Face Recognition: From traditional to learning-based methods

Giuliano Grossi

Department of Computer Science - University of Milan (Italy)

ICPRS, July 2019

Face Recognition...

still remains (after 50 years!) a hot topic because of

- *Lots of applications: biometrics recognition, military, finance, public security and daily life...*
- *The ongoing development is still far from coming to an end*
- *Technology breeding biases: see BBC News*

BBC - Technology

BBC | Sign in | News | Sport | Weather | Shop | Reel | Travel | More | Search

NEWS

Home | Video | World | UK | Business | Tech | Science | Stories | Entertainment & Arts | Health | World News TV | More

Technology

Use of facial recognition tech 'dangerously irresponsible'

By Geoff White
BBC Click

13 May 2019

Share

Police facial recognition systems have been shown to do poorly when analysing non-white faces

Black and minority ethnic people could be falsely identified and face questioning because police have failed to test how well their systems deal with non-white faces, say campaigners.

Emerging Technology

EMERGING TECH

Amazon's facial ID incorrectly identifies members of Congress as criminals

Congress members want a meeting with Jeff Bezos to discuss Rekognition mistakes

BBC - Technology

BBC | Sign in | News | Sport | Weather | Shop | Reel | Travel | More | Search

NEWS

Home | Video | World | UK | Business | Tech | Science | Stories | Entertainment & Arts | Health | World News TV | More

Technology

Apple AI accused of leading to man's wrongful arrest

23 April 2019

Share

Mr Bah claims Apple's facial recognition technology has mistakenly identified him as a thief

A student is suing Apple Inc for $1bn (£0.77bn), claiming that its in-store AI led to his mistaken arrest.

BBC - Technology

BBC | Sign in | News | Sport | Weather | Shop | Reel | Travel | More | Search

NEWS

Home | Video | World | UK | Business | Tech | Science | Stories | Entertainment & Arts | Health | World News TV | More

Technology

Face recognition police tools 'staggeringly inaccurate'

By Chris Fox
Technology reporter

15 May 2018 | Share

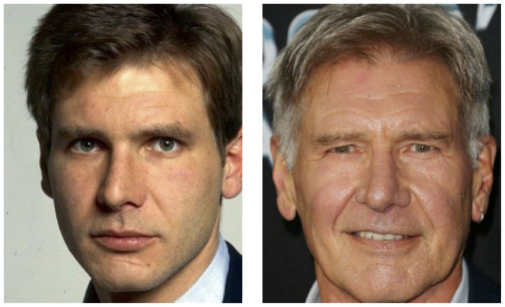
Police must address concerns over the use of facial recognition systems or may face legal action, the UK's privacy watchdog says.

Outline

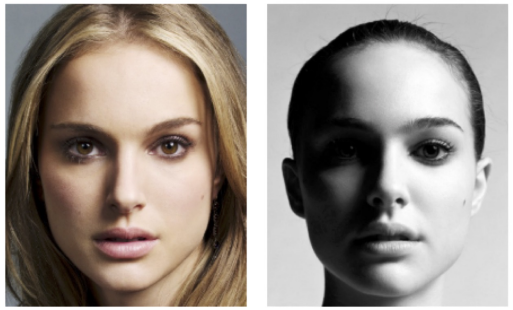
- Data collection issues
- Traditional methods, holistic, feature based...
- Deep learning based approach
- Sparse dictionary learning based approach
- Open problems, challenges and future research

Dataset & Face Variability

Head pose



Age



Illumination

Facial expression

Occlusion

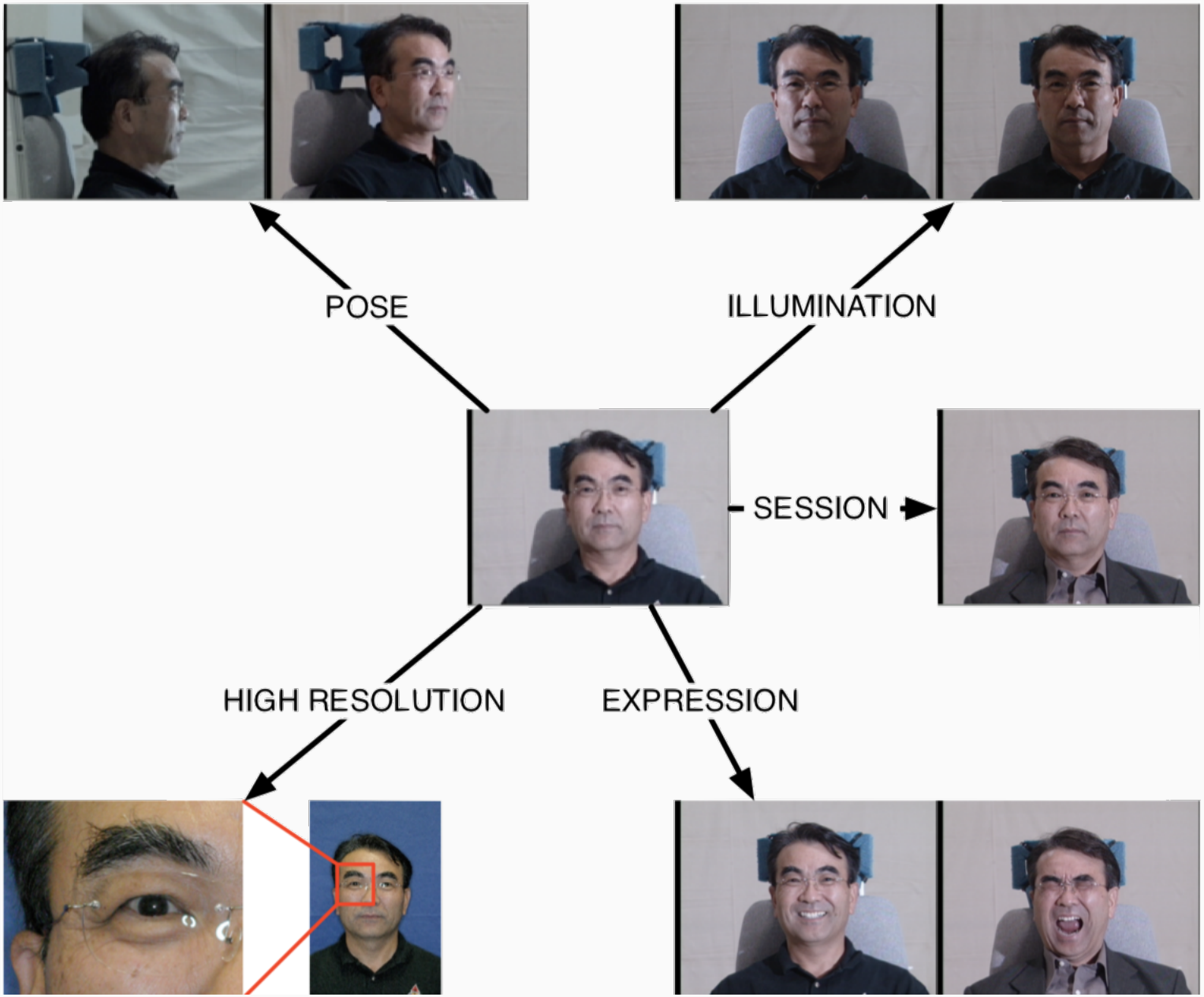
Disguise

Main Challanges... since 1966

> *"This recognition problem is made difficult by the* **great variability** *in* **head rotation** *and* **tilt**, **lighting intensity** *and* **angle**, **facial expression**, **aging**, *etc. Some other attempts at facial recognition by machine have allowed for little or no variability in these quantities. Yet the method of correlation (or* **pattern matching***) of unprocessed optical data, which is often used by some researchers,* **is certain to fail** *in cases where the* **variability is great**. *In particular, the correlation is very low between two pictures of the same person with two different head rotations."*

W. W. Bledsoe, *"The model method in facial recognition"*, Panoramic Research Inc., Palo Alto, CA, vol. 15, no. 47, 1966.

Dataset multi-PIE - Controlled Conditions (2000)



Dataset multi-PIE - Samples

multi-PIE (**multi Pose, Illumination, Expression**) dataset:

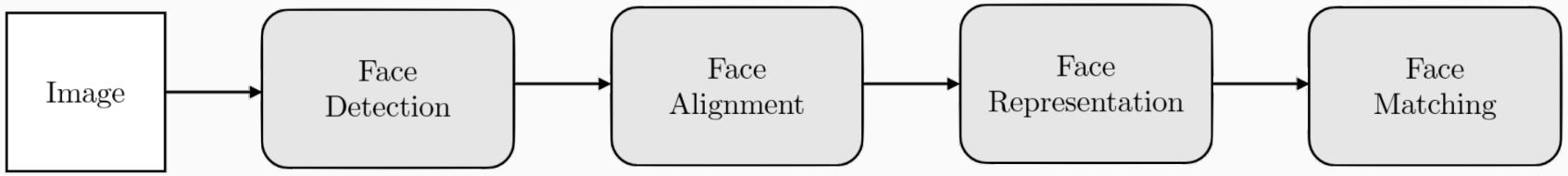
- 337 subjects, imaged under 15 view points
- 19 illumination conditions in up to 4 recording sessions
- **Controlled variations** is a shortcoming!

