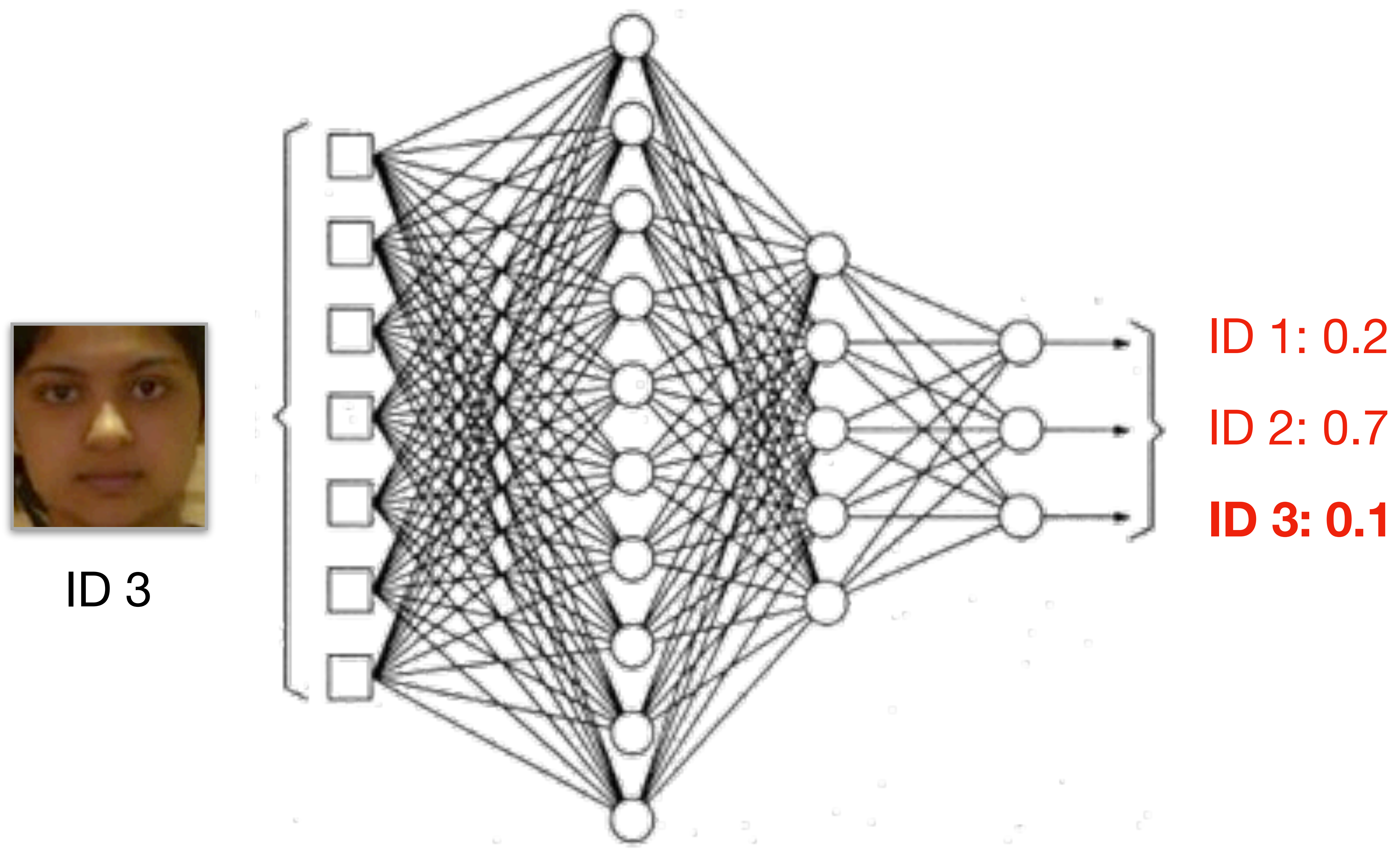
Cross-entropy Loss (CE)

#training faces

#people's IDs

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

CNN output for ID



Random weights in the beginning.

LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

Deep Learning

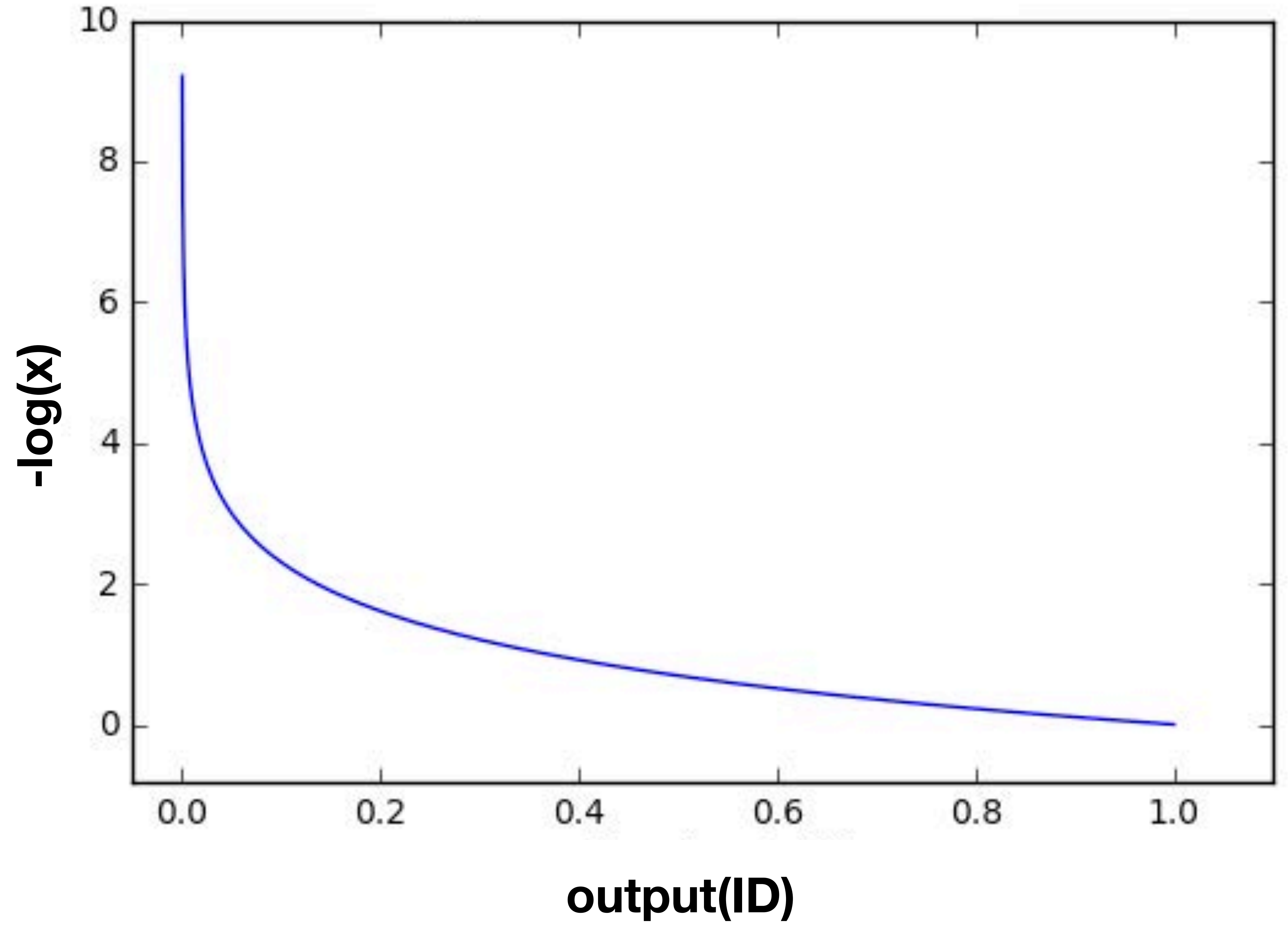
Cross-entropy Loss (CE)

#training faces

#people's IDs

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

CNN output for ID



LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

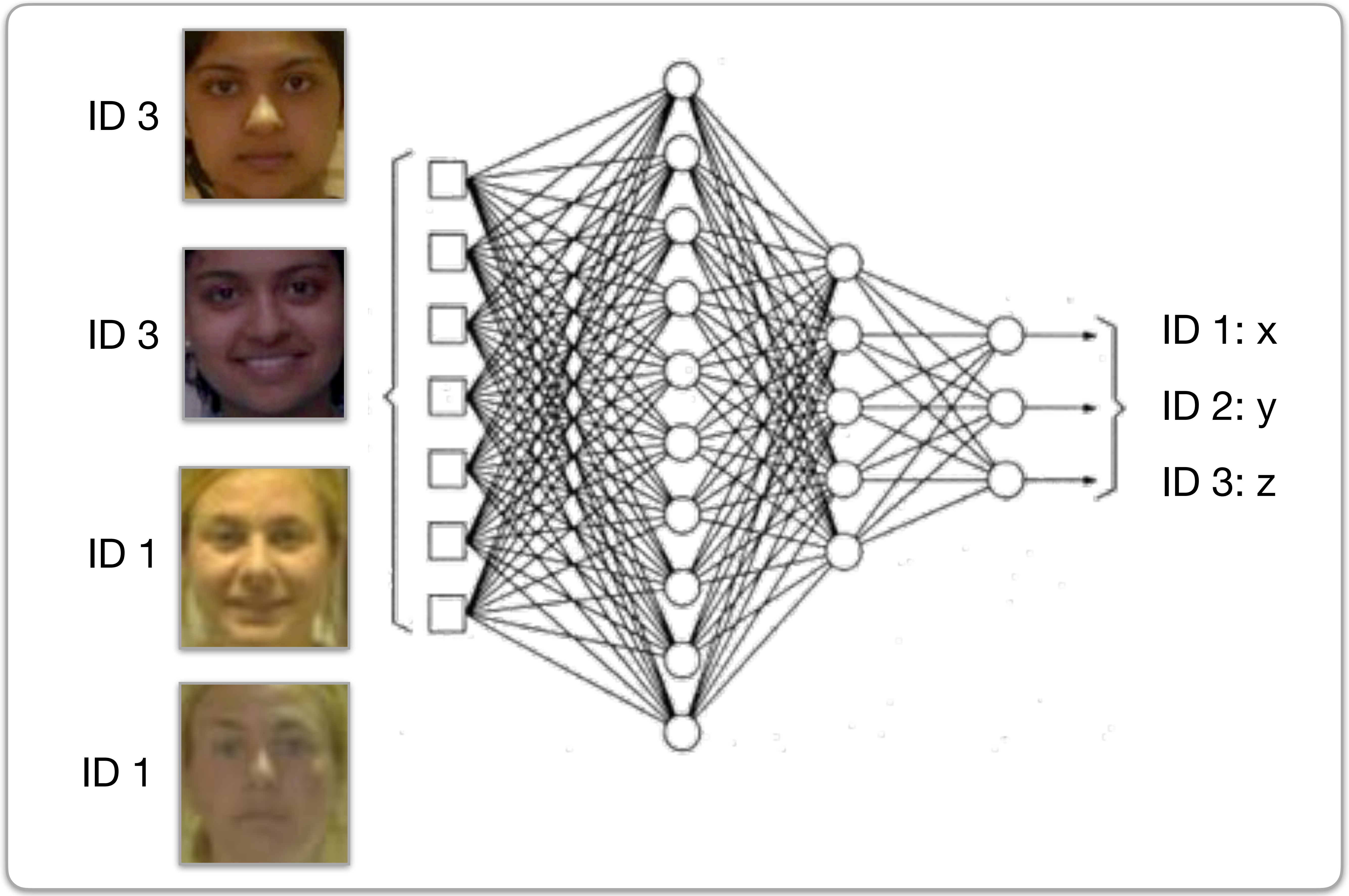
RECAP

Deep Learning

What are the cons here?
How to enroll a new person?

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

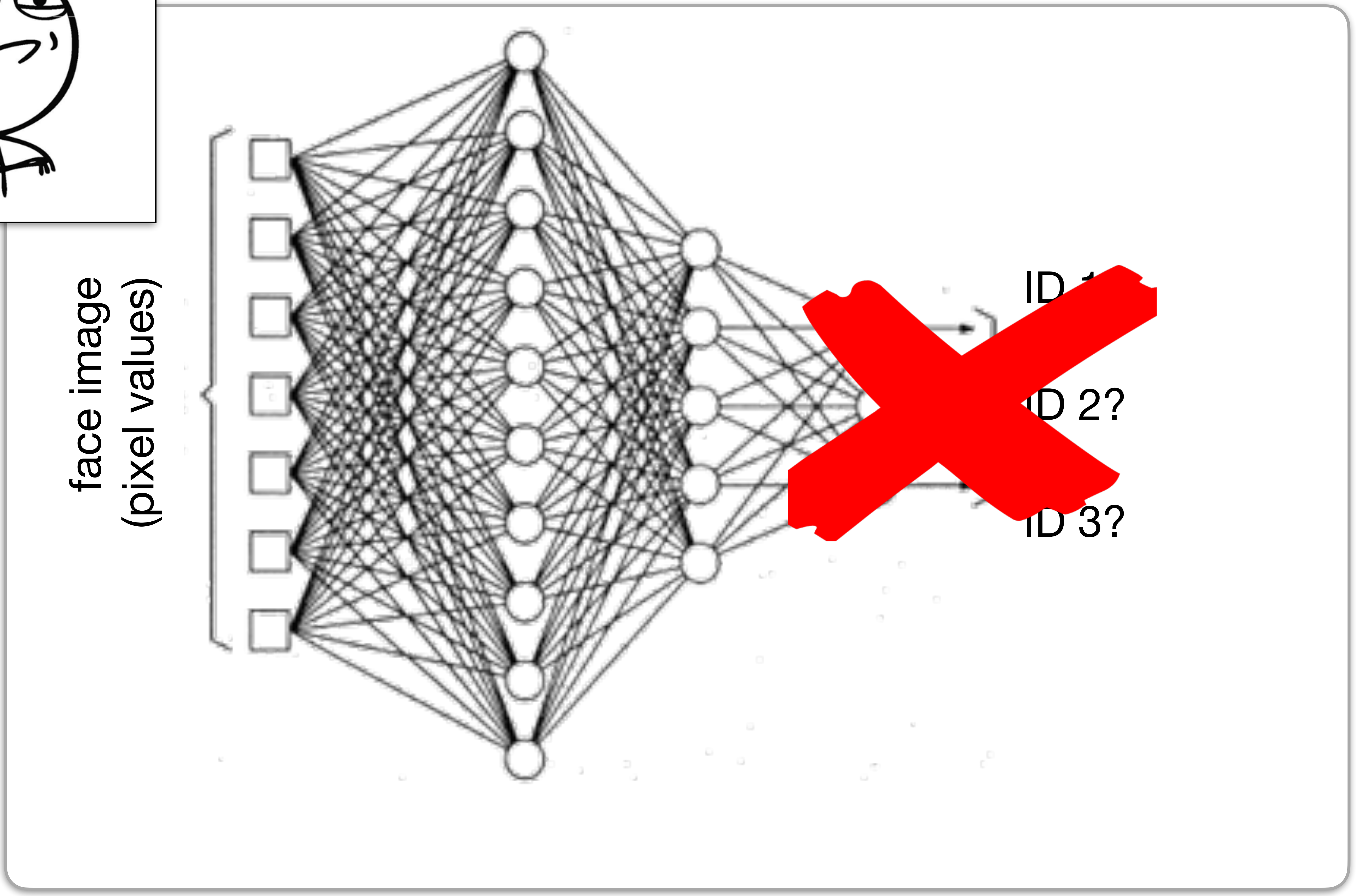
Need for large training dataset (thousands of sample per class).



LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

Deep Learning

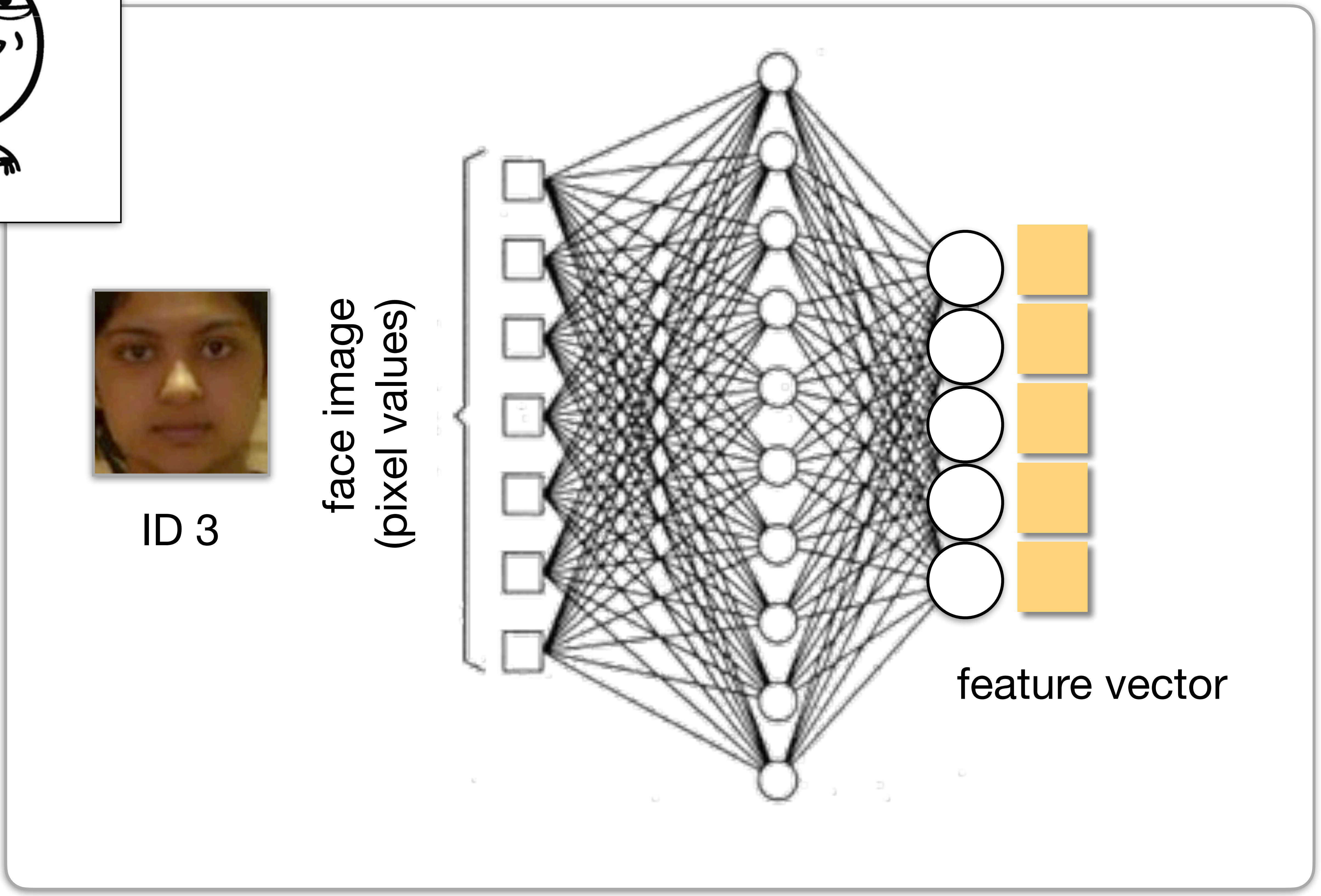


How to make CNN more flexible?
Remove fully connected layer and use last convolutional layers as a feature descriptor.

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.

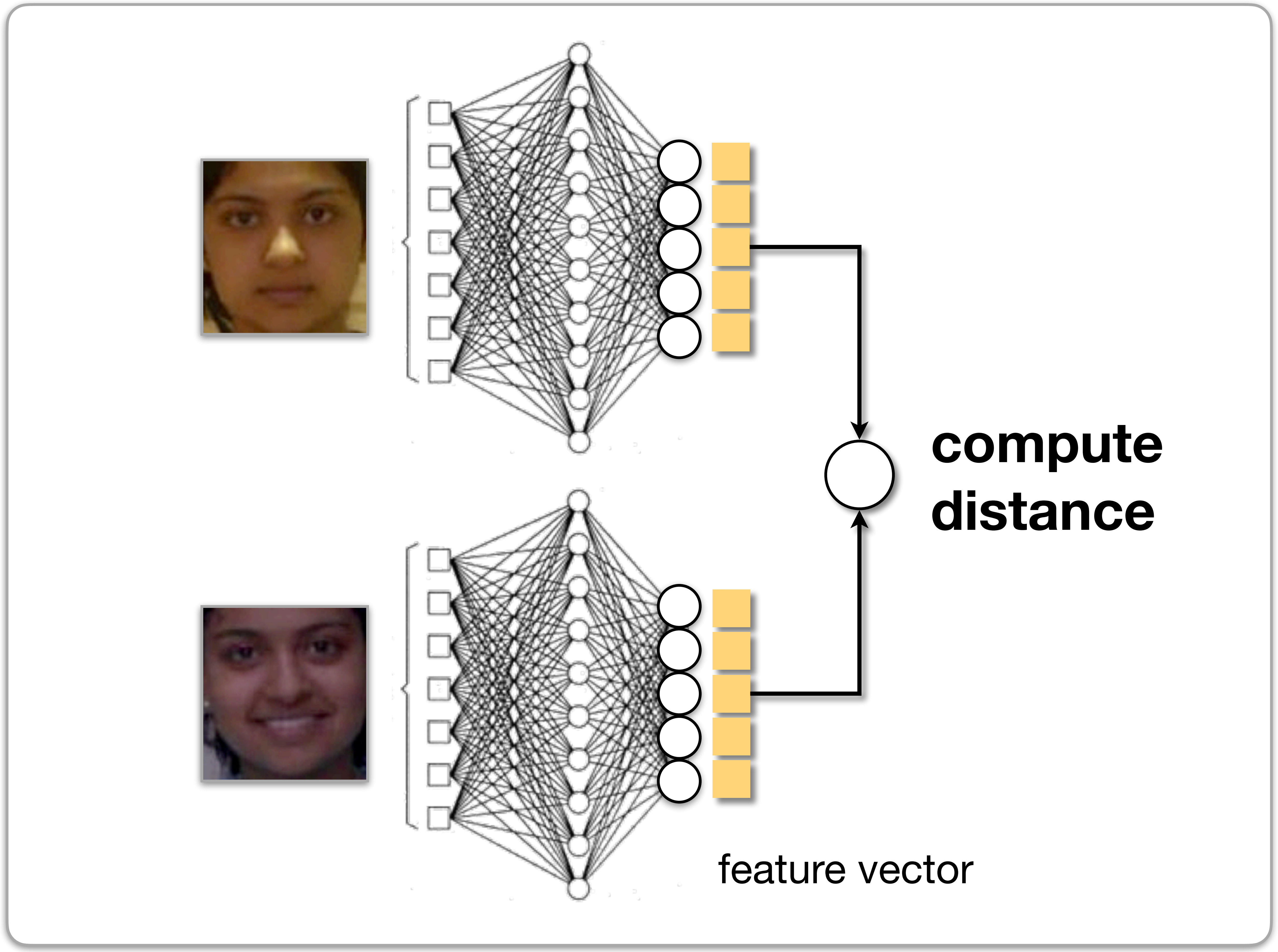


LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

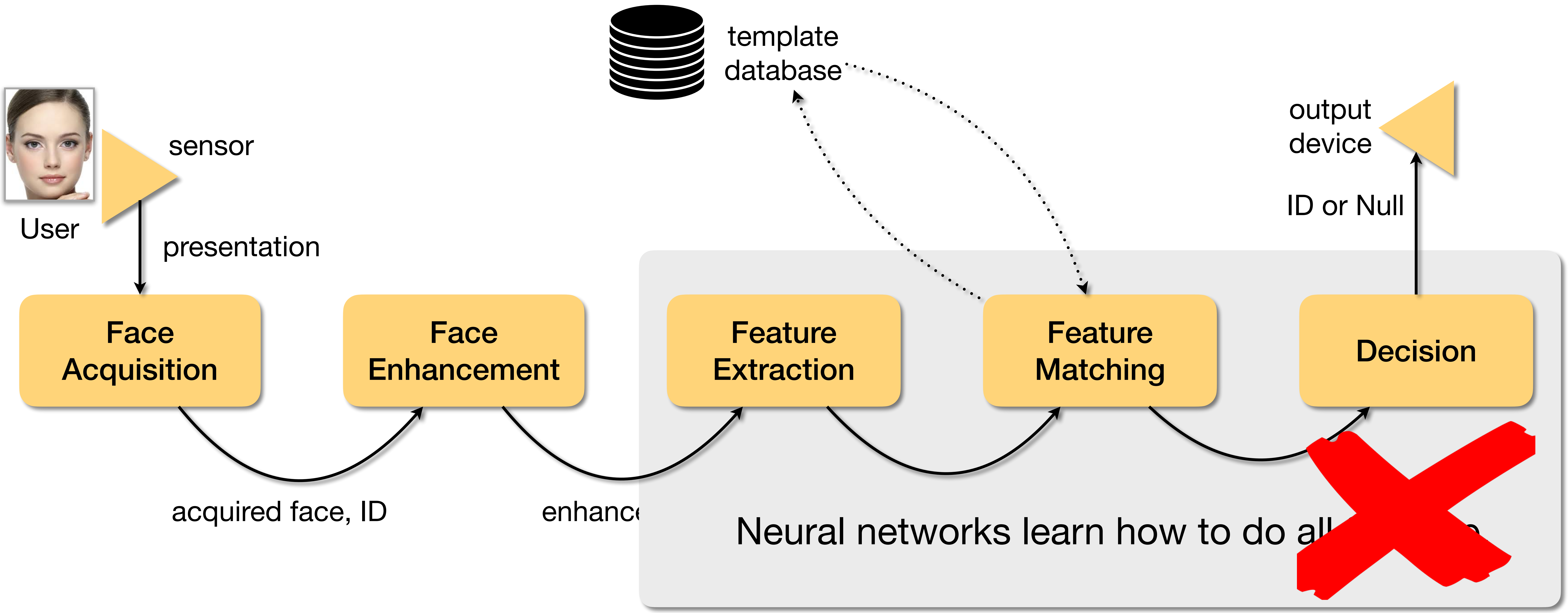
Deep Learning

How to make CNN more flexible?
To speed up training, use **siamese networks** (same architecture, same weights).



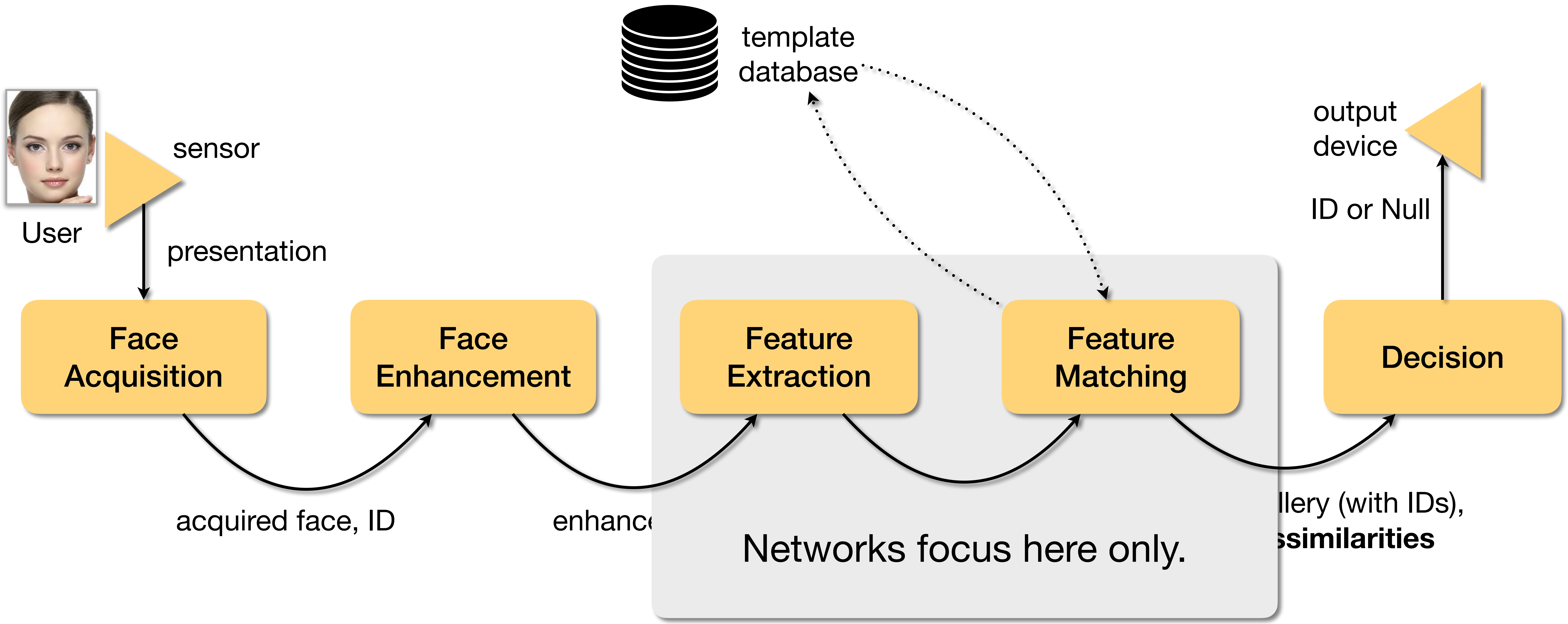
LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition



LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

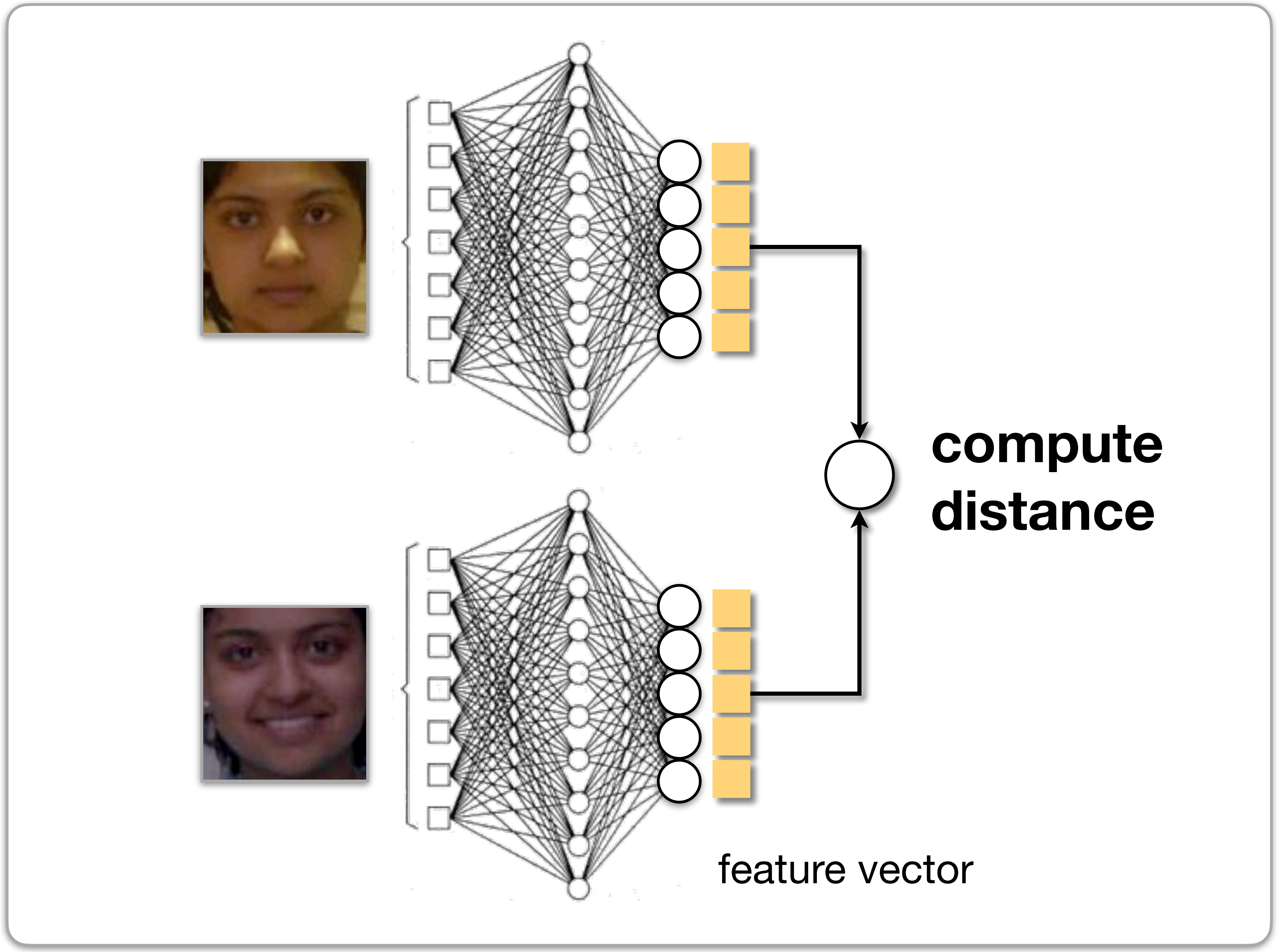


LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

Deep Learning

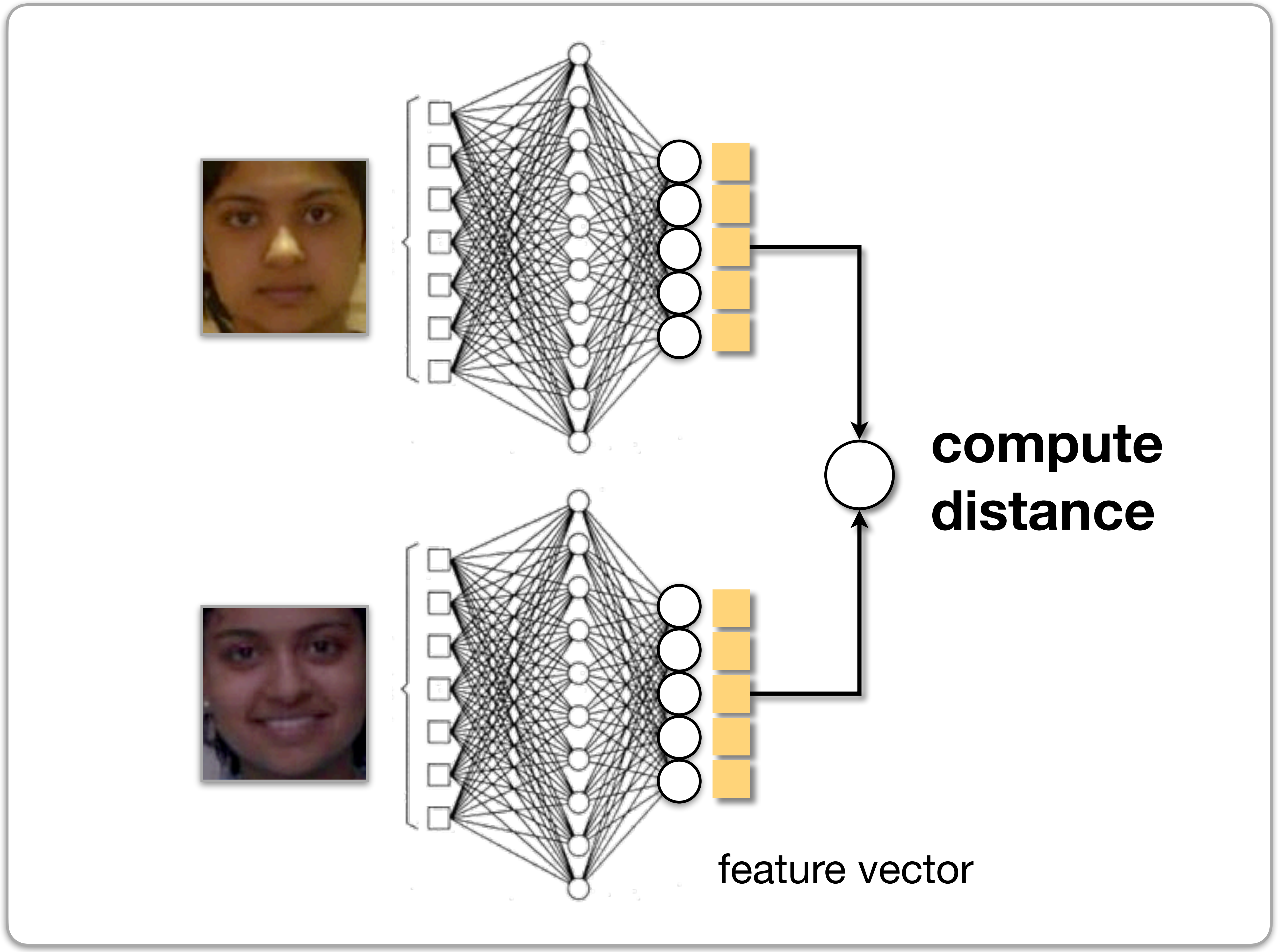
Training Approaches
Pairwise-loss-based
Triplet-loss-based



Data-driven Face Recognition

Deep Learning

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

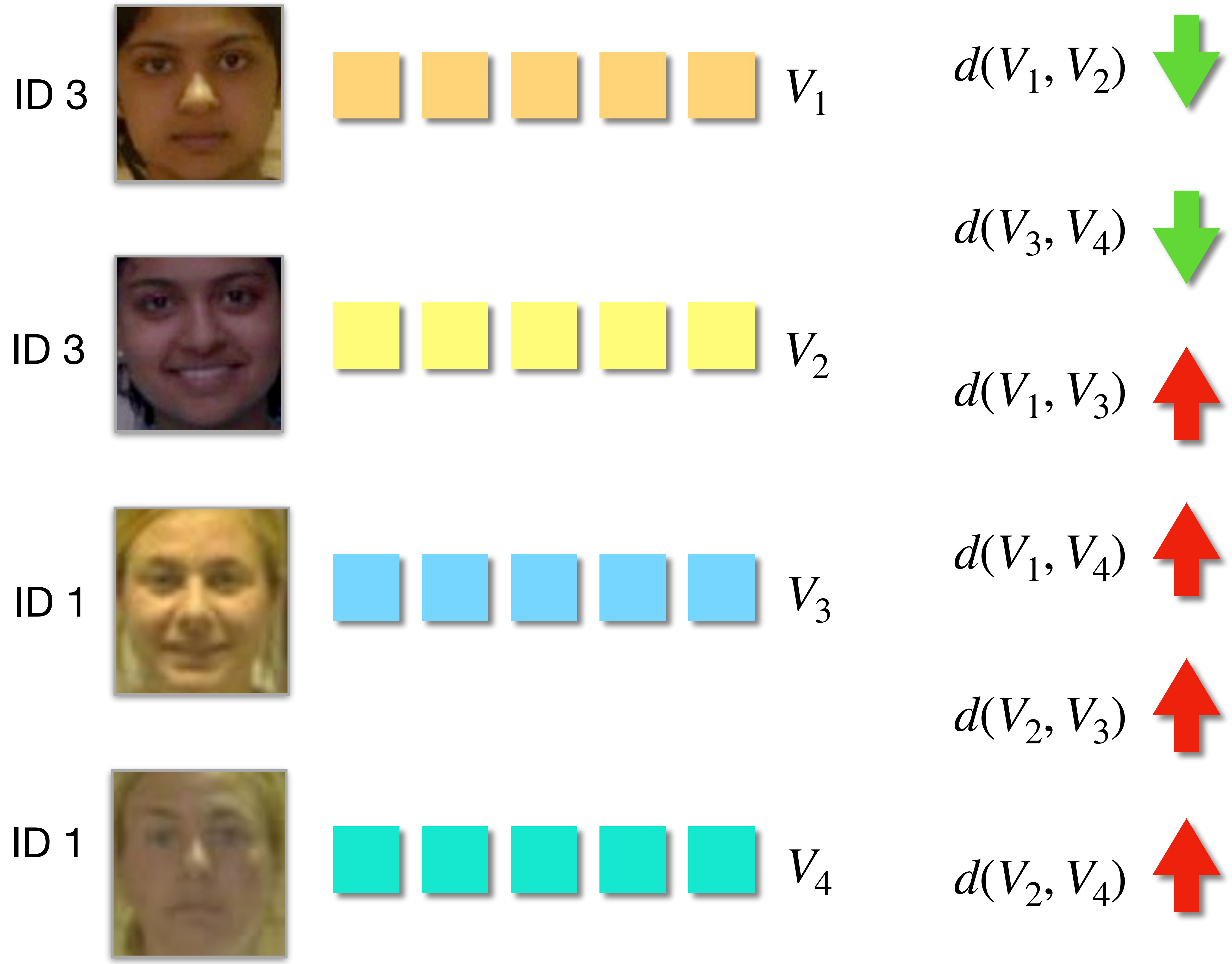


LOYOLA
UNIVERSITY CHICAGO

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.



LOYOLA UNIVERSITY CHICAGO

Pairwise Face Recognition

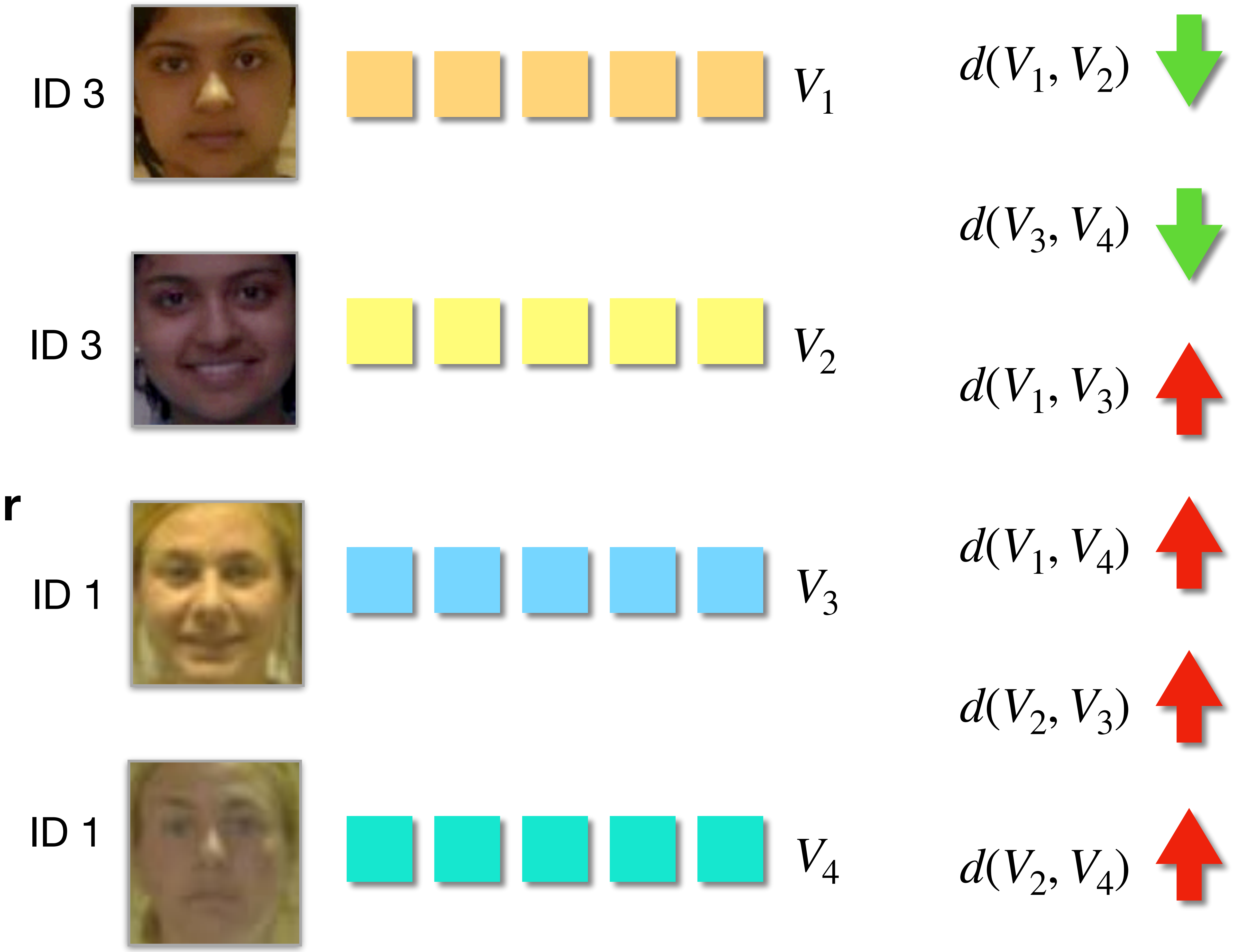
Pairwise Loss (PL)

