Face Recognition: Past, Present and Future

Massimo Tistarelli

Computer Vision Laboratory

University of Sassari – Italy

tista@uniss.it

Credits

From the laboratory staff:

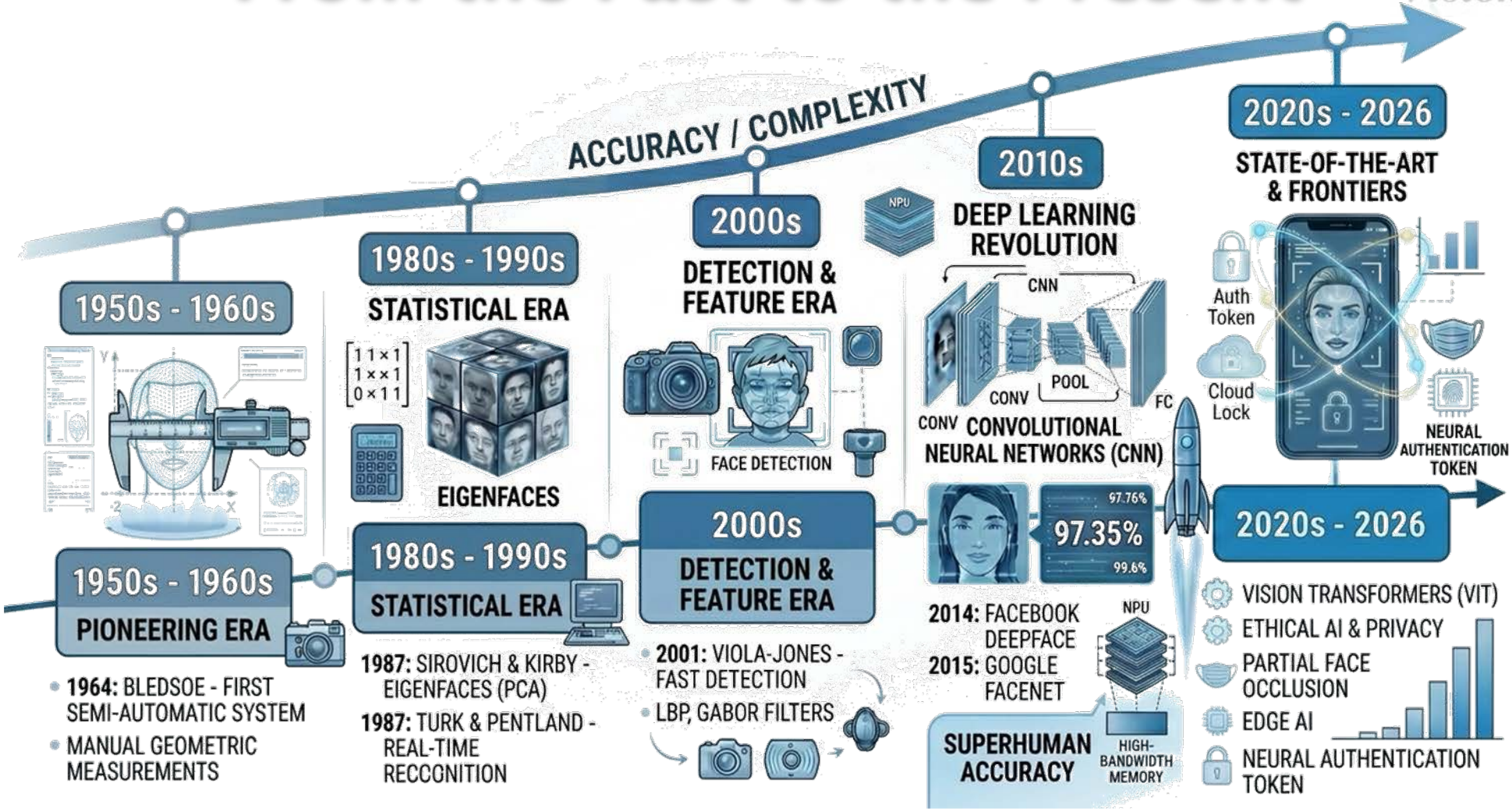
Linda Brodo
Marinella Cadoni
Filippo Casu
Enrico Grosso
Alessandra Ibba
Andrea Lagorio
Seth Nixon
Pietro Ruiu
Amhad Waseem
Souad Khellat Khiel (past visiting)
Gianluca Masala (past visiting)
Norman Poh (past visiting)
Ajita Rattani (past visiting)
Yunlian Sun (past visiting)
Daksha Yadav (past visiting)
Yu Guan (past visiting)
Marcos Ortega Hortas (past visiting)

Credits

…and other labs:

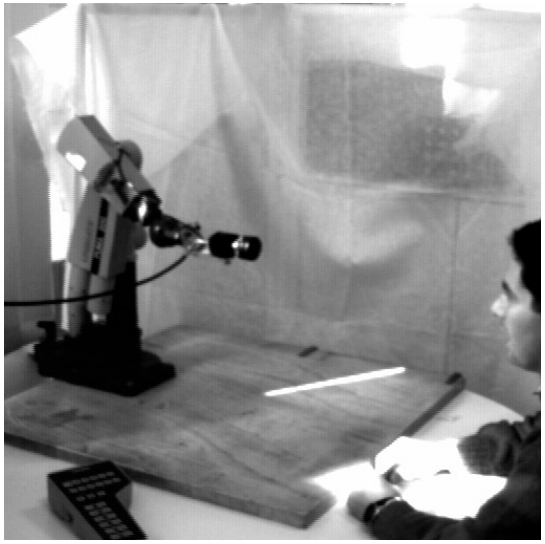
Manuele Bicego – University of Verona
Rama Chellappa – University of Maryland
Anil Jain – Michigan State University
Zhe Jin – Anhui University
Chang-Tsun Li – Deakin University
Alice O'Toole – University of Texas at Dallas
Jonathon Phillips – NIST
Norman Poh – University of Surrey

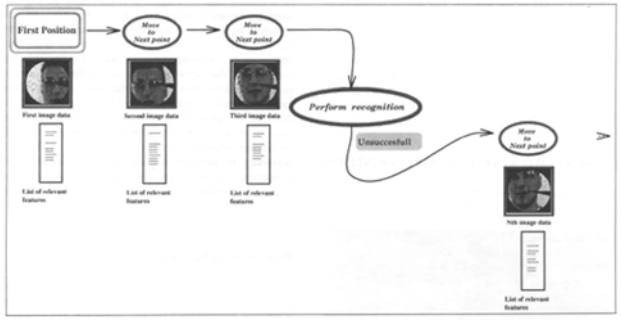
From the Past to the Present



Once upon a time…

CVPR June 1994… Seattle, WA





Recognition by Using an Active/Space-Variant Sensor

M. Tistarelli
DIST - University of Genoa
Laboratory for Integrated Advanced Robotics (LIRA - Lab)
via Opera Pia 11a - 16145 Genoa, Italy

Abstract

The problem of object recognition is addressed. In the literature this task has been generally considered in a "passive" perspective, where everything is static and there is no definite relation between the object and its environment. We propose an "active" approach for object recognition, based on the capability of the observer to move and give a better description of the object under consideration and also to take advantage of the relations between the objects and the environment. This can be accomplished at the task level and at the sensor level.

The face recognition problem, based on the face-space approach, is considered to demonstrate the advantage of adopting an active retina to sample the face, build a database and perform the recognition task. By using an active space-variant retina the size of the database is considerably reduced and consequently also the processing time for recognition.

A comparative experiment using the active and static approach is presented.

1 Introduction

Object recognition is one of the most "classical" themes in artificial intelligence applied to vision. Nonetheless up to now it is not a solved problem at all, but many different systems and methods have been investigated with limited success [1]. Certainly the reason of such effort is the formidable complexity of the recognition problem and the ability of humans to recognize objects quite quickly. But, is this ability due to a particular efficency of the search strategy in the model database? or is it due to the computational power of the inference engine (the brain)? without any doubt these are two relevant characteristics of the human brain, but these are not necessarily the primary reasons for the efficency of the human visual system. On the other hand we can consider that all the research carried out in the past, along these directions, did not obtain the expected results.

2 Fixation and recognition

What is the role of fixation in the recognition process? Yarbus, in his work on ocular movements [8], demonstrated that the sequence of fixations performed by the human oculo motor system, strongly depends on the task (in this case the question asked to the subject). He also showed that the eyes perform a particular sequence of fixations, if the subject has to recognize a part or a person in the scene. The eyes are successively directed toward the parts of the scene containing the most relevant features [2]. This motion strategy suggests that the motion of the eyes is particularly important for recognition (at least in the human visual system).

It is generally assumed that, for recognition, it is desirable to have a high resolution description of the most salient features of the interest object. This can be accomplished either by "foveating", in rapid succession, these parts of the scene or moving an interest window on a high resolution image [10].

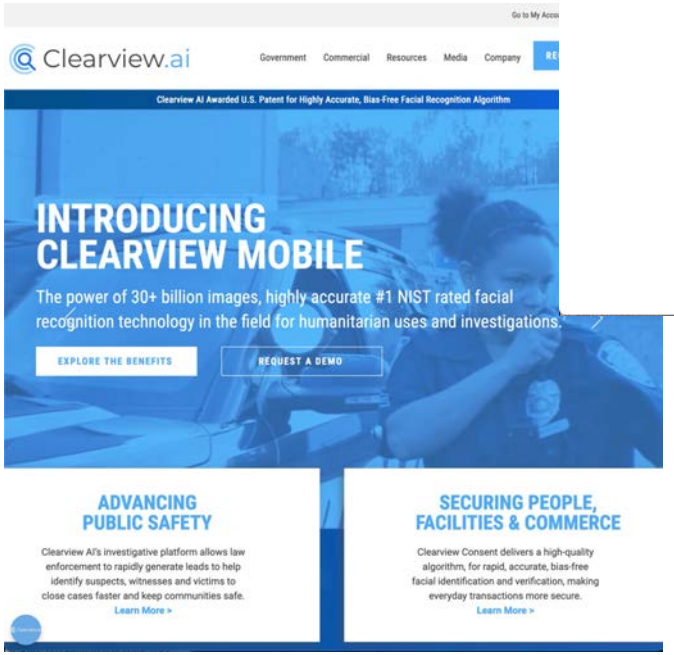
Certainly object features are important for recognition, but the context, or the peripheral part of the visual field, allows to define a spatial relation among the object features which really characterize the object itself. A way to meet these requirements is to adopt a space-variant sampling strategy of the image, where the central part of the visual field is sampled at a higher resolution than the periphery, with a linear variation in resolution from the center to the periphery. An advantage of this approach is the great data

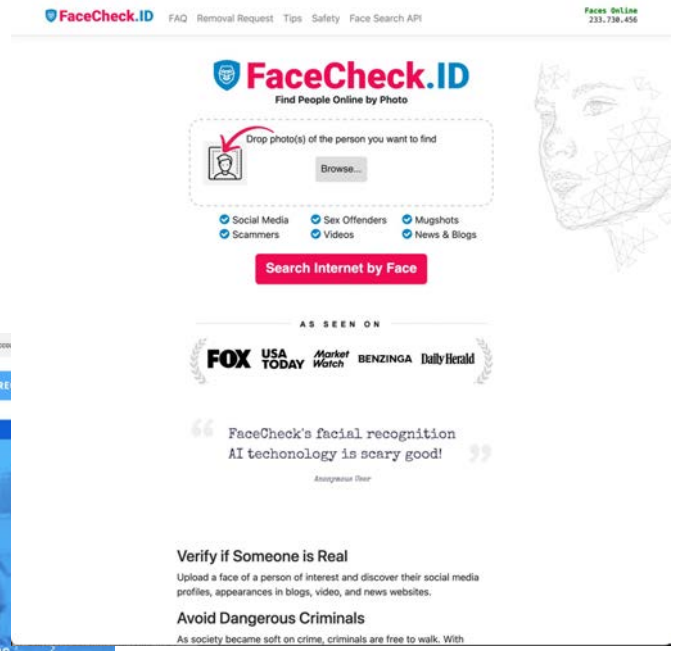
[1] The success of the techniques is limited in the sense that the generality of the solutions is not even comparable to the aims, which is to develop a fully general object recognition system, working in real world environments.

[2] Eklundh [9] demonstrated that these points can be recovered through a scale-space analysis of the image.

Tistarelli, M (1994) "**Recognition by using an active/space-variant sensor**" *IEEE CVPR*, 1994
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

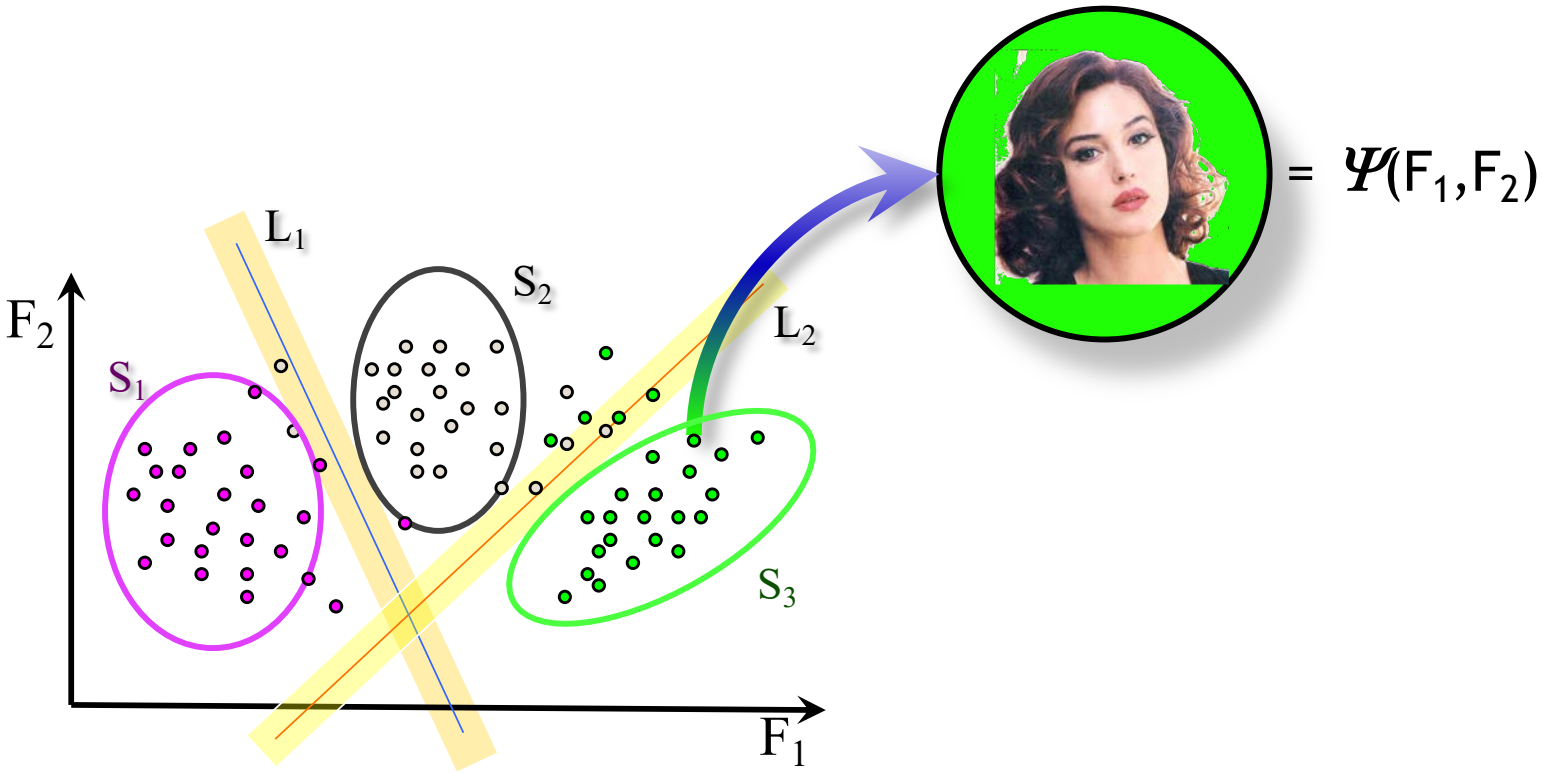
Faces in the market





Face Recognition

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



Two *simple* blocks

- **Feature representation**:
  - **Handcarfted**: SIFT, SURF, HCD, Edges, Wavelets,…
  - **Learned**: CNNs, PCA, LDA, FFT,…
- **Classifier**:
  - Cross-entropy loss, Kernel Machines, Generative/Discriminative Models, MCS, MRF, NNs, SVM, …

Many approaches to achieve two *simple tasks* and *easily* perform face recognition

Question 1

- **If the solution is so simple, why do we still need to pursue research on this topic?**

Question 1

- **If the solution is so simple, why do we still need to pursue research on this topic?**

1. **We all have to pay our bills… from research or company grants**

Question 1

- **If the solution is so simple, why do we still need to pursue research on this topic?**

1. We all have to pay our bills… from research grants
2. **We are not so clever**

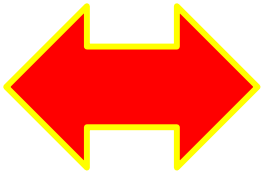
Question 1

- **If the solution is so simple, why do we still need to pursue research on this topic?**

