http://bit.ly/dlcv2018

#DLUPC

Day 2 Lecture 2

Face Recognition

Elisa Sayrol Clols
elisa.sayrol@upc.edu

Associate Professor
Universitat Politecnica de Catalunya
Technical University of Catalonia

Acknowledgments
Working in this area @UPC (GPI)

Ramon Morros
ramon.morros@upc.edu

.... and plenty of students

Face Recognition

- Face Detection & Alignment/ Frontalization
- Face Recognition
    - Face Identification (Classification)
    - Face Verification (Binary Decision)
- Data Bases
- Main Contributions
    - DeepFace
    - FaceNet
    - DeepID
    - Sphere Faces
- Challenges
- UPC Opportunities

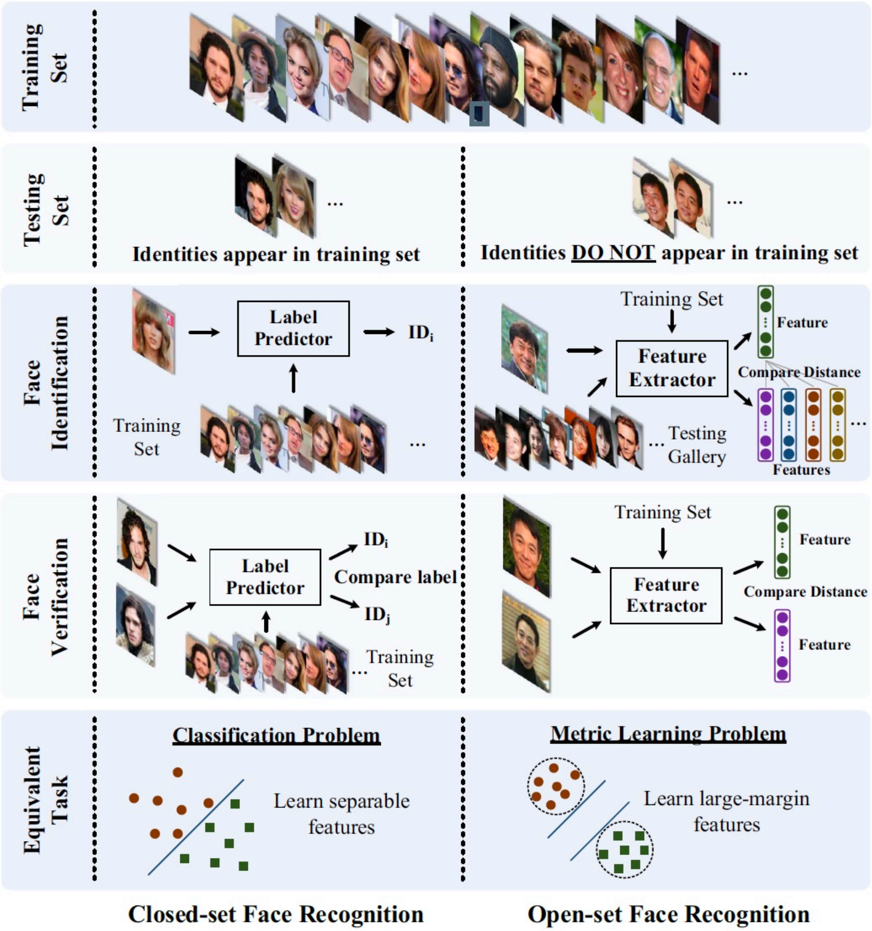
Survey: Wang A. et al., Deep Face Recognition: A Survey, arXiv June 2018

Face Detection and Recognition

[NEC’s NoeFace]

Identification

Verification



Liu W. et al., [SphereFace: Deep Hypersphere Embedding for Face Recognition](), CVPR 2017

Databases

Labeled Faces in the Wild (LFW)
13.000 images of faces collected from the web,

1.680 of the people pictured have two or more
Labeled Faces in the Wild: A Survey.
In *Advances in Face Detection and Facial Image Analysis*,
edited by Michal Kawulok, M. Emre Celebi,
and Bogdan Smolka, Springer, pages 189-248, 2016.

