Constrained and Cooperative Face Recognition

Massimo Tistarelli

Computer Vision Laboratory

University of Sassari – Italy

tista@uniss.it

The Computer Vision Lab

Since 2003 hosting the Int.l Summer School on Biometrics

Credits

Vision Lab

- 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
  Humera Tariq
  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

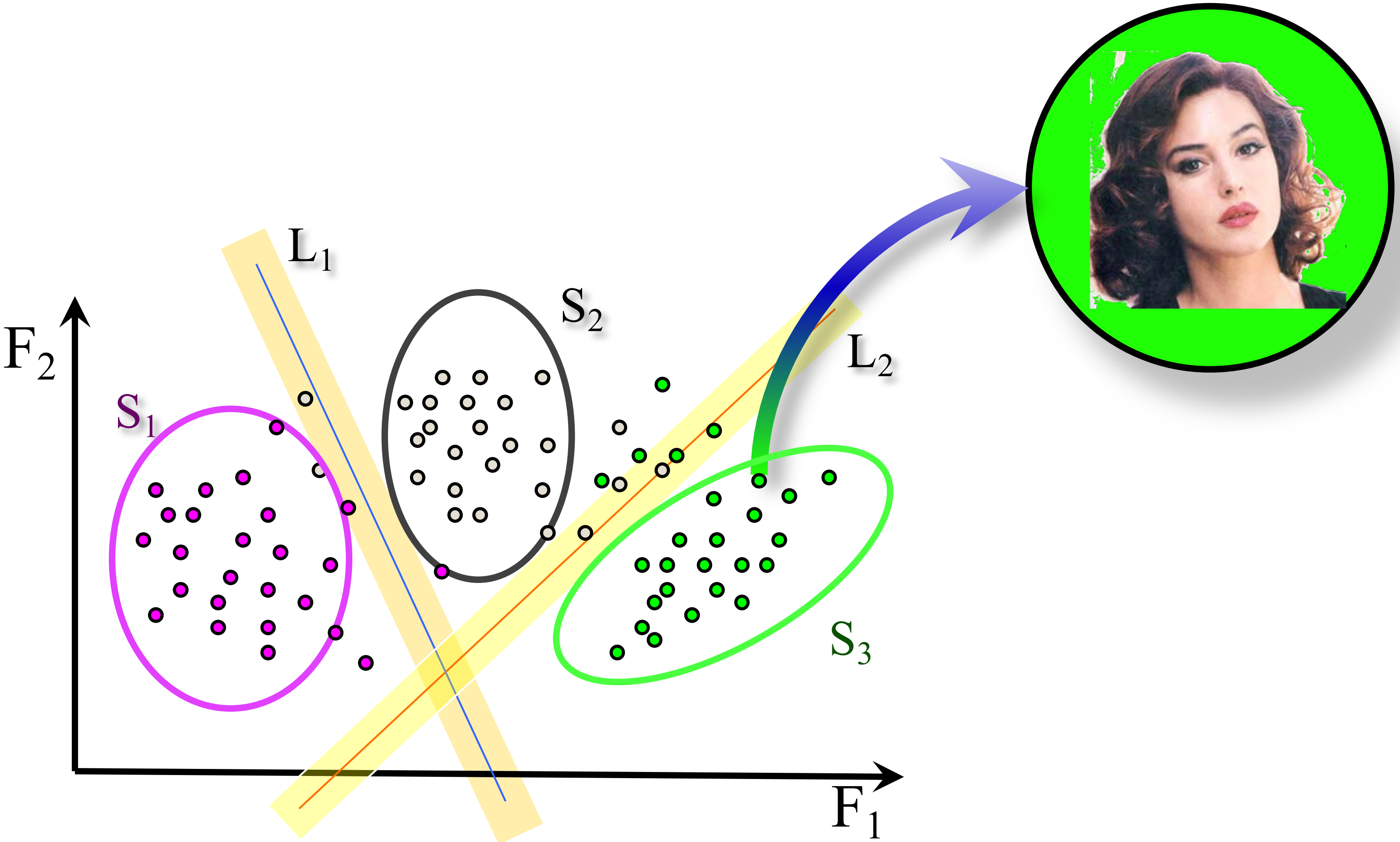
Media Advertisements

Media Advertisements

Face Recognition

Vision Lab

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



$= \Psi(F_1, F_2)$

Genuine and Impostor scores

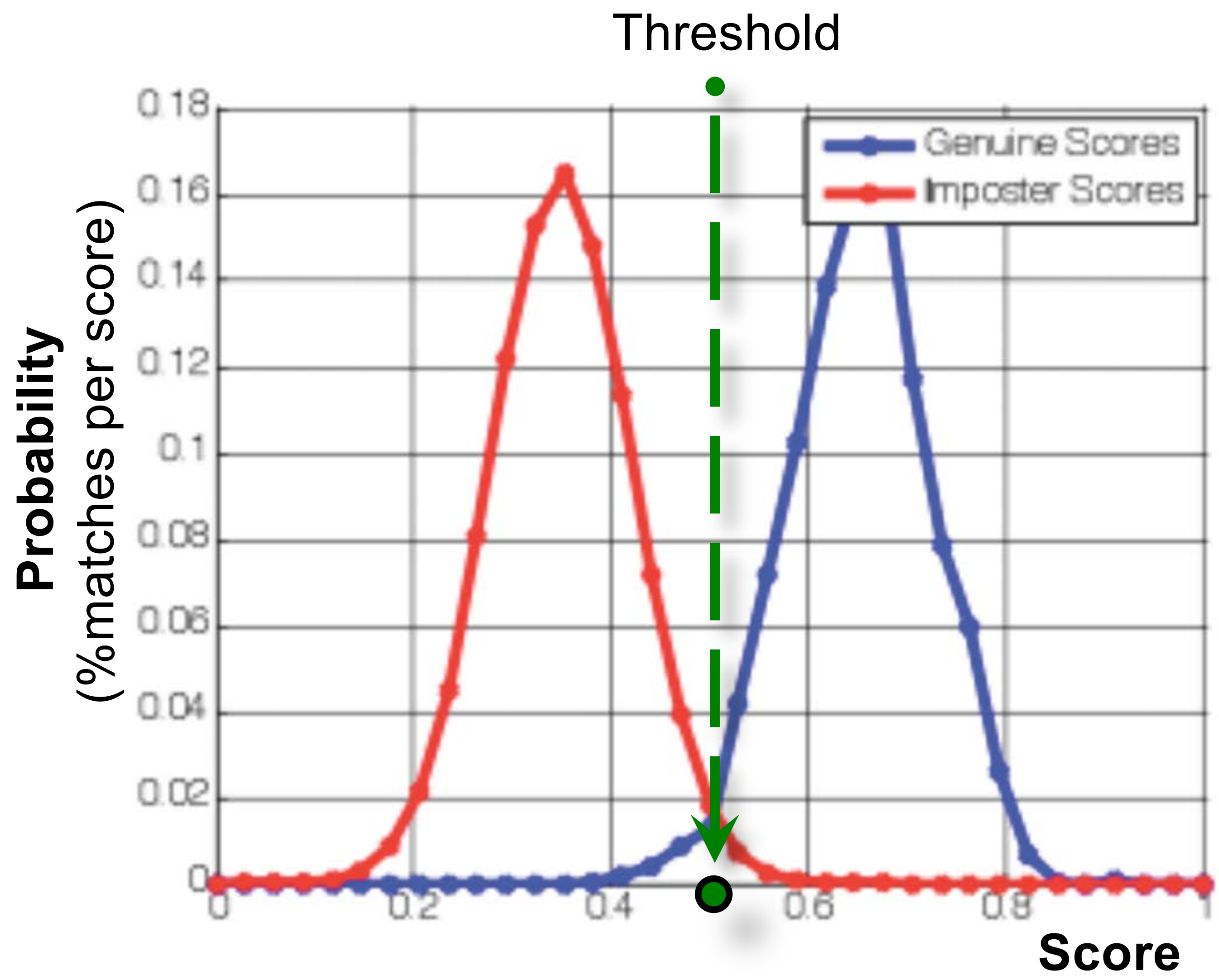
- **Genuine score**: Match score (or *distance*) computed when two biometric samples from the **same** individual are compared.

- **Impostor score**: Match score (or *distance*) computed when two biometric samples originating from **different** individuals are compared.

Therefore, **a genuine user score should be always greater than an impostor score**.

- A **threshold** (or **classifier**) is used to determine if a score is related to a genuine user or an impostor.

Match score distributions



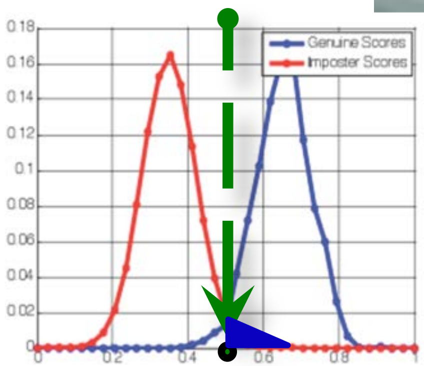
Inter-class similarity

Two different people with very similar appearance

FALSE MATCH

www.marykateandashley.com

Twins



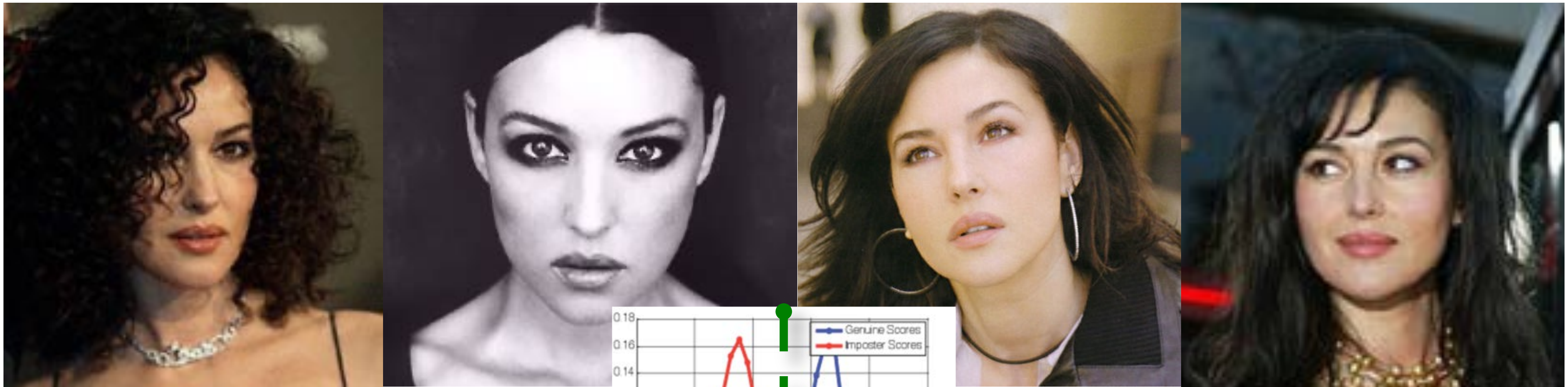
i/english/in_depth/americas/2000/us_elections

Father and son

Intra-class *variability*

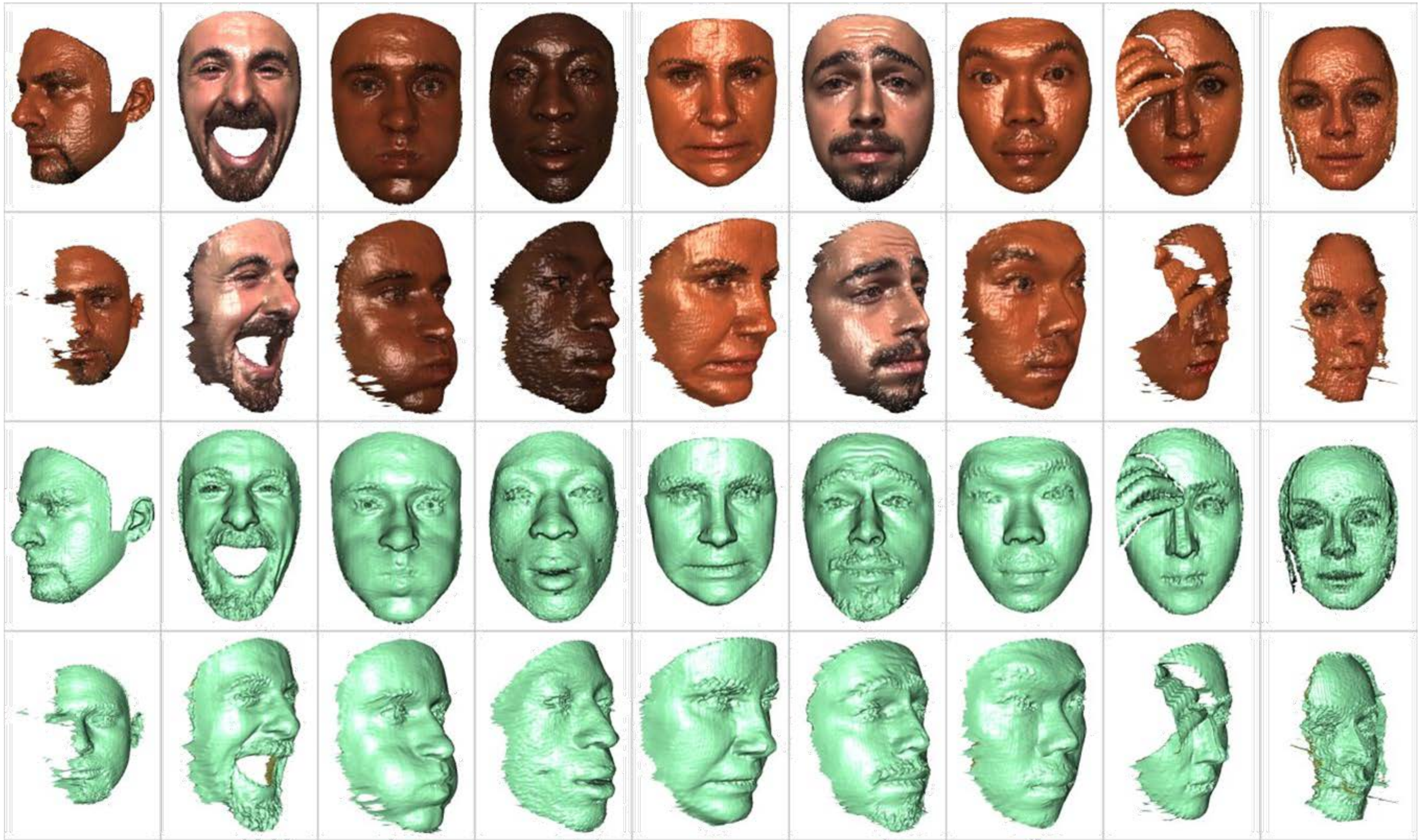
The same person with very different biometric samples

FALSE <u>NON MATCH</u>



Monica Bellucci

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

UMD-AA Mobile Device Database

U. Mahbub, S. Sarkar, V. M. Patel and R. Chellappa, "**Active user authentication for smartphones: A challenge data set and benchmark results**," 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, NY, 2016, pp. 1-8..

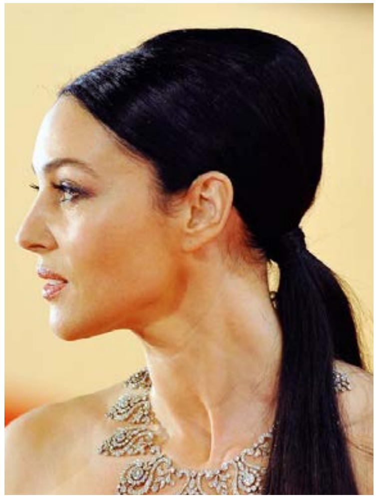
Visual challenges

A – Aging

P – Pose

I – Illumination

E - Expression



An inverse problem

Jacques Hadamard

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

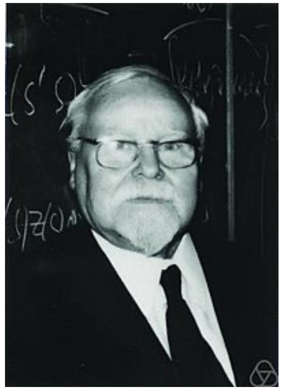
1) a ***unique*** solution exists and
2) it depends ***continuously*** upon the data.

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

An ill-posed problem

Vision Lab

Jacques Hadamard



Andrej Tikhonov

Two adverse conditions:

1) **Noise** in the data (many sources, including **A.P.I.E.**)
2) **Dimensionality** of the data (from 4D to 2D)

Solution: Regularization

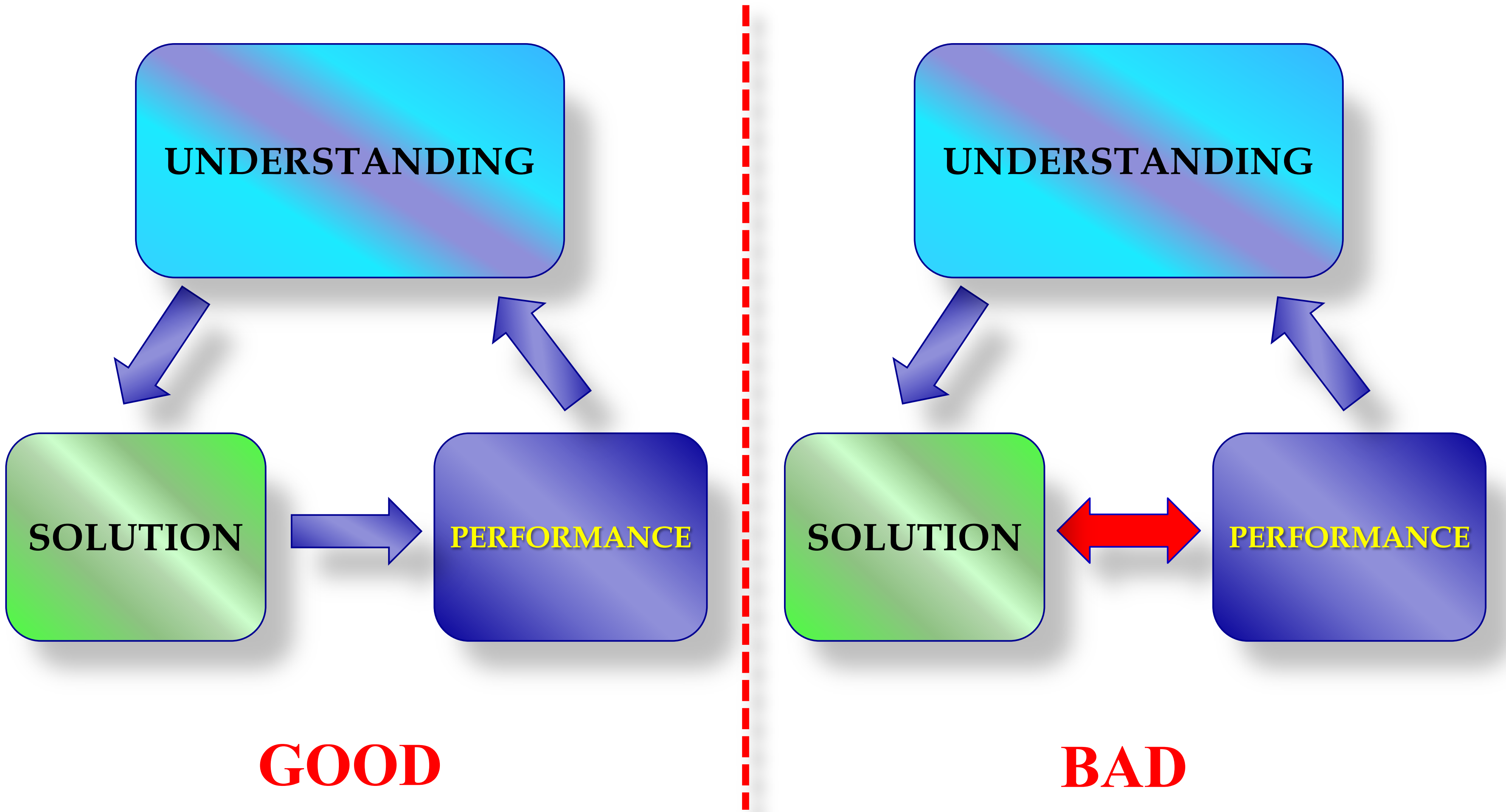
A.N. Tikhonov, "**On the stability of inverse problems**". Doklady Acad. Sci. USSR 39 (1943), 176–179.

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.

Good research or bad research?