1. We all have to pay our bills… from research grants

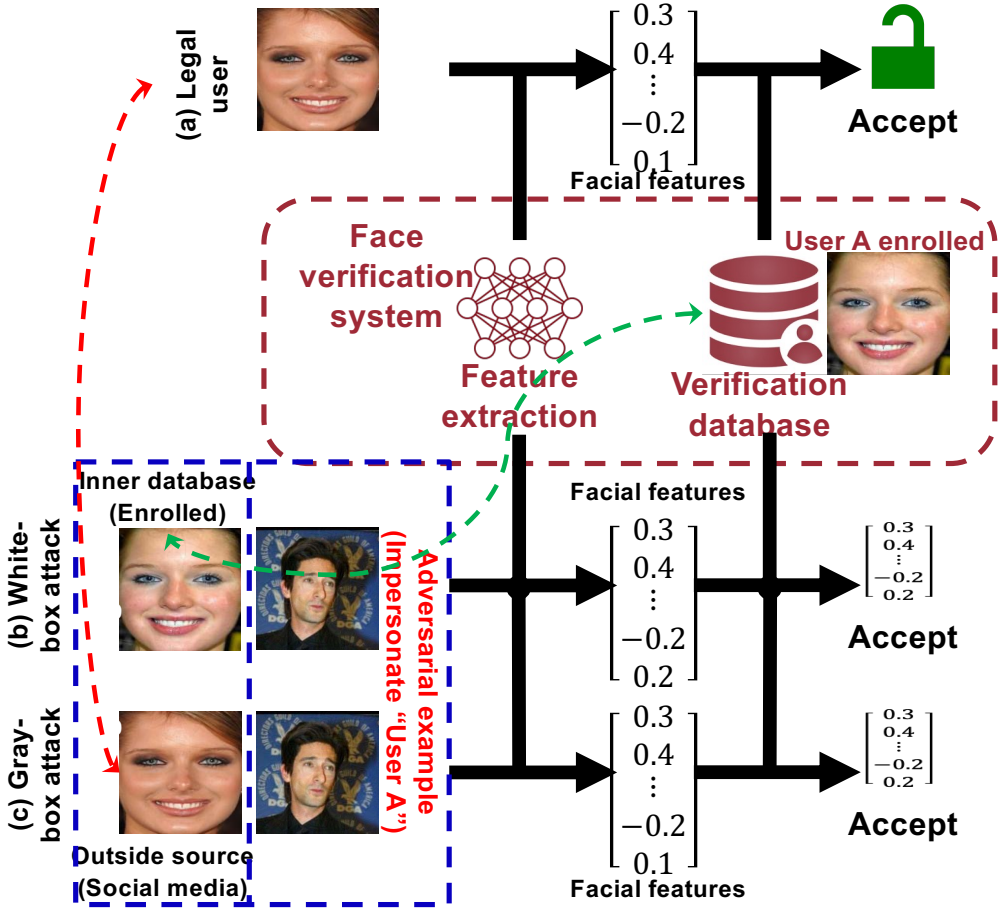
2. **We are not so clever**
   - **Proof**: regardless the huge efforts in terms of human and computational resources, the systems we have built in 30+ years are still prone to errors:
     - **Adversarial examples**
     - **Data alignement**
     - **Estimation bias**

Who is who?



Mahmood Sharif , Sruti Bhagavatula, Lujo Bauer, Michael 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

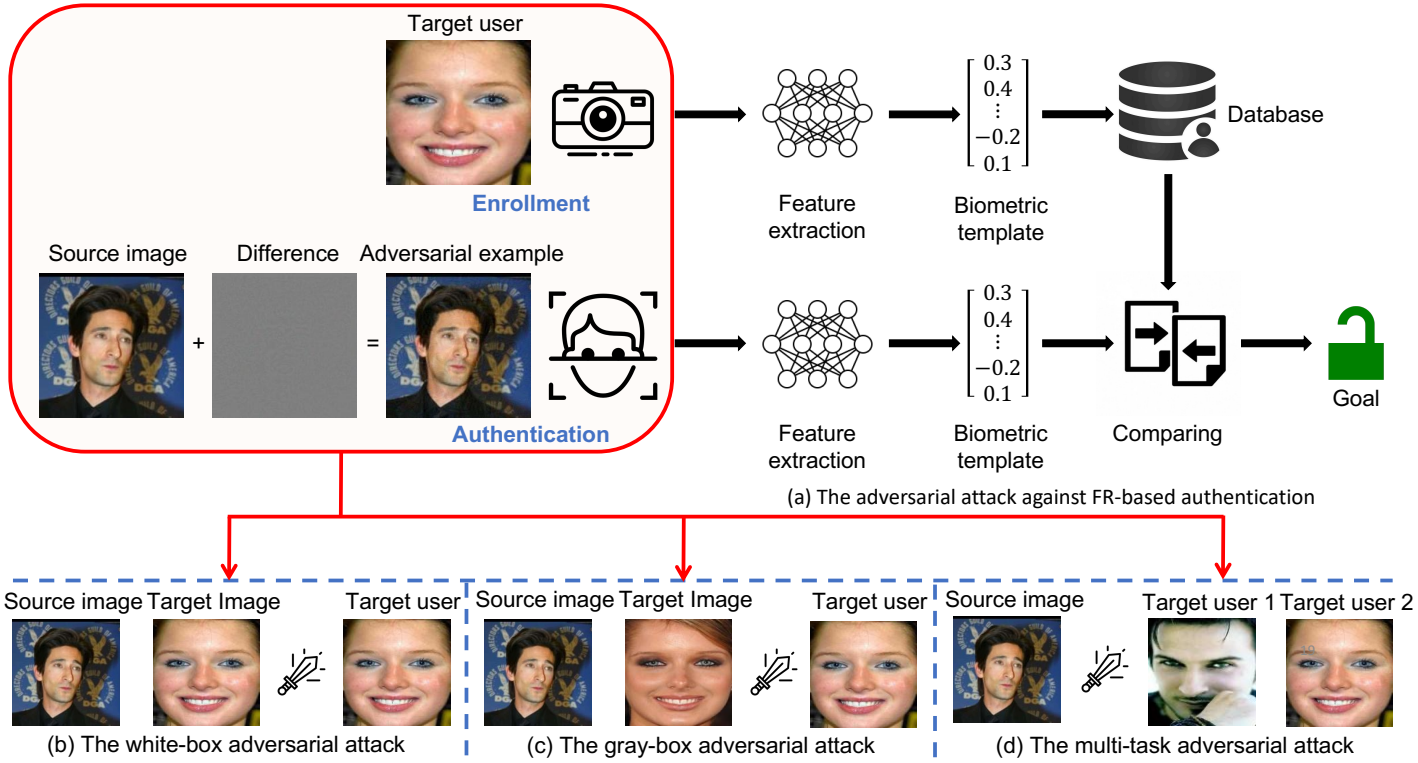
Adversarial Attacks



- **White-box attack**:
  1. The network model for face verification (feature extraction) is **KNOWN**.
  2. The original enrolled face image is **KNOWN**.
- **Gray-box attack**:
  1. The feature extraction model is **KNOWN** to the attacker.
  2. The original enrolled face image is **UNKNOWN** to the attacker.
- **Black-box attack:**
  Both the feature extraction model **and** the original enrolled face image are **UNKNOWN**.

H. Wang, S. Wang, Z. Jin, Y. Wang, C. Chen, and M. Tistarelli, **Similarity-based gray-box adversarial attack against deep face recognition**, in IEEE International Conference on Automatic Face and Gesture Recognition 2021 (FG2021), 2021

Adversarial Attacks



H. Wang, S. Wang, Z. Jin, Y. Wang, C. Chen, and M. Tistarelli, **Similarity-based gray-box adversarial attack against deep face recognition**, in IEEE International Conference on Automatic Face and Gesture Recognition 2021 (FG2021), 2021

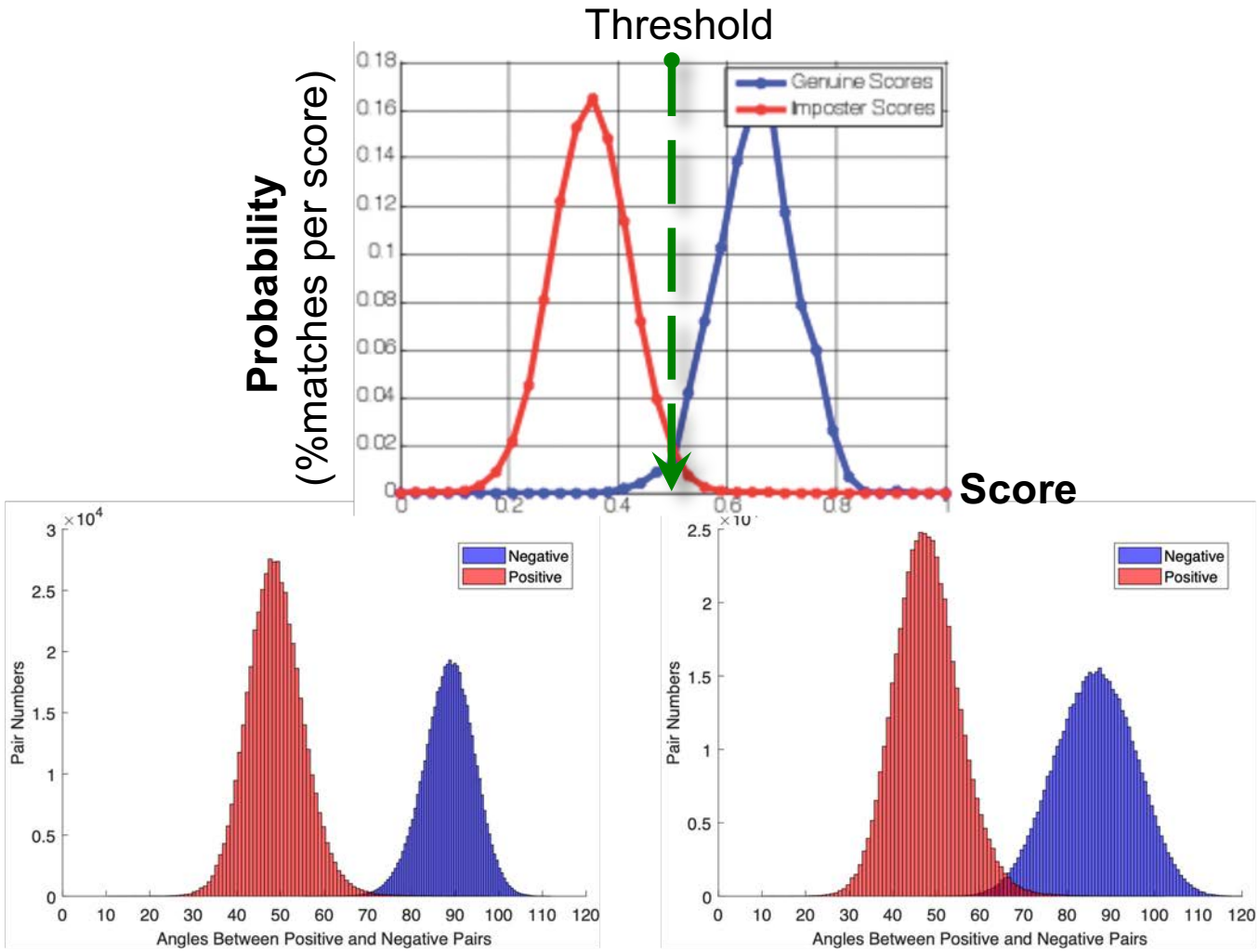
H. Wang, S. Wang, Z. Jin, Y. Wang, C. Chen, and M. TistarelliA **Multi-Task Adversarial Attack against Face Authentication**. ACM Trans. Multimedia Comput. Commun. Appl. 20:11, Article 332, 2024

Question 1

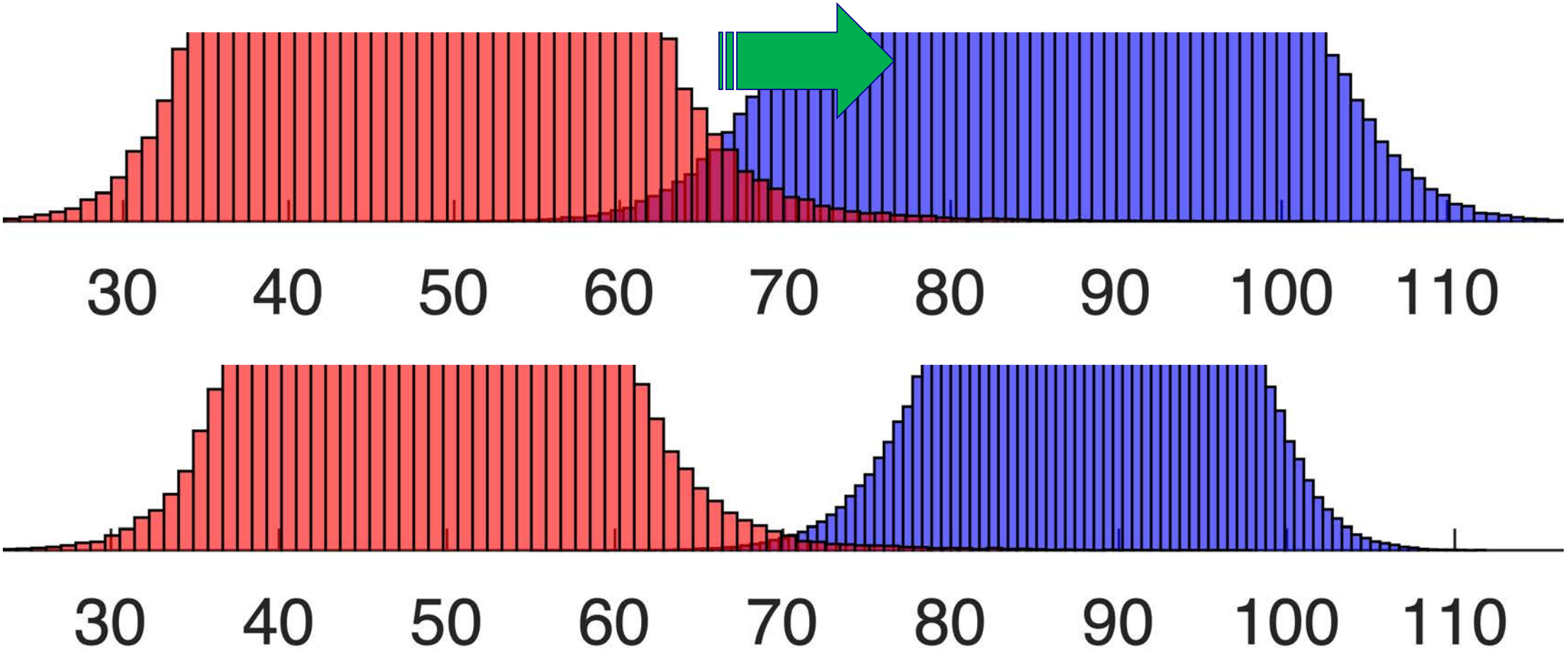
- **If the solution is so simple, why do we still need to pursue research on this topic?**

1. We all have to pay our bills… from research grants
2. We are not so clever
3. **There is something making the problem harder than we thought**

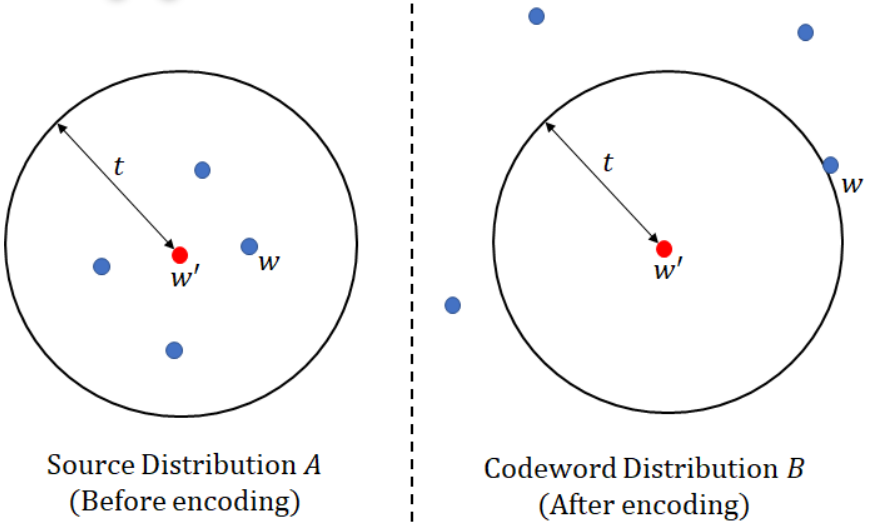
Match score distributions



Match score distributions



Crypto-biometrics



Problem: *Biometric sources generate random distributions of data. Error correction encoding may project the data far apart in the feature space, enhancing the error.*

Example: *Source distribution A and codeword distribution B may tolerate the same error* ***t****. However, in the codeword domain neighboring data points may be pushed apart and unsecure.*

Significance of the problem: *Tolerating errors in biometric data allow biometric cryptosystem to achieve a higher security level and privacy protection, as well as a higher convenience, i.e., usability.*

Yen-Lung Lai, Xingbo Dong, Zhe Jin, Massimo Tistarelli, Wun-She Yap, Bok-Min Goi: **Breaking Free From Entropy's Shackles: Cosine Distance-Sensitive Error Correction for Reliable Biometric Cryptography**. IEEE Trans. Inf. Forensics Secur. 18: 3101-3115 (2023)