Dataset LFW - Unconstrained Conditions (2007)

LFW (**Labeled Faces in the Wild**) dataset:

- More than 13.000 images of 5749 different individuals
- 1680 subjects with at least 2 images
- **Natural variations** in pose, illumination, expression, presence of partial occlusions

FR Building Blocks



Face detection: Find the position of the faces in an image

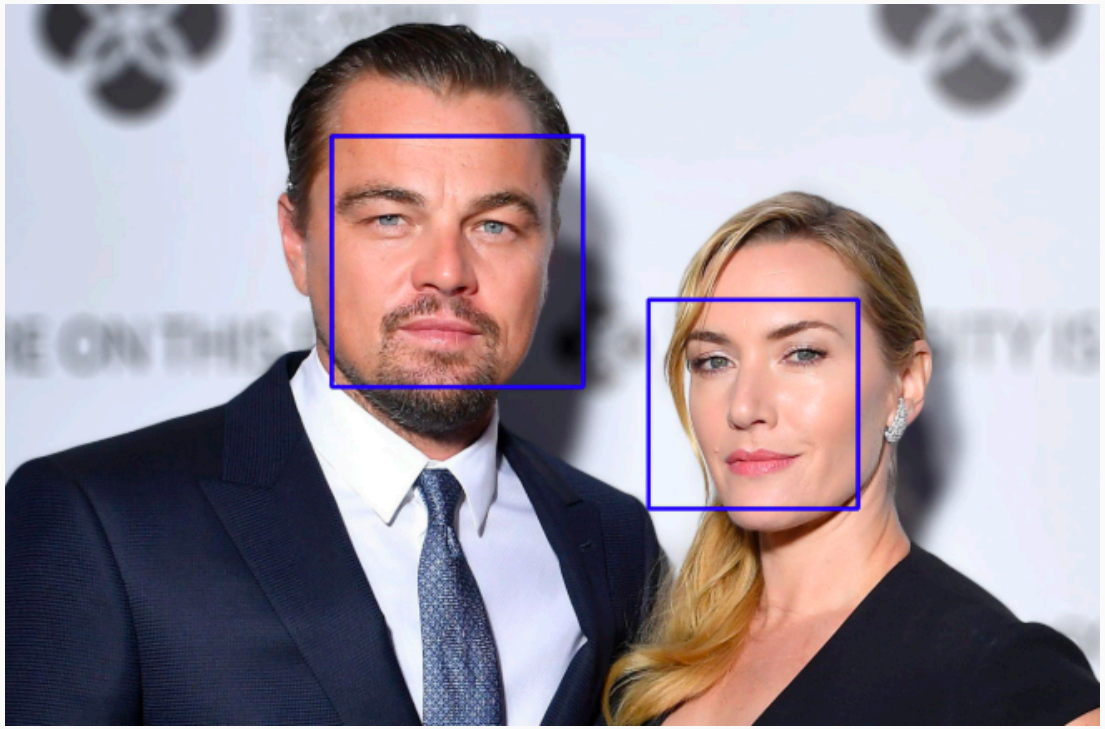
Face alignment: Align to fixed reference points

Face representation: Map image into a feature vector

Face matching: Provide a likelihood measure for each subject

Face Detection Step

- Find the **position** of the faces in an image (if any)
- Return the coords of a **bounding box** for each one
- Known face detector algorithms: '**Viola-Jones**'...

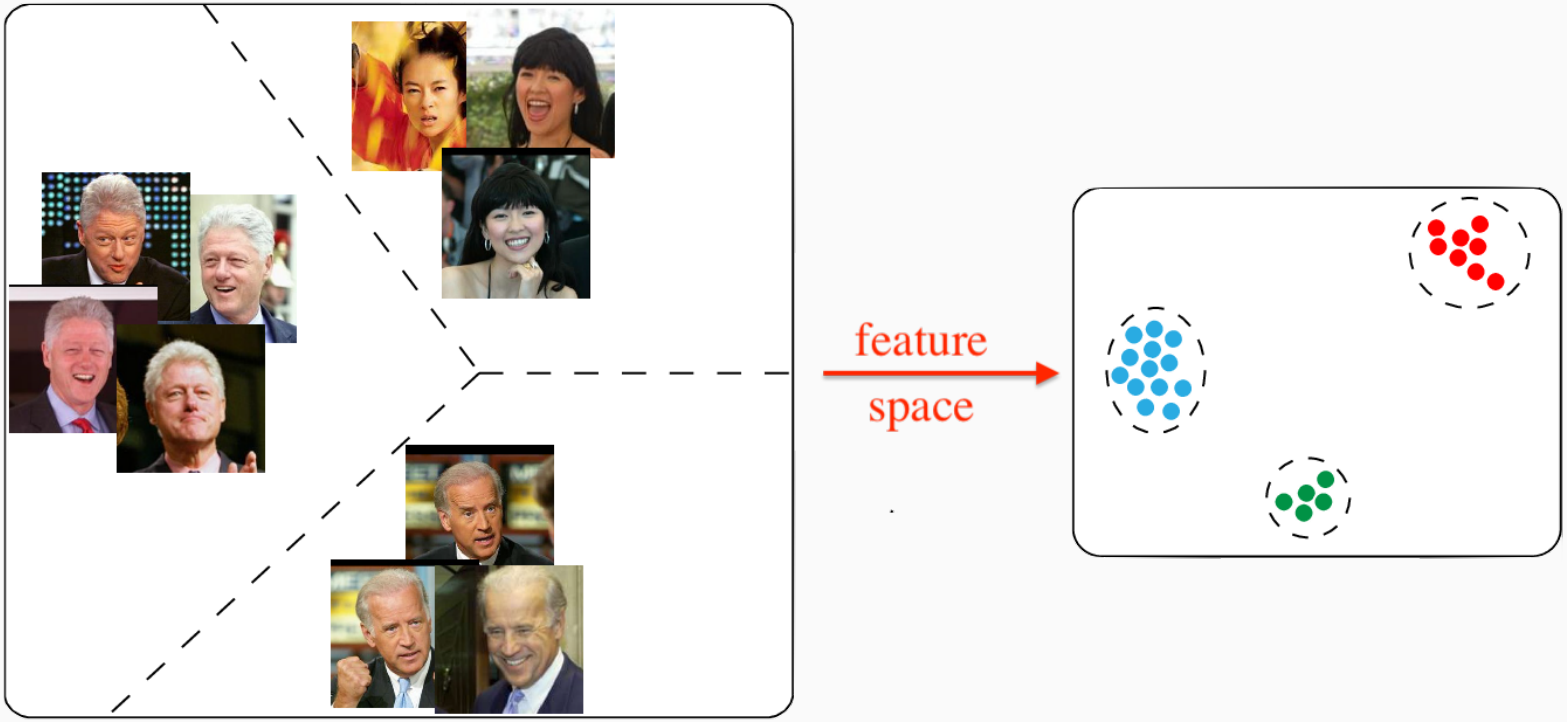


Face Alignment Step

- Find a set of **facial landmarks** (by a landmark detector)
- Use the best **affine transformation** to fit reference points

- 3D alignment algorithms can also achieve **face frontalisation**

Face Representation Step



- Pixel values of a face image are transformed into a compact and discriminative **feature vector**
- Ideally all the faces of a same subject should map to similar feature vectors

From traditional methods...

Traditional Methods i.e. Pre-Deep Learning

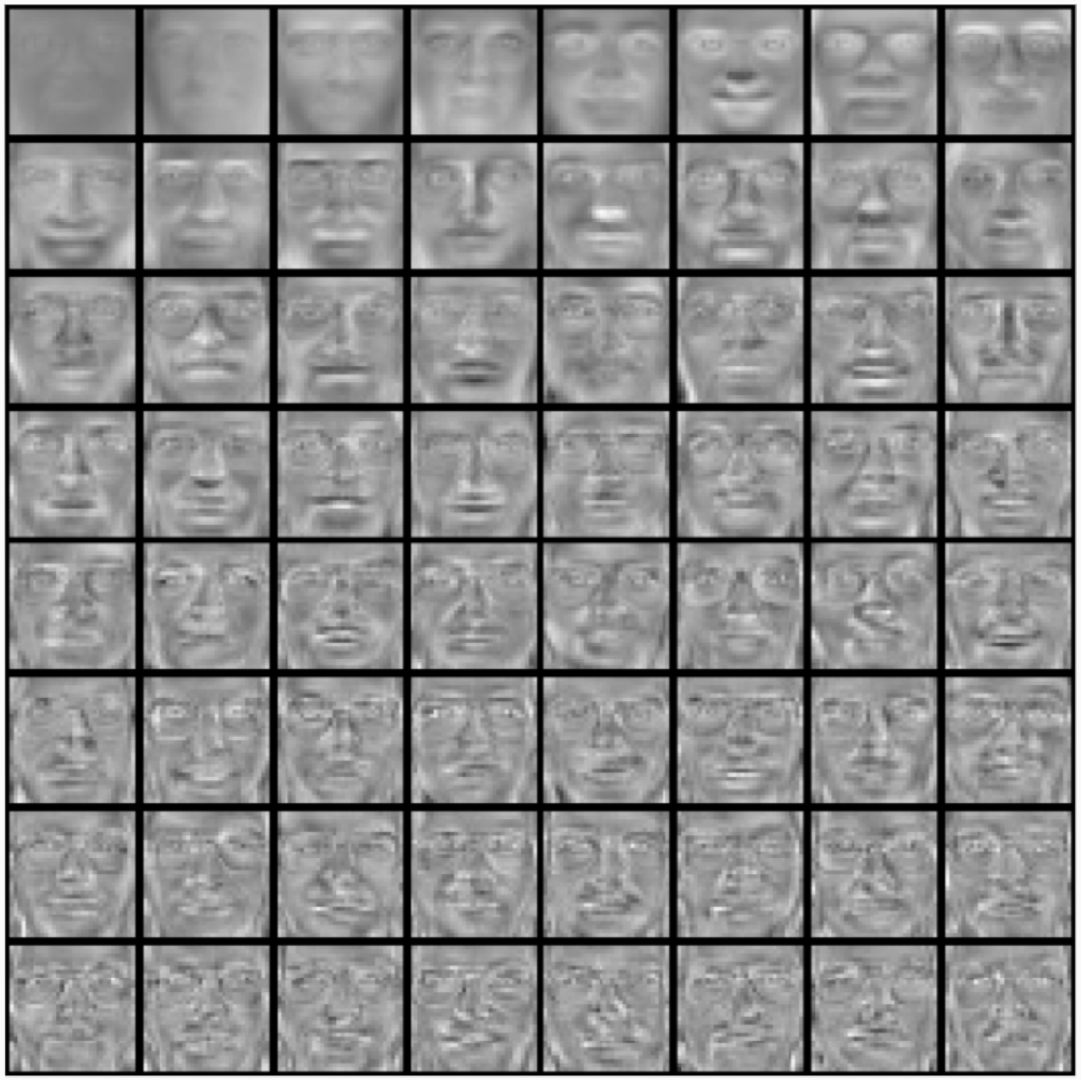
- **Geometry-based**: measuring relative positions and distances between set of facial landmarks, ...
- **Holistic**: projecting face images onto low-dimensional spaces, e.g. PCA Eigenfaces, LDA Fisher Faces, Bayesian analysis, SVM, ...
- **Feature-based**: histograms of LBP descriptors, Gabor feature maps, SIFT descriptor, ...
- **Hybrid**: combining techniques from holistic and feature-based methods, local features (e.g. LBP, SIFT) + projections on lower-dimensional subspaces (e.g. PCA or LDA)