Common mistakes

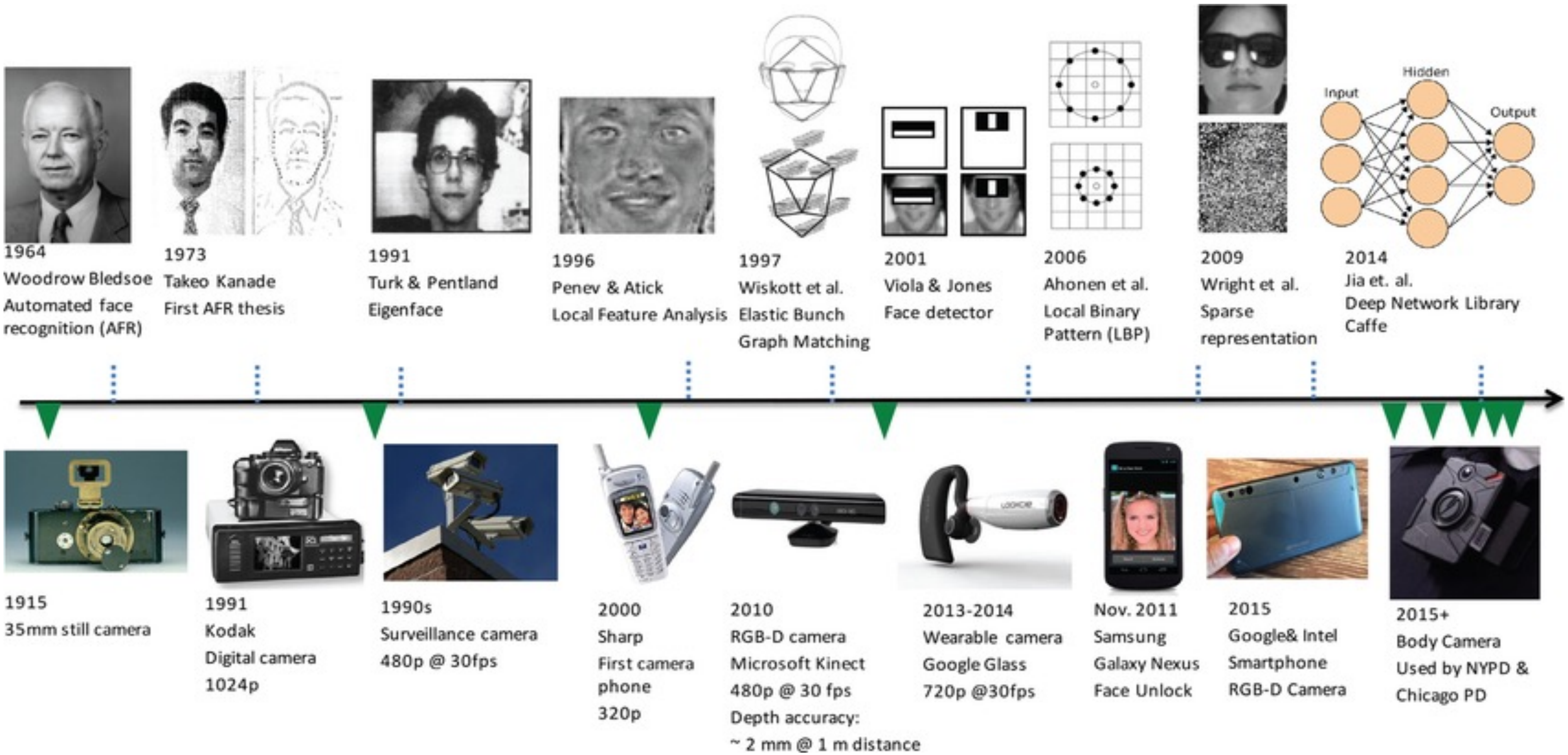
Vision Lab

1. Start **programming** before **thinking**.
2. Building a system **blindly** combining a number of already available algorithms.
3. Performing **blind tests** with available tools and datasets *(«Quick prototyping»*?).
4. Twickling the **parameters** until you obtain the **desired performance**.
5. Arbitrarily **selecting the data** from the available datasets **after** performing the initial testing.
6. Making **strong statements** without a solid proof.
7. Making **unrealistic assumptions**.

Addressing the problem

1. Analyze the **problem**, the available **data** and the **constraints**.
2. Make a **bibliographical search** (don't try to re-invent the wheel… one is enough).
3. Define a **model** describing the **physics** of the **event.**
4. Find a **mathematical framework** which may bring to a solution.
5. Carefully **design** an **experimental set-up**.
6. Collect or acquire a **statistically meaningful dataset**.
7. Start **programming**.
8. Perform an **evaluation test** to define the **parameters space**.
9. Start testing and collecting results, especially the **failing modes**.
10. Perform a **comparative analysis** of the results with other approaches at the ***current*** state of the art.
11. **Go back to item 3**.

Face recognition milestones

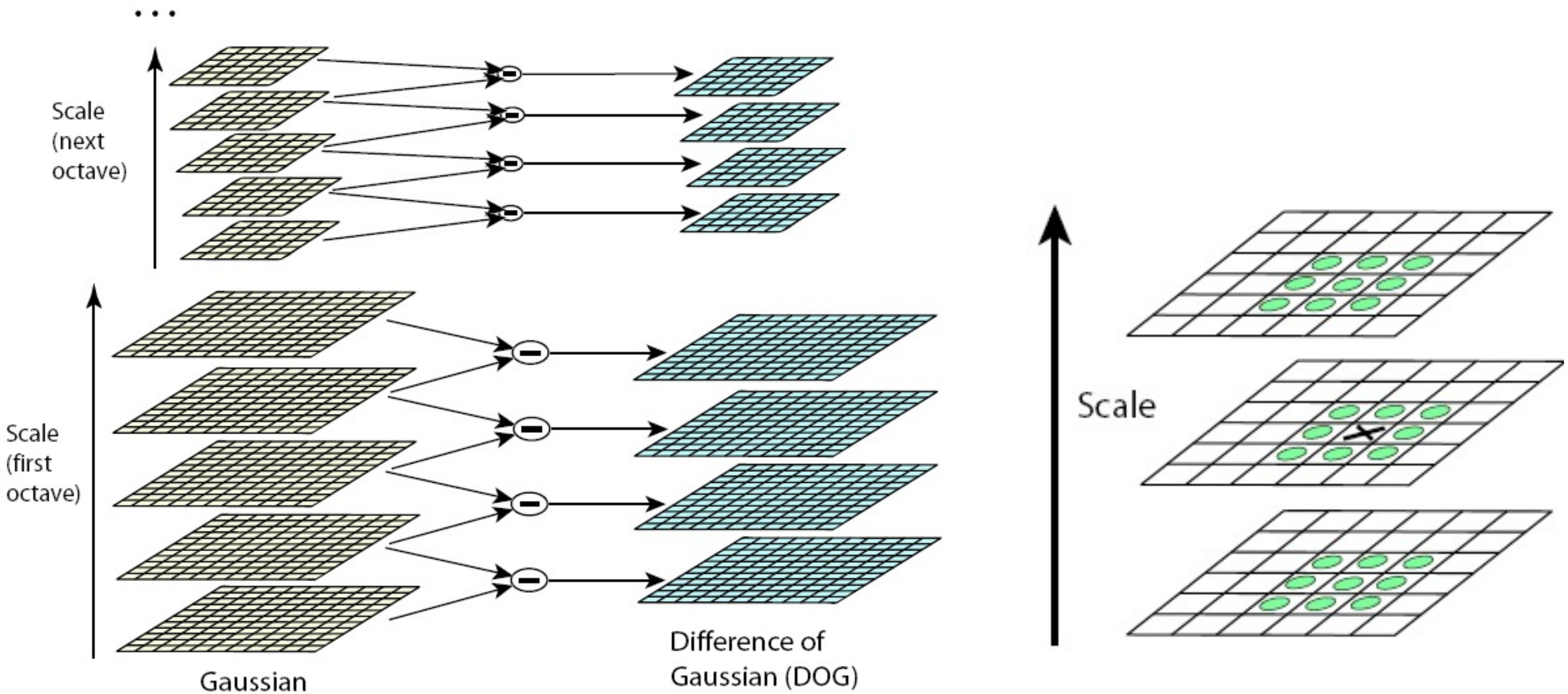


A. Jain, K, Nandakumar, A. Ross, "50 Years of Biometric Research: Accomplishments, Challenges, and Opportunities", Pattern Recognition Letters 79:80-105, 2016.

Scale Invariant Features

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

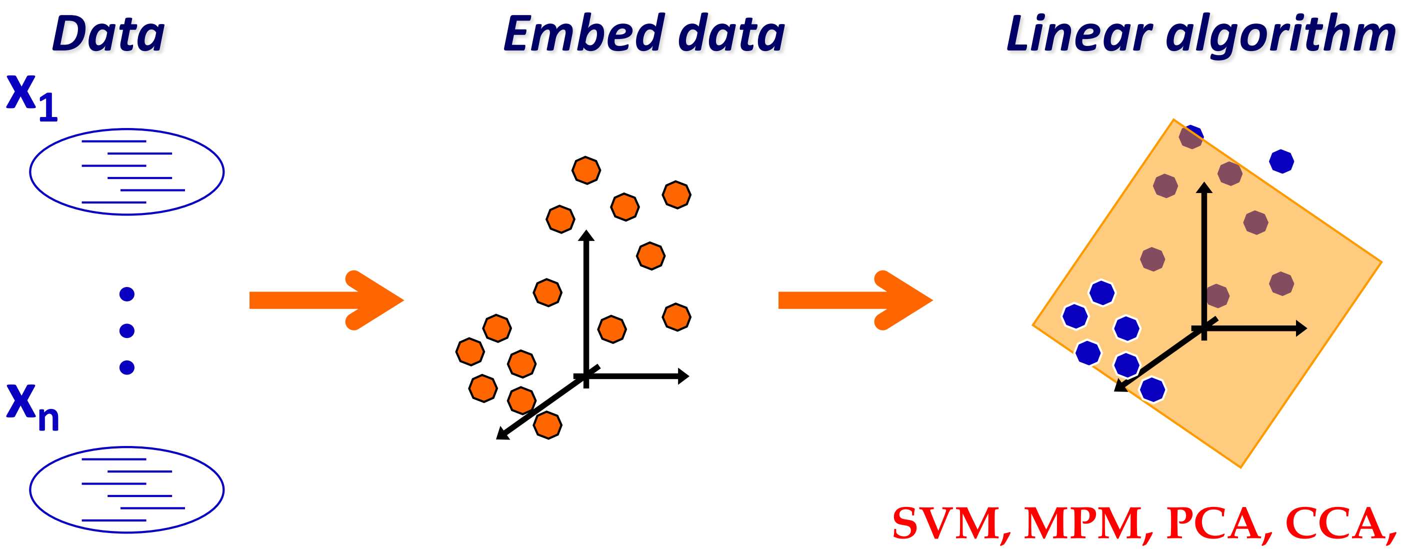
$$\mathbf{D}(x, y, \sigma, k) = \mathbf{L}(x, y, k\sigma) - \mathbf{L}(x, y, \sigma)$$



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

Kernel methods

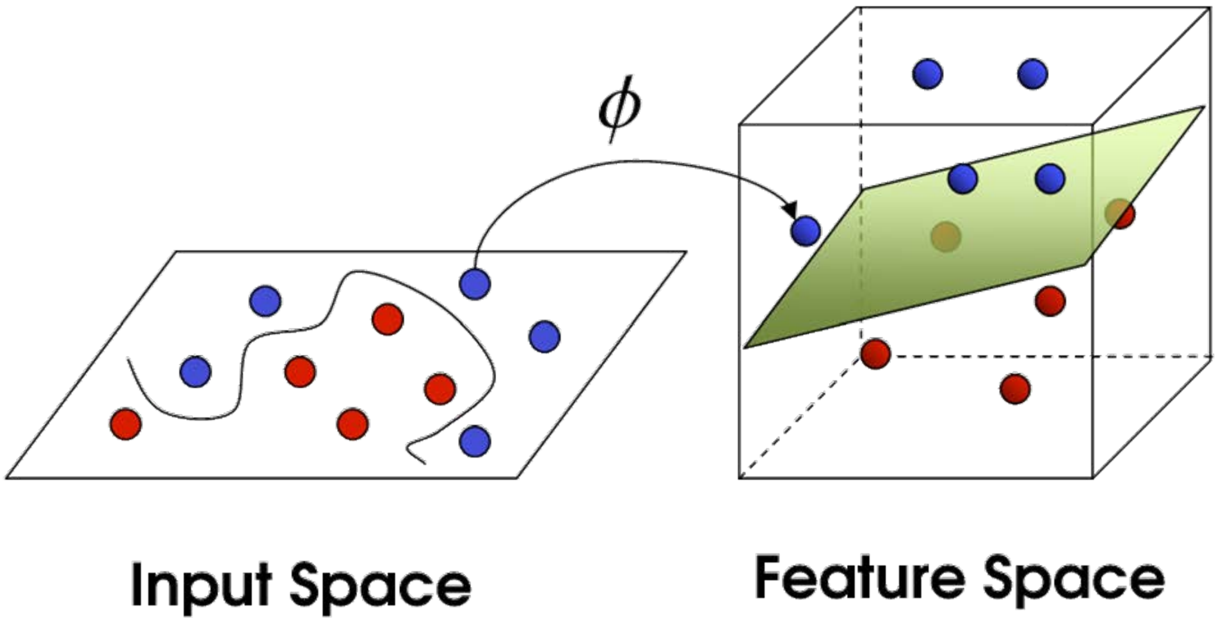
- **K-PCA**; **K-ICA**; **K-LDA**… (B. Schölkopf et al. 1998)
- Are all **variations of existing face-space representations**. The transformation is mediated by a **kernel function** such as Gaussian, polinomial, sigmoid and Radial Basis Functions.
- More **robust to noise and discretization** - Better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



SVM, MPM, PCA, CCA, FDA…

Support vectors

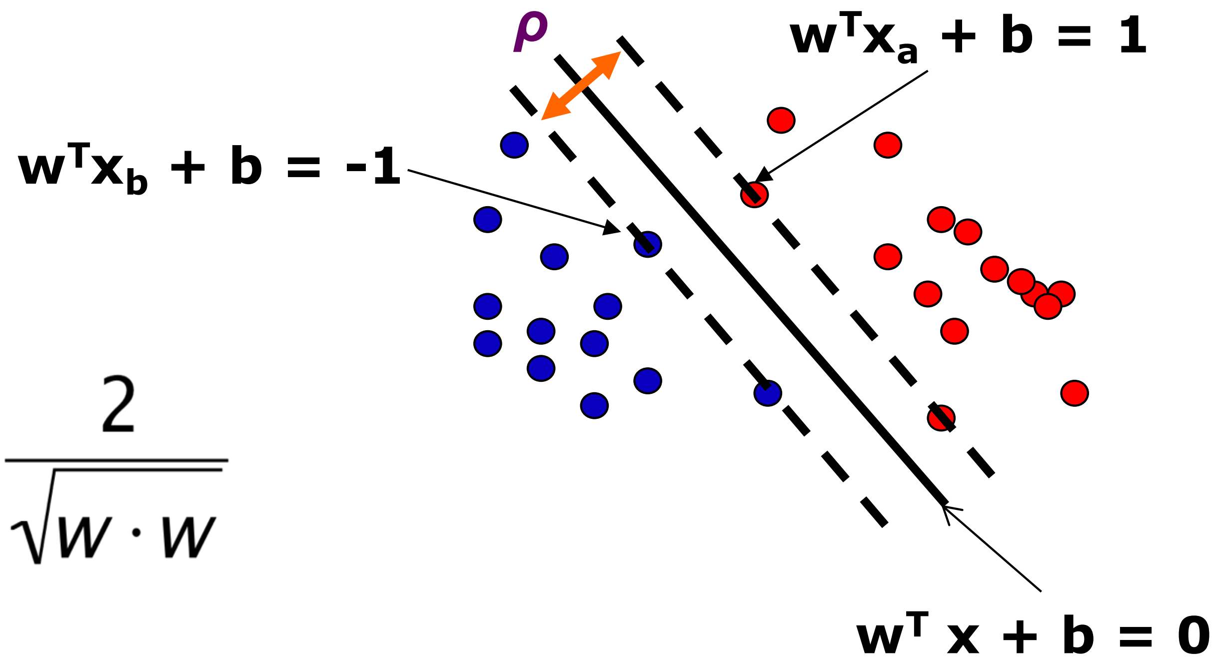
- Solves linearly separable problems
  1. **Data projection**: Input data are transformed **mapping into higher dimensions**



Support vectors

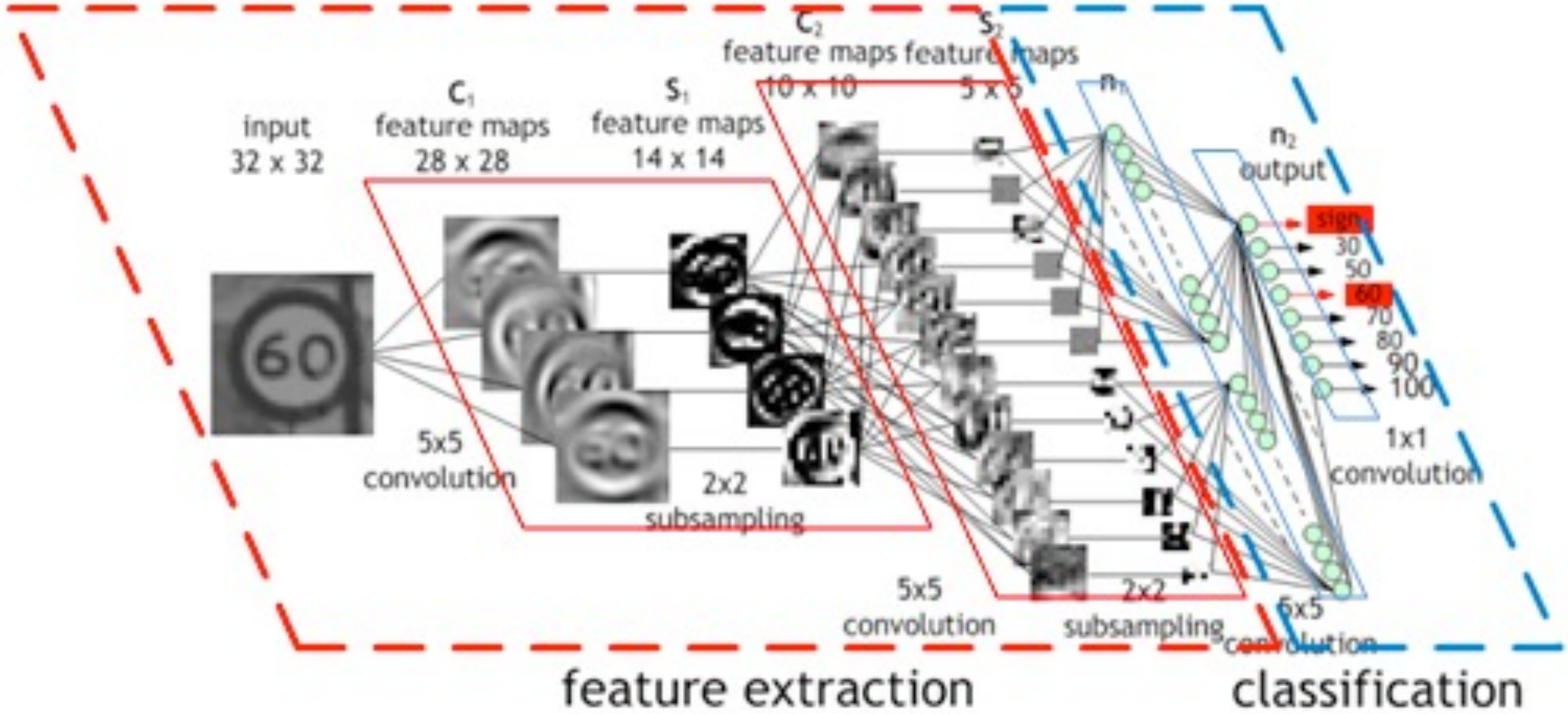
2. **Training**: find optimal **hyperplane** $\mathbf{w}^T\mathbf{x}_i + \mathbf{b} = 0$

margin maximisation $\min_{i=1,\dots,n} |\mathbf{w}^T\mathbf{x}_i + \mathbf{b}| = 1$

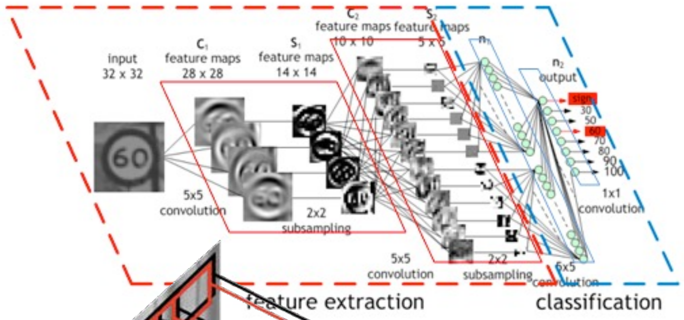


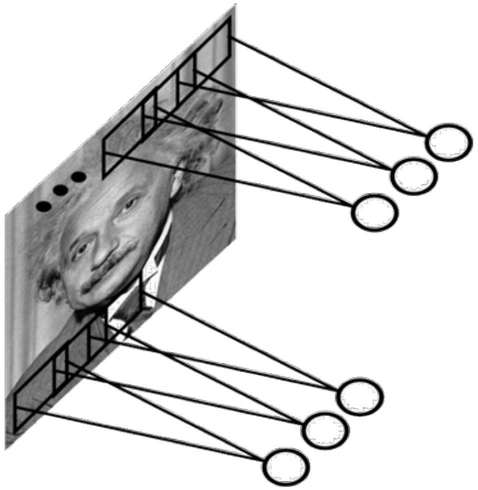
$$\rho(w) = \frac{2}{|w|} = \frac{2}{\sqrt{w \cdot w}}$$

Convolutional Neural Networks

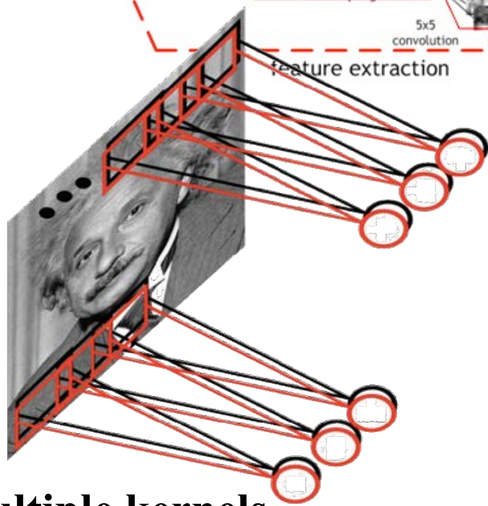


Convolutional Neural Networks

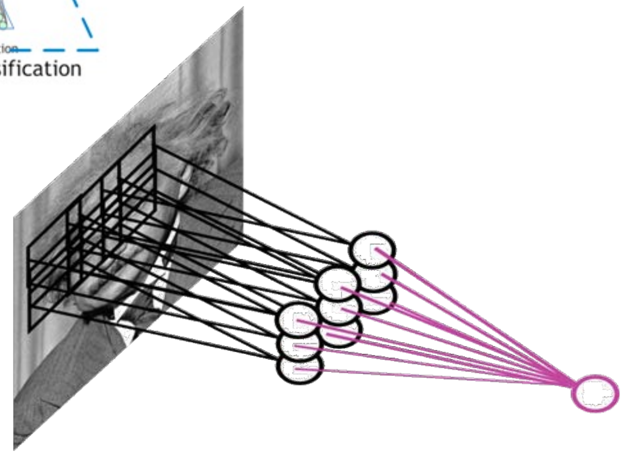




Single kernel Convolution



Multiple kernels Convolution



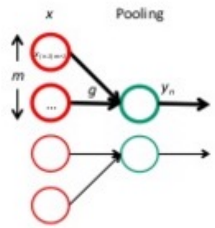
Spatial Pooling

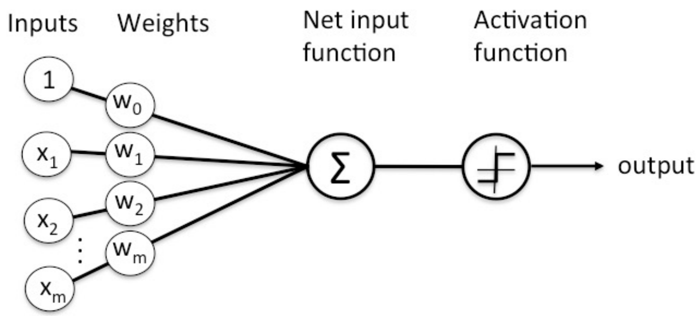
Let m be the size of pooling region, x be the input, and y be the output of the pooling layer. subsample(f, g)[n] denotes the n-th element of subsample(f, g).