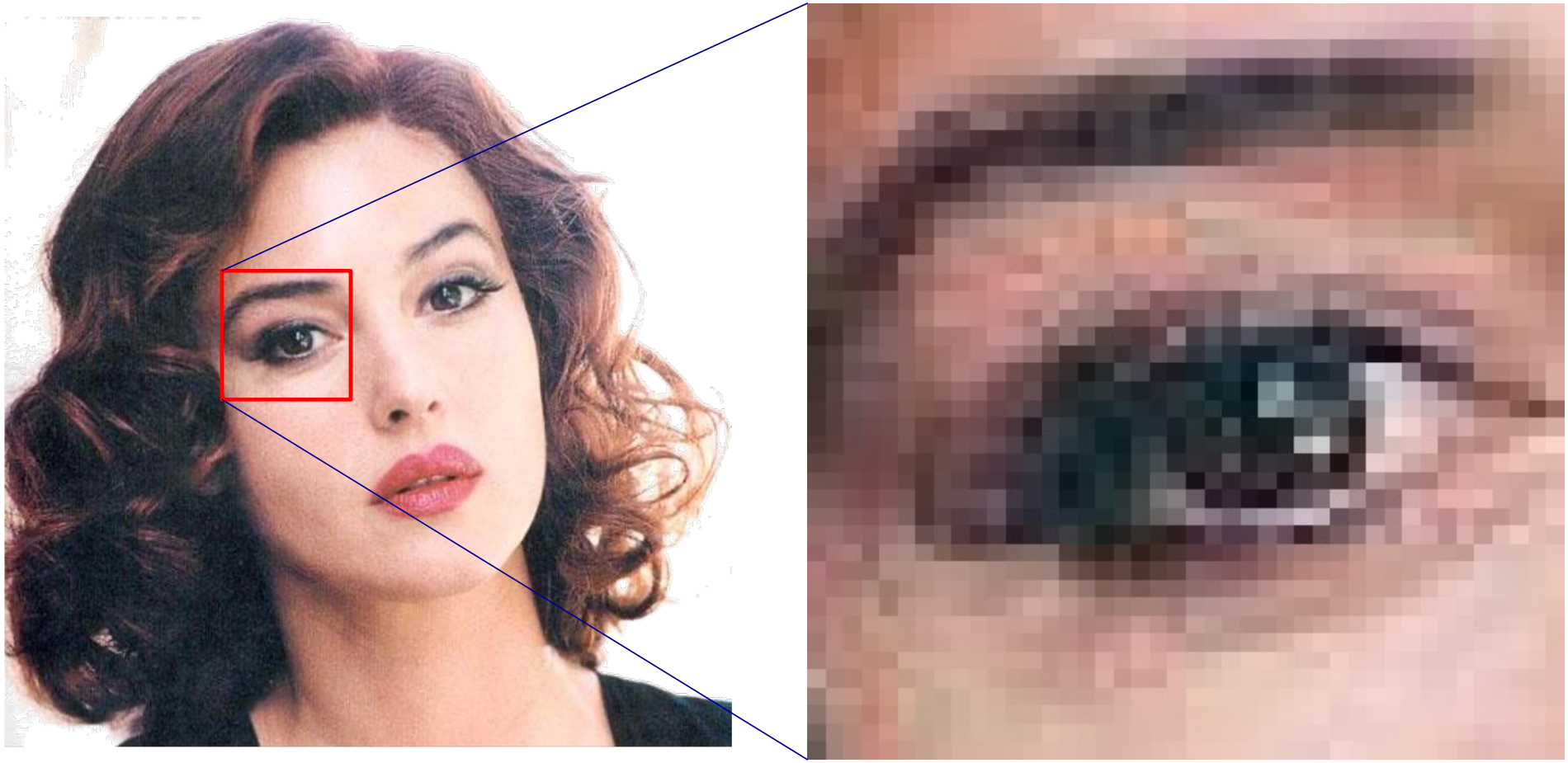
Question 1

3. There is something making the problem harder than we thought

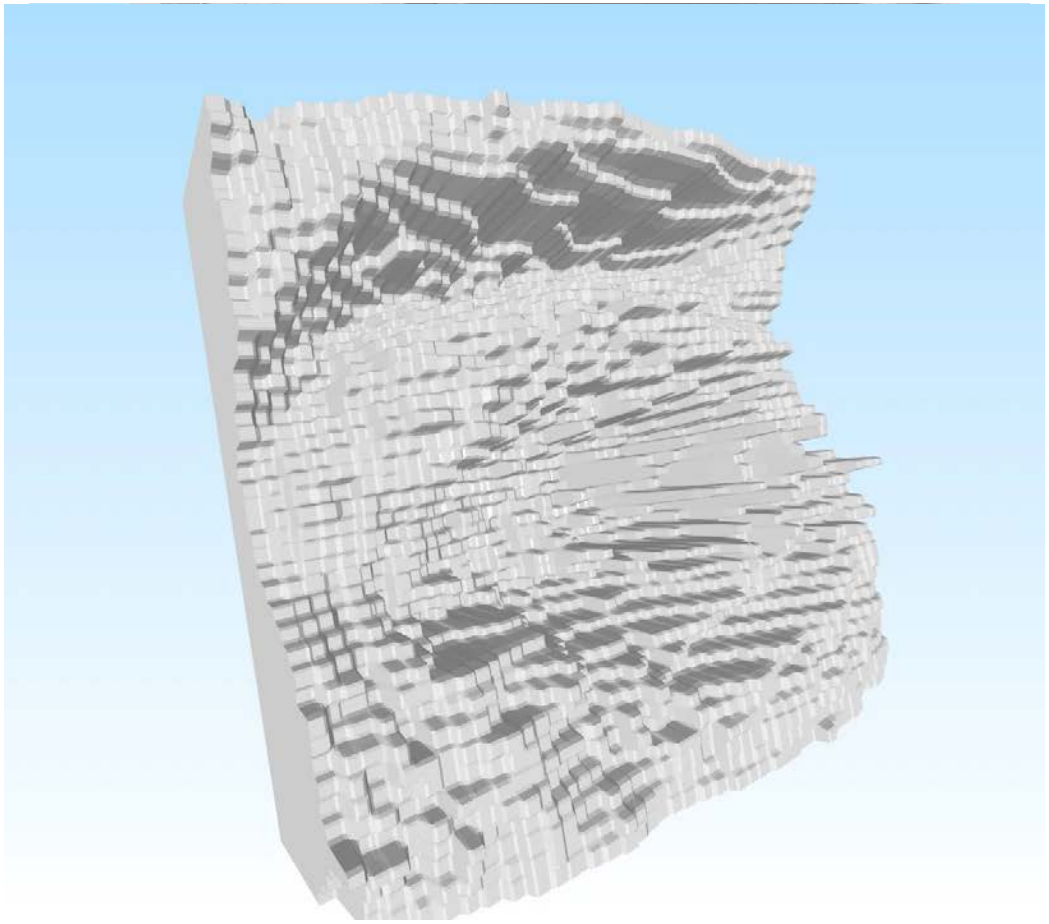
Answer 3 has a strong basis:

- **Fully automatic, data-driven learning, may drive to loose touch with the *physical nature* of the problem.**

Pixels



Pixels



Question 1

3. There is something making the problem harder than we thought

Answer 3 has a strong basis:

- Fully automatic, data-driven learning, may drive to loose touch with the physical nature of the problem.
- **The complexity of the visual world, even for one class of objects such as faces, is far greater than any canned face dataset scraped from the web.**

Visual complexity

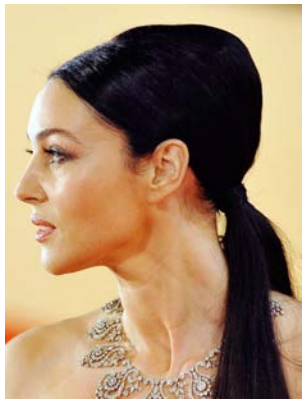
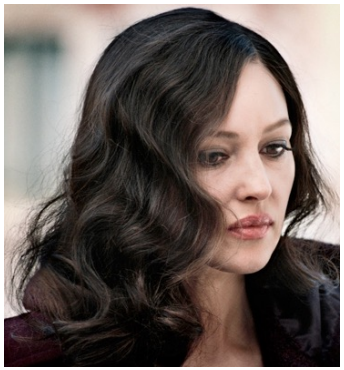
A – Aging

P – Pose

I – Illumination

E - Expression




Visual complexity

Esthetic surgery

Make up

Visual complexity

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.

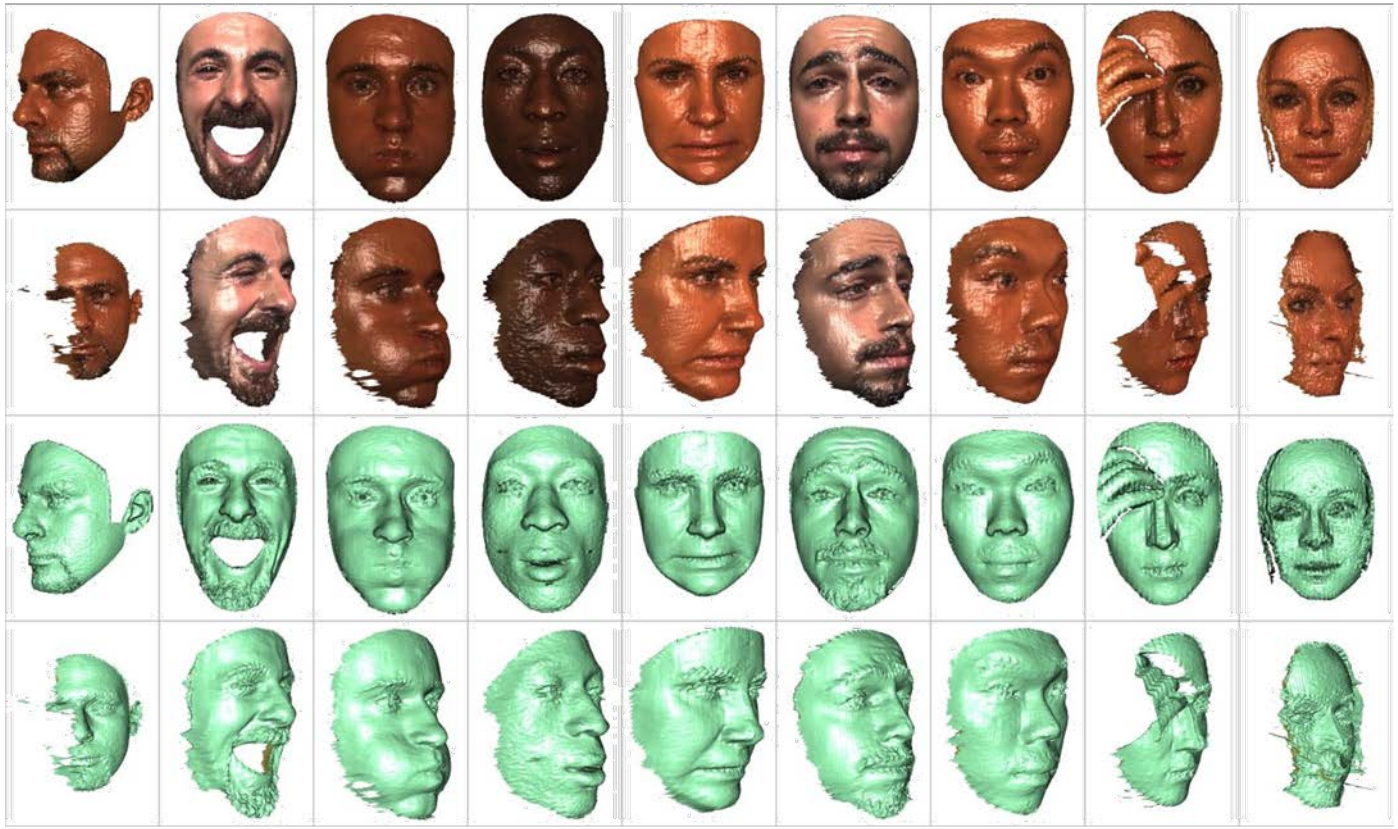
Question 1

3. There is something making the problem harder than we thought

Answer 3 has a strong basis:

- Being fascinated by the lure of automatic, data-driven learning, we lost touch with the physical nature of the problem.
- The complexity of the visual world, even for one class of objects such as faces, is far greater than any canned face dataset scraped from the web.
- **The plasticity of the face due to natural movements makes it a fully deformable object.**

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.

3D faces from mobile devices

- More than 100 subjects
- Raw data and reconstructed 3D models
- Different 3D scans acquired in 2 sessions

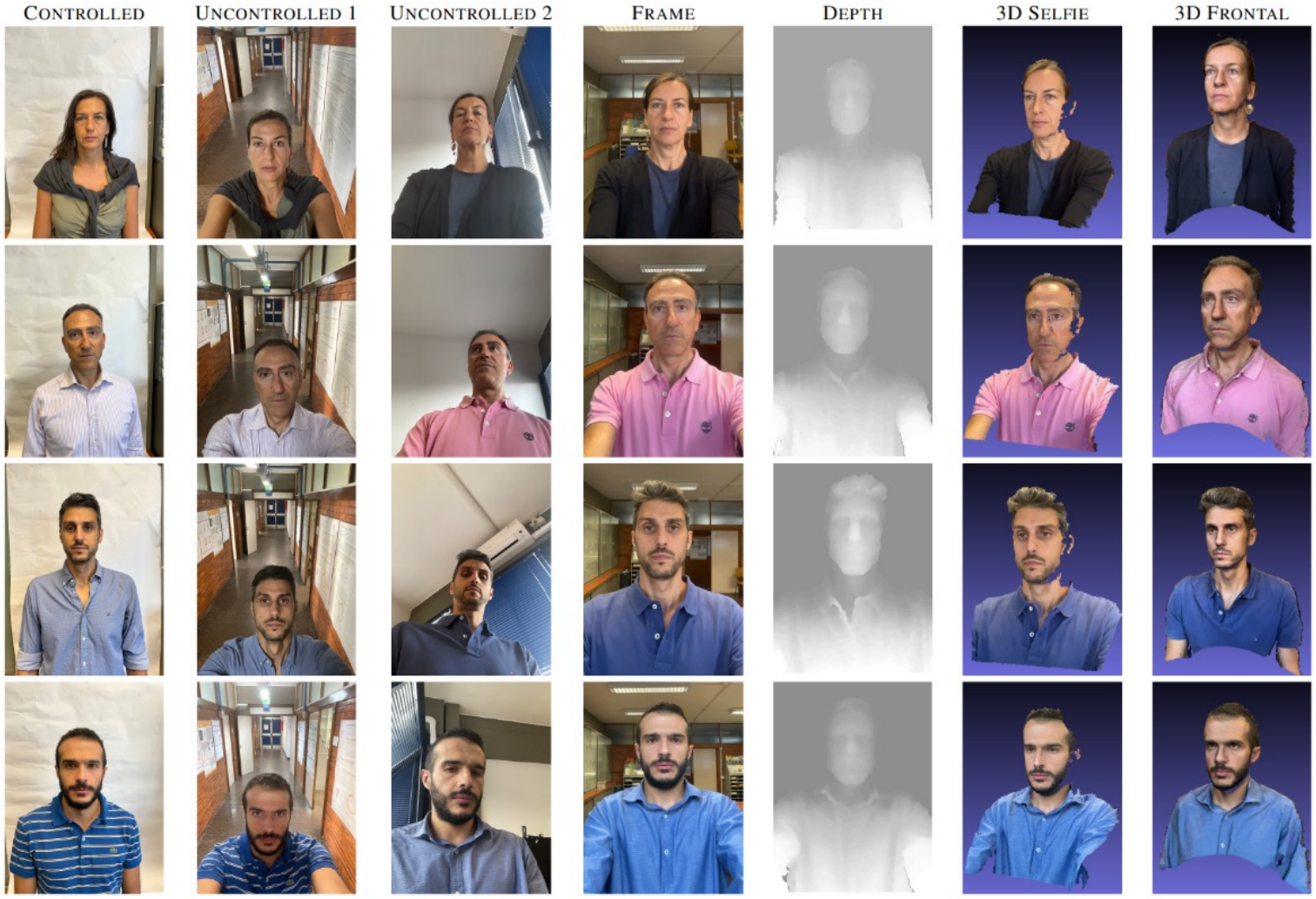
TABLE II
OVERVIEW OF TYPE OF DATA COLLECTED IN DIFFERENT ENVIRONMENTS.