the smaller, the better

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

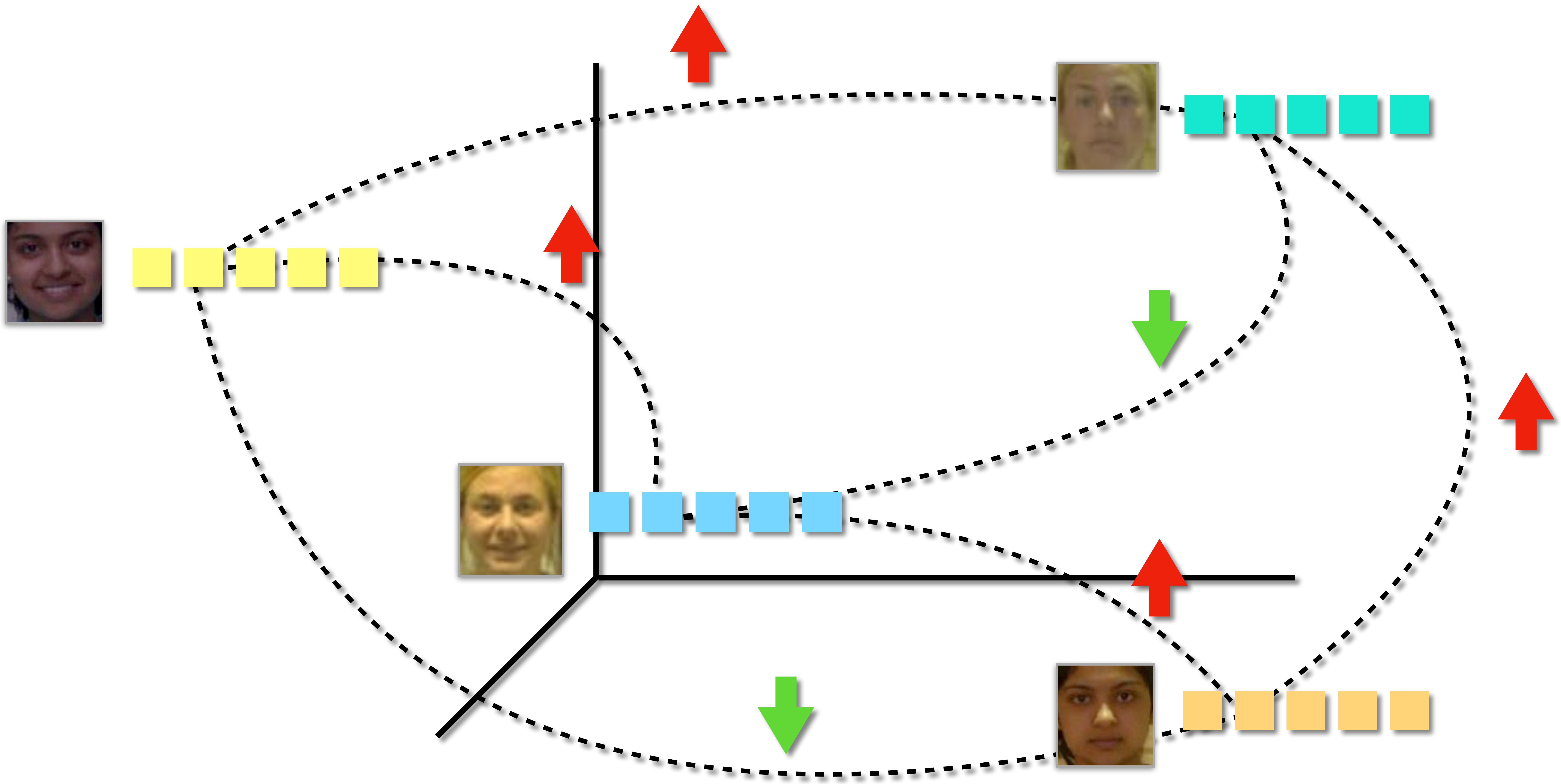
enforced margin

it must be larger than *m*



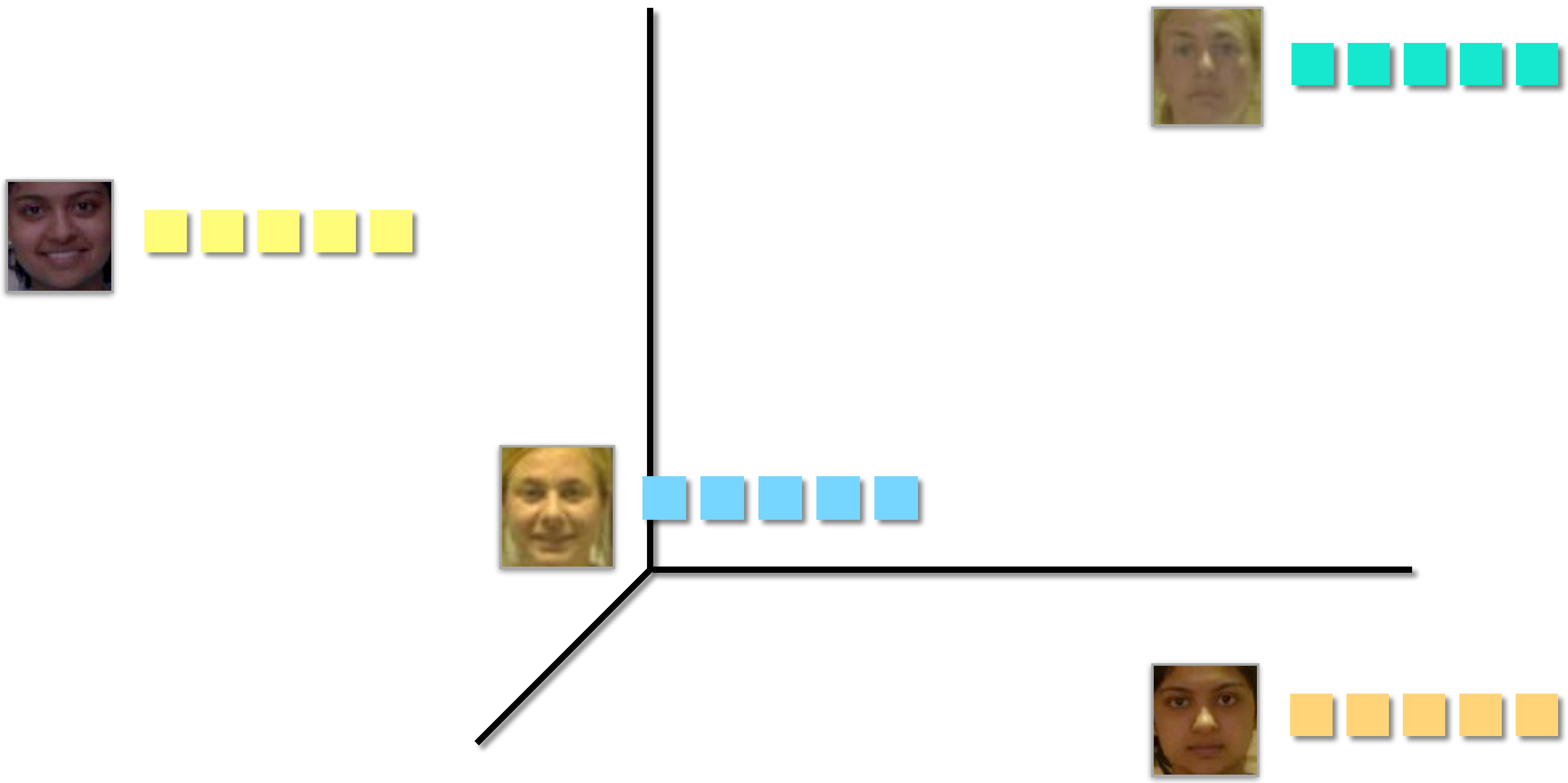
LOYOLA UNIVERSITY CHICAGO

Pairwise Face Recognition

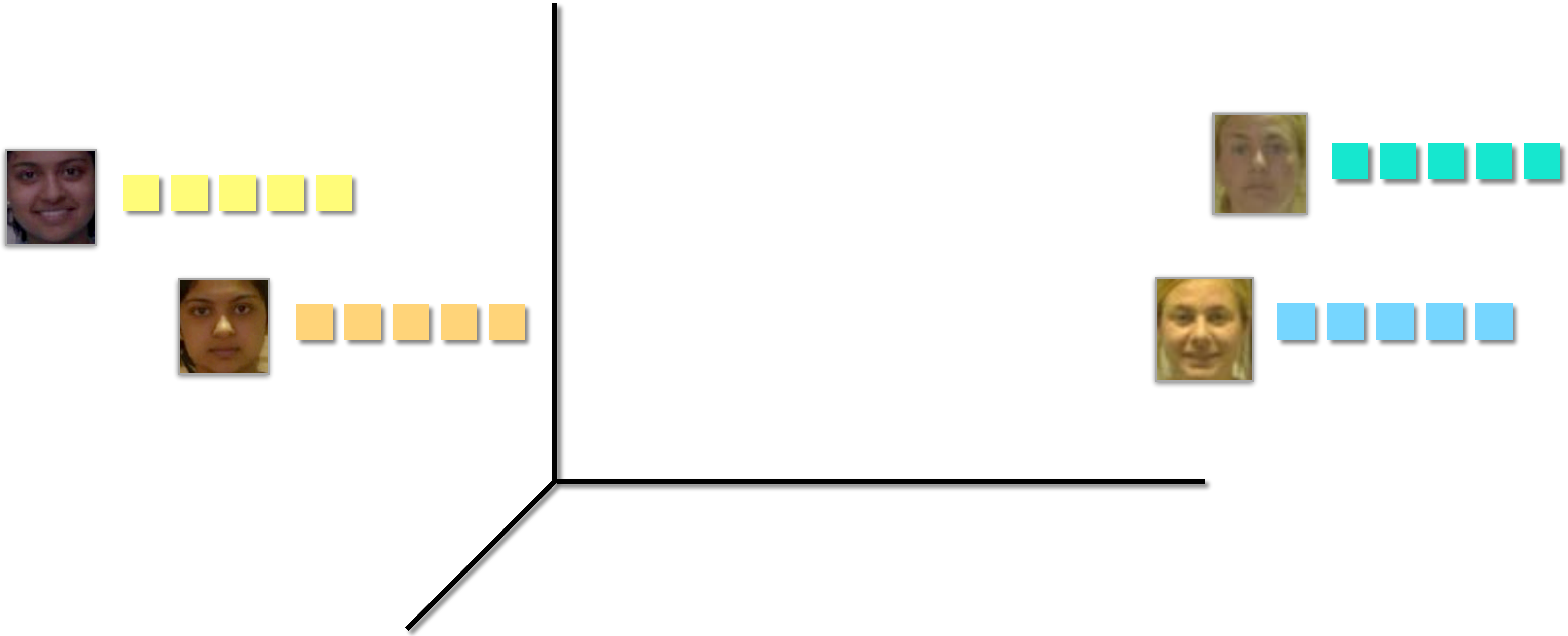


LOYOLA UNIVERSITY CHICAGO

Pairwise Face Recognition



Pairwise Face Recognition



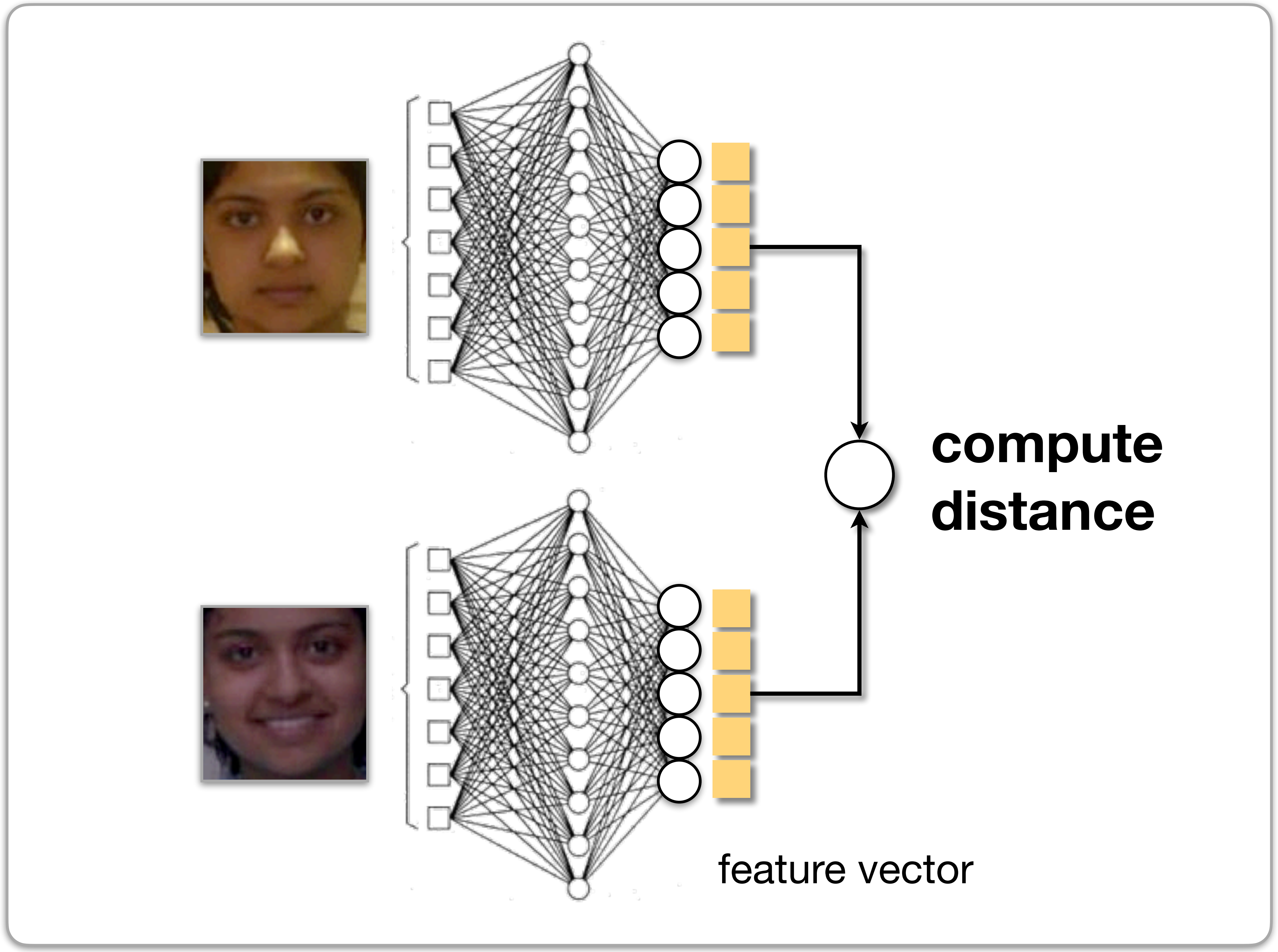
Data-driven Face Recognition

Deep Learning

Training Approaches

Pairwise-loss-based

Triplet-loss-based

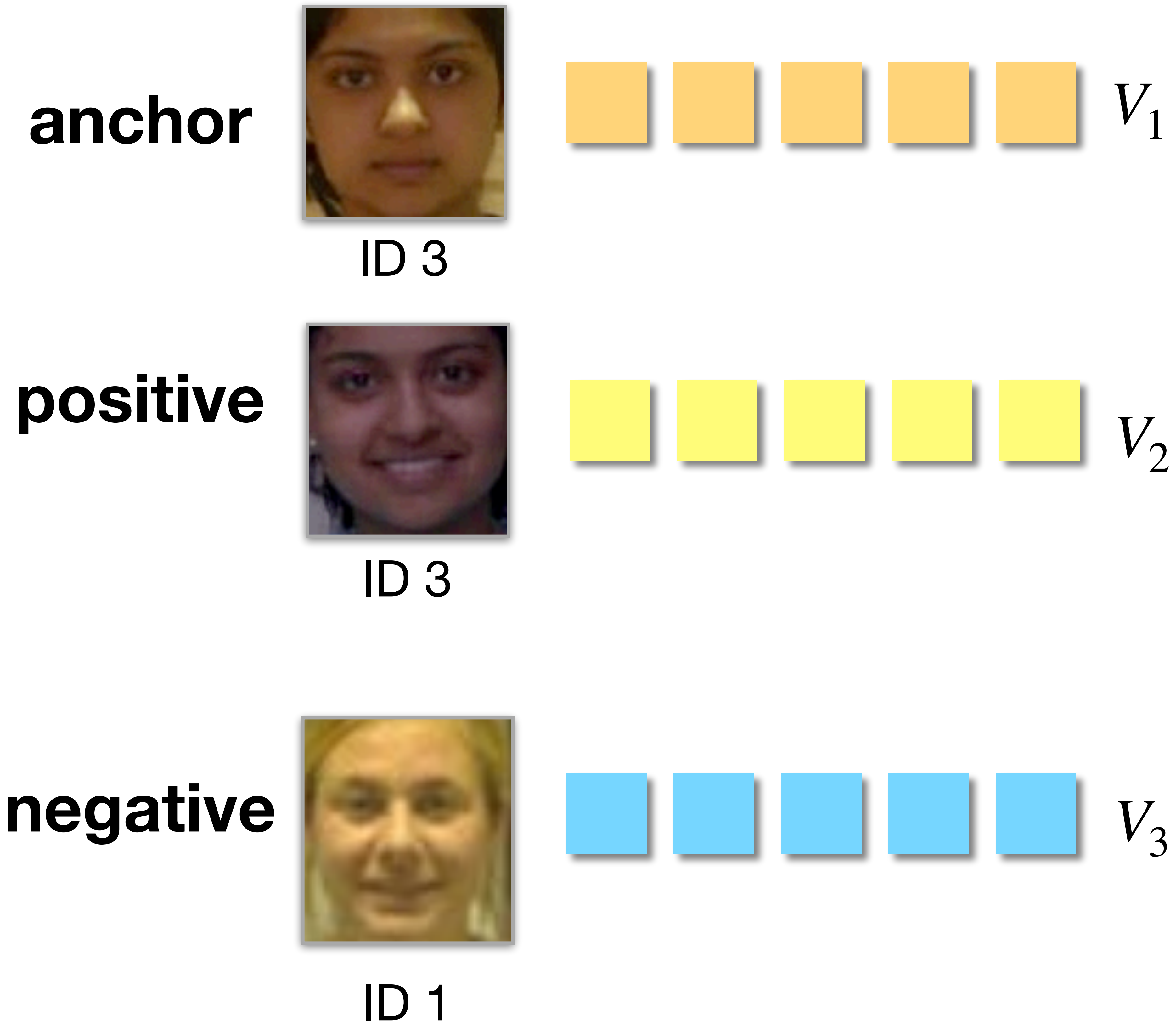


LOYOLA UNIVERSITY CHICAGO

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.



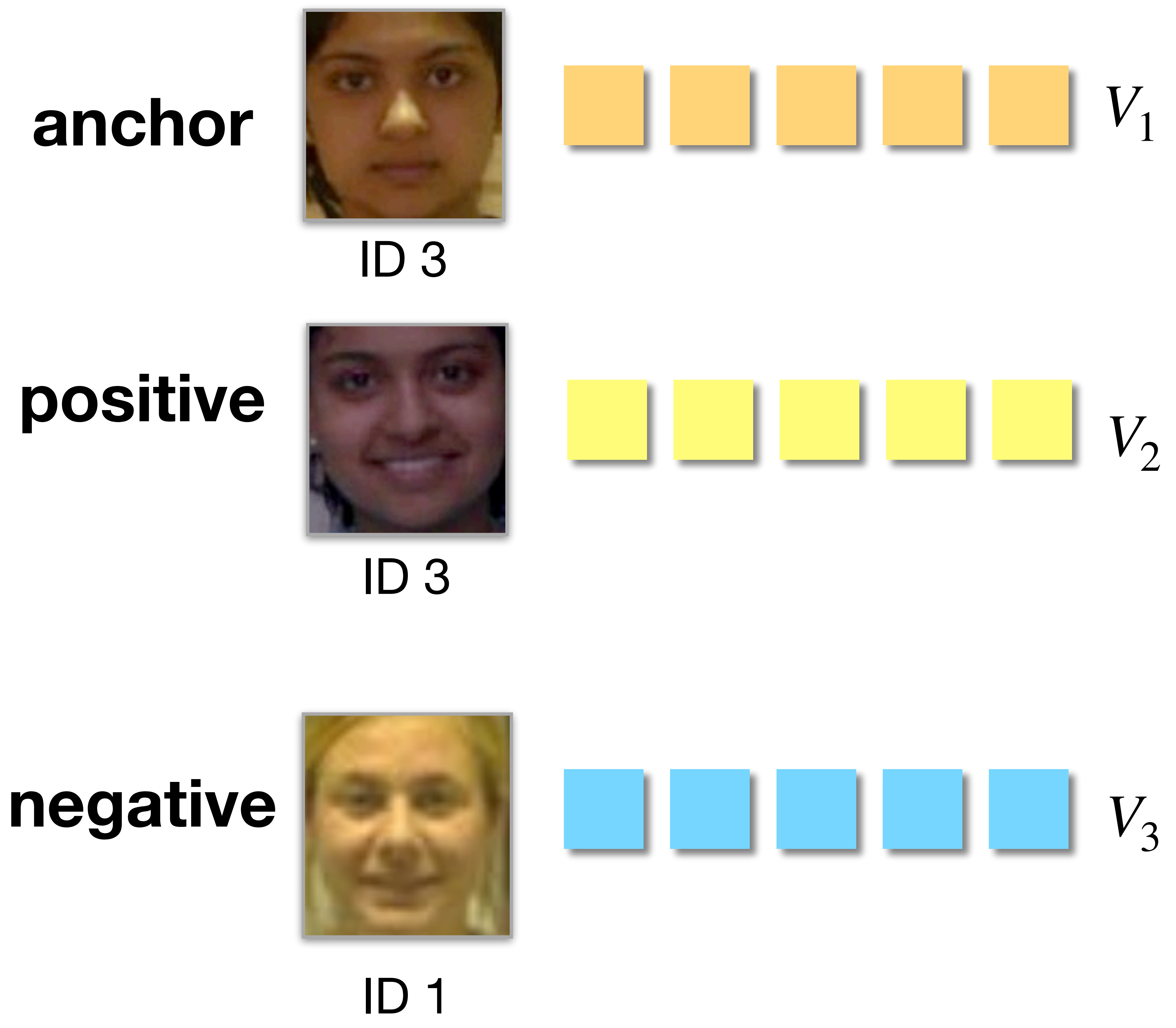
LOYOLA UNIVERSITY CHICAGO

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

LOYOLA UNIVERSITY CHICAGO

Triplet Face Recognition

Triplet Loss (TL)

the smaller, the better

$$TL = max(0, m + d(V_{\text{anchor}}, V_{\text{positive}}) - d(V_{\text{anchor}}, V_{\text{negative}}))$$