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

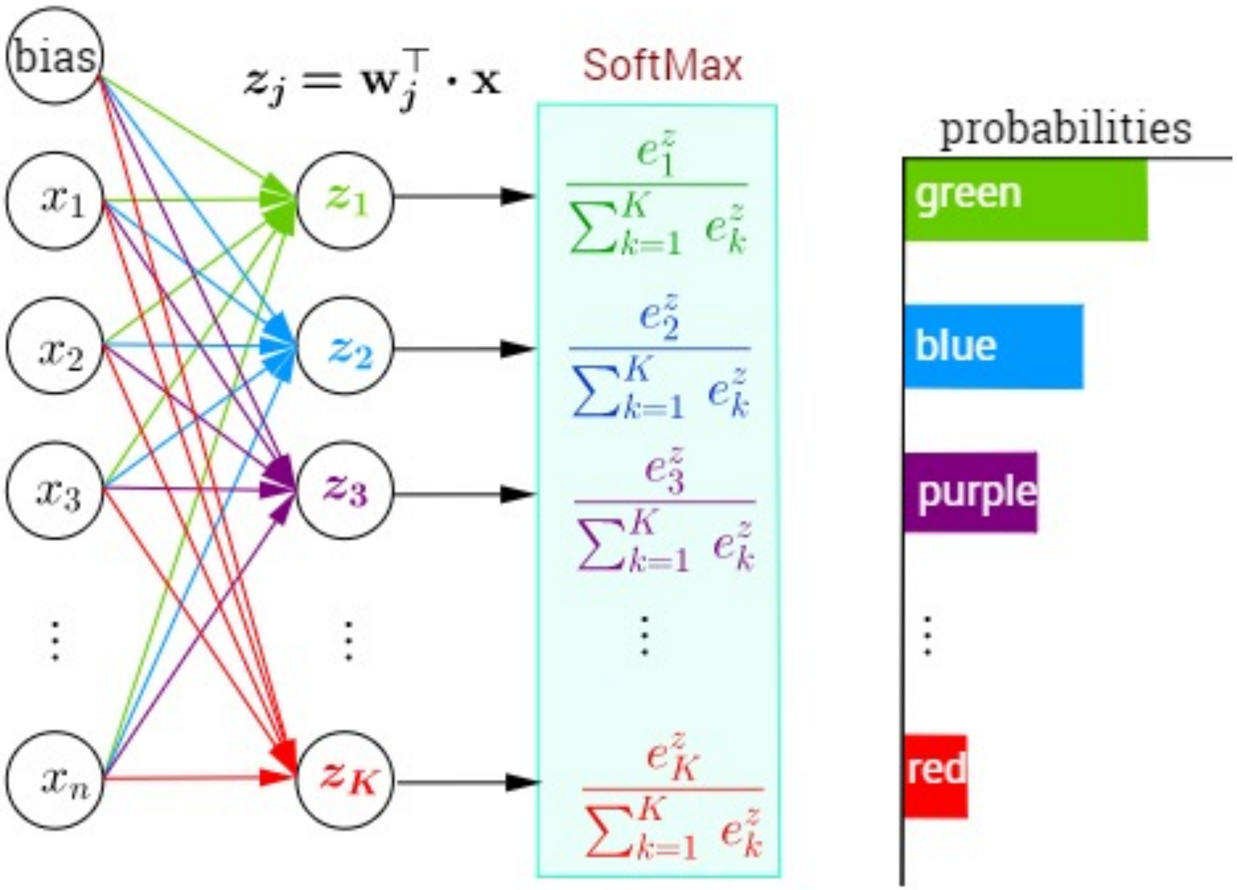
$$y = \mathrm{subsample}(x,g) = [y_n]$$

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





Convolutional Neural Networks




```
def softmax(X):
        exps = np.exp(X)
        return exps / np.sum(exps)
```

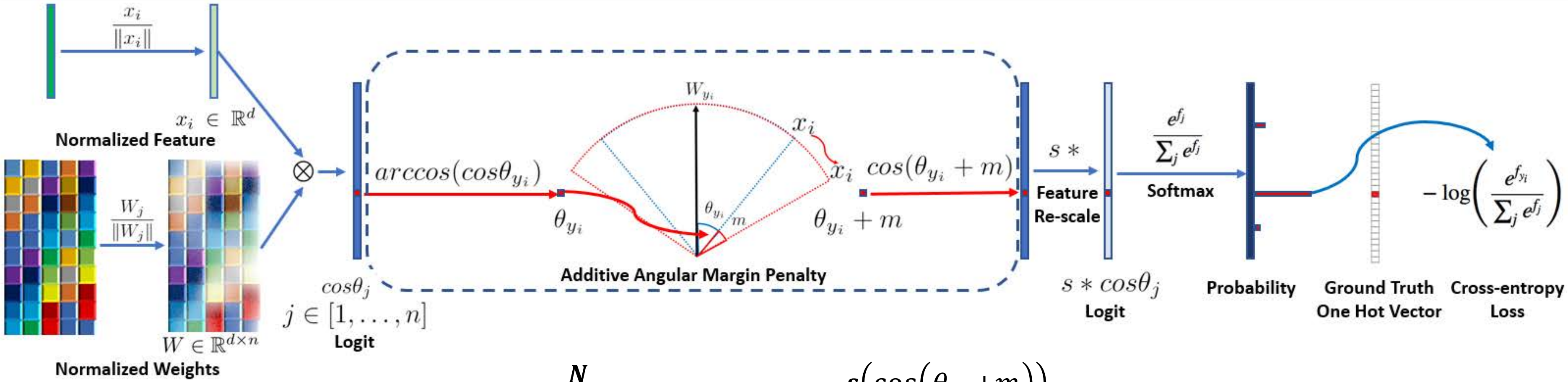
Convolutional Neural Networks

Cross entropy indicates the distance between what the model believes the output distribution should be, and what the original distribution really is:

$$\boldsymbol{H}(\boldsymbol{y},\boldsymbol{p}) = -\sum_{\boldsymbol{i}} \boldsymbol{y}_{\boldsymbol{i}} \log(\boldsymbol{p}_{\boldsymbol{i}})$$

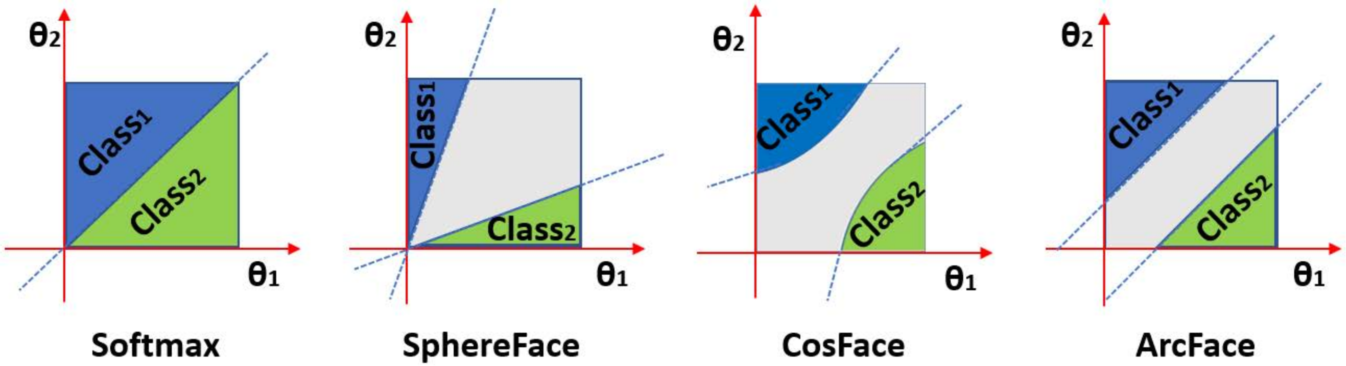
```
def cross_entropy(X,y):

""" X is the output from a fully connected layer (num_examples x num_classes)
y is labels (num_examples x 1)
Note that y is not one-hot encoded vector. It can be computed as y.argmax(axis=1) from one-hot encoded vectors of labels if required.
"""
    m = y.shape[0]          # We use multidimensional array indexing to extract
    p = softmax(X)          # softmax probability of the correct label for each sample.

    log_likelihood = -np.log(p[range(m),y])
    loss = np.sum(log_likelihood) / m
    return loss
```

Loss functions



$$L(s) = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}}{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}+\sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}$$

θ_j is the angle between the weight W_j and the feature x_i ; $s = \|x_i\|$



Deng J, Guo J, Yang J, Xue N, Cotsia I, Zafeiriou SP. **ArcFace: Additive Angular Margin Loss for Deep Face Recognition**. IEEE Trans PAMI. 2021 Jun 9; doi: 10.1109/TPAMI.2021.3087709. https://github.com/deepinsight/insightface

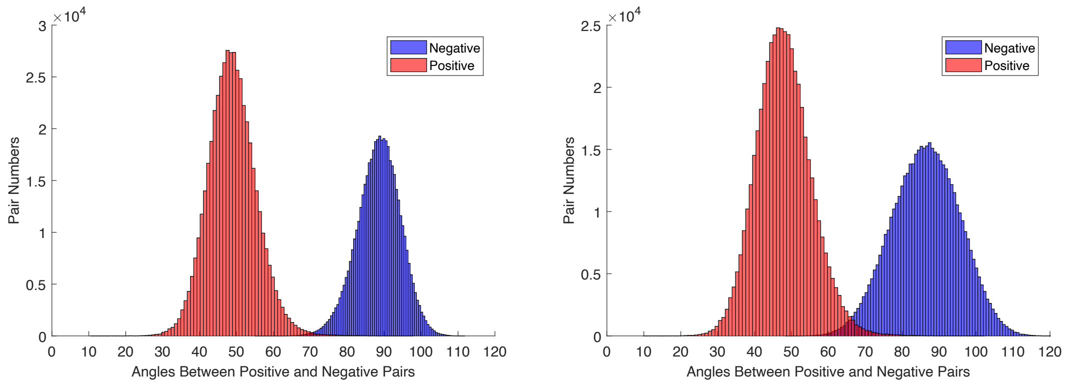
Loss functions

Vision Lab

Loss Functions	LFW	CFP-FP	AgeDB-30
ArcFace (0.4)	99.53	95.41	94.98
ArcFace (0.45)	99.46	95.47	94.93
ArcFace (0.5)	**99.53**	**95.56**	**95.15**
ArcFace (0.55)	99.41	95.32	95.05
SphereFace [18]	99.42	-	-
SphereFace (1.35)	99.11	94.38	91.70
CosFace [37]	99.33	-	-
CosFace (0.35)	99.51	95.44	94.56
CM1 (1, 0.3, 0.2)	99.48	95.12	94.38
CM2 (0.9, 0.4, 0.15)	99.50	95.24	94.86
Softmax	99.08	94.39	92.33
Norm-Softmax (NS)	98.56	89.79	88.72
NS+Intra	98.75	93.81	90.92
NS+Inter	98.68	90.67	89.50
NS+Intra+Inter	98.73	94.00	91.41
Triplet (0.35)	98.98	91.90	89.98
ArcFace+Intra	99.45	95.37	94.73
ArcFace+Inter	99.43	95.25	94.55
ArcFace+Intra+Inter	99.43	95.42	95.10
ArcFace+Triplet	99.50	95.51	94.40

Table 2. Verification results (%) of different loss functions ([CASIA, ResNet50, loss*]).

Method	#Image	LFW	YTF
DeepID [32]	0.2M	99.47	93.20
Deep Face [33]	4.4M	97.35	91.4
VGG Face [24]	2.6M	98.95	97.30
FaceNet [29]	200M	99.63	95.10
Baidu [16]	1.3M	99.13	-
Center Loss [38]	0.7M	99.28	94.9
Range Loss [46]	5M	99.52	93.70
Marginal Loss [9]	3.8M	99.48	95.98
SphereFace [18]	0.5M	99.42	95.0
SphereFace+ [17]	0.5M	99.47	-
CosFace [37]	5M	99.73	97.6
MS1MV2, R100, ArcFace	5.8M	**99.83**	**98.02**

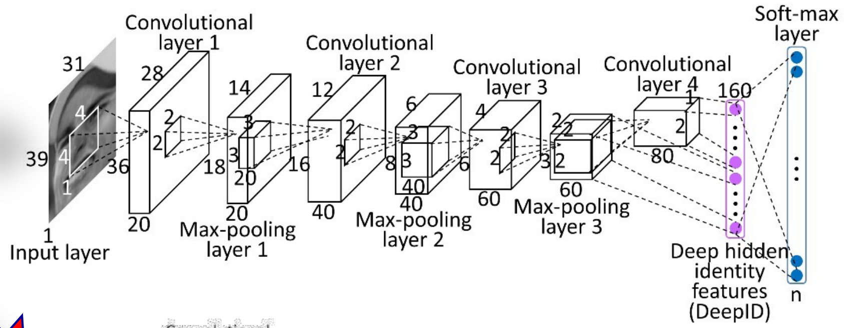


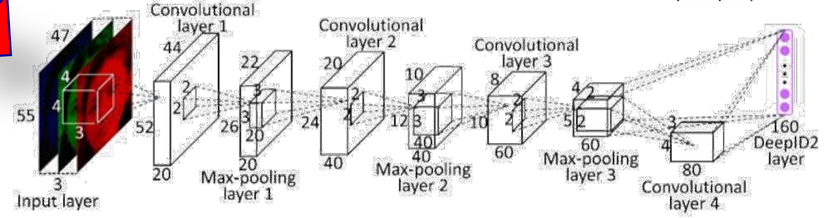
(a) ArcFace (b) Triplet-Loss

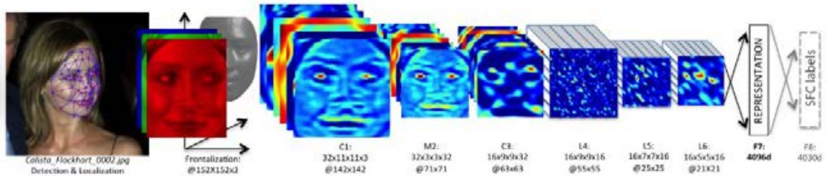
Deng J, Guo J, Yang J, Xue N, Cotsia I, Zafeiriou SP. **ArcFace: Additive Angular Margin Loss for Deep Face Recognition**. IEEE Trans PAMI. 2021 Jun 9; doi: 10.1109/TPAMI.2021.3087709. https://github.com/deepinsight/insightface

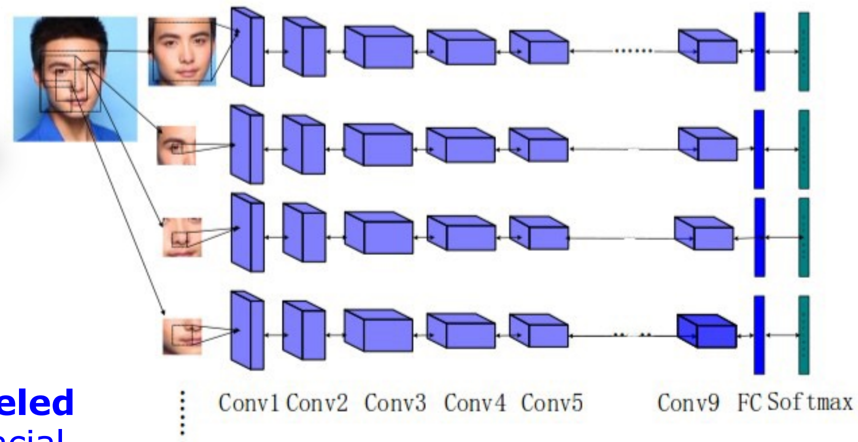
State of the art

- DeepID (Y. Sun, X.Wang, X. Tang – CVPR 2014)
- DeepID2 (Y. Sun, X.Wang, X. Tang - NIPS 2014)
- DeepID2+
- DeepID3
- DeepFace (Y. Taigman, M. Yang, M. Ranzato, L. Wolf CVPR 2015)
- Face++; FaceNet
- VGG (M. Parkhi, A. Vedaldi, A. Zissermann - BMVC 2015)
- Baidu (J.Liu, Y.Deng, T.Bai, Z.Wei, C.Huang CVPR 2015)
- GANs, ArcFace, ResNet… **What’s next?**









E. Learned-Miller, G. Huang, A. RoyChowdhury, H. Li, G. Hua, “**Labeled Faces in the Wild: A Survey**”, Advances in Face Detection and Facial Image Analysis, pp 189-248, Springer 2016.

State of the art

Vision Lab

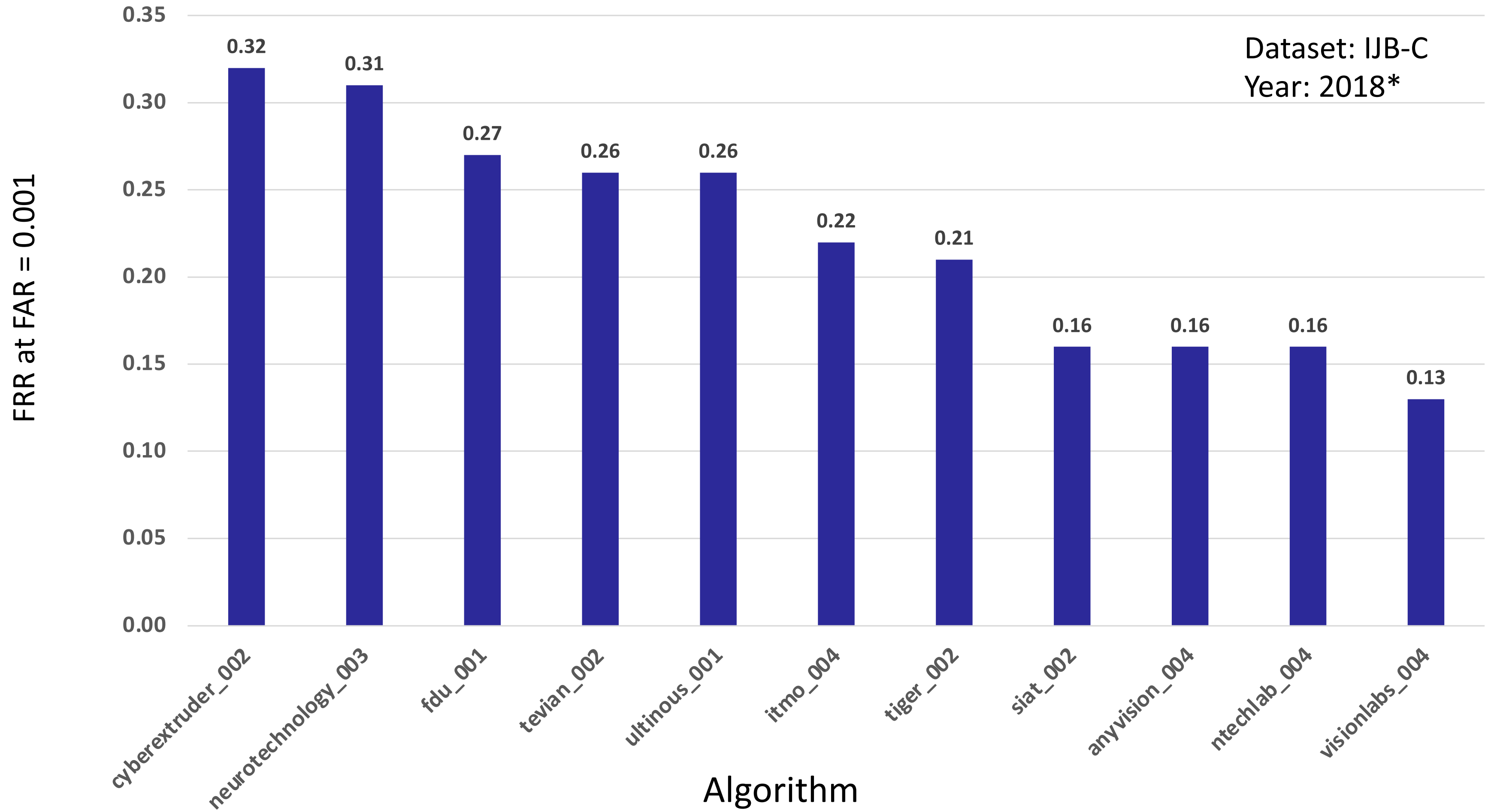
Dataset	Available	#Photos and #people
LFW	Public	13K of 5K people
CelebFaces 2014	Private	202K of 10K people
CASIA-WebFace 2014	Public	500K of 10K people
FaceScrub 2014	Public	100K of 500 people
YouTube Faces	Public	3425 videos of 1595 people
DeepFace (Facebook) 2014	Private	4.4 Million of 4K people
FaceNet (Google) 2015	Private	100-200 Million of 8M people
MegaFace	**Public**	**1 Million**

Figure 2: Representative sample of face recognition datasets that were created in the recent years (in addition to LFW). All the public datasets are small scale, and all the large scale datasets are mainly used for training rather than testing and are not publicly available. MegaFace (this paper) is the first large scale unconstrained dataset. It is collected from Flickr and will be available publicly.

