Human Face Recognition: Learning from Biological Deep Networks

Massimo Tistarelli

Computer Vision Laboratory

University of Sassari – Italy

tista@uniss.it

Credits

- From the laboratory staff:

Linda Brodo
Marinella Cadoni
Filippo Casu
Massimo Gessa
Enrico Grosso
Souad Khellat Khiel
Andrea Lagorio
Ludovica Lorusso
Gianluca Masala
Norman Poh (past visiting)
Luca Pulina
Ajita Rattani
Elif Surer
Yunlian Sun
Daksha Yadav (past visiting)
Yu Guan (past visiting)
Marcos Ortega Hortas (past visiting)
Albert Ali Salah (past visiting)

Credits

- …and other labs:

Manuele Bicego – University of Verona
Rama Chellappa – University of Maryland
Anil Jain – Michigan State University
Alice O'Toole – University of Texas at Dallas
Chang-Tsun Li – University of Warwick
Jonathon Phillips – NIST
Norman Poh – University of Surrey

IC1106 - Integrating Biometrics and Forensics for the Digital Age

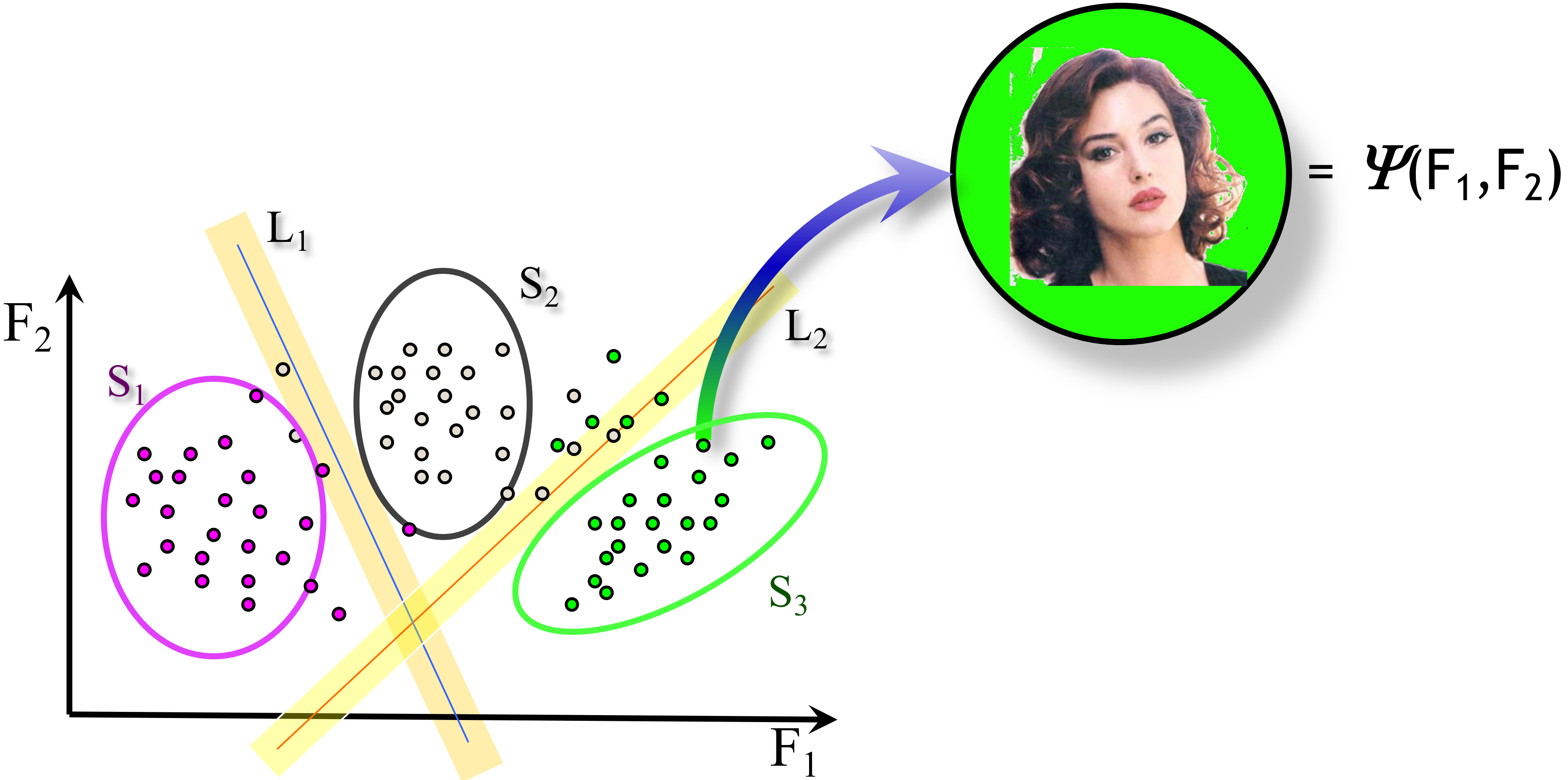
Multi-Foresee

IDENTITY

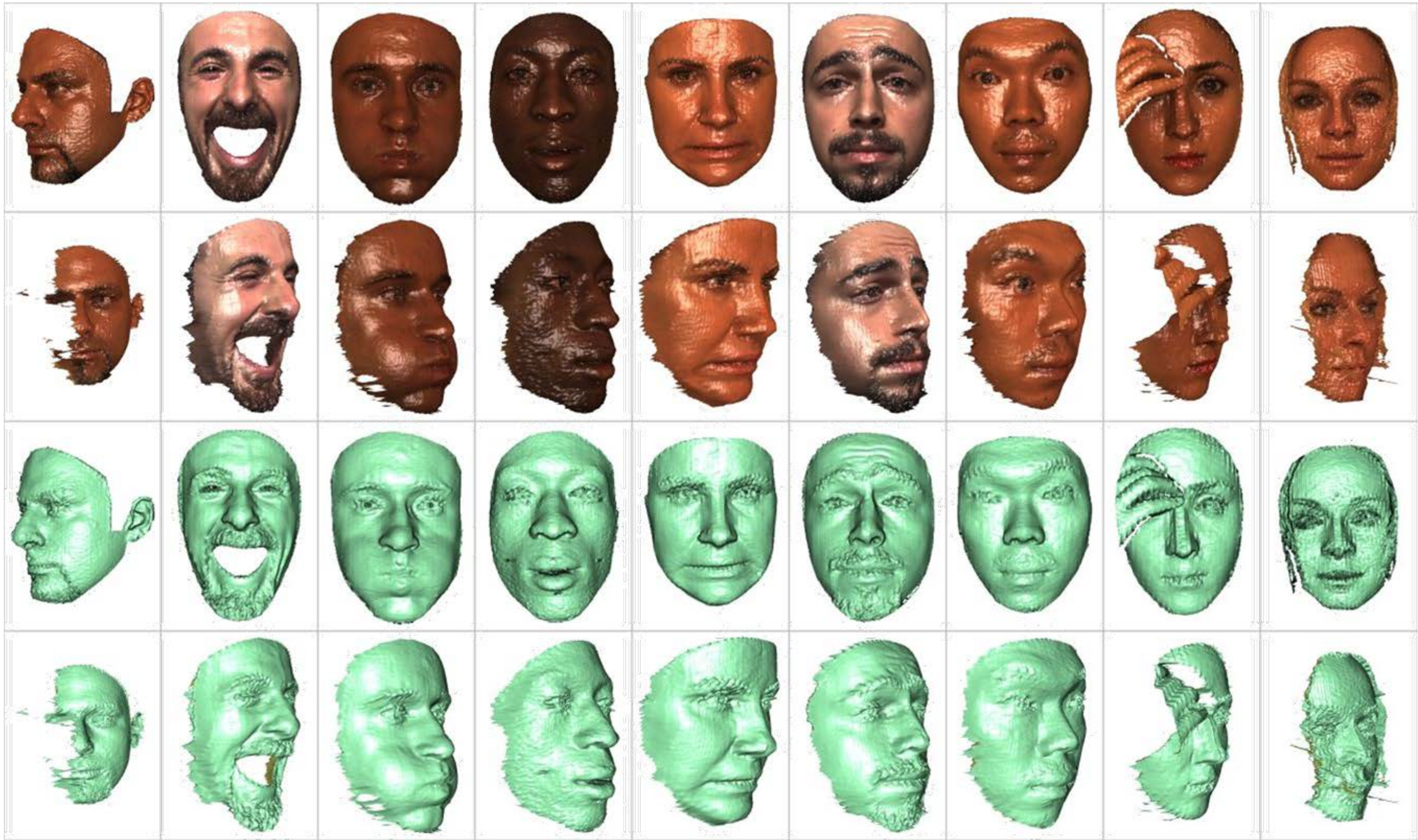
Computer Vision Enabled Multimedia Forensics and People Identification

Face Recognition

A class (***identity***) separation problem



Face shape and texture



A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, "**Bosphorus Database for 3D Face Analysis**", The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008) Roskilde University, Denmark, May 2008.

Visual challenges

Visual challenges

Visual challenges

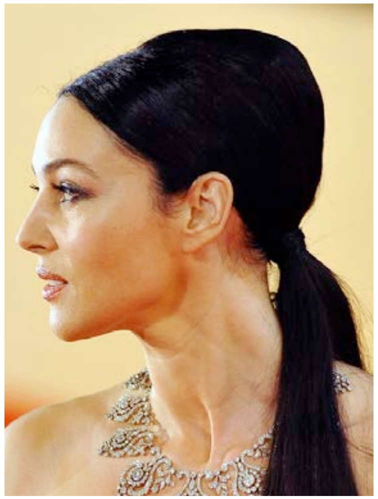
A – Aging

P – Pose

I – Illumination

E - Expression



An ill-posed problem

Vision Lab

An inverse problem is *well-posed* in the sense of Hadamard when:

1) a ***unique*** solution exists and
2) it depends ***continuously*** upon the data.

J. Hadamard, "**Sur les problemes aux derivees partielles et leur signification physique**". In: Princeton University Bulletin, 1902, 49–52.

An ill-posed problem

Two adverse conditions:

1) **Noise** in the data (many sources)
2) **Dimensionality** of the data (from 4D to 2D)

Solution: Regularization

A.N. Tikhonov, "**On the stability of inverse problems**". Doklady Acad. Sci. USSR 39 (1943), 176–179.

A.N. Tikhonov, "**On the solution of ill-posed problems and the method of regularization**". Dokl. Akad. Nauk SSSR 151(3) (1963), 501–4.

A.N. Tikhonov, "**On the regularization of ill-posed problems**". Dokl. Akad. Nauk SSSR 153(1) (1963), 49–52 (in Russian).

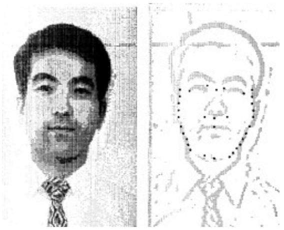
A. N. Tikhonov and V. Ya. Arsenin, "**Solutions of Ill-Posed Problems**". Wiley, New York, 1977.

Face recognition milestones

1964

Woodrow Bledsoe

Face recognition



1973

Takeo Kanade

First FR thesis

1991

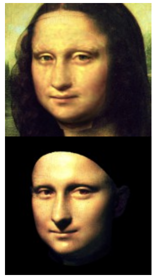
Turk & Pentland

Eigenface

1997

Belhumeur et al.

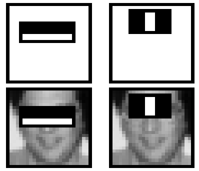
Fisherface



1999

Blanz & Vetter

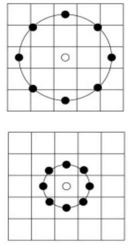
Morphable face



2001

Viola & Jones

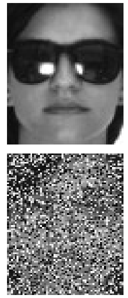
Face detector



2006

Ahonen et al.

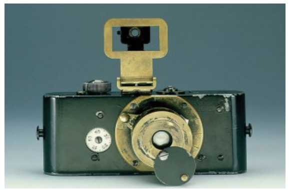
LBP



2009

Wright et al.

Sparse rep.



1915

35mm still camera

1991

Kodak

Digital camera

1024p

1990s

Surveillance camera

480p @ 30fps

2000

Sharp

First camera phone

320p

2010-2013

Wearable camera

480P @ 30fps

Google Glass

720p @30fps

April 2013

Samsung Galaxy S4

1080p @ 30fps

Bledsoe, W. W. 1964. The Model Method in Facial Recognition, TR PRI 15, Panoramic Research, Inc., California.

Takeo Kanade, Picture Processing System by Computer Complex and Recognition of Human Faces, Kyoto Univ., 1973.

M. Turk and A. Pentland, Eigenfaces for recognition. Journal of Cognitive Neuroscience 3 (1): 71–86, 1991.

Belhumeur, P.N. et al., Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, PAMI, 19-7, 1997.

V. Blanz and T. Vetter, A morphable model for the synthesis of 3D faces, SIGGRAPH 1999.

http://photodoto.com/camera-history-timeline/

Viola, Jones: Robust Real-time Object Detection, IJCV 2001.

Ahonen,et al. Face Description with Local Binary Patterns: Application to Face Recognition, PAMI, 2006.

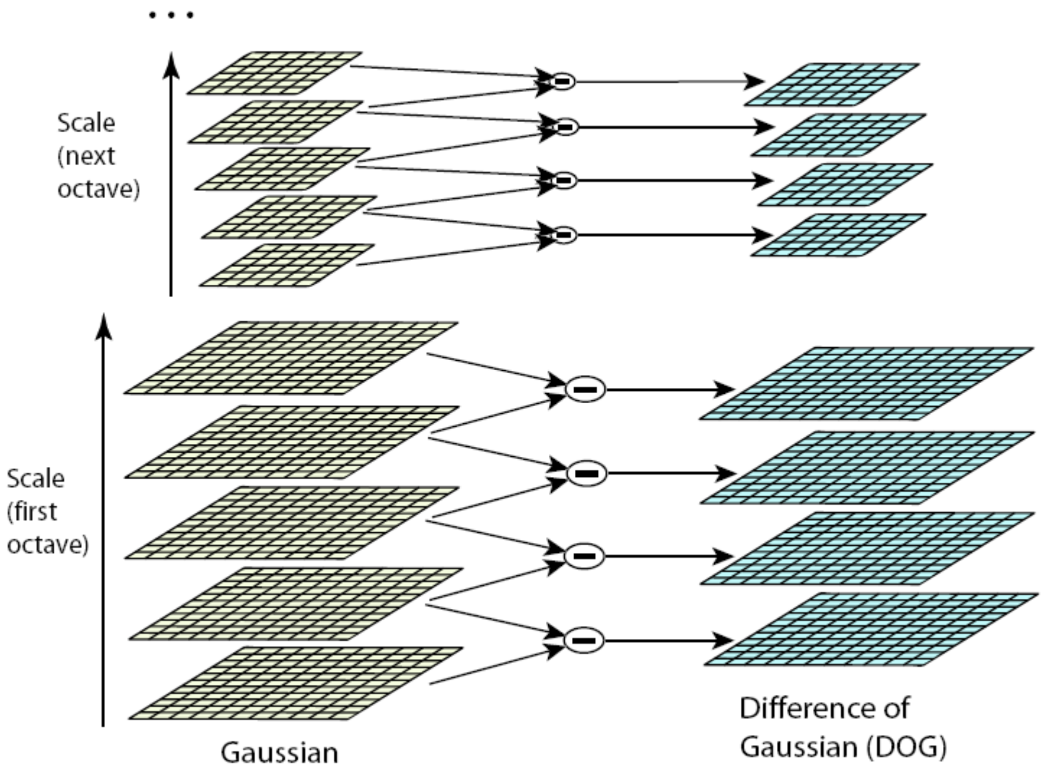
J. Wright et al. Robust Face Recognition via Sparse Representation, PAMI, 31-2, 2009.