MS-Celeb-1M about 10M images for 100K celebrities,

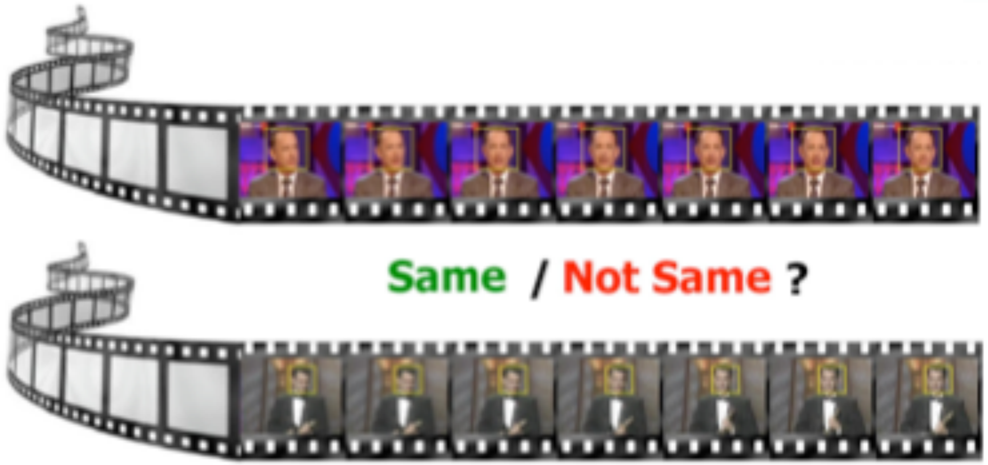
from MicroSoft

MegaFace and MF2: 4.7 milion faces, 672.057 unique
people, 1M distractors. Uses FaceScrub and FG-NET

CelebFaces+

202.599 face images of 10.177 identities (celebrities).
People in LFW and CelebFaces+ are mutually exclusive.

Databases



YouTube Faces:

3.425 videos, 1.595 identities (celebrities).

621.126 picture images that come from the videos so there is not a lot of variability between them.

May overlap with other celebrity databases

Available info: Original frames, cropped faces, aligned faces.

Head-pose angles for all the faces

FaceScrub

106.863 photos of 530 celebrities, 265 whom are male (55.306 images), and 265 female (51.557 images). Face bounding boxes provided. Full frame and cropped version available.

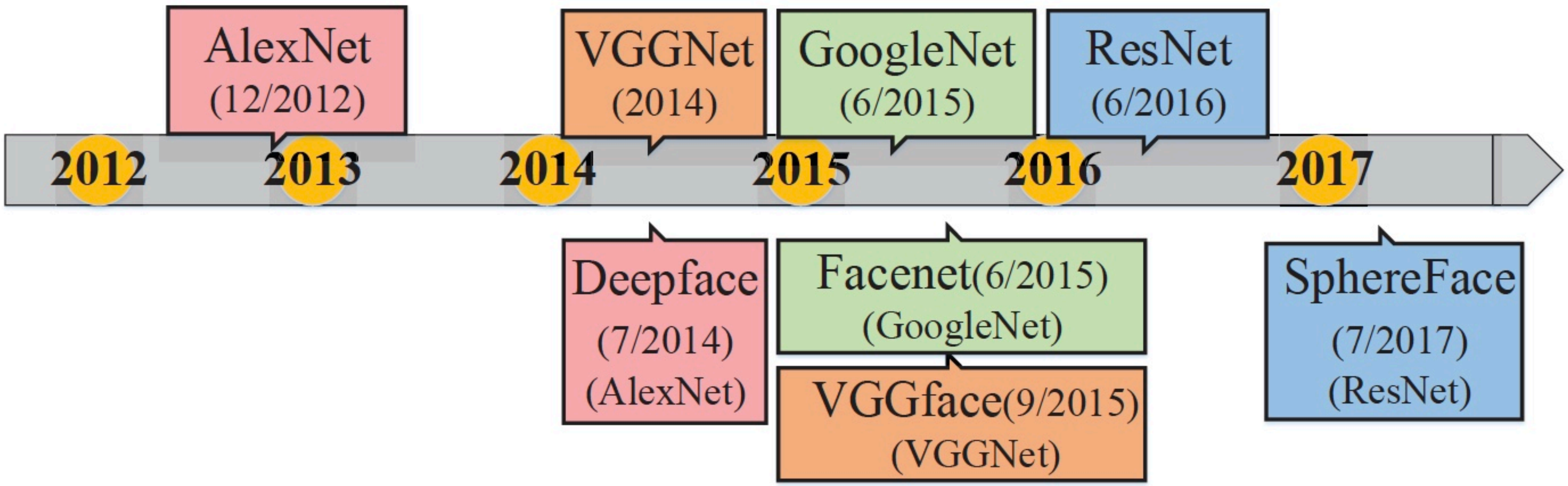
GoogleUPC !!!