Miller et al. (2015) Mega-Face: A million faces for recognition at scale.

State of the art

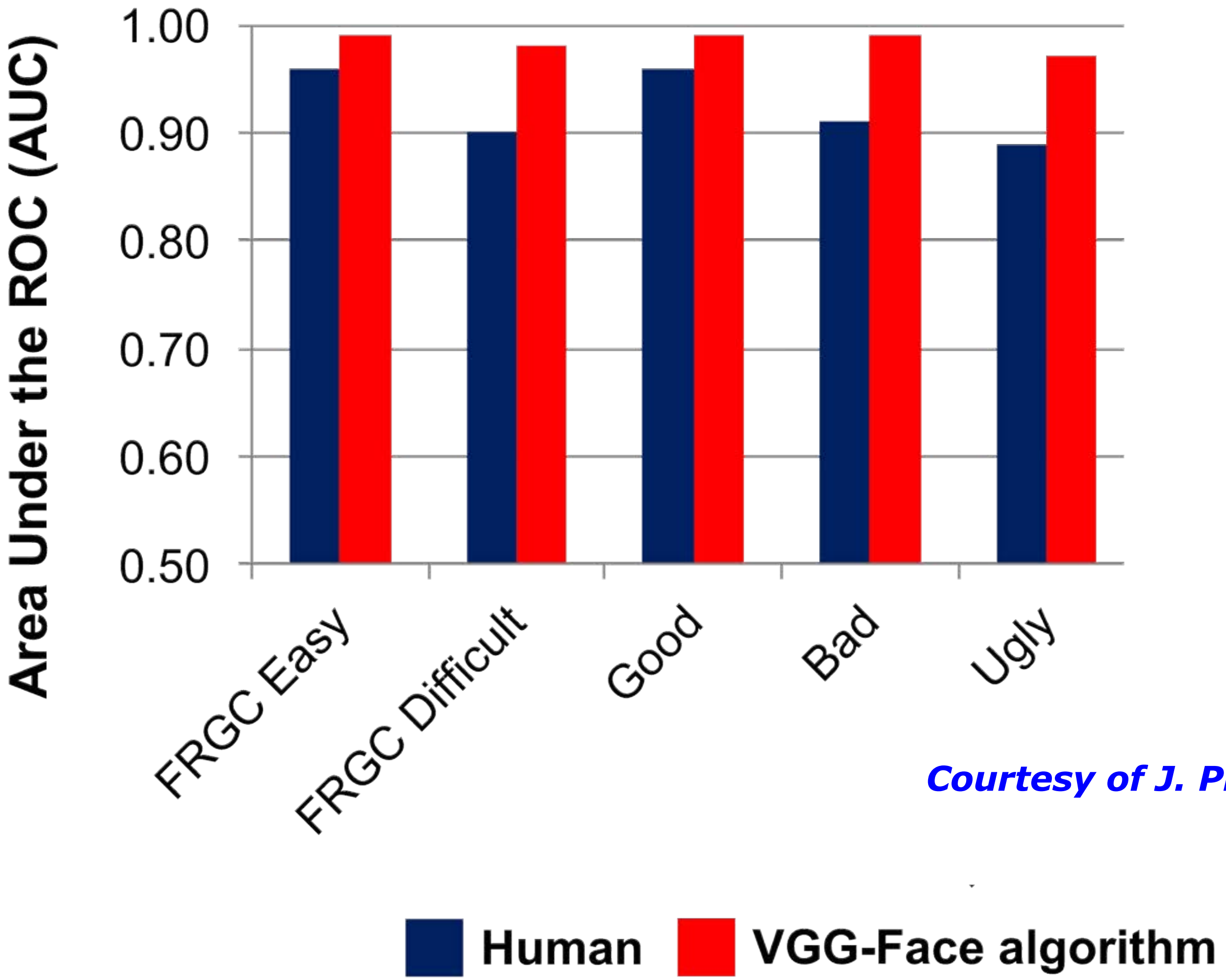
FRVT 1:1 Wild-to-wild comparisons



Courtesy of J. Phillips (2021)

Face Recognition Performance

❖ **How do machines vs humans perform**

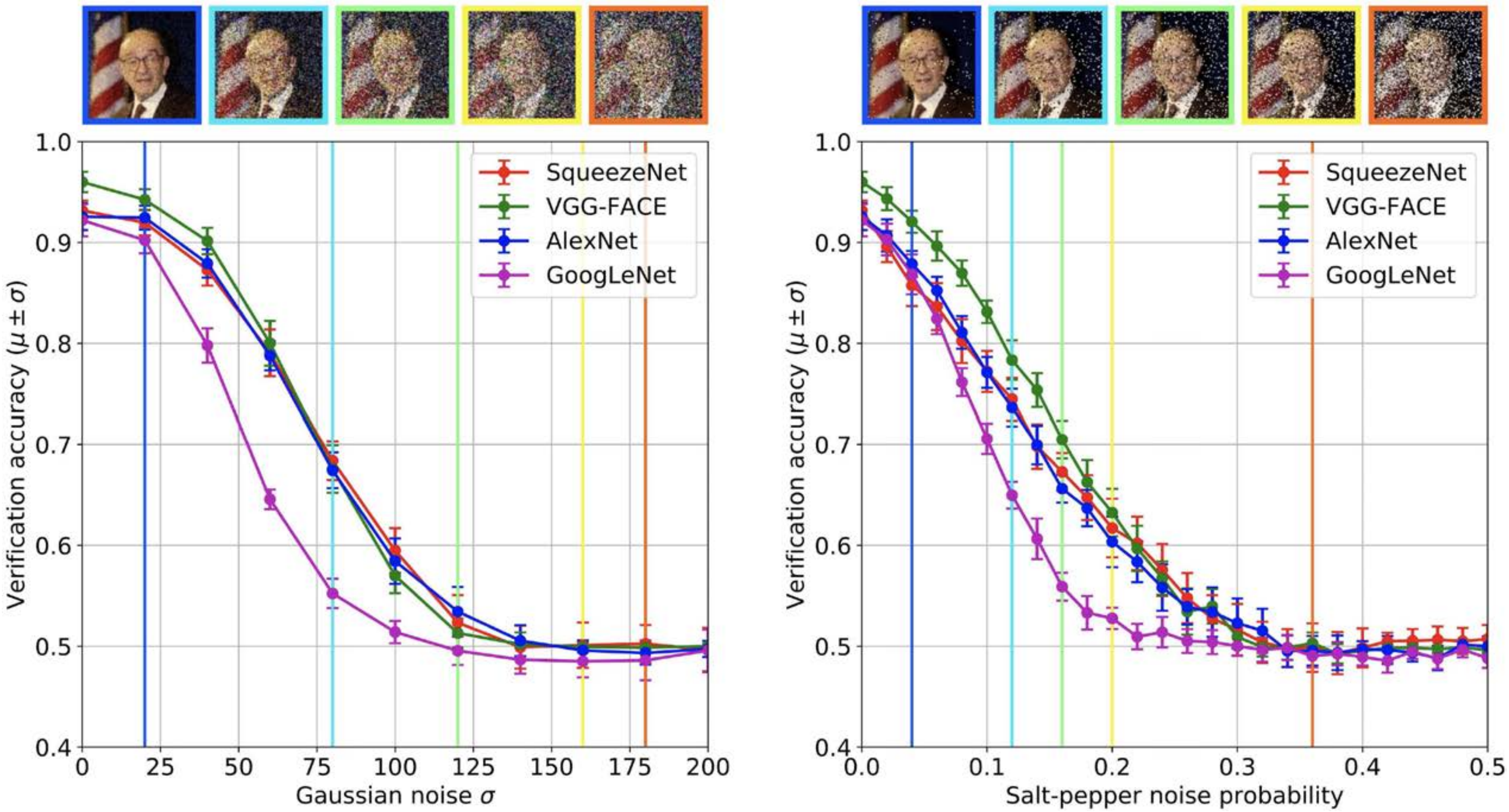


Courtesy of J. Phillips et al. (2018)

CNN Performance

Vision Lab

- **However, we're not done yet...**

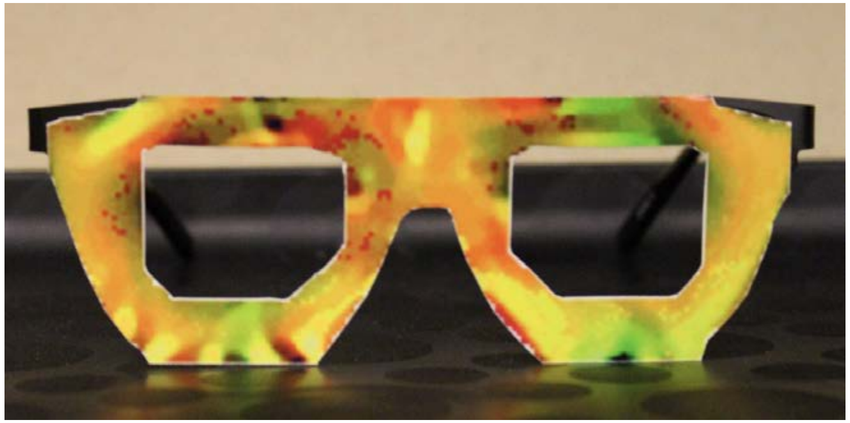


K. Grm , V. Štruc, A. Artiges, M. Caron, H. K. Ekenel, "**Strengths and weaknesses of deep learning models for face recognition against image degradations**" IET Biometrics, 7(1):81-89, 2018

CNN Performance

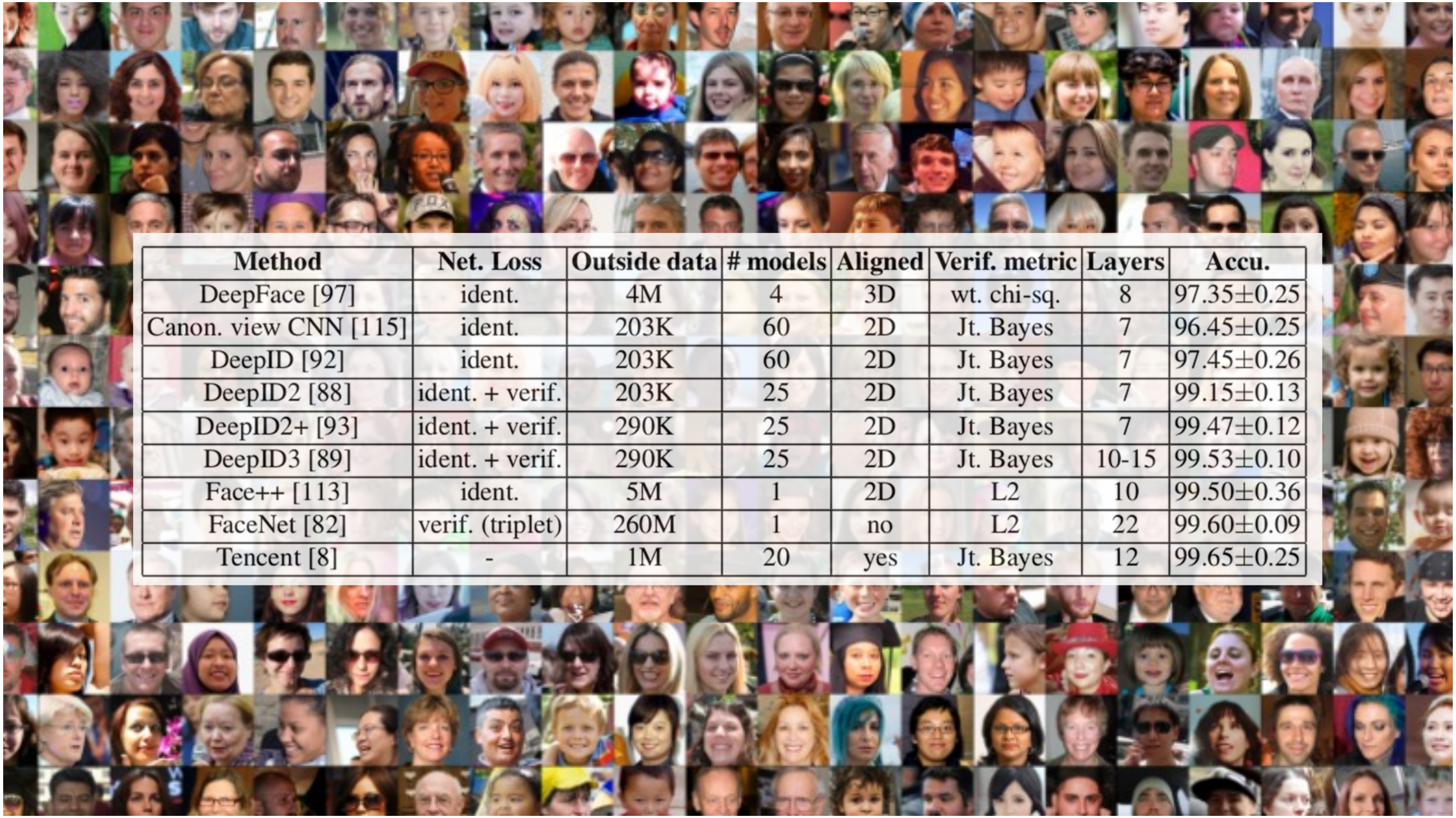
Vision Lab

- 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"



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 concerns

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

CNNs: Where are we going?

International Journal of Computer Vision (2021) 129:781–802
https://doi.org/10.1007/s11263-020-01405-z

Check for updates

Deep Nets: What have They Ever Done for Vision?

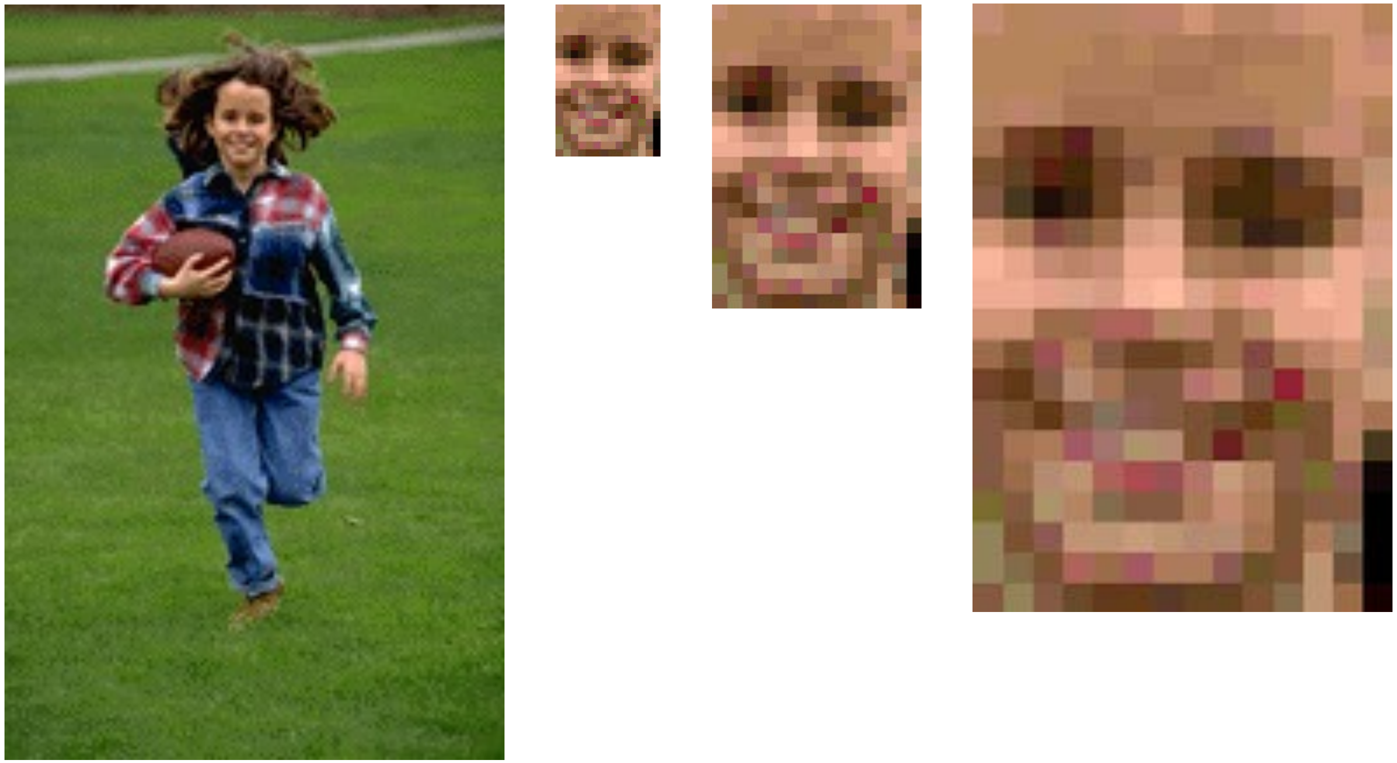
Alan L. Yuille[1] · Chenxi Liu[1]

Received: 10 January 2019 / Accepted: 9 November 2020 / Published online: 27 November 2020

Abstract
This is an opinion paper about the strengths and weaknesses of Deep Nets for vision. They are at the heart of the enormous recent progress in artificial intelligence and are of growing importance in cognitive science and neuroscience. They have had many successes but also have several limitations and there is limited understanding of their inner workings. At present Deep Nets perform very well on specific visual tasks with benchmark datasets but they are much less general purpose, flexible, and adaptive than the human visual system. We argue that Deep Nets in their current form are unlikely to be able to overcome the fundamental problem of computer vision, namely how to deal with the combinatorial explosion, caused by the enormous complexity of natural images, and obtain the rich understanding of visual scenes that the human visual achieves. We argue that **this combinatorial explosion takes us into a regime where "big data is not enough" and where we need to rethink our methods for benchmarking performance and evaluating vision algorithms**. We stress that, as vision algorithms are increasingly used in real world applications, that performance evaluation is not merely an academic exercise but has important consequences in the real world. It is impractical to review the entire Deep Net literature so we restrict ourselves to a limited range of topics and references which are intended as entry points into the literature. The views expressed in this paper are our own and do not necessarily represent those of anybody else in the computer vision community.

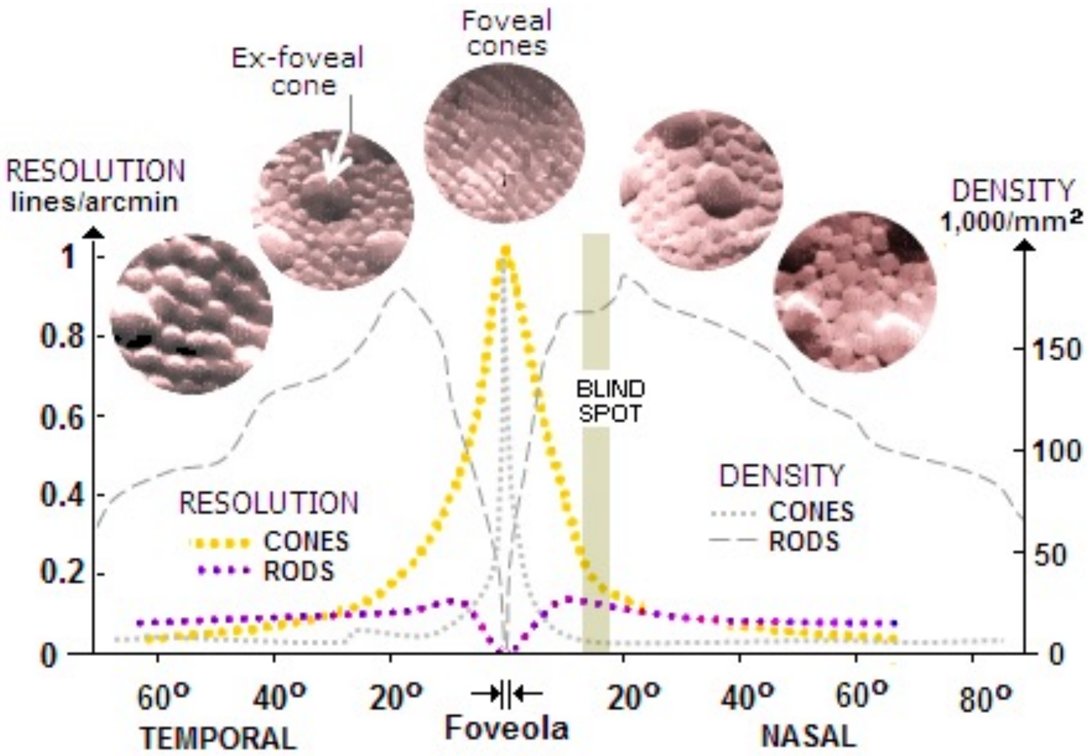
Keywords Deep neural networks · Computer vision · Success · Limitation · Cognitive science · Neuroscience