Original Faces - Dataset

Eigenfaces - Eigenvectors [1991]

Mean: μ

Top eigenvectors: $\mathbf{u}_1, \ldots \mathbf{u}_k$



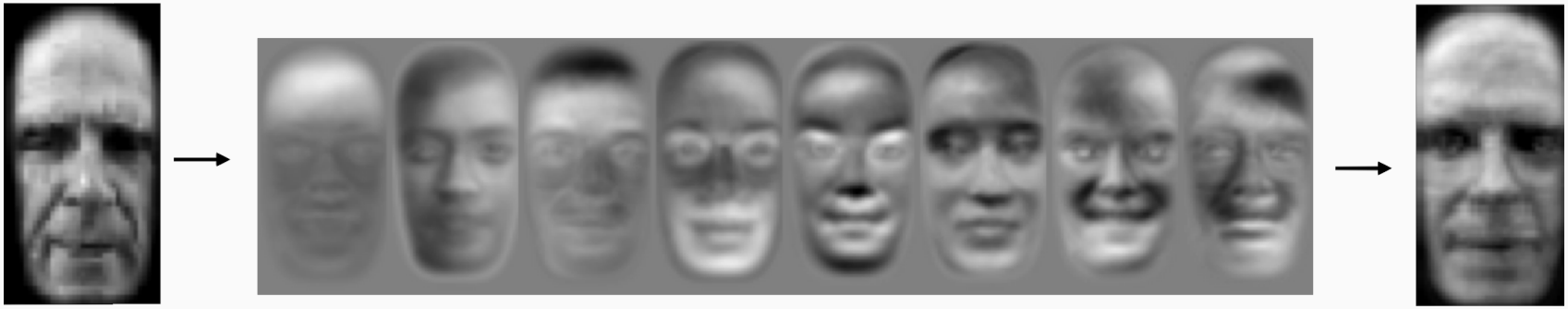
Eigenfaces Classification

- Compute **eigenfaces** $u_1, \ldots, u_k$ that span the space of faces

$$x \mapsto (\underbrace{(x-\mu)\cdot u_1}_{a_1}, \ldots, \underbrace{(x-\mu)\cdot u_k}_{a_k})$$

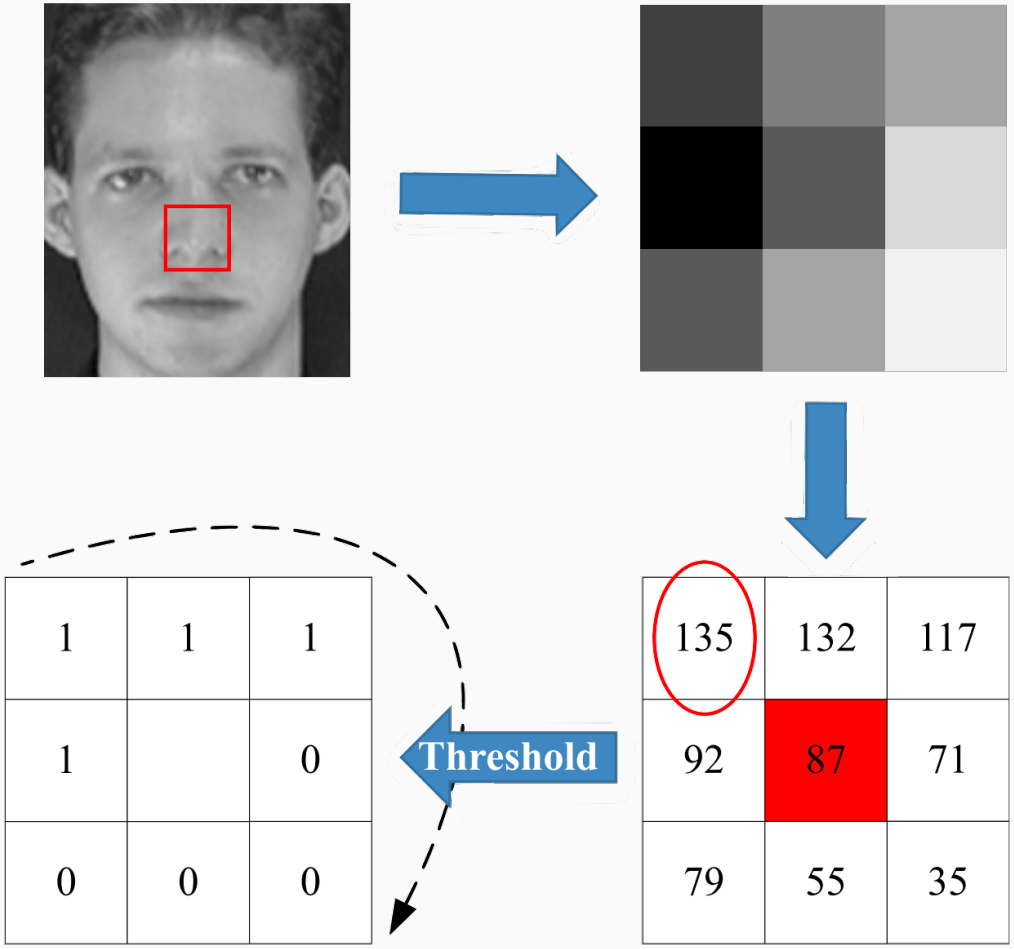
- **Representation**: using coordinates $a_1, \ldots, a_k$

$$x \approx \mu + a_1 u_1 + \cdots + a_k u_k$$



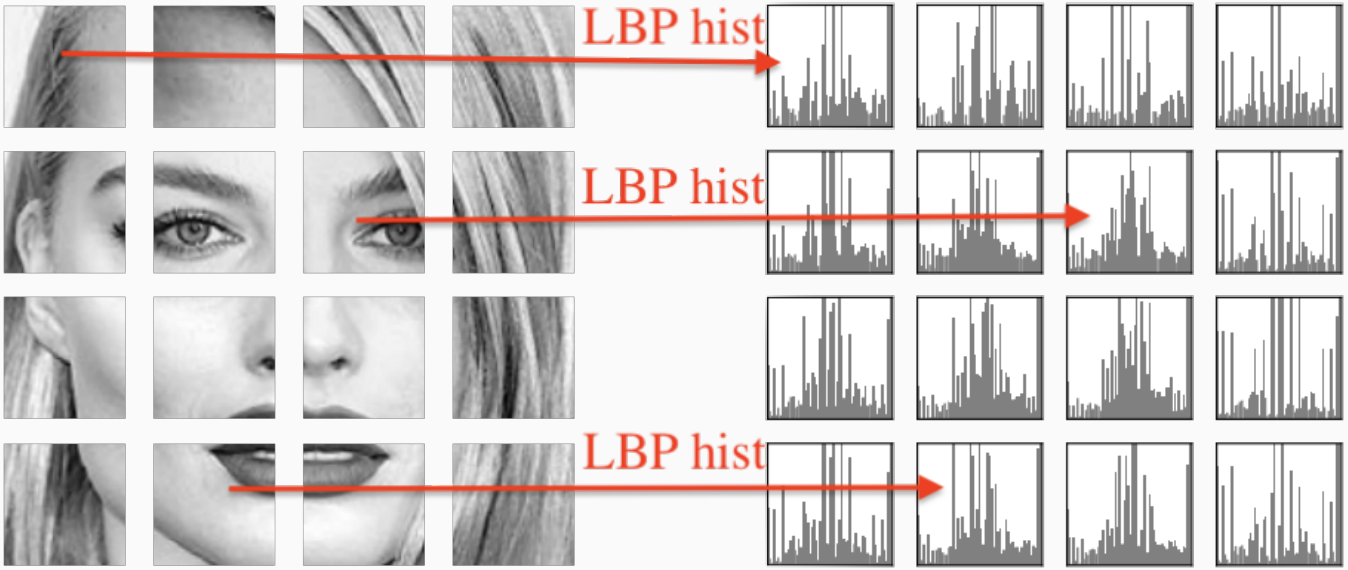
- **Identity**: $x \mapsto (a_1, \ldots, a_k)$ to find closest labeled face in database (e.g. Nearest-Neighbor in k-dim space)