FaceBook, Google, VGGFaces are private Databases

Reference: Wang A. et al., Deep Face Recognition: A Survey, arXiv June 2018

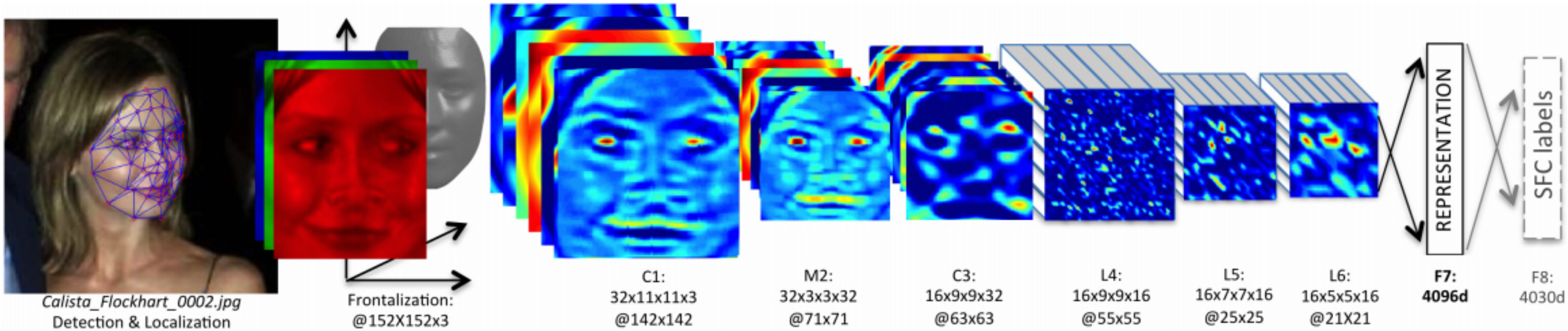
Other databases: https://deeplearning4j.org/opendata

Evolution of Main Contributions



Reference: Wang A. et al., [Deep Face Recognition: A Survey](), arXiv June 2018

DeepFace Architecture



Yaniv Taigman, etc (Facebook) . [DeepFace: Closing the Gap to Human-Level Performance in Face Verification](), CVPR 2014

