Day 3 Lecture 5

Face Recognition

Acknowledments:
Prof. Ramon Morros
Students: Gerard Martí (Ms CV), Carlos Roig (Bs Tel) and Alessandro Vilardi (Ms EE)

Elisa Sayrol

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Department of Signal Theory and Communications
Image Processing Group

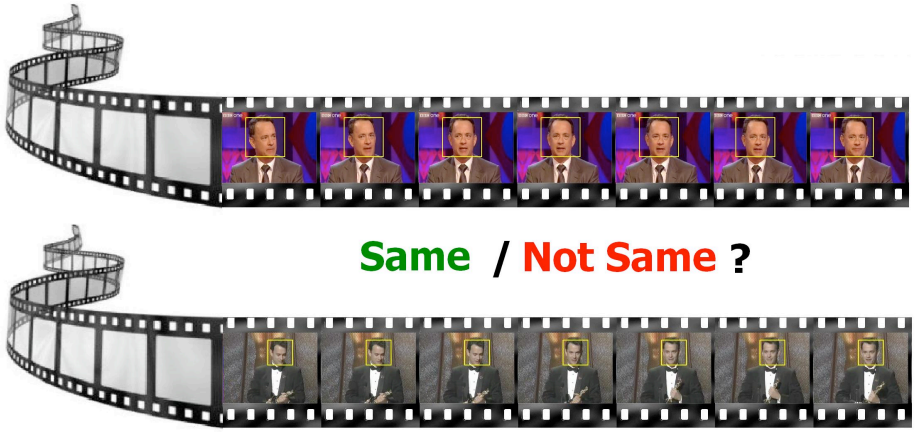
Face Recognition

- Face Detection
- Face Alignment/ Frontalization
- Face Recognition
    - Face Identification (Classification)
    - Face Verification(Binary Decision)
- People Recognition in videos (Camomile Project at UPC)
    - Speech
    - Face
    - Text

Face Recognition

- Databases
- Well-Known Systems
  - Deep Face (FaceBook)
  - FaceNet (Google)
  - Deep ID
- Some experiments at UPC

Databases



YouTube Faces: [http://www.cs.tau.ac.il/~wolf/ytfaces/]
621126 pictures, 1595 identities (celebrities). Images come from 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: [http://vintage.winklerbros.net/facescrub.html]
[http://megaface.cs.washington.edu/participate/challenge.html]
106863 photos of 530 celebrities, 265 whom are male (55306 images), and 265 female (51557 images). Face bounding boxes provided. Full frame and cropped version available.

MegaFace: [http://megaface.cs.washington.edu/]
1 milion faces, 690572 unique people

MSRA-CFW [http://research.microsoft.com/en-us/projects/msra-cfw/]
202792 faces, 1583 people (celebrities). May overlap with other celebrity databases.
Links, has to be downloaded (downloading!).

Databases

Labeled Faces in the Wild (LFW)
[http://vis-www.cs.umass.edu/lfw/]
13,000 images of faces collected from the web,
1680 of the people pictured have two or more distinct photos in the data set.

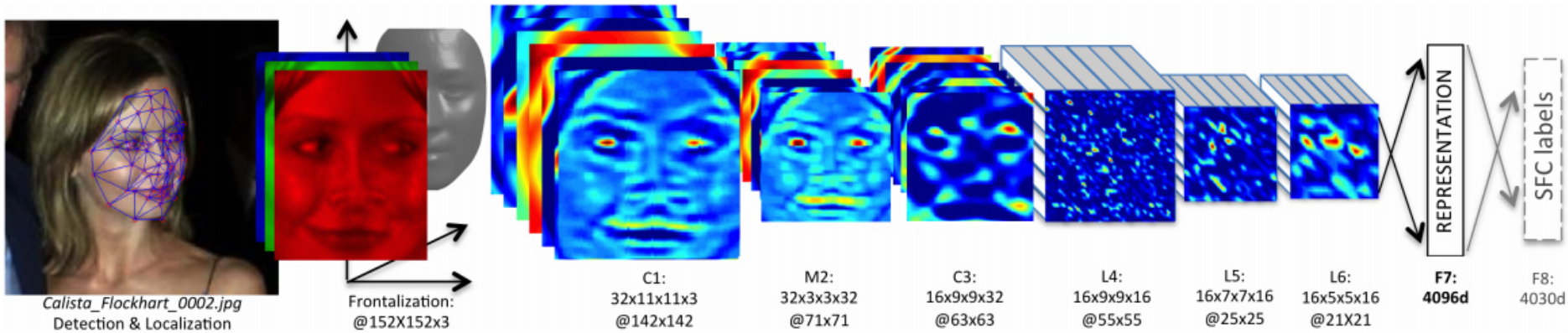
CelebFaces
[http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html]
202599 face images of 10177 identities (celebrities).
People in LFW and CelebFaces+ are mutually exclusive.