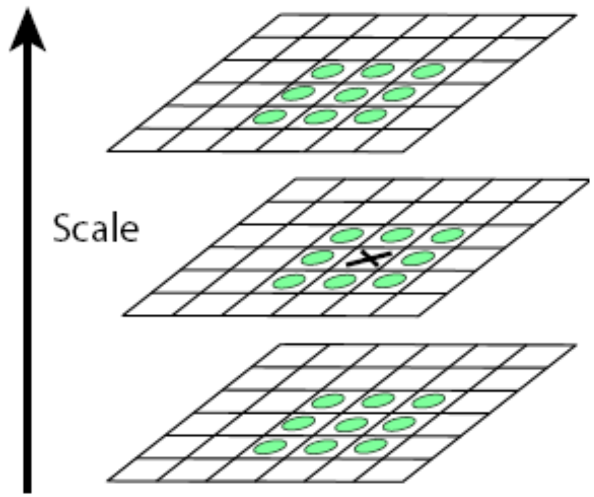
http://static7.businessinsider.com/image/4d013ea7cadcbb7033010000/looxcie-video-camera.jpg

Scale Invariant Features

$$\mathbf{D}(x, y, \sigma, k) = (\mathbf{G}(x, y, k\sigma) - \mathbf{G}(x, y, \sigma)) * \mathbf{I}(x, y)$$

$$\mathbf{D}(x, y, \sigma, k) = \mathbf{L}(x, y, k\sigma) - \mathbf{L}(x, y, \sigma)$$

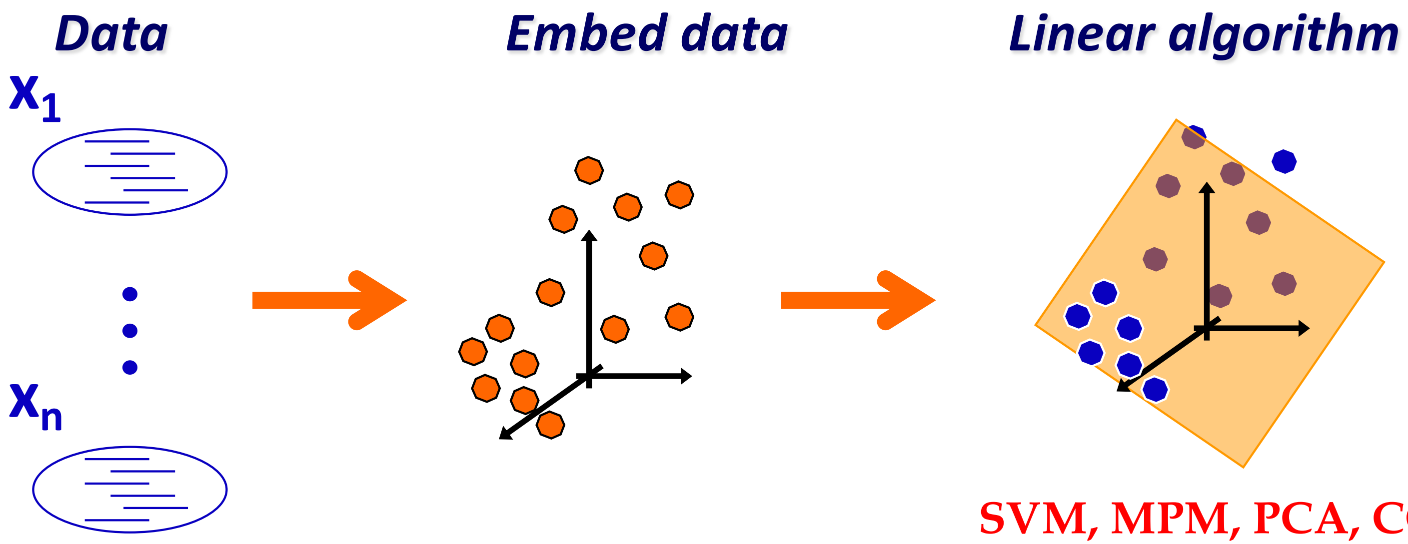




G. Lowe, “Object recognition from local scale invariant features”, International Conference on Computer Vision , 1999.

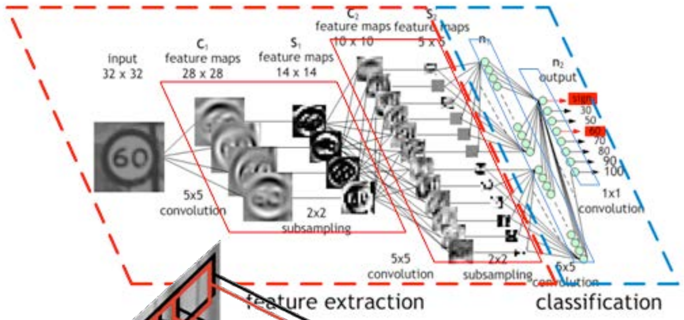
Kernel methods

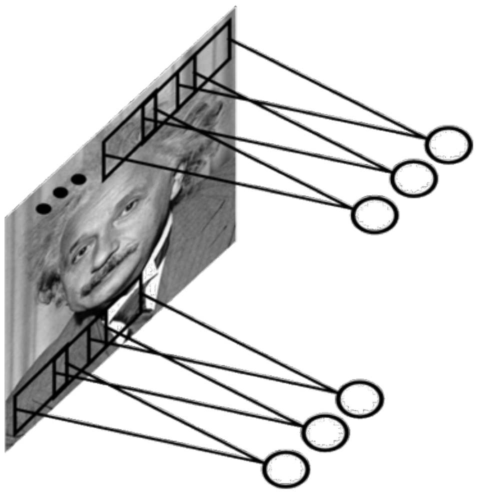
- **K-PCA**; **K-ICA**; **K-LDA**… (B. Schölkopf et al. 1998)
- Are all **variations of existing face-space representations**. The transformation is mediated by a **kernel function** such as Gaussian, polinomial, sigmoid and Radial Basis Functions
- More **robust to noise and discretization** - Better separation of classes
- Related to the general ***Learning Theory***



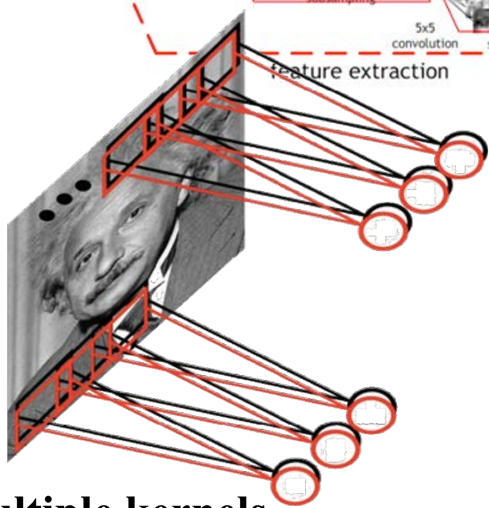
SVM, MPM, PCA, CCA, FDA…

Convolutional Neural Networks

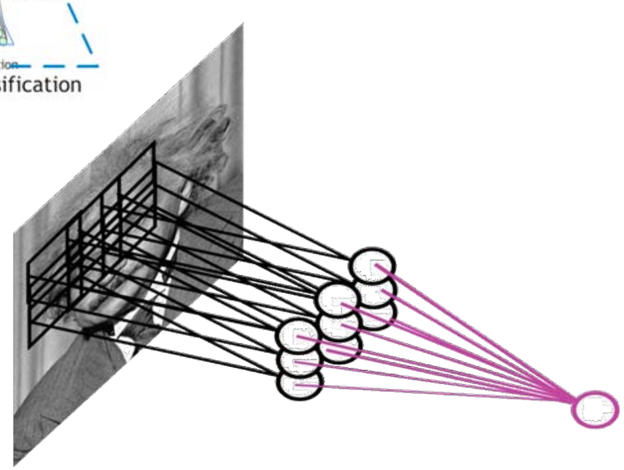




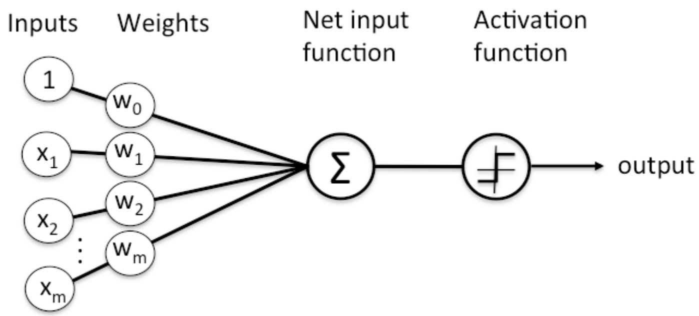
Single kernel Convolution



Multiple kernels Convolution



Spatial Pooling

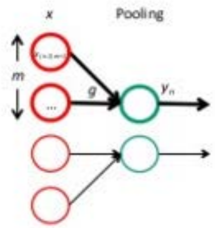


Let m be the size of pooling region, x be the input, and y be the output of the pooling layer. subsample$(f, g)[n]$ denotes the n-th element of subsample(f, g).

$$y_n = \text{subsample}(x,g)[n] = g\left(x_{(n-1)m+1:nm}\right)$$

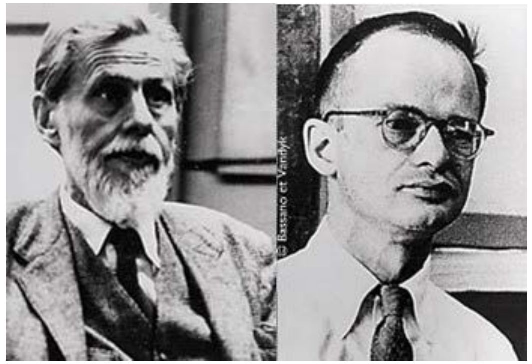
$$y = \text{subsample}(x,g) = [y_n]$$

$$g(x) = \begin{cases} \dfrac{\sum_{k=1}^{m} x_k}{m}, \quad \dfrac{\partial g}{\partial x} = \dfrac{1}{m} & \text{mean pooling} \\ \max(x), \quad \dfrac{\partial g}{\partial x_i} = \begin{cases} 1 \text{ if } x_i = \max(x) \\ 0 \text{ otherwise} \end{cases} & \text{max pooling} \\ \|x\|_p = \left(\sum_{k=1}^{m} |x_k|^p\right)^{1/p}, \quad \dfrac{\partial g}{\partial x_i} = \left(\sum_{k=1}^{m} |x_k|^p\right)^{1/p-1} |x_i|^{p-1} & L^p \text{ pooling} \\ \text{or any other differentiable } \mathbf{R}^m \to \mathbf{R} \text{ functions} & \end{cases}$$

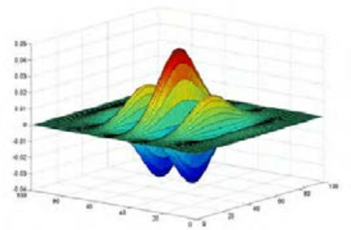


Why CNNs… today?

Neural networks have been proposed since the early '40s:
In **1943**, neurophysiologist **Warren McCulloch** and mathematician **Walter Pitts** wrote a paper on how neurons might work. **They modeled a simple neural network using electrical circuits**.



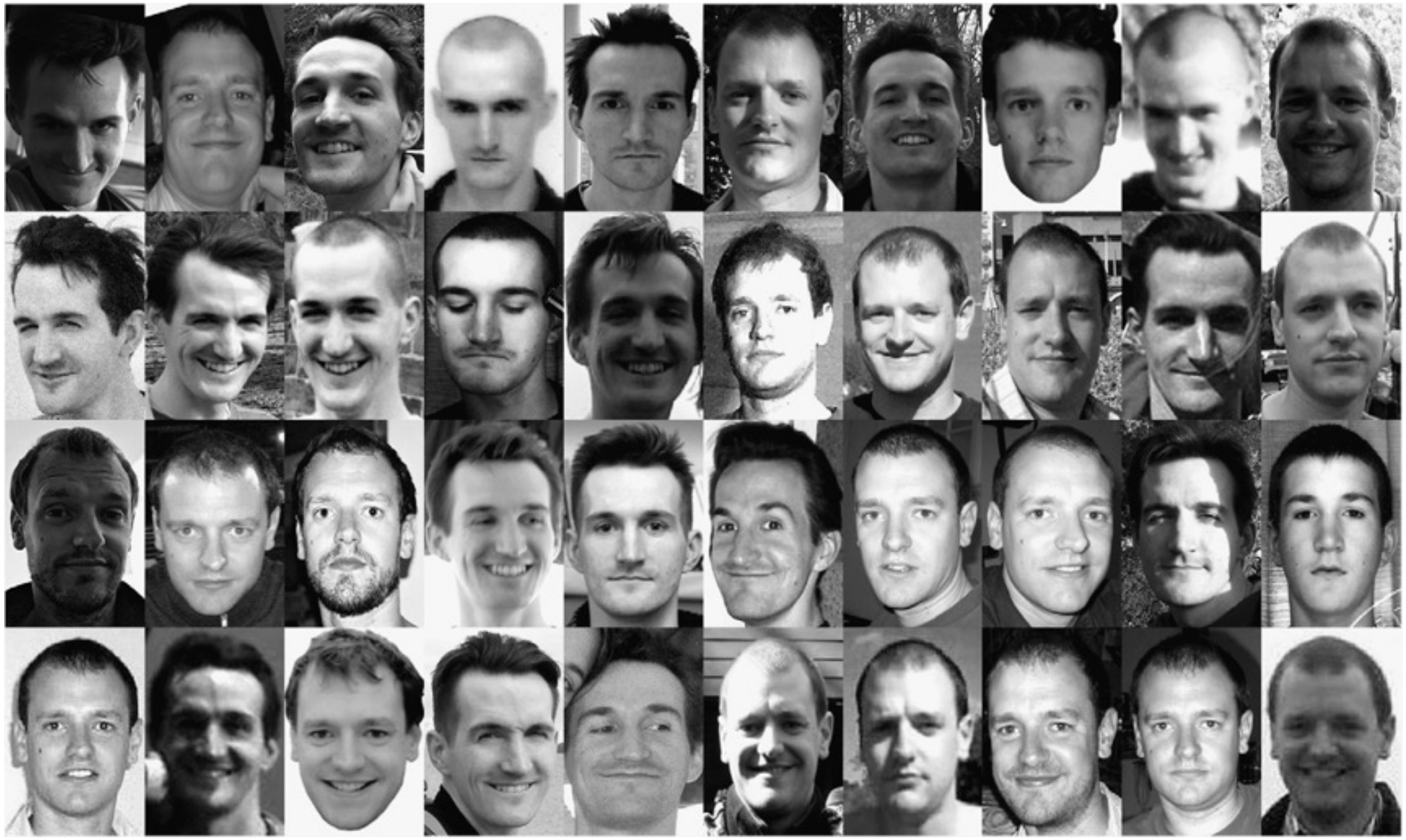
Convolutions or **digital filtering** have been used since the 50's for several vision tasks, including face recognition.



The progress in the **Theory of Learning** and of **computing power** allowed to implement more efficient and complex neural networks with multiple hidden layers…

Human Performance

- How do <u>humans</u> perform in recognizing faces?