DeepFace, Verification

A) Weighted χ2 distance

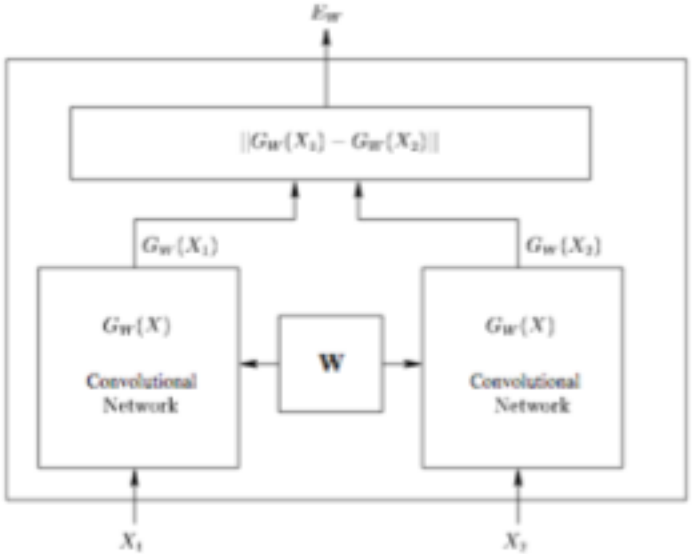
$$\chi^2(f_1,f_2) = \sum_i w_i \, {(f_1[i] - f_2[i])^2} \Big/ {(f_1[i] + f_2[i])}$$

where f_1 and f_2 are the DeepFace Representations.

The weights parameters w_i are learned using a linear SVM

*S. Chopra, R. Hadsell, and Y. LeCun. [Learning a similarity metric discriminatively, with application to face verification](), CVPR,2005.

B) Use of **Siamese Networks** inspired

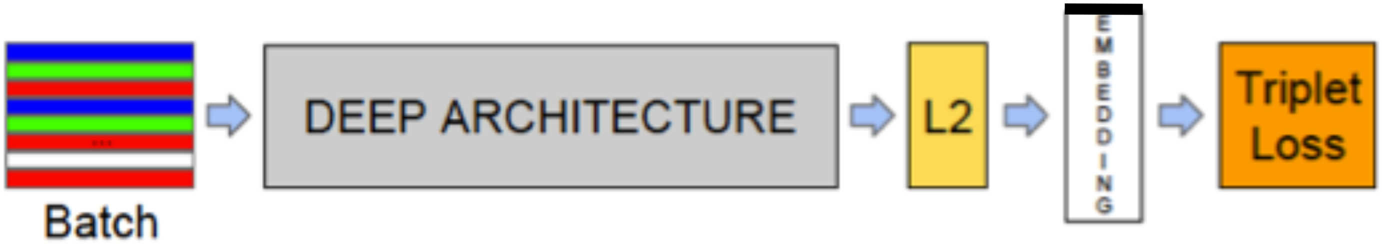


In DeepFace: $d(f_1,f_2) = \sum_i \alpha_i \left| f_1[i] - f_2[i] \right|$

α_i are the trainable parameters with standard cross-entropy loss and backward propagation

FaceNet

This Face recognition/verification/clustering model learns a mapping from face images to a compact **Euclidean space** where distances directly correspond to a measure of face similarity.