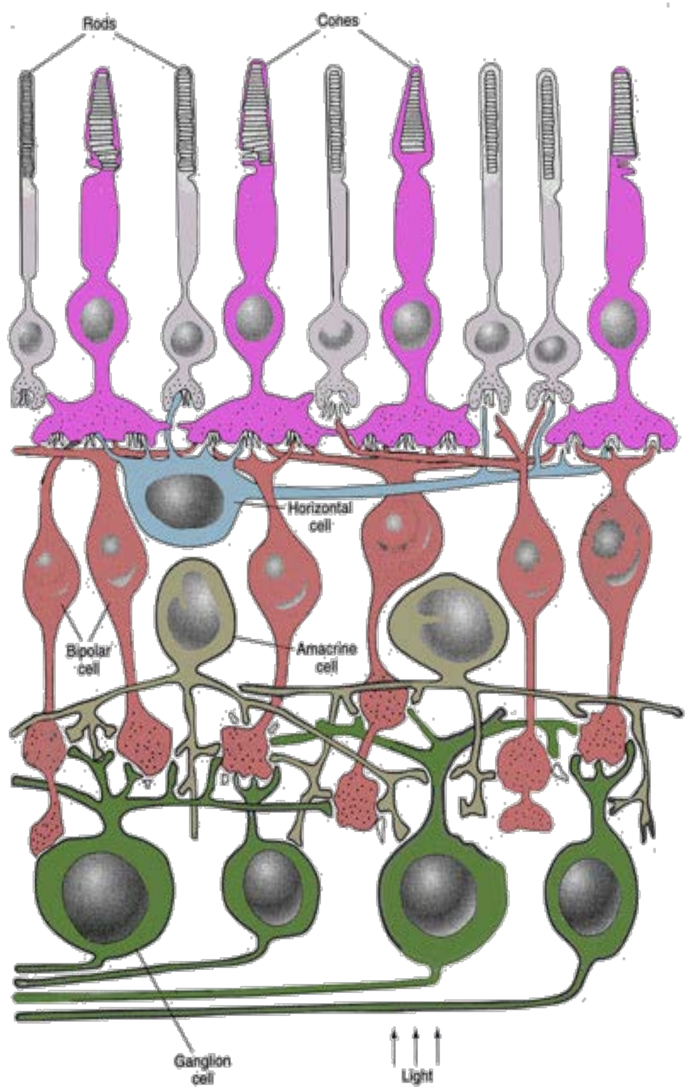
A different "perspective"



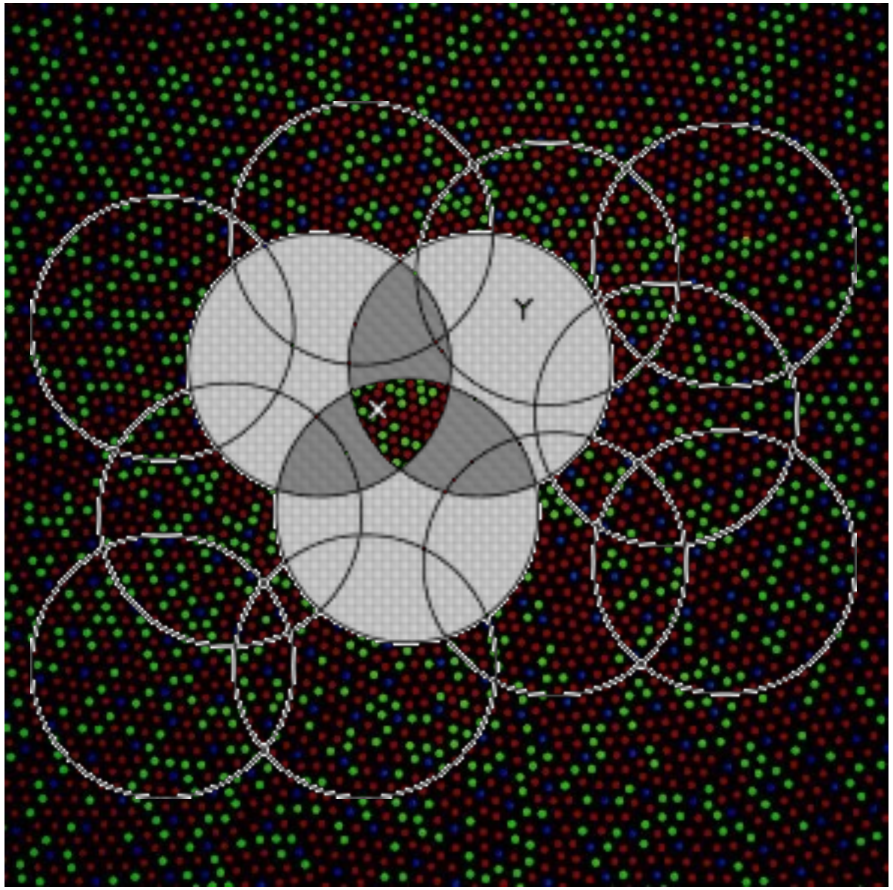
Spatial distribution and **Frequency tuning**

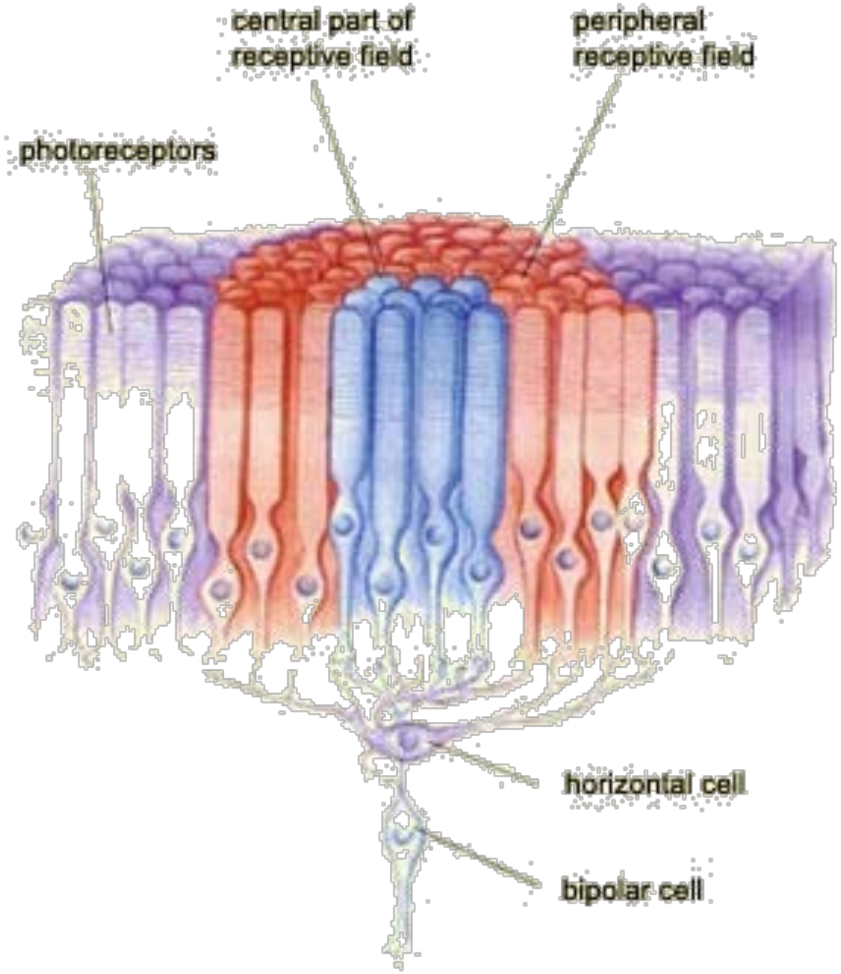
The human retina





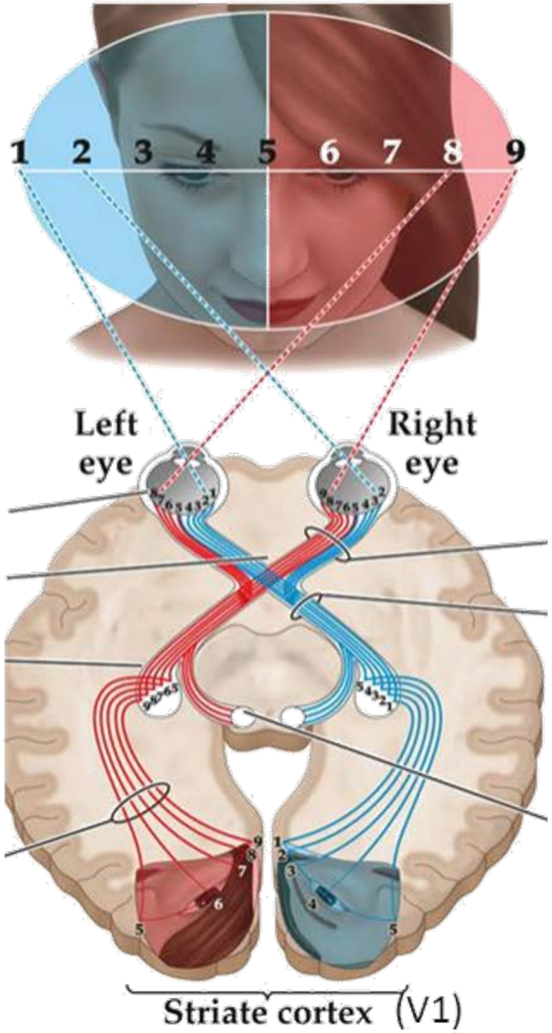
Receptive fields

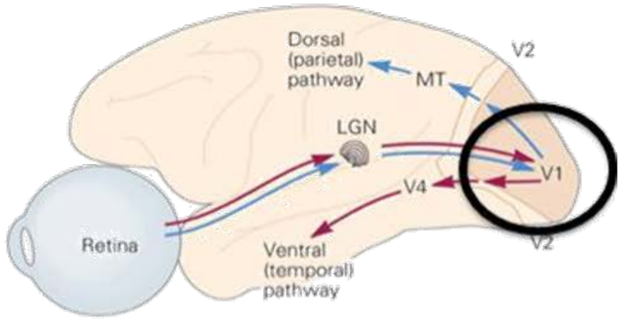




Retinotopic mapping

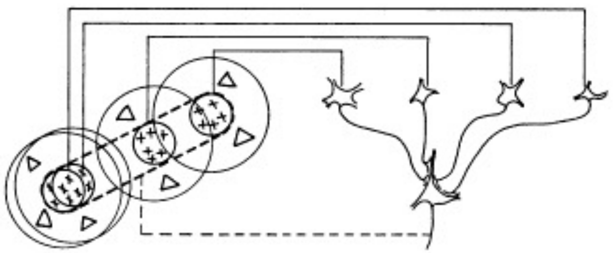
V1 retinotopic maps



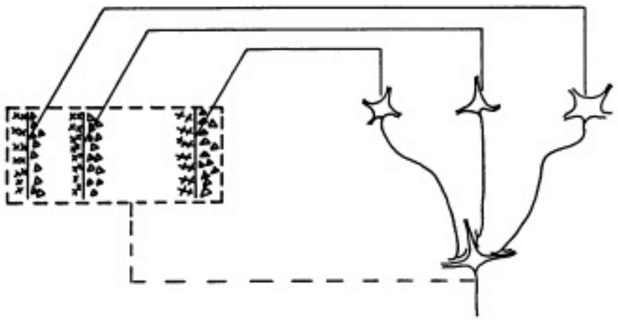


- Each point of the visual field maps on to a local group of neurons in V1.
- Retinotopy = Remapping of retinal image onto cortical surface
- Foveal region uses more of V1 (greater magnification factor)

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.

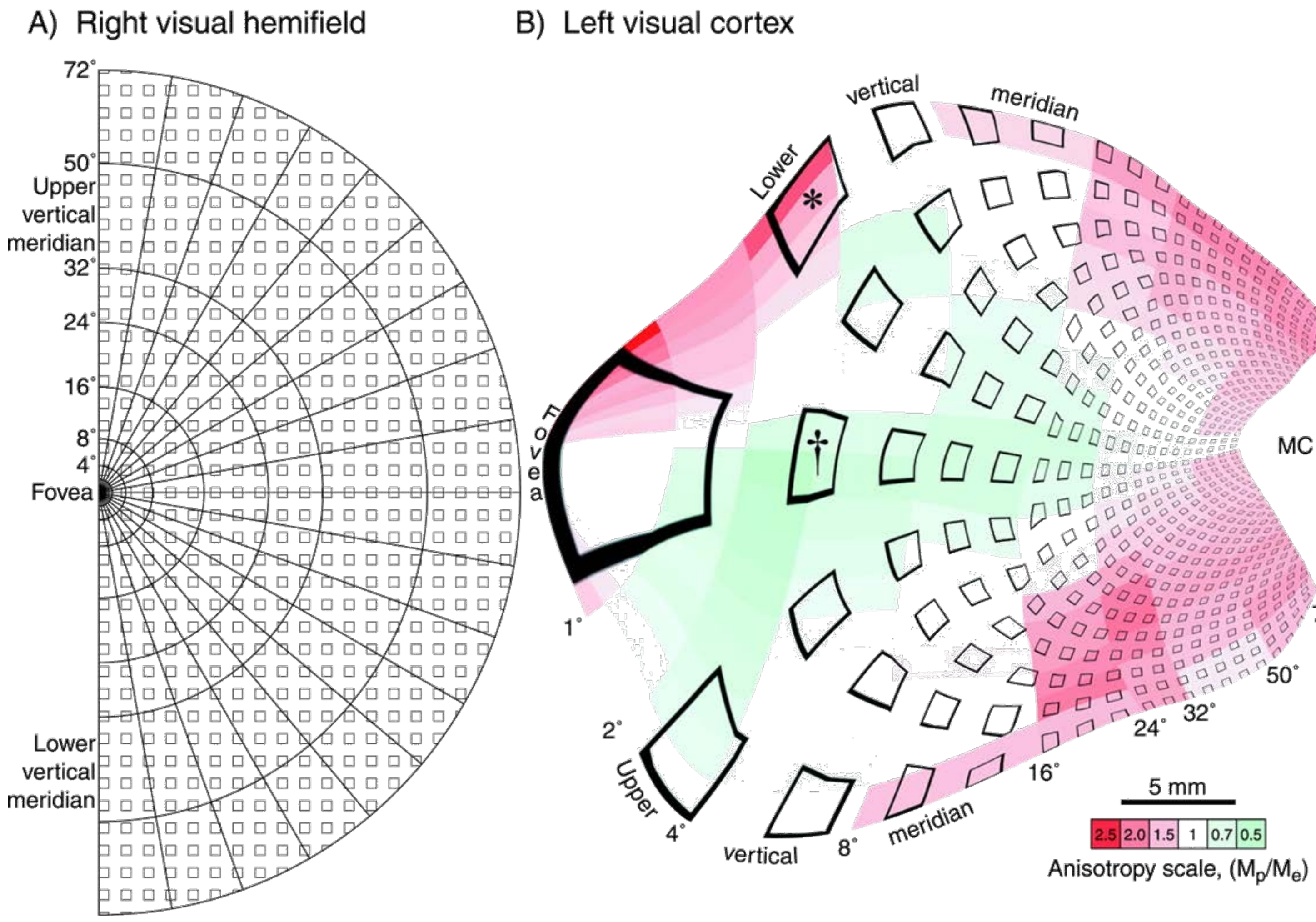


Text-fig. 20. Possible scheme for explaining the organization of complex receptive fields. A number of cells with simple fields, of which three are shown schematically, are imagined to project to a single cortical cell of higher order. Each projecting neurone has a receptive field arranged as shown to 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.

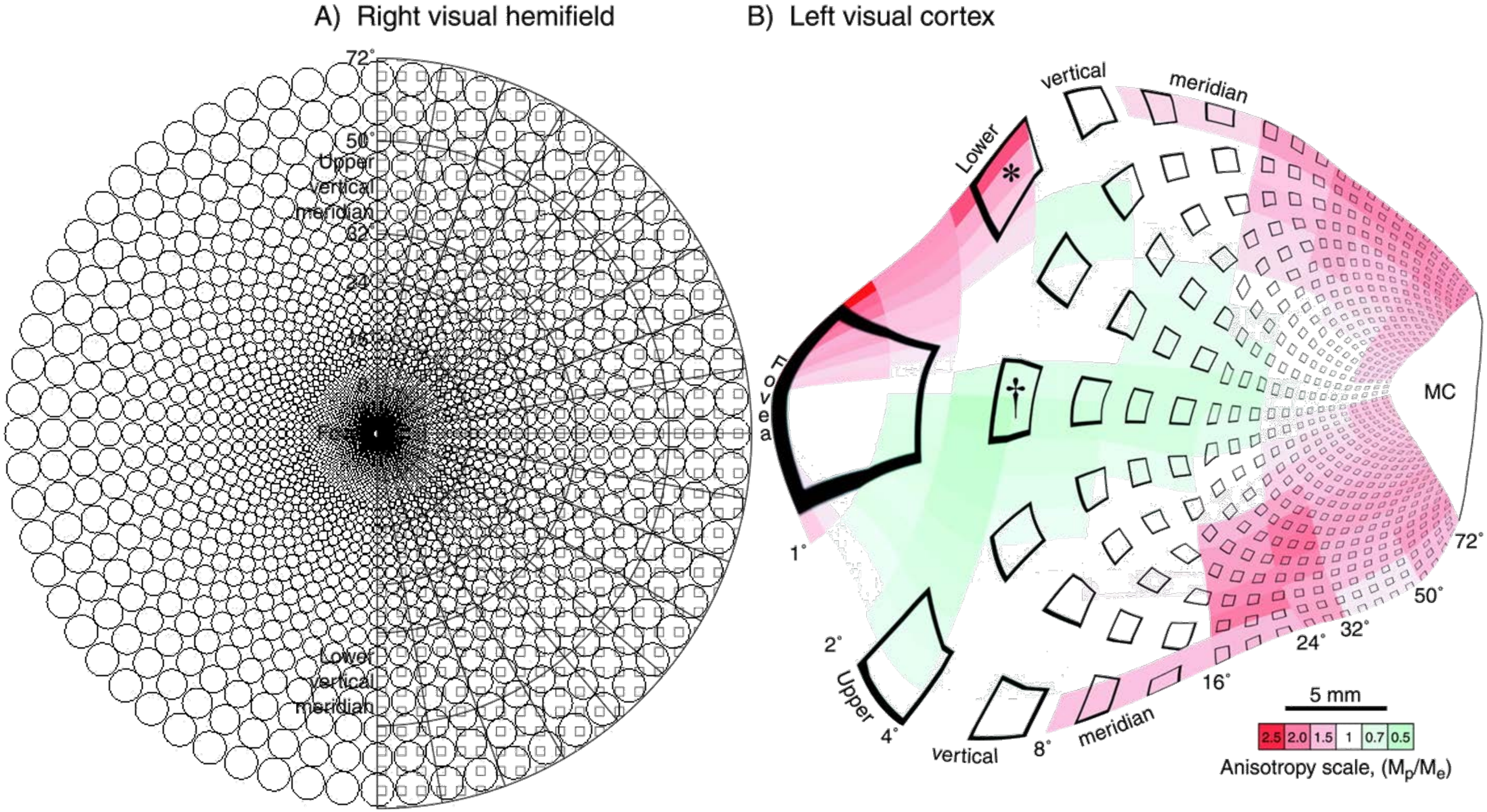
Simple and Complex cells

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

Retinotopic mapping



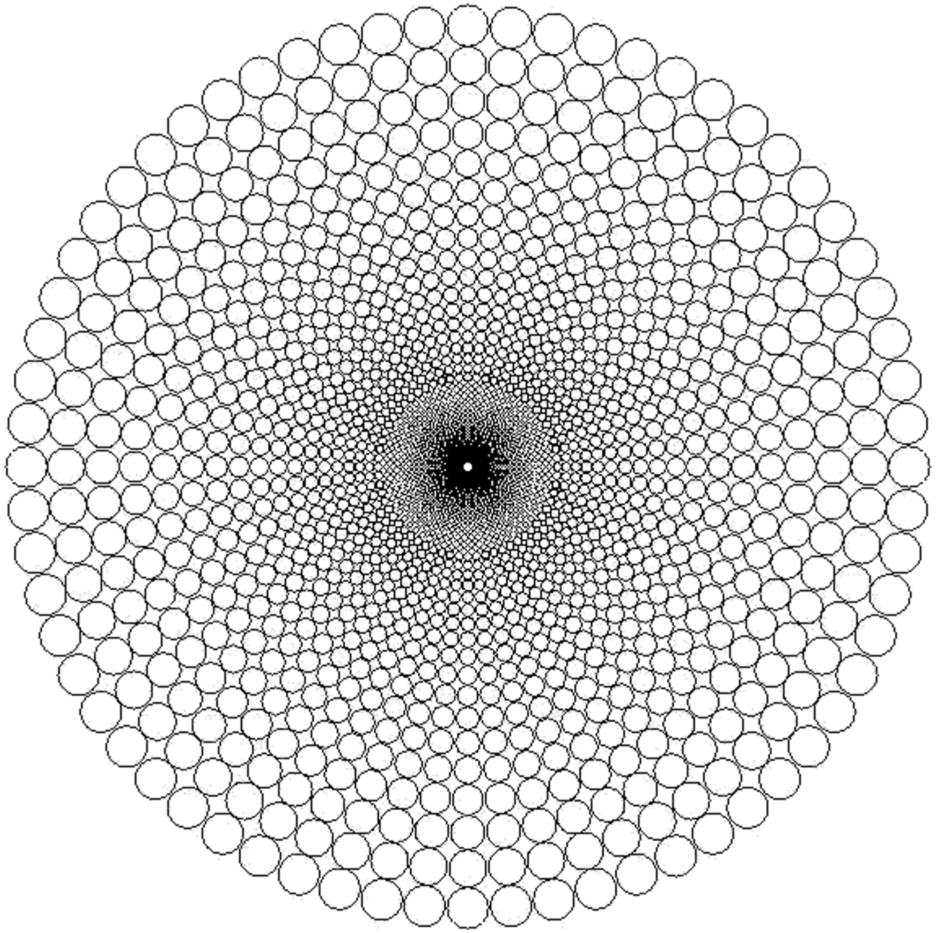
Retinotopic mapping



Log-Polar mapping

Vision Lab

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



$$\begin{cases} x = \rho \sin\theta \\ y = \rho \cos\theta \end{cases}$$

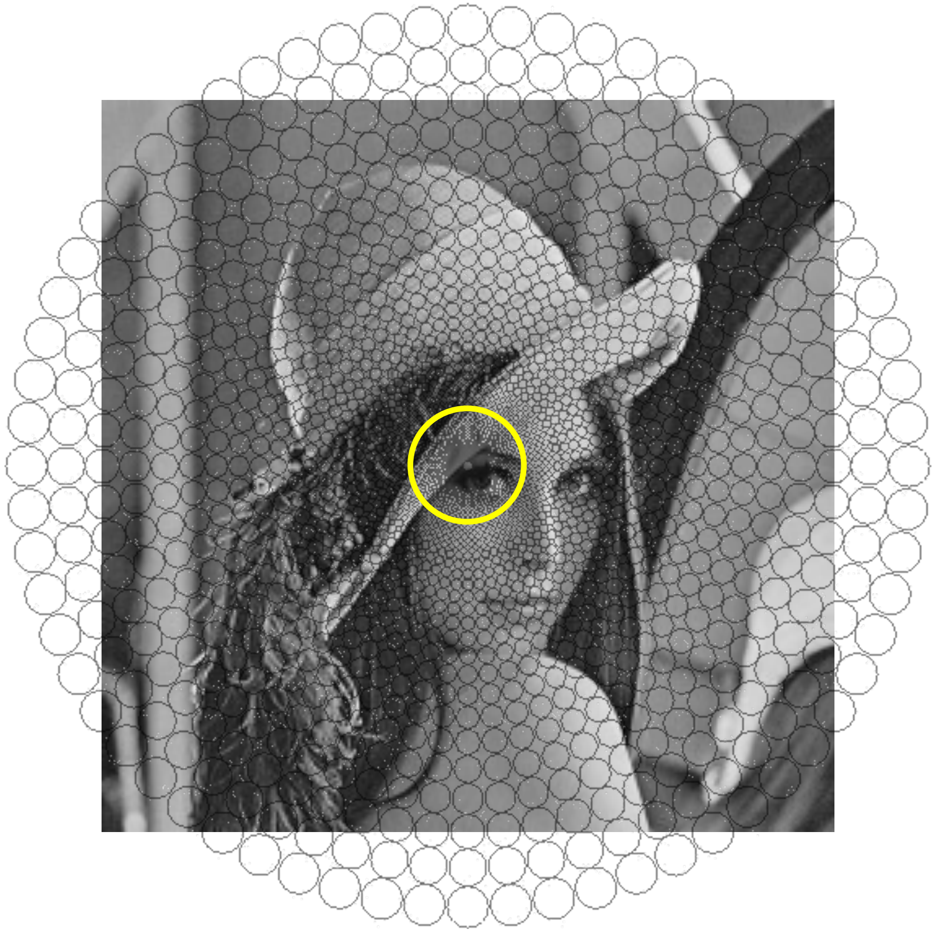
$$\begin{cases} \xi = \log_a \left(\rho / \rho_0 \right) \\ \eta = q\theta \end{cases}$$