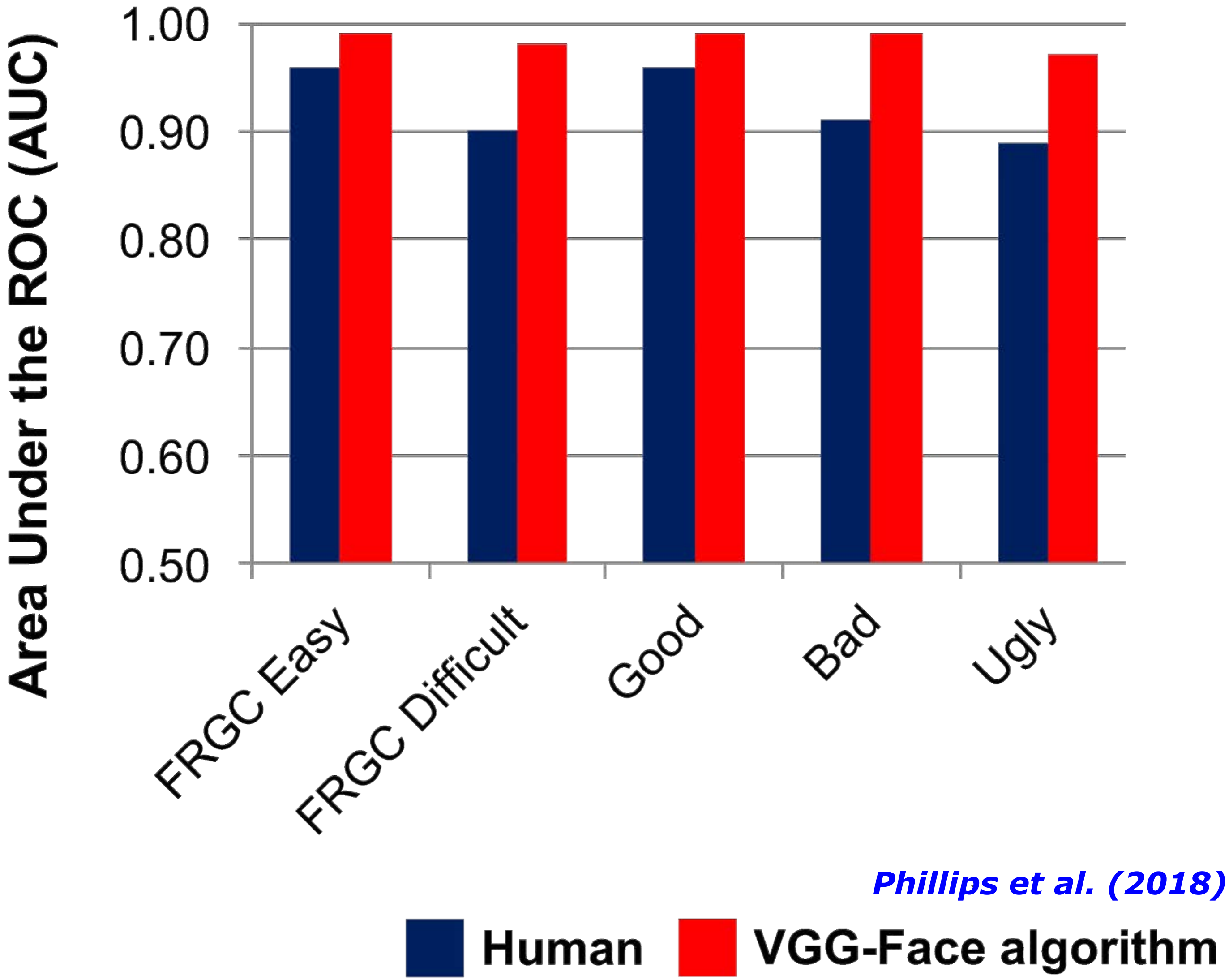


Jenkins, White, Burton (2011)

CNN Performance

Vision Lab

- How do <u>machines</u> perform in recognizing faces?

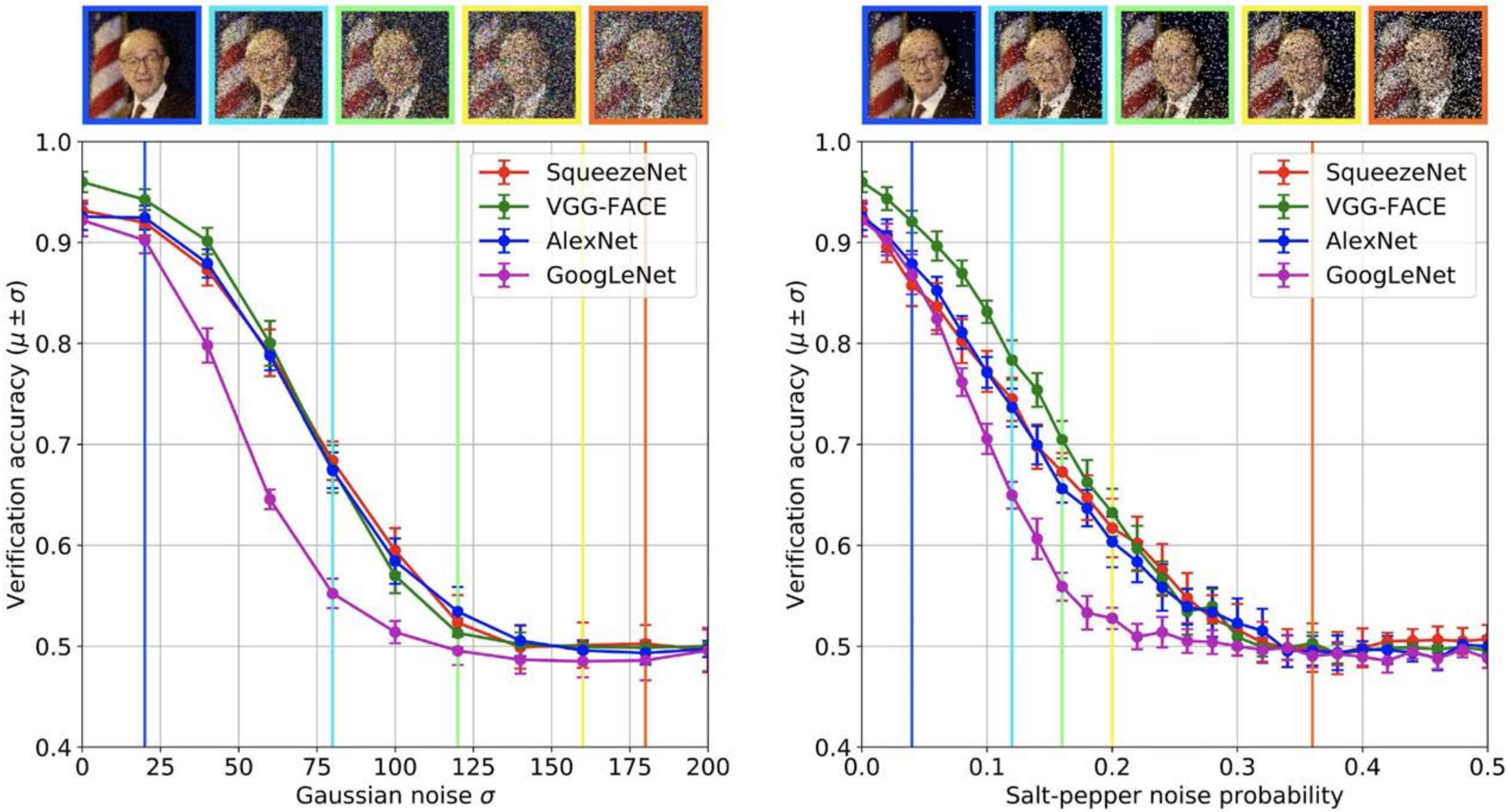


Phillips et al. (2018)

CNN Performance

Vision Lab

- However, we're not done yet…

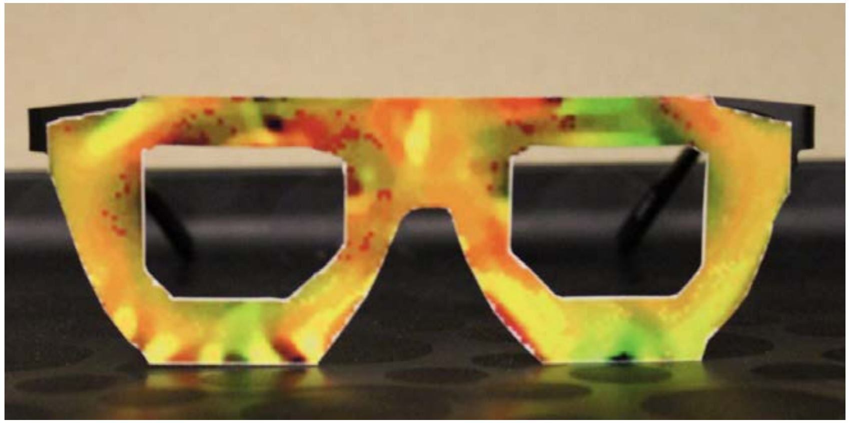


K. Grm , V. Štruc, A. Artiges, M. Caron, H. K. Ekenel, "**Strengths and weaknesses of deep learning models for face recognition against image degradations**" IET Biometrics, 7(1):81-89, 2018

CNN Performance

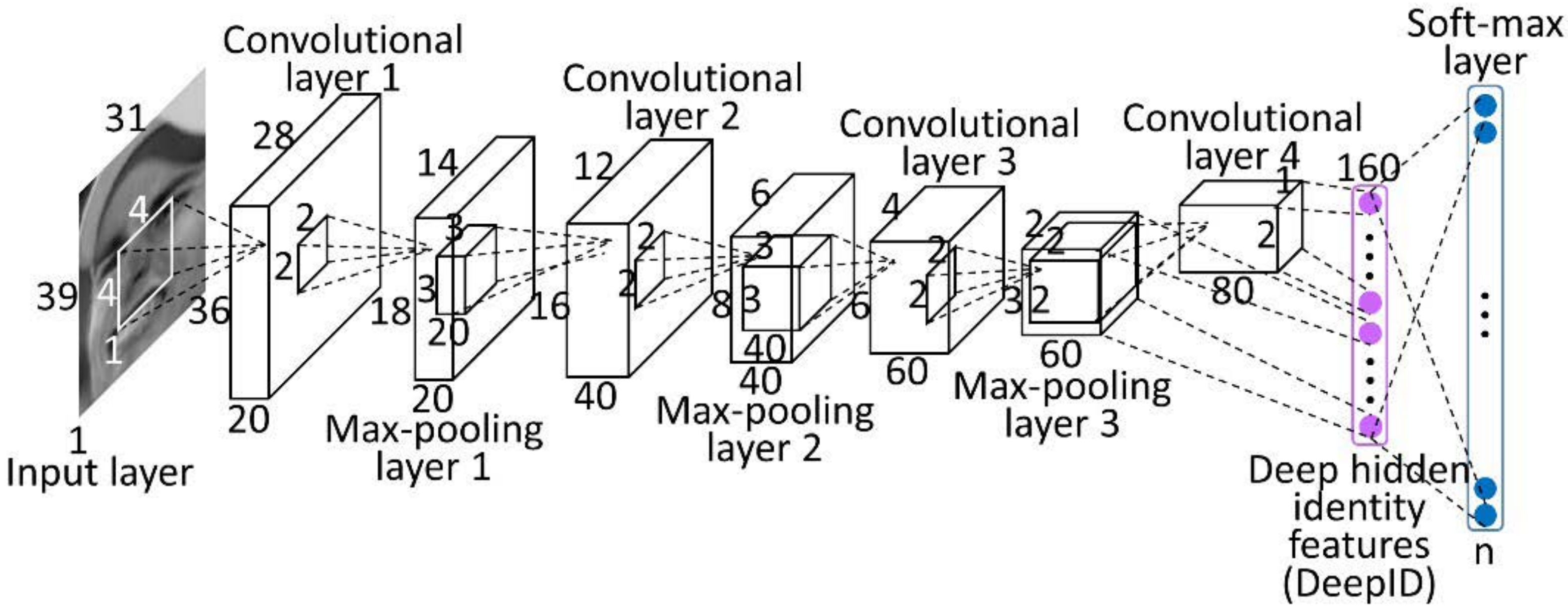
Vision Lab

- **The “magic glasses”**



M. Sharif , S. Bhagavatula, L. Bauer, M. K. Reiter, "**Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition**", CCS’16 October 24-28, 2016, Vienna, Austria

The "curse of training"

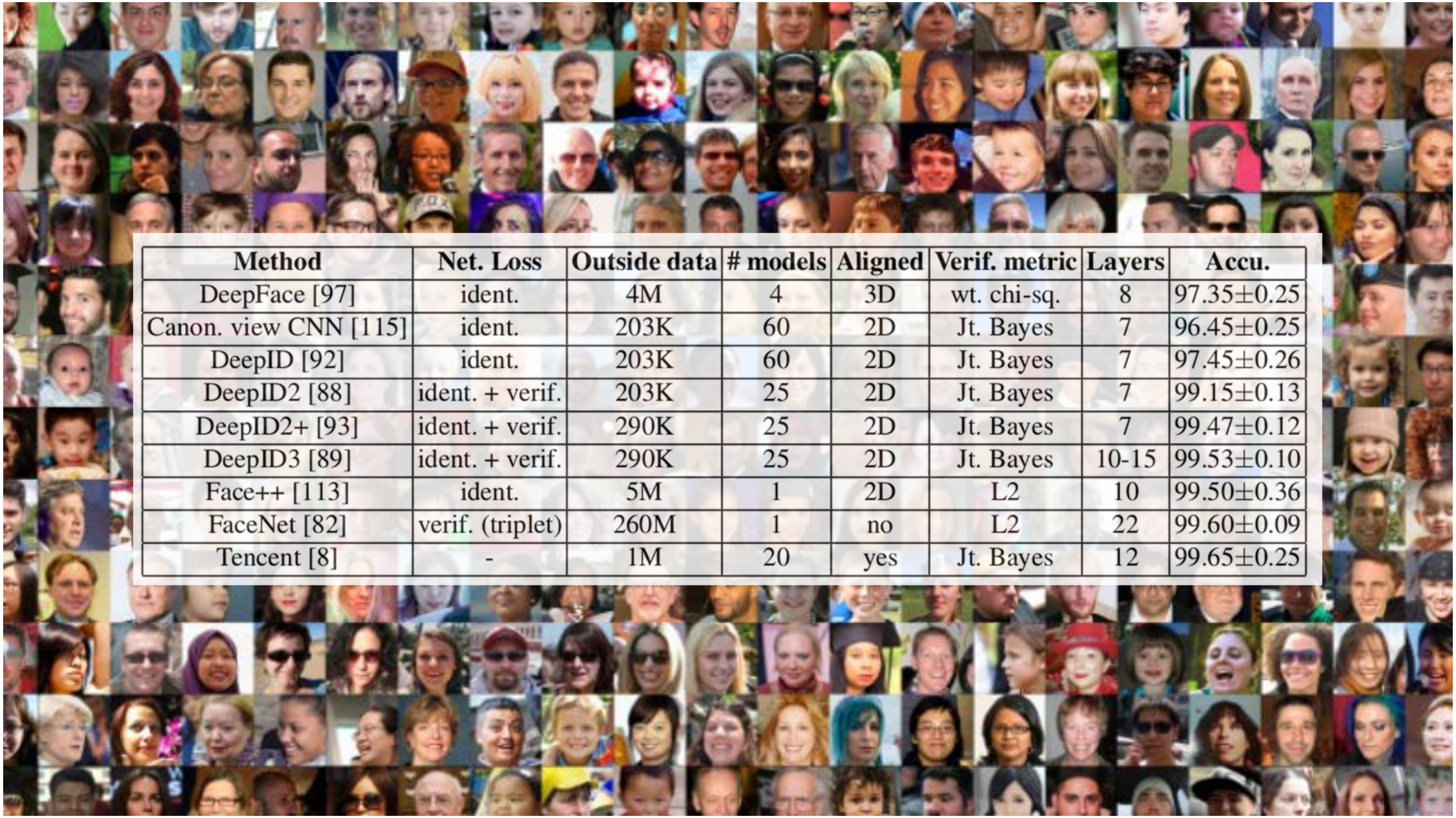


A deep CNN is used to extract a feature vector with relatively high dimension. The network can be supervised by multiclass loss and verification loss

PCA, Joint Bayesian or metric-learning methods are used to learn a more efficient low dimensional representation

The amount of training data can range from **100K** up to **260M**

The "curse of training"



Method	Net. Loss	Outside data	# models	Aligned	Verif. metric	Layers	Accu.
DeepFace [97]	ident.	4M	4	3D	wt. chi-sq.	8	97.35±0.25
Canon. view CNN [115]	ident.	203K	60	2D	Jt. Bayes	7	96.45±0.25
DeepID [92]	ident.	203K	60	2D	Jt. Bayes	7	97.45±0.26
DeepID2 [88]	ident. + verif.	203K	25	2D	Jt. Bayes	7	99.15±0.13
DeepID2+ [93]	ident. + verif.	290K	25	2D	Jt. Bayes	7	99.47±0.12
DeepID3 [89]	ident. + verif.	290K	25	2D	Jt. Bayes	10-15	99.53±0.10
Face++ [113]	ident.	5M	1	2D	L2	10	99.50±0.36
FaceNet [82]	verif. (triplet)	260M	1	no	L2	22	99.60±0.09
Tencent [8]	-	1M	20	yes	Jt. Bayes	12	99.65±0.25

Is this *you*?

Face recognition under ban

CNN BUSINESS

San Francisco just banned facial-recognition technology

By Rachel Metz, CNN Business

Updated 2315 GMT (0715 HKT) May 14, 2019

Microsoft CEO says self regulation needed with new technologies

US Steel announces temporary layoffs

Company is growing steak without the cow

Carlo: dema husba

TOP STORIES

What we learned from one of Jeffrey Epstein's final interviews with a...