10k US Adult Faces [http://www.wilmabainbridge.com/facememorability2.html]
~10000 images, ?? people (celebrities excluded manually).
/work/morros/faces/facedatabase/

CASIA [http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html]
494,414 images, 10,575 subjects.

GoogleUPC !!!

DeepFace Architecture



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

DeepFace, Verification

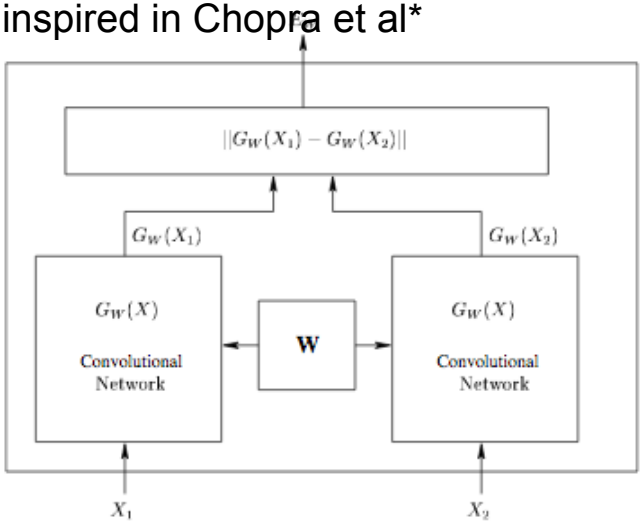
A) Weighted χ2 distance

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

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 met-ric discriminatively, with application to face verification, CVPR,2005.

B) Use of **Siamese Networks** inspired in Chopra et al*

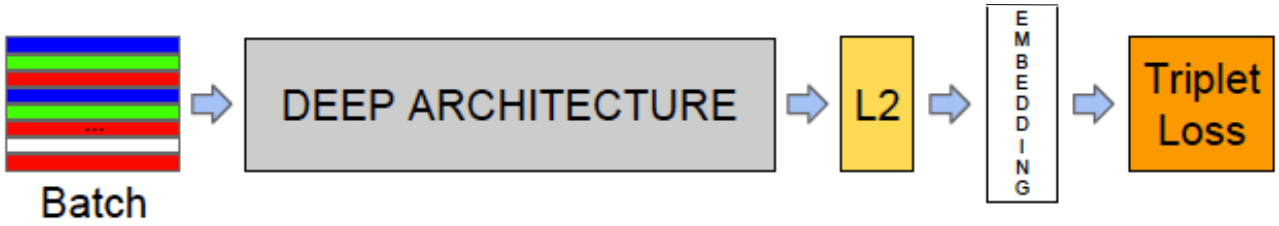


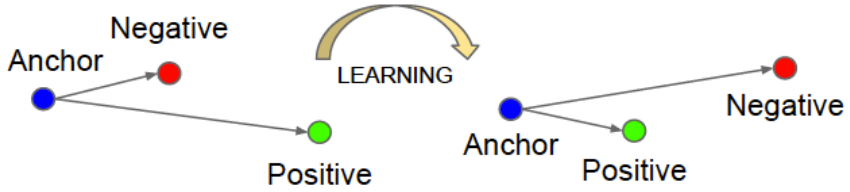
In DeepFace: $d(f_1, f_2) = \sum_i \alpha_i |f_1[i] - f_2[i]|$

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.



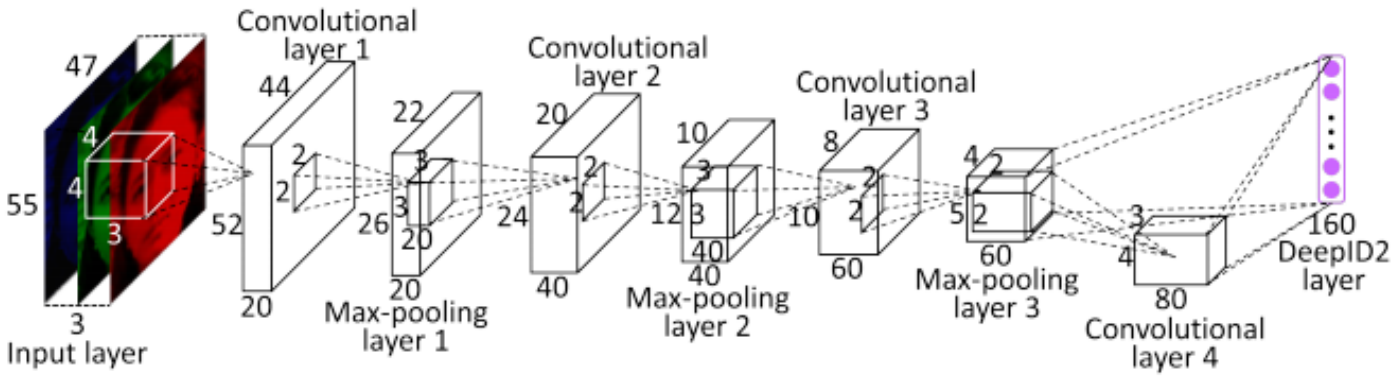


$$\|x_i^a - x_i^p\|_2^2 + \alpha < \|x_i^a - x_i^n\|_2^2, \forall (x_i^a, x_i^p, x_i^n) \in \mathcal{T} .$$