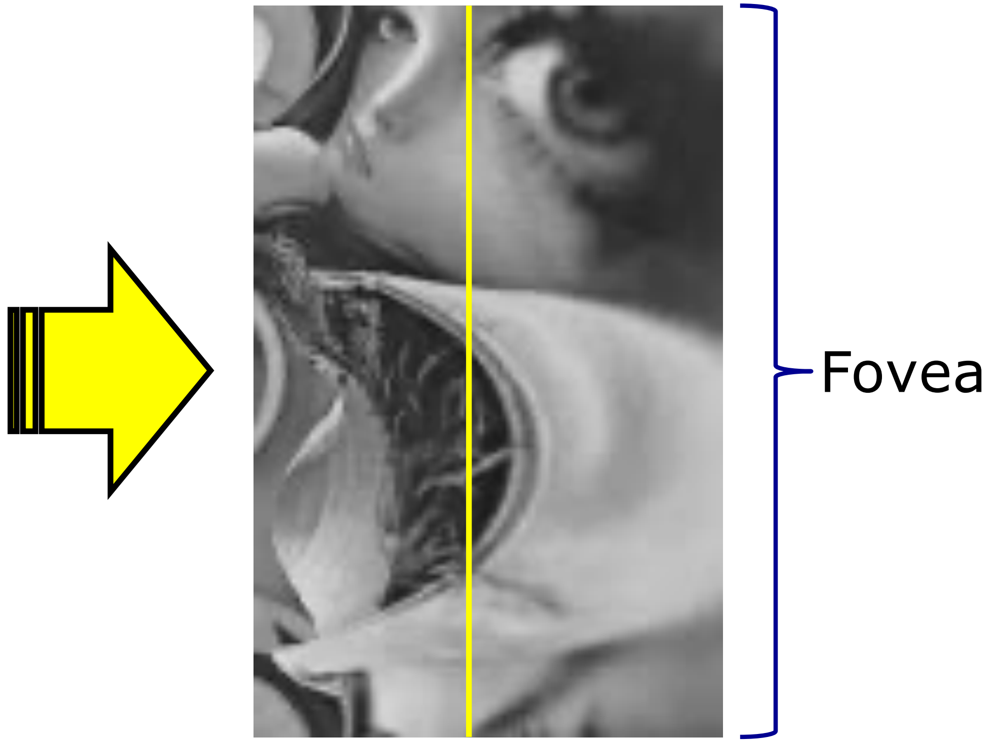
Massone, L., Sandini,G. and Tagliasco, V. "**Form-invariant topological mapping strategy for 2-d shape recognition**", CVGIP, vol. 30 No.2, pp. 169-188, 1985

Log-Polar mapping

Vision Lab

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

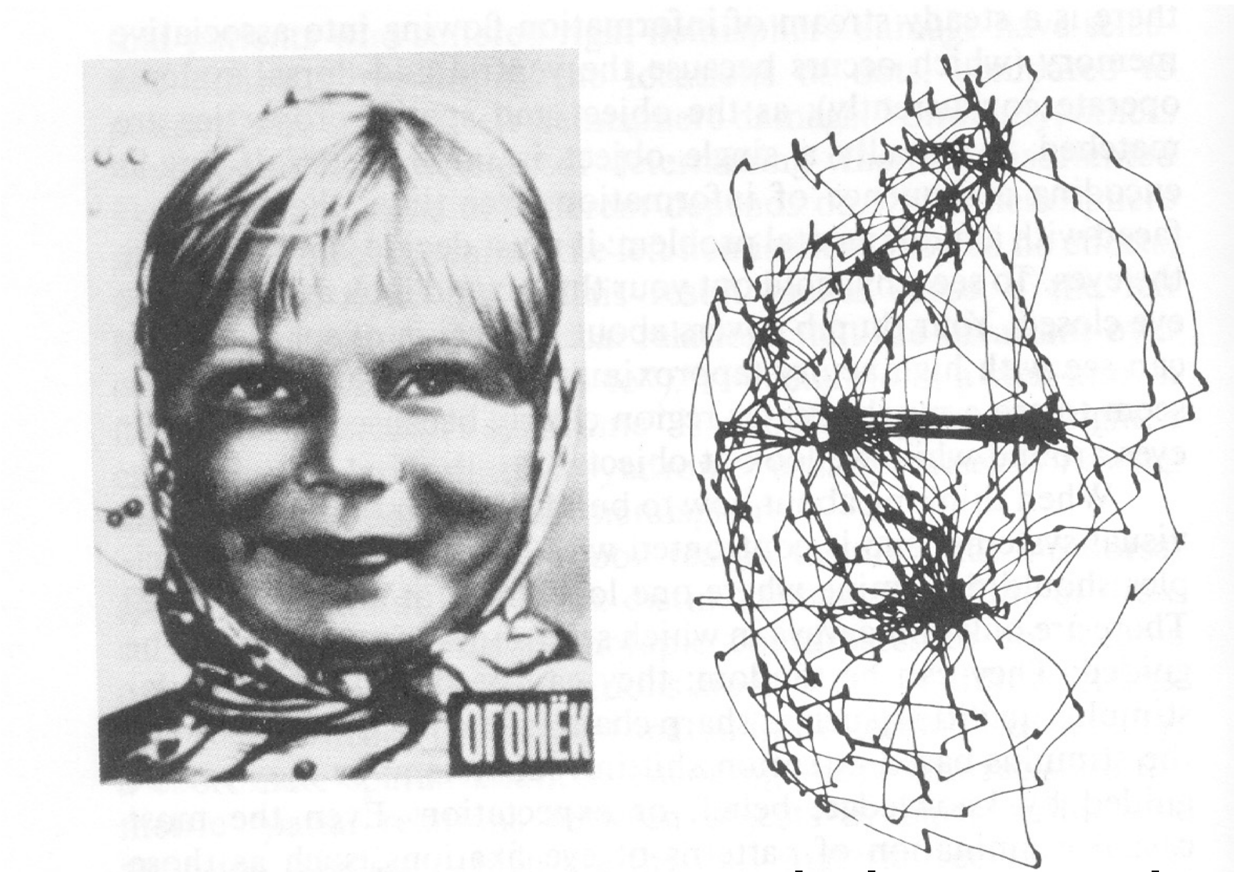


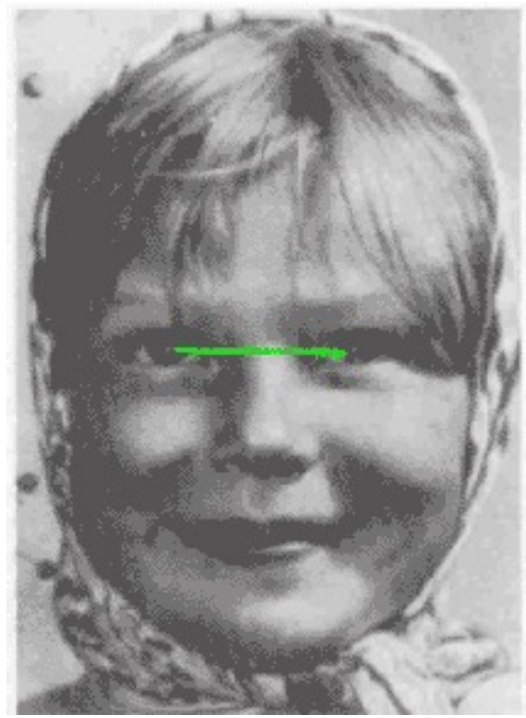


Massone, L., Sandini, G. and Tagliasco, V. "**Form-invariant topological mapping strategy for 2-d shape recognition**", CVGIP, vol. 30 No.2, pp. 169-188, 1985

Visual attention

Visual attention





Eye movements while watching a girl’s face

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

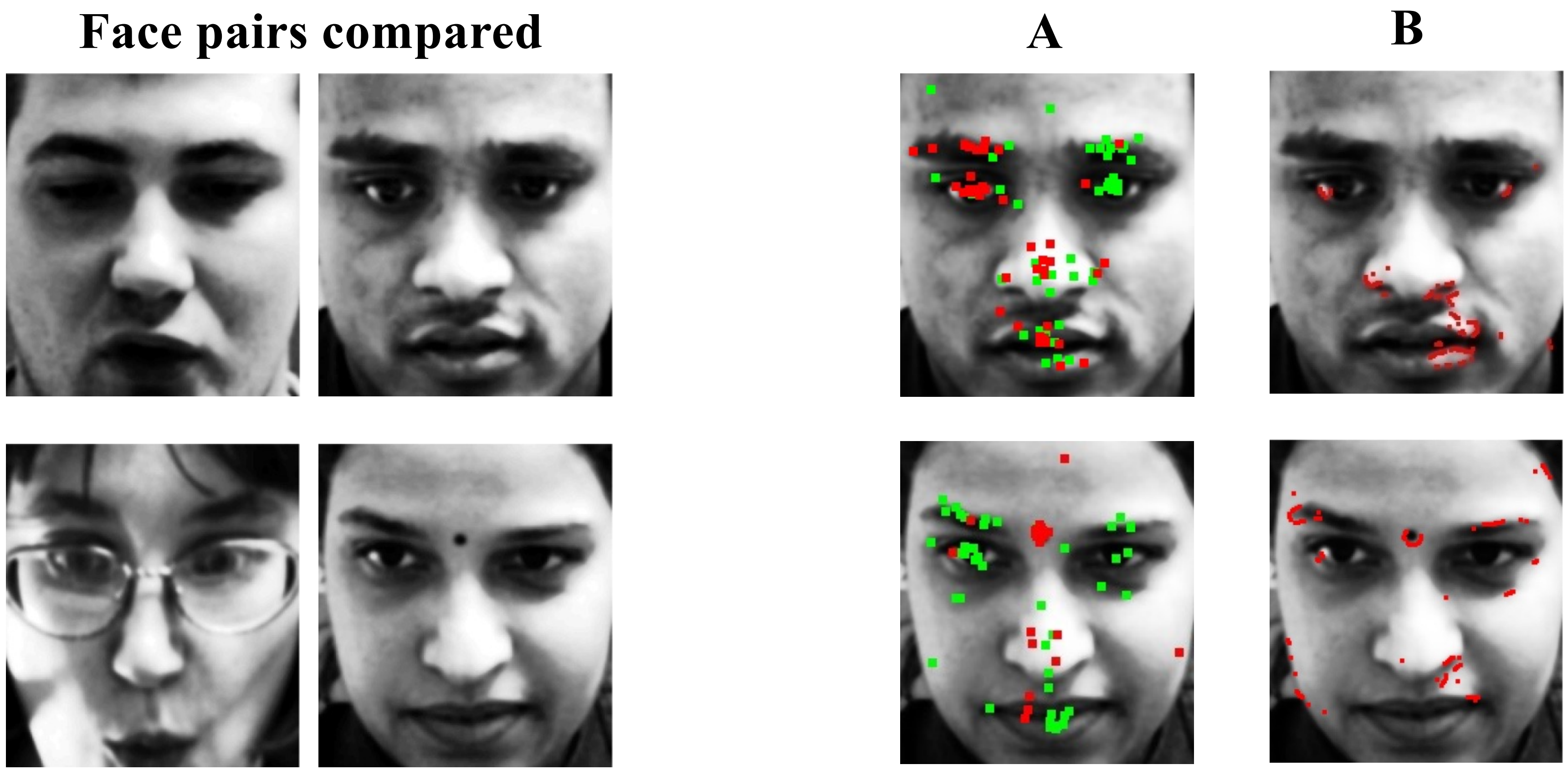
Visual attention

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

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

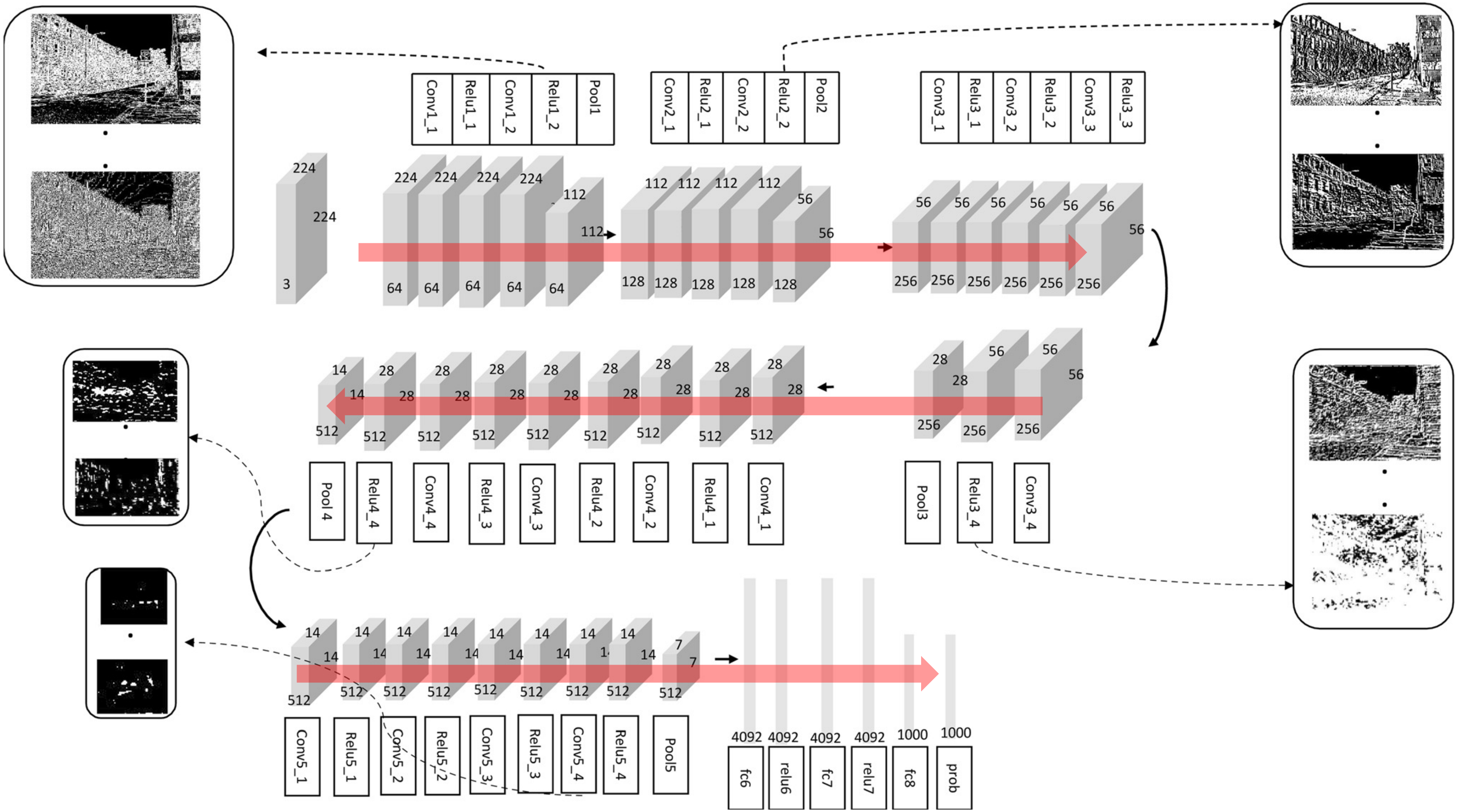


(A) perceptual and (B) computational results of saliency of local facial features, demonstrate the relevance of *non-standard* facial landmarks

Bicego M., Brelstaff G., Brodo L., Grosso E., Lagorio A. and Tistarelli M. (2007) "**Distinctiveness of faces: a computational approach**", ACM Transactions on Applied Perception, Vol. 5, n. 2, 2008.

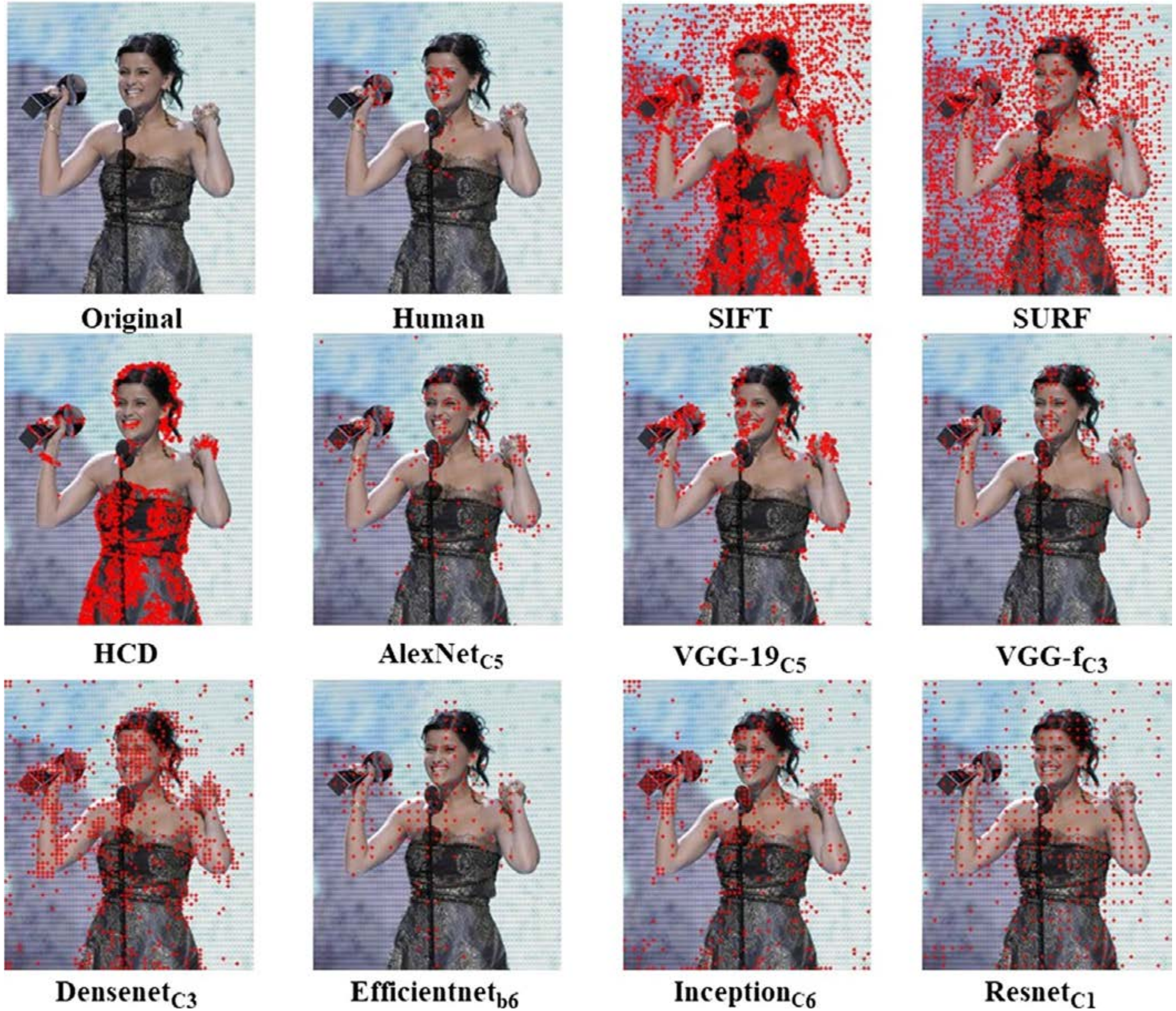
Visual attention

Vision Lab



M. Cadoni, A. Lagorio, S. Khellat-Kihel, E. Grosso (2021) "**On the correlation between human fixations, handcrafted and CNN features**", Neural Computing and Applications https://doi.org/10.1007/s00521-021-05863-5.