A 3-year-old was found alone and adrift in a boat in Texas. A man's...

Recommended by Outbrain

San Francisco (CNN Business) — San Francisco, long one of the most tech-friendly and tech-savvy cities in the world, is now the first in the United States to prohibit its government from using facial-recognition technology.

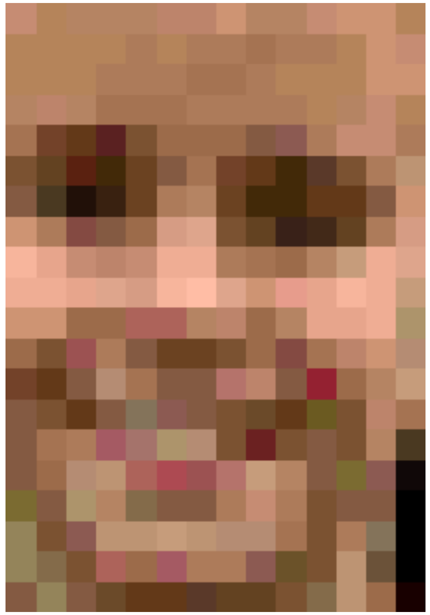
The ban is part of a broader anti-surveillance ordinance that the city's Board of Supervisors approved on Tuesday. The ordinance, which outlaws the use of facial-recognition technology by police and other government departments, could also spur other local governments to take similar action. Eight of the board's 11 supervisors voted in favor of it; one voted against it, and two who support it were absent.

…The ordinance adds yet more fuel to the fire blazing around facial-recognition technology. While the technology grows in popularity, it has come under increased scrutiny as **concerns mount regarding its deployment, accuracy, and even where the faces come from** that are **used to train the systems**.

https://edition.cnn.com/2019/05/14/tech/san-francisco-facial-recognition-ban/index.html

Face perception

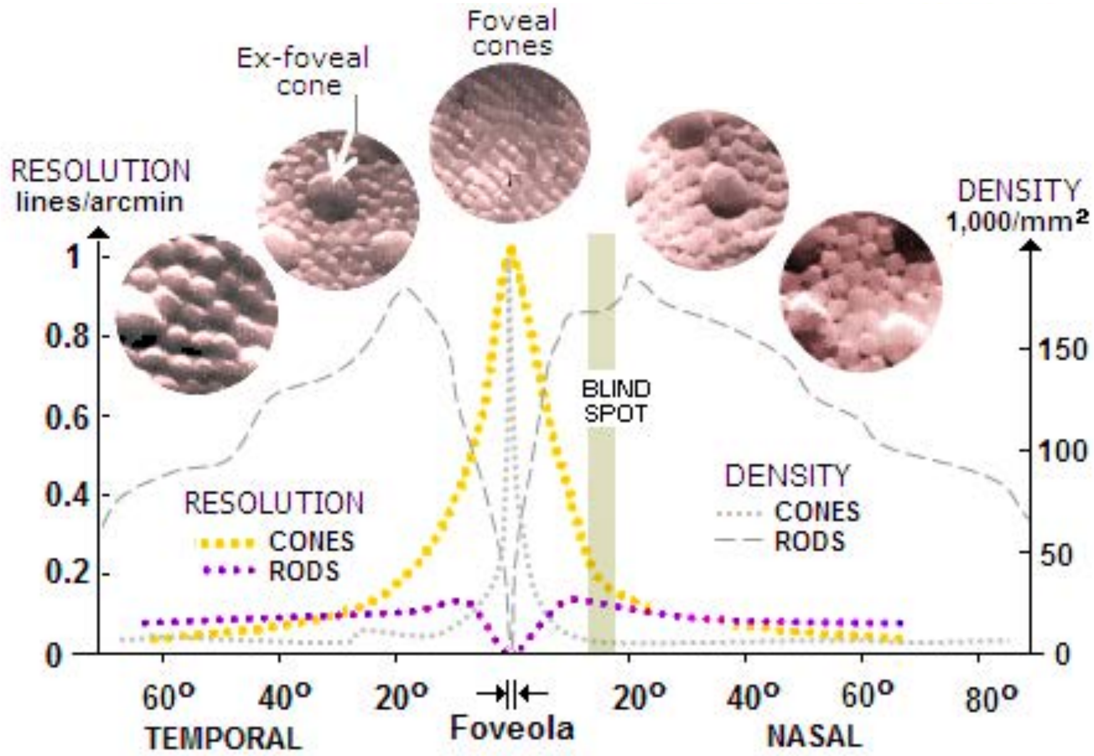
How many pixels to detect/recognize a face?

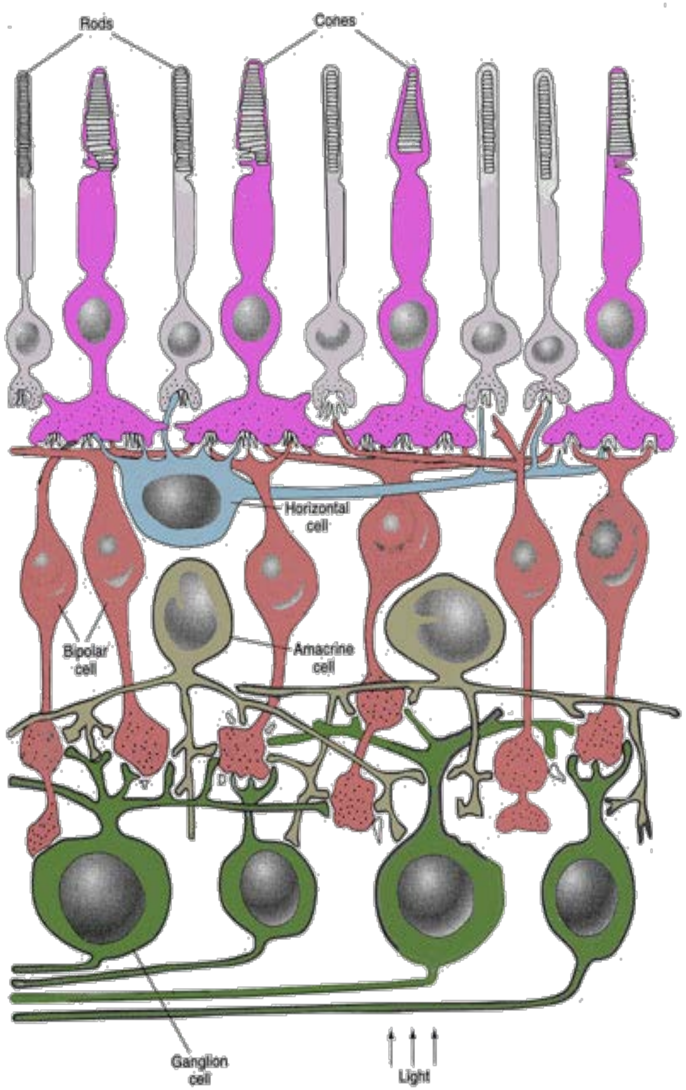


... Not many ... *(20x14)*

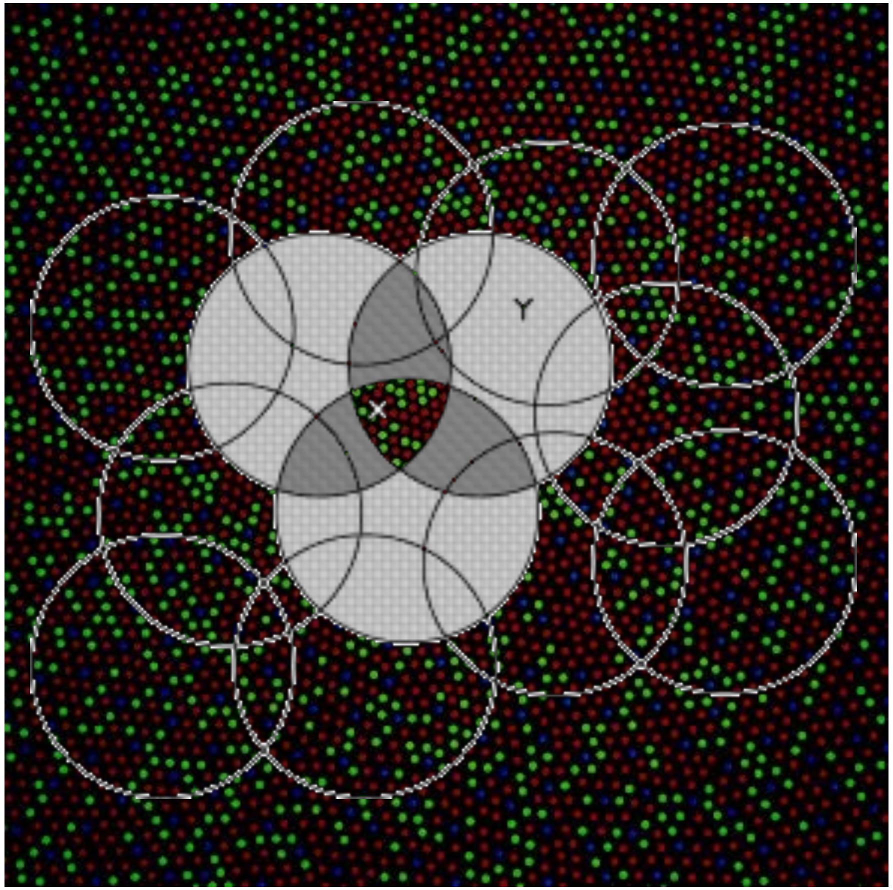
It's more a question of **spatial distribution** and …proper **frequency tuning**

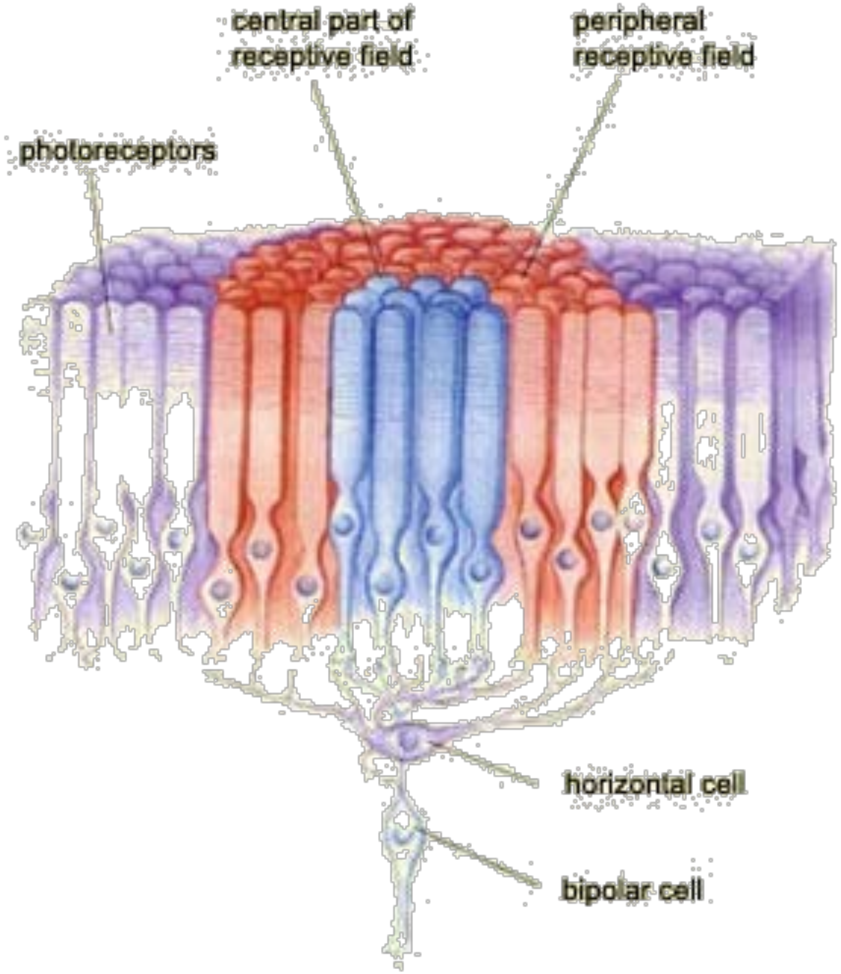
The human retina



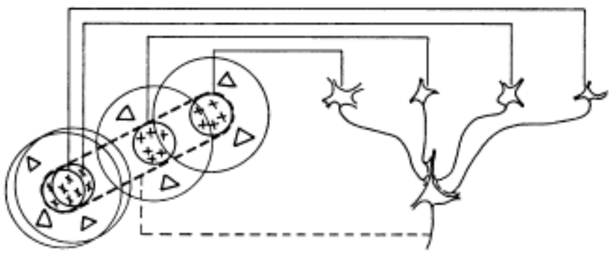


Receptive fields

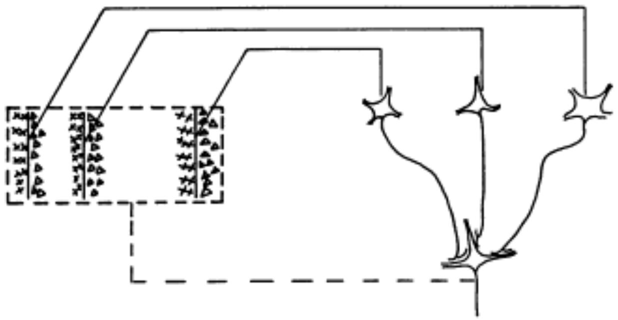




Hubel & Wiesel 1962



Text-fig. 19. Possible scheme for explaining the organization of simple receptive fields. A large number of lateral geniculate cells, of which four are illustrated in the upper right in the figure, have receptive fields with 'on' centres arranged along a straight line on the retina. All of these project upon a single cortical cell, and the synapses are supposed to be excitatory. The receptive field of the cortical cell will then have an elongated 'on' centre indicated by the interrupted lines in the receptive-field diagram to the left of the figure.



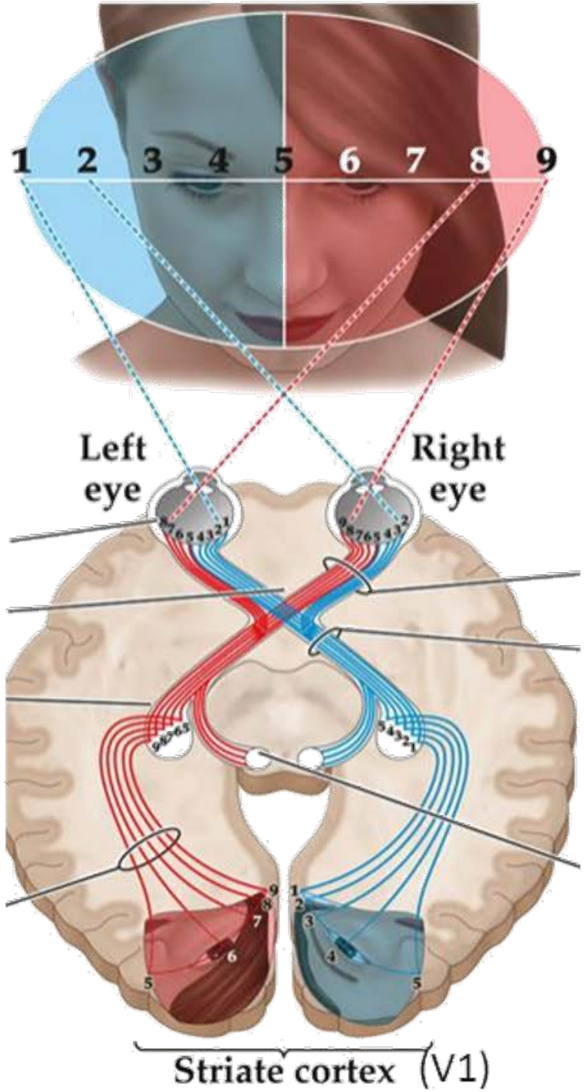
Text-fig. 20. Possible scheme for explaining the organization of complex receptive fields. A number of cells with simple fields, of which three are shown schematically, are imagined to project to a single cortical cell of higher order. Each projecting neurone has a receptive field arranged as shown to the left: an excitatory region to the left and an inhibitory region to the right of a vertical straight-line boundary. The boundaries of the fields are staggered within an area outlined by the interrupted lines. Any vertical-edge stimulus falling across this rectangle, regardless of its position, will excite some simple-field cells, leading to excitation of the higher-order cell.