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

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

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

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

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

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

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

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

DeepID2 Uses a Joint Bayesian model in top of the network for face verification.
If we model a face as $x = \mu + \varepsilon$

- μ Interpersonal variations
- ε Intrapersonal variations

  Both Gaussian Distributed, estimated during Training

Verification is achieved through Log-Likelihood Ratio Test:

$$r(x_1, x_2) = \log \frac{P(x_1, x_2|H_I)}{P(x_1, x_2|H_E)} = x_1^T A x_1 + x_2^T A x_2 - 2x_1^T G x_2,$$

where

$$A = (S_\mu + S_\varepsilon)^{-1} - (F + G),$$

$$\begin{pmatrix} F+G & G \\ G & F+G \end{pmatrix} = \begin{pmatrix} S_\mu + S_\varepsilon & S_\mu \\ S_\mu & S_\mu + S_\varepsilon \end{pmatrix}^{-1}.$$

Chen, et al. Bayesian Face Revisited: A Joint Formulation, ECCV 2012

Experiments at UPC Face recognition (2015)

Comparing Face recognition

Deep Face CNN
4,4 million images
4030 people
93% accuracy

Deep Face results

DeepID2
202,599 images
10,177 people
98,9% accuracy

DeepID2 results

Imagenet Fine-Tuning
10.422 images
520 people
97,3% accuracy

Imagenet Fine-Tuning results (Msc Sergi Delgado)

Experiments at UPC Face recognition (2015)

Our own database to be used in the Camomile EU Project
- 520 instances composed by 10.422 images.
- 8858 images were used in training stage, and 1.564 for testing.

Essex Dataset

Crops from TV show videos

Ongoing experiments at UPC Face recognition (2016)

Ramon Morros

Students Carlos Roig (Bs in Tel), Alessandro Vilardi (Ms in EE), Gerard Martí (Ms in CV)

Face Recognition using Very Deep Neural Networks

- VGG
- GoogleNet
- ResNet
- Ensenble VGG+GoogleNet

Pre-trained Networks with VGG-Imagenet or VGG-Faces. Google Net and ResNet pretrained over Imagenet.
Experiments with YouTube Faces, FaceScrub and Google UPC Faces

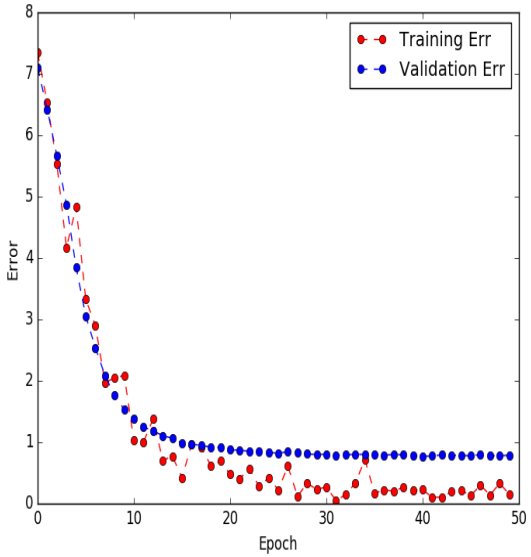
Ongoing experiments at UPC Face recognition (2016)

Ramon Morros

Students Carlos Roig (Bs in Tel), Alessandro Vilardi (Ms in EE), Gerard Martí (Ms in CV)

Experiments with a compound DataBase with YouTube Faces, FaceScrub and LFW
3.500 identities
100.000 images

With VGG



Ongoing experiments at UPC Face recognition (2016)

Ramon Morros

Students Carlos Roig (Bs in Tel), Alessandro Vilardi (Ms in EE), Gerard Martí (Ms in CV)

Verification with:
VGG+Autoencoder with 8 hidden layers to reduce dimensionality, from 4096 to 256 vector+ Joint Bayesian

Results with DataBase 1 (the previous one without YTF, and a Test set with FaceScrub and LFW)

Precision	Recall	f1-score	support (pairs of the dataset)
0.97	**0.95**	**0.96**	**2288**

Results with DataBase 2 (the previous one without YTF, and a Validation set of LFW)

Precision	Recall	f1-score	support(pairs of the dataset)
0.80	**0.80**	**0.79**	**998**

Undergoing Experiments also with Advanced Joint Bayesian, Siamese networks, Triplets….