ENVIRONMENT	DATA COLLECTED	NO.
Controlled	- face scan with TrueDepth sensor, frontal ill.	1
	- face scan with TrueDepth sensor, lateral ill.	1
	- high-quality photo (12 MP, frontal camera)	1
Uncontrolled	- frontal selfie scan with TrueDepth sensor	1
	- frontal selfie photo, frontal camera	2
	- selfie video by rotating around the face (1080x1920 @ 30 FPS, frontal camera)	1

TABLE III
TECHNICAL SPECIFICATION OF THE DATA COLLECTED

DATA TYPE	QUANTITY	FORMAT	RESOLUTION [HxW pixels]
Scan frame	73.8k	JPEG	1920x1080
Depth maps	73.8k	PNG	640x360
IMU data	73.8k	JSON	-
3D models	600	OBJ	90k x 150k (vert x faces)
Texture files	600	JPEG	2048x2048
Uncont images	400	JPEG	4032x3024
Cont images	200	JPEG	4032x3024
Videos	200	MOV	1920x1080@30FPS

P. Ruiu, M. Cadoni, A. Lagorio, S. Nixon, F. Casu, M. Farina, M. Fadda, G. A. Trunfio, M. Tistarelli, E. Grosso, **Uniss-MDF: A Multidimensional Face dataset for assessing face analysis on the move**, Computer Vision and Image Understanding, Volume 258, 2025,



P. Ruiu, M. Cadoni, A. Lagorio, S. Nixon, F. Casu, M. Farina, M. Fadda, G. A. Trunfio, M. Tistarelli, E. Grosso, **Uniss-MDF: A Multidimensional Face dataset for assessing face analysis on the move**, Computer Vision and Image Understanding, Volume 258, 2025,

Question 1

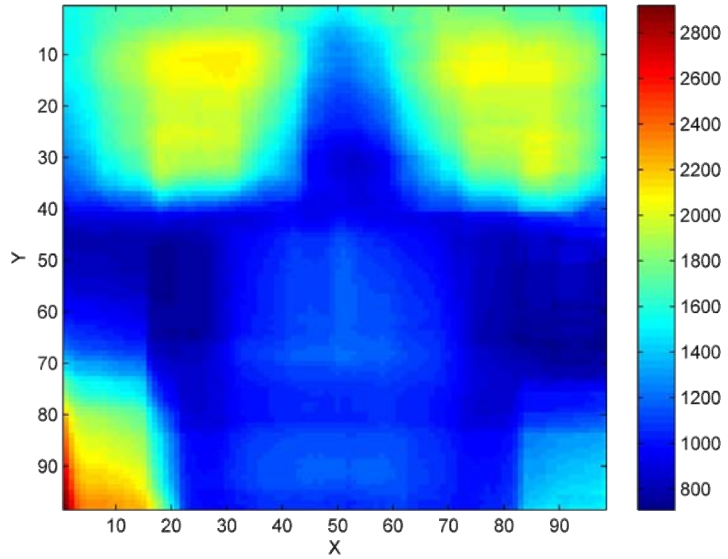
Vision Lab

3. There is something making the problem harder than we thought

Answer 3 has a strong basis:

- Being fascinated by the lure of automatic, data-driven learning, we lost touch with the physical nature of the problem.
- The complexity of the visual world, even for one class of objects such as faces, is far greater than any canned face dataset scraped from the web.
- The plasticity of the face due to natural movements makes it a truly deformable object.
- **There are subtile, even coarse, changes over small time spans.**

Short-time aging



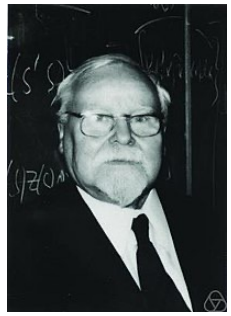
$$d_{I_1,I_2}(x,y) = \frac{1}{2}\left(\frac{1}{|P_{I_1}|}\sum_{p\in P_{I_1}}\omega(p) + \frac{1}{|P_{I_2}|}\sum_{q\in P_{I_2}}\omega(q)\right)$$

$$E(x,y,t) = \frac{1}{T}\sum_{i=1}^{T} d_{I_i,I_{i+t}}(x,y)$$

M. Ortega, L. Brodo, M. Bicego, M. Tistarelli (2009) "**Measuring changes in face appearance through aging**", in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR 2009, pp.107-113, 2009.

An hard, ill-posed problem

Jacques Hadamard



Andrej Tikhonov

Two adverse conditions:

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

Solution: Regularization

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

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

Question 1

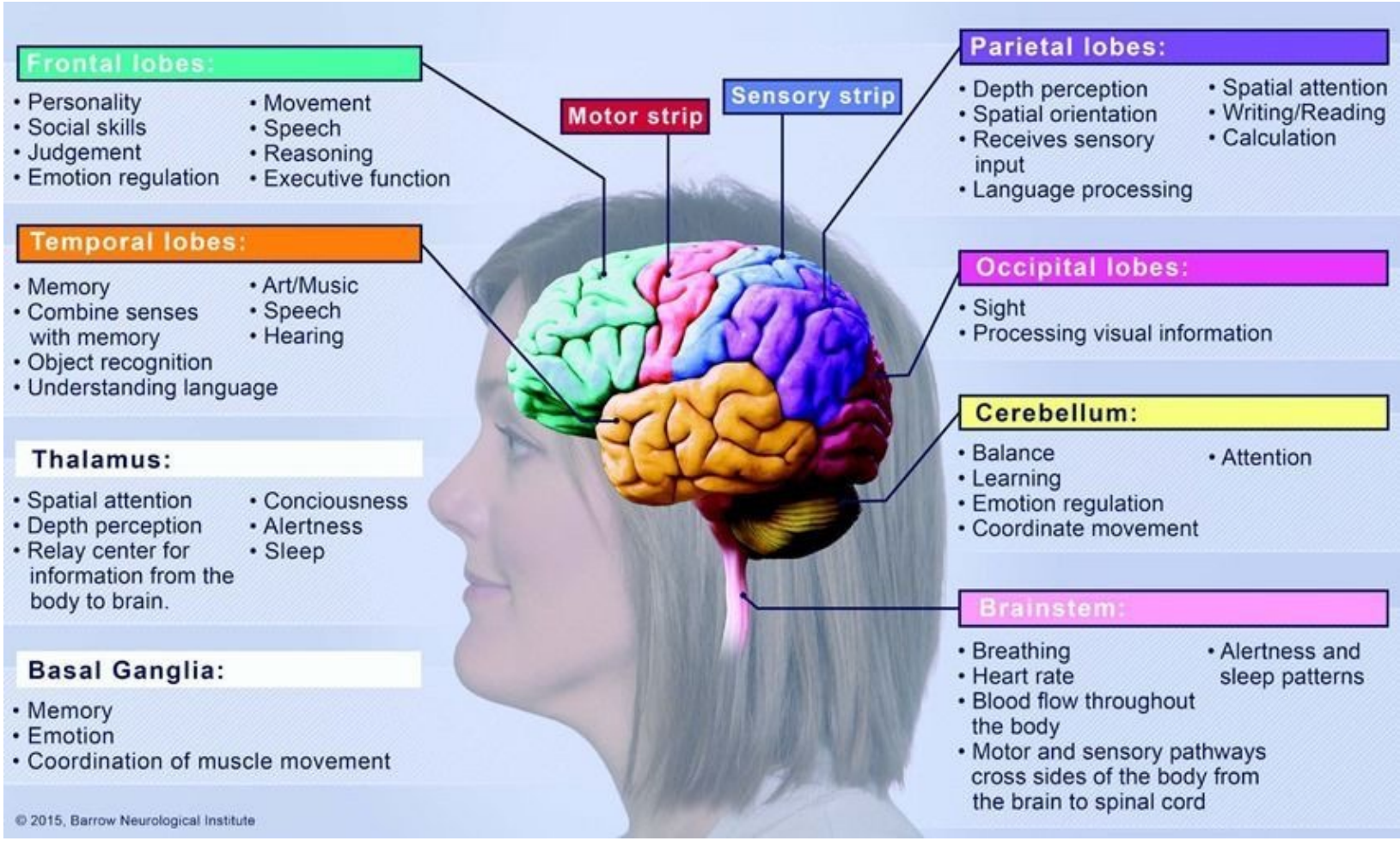
Vision Lab

3. There is something making the problem harder than we thought

Answer 3 has a strong basis:

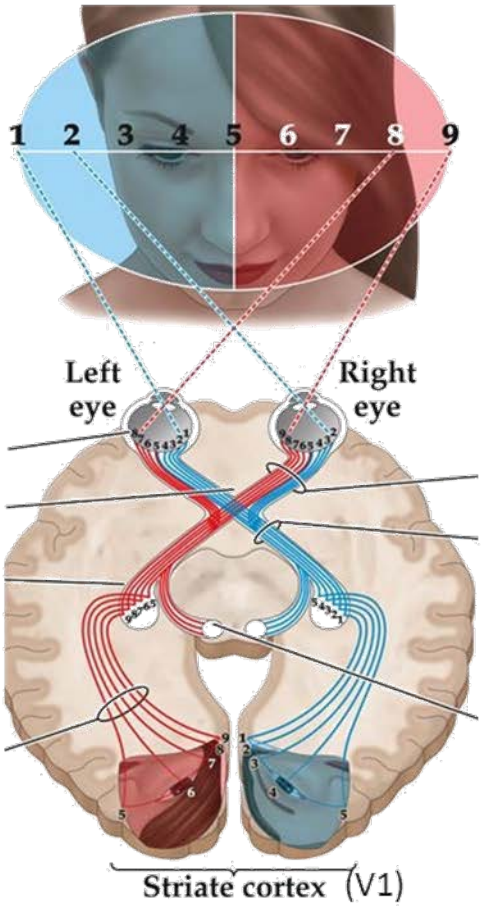
- Fully automatic, data-driven learning, may drive to loose touch with the physical nature of the problem.
- The complexity of the visual world, even for one class of objects such as faces, is far greater than any canned face dataset scraped from the web.
- The plasticity of the face due to natural movements makes it a truly deformable object.
- There are subtile, even coarse, changes over small time spans.
- **We even *MAKE the problem harder.* The architecture of the HSV is designed to recognize objects: we don't *learn* to recognize.**

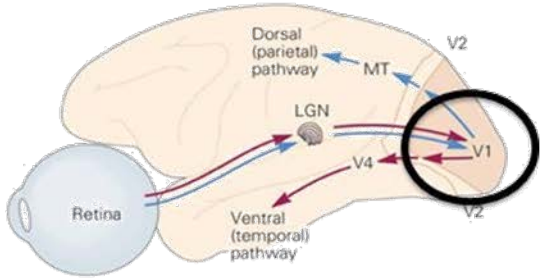
Brain models



Retinotopic mapping

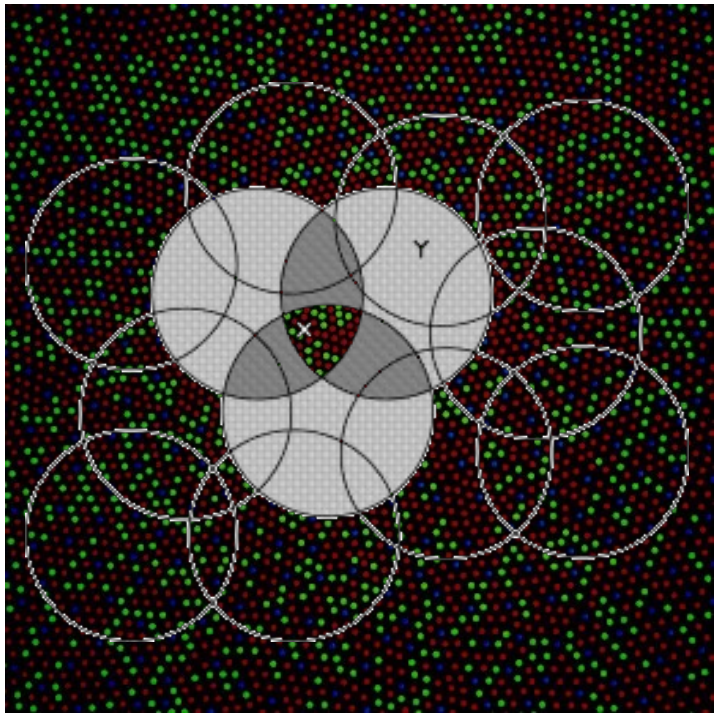
V1 retinotopic maps

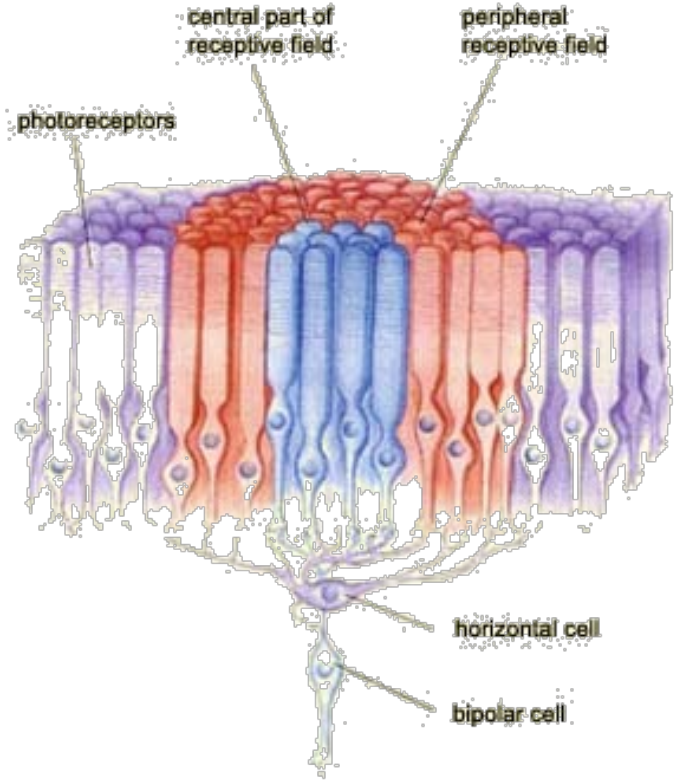




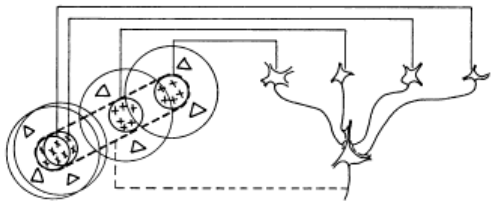
- Each point of the visual field maps on to a local group of neurons in V1.
- Retinotopy = Remapping of retinal image onto cortical surface
- Foveal region uses more of V1 (greater magnification factor)

Receptive fields

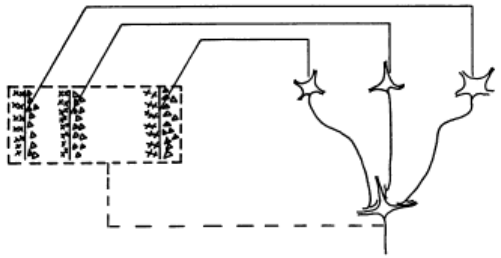




Receptive fields



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.

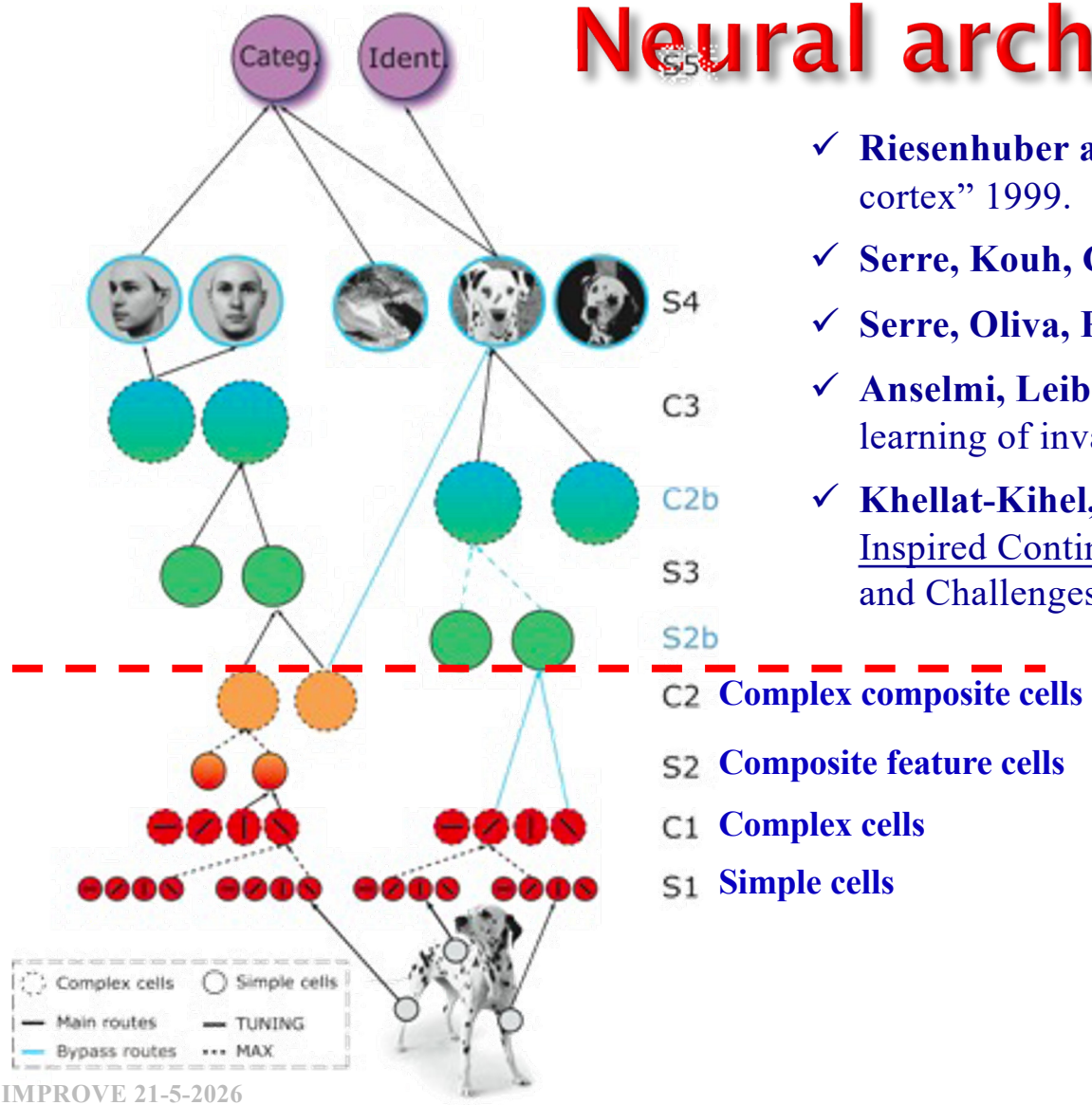


Text-fig. 20. Possible scheme for explaining the organization of complex receptive fields. A number of cells with simple fields, of which three are shown schematically, are imagined to project to a single cortical cell of higher order. Each projecting neurone has a receptive field arranged as shown to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

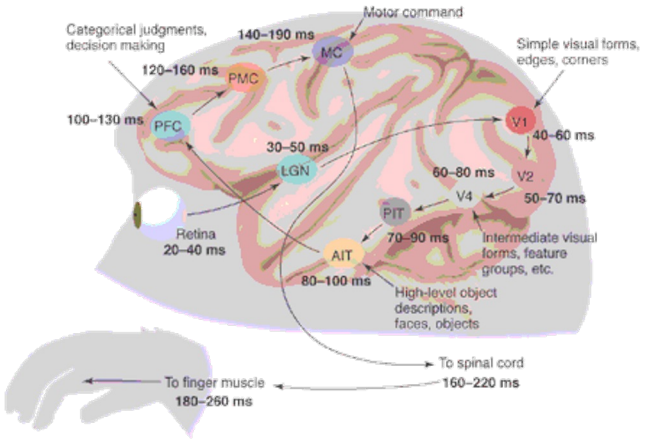
Simple and Complex cells

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

Neural architectures



- ✓ **Riesenhuber and Poggio** " Hierarchical models of object recognition in cortex" 1999.
- ✓ **Serre, Kouh, Cadieu, Knoblich, Kreiman & Poggio** 2005.
- ✓ **Serre, Oliva, Poggio** 2007.
- ✓ **Anselmi, Leibo, Rosasco, Mutch, Tacchetti, and Poggio**, "Unsupervised learning of invariant representations", Theoretical Computer Science, 2015.
- ✓ **Khellat-Kihel, Lagorio, Tistarelli** "Foveated Vision for Biologically Inspired Continuous Face Authentication" - Selfie Biometrics: Advances and Challenges, 2019



Question 1

- **If the solution is so simple, why do we still need to pursue research on this topic?**

1. We all have to pay our bills… from research grants
2. We are not so clever
3. There is something making the problem harder than we thought
4. **There is a strong demand to build more *accurate*, *robust* and *efficient* systems**

Question 1

4. There is a strong demand to build more accurate, robust and efficient systems

- This has been the **success** of biometrics, but also its **suicide**.
- Probably the best success story of **CV** and **PR**, risks to revert **scientific investigation** into **engineering design**.
- However, there are still several challenging **applications** requiring ***clever scientific answers***.
  - **Examples**: multi-modal integration; fake detection; adversarial learning; continual learning; behavior analysis; etc.