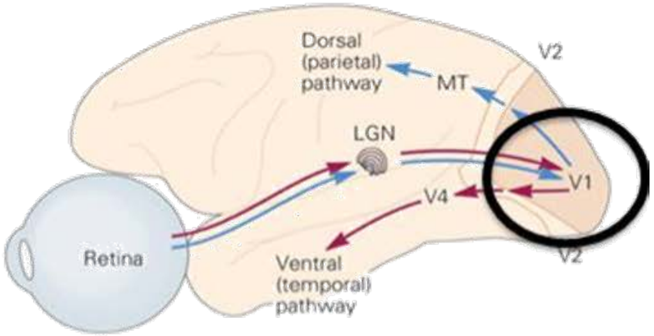
Simple and Complex cells

Hubel DH & Wiesel TN (1962). "**Receptive fields, binocular interaction and functional architecture in the cat's visualcortex**". JPhysiol160, 106–154

Retinotopic mapping

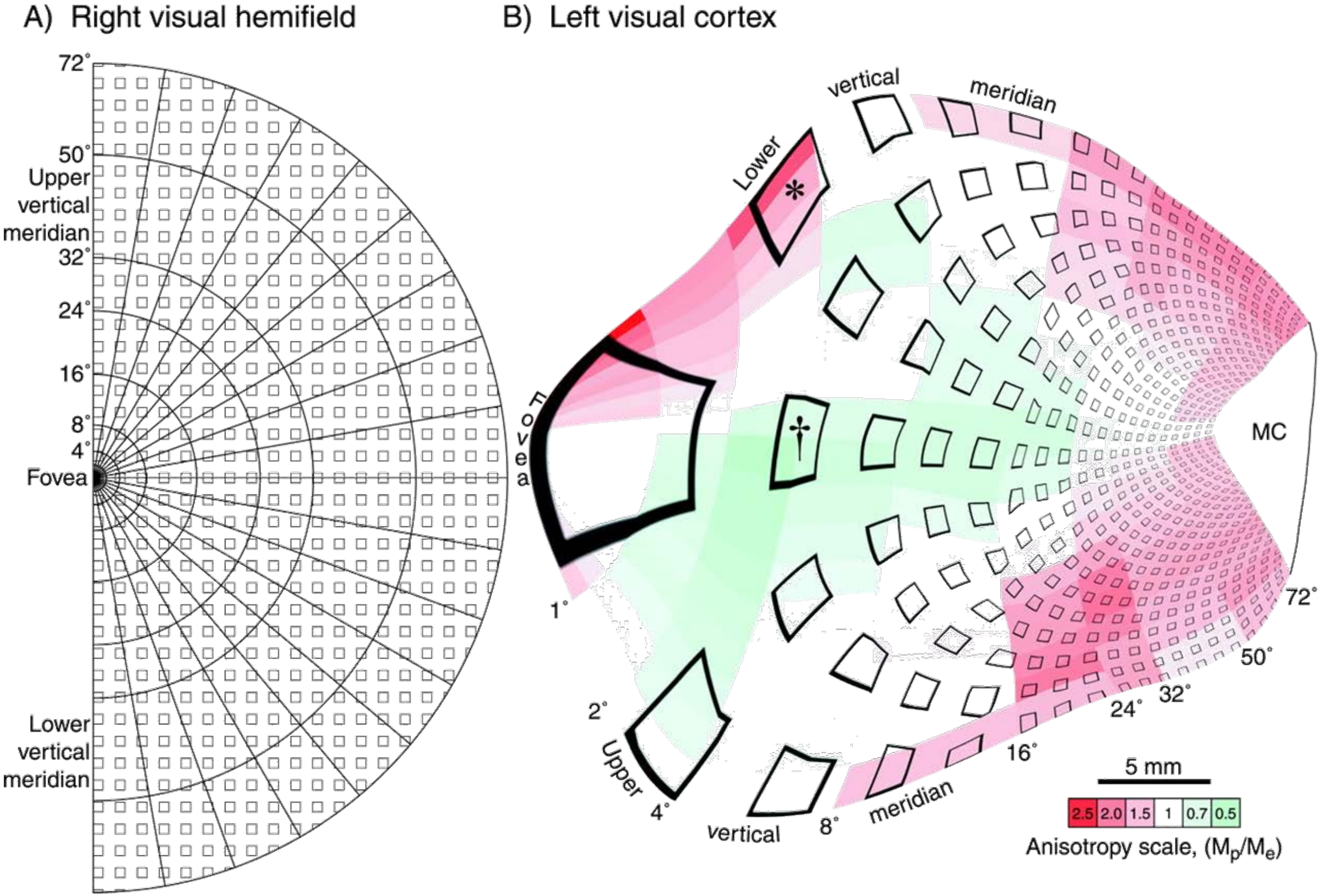
V1 retinotopic maps





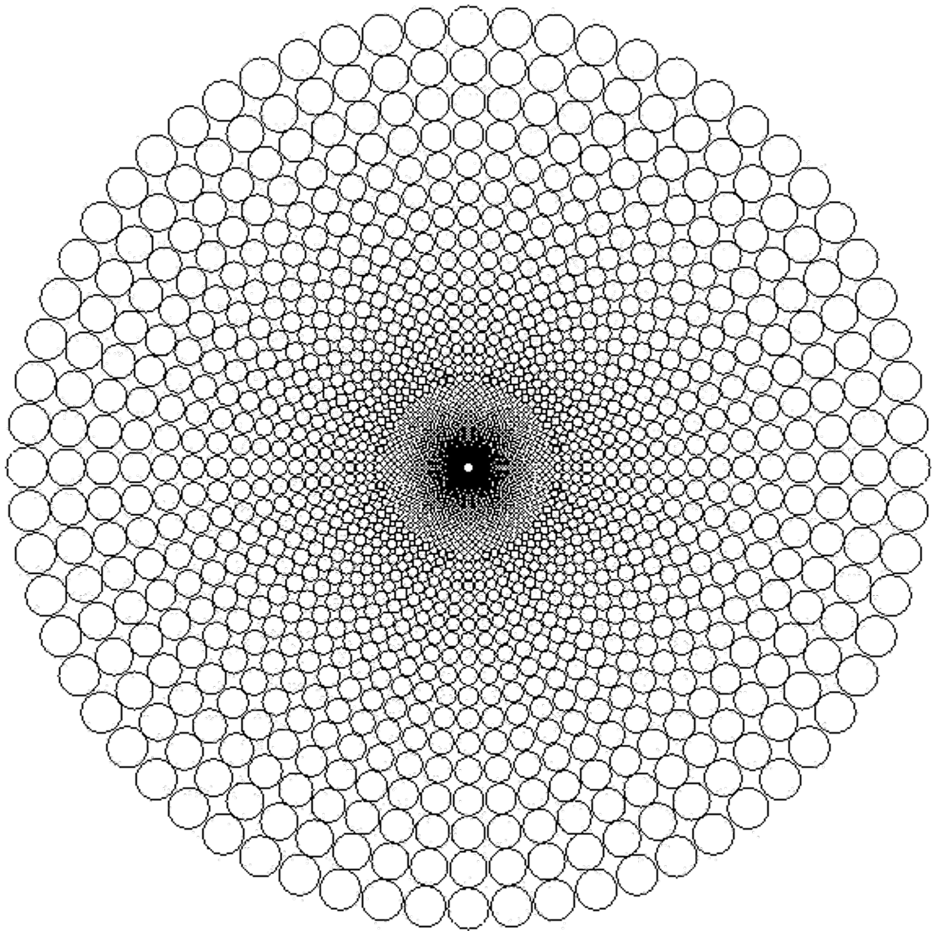
- Each point of the visual field maps on to a local group of neurons in V1.
- Retinotopy = Remapping of retinal image onto cortical surface
- Foveal region uses more of V1 (greater magnification factor)

Retinotopic mapping



Log-Polar mapping

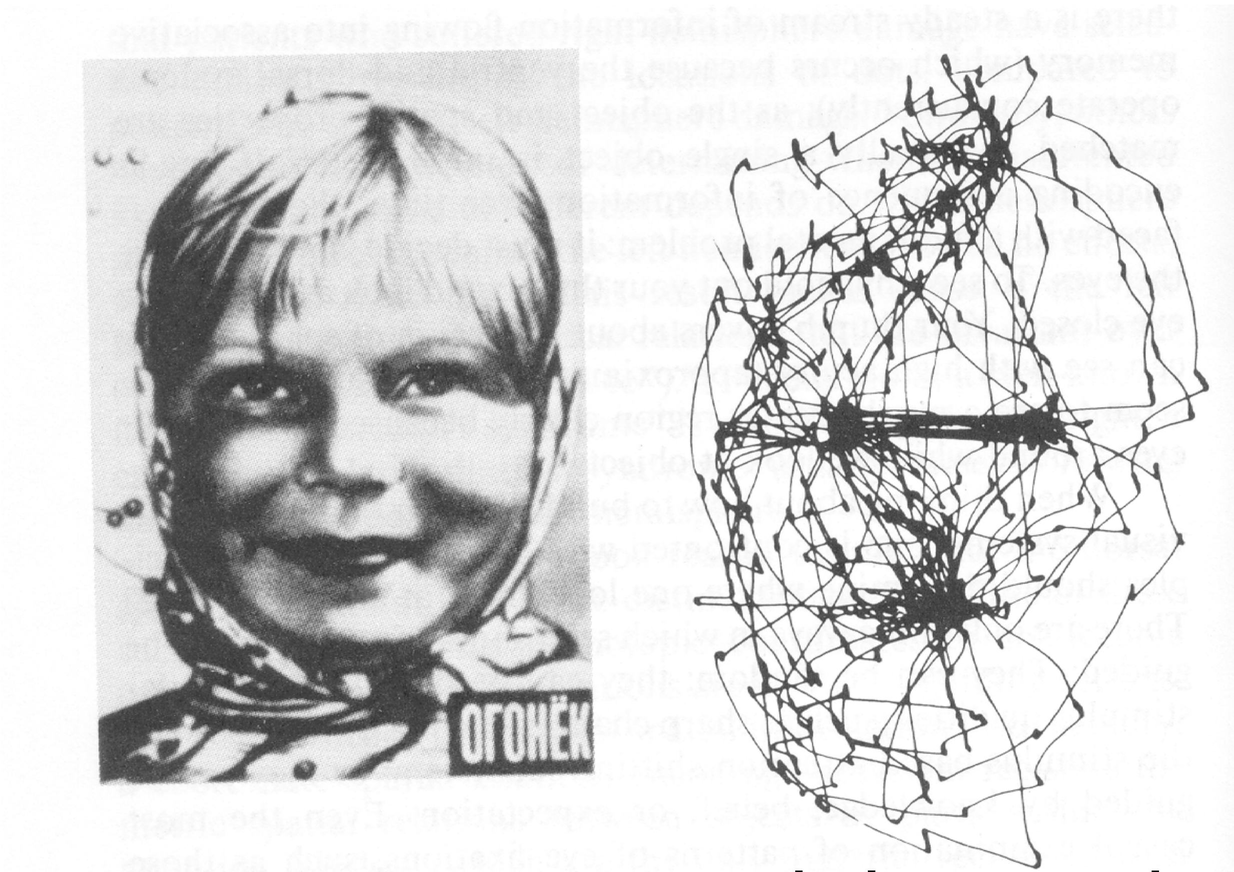
The **complex log-polar transform** is a good approximation of the retinal sampling

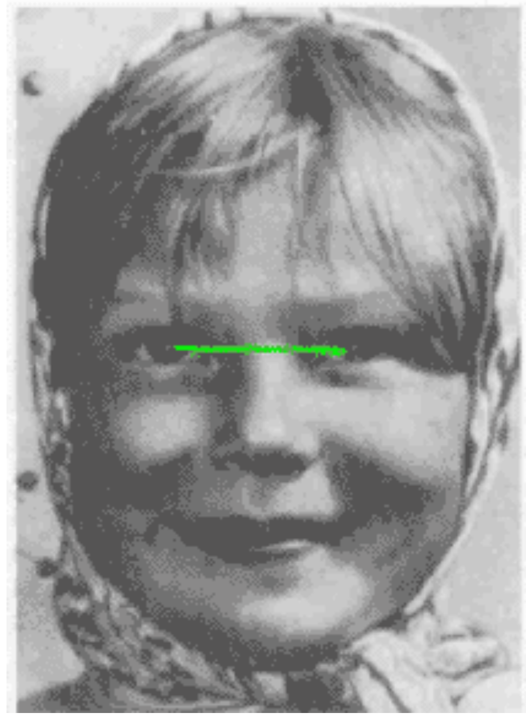


$$\begin{cases} x = \rho \sin\theta \\ y = \rho \cos\theta \end{cases} \qquad \begin{cases} \xi = \log_a\left(\rho / \rho_0 \right) \\ \eta = q\theta \end{cases}$$

Visual attention

Vision Lab

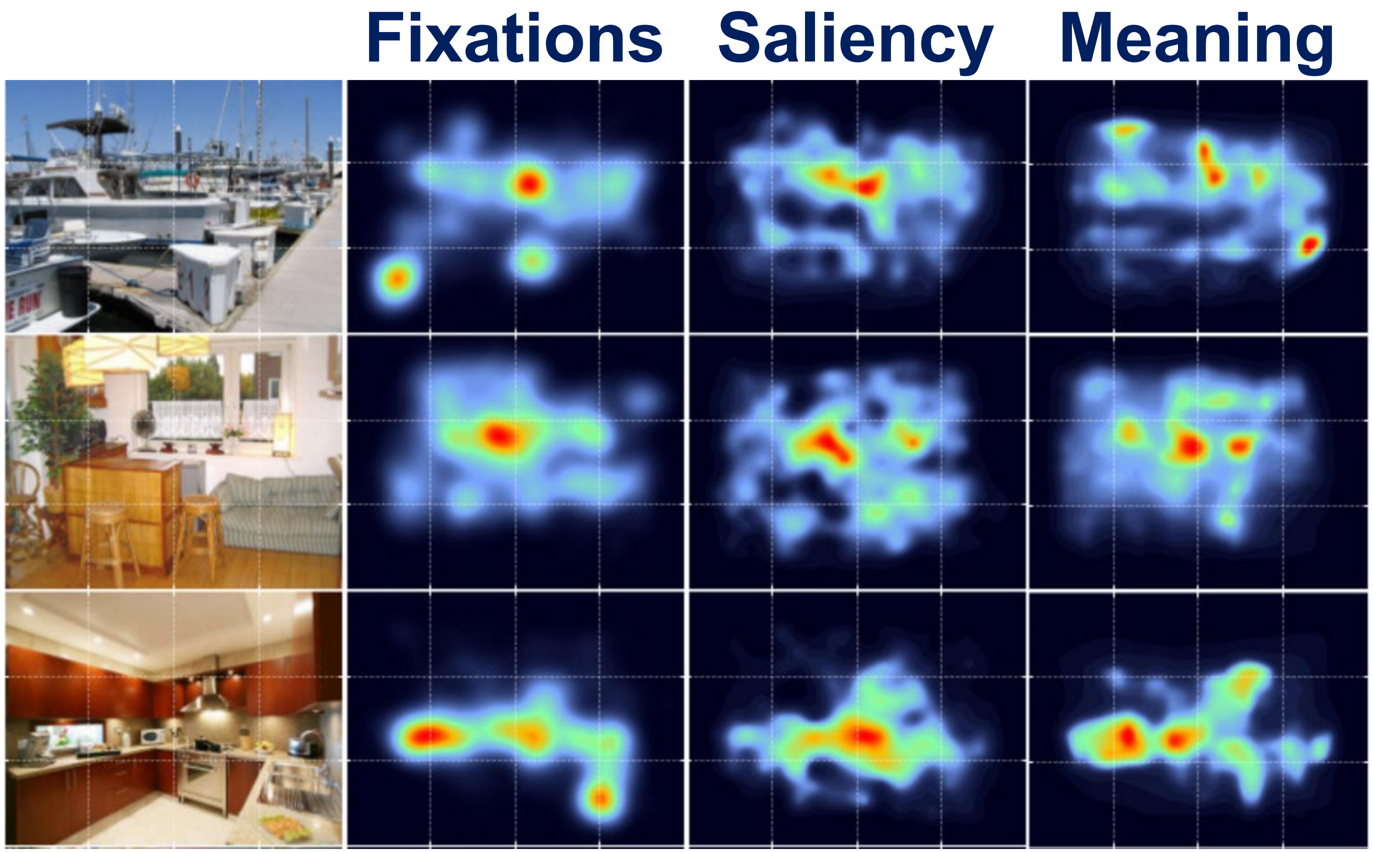




Eye movements while watching a girl's face

A.L. Yarbus, "**Eye Movements and Vision**", Plenum Press, 1967

Visual attention



- Attention is driven by **utilitarian features** related to the **objects' meaning**