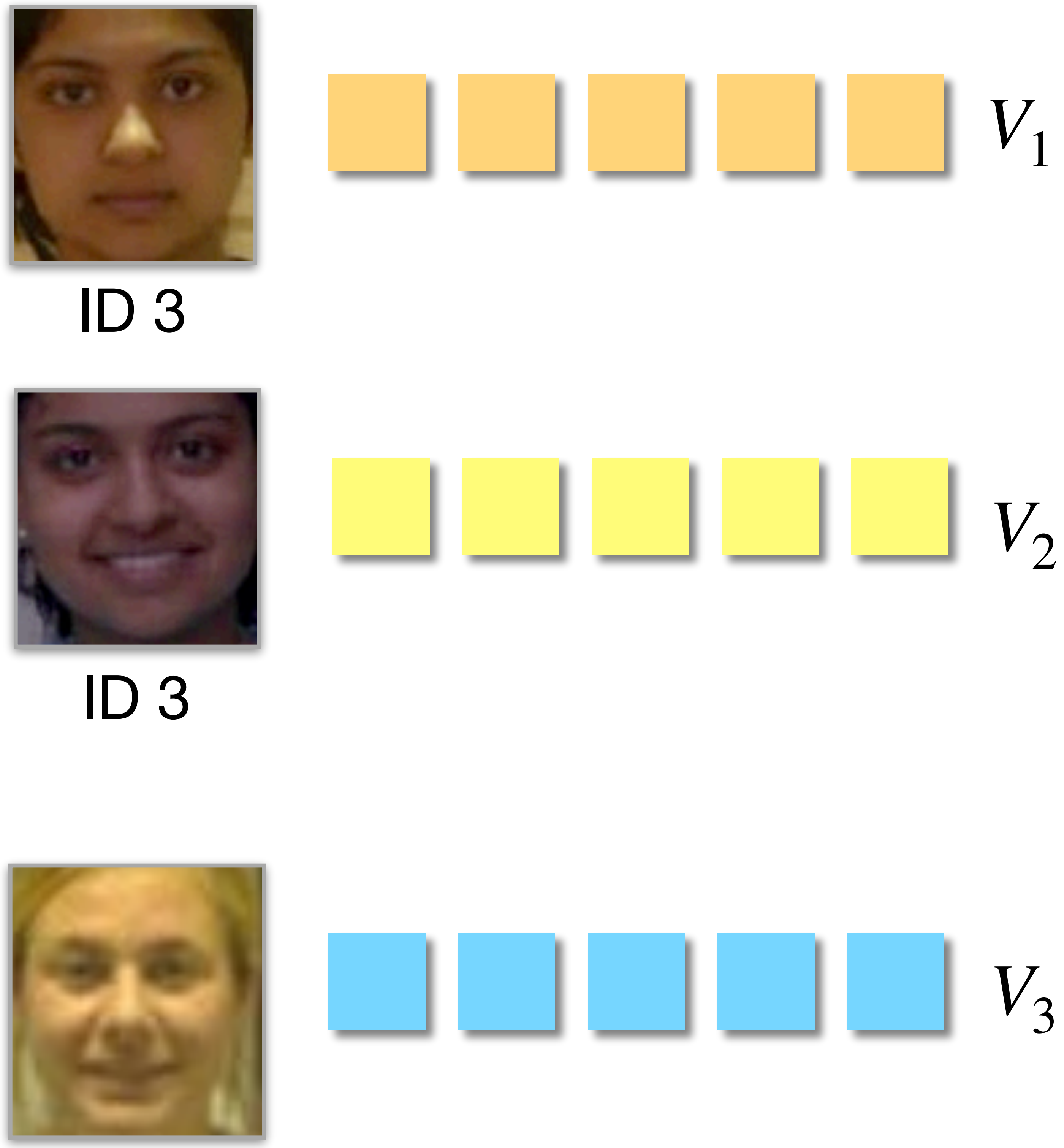
enforced margin

the larger, the better

anchor

positive

negative



ID 3

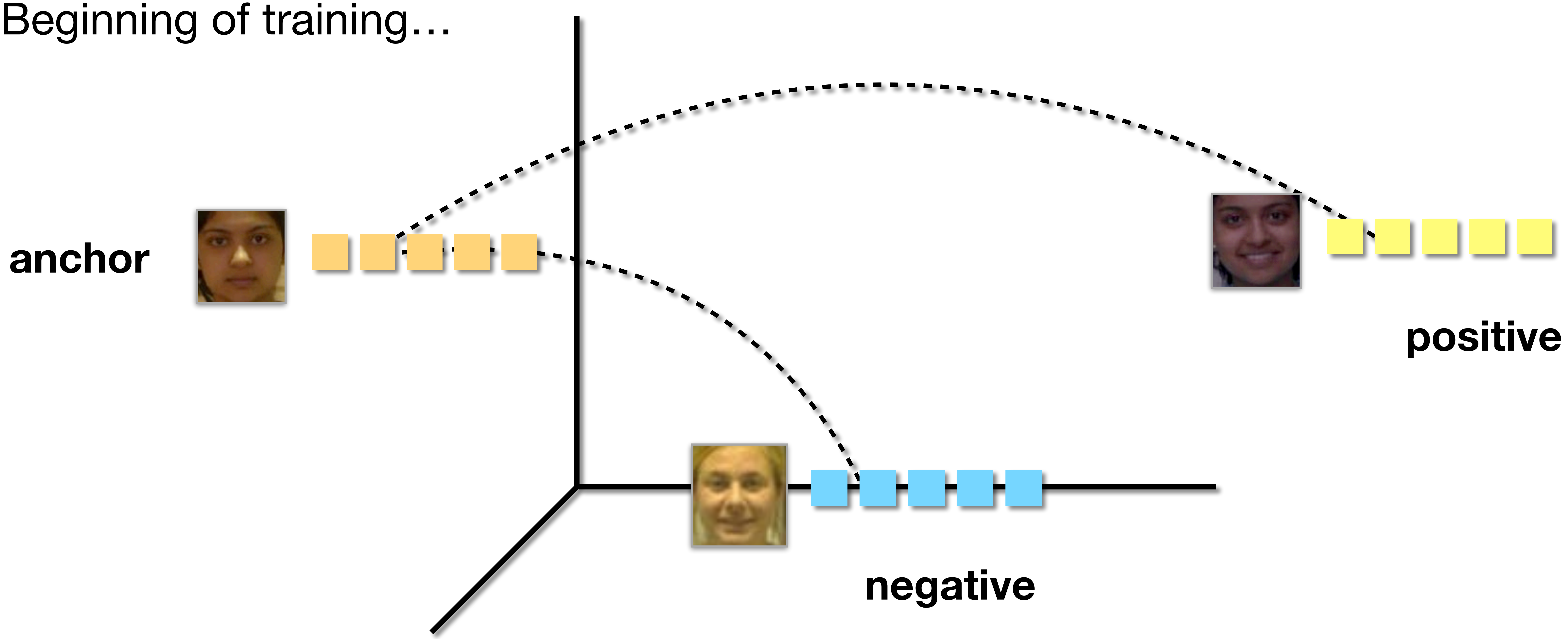
ID 3

ID 1

LOYOLA UNIVERSITY CHICAGO

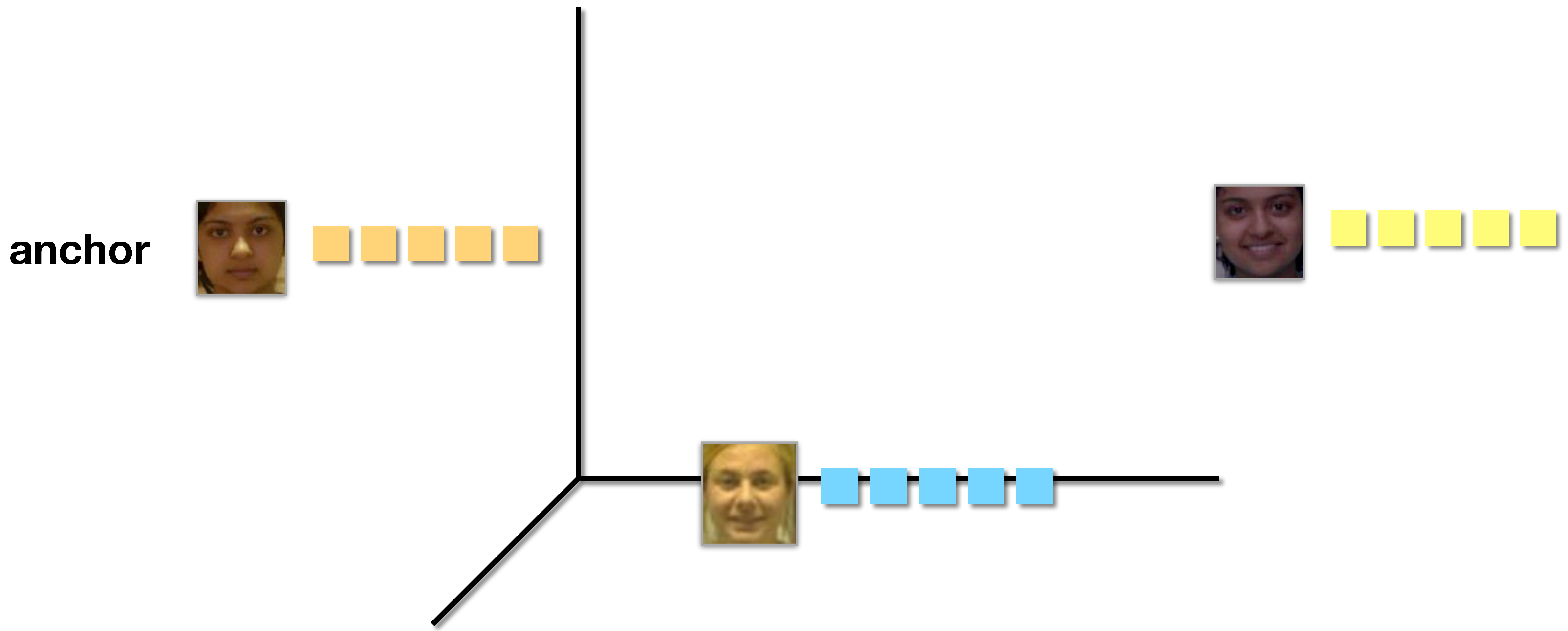
Triplet Face Recognition

Beginning of training…



LOYOLA UNIVERSITY CHICAGO

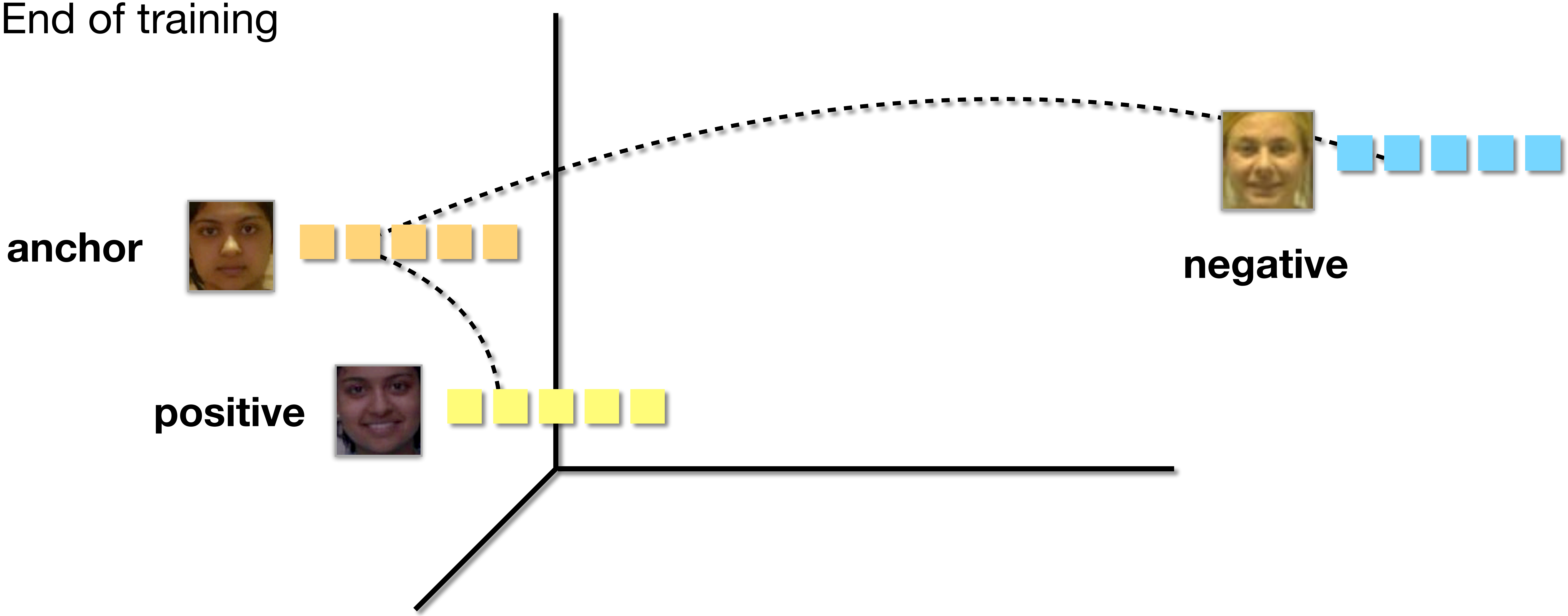
Triplet Face Recognition



LOYOLA UNIVERSITY CHICAGO

Triplet Face Recognition

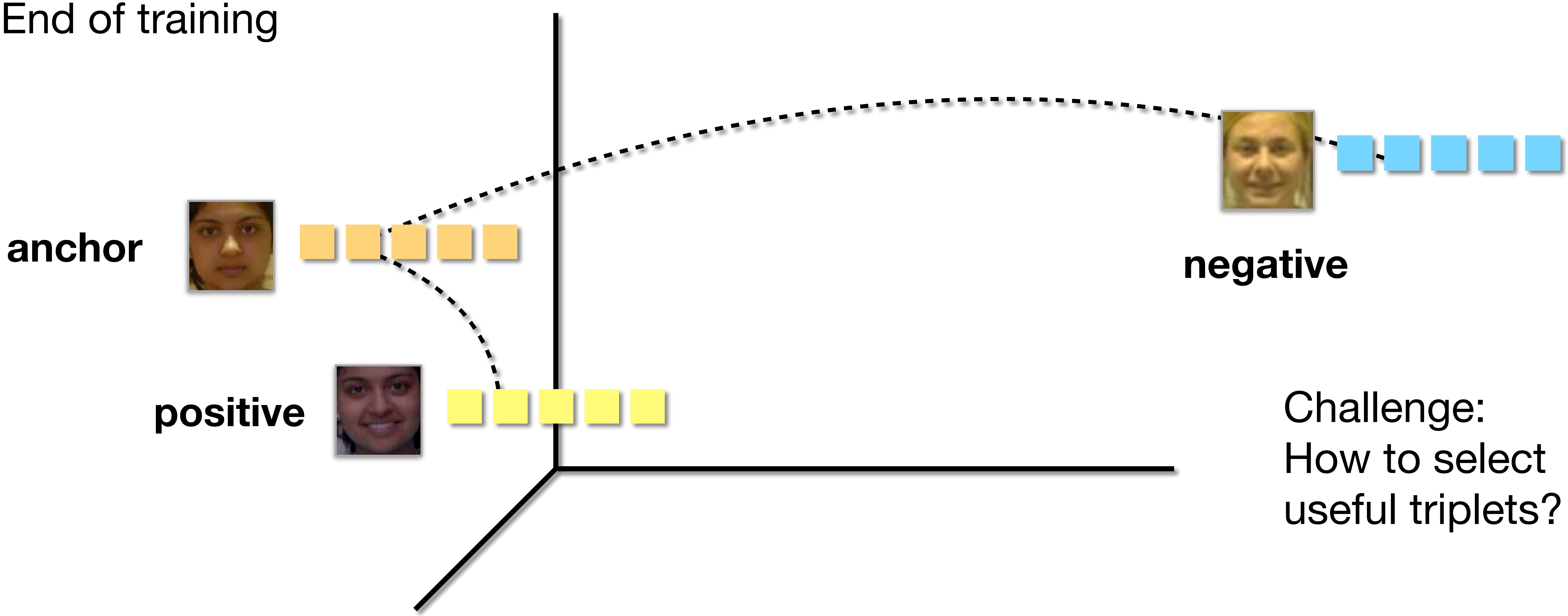
End of training



LOYOLA UNIVERSITY CHICAGO

Triplet Face Recognition

End of training

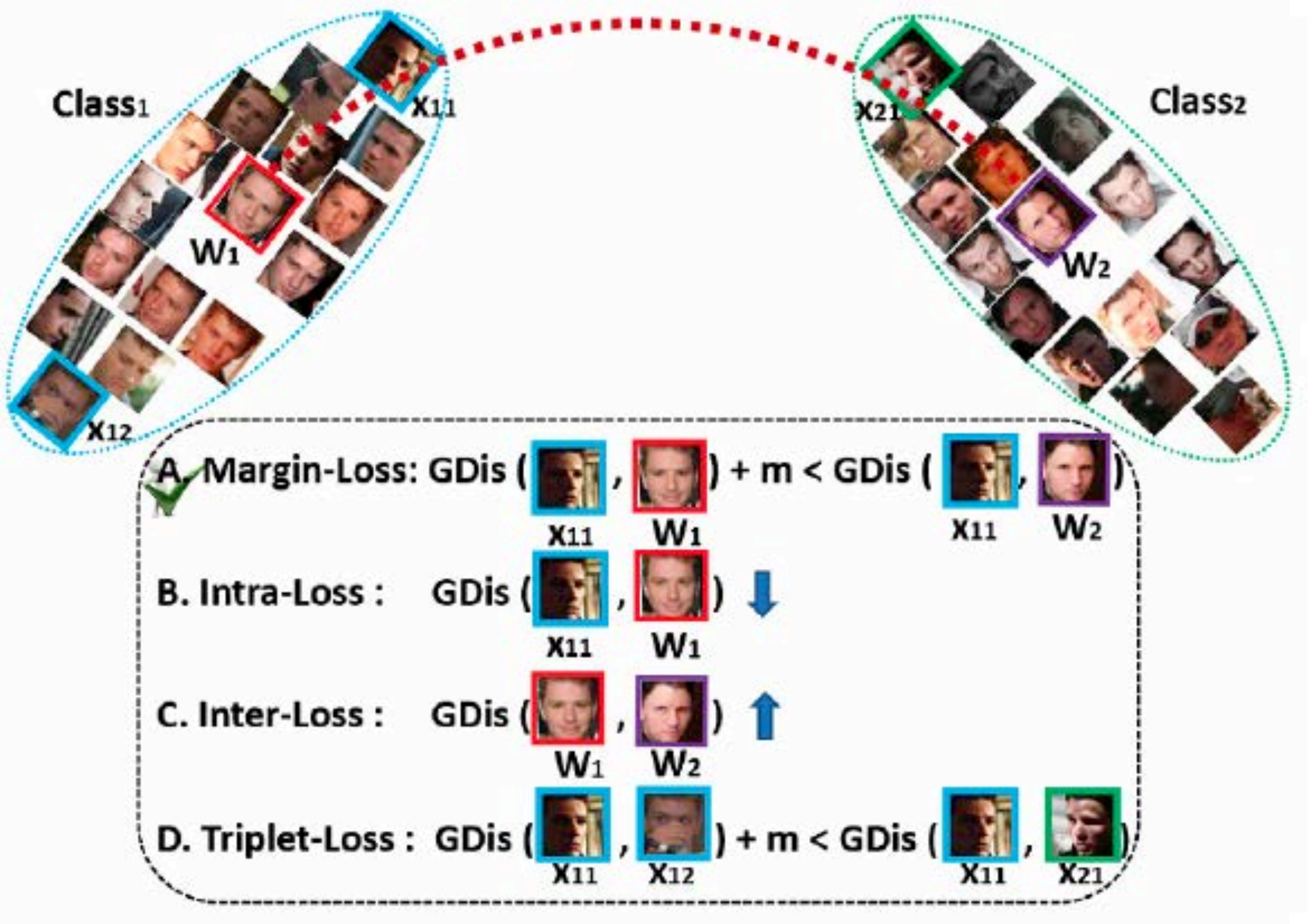


Challenge:
How to select
useful triplets?