Question 1

- **If the solution is so simple, why do we still need to pursue research on this topic?**

1. We all have to pay our bills… from research grants
2. We are not so clever
3. There is something making the problem harder than we thought
4. There is a strong demand to build more accurate, robust and efficient systems
5. **We are all driven by curiosity**

Question 1

5. We are all driven by curiosity

Answer 5 is the last, but not the least important:

- **There is no scientific research without curiosity**.
- This requires to make **questions** and look for **answers**.
- I have my own **questions** and I gladly share with you…

Question 2

- **What are the drawbacks and limitations of current deep learning models? How far can we go by exploiting increasing amounts of data?**

Always more data… and energy

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.

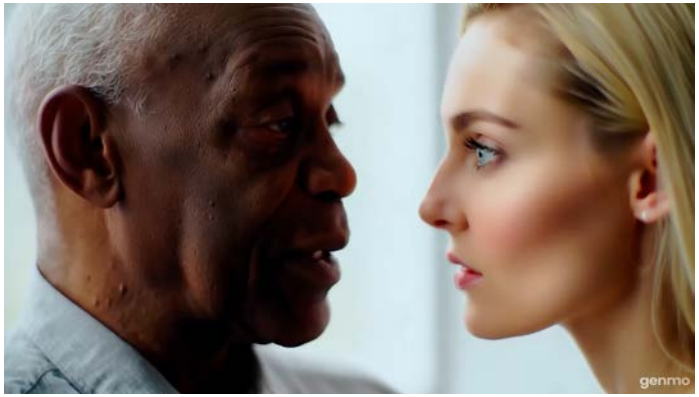
Method	#Image	LFW	YTF
DeepID [32]	0.2M	99.47	93.20
Deep Face [33]	4.4M	97.35	91.4
VGG Face [24]	2.6M	98.95	97.30
FaceNet [29]	200M	99.63	95.10
Baidu [16]	1.3M	99.13	-
Center Loss [38]	0.7M	99.28	94.9
Range Loss [46]	5M	99.52	93.70
Marginal Loss [9]	3.8M	99.48	95.98
SphereFace [18]	0.5M	99.42	95.0
SphereFace+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

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

Face generation

Please create a highly realistic, photographic style image of an adult/young man/woman. The image contains the face smiling. Natural window lighting, photorealistic, shot on Canon EOS R5, 85mm lens, f/1.8 aperture, UHD resolution, professional photography, hyper-detailed skin texture, volumetric lighting, HDR, cinematic depth.

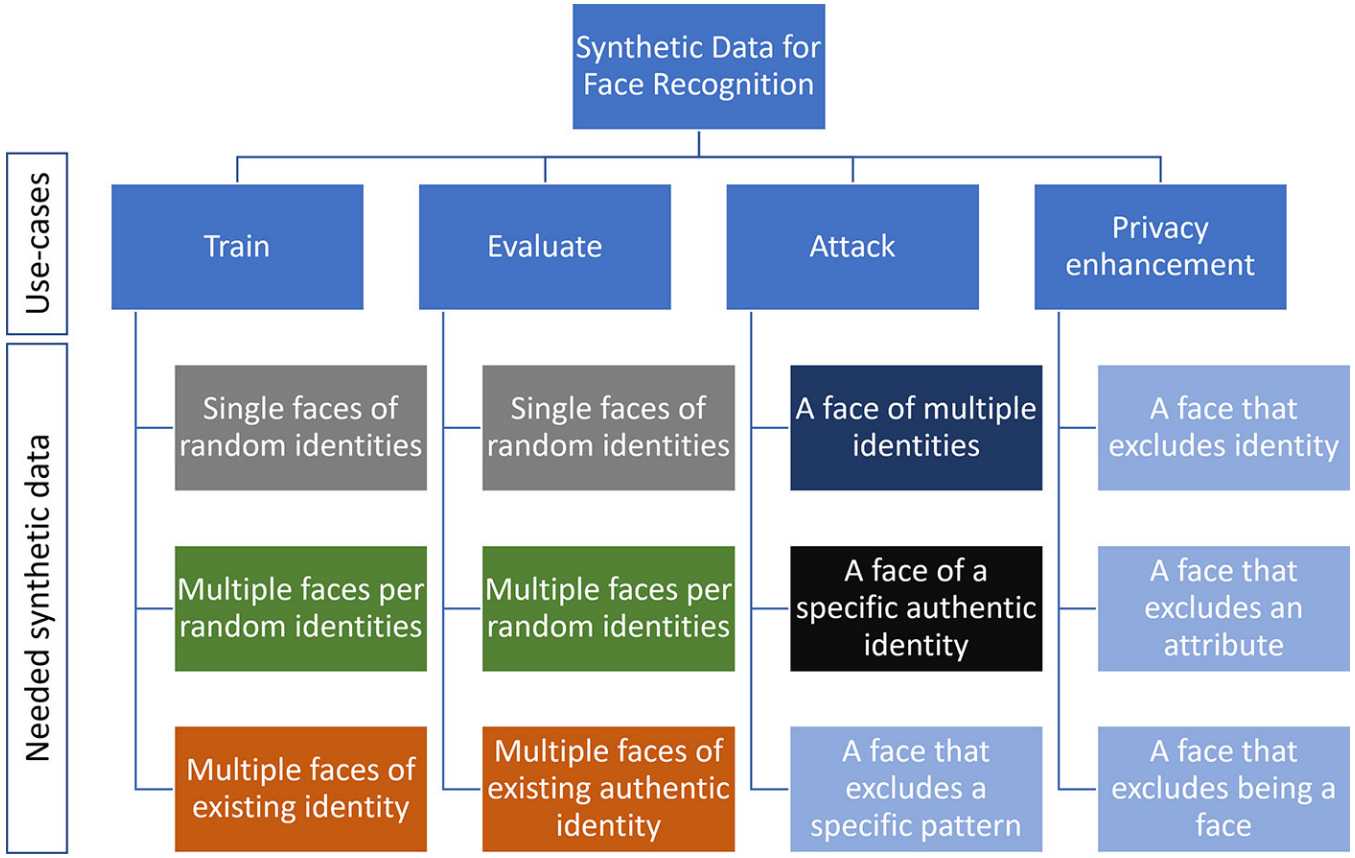
Leonardo.AI



Genmo.AI

AWS Nova Reel

Face generation



Fadi Boutros, Vitomir Struc, Julian Fierrez, Naser Damer, **Synthetic data for face recognition: Current state and future prospects**, Image and Vision Computing, Volume 135, 2023.

Face generation

- **Side effects**:
  - Identity **leakage** & **duplication**
  - **Biases** inherited from training data
  - Potential **overfitting** to synthetic features
  - Dependence on the **quality** of the *realism transfer model*
  - Potential residual **artifacts** affecting model training
  - Ensuring **consistency** across generated images
  - Ensuring the synthetic data accurately **represents** real-world complexities
  - Accuracy of **control** over synthetized faces

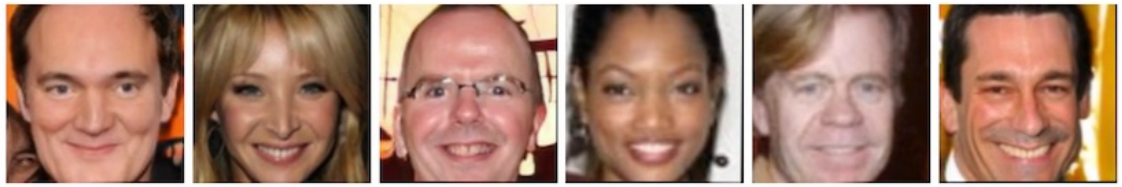
Fadi Boutros, Vitomir Struc, Julian Fierrez, Naser Damer, **Synthetic data for face recognition: Current state and future prospects**, Image and Vision Computing, Volume 135, 2023.

Face generation

- **Identity leakage**

Synthetic

Synthetic



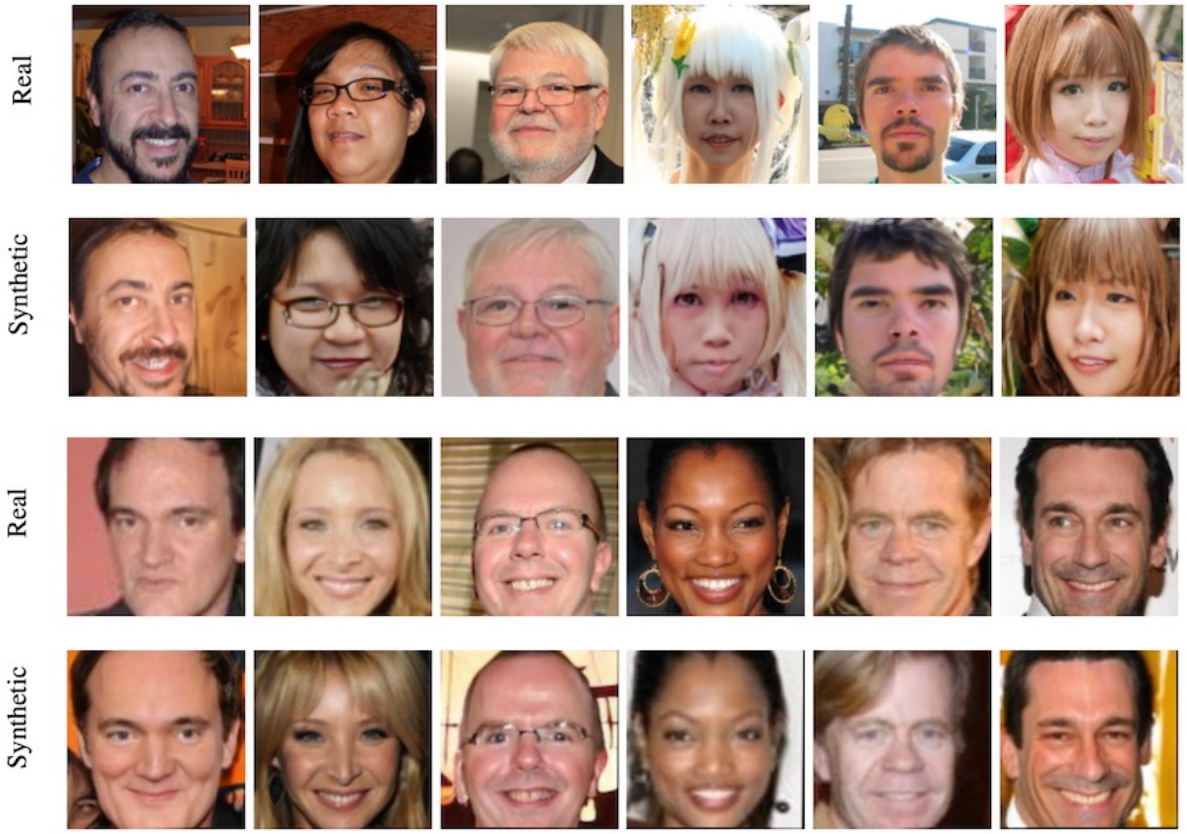
Generation models

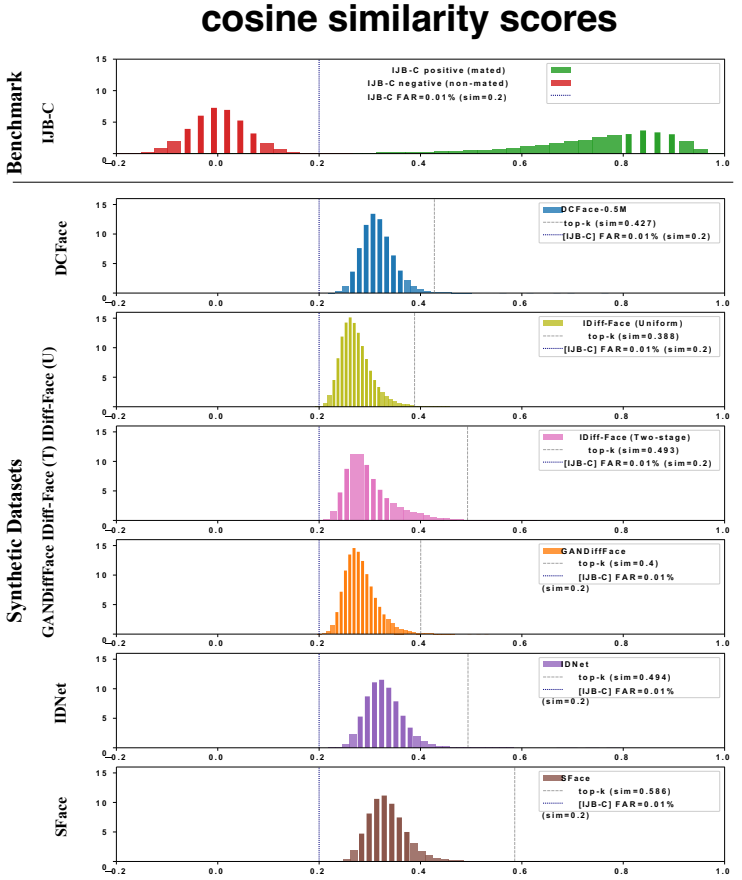
Reference	Synthetic Dataset	Generator	Training Dataset
[8]	SFace	StyleGAN-ADA (identity-conditioned)	CASIA-WebFace
[9]	IDNet	StyleGAN-ADA (identity-conditioned)	CASIA-WebFace
[10]	DCFace	new diffusion model (identity and style conditioned)	CASIA-WebFace
[11]	IDiff-Face (Uniform) IDiff-Face (Two-stage)	new diffusion model (identity-conditioned)	FFHQ
[12]	GANDiffFace	StyleGAN (pretrained) DreamBooth (pretrained)	FFHQ LAION

Hatef Otroshi Shahreza and Sébastien Marcel, **Unveiling Synthetic Faces: How Synthetic Datasets Can Expose Real Identities**, AdvML-Frontiers'24: The 3rd Workshop on New Frontiers in Adversarial Machine Learning@NeurIPS'24, Vancouver, CA.