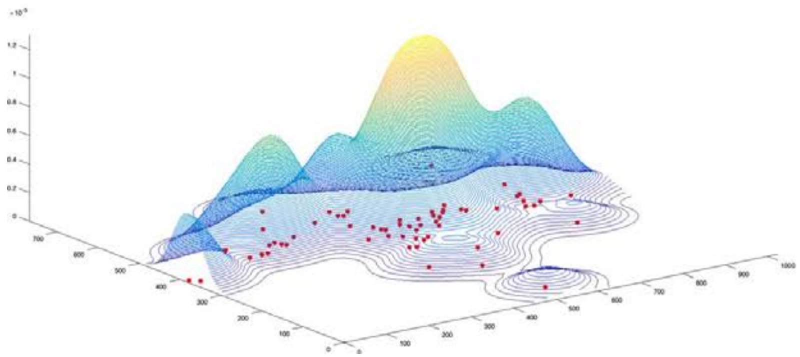
Visual attention

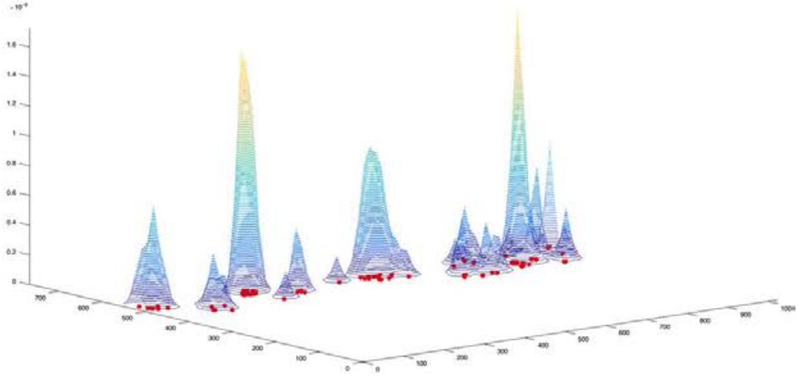


Visual attention

Fixation points

AlexNet interest points.

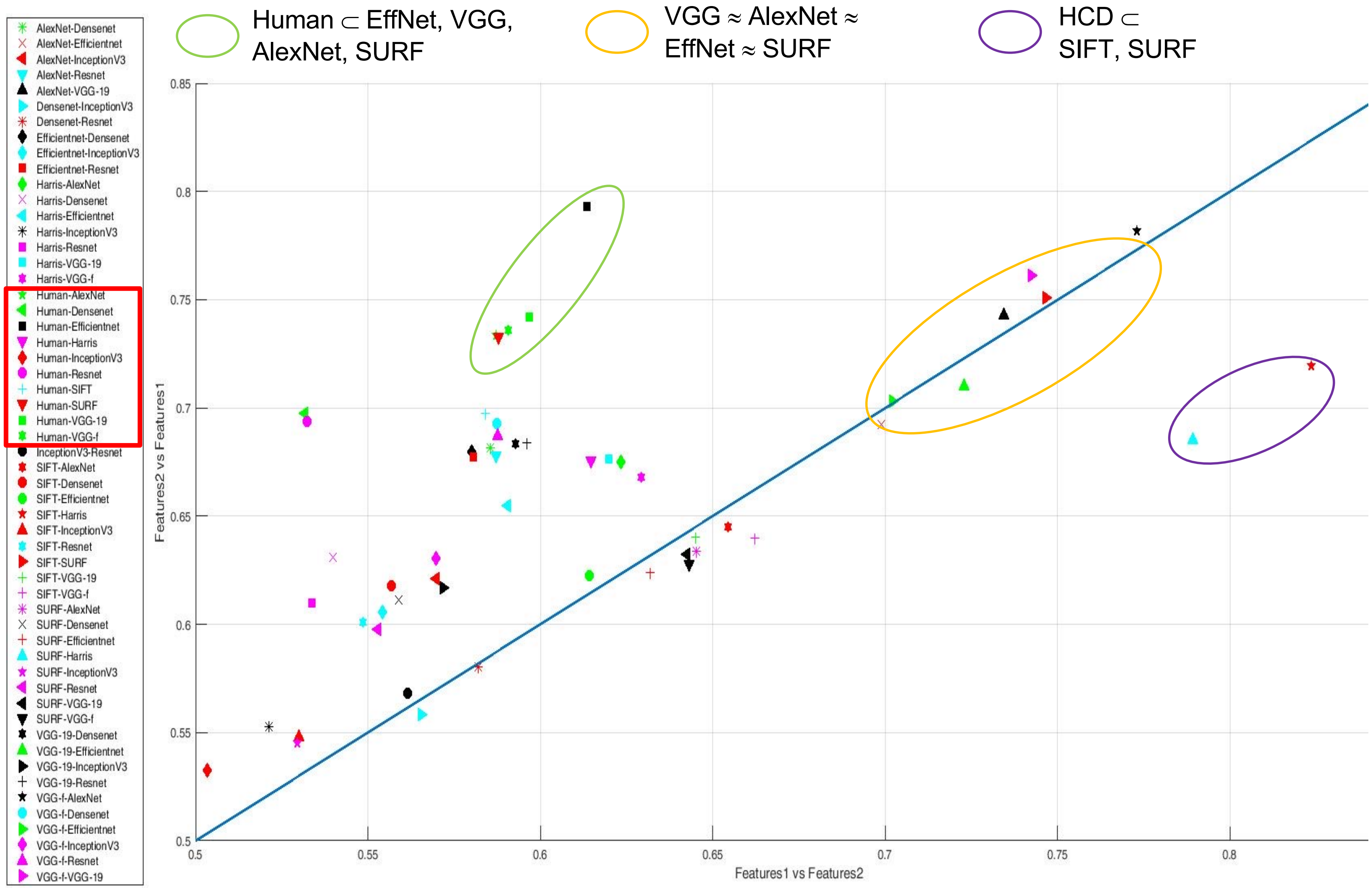




Interest regions are modeled via **K**erne**l** **D**ensity **E**stimation.

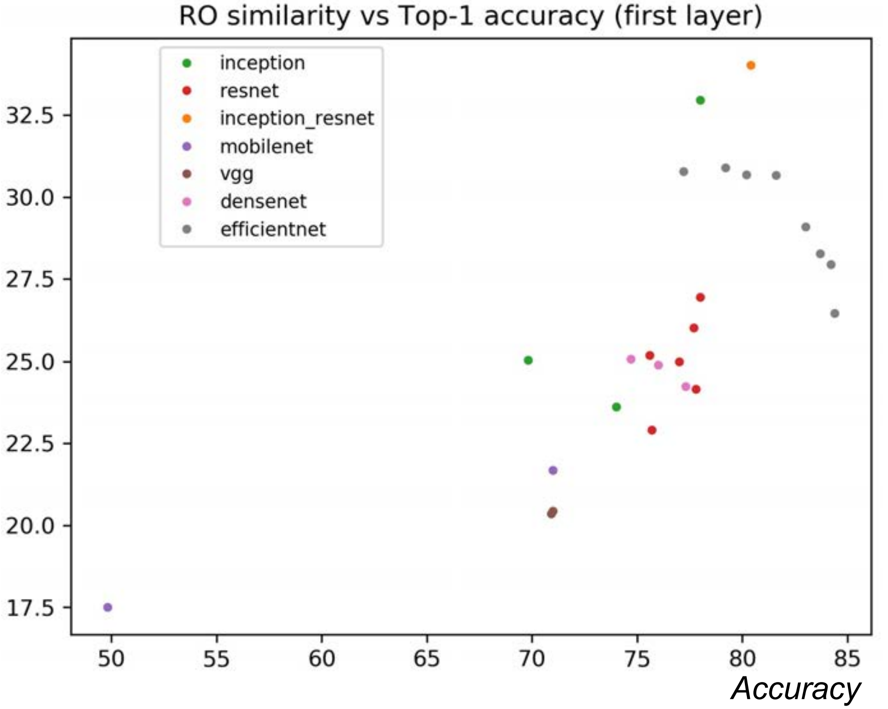
Visual attention

Vision Lab

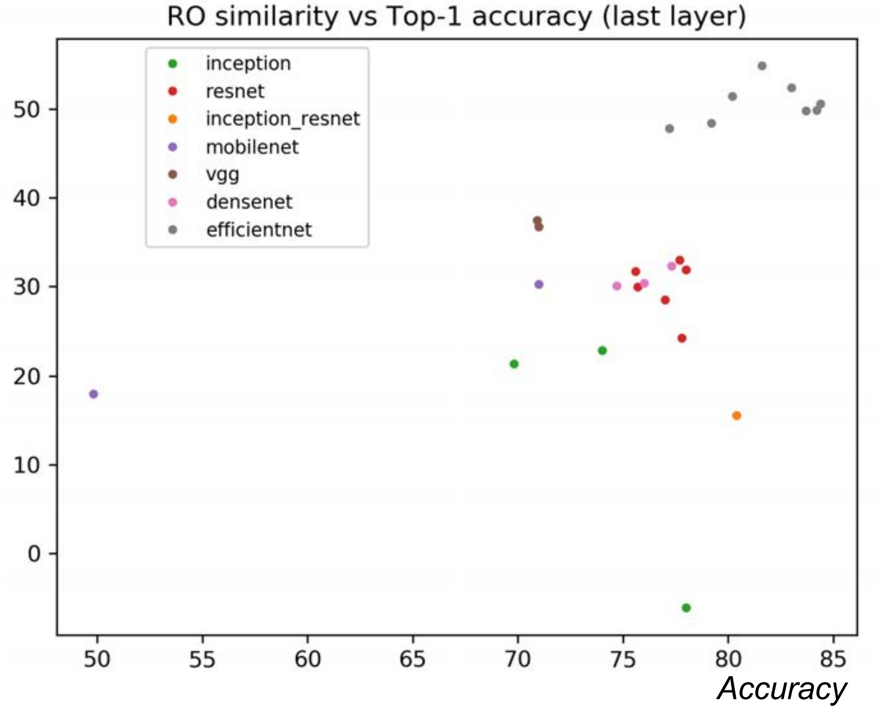


Local similarity between human fixations, CNNs and handcrafted features

Visual attention



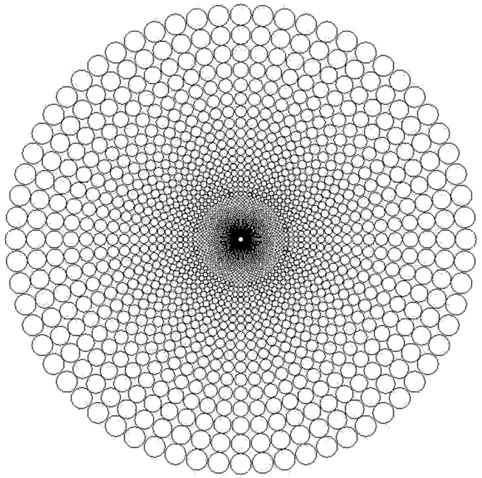
Scatter plot of CNN first layer similarity to fixations vs CNN classification performance.
Spearman rank correlation coefficient $\rho = 0.76$



Scatter plot of CNN last layer similarity to fixations vs CNN classification performance.
Spearman rank correlation coefficient $\rho = 0.54$

M. Cadoni, A. Lagorio, E. Grosso, T. Jia Huei, C. Chee Seng (2021) "**From early biological models to CNNs: do they look where humans look?**", 25th Int.l Conference on Pattern Recognition ICPR 2020, pp. 6313-6320
doi: 10.1109/ICPR48806.2021.9412717.

Space-variant imaging



Sandini, G. , Tistarelli, M. "**Vision and space-variant sensing**", in Neural Networks for Perception: Human and Machine Perception, H. Wechsler, Ed. Academic Press, 1991.

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

Sandini, G. , Tistarelli, M. "**Vision and space-variant sensing**", in Neural Networks for Perception: Human and Machine Perception, H. Wechsler, Ed. Academic Press, 1991.

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

Sandini, G. , Tistarelli, M. "**Vision and space-variant sensing**", in Neural Networks for Perception: Human and Machine Perception, H. Wechsler, Ed. Academic Press, 1991.

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

Sandini, G. , Tistarelli, M. "**Vision and space-variant sensing**", in Neural Networks for Perception: Human and Machine Perception, H. Wechsler, Ed. Academic Press, 1991.

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

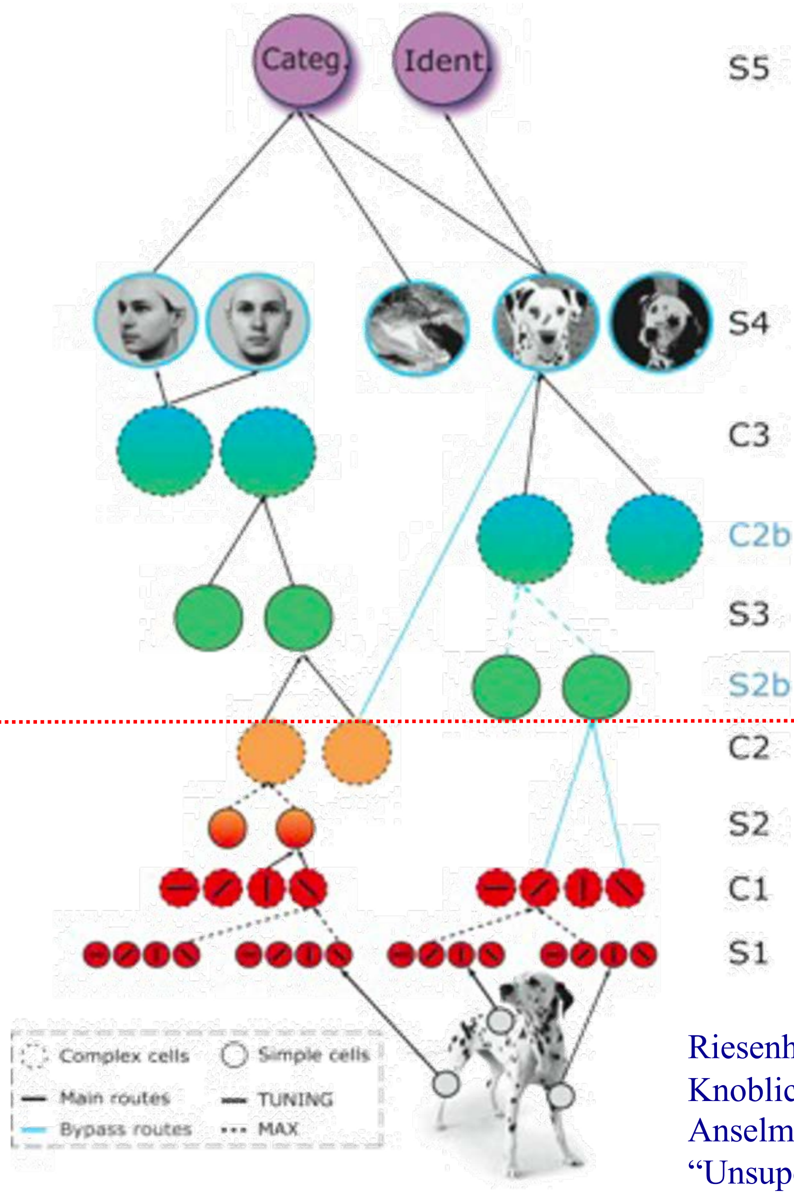
Space-variant imaging

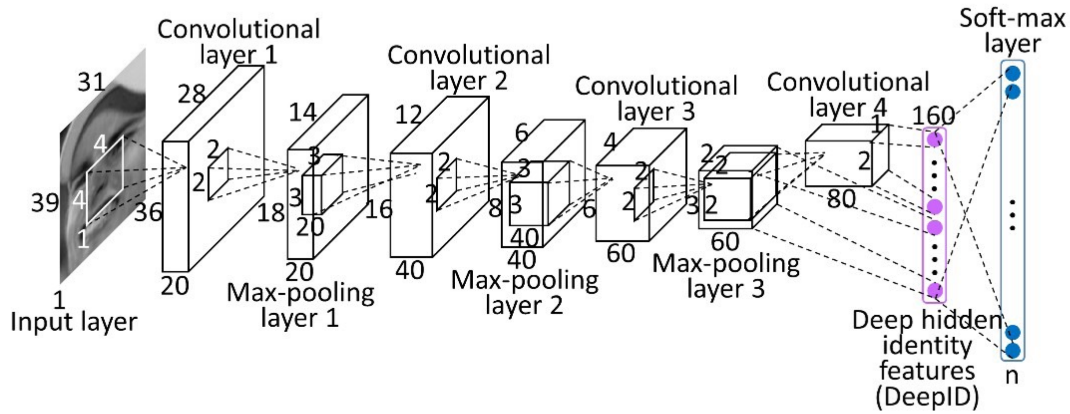
Sandini, G. , Tistarelli, M. "**Vision and space-variant sensing**", in Neural Networks for Perception: Human and Machine Perception, H. Wechsler, Ed. Academic Press, 1991.