Improvements

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

Centre Loss



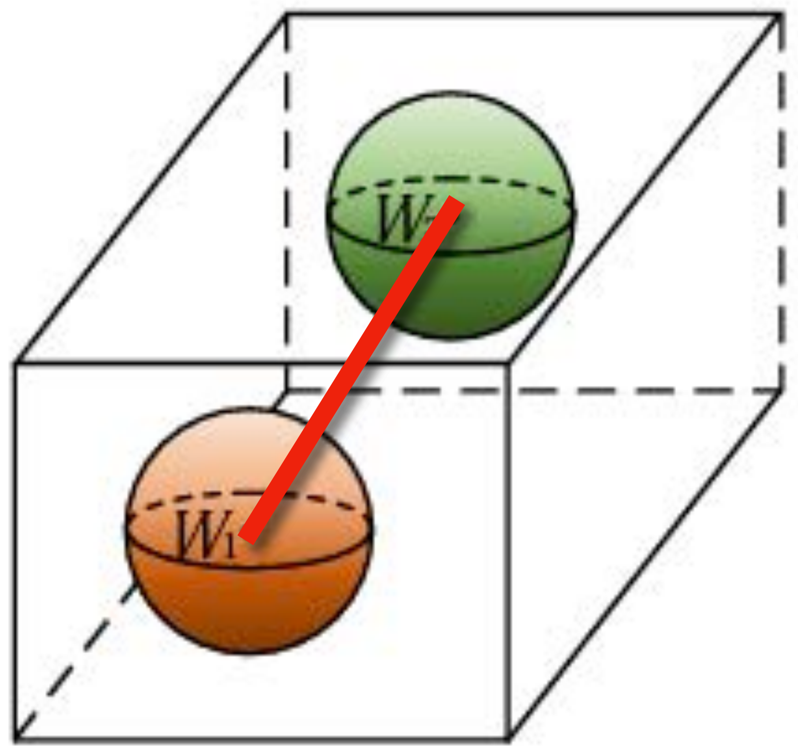
Use class clusters' centers to improve the convergence of the learning process.

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.



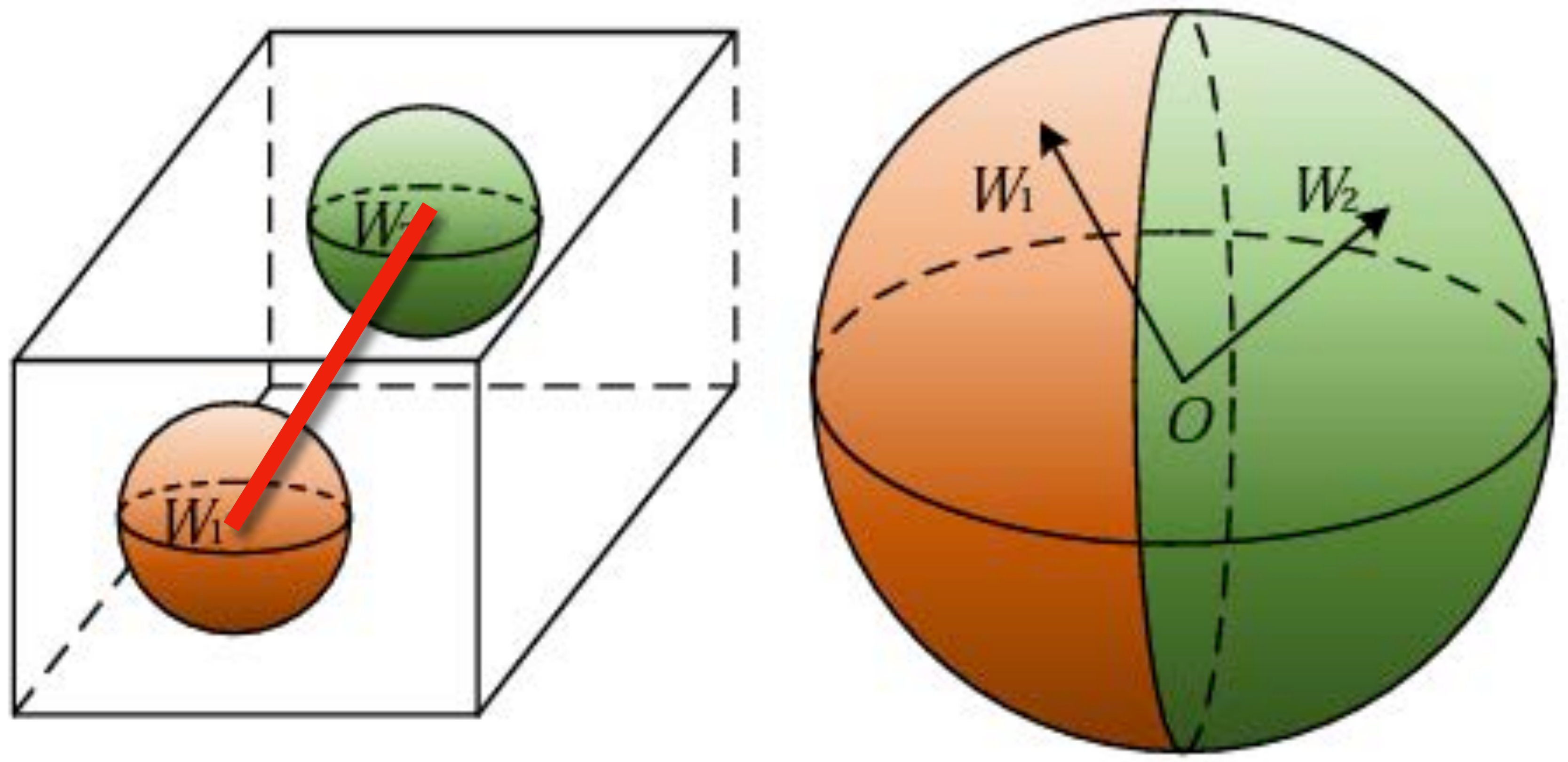
Euclidean space

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.



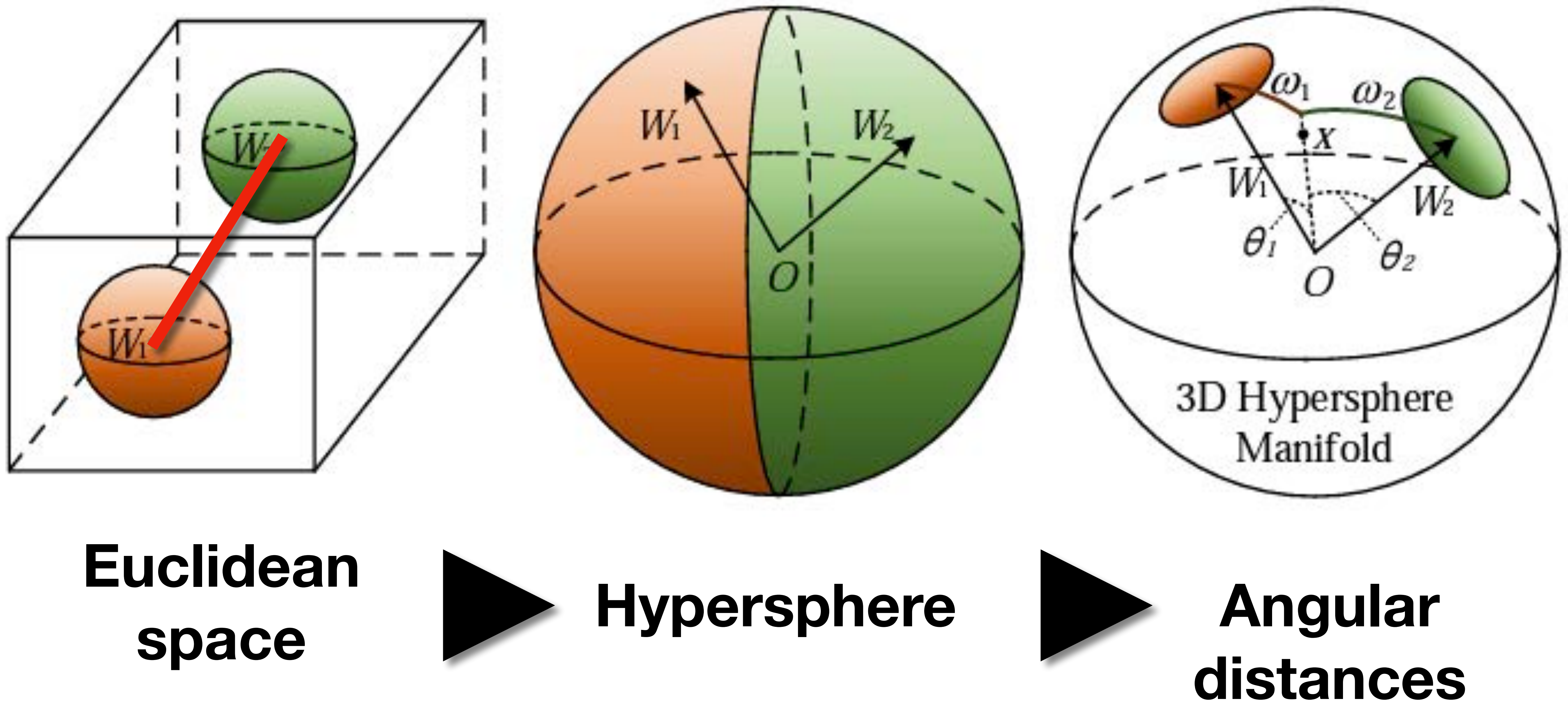
Euclidean space ▶ **Hypersphere**

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

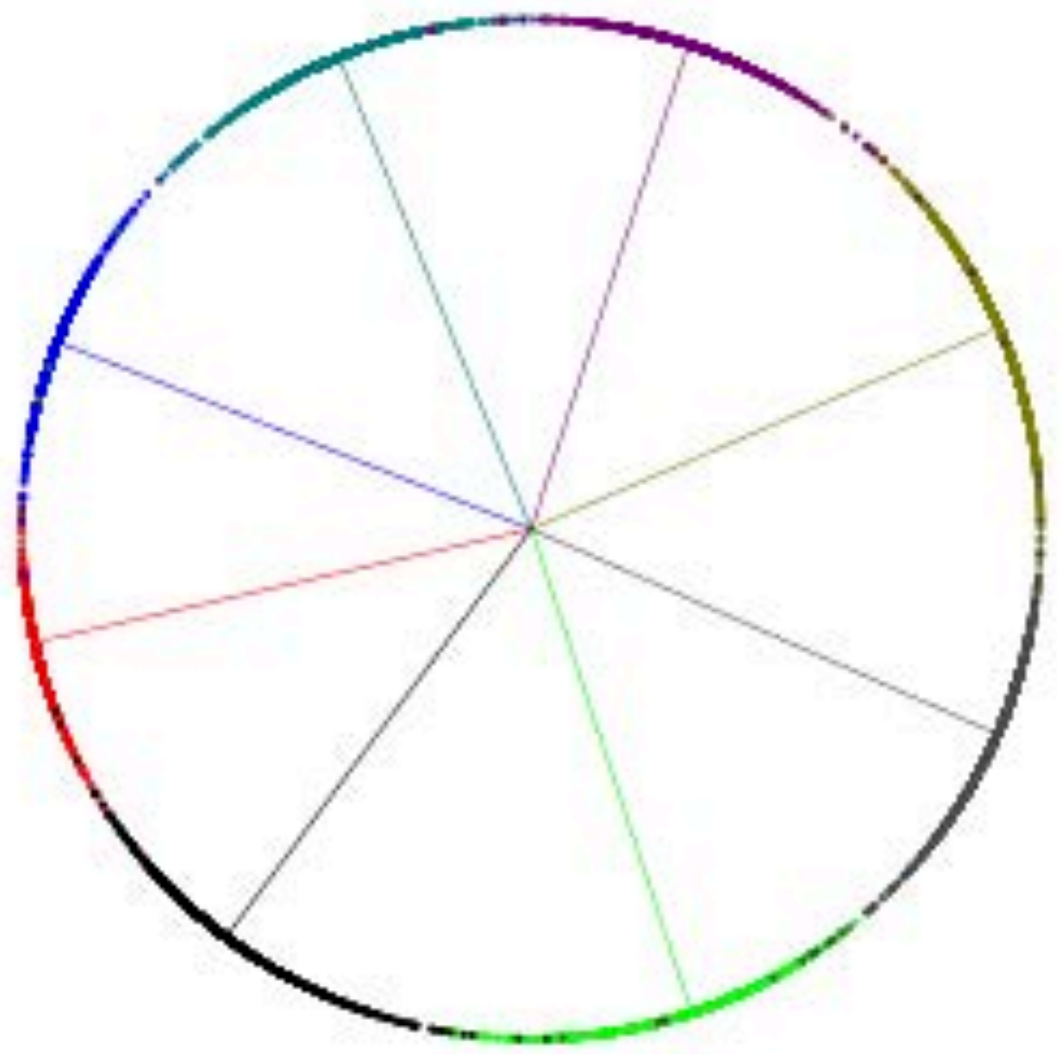
LOYOLA UNIVERSITY CHICAGO

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



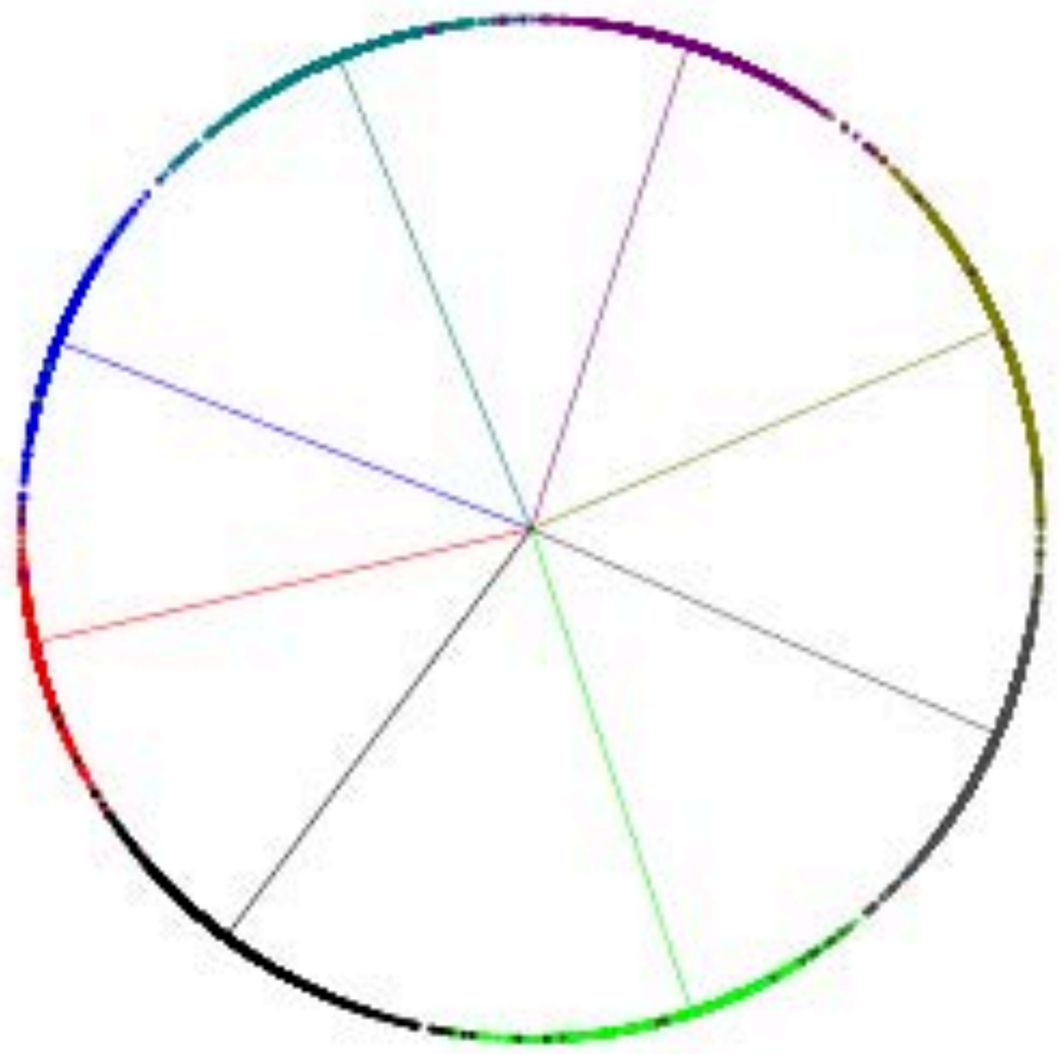
Margin-less class separation

Improvements

ArcFace
Current state of the art.