Face generation

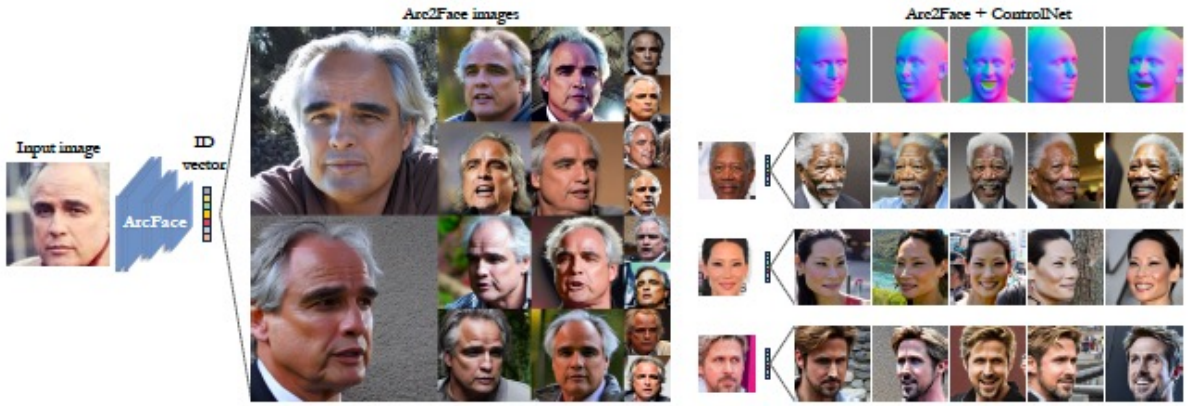
Identity leakage





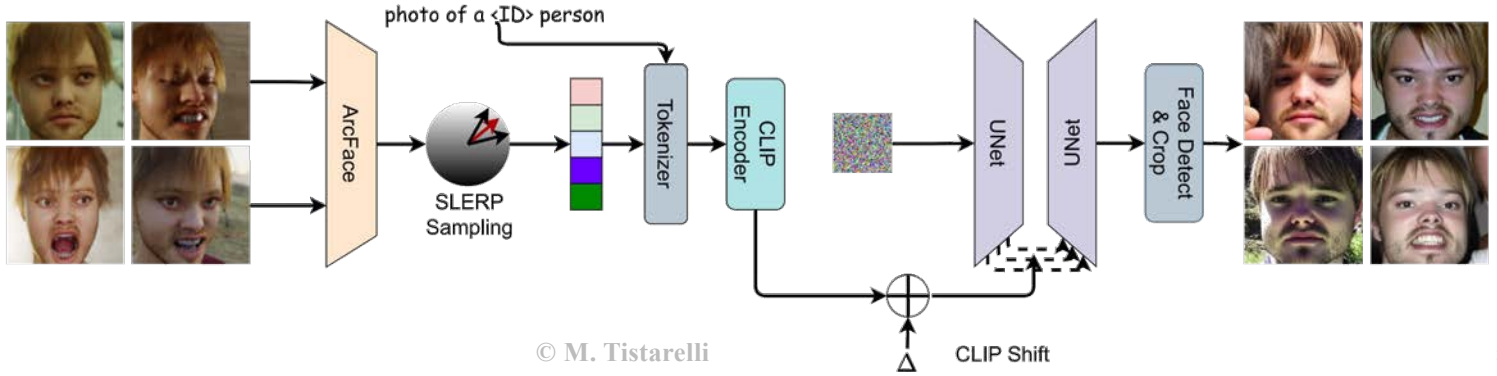
Hatef Otroshi Shahreza and Sébastien Marcel, **Unveiling Synthetic Faces: How Synthetic Datasets Can Expose Real Identities**, AdvML-Frontiers'24: The 3rd Workshop on New Frontiers in Adversarial Machine Learning@NeurIPS'24, Vancouver, CA.

Foundation models

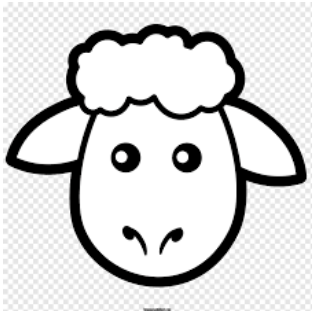


F. P. Papantoniou, A. Lattas, S. Moschoglou, J. Deng1, B. Kainz and S. Zafeiriou, **Arc2Face: A Foundation Model for ID-Consistent Human Faces**, In ECCV, volume 1, page 3, 2024.

Anjith George and Sebastien Marcel, **Digi2Real: Bridging the Realism Gap in Synthetic Data Face Recognition via Foundation Models**, WACV 2025 (Workshop on SynRDinBAS).



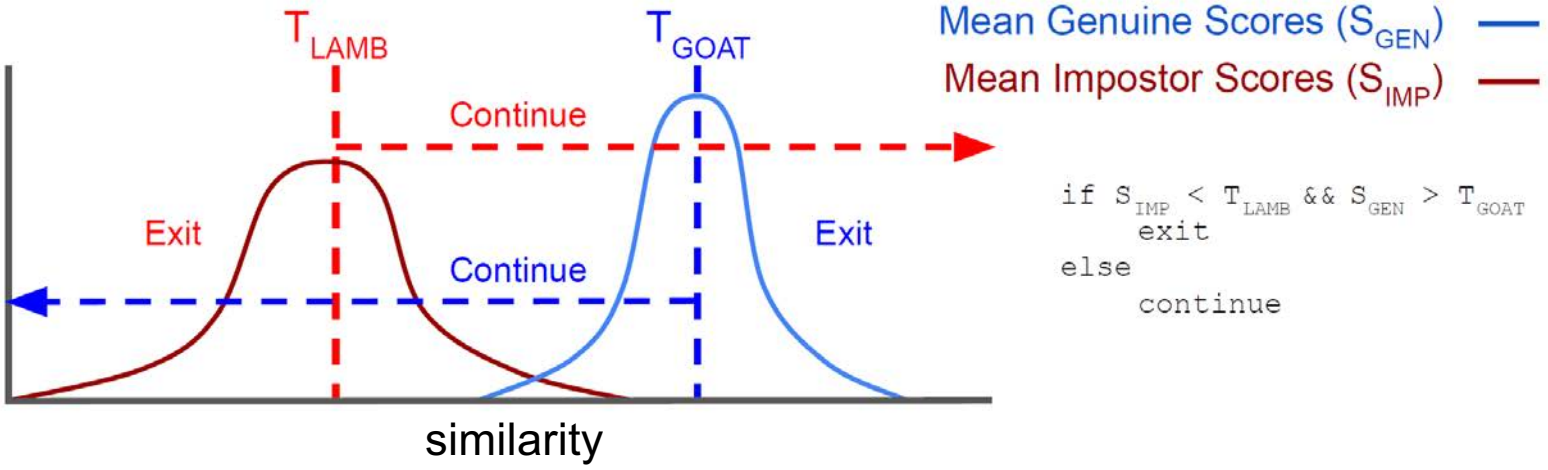
The Doddington zoo



- Sheep -> easy to be recognised
- Goats -> less likely to be recognised correctly
- Lambs -> more likely to be recognised incorrectly

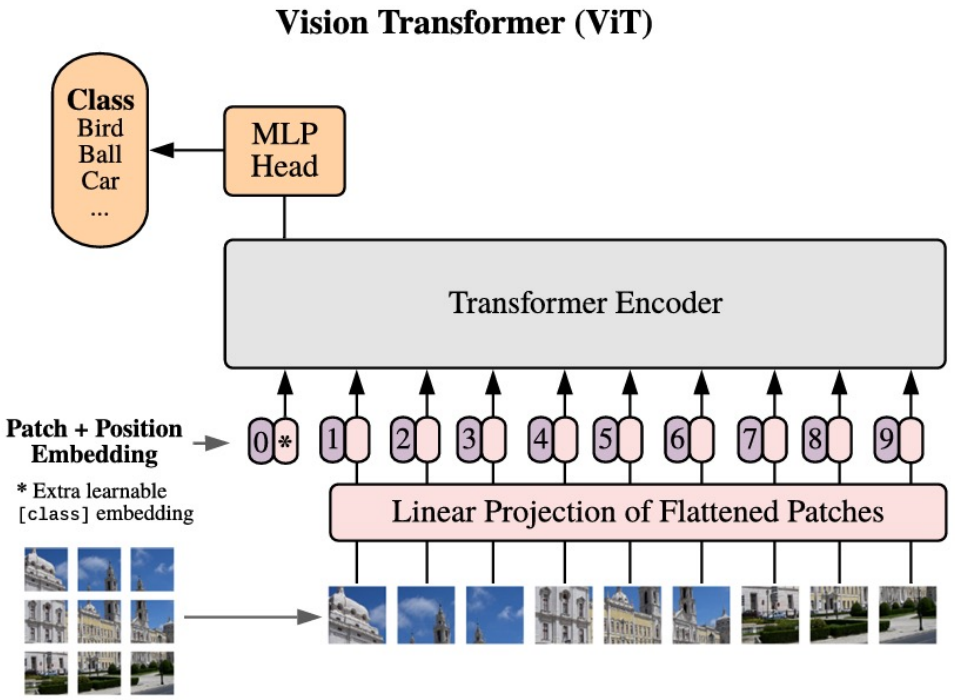
G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. Technical report, National Inst of Standards and Technology Gaithersburg Md, 1998.

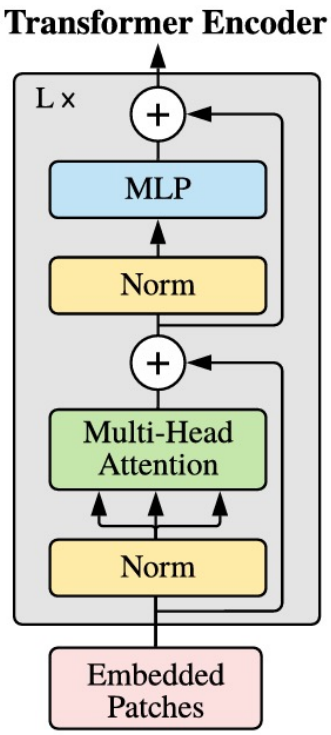
$$cos(\theta) = \frac{A \cdot B}{\|A\| \, \|B\|}$$



Ross, A.; Rattani, A.; Tistarelli, M.: **Exploiting the "doddington zoo" effect in biometric fusion**. In: IEEE 3rd International Conf. on Biometrics: Theory, Applications, and Systems. IEEE, pp. 1–7, 2009.

Vision Transformers

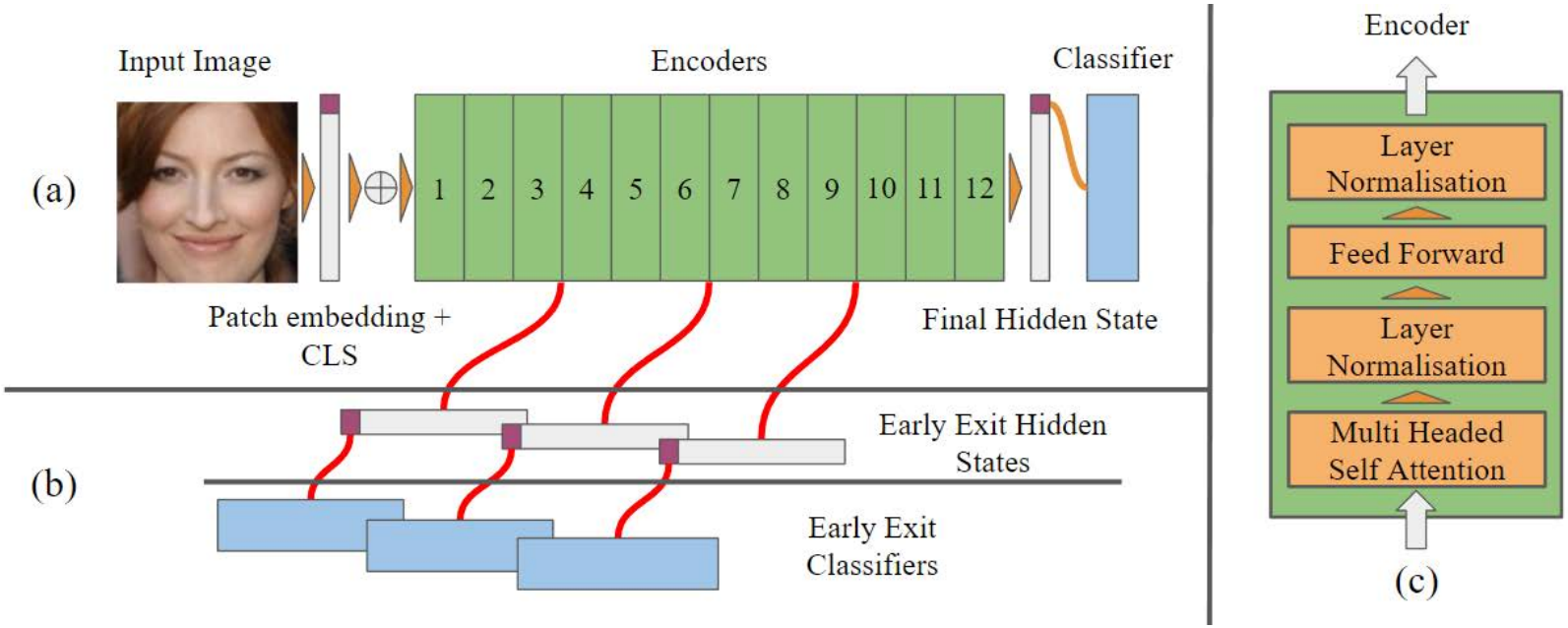




$$L_{CE} = -\sum_{i=1}^{n} t_i log(p_i)$$

Dosovitskiy, Alexey, et al. **An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale**. International Conference on Learning Representations. arXiv preprint arXiv:2010.11929, 2020.

ViT Early Exit



$$L_{CE}^{EE} = -\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n} t_i^j log(p_i^j)$$

S. Nixon, P. Ruiu, M. Cadoni, A. Lagorio and M. Tistarelli, **Exploiting Face Recognizability with Early Exit Vision Transformers**, 2023 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2023, pp. 1-7

S. Nixon, P. Ruiu, M. Cadoni, A. Lagorio and M. Tistarelli, **Assessing bias and computational efficiency in vision transformers using early exits**, EURASIP Journal on Image and Video Processing, 2025(1), 2. https://doi.org/10.1186/s13640-024-00658-9

Results

- Early Exits **reduce computational cost** by up to ~18% with ~1% accuracy degradation.
- Configurations with **fewer exits** provide the **best accuracy–efficiency** trade-off.
- **Demographic bias** differs across cohorts and is amplified at earlier exits.
- **Balanced training data** reduces bias more effectively than loss re-weighting.
- **Early Exits enable efficient face recognition** but require explicit bias mitigation.

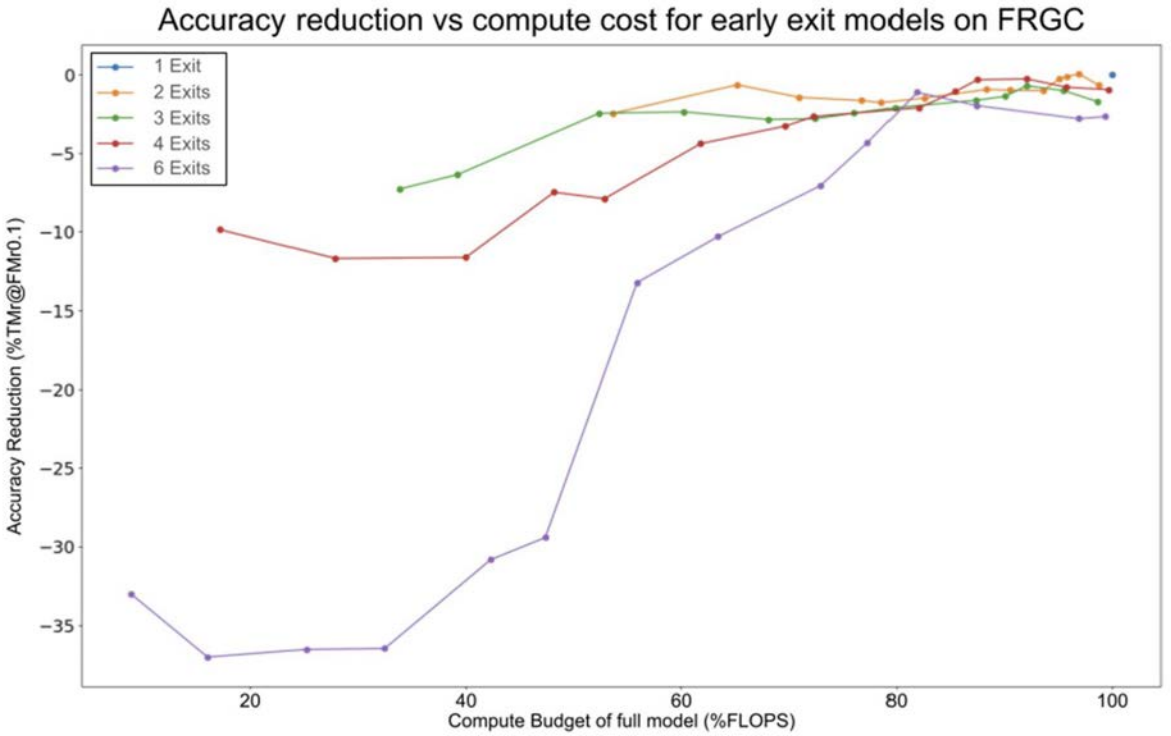


Fig. 5 Graph showing TMr@FMr0.1% for various thresholds

S. Nixon, P. Ruiu, M. Cadoni, A. Lagorio and M. Tistarelli, **Assessing bias and computational efficiency in vision transformers using early exits**, EURASIP Journal on Image and Video Processing, 2025(1), 2. https://doi.org/10.1186/s13640-024-00658-9

Question 2

- **What are the drawbacks and limitations of current deep learning models? How far can we go by exploiting increasing amounts of face data?**

Didn’t we forget something?