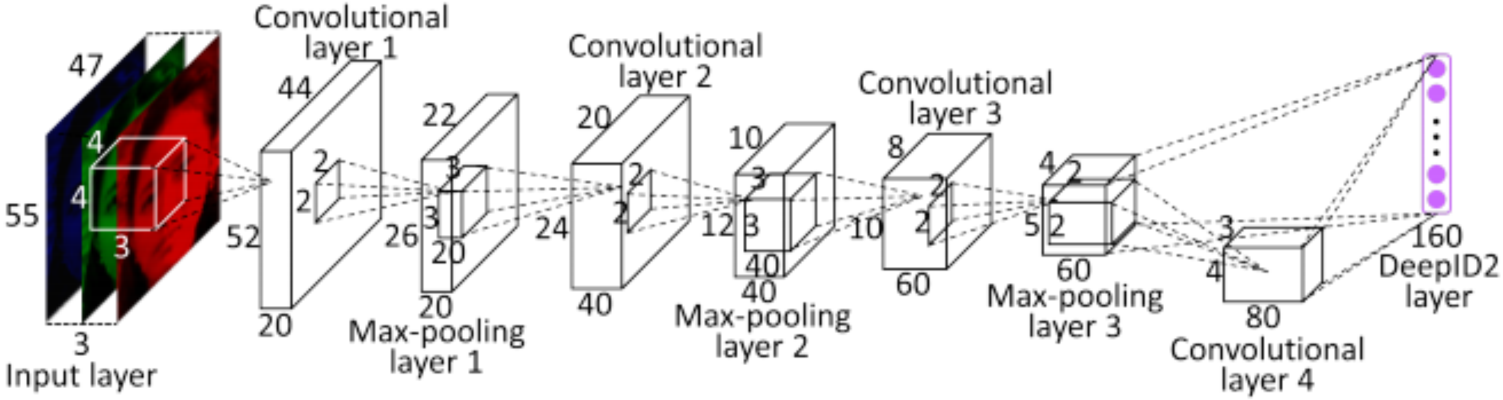


$$\|x_i^a - x_i^p\|_2^2 + \alpha < \|x_i^a - x_i^n\|_2^2, \ \forall (x_i^a, x_i^p, x_i^n) \in \mathcal{T}.$$

Triplet Loss function $\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$ Where f is the embedding

Florian Schroff et al. (Google) FaceNet: A Unified Embedding for Face Recognition and Clustering, CVPR 2015

Deep ID2



Parameters of the Network: $f = \mathrm{Conv}(x, \theta_c),$

But you compute parameters From a Verification loss function and an Identification loss Function

$$\mathrm{Verif}(f_i, f_j, y_{ij}, \theta_{ve}) = \begin{cases} \frac{1}{2}\|f_i - f_j\|_2^2 & \text{if } y_{ij} = 1 \\ \frac{1}{2}\max\left(0, m - \|f_i - f_j\|_2\right)^2 & \text{if } y_{ij} = -1 \end{cases}$$

$$\mathrm{Ident}(f, t, \theta_{id}) = -\sum_{i=1}^{n} -p_i \log \hat{p}_i = -\log \hat{p}_t ,$$

Yi Sun, etc. Deep Learning Face Representation by Joint Identification-Verification. NIPS 2014

Deep ID2

When you backprop you backprop gradients of verification and identification parameters and you also update the weight of the convolutional layers

Table 1: The DeepID2 learning algorithm.

input: training set $\chi = \{(x_i, l_i)\}$, initialized parameters θ_c, θ_{id}, and θ_{ve}, hyperparameter λ, learning rate $\eta(t)$, $t \leftarrow 0$

while not converge **do**

$t \leftarrow t+1$ sample two training samples (x_i, l_i) and (x_j, l_j) from χ

$f_i = \mathrm{Conv}(x_i, \theta_c)$ and $f_j = \mathrm{Conv}(x_j, \theta_c)$

$\nabla\theta_{id} = \frac{\partial \mathrm{Ident}(f_i, l_i, \theta_{id})}{\partial \theta_{id}} + \frac{\partial \mathrm{Ident}(f_j, l_j, \theta_{id})}{\partial \theta_{id}}$

$\nabla\theta_{ve} = \lambda \cdot \frac{\partial \mathrm{Verif}(f_i, f_j, y_{ij}, \theta_{ve})}{\partial \theta_{ve}}$, where $y_{ij} = 1$ if $l_i = l_j$, and $y_{ij} = -1$ otherwise.

$\nabla f_i = \frac{\partial \mathrm{Ident}(f_i, l_i, \theta_{id})}{\partial f_i} + \lambda \cdot \frac{\partial \mathrm{Verif}(f_i, f_j, y_{ij}, \theta_{ve})}{\partial f_i}$

$\nabla f_j = \frac{\partial \mathrm{Ident}(f_j, l_j, \theta_{id})}{\partial f_j} + \lambda \cdot \frac{\partial \mathrm{Verif}(f_i, f_j, y_{ij}, \theta_{ve})}{\partial f_j}$