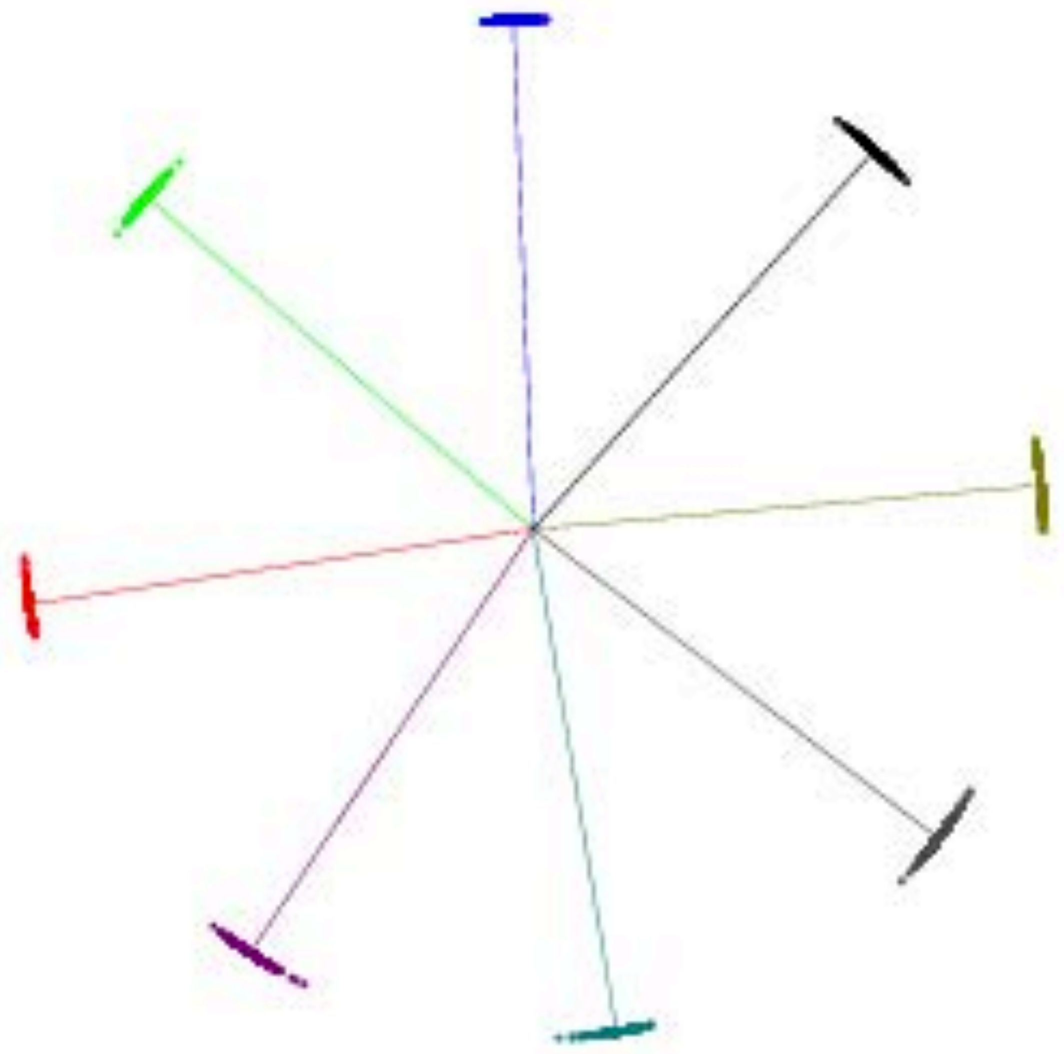
Deng et al. proposed the **additive angular margin loss** to the problem of face recognition.