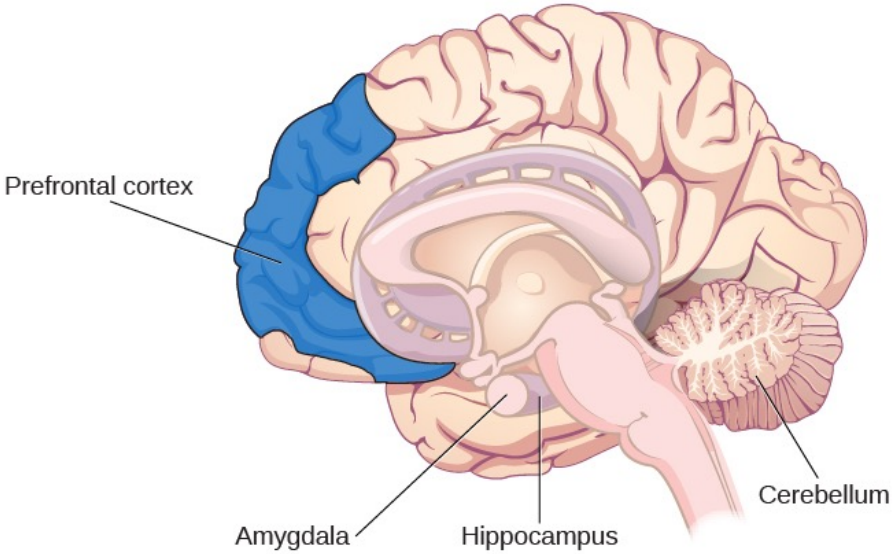
Question 2

- **What are the drawbacks and limitations of current deep learning models? How far can we go by exploiting increasing amounts of face data?**

Didn’t we forget something?

Yes! …Memory!

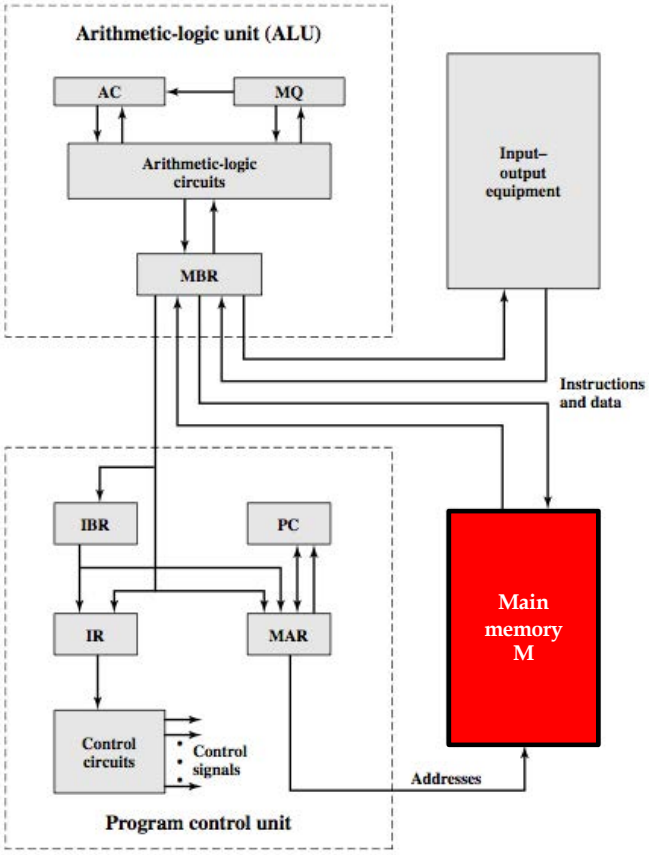
Something to remember



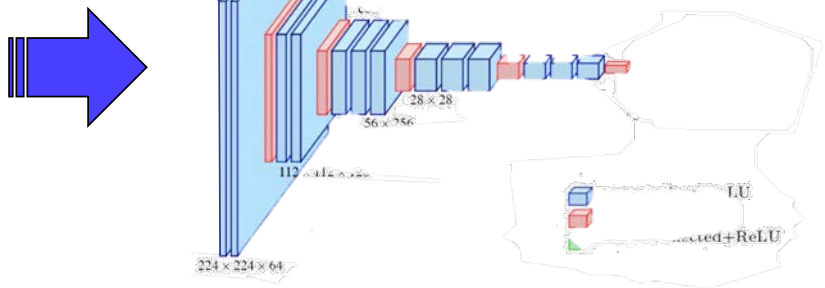
1. The **amygdala** is involved in fear and **fear memories**.
2. The **hippocampus** is associated with **declarative and episodic memory** as well as **recognition memory**.
3. The **cerebellum** plays a role in processing **procedural memories**, such as how to play the piano.
4. The **prefrontal cortex** appears to be involved in **remembering semantic tasks**.

IAS computer

(1952, Princeton USA)

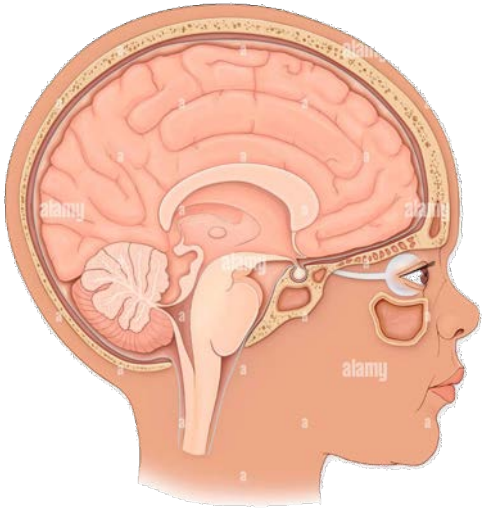


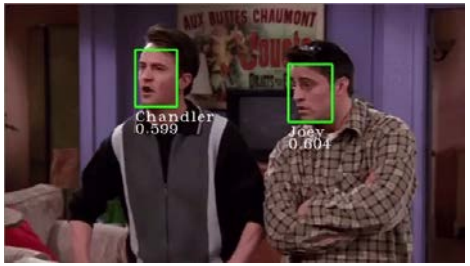
Memory models



Pre-canned memory

Continuous memory





https://towardsdatascience.com/s01e01-3eb397d458d

Kramer RSS. 2021. **Forgetting faces over a week: investigating self-reported face recognition ability and personality**. PeerJ 9:e11828 http://doi.org/10.7717/peerj.11828

Memory models

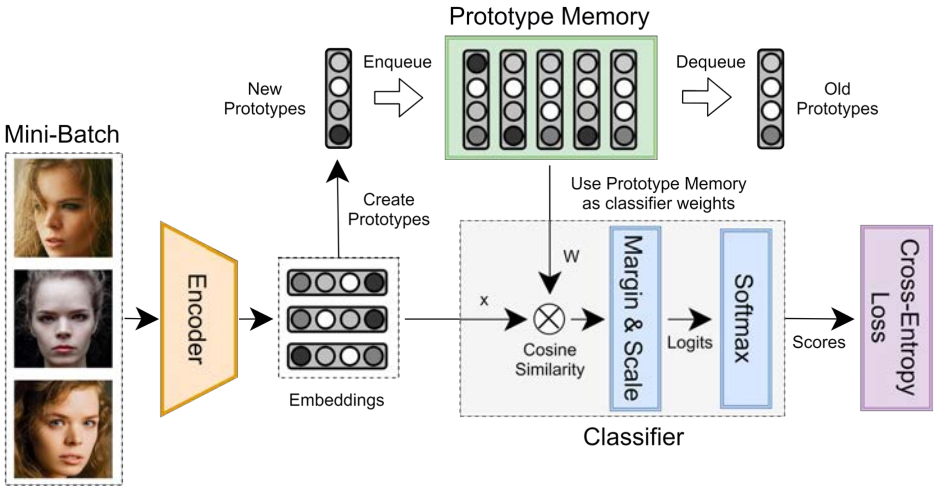


FIGURE 1. Scheme of the face representation learning with Prototype Memory. Mini-batch with several images per class is passed through the encoder, producing exemplar embeddings. Same-class embeddings are used to generate new class prototypes. These prototypes are enqueued to the memory module and used as the classifier weights in softmax classifier-based training. When the memory is full, the oldest prototypes are dequeued and disposed of.

E Smirnov, N Garaev, V Galyuk, E Lukyanets, "**Prototype Memory for Large-Scale Face Representation Learning**" IEEE Access, 2022

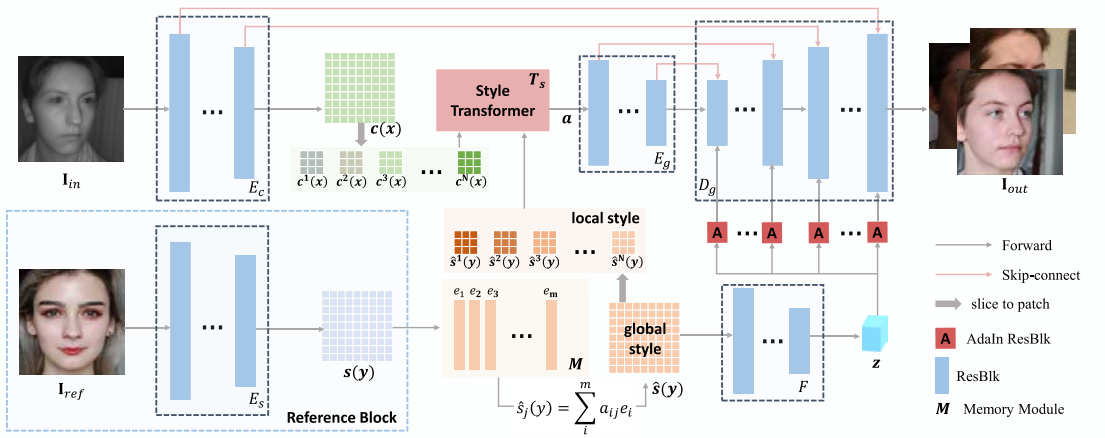
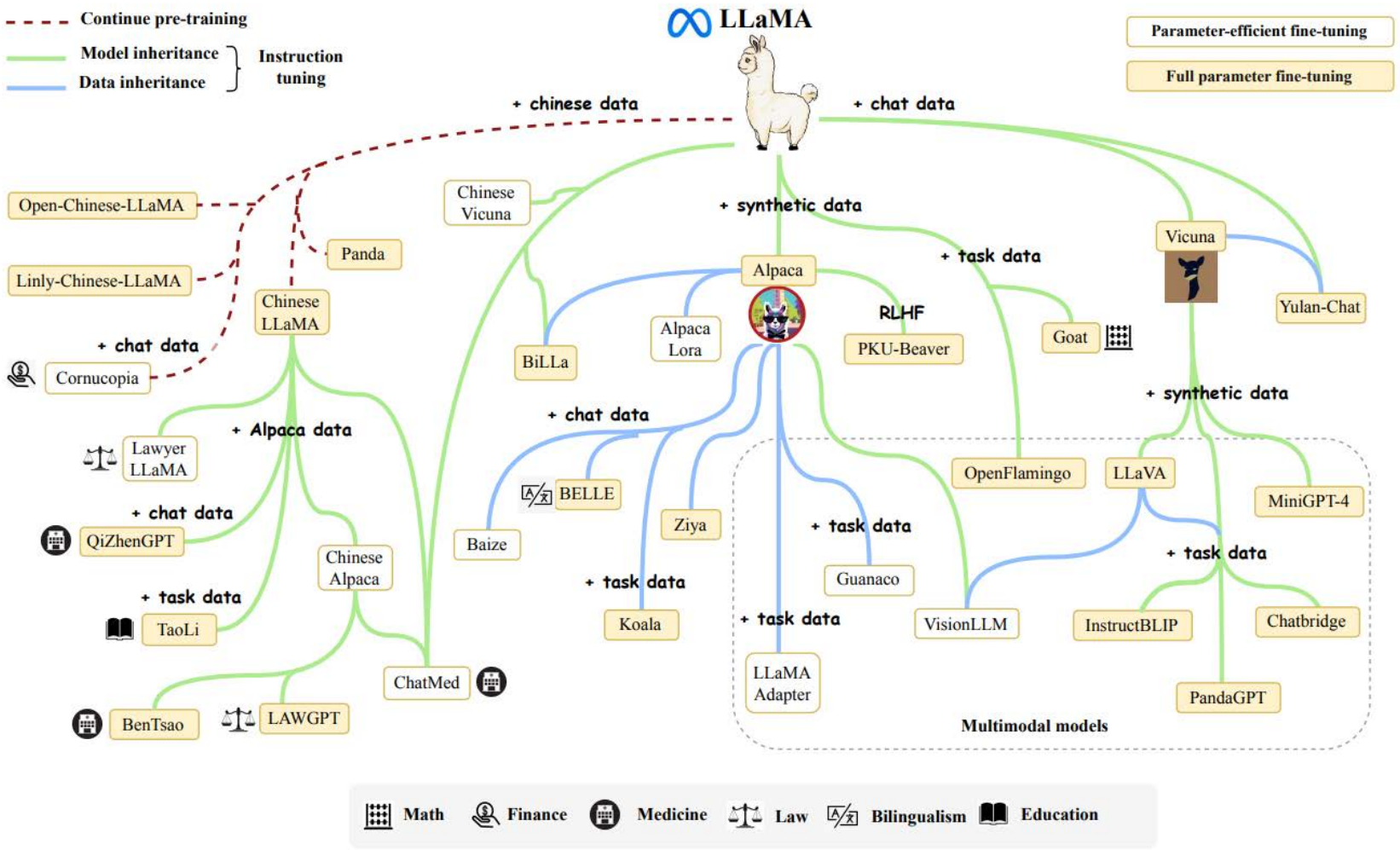


Fig. 1. Overall architecture of the proposed memory-modulated transformer network (MMTN). MMTN is composed of two branches that are responsible for content and style encoding, respectively. The content encoder *Ec* first encodes the input image I*in* into the content feature map *c*(*x*). Similarly, the style feature map *s*(*y*) is produced by the style encoder *Es* with the reference image I*ref* as the input. This feature map *s*(*y*) is further fed into the memory module *M* to produce the reconstructed feature *s*ˆ(*y*), which is utilized in two ways. For one way, *s*ˆ(*y*) is fed into a mapping network *F* to generate the global style code *z*, which is integrated into the decoder *Dg* with AdaIn. For the other way, *s*ˆ(*y*) and *c*(*x*) are cropped into patches to be fed into the style transformer module *Ts*. The output of *Ts* is further fed into the encoder *Eg* and decoder *Dg* to generate the output images I*out*.

M. Luo, H. Wu, H. Huang, W. He and R. He, "**Memory-Modulated Transformer Network for Heterogeneous Face Recognition**," in *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2095-2109, 2022, doi: 10.1109/TIFS.2022.3177960.

Memory for everybody

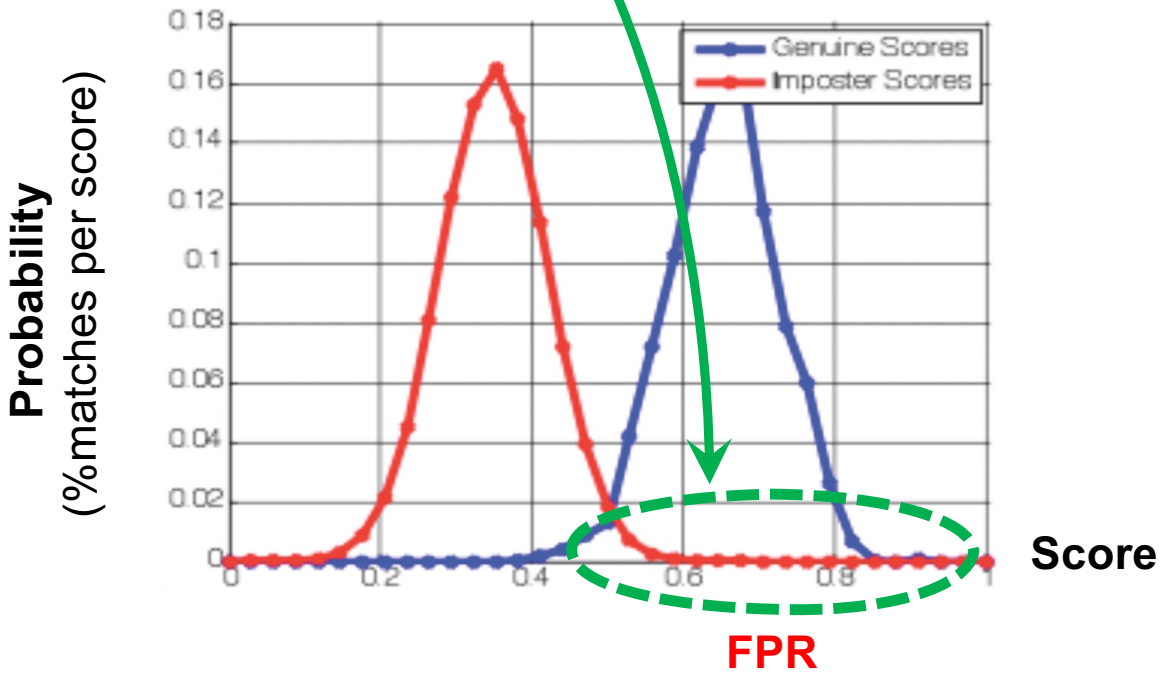


Exclusive vs Inclusive

- Biometrics and person identification has been largely dominated by **security** and **law enforcement (exclusive)** applications.
- **By design** these are applications requiring ***exclusive*** identification to avoid unlegitimate user access.
- Decisions are to be made against **large amounts of data and classes**.
- This requires to "***push up***" the limit of the **True Positive Matching Rate**.

Exclusive vs Inclusive

- This **segment** of the bimodal distribution has been always addressed… and stressed (**exclusive**).
- **TPR** is maximized against **FPR**.