J.M,. Henderson, T.R. Hayes, "**Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps**", Journal of Vision 18(6):1-18, June 2018

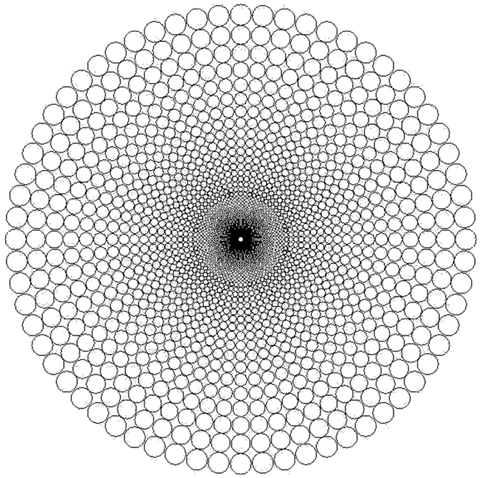
Visual attention

- Body parts have a **meaning**

Visual attention

- Body parts have a **meaning**

Space-variant imaging



Tistarelli, M. and Grosso, E. (1997) "**Active face recognition with an hybrid approach**" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997

Tistarelli, M. and Grosso, E. (2000) "**Active vision-based face authentication**" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space-variant imaging

Tistarelli, M. and Grosso, E. (1997) "**Active face recognition with an hybrid approach**" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997

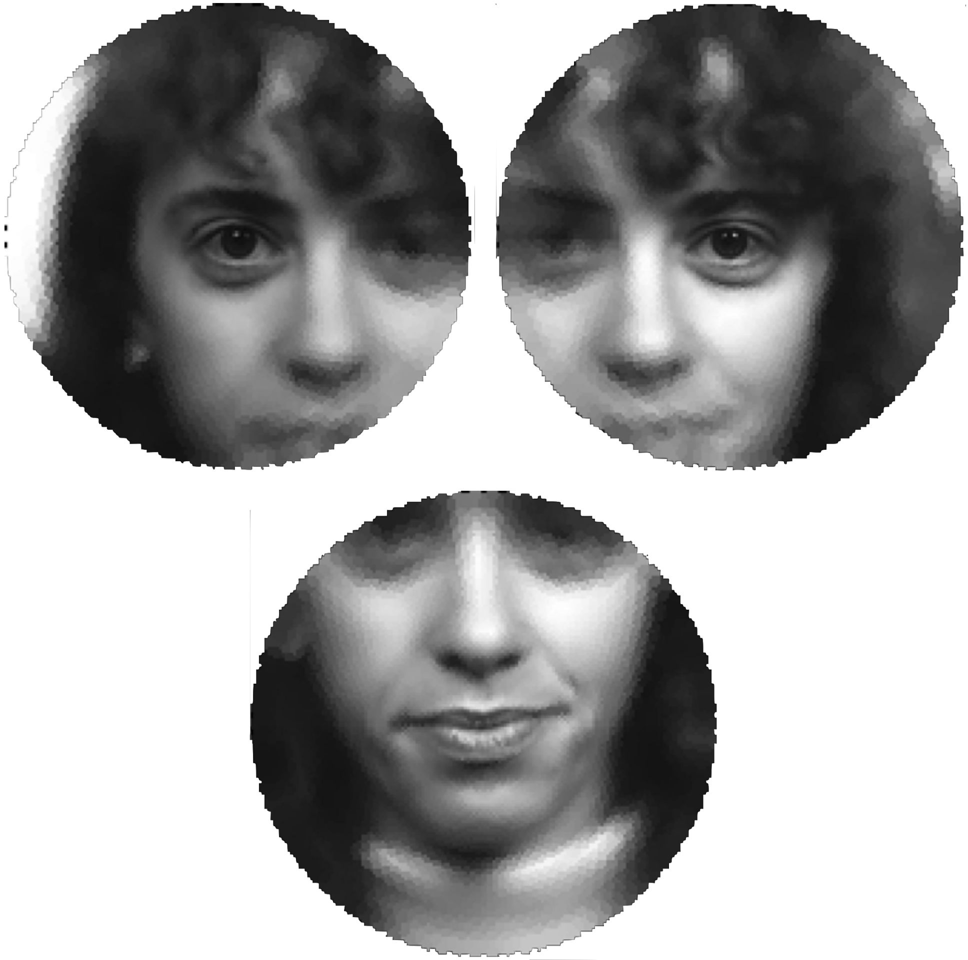
Tistarelli, M. and Grosso, E. (2000) "**Active vision-based face authentication**" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space-variant imaging

Tistarelli, M. and Grosso, E. (1997) "**Active face recognition with an hybrid approach**" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997

Tistarelli, M. and Grosso, E. (2000) "**Active vision-based face authentication**" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space-variant imaging



Tistarelli, M. and Grosso, E. (1997) "**Active face recognition with an hybrid approach**” *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997

Tistarelli, M. and Grosso, E. (2000) "**Active vision-based face authentication**" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space-variant imaging

Tistarelli, M. and Grosso, E. (1997) "**Active face recognition with an hybrid approach**" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997

Tistarelli, M. and Grosso, E. (2000) "**Active vision-based face authentication**" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Space-variant imaging

Tistarelli, M. and Grosso, E. (1997) "**Active face recognition with an hybrid approach**" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997

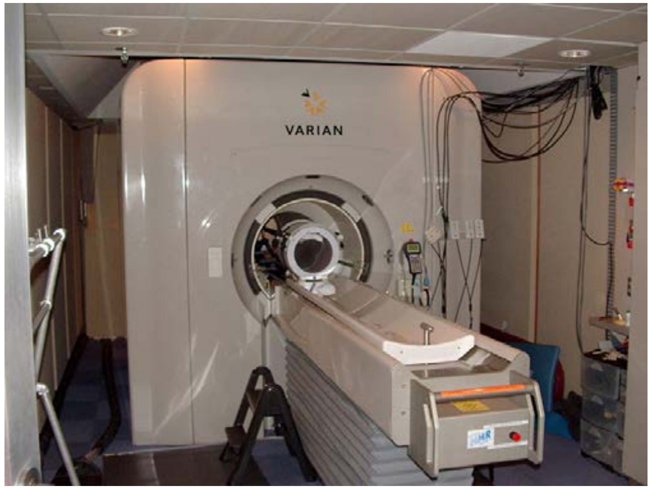
Tistarelli, M. and Grosso, E. (2000) "**Active vision-based face authentication**" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

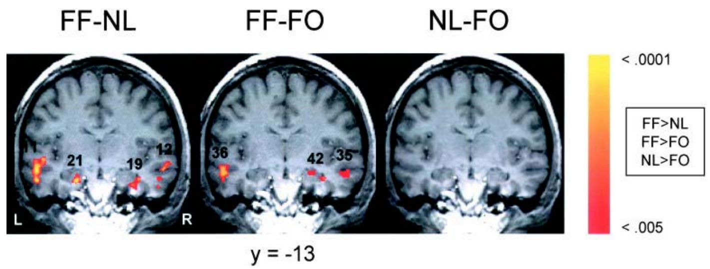
Space-variant imaging

Tistarelli, M. and Grosso, E. (1997) "**Active face recognition with an hybrid approach**" *Pattern Recognition Letters*, Vol. 18, pp 933-946, 1997

Tistarelli, M. and Grosso, E. (2000) "**Active vision-based face authentication**" *Image and Vision Computing*, Vol. 18, no. 4, pp 299-314, 2000

Functional MRI





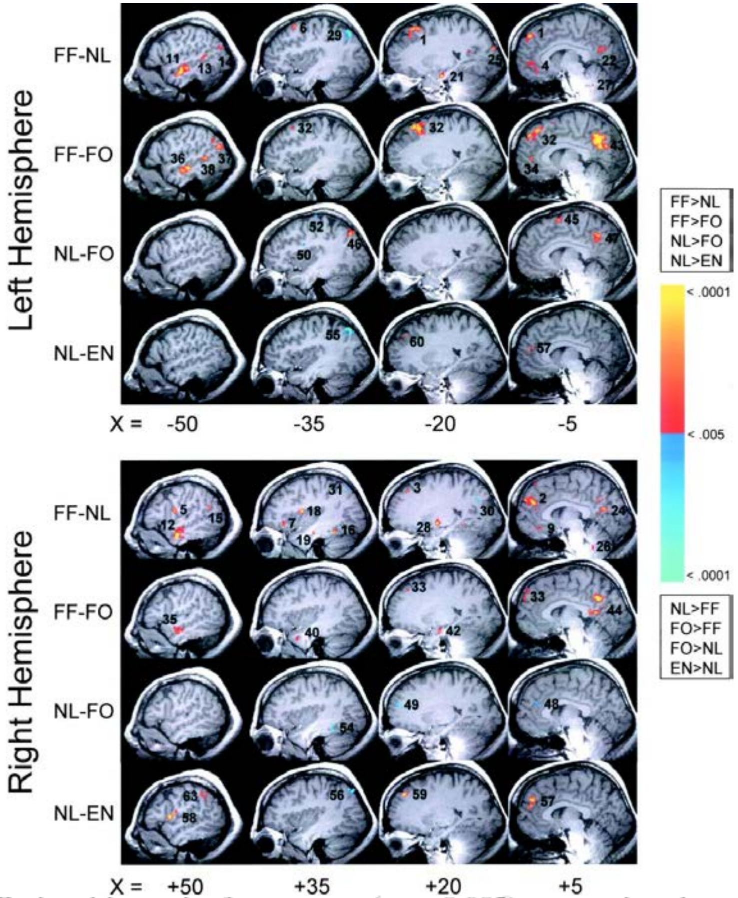
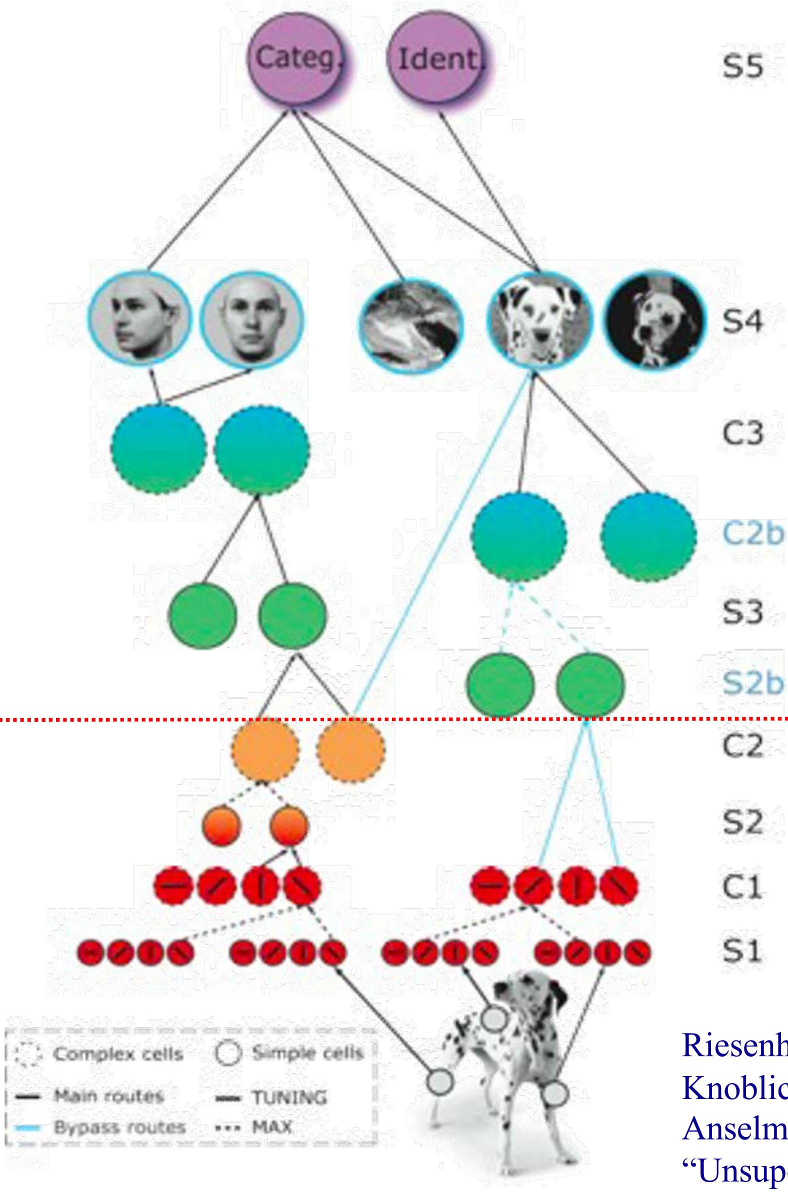


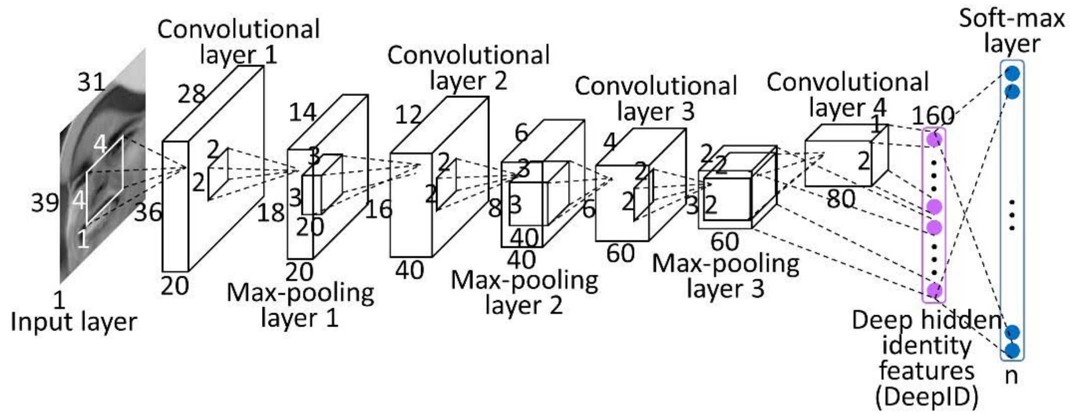
Figure 2. Areas of significantly increased (*red–yellow* scale) and decreased (*blue–cyan* scale) MR signal intensity from *t* tests ($p < 0.005$) comparing the three conditions: FF minus NL, FF minus FO, and NL minus FO. Numbers below each image represent millimeters from the interhemispheric fissure (−, left; +, right). Numbers adjacent to activated foci correspond to location numbers (first column) of Tables 1, 2, and 3.

Recognition of 50 Familiar Faces (**FF**) vs 50 Newly Learned Faces (**NL**) and compared to rejection of 50 Foil (**FO** -False Objective) faces. Encoding (**EN**) session for learning new faces.