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



Margin-less class separation

VS



Additive angular margin loss

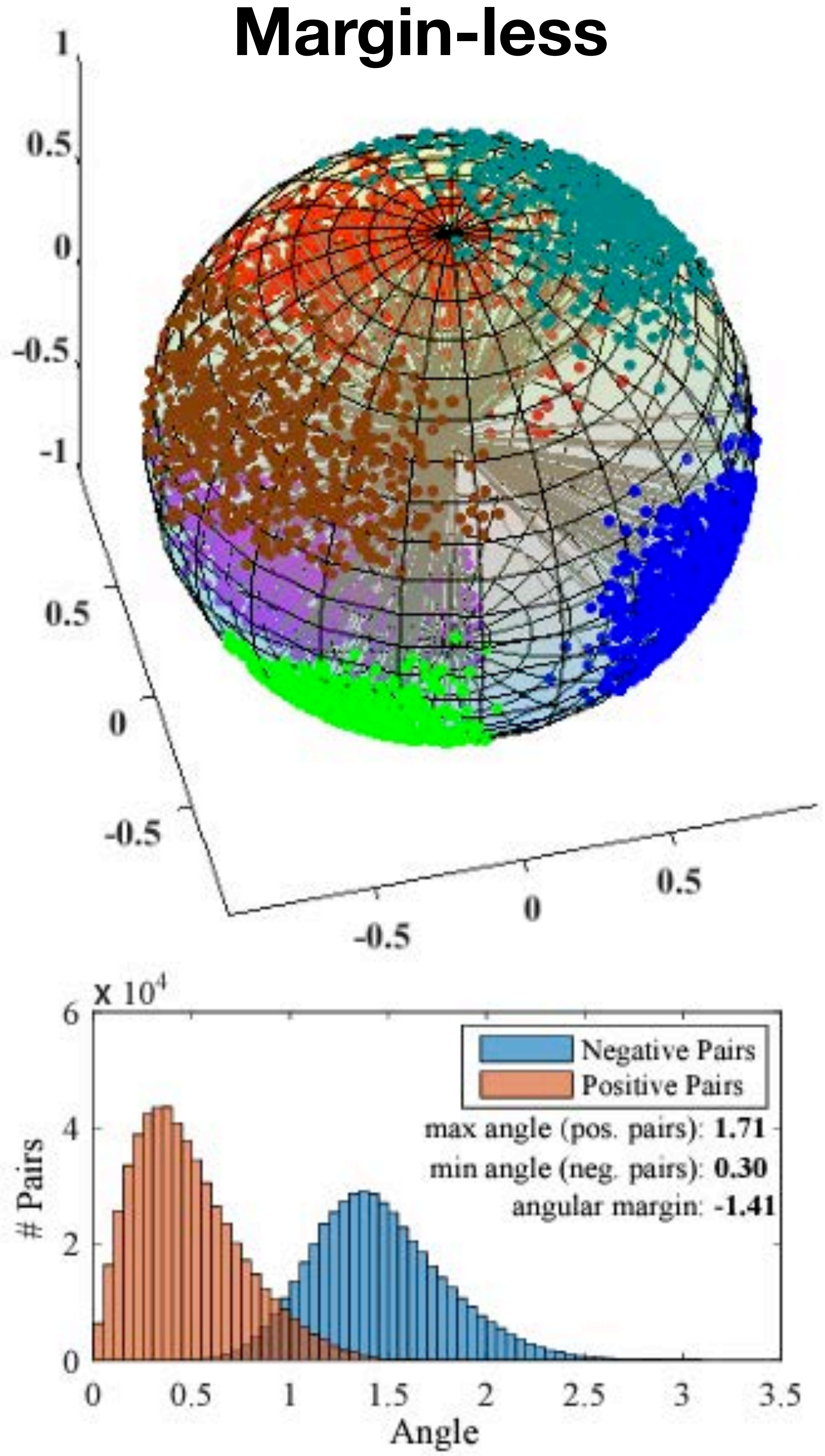
LOYOLA UNIVERSITY CHICAGO

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



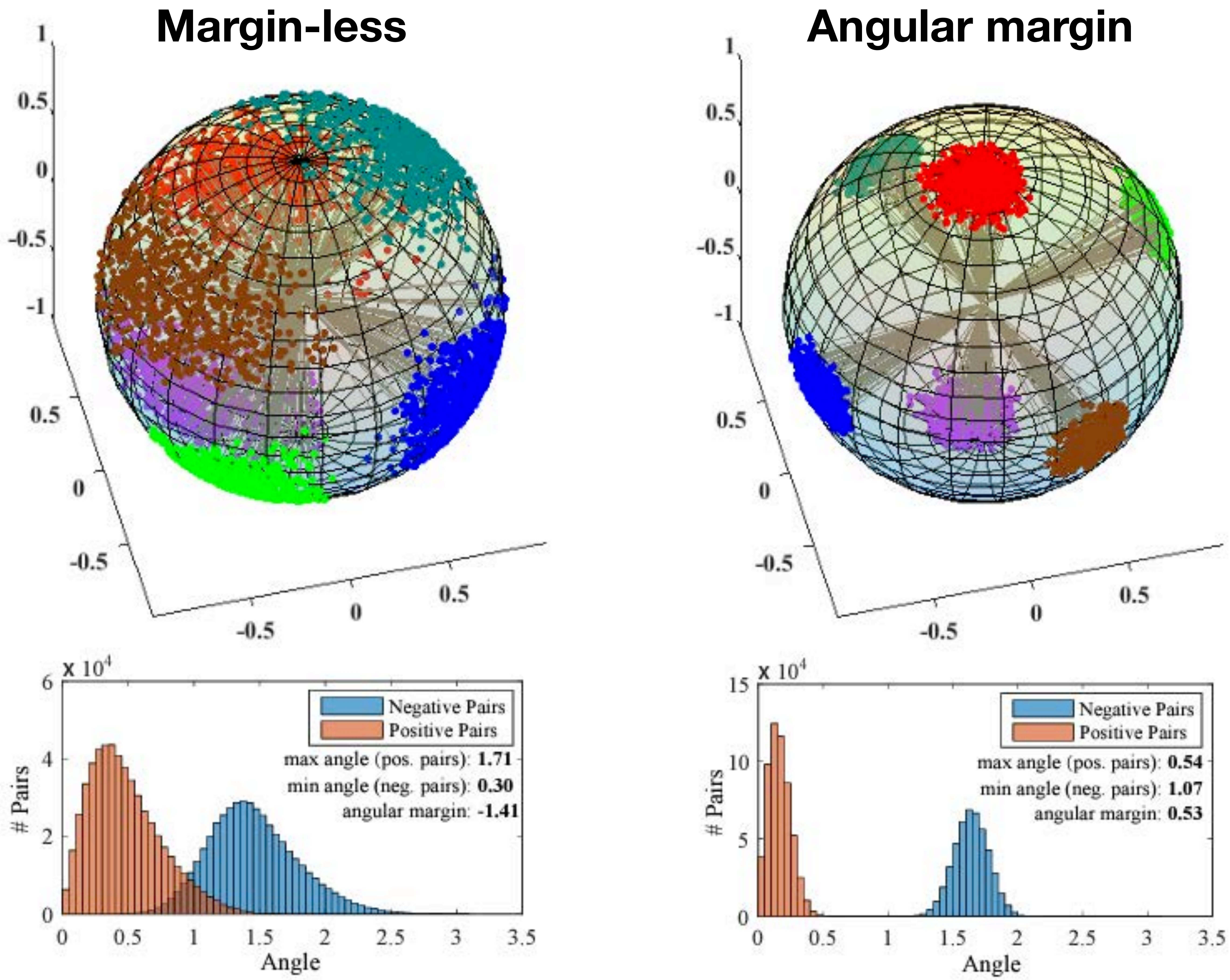
Improvements

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



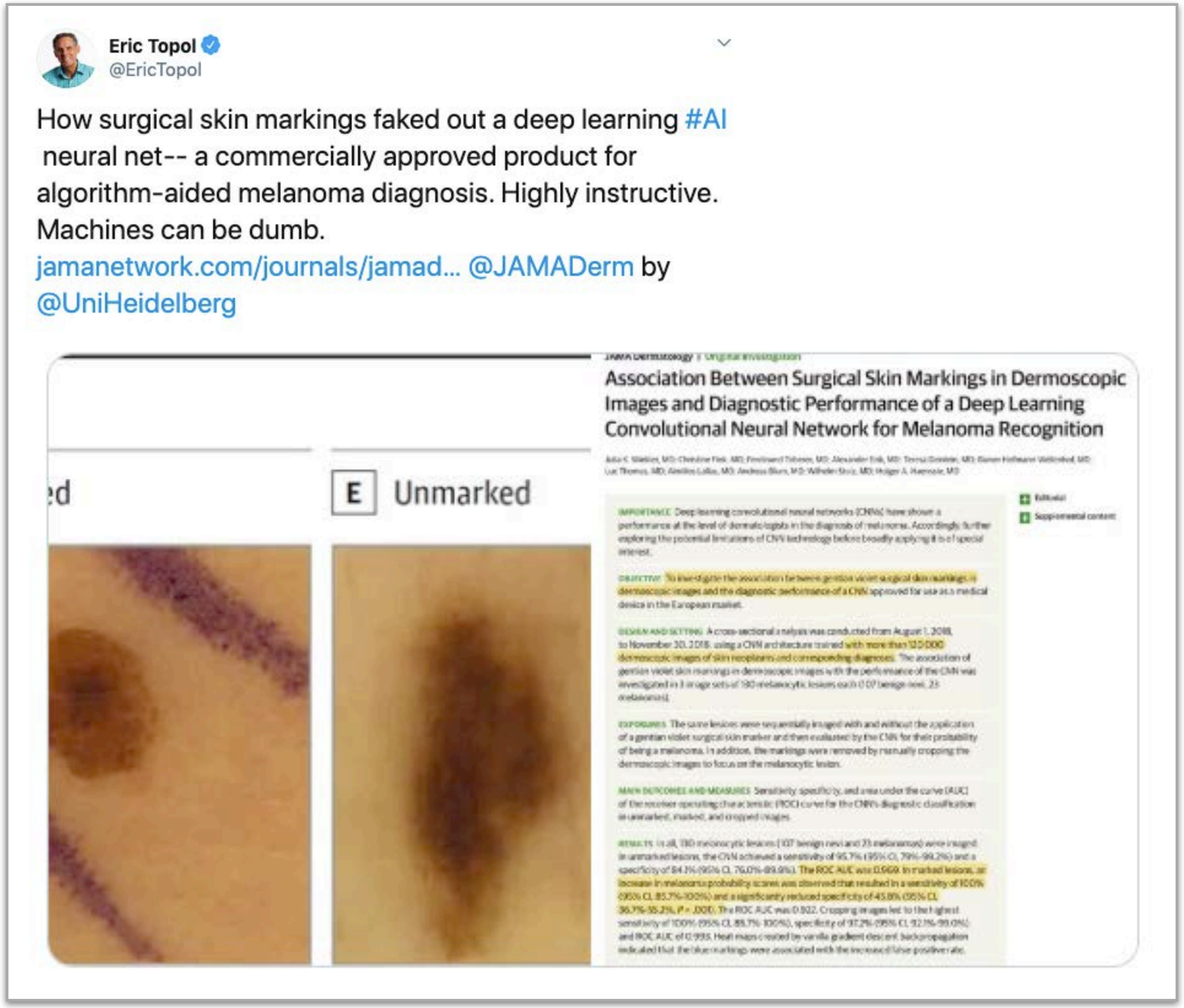
Data-driven Face Recognition

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



LOYOLA UNIVERSITY CHICAGO

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 crim-

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

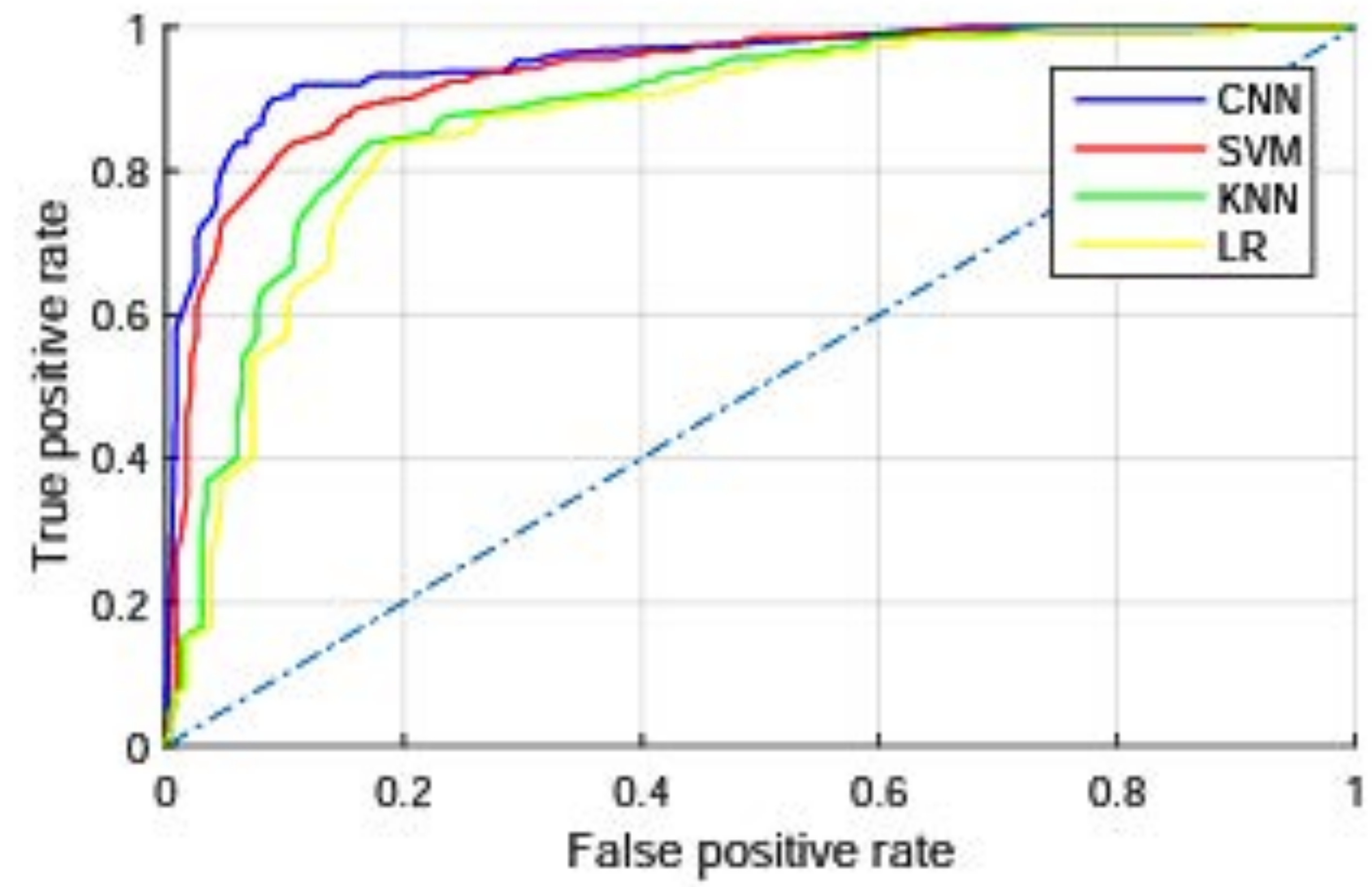


Figure 3. The ROC curves of the four tested binary face classifiers on criminality.

Classifiers	CNN	SVM	KNN	LR
AUC	0.9540	0.9303	0.8838	0.8666

Table 1. The AUC results for the four tested face classifiers on criminality.

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.

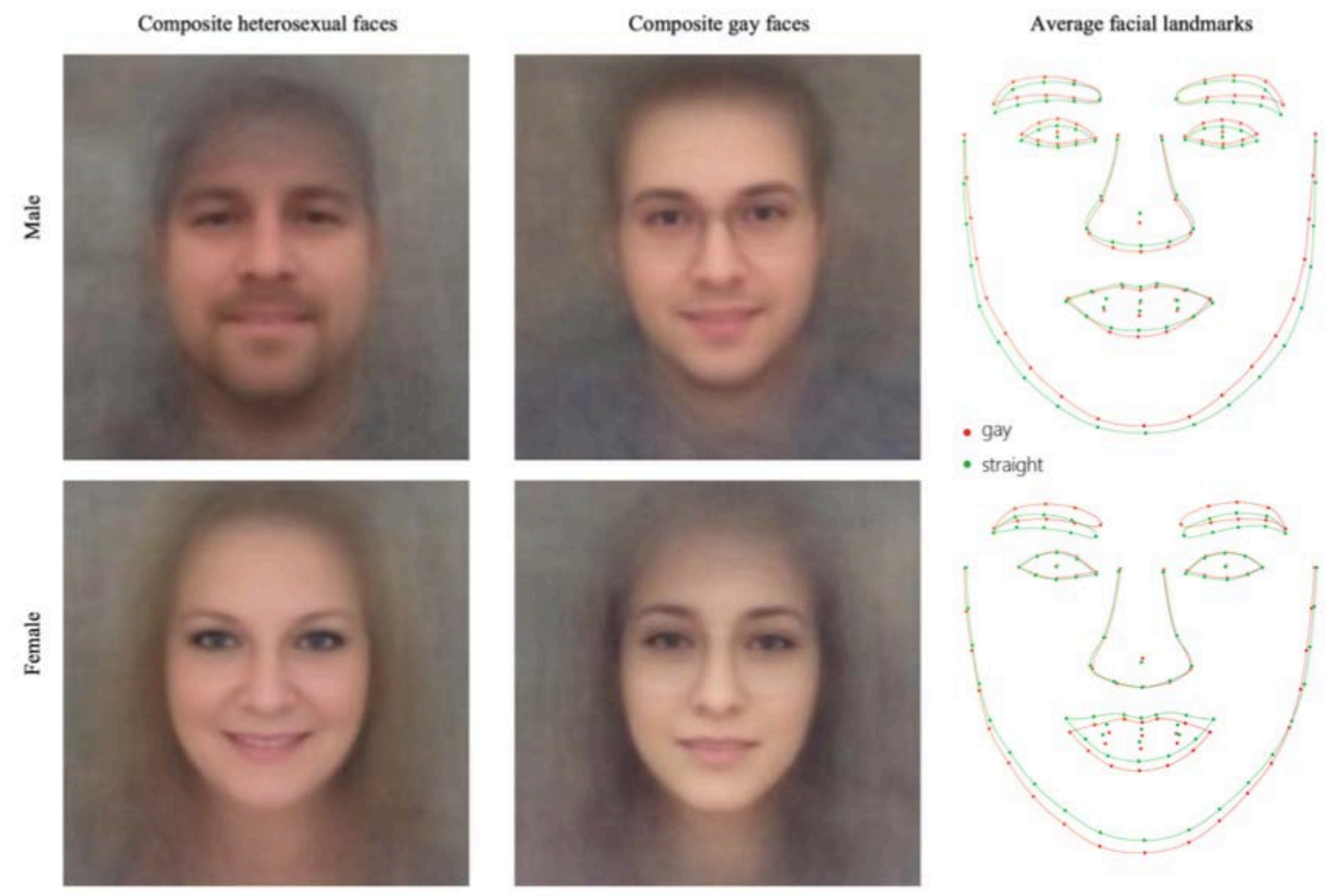


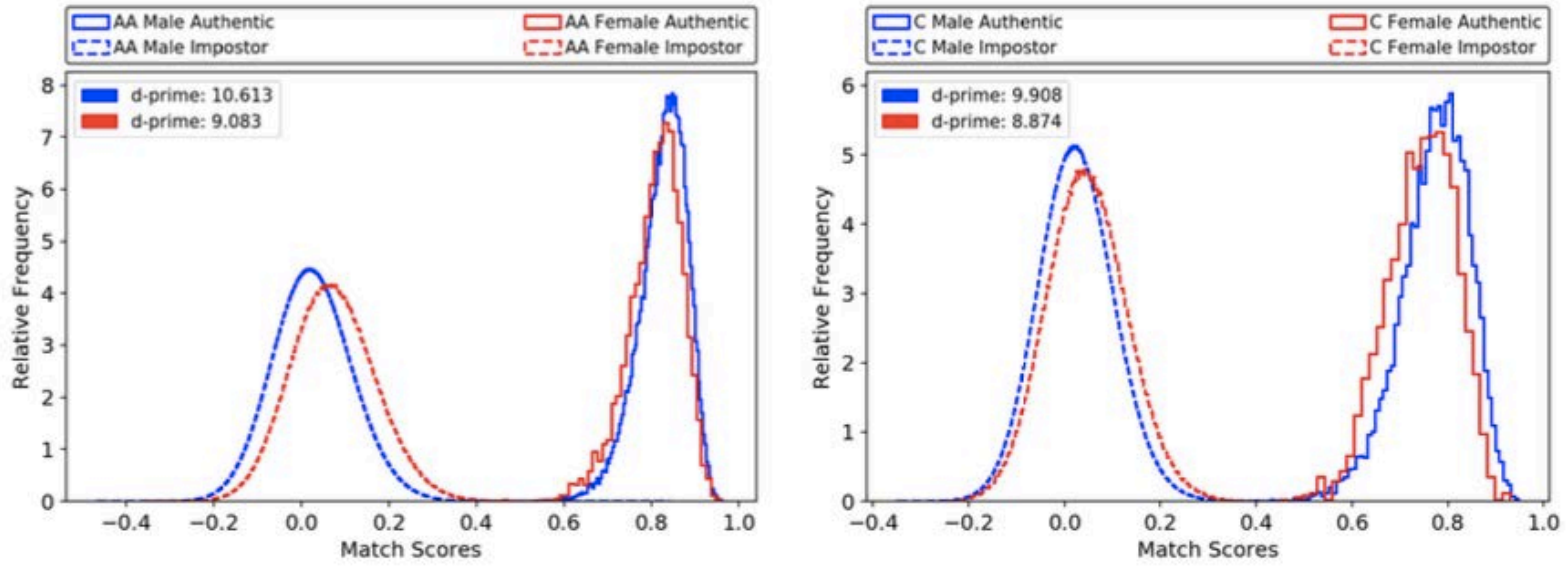
Figure 4. Composite faces and the average facial landmarks built by averaging faces classified as most and least likely to be gay.

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

LOYOLA UNIVERSITY CHICAGO

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

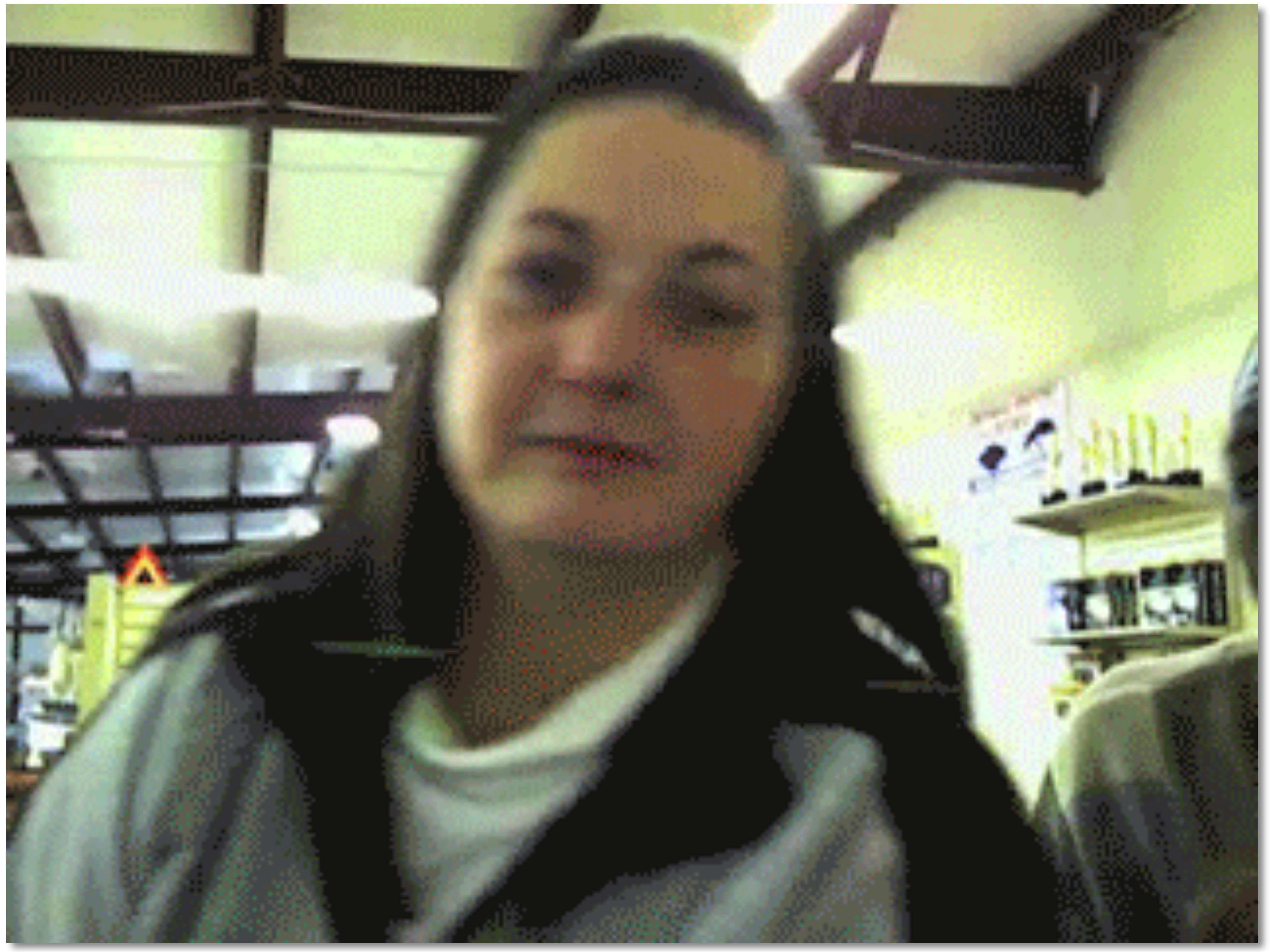
As of this writing, the database contains 1,724 face images of 515 individuals. These images represent a diverse population with respect to age, gender, and ethnicity. There are 1,278 images of individuals of African-American decent, 433 images of individuals of Caucasian decent and 3 images classified 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)?



LOYOLA UNIVERSITY CHICAGO

Data-driven Face Recognition

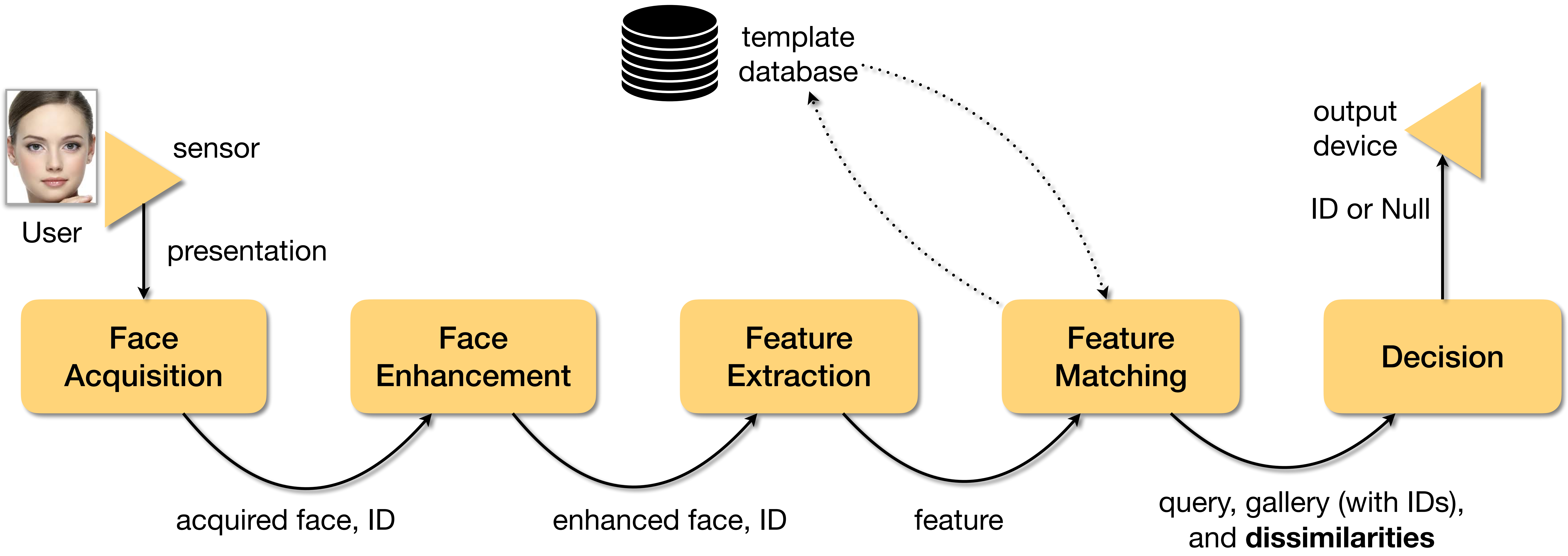
Problems

Avoid Bias
Diversify the training dataset.

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

LOYOLA UNIVERSITY CHICAGO

What’s Next?



LOYOLA UNIVERSITY CHICAGO

What’s Next?

Face Recognition Coding Class
Please bring your computers.

Fill out your
Today-I-missed **Statement**
Please visit
https://sakai.luc.edu/x/PnQvIG.

LOYOLA UNIVERSITY CHICAGO