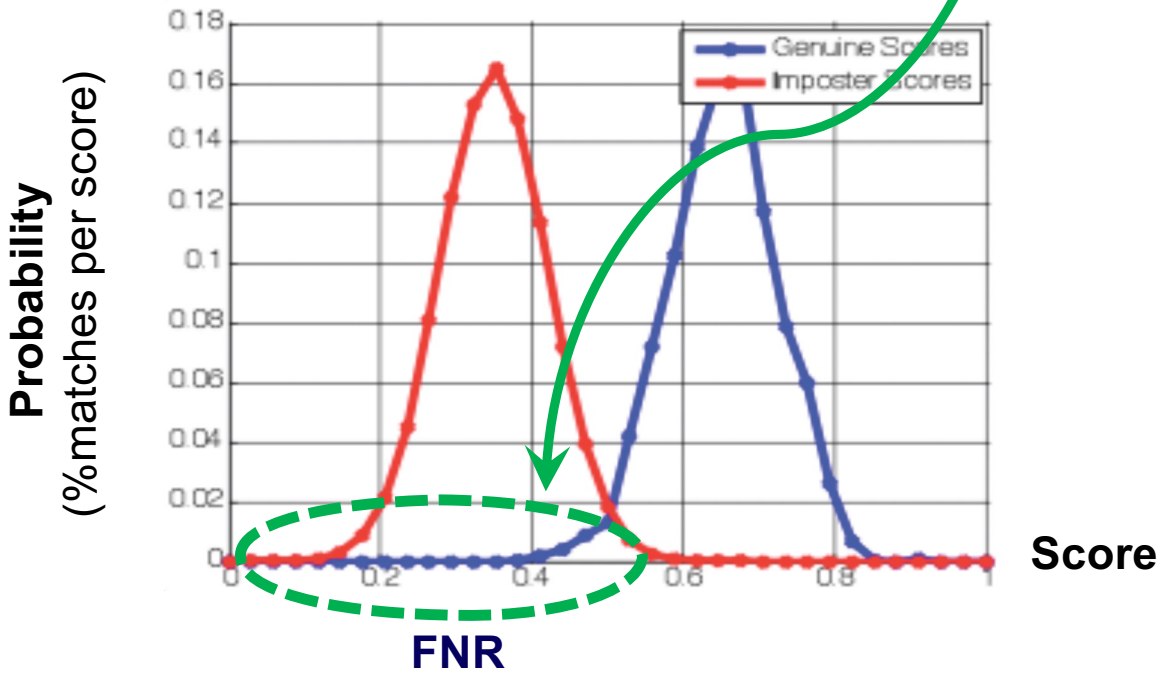


Exclusive vs Inclusive

- ***Inclusive*** applications generally do not require to deal with a very large number of classes.
- By design they need a **quick memory update** to include new subjects.
- Decisions are not as critical, and accurate authentication can be left to another processing stage.
- Uncertainties are mitigated by ***extended interactions***.
- Instead of "***pushing up***" the limit of the True Positive Matching Rate, they minimize the **False Negative Matching Rate**.

Exclusive vs Inclusive

- In more inclusive applications this **segment** is more critical (**inclusive**).
- **TPR** is maximized against **FNR**.



Exclusive vs Inclusive

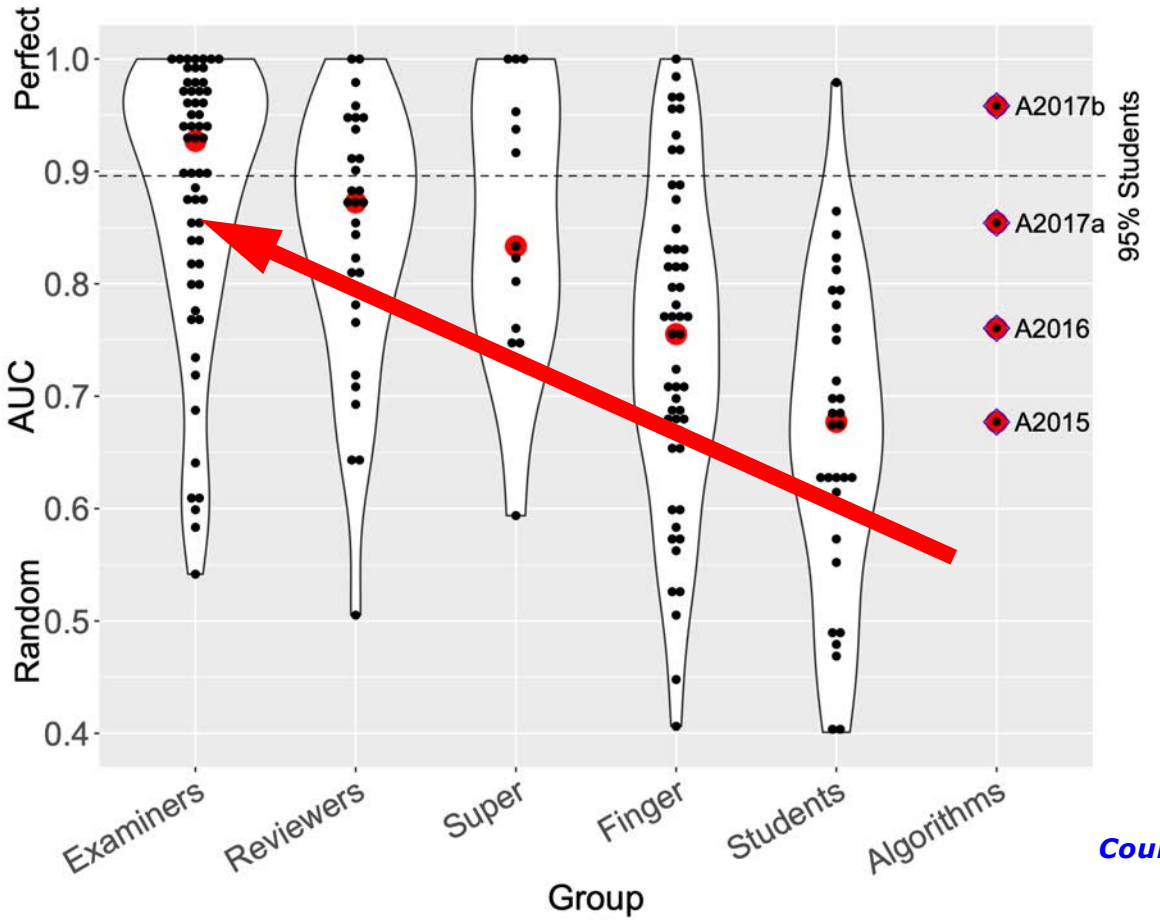
Examples of inclusive applications:

- Nursing robotics
- Automated *meet and greet*
- Customer service (physical or online)
- Virtual presence
- Automated cashier
- Adaptive environments

Question 3

- **Is the Human Visual System still the best comparative face recognition model? If so, what can we learn from the way humans recognize faces?**

The performance arena

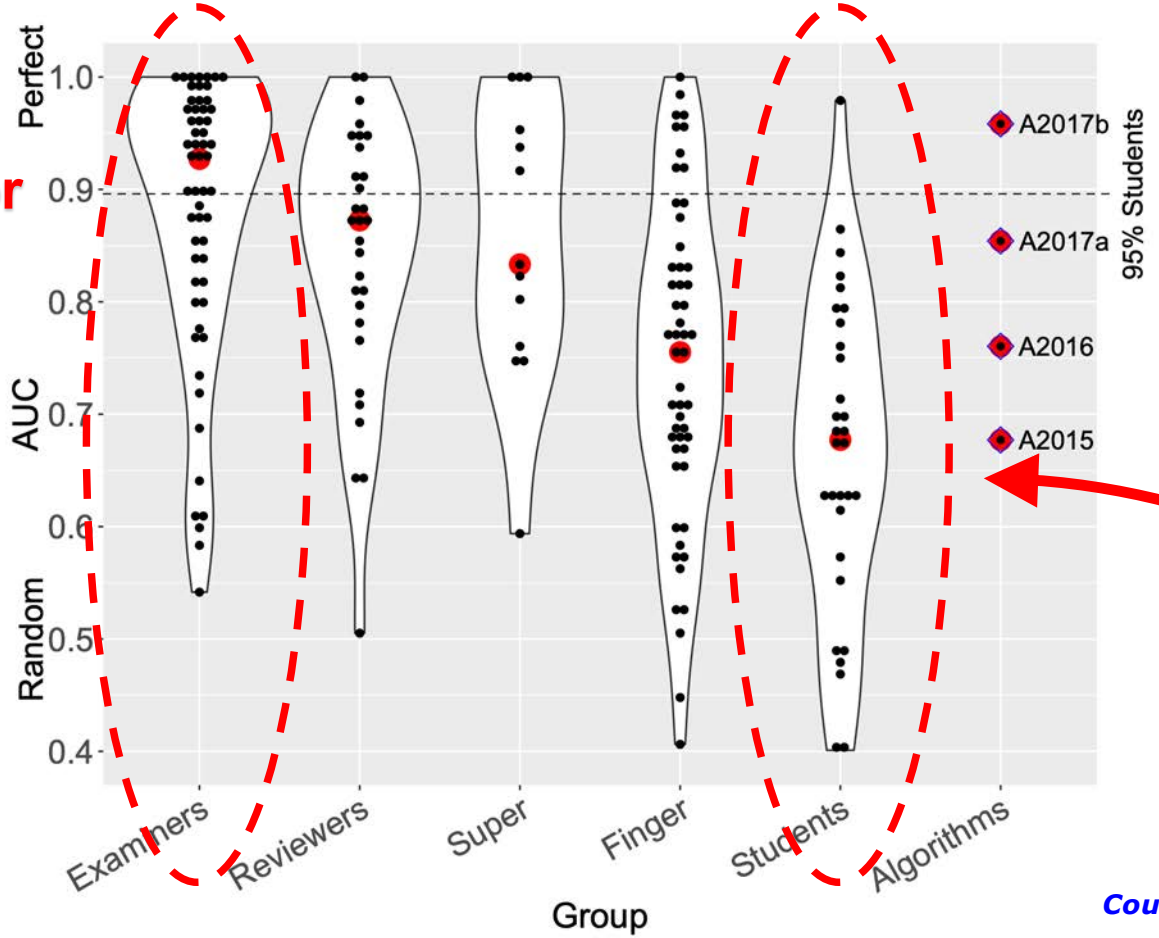


Courtesy of J. Phillips et al. (2022)

Phillips, Yates, Hu, Hahn, Noyes, Jackson, Jeckln, Ranjan, Sankaranarayanan, Chen, Castillo, Chellappa, White, O'Toole,
"**Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms**", Proc. *National Academy of Sciences*, 2018.

Natural vs Artificial Intelligence

Trained behavior



Normal/designed behavior

Courtesy of J. Phillips et al. (2022)

Phillips, Yates, Hu, Hahn, Noyes, Jackson, Jeckln, Ranjan, Sankaranarayanan, Chen, Castillo, Chellappa, White, O'Toole,
"**Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms**", Proc. *National Academy of Sciences*, 2018.

Perception of familiar faces

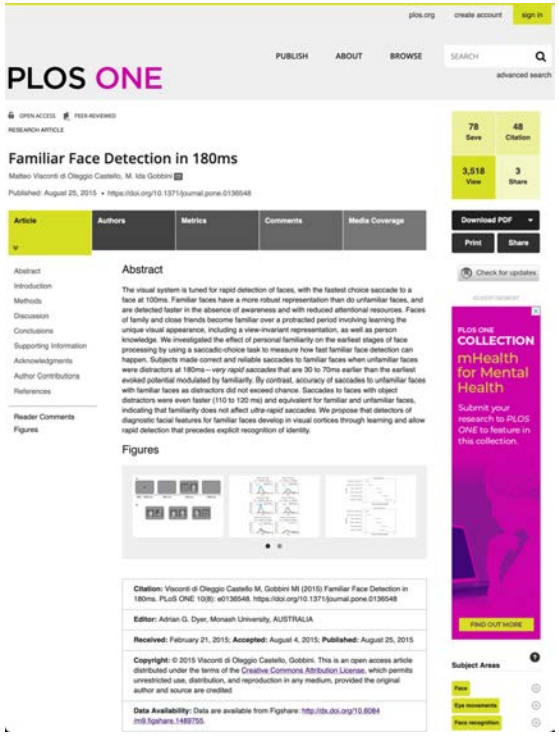
Vision Lab

Perception

Ellis, H.D., Shepherd, J.W., Davies, G.M., 1979. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. Perception 8 (4), 431–439.

Impact Factor: 1.695
5-Year Impact Factor: 1.785

Restricted access | Research article | First published online June 25, 2016


PLOS ONE

Familiar Face Detection in 180ms

Matteo Visconti di Oleggio Castello, M. Ida Gobbini

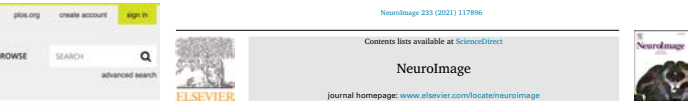
Published: August 25, 2015 • https://doi.org/10.1371/journal.pone.0136548

Abstract

The visual system is tuned for rapid detection of faces, with the fastest choice saccade to a face at 100ms. Familiar faces have a more robust representation than do unfamiliar faces, and are detected faster in the absence of awareness and with reduced attentional resources. Faces of family and close friends become familiar over a protracted period involving learning the unique visual appearance, including a view-invariant representation, as well as person knowledge. We investigated the effect of personal familiarity on the earliest stages of face processing by using a saccadic-choice task to measure how fast familiar face detection can happen. Subjects made correct and reliable saccades to familiar faces when unfamiliar faces were distractors at 180ms—*very rapid saccades* that are 30 to 70ms earlier than the earliest evoked potential modulated by familiarity. By contrast, accuracy of saccades to unfamiliar faces with familiar faces as distractors did not exceed chance. Saccades to faces with object distractors were even faster (110 to 120 ms) and equivalent for familiar and unfamiliar faces, indicating that familiarity does not affect ultra-rapid saccades. We propose that detectors of diagnostic facial features for familiar faces develop in visual cortices through learning and allow rapid detection that precedes explicit recognition of identity.

Figures

PLOS ONE COLLECTION mHealth for Mental Health


NeuroImage 233 (2021) 117896

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Perceptual difficulty modulates the direction of information flow in familiar face recognition

Hamid Karimi-Rouzbahani[a,b,*], Farzad Ramezani[c], Alexandra Woolgar[a,b], Anina Rich[b], Masoud Ghodrati[d,*]

[a] Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, United Kingdom
[b] Perception in Action Research Centre and Department of Cognitive Science Macquarie University, Australia
[c] Department of Computer Science, School of Mathematics, Statistics, and Computer Science, University of Tehran, Iran
[d] Neuroscience Program, Biomedicine Discovery Institute, Monash University, Australia

ARTICLE INFO

Keywords:
Face recognition
Familiar faces
Multivariate pattern analysis (MVPA)
Representational similarity analysis (RSA)
Informational brain connectivity

ABSTRACT

Humans are fast and accurate when they recognize familiar faces. Previous neurophysiological studies have shown enhanced representations for the dichotomy of familiar vs. unfamiliar faces. As familiarity is a spectrum, however, any neural correlate should reflect graded representations for more vs. less familiar faces along the spectrum. By systematically varying familiarity across stimuli, we show a neural familiarity spectrum using electroencephalography. We then evaluated the spatiotemporal dynamics of familiar face recognition across the brain. Specifically, we developed a novel informational connectivity method to test whether peri-frontal brain areas contribute to familiar face recognition. Results showed that feed-forward flow dominates for the most familiar faces and top-down flow was only dominant when sensory evidence was insufficient to support face recognition. These results demonstrate that perceptual difficulty and the level of familiarity influence the neural representation of familiar faces and the degree to which peri-frontal neural networks contribute to familiar face recognition.

Introduction

Faces are crucial for our social interactions, allowing us to extract information about identity, gender, age, familiarity, intent and emotion. Humans categorize familiar faces more quickly and accurately than unfamiliar ones, and this advantage is more pronounced under difficult viewing conditions, where categorizing unfamiliar faces often fails (Ramon and Gobbini, 2018; Young and Burton, 2018). The neural correlates of this behavioral advantage suggest an enhanced representation of familiar over unfamiliar faces in the brain (Dobs et al., 2019; Landi and Freiwald, 2017). Here, we focus on addressing two major questions about familiar face recognition. First, whether there is a "familiarity spectrum" for faces in the brain with enhanced representations for more vs. less familiar faces along the spectrum. Second, whether higher-order frontal brain areas contribute to familiar face recognition, testing previous suggestions about their role in visual recognition (Bar et al., 2006; Goddard et al., 2016; Karimi-Rouzbahani et al., 2019; Polyn et al., 2005; Summerfield et al., 2006; Todorov et al., 2007), and whether levels of face familiarity and perceptual difficulty (as has been suggested previously (Woolgar et al., 2011, 2015)) impact the involvement of frontal cognitive areas in familiar face recognition.

One of the main limitations of previous studies, which hinders our progress in answering our first question, is that they mostly used celebrity faces as the familiar category (Ambrus et al., 2019; Collins et al., 2018; Dobs et al., 2019). As familiar faces can range widely from celebrity faces to highly familiar ones such as family members, relatives, friends, and even one's own face (Ramon and Gobbini, 2018), these results might not reflect the full familiarity spectrum. A better understanding of familiar face recognition requires characterizing the computational steps and representations for sub-categories of familiar faces, including personally familiar, visually familiar, famous, and experimentally learned faces. Such face categories do not only differ in terms of how much exposure the individual has had to them, but also the availability of personal knowledge, relationships, and emotions associated with the identities in question (Leppänen and Nelson, 2009; Ramon and Gobbini, 2018; Kovács, 2020). However, we still expect that potentially enhanced representations for more vs. less familiar faces, as they modulate the behavior, can also be detected using neuroimaging analysis. Moreover, these categories may vary in terms of the potential for top-down influences in the process. Importantly, while a few functional magnetic resonance imaging (fMRI) studies have investigated the differences between different levels of familiar faces (Gobbini et al., 2004; Landi and Freiwald, 2017