C. L. Leveroni et al. **"Neural Systems Underlying the Recognition of Familiar and Newly Learned Faces",** The Journal of Neuroscience, January 15, 2000, *20*(2):878–886

Brain models





C2 **Complex composite cells**

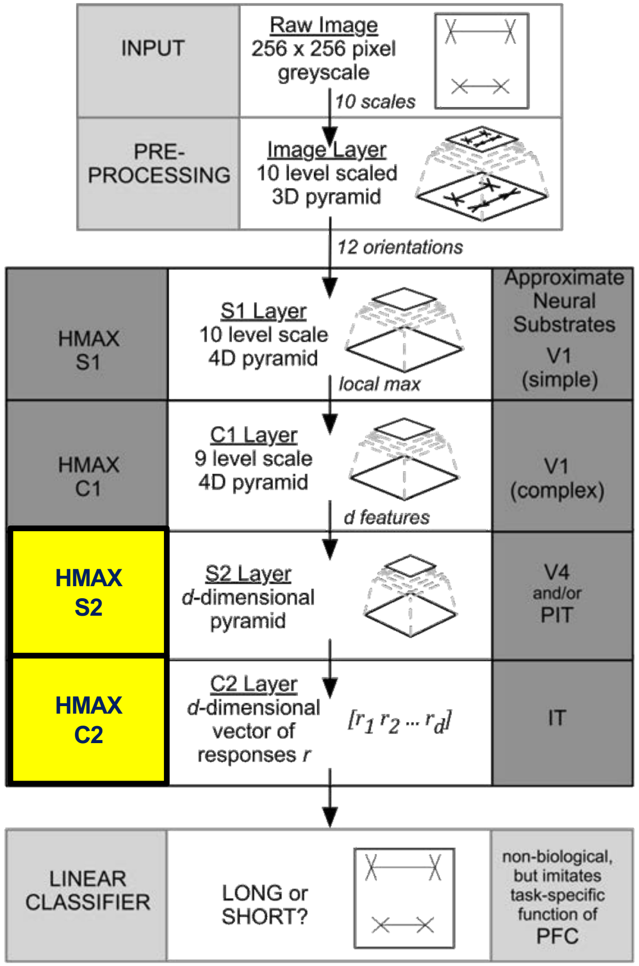
S2 **Composite feature cells**

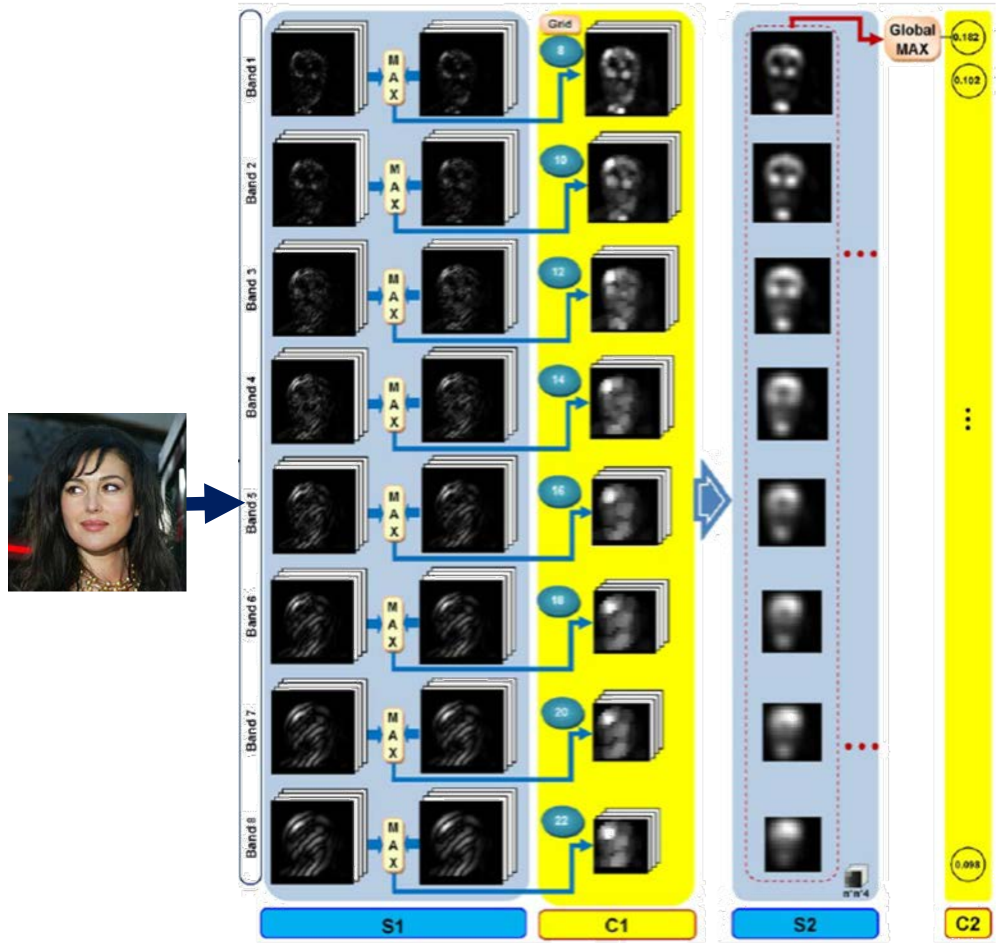
C1 **Complex cells**

S1 **Simple cells**

Riesenhuber & Poggio 1999, 2000; Serre Kouh Cadieu
Knoblich Kreiman & Poggio 2005; Serre Oliva Poggio 2007
Anselmi, F., Leibo, J. Z., Rosasco, L., Mutch, J., Tacchetti, A., and Poggio, T., "Unsupervised learning of invariant representations", Theoretical Computer Science, 2015.

The HMAX model

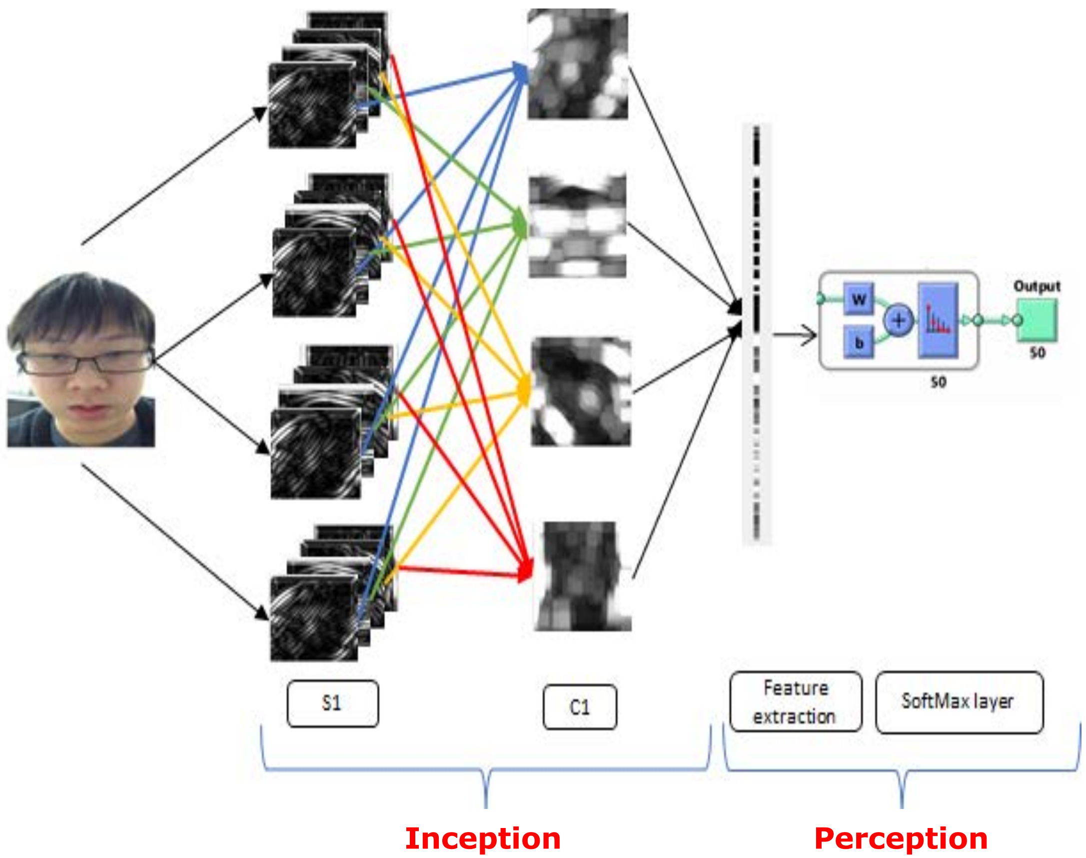




Riesenhuber, M. & Poggio, T. (1999). Hierarchical Models of Object Recognition in Cortex. Nature Neuroscience 2: 1019-1025.

(S1) In this layer an input image is analyzed with a pyramid of filters (16 filter sizes×4 orientations = 64 images)
(C1) In this layer, the local maximum between 2 adjacent scales with the same orientation is taken.
(S2) The Euclidean distances between stored prototypes, which are obtained in the learning stage, and new input is computed. This process occurs for all bands in C1 and as a result, S2 maps are obtained.
(C2) The global maximum is computed over all S2 responses in all positions and scales in this layer.

Face recognition with HMAX

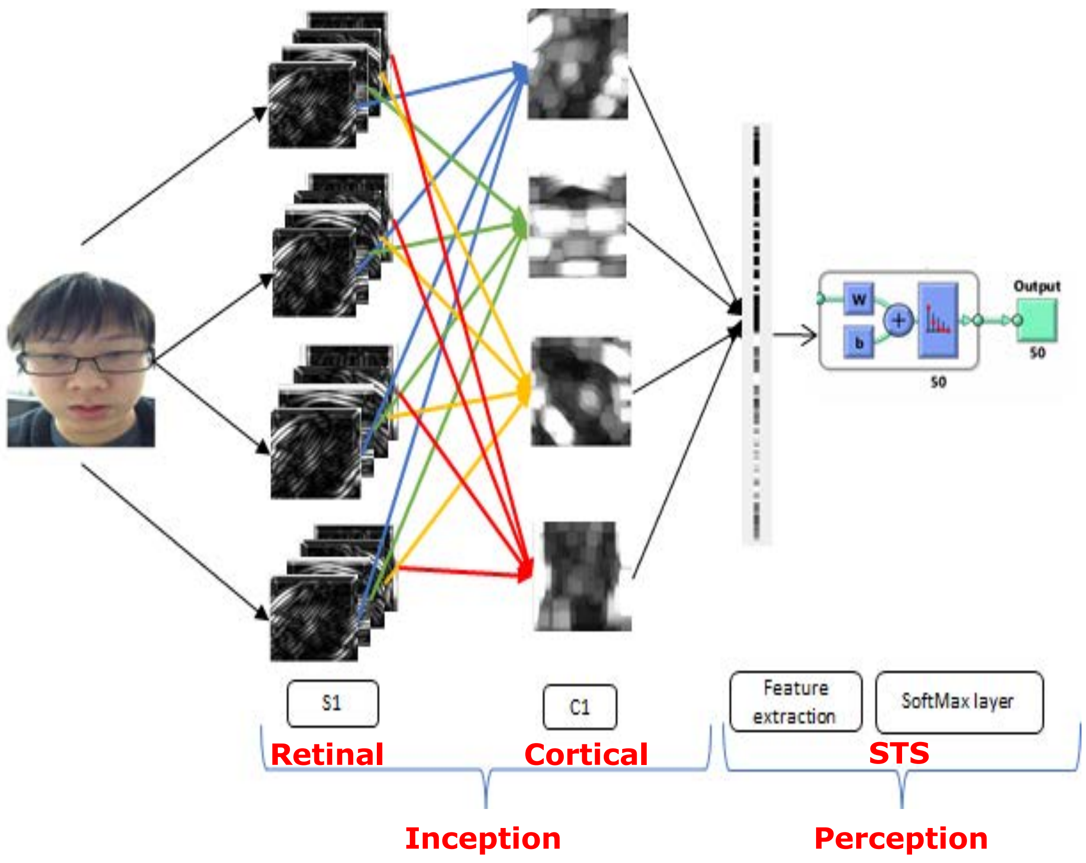


The architecture is inspired by the concept of ***Inception-Perception***.

The ***Inception*** part is implemented by the **S1** and **C1** layers of the **HMAX** network, followed by a down-sampling operator to build the feature vectors.

The ***Perception*** part is implemented by a **SoftMax** layer.

Face recognition with HMAX

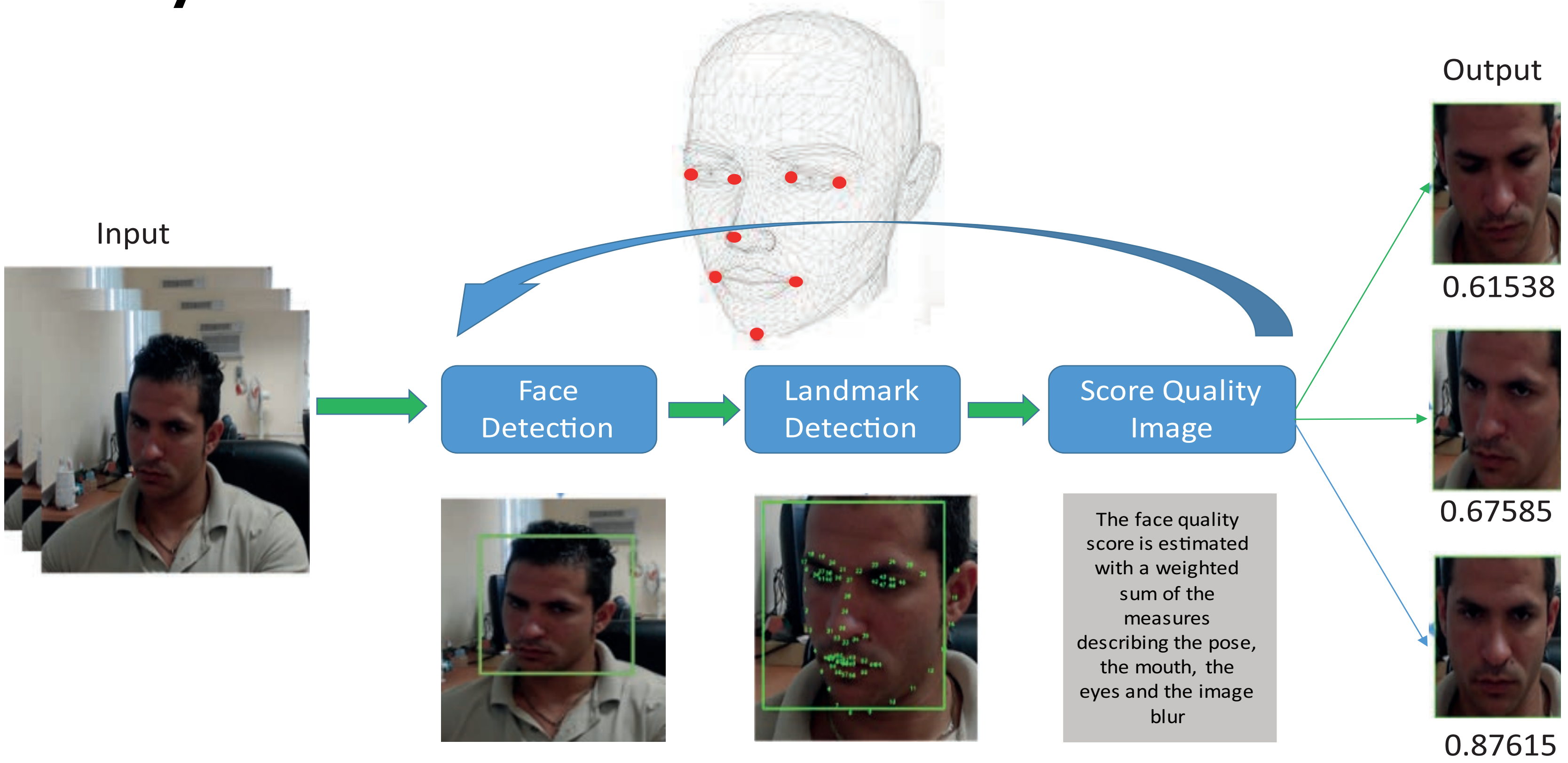


The **Gabor** and **max pooling** layers encode the face images based on a biologically-inspired chain running from the **retinal** stage to the **V1 cortex**.

The connections between the **V1 cortex** and the **Superior Temporal Sulcus**, the face-selective area, is simulated by a network whose neurons are activated by a ***SoftMax*** function.

Visual attention

- ***Meaningful* facial regions** are extracted according to the position of facial landmarks
- Images are clustered in different categories, according to the approximate **head rotation** along the vertical axis.
- Regions are associated to each pose category according to their **visibility**



Feature extraction and fusion

- The **S1** and **C1** layers in the HMAX are used.
  - The **S1** layer performs a band-pass filtering with a bank of Gabor kernels.
  - At the local invariance layer (**C1**), a local maximum is computed for each orientation.
- The final feature vector is built by down-sampling the output by 8, obtaining a 256-dimensional feature vector.
- The feature vectors, extracted from different facial regions, are concatenated into a single feature vector of fixed size, according to the head rotation. For example, the feature vector for head right rotation is:

$$F = [F_{le}; F_m; F_c; F_a]$$

F_{le} ; F_m; F_c *and* F_a are the feature vectors obtained from the face regions extracted from the left eye, mouth, chin and forehead.