$\nabla\theta_c = \nabla f_i \cdot \frac{\partial \mathrm{Conv}(x_i, \theta_c)}{\partial \theta_c} + \nabla f_j \cdot \frac{\partial \mathrm{Conv}(x_j, \theta_c)}{\partial \theta_c}$

update $\theta_{id} = \theta_{id} - \eta(t) \cdot \theta_{id}$, $\theta_{ve} = \theta_{ve} - \eta(t) \cdot \theta_{ve}$, and $\theta_c = \theta_c - \eta(t) \cdot \theta_c$.

end while

output θ_c

Deep ID2

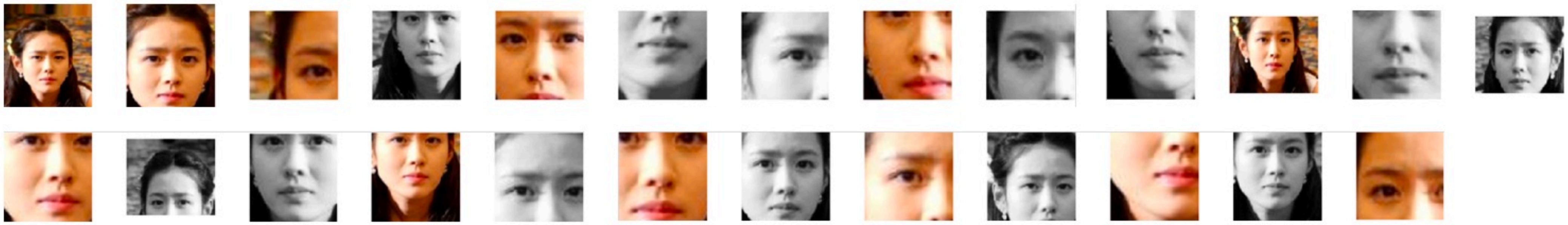


Figure 2: Patches selected for feature extraction.

25 patches generate 25 160-dimensional DeepID vectors
4000 dimensional DeepID2 vector
Compressed by PCA

Deep ID2

DeepID2 Uses a Joint Bayesian model in top of the network for face verification.

If we model a face as $x = \mu + \varepsilon$

μ Represents the Identity. Extrapersonal variations

ε Intrapersonal variations

Both Gaussian Distributed, estimated during training

Verification is achieved through Log-Likelihood Ratio Test:

$$r(x_1, x_2) = \log \frac{P(x_1, x_2 | H_I)}{P(x_1, x_2 | H_E)}$$

H_I Interpersonal hypothesis
H_E Extrapersonal hypothesis

Chen, et al. [Bayesian Face Revisited: A Joint Formulation](), ECCV 2012

SphereFace

1) From the SoftMax Loss Function

$$L_i = -\log\left(\frac{e^{\boldsymbol{W}_{y_i}^T \boldsymbol{x}_i + b_{y_i}}}{\sum_j e^{\boldsymbol{W}_j^T \boldsymbol{x}_i + b_j}}\right)$$
$$= -\log\left(\frac{e^{\|\boldsymbol{W}_{y_i}\|\|\boldsymbol{x}_i\|\cos(\theta_{y_i,i}) + b_{y_i}}}{\sum_j e^{\|\boldsymbol{W}_j\|\|\boldsymbol{x}_i\|\cos(\theta_{j,i}) + b_j}}\right)$$

2) To the modified SoftMax Loss Function

$$L_{\text{modified}} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\|\cos(\theta_{y_i,i})}}{\sum_j e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\right)$$

3) With the idea of increasing distances between clases, an ***m*** factor is introduced.