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

Brain models





Complex composite cells

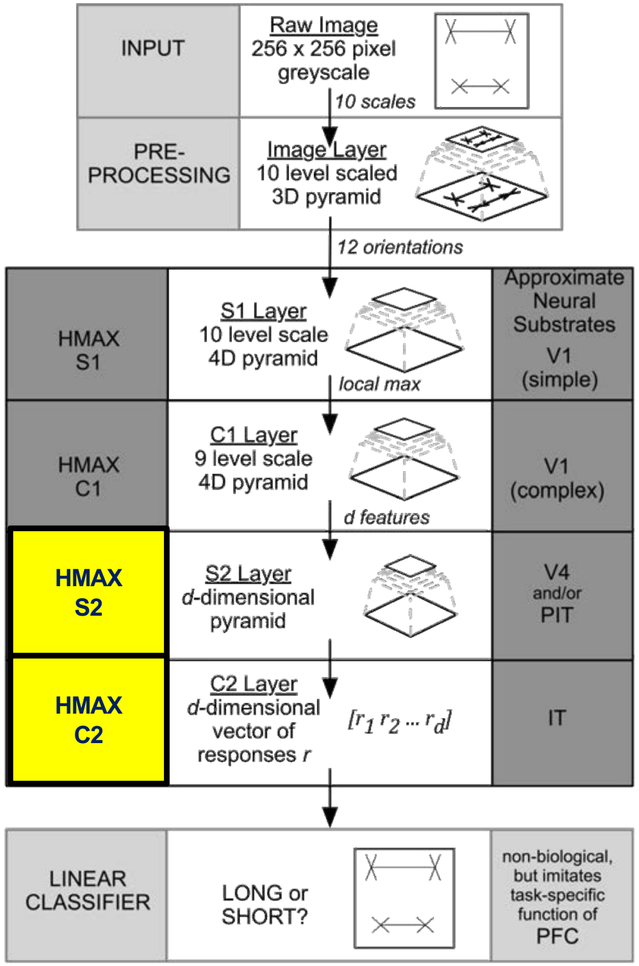
Composite feature cells

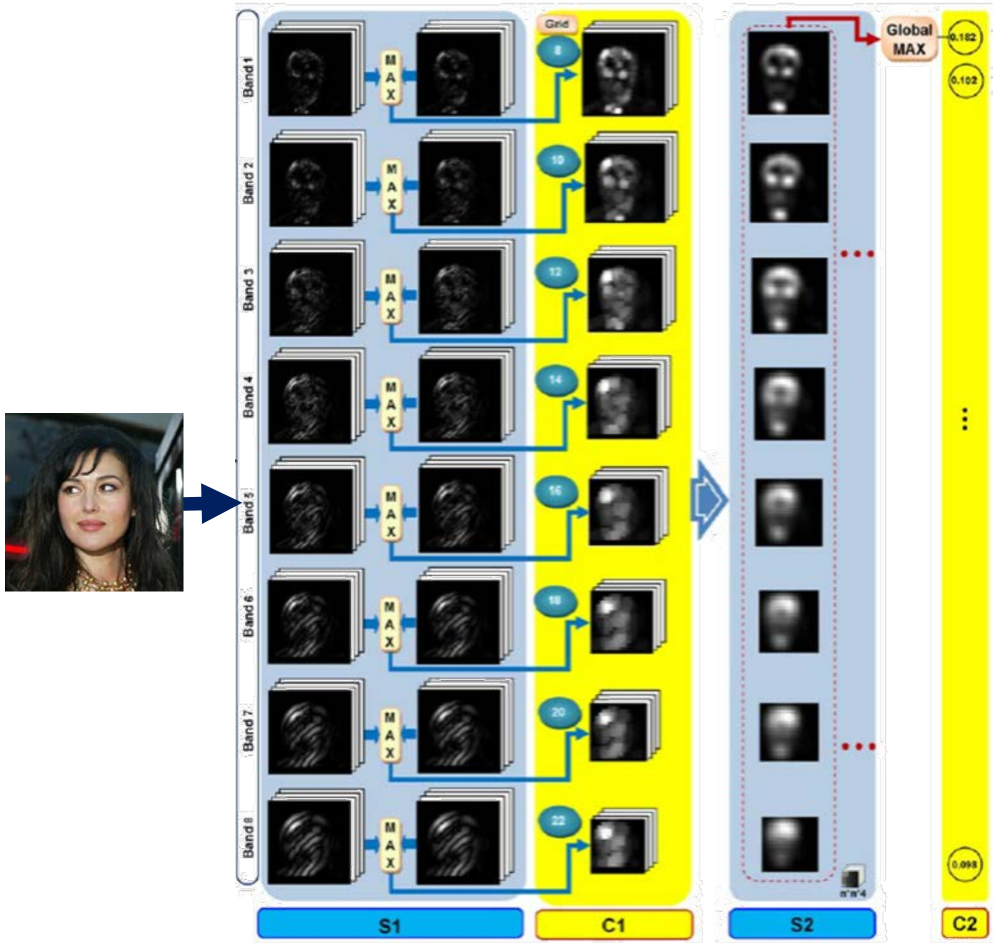
Complex cells

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

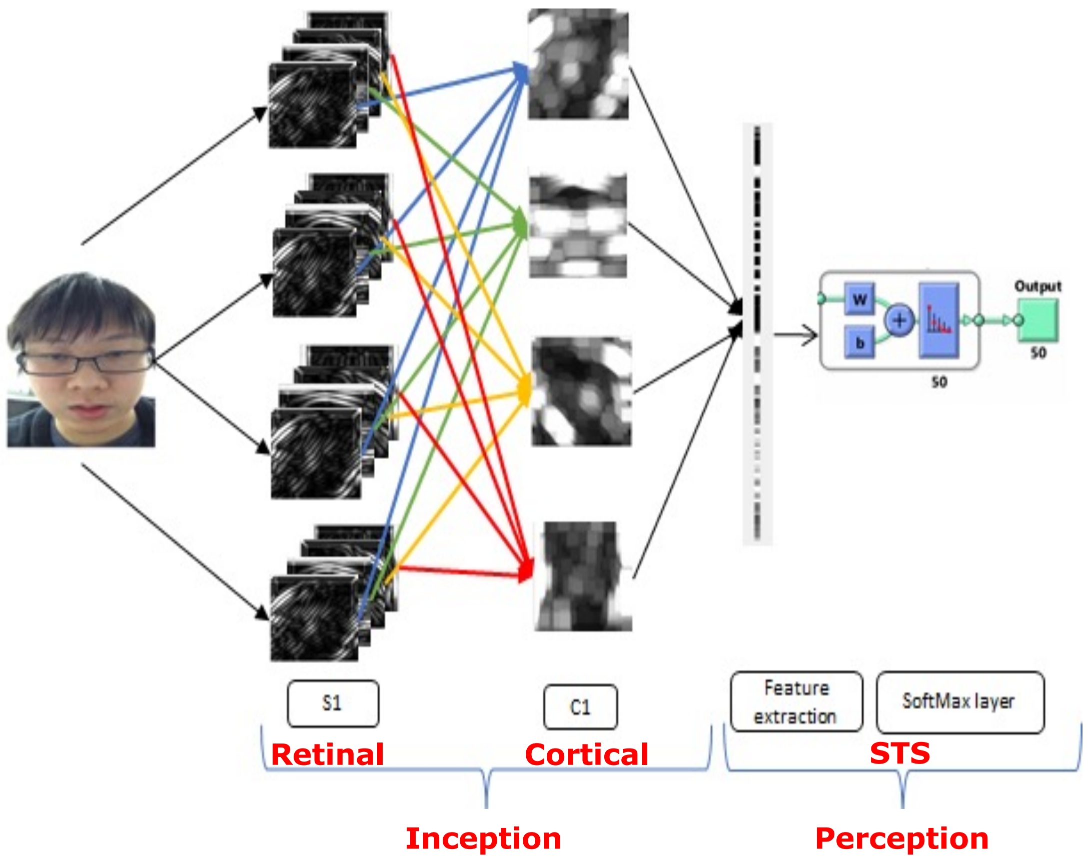




Riesenhuber, M. & Poggio, T. (1999). Hierarchical Models of Object Recognition in Cortex. Nature Neuroscience 2: 1019-1025.

(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

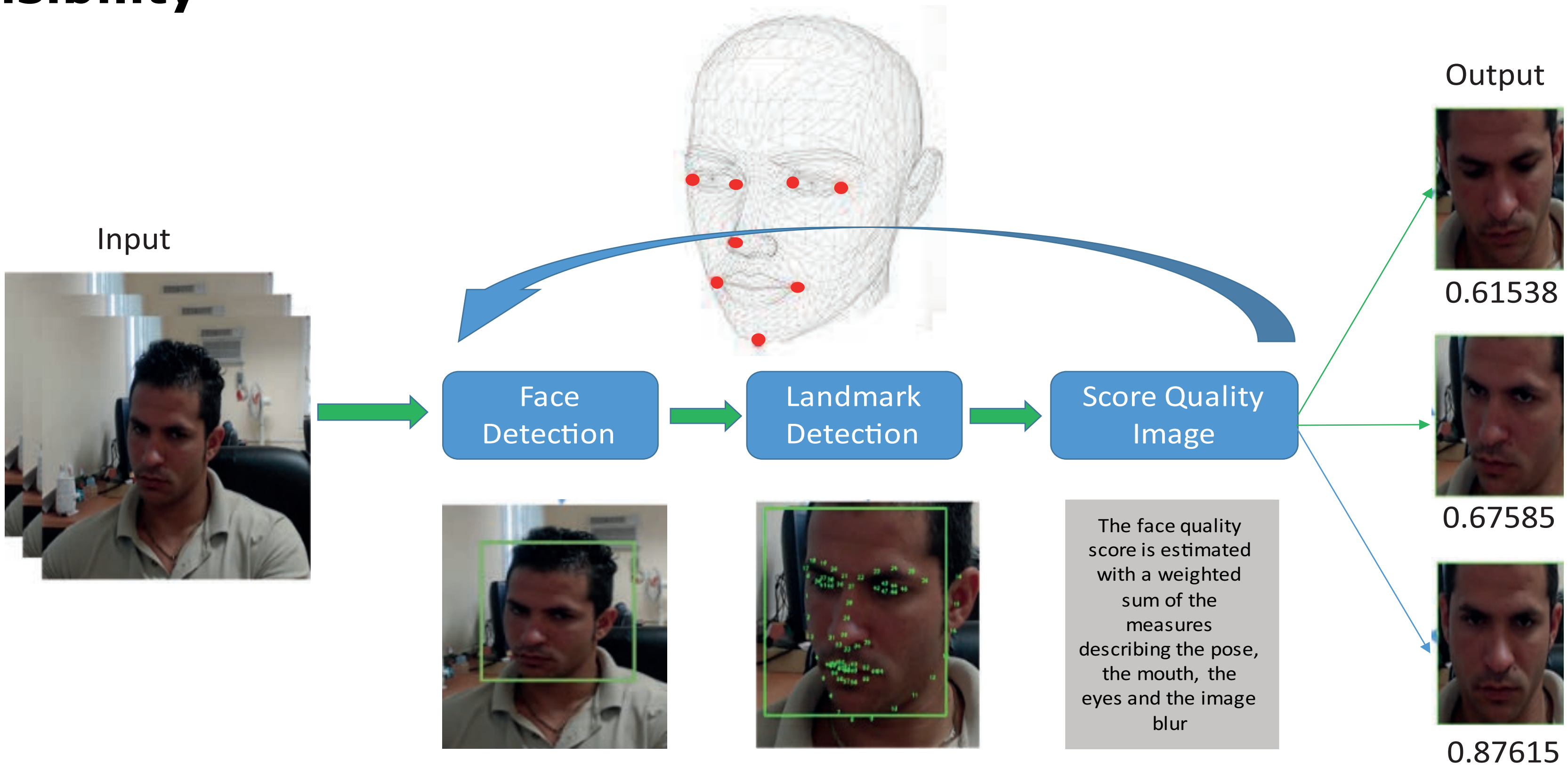


The **Gabor** and **max pooling** layers encode the face images based on a biologically-inspired chain running from the **retinal** stage to the **V1 cortex**.

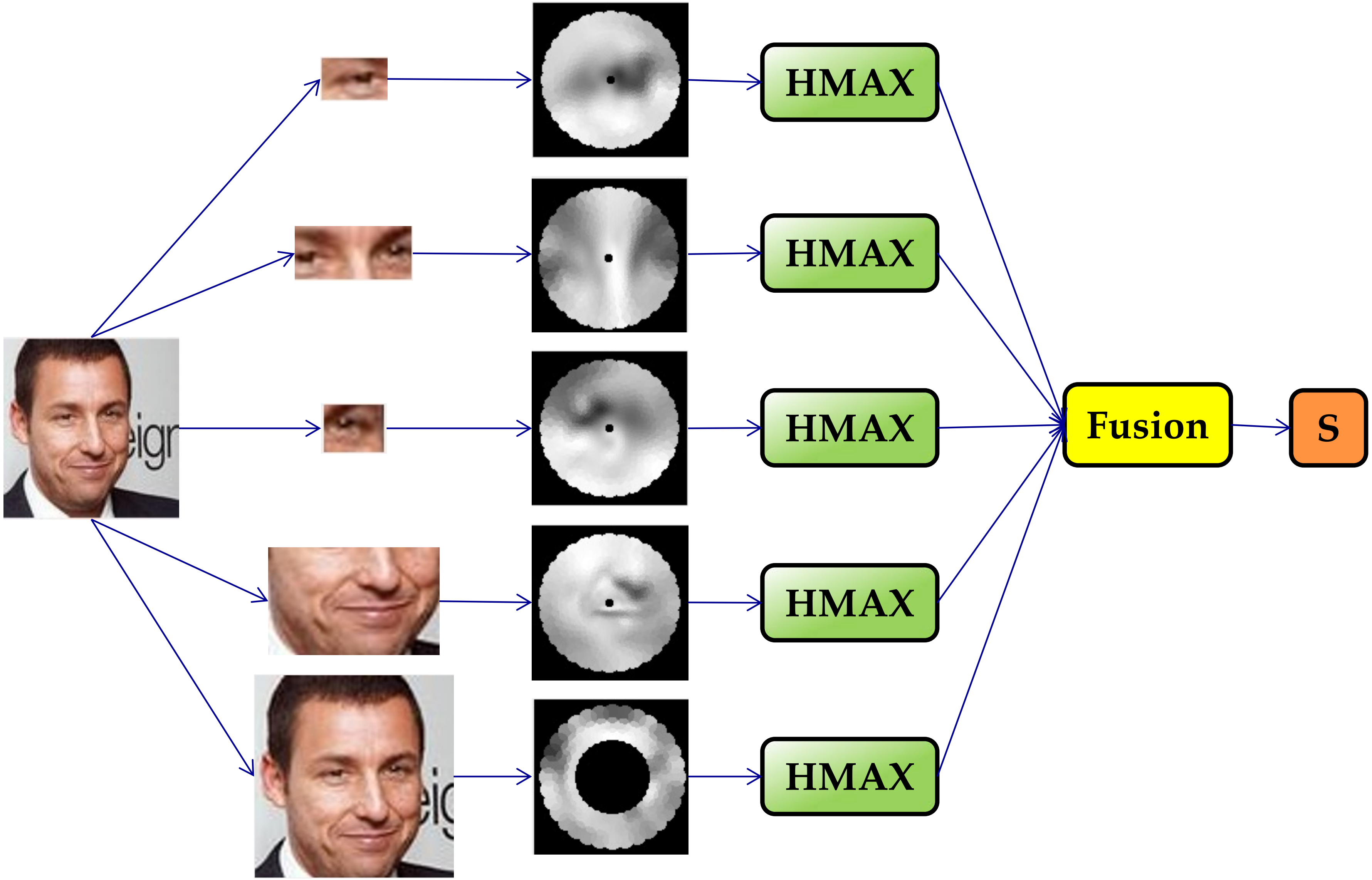
The connections between the **V1 cortex** and the **Superior Temporal Sulcus**, the face-selective area, is simulated by a network whose neurons are activated by a ***SoftMax*** function.

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**



Foveated HMAX



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***:

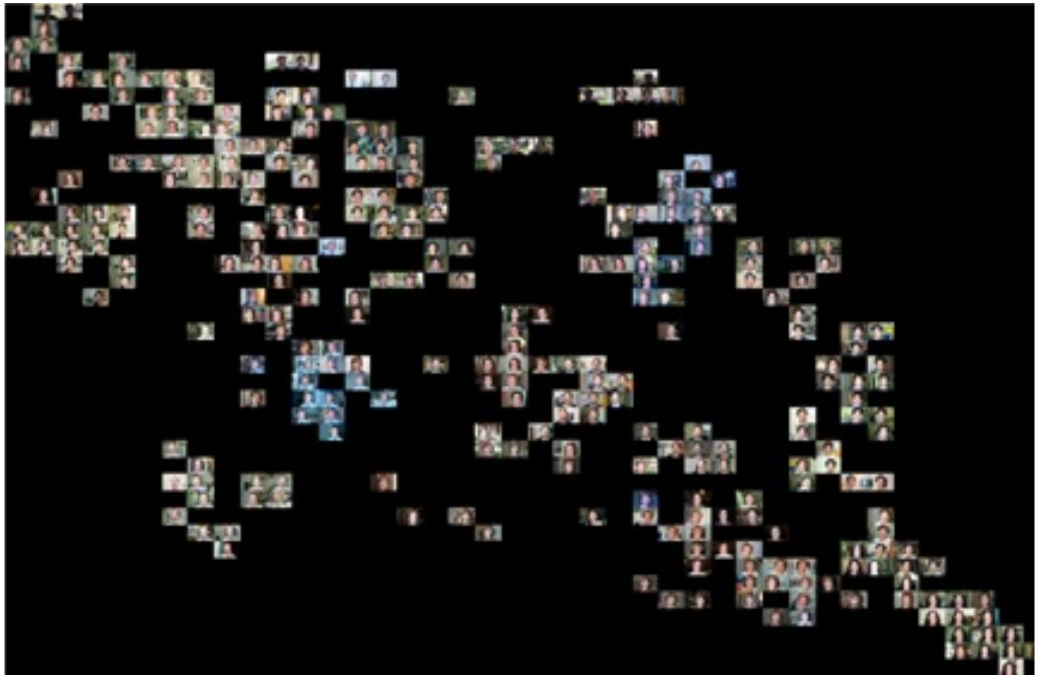
$$\boldsymbol{H}(\boldsymbol{y},\boldsymbol{p}) = \sum_i L_i(p_i)\,; \qquad L_i = -\log\left(\frac{e^{f_i}}{\sum j e^{f_j}}\right)$$

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.

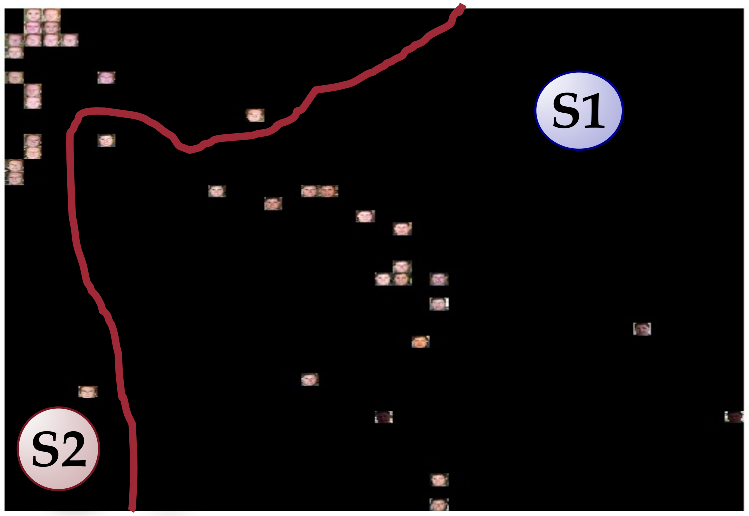
Foveated face recognition

HMAX Space representation on uniformly sampled face images

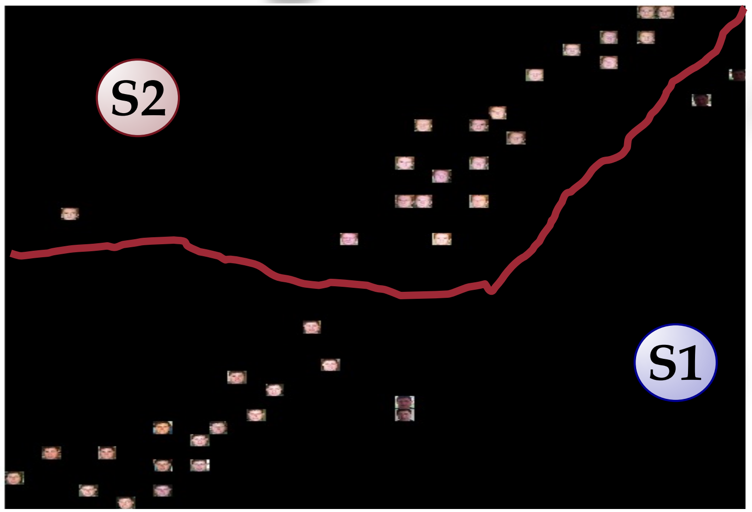


HMAX Space representation on log-polar sampled face images

Foveated face recognition



Uniform resolution



Log-polar mapping

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

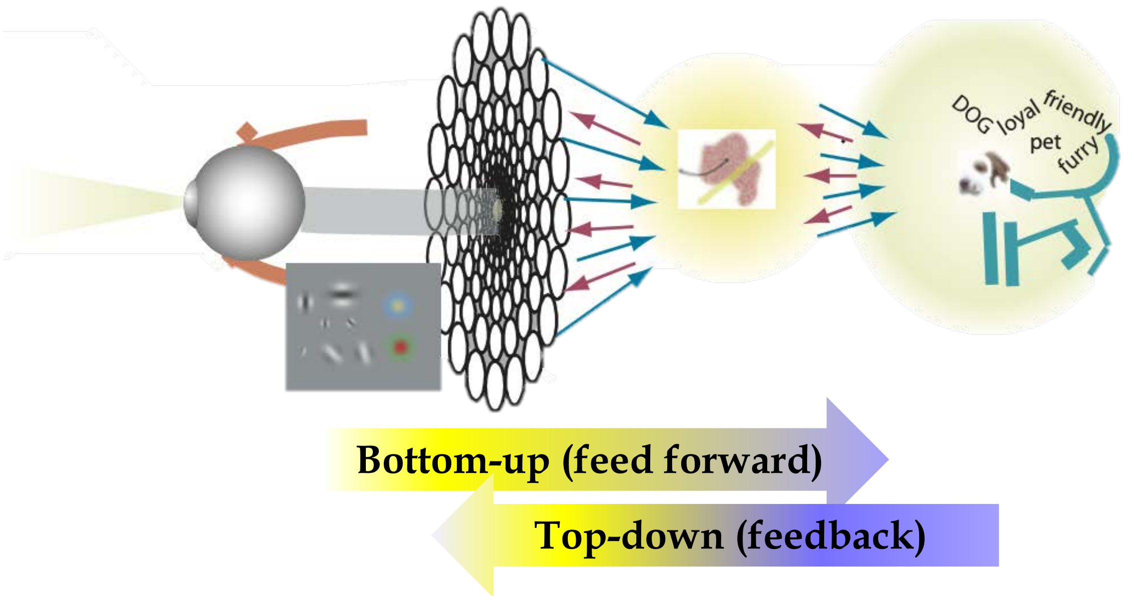
S. Khellat Khiel, A. Lagorio, M. Tistarelli. "**Foveated vision for biologically-inspired continuous face authentication**". In A. Rattani Ed. *Selfie Biometrics: Methods and Challenges*, Springer 2019.

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 can be 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.

What about the future?

- **Learn more from biological neural architectures to build network models:** Beyond the retino-cortical topological mapping
- **Learn from human perceptual behaviors:** Improve attention mechanisms; make networks more *curious*
- **Change the learning paradigm:** Exploit interactions; incremental and continuous learning
- **Adversarial attacks and robustness**: Interpolation/ approximation mistakes? How do they compare to optical illusions?
- **Add feedback to the system**: Reinforcement learning?



Vision Lab

THANK YOU FOR YOUR

19th Int.l Summer School for Advanced Studies on Biometrics for secure authentication:

"CONTINUALLY LEARNING BIOMETRICS "

Alghero, Italy - June, 6-10 2022

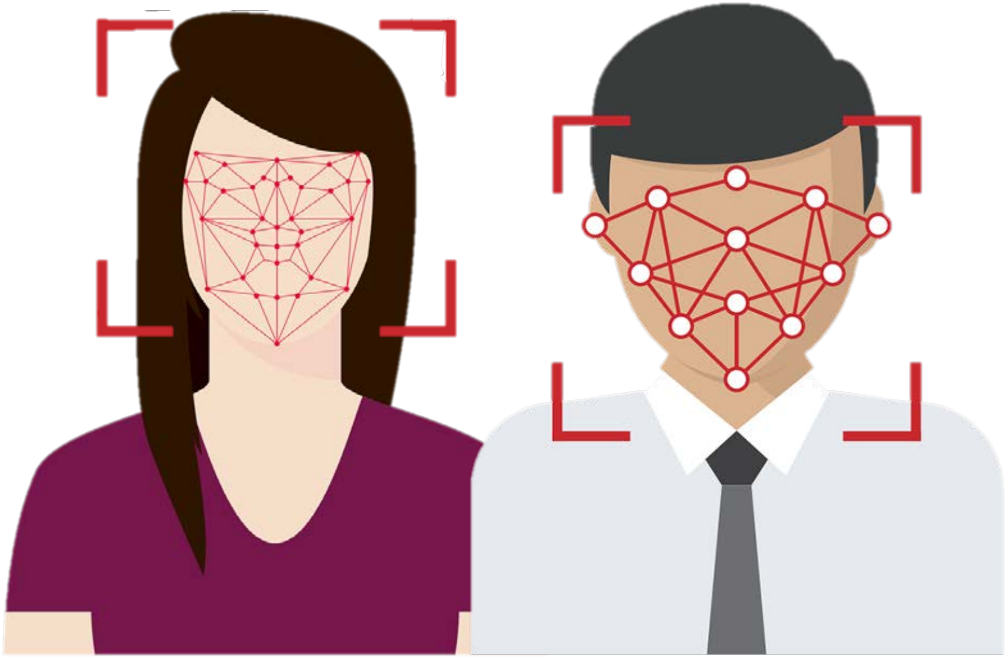
http://biometrics.uniss.it

Contact: tista@uniss.it

IAPR

IDENTITY

IDEMIA augmented identity



IEEE

EURASIP

450 UNISS

SVSCEPTVM PERFICE MVNVS