LBP Operator



Binary pattern: $11100001_2 = 225_{10}$

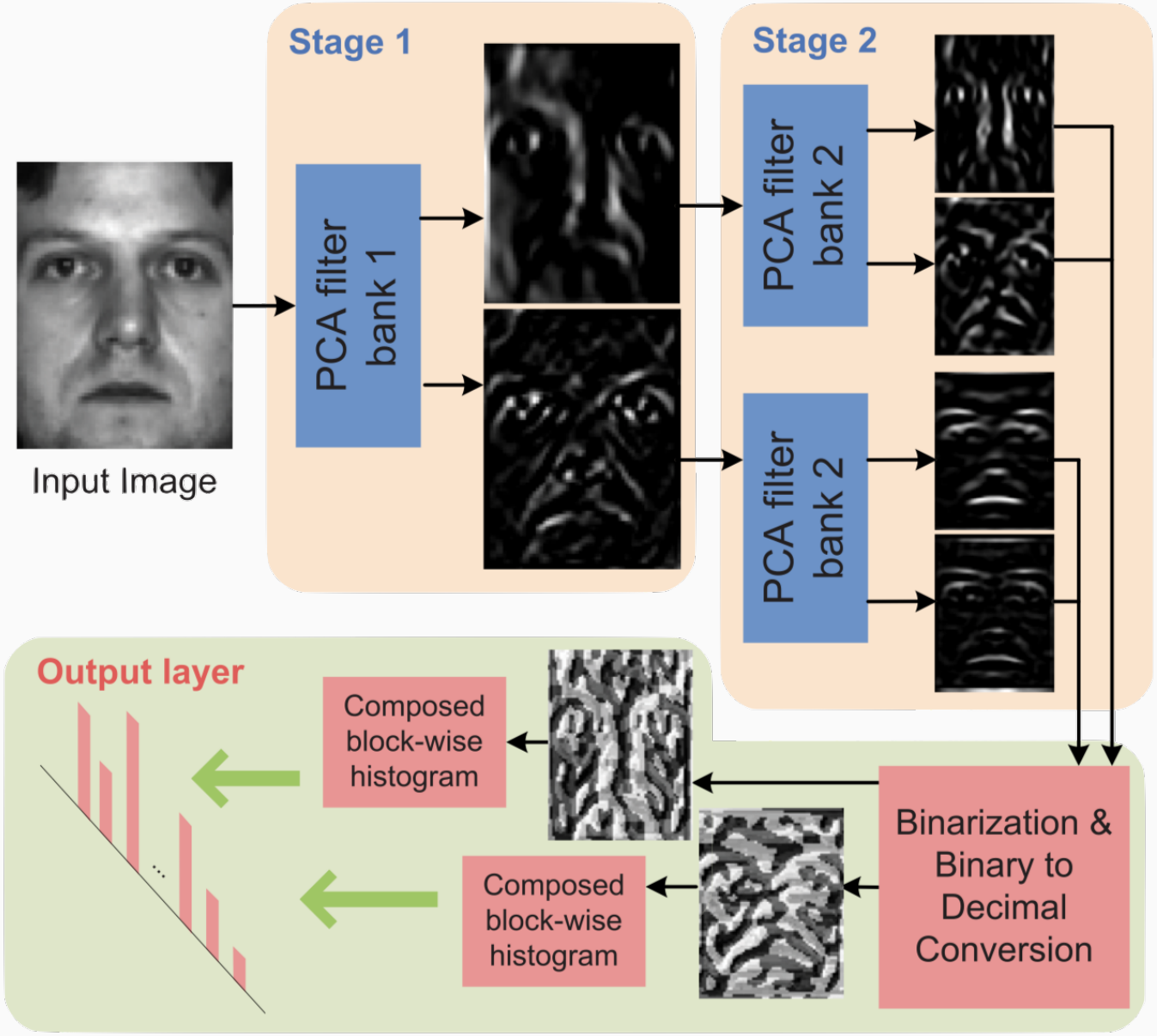
LBP Classification [2006]



- **Histograms** of **LBP descriptors** are extracted from local regions independently
- Difference among **histograms** of **feature vectors** by weighted Chi-square distance:

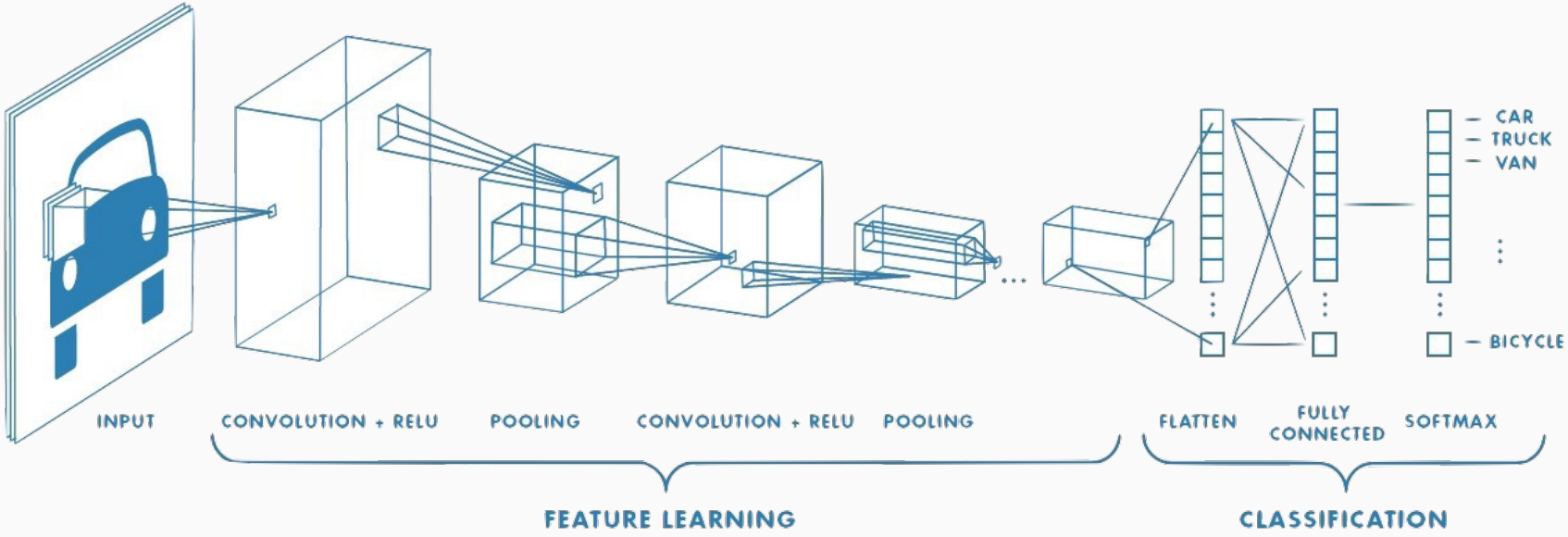
$$\chi^2(\mathbf{a}, \mathbf{b}) = \sum_i \frac{w_i (a_i - b_i)^2}{a_i + b_i}$$

PCANet


Input Image
Stage 1
PCA filter bank 1
Stage 2
PCA filter bank 2
PCA filter bank 2
Output layer
Composed block-wise histogram
Composed block-wise histogram
Binarization & Binary to Decimal Conversion

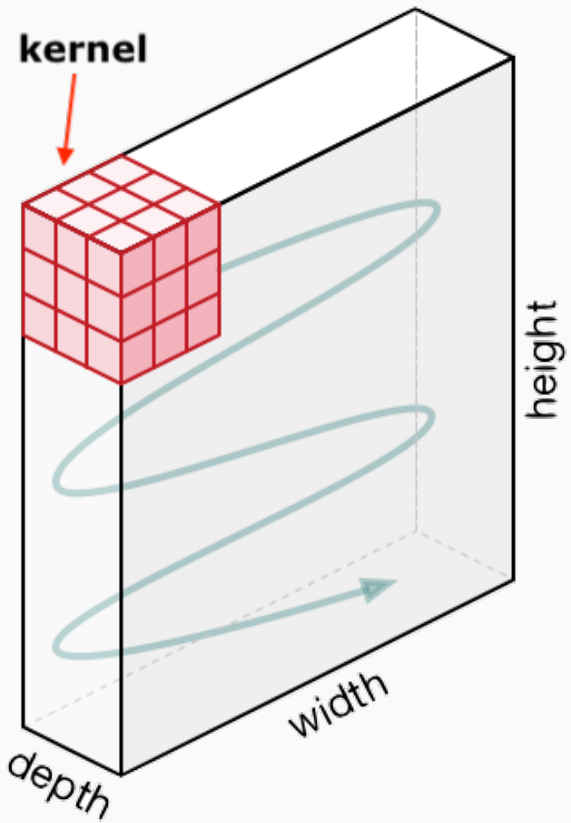
...to Deep Learning (CNNs)

CNN Structure



Convolution Layer

- **Convolution**: kernel shifts + mat multiplications
- **Depth**: kernel depth = number image channels (e.g. RGB)
- **Training**: the kernel weights are updated using the usual back-propagation algorithm (backward pass)
- **Testing**: kernels are used to extract features starting from test data and traversing through all neurons from first to last layer (forward pass)



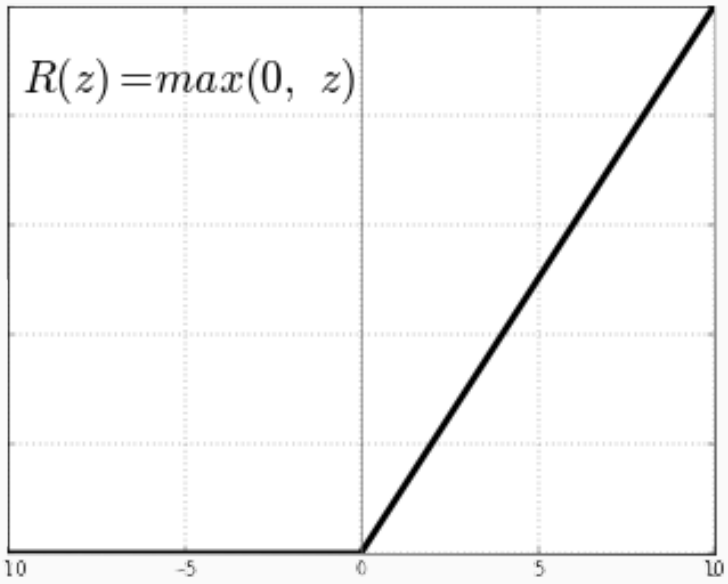
The ReLu (Rectified Linear Unit) Layer

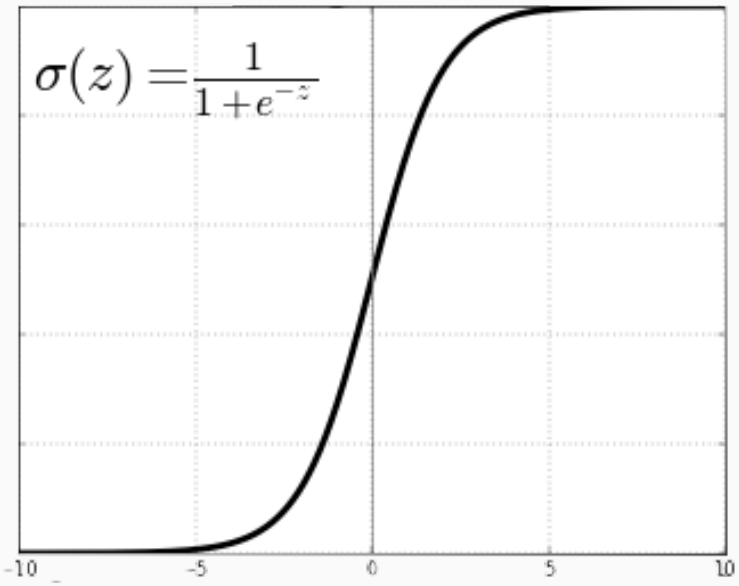
- **ReLu**: activation function for the outputs of the CNN neurons

$$R(z) = \max(0, z)$$

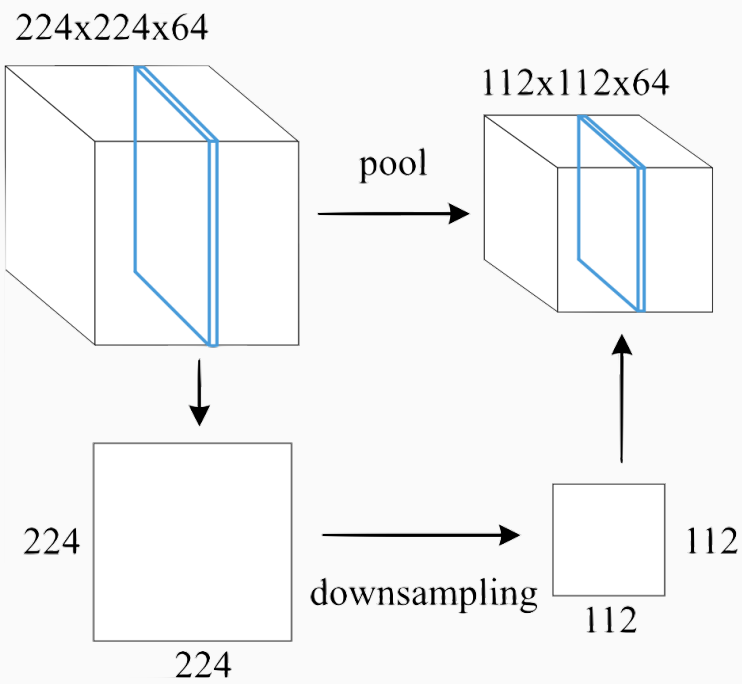
- **Softplus**: differentiable smoothed version of ReLu

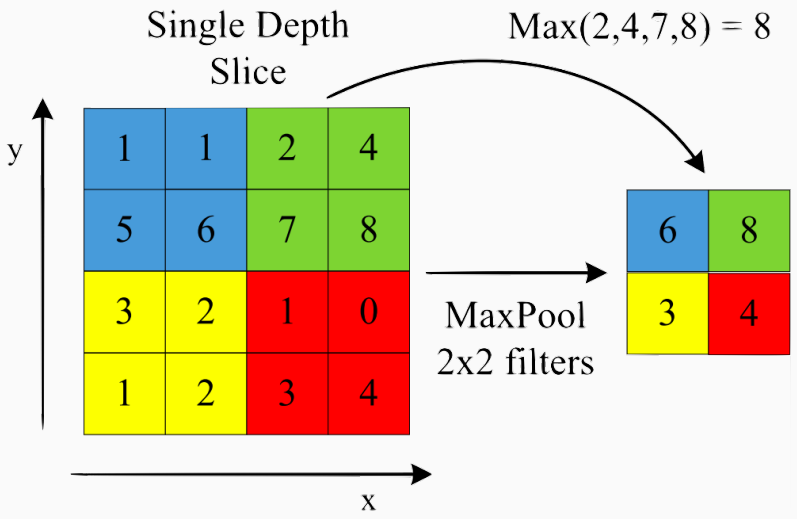
$$S(z) = \ln(1 + e^z), \qquad S'(z) = \sigma(z) = \frac{1}{1 + e^{-z}}$$





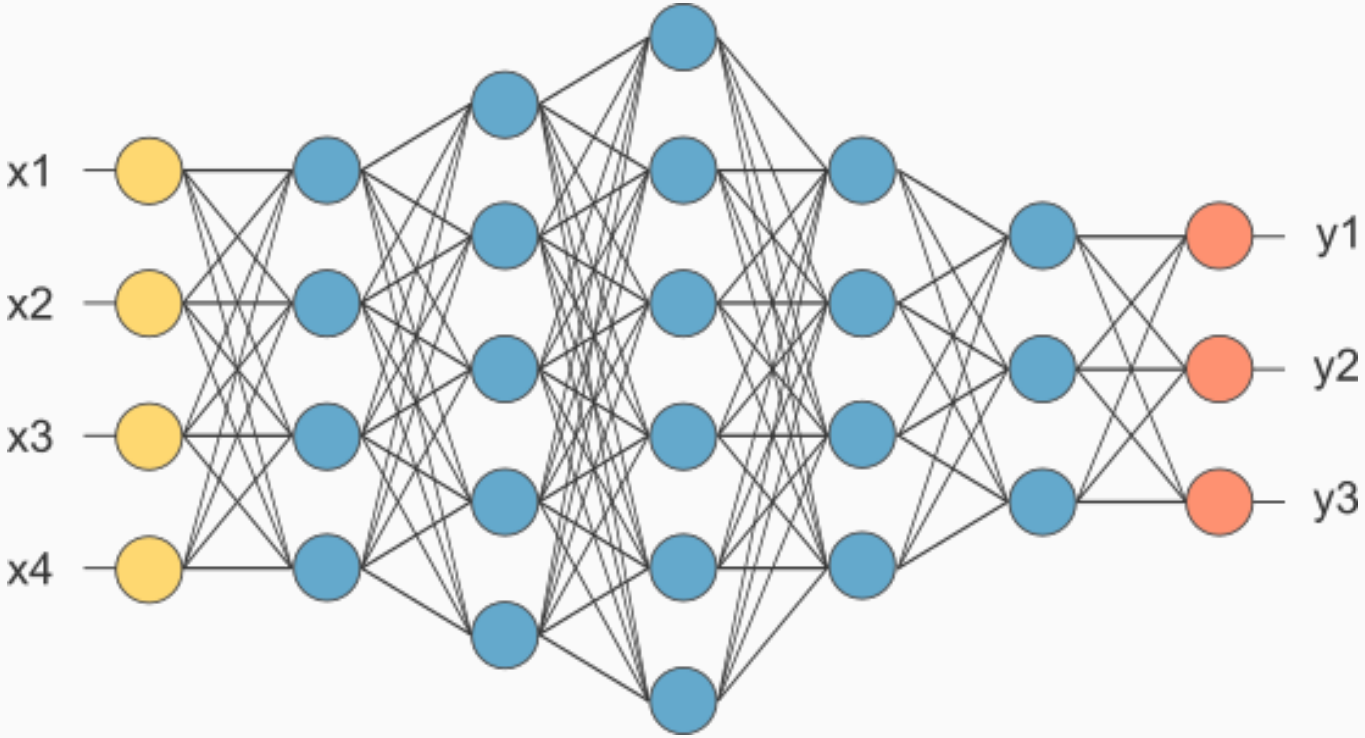
Pooling Layer





- **MAX pooling** takes the largest element from the rectified feature map
- Reduces the dimensionality by subsampling
- Retains the most important information (remove noise)
- Possible choices: {**MAX**, **AVERAGE**, **SUM**} pooling

Fully Connected Layer



- Like ordinary **Neural Networks** (feedfor/back-ward propag.)
- Neurons in a layer are connect. to all neurons in the prev layer
- Computes the **class scores** $\Leftrightarrow$ volume of size $[1 \times 1 \times N]$

SoftMax Loss

- **Goal**: to compress intra-variance and enlarge inter-variance classes

- **Softmax-based** methods usually accept identity labels as supervision: posterior prob of embedding y_i to belong to the identity c is:

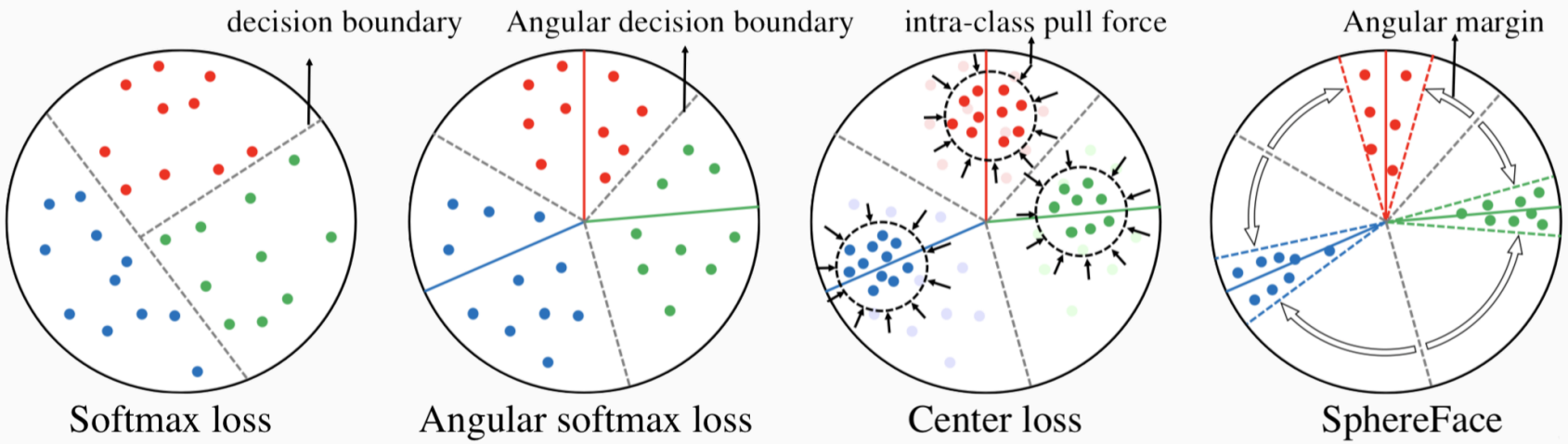
$$p_c(y_i) = \frac{e^{\mathbf{w_c}^T \cdot y_i + b_c}}{\sum_{j=1}^{C} e^{\mathbf{w_j}^T \cdot y_i + b_j}}$$

  $W \in \mathbb{R}^{K \times C}$ matrix mapping to posterior prob, b is the bias

- Given the **identity label** ℓ_i the softmax loss is

$$\mathcal{L}_c(y_i, \ell_i) = -\sum_{c=1}^{C} \mathbf{1}(\ell_i = c) \log p_c(y_i)$$

SoftMax Loss Variant



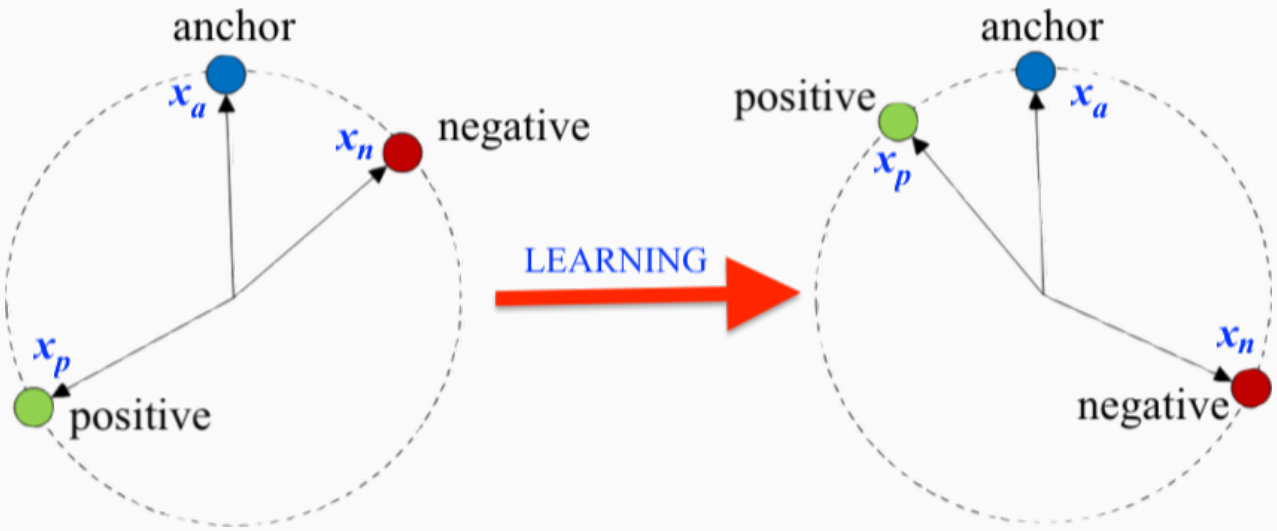
- **Angular SoftMax** uses angular decision boundaries

$$p_c(y_i) = \frac{e^{\|y_i\| \cos(\theta_{i,c})}}{\sum_{j=1}^{C} e^{\|y_i\| \cos(\theta_{j,c})}}$$

- **SphereFace** uses more stringent decision boundaries:

$$p_c(y_i) = \frac{e^{\|y_i\| \cos(m\,\theta_{i,c})}}{e^{\|y_i\| \cos(m\,\theta_{i,c})} + \sum_{j \neq i}^{C} e^{\|y_i\| \cos(\theta_{j,c})}}, \quad m = 1, 2, \ldots$$

Triplet Loss

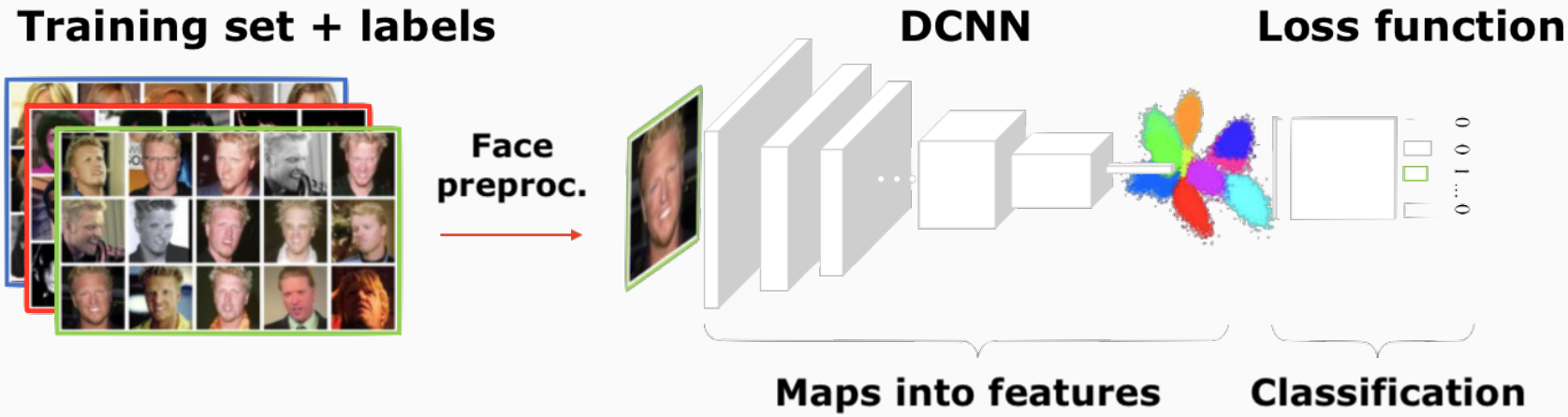


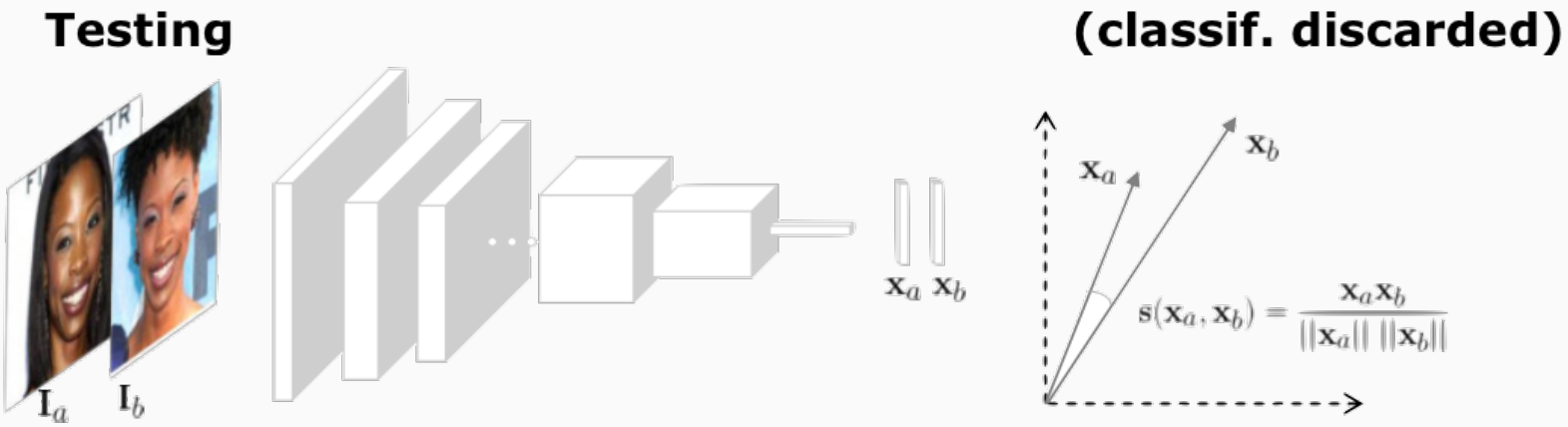
- Separate the distance between positive pairs from the distance between negative pairs by a margin

$$\|f(x_a) - f(x_p)\|_2^2 < \alpha \|f(x_a) - f(x_n)\|_2^2$$

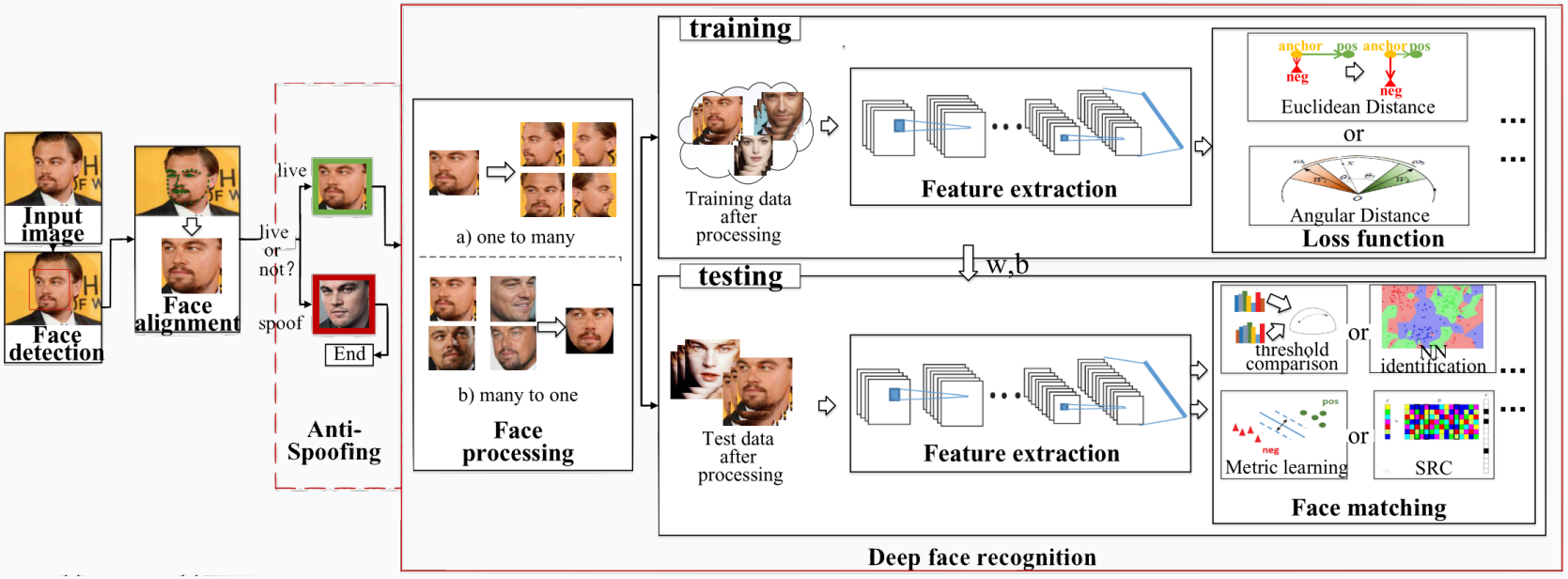
x_a is an anchor image x_p is an image of the same subject
x_n is an image of a different subject α enforces difference

CNN Training & Testing Pipeline



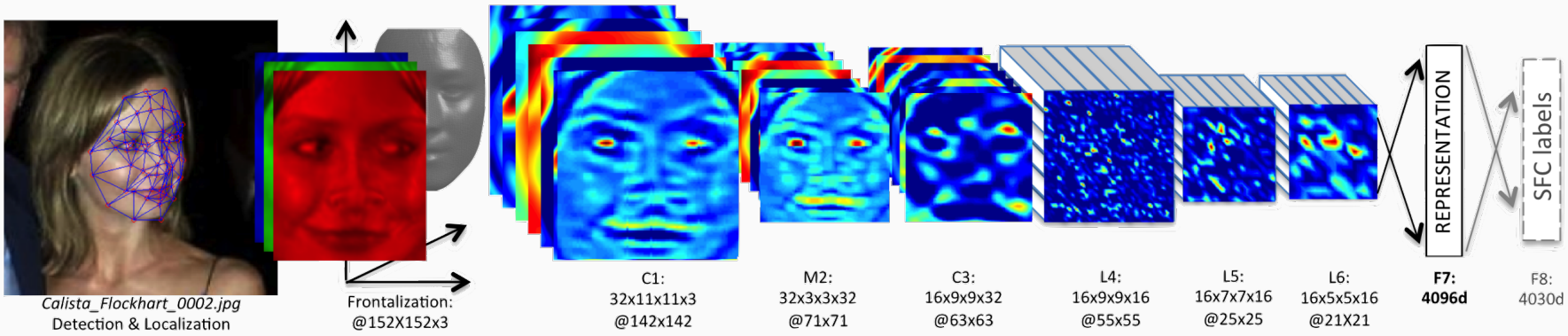


CNN Classification Process

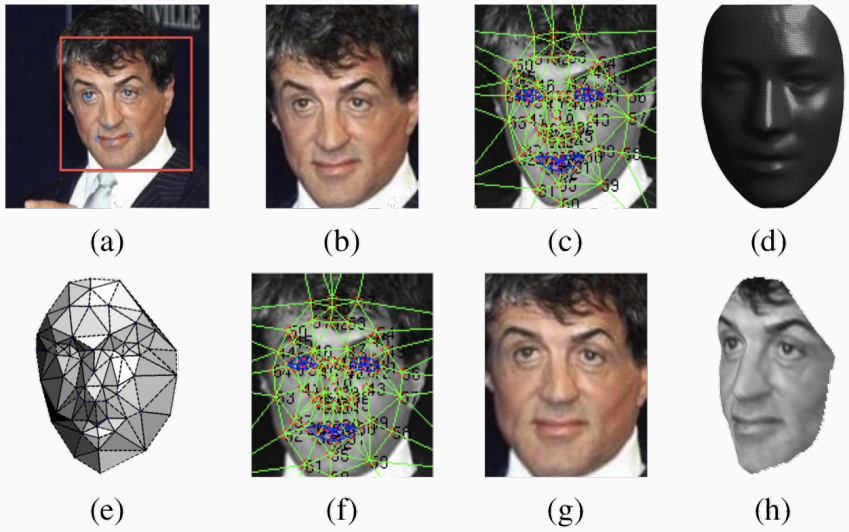


- **Preprocessing**: Detection, alignment, anti-sfoofing to get canonical coordinates & form
- **Feature extraction** in training and testing steps using different architectures and loss functions
- **Face matching** methods are used to do feature classification

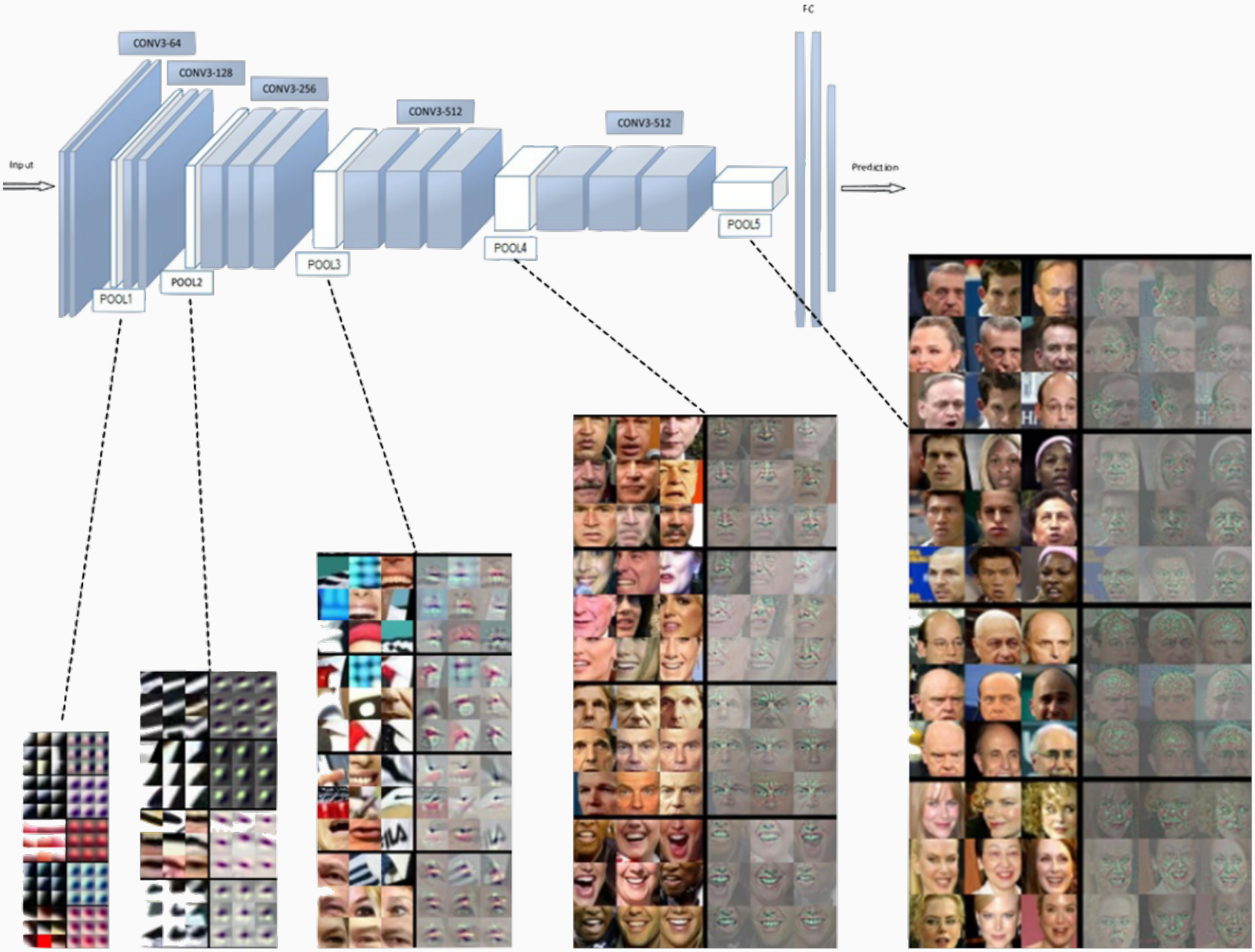
DeepFace CNN [2014]



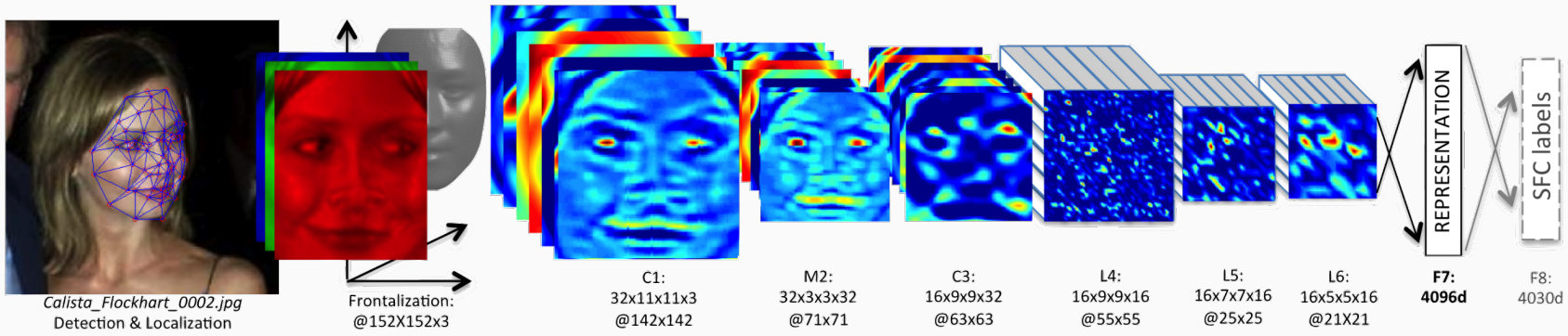
- **3D Frontalization:**
  - (a) detection
  - (b) 2D alignment
  - (c) fiducial points
  - (d) reference 3D shape
  - (e) Triangle visibility
  - (f) 67 fiducial points of 3D model
  - (g) final frontalized crop



DeepFace: Convolutional Levels

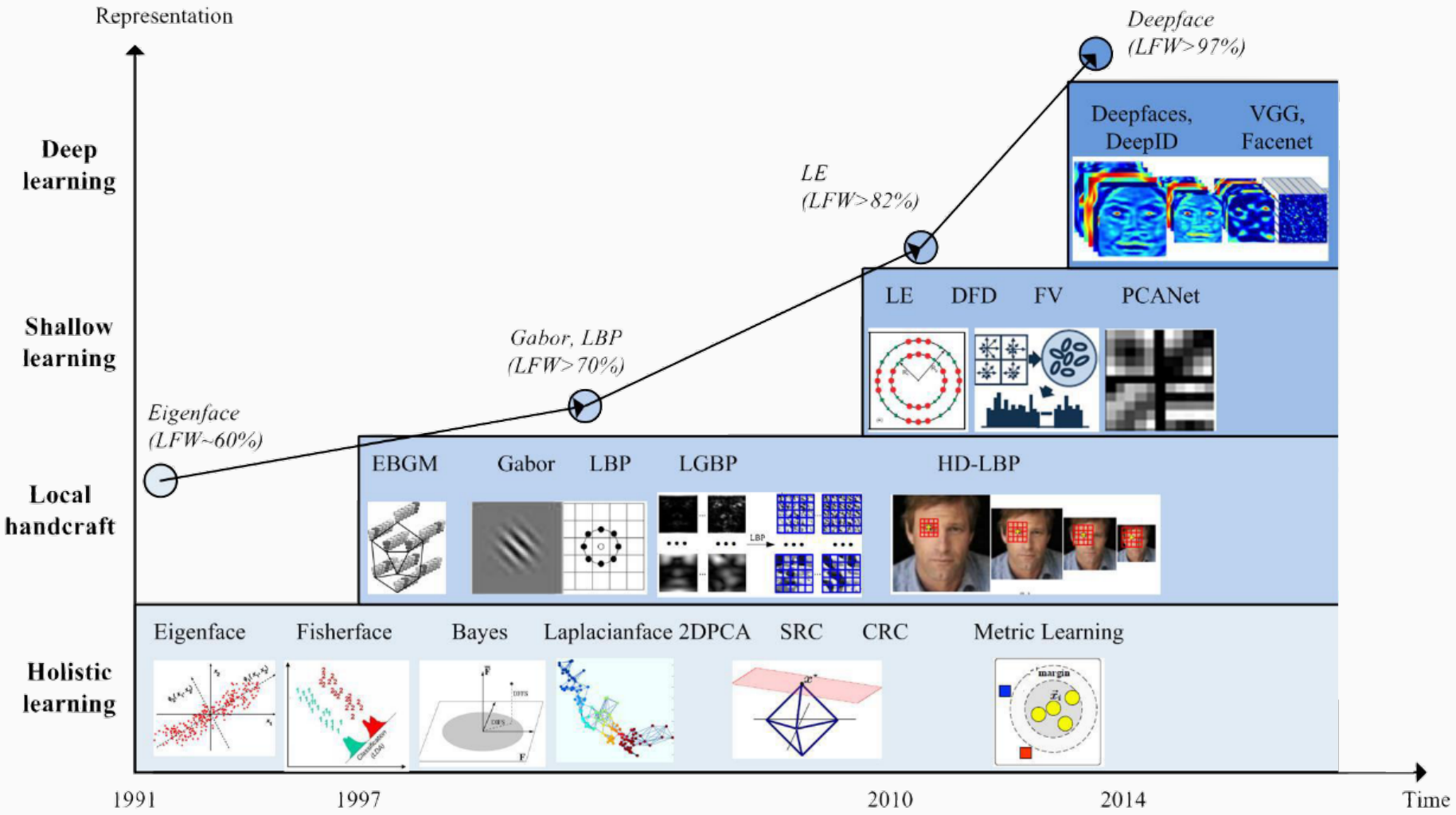


DeepFace: SoftMax classification



- The top two layers (F7 and F8) are **fully connected**: each output unit is connected to all inputs
- The output o_c of the last fully-connected layer is fed to a **C-way softmax**: $p_c = \exp(o_c) / \sum_h \exp(o_h)$
- Trained on ∼4 million facial images, 4400 individuals. Verification is done by comparing features for two faces
- DeepFace achieved the state-of-the-art accuracy on **LFW** approaching human performance for the first time... **DeepFace**: **97.35%** vs. **Human**: **97.53%**

FR Performances on LFW over Past Years



CNN Performances on LFW

Method	Year	Arch	Training Set	Accuracy
DeepFace	2014	Alexnet	Facebook (4.4M,4K)	97.35
DeepID3	2015	VGGNet-10	CelebFaces+ (0.2M,10K)	99.53
FaceNet	2015	GoogleNet-24	Google (500M,10M)	99.63
VGGface	2015	VGGNet-16	VGGface (2.6M,2.6K)	98.95
L2-softmax	2017	ResNet-101	MS-Celeb-1M (3.7M,58K)	99.78
CoCo loss	2017	ResNet-28	MS-Celeb-1M (3M,80K)	99.86
Arcface	2018	ResNet-100	MS-Celeb-1M (3.8M,85K)	99.83

Conclusion and future work

- We all understand that this is an ongoing development, which is far from its end
- Remaining challenges defined by non-saturated benchmark datasets: very large number of candidates, low/one-shot and large pose-variance, highly noisy images
- Understanding deep face recognition: what is the "identity capacity" of a deep representation... What's the role of augmented images? (GAN, generative adversarial are power tools... how can they help?)
- Pursuit of extreme accuracy and efficiency: Surveillance or financial identity verification, require high matching accuracy... It is still a big challenge even with deep learning on massive training data

Thanks!

Bibliography

1. M. Turk and A. Pentland. Eigenfaces for recognition. Journal of cognitive neuroscience, vol. 3, no. 1, pp. 71–86, 1991.
2. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In CVPR, pages 1701–1708, 2014
3. S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. SIAM J. Sci. Comp., 20(1):33–61, 1998
4. R. Gross, I. Matthews, J. F. Cohn, T. Kanade, S. Baker. Multi-PIE. Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition, 2008
5. Cuculo, V.; D'Amelio, A.; Grossi, G.; Lanzarotti, R.; Lin, J. Robust Single-Sample Face Recognition by Sparsity-Driven Sub-Dictionary Learning Using Deep Features. Sensors 2019
6. Yang, M., Wang, X., Zeng, G., Shen, L.: Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person. Pattern Recognition, 2017