That is correctly, for example, classifying identity 2 with respect to identity 1 will require $\theta_2 > \frac{\theta_1}{m}$ instead of $\theta_2 > \theta_1$

$$L_{\text{ang}} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\|\cos(m\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\|\cos(m\theta_{y_i,i})} + \sum_{j\neq y_i} e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\right)$$

4) Finally the proposed Angular–SoftMax Loss is

$$L_{\text{ang}} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\|\psi(\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\|\psi(\theta_{y_i,i})} + \sum_{j\neq y_i} e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\right)$$

with

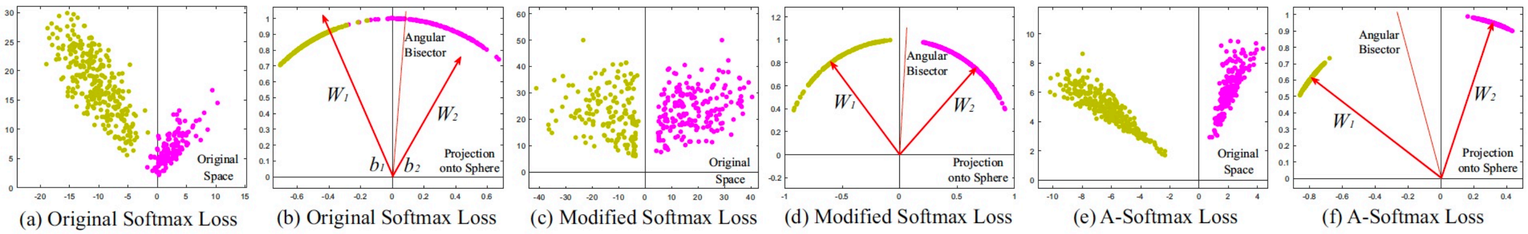
$$\psi(\theta_{y_i,i}) = (-1)^k \cos(m\theta_{y_i,i}) - 2k$$

$$\theta_{y_i,i} \in \left[\frac{k\pi}{m}, \frac{(k+1)\pi}{m}\right] \text{ and } k \in [0, m-1].$$

Liu W. et al., SphereFace: Deep Hypersphere Embedding for Face Recognition, CVPR 2017