Classification

- During the learning phase, a neural network, with a ***SoftMax*** activation, is trained from a subset of the available sample data (disjoint from the test data).
- The loss function for the ***SoftMax*** layer is based on the computation of the ***crossentropy***:

$$L_i = -\log(\frac{e^{f_i}}{\sum j e^{f_j}})$$

Where $\boldsymbol{f_j}$ is the *j-th* element of the feature vector representing subject ***f***, while $\boldsymbol{L_i}$ is the full loss over the training examples.

- The concatenated feature vectors are fed to the classification network. The scores obtained from each image group are fused by applying a mean rule.

Experimental results

TABLE VI
RECOGNITION RATE OBTAINED BY FUSING THE FEATURES EXTRACTED FROM DIFFERENT FACIAL REGIONS.

Session	Best frames	Average frames	Bad frames
1	**96.36**	72.73	58.18
2	**87.27**	34.55	74.55
3	**80.00**	50.91	54.55

- In this experiment different regions are fused from each frame category.
- The features extracted from the fiducial regions are concatenated into a single feature vector for classification.

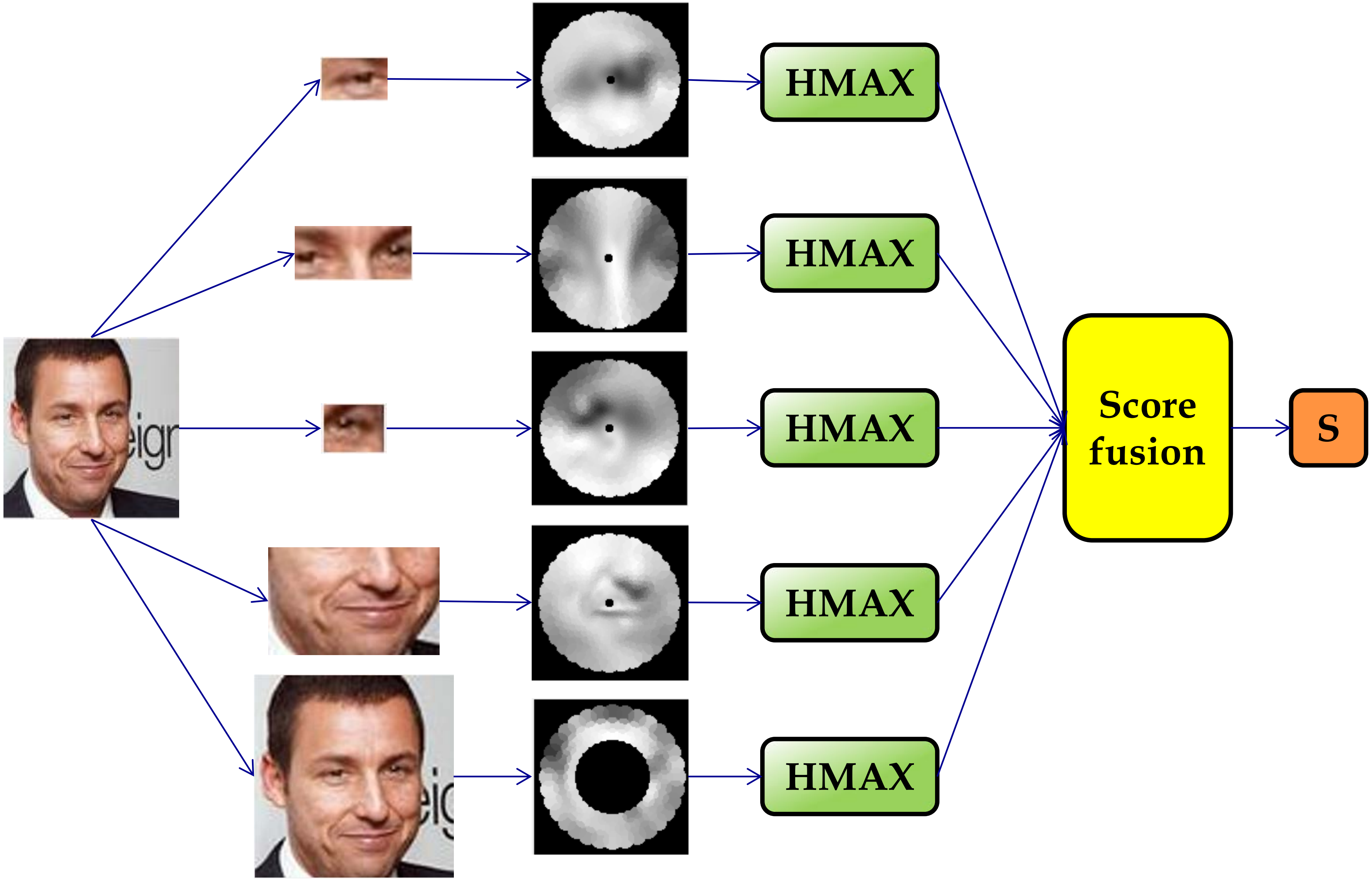
TABLE VII
COMPARISON WITH THE METHODS DESCRIBED IN [1] AND FOLLOWING THE TESTING PROTOCOL 1. THE RECOGNITION RATE FOR THE PROPOSED METHOD WAS OBTAINED FROM THE THREE HEAD POSE GROUPS.

Training	Testing	FF	SRC	MSSRC	S1C1	VGG	Proposed approach
1	2	54.48	52.79	47.21	7.27	**67.27**	61.82
1	3	45.27	51.18	46.15	16.36	49.09	**52.73**
2	1	25.52	44.18	43.06	20.00	50.91	**65.45**
2	3	56.8	58.58	60.36	52.73	38.18	**72.73**
3	1	24.77	17.64	17.64	20.00	**47.27**	36.36
3	2	**56.01**	51.95	45.85	51.82	33.64	50.91

- In this experiment the protocols defined for the UMDAA database were applied.
- Performances are compared with Fisher Faces (FF), Sparse Representation based classification (SRC) and Mean-Sequence SRC (MSSRC) and the VGG deep network model.

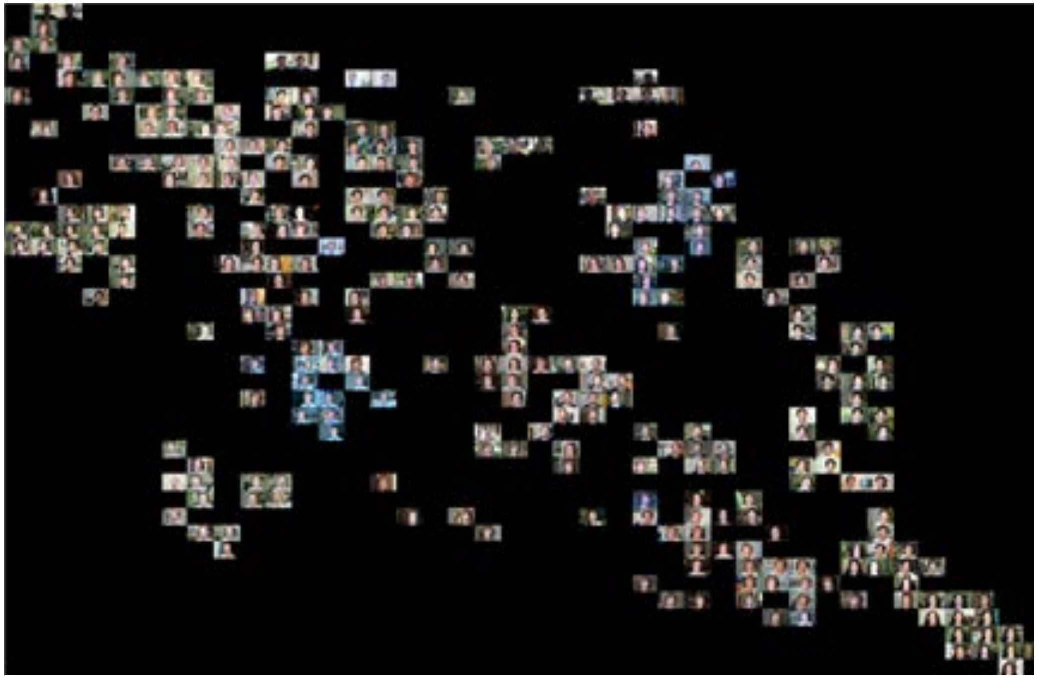
Khellat Kiehl, S, Lagorio, A and Tistarelli, M (2019) "**A Biologically-Inspired Attentional Approach for Face Recognition**" *Proc. of IAPR/IEEE Int.l Workshop on Biometrics and Forensics* – IWBF 2019, Cancun, Mexico, May 2019.

Foveated face recognition



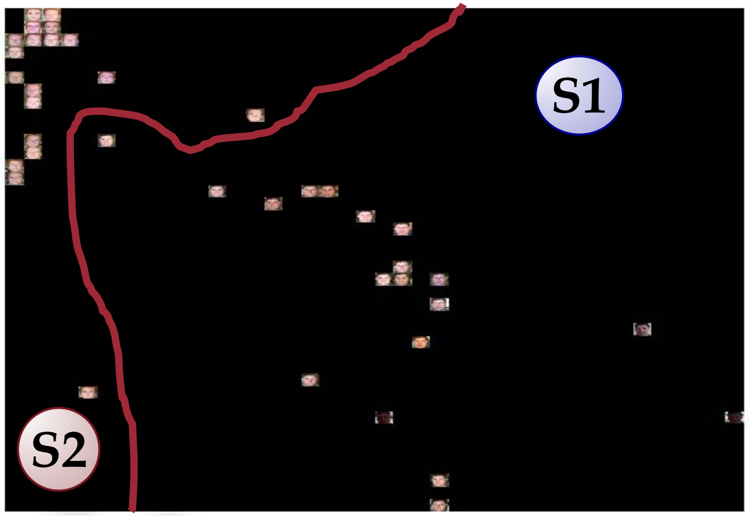
Foveated face recognition

HMAX Space representation on uniformly sampled face images

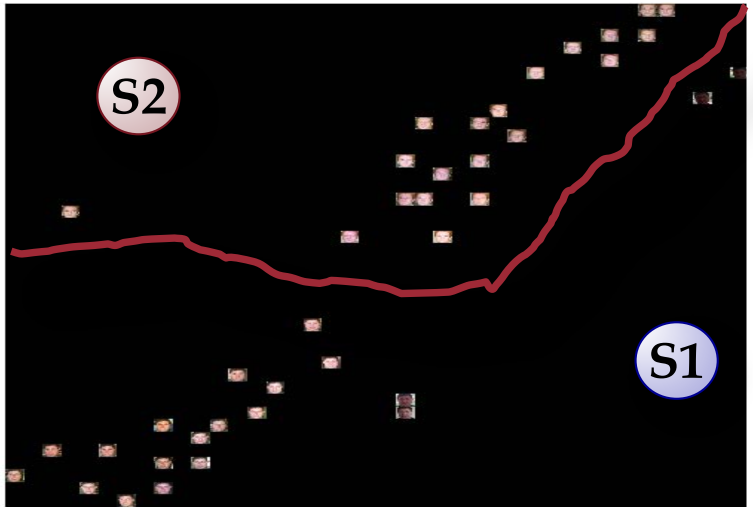


HMAX Space representation on log-polar sampled face images

Foveated face recognition



Uniform resolution



Log-polar mapping

Training	Testing	FF	SRC	MSSRC	VGG	Outer face	Ocular regions	Fusion
1 Lab light	2 Dim light	54.48	52.79	47.21	**62.27**	53.15	33.33	54.95
1 Lab light	3 Sun light	45.27	51.18	46.15	49.09	94.31	91.87	**95.12**
2 Dim light	1 Lab light	25.52	44.18	43.06	50.91	56.76	66.67	**78.38**
2 Dim light	3 Sun light	56.80	58.58	60.36	38.18	84.68	73.87	**84.68**
3 Sun light	1 Lab light	24.77	17.64	17.64	47.27	48.78	73.17	**73.98**
3 Sun light	2 Dim light	**56.01**	51.95	45.85	33.64	48.65	31.53	50.45

Performances are compared with Fisher Faces (FF),
Sparse Representation based Classification (SRC), Mean-Sequence SRC (MSSRC) and VGG deep CNN.

S. Khellat Khiel, A. Lagorio, M. Tistarelli. "**Face Recognition 'On the Move' Combining Incomplete Information**". Proc. of 6th Int.l Workshop on Biometrics and Forensics, June 7,8 2018, Alghero, Italy. IEEE 2018.

S. Khellat Khiel, A. Lagorio, M. Tistarelli. "**Foveated vision for biologically-inspired continuous face authentication**". In A. Rattani Ed. *Selfie Biometrics: Methods and Challenges*, Springer 2019.

Conclusion

- **Deep neural architectures** provide today the current state of the art performance of face recognition *in the wild.*
  - The large number of layers requires a huge amount of data for training to reach a stable configuration of the neural connectivity.
  - They are sensitive to unexpected changes in the spatial frequencies of the input patterns.
- **Simple biologically-inspired networks** may allow to perform very complex visual tasks.
- In biological systems **attention** drives **recognition**.
  - A space-variant **scale-space decomposition** of the input signal allows to select the most informative data.
- The **S1C1** neural architecture, derived from the ***HMAX*** model, with face quality, **outperforms the deep VGG model**.
  - The **peripheral area of the face** (face outline and hair dressing) proved to be very distinctive for recognition.

18th Int.l Summer School for Advanced Studies on Biometrics for secure authentication:

“AI FOR BIOMETRICS / BIOMETRICS FOR AI”

Alghero, Italy - May, 31 – June 5 2021

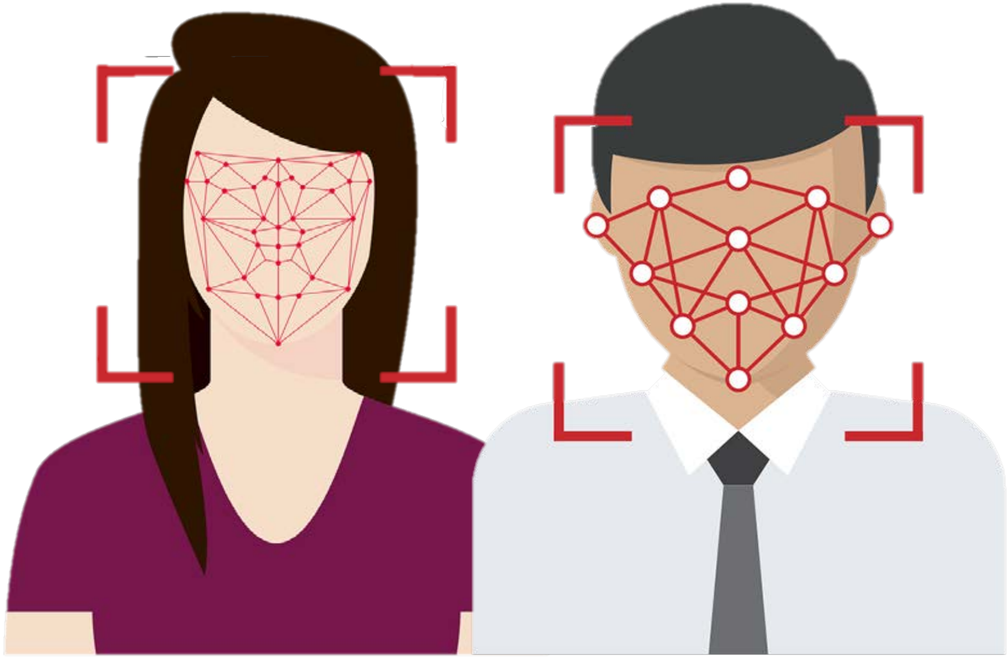
http://biometrics.uniss.it

Contact: tista@uniss.it

IAPR

IDENTITY

IDEMIA augmented identity



IEEE

EURASIP

450 UNISS

SVSCEPTVM PERFICE MVNVS

Vision Lab

THANK YOU
FOR YOUR *ATTENTION*