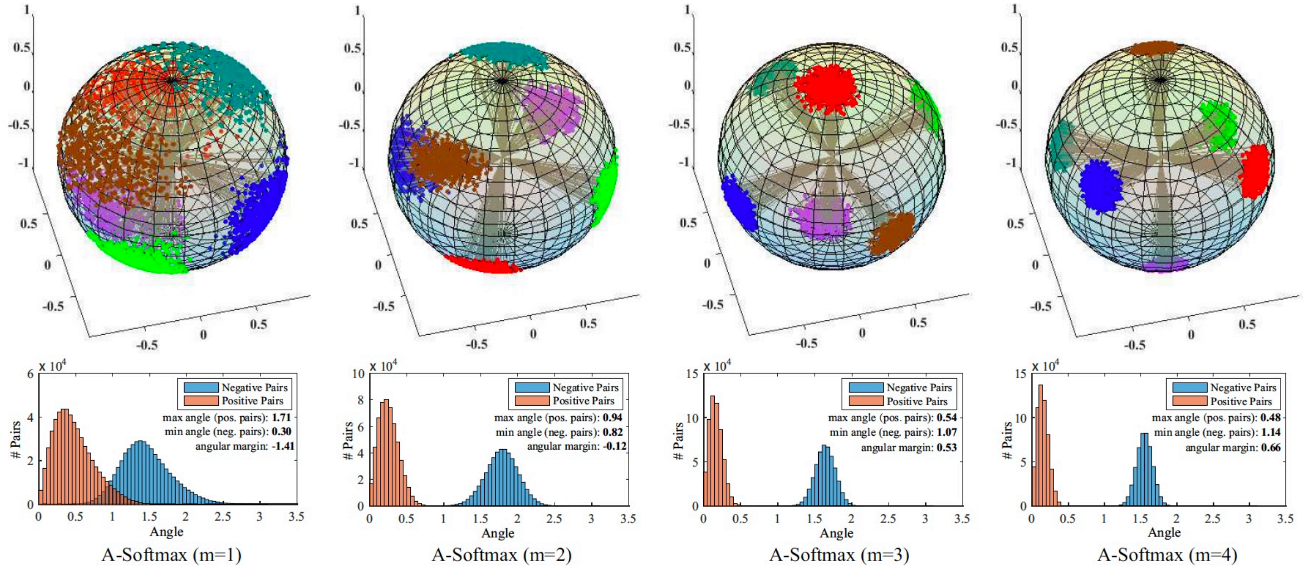
SphereFace

Example of 2-D features and its projections onto a sphere



Example of 3-D features and its projections onto the unit sphere

Angle distribution of positive and negative países as a function of m

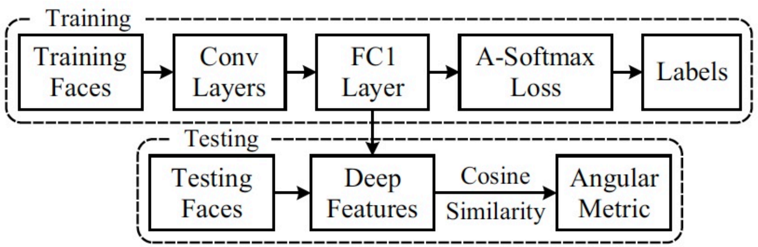


Liu W. et al., [SphereFace: Deep Hypersphere Embedding for Face Recognition], CVPR 2017

SphereFace

Accuracy Results and Comparisons On LFW and YTF

CNN architecture with upto 64 layers with residual units:



Method	Models	Data	LFW	YTF
DeepFace	3	4M*	97.35	91.4
FaceNet	1	200M*	**99.65**	95.1
Deep FR	1	2.6M	98.95	**97.3**
DeepID2+	1	300K*	98.70	N/A
DeepID2+	25	300K*	99.47	93.2
Baidu	1	1.3M*	99.13	N/A
Center Face	1	0.7M*	99.28	94.9
Yi et al.	1	WebFace	97.73	92.2
Ding et al.	1	WebFace	98.43	N/A
Liu et al.	1	WebFace	98.71	N/A
Softmax Loss	1	WebFace	97.88	93.1
Softmax+Contrastive	1	WebFace	98.78	93.5
Triplet Loss	1	WebFace	98.70	93.4
L-Softmax Loss	1	WebFace	99.10	94.0
Softmax+Center Loss	1	WebFace	99.05	94.4
SphereFace	1	WebFace	**99.42**	**95.0**

Liu W. et al., [SphereFace: Deep Hypersphere Embedding for Face Recognition](), CVPR 2017

Open Challenges in Face Recognition

- Cross-Pose Face Recognition. Many existing algorithms suffer a decrease of over 10% from frontal to profile verification
- Cross-Age Face Recognition: Cross-age FR is extremely challenging due to the changes in facial appearance by the aging process over time.
- Makeup Face Recognition
- NIR-VIS Face Recognition: Due to the excellent performance of the near-infrared spectrum (NIS) images under lowlight scenarios, NIS images are widely applied in surveillance systems.
- Low-Resolution Face Recognition:
- Photo-Sketch Face Recognition

Open Challenges in Face Recognition

- Low-Shot Face Recognition. Limited number of shots and with blur, occlusion, and different expressions
- Set/Template-Based Face Recognition: FR problems assume that both probe and gallery sets are represented using a set of media, e.g., images and videos, rather than just one.
- Video Face Recognition
- 3D Face Recognition
- Face Anti-spoofing
- Face Recognition for Mobile Devices

Face Recognition @UPC with Ramon Morros

- Participation in Projects and Challenges, Camomile & Mediaeval, to identify People in Video Sequences with Audio
- Advisors of TFGs, Introduction to Research and TFM since 2015
- Currently working on Incremental Learning, MultiTask Learning, Progressive Networks and Open Set Deep Learning (OpenMax)

Selection of Papers: DNN Face Recognition Papers

Thank You!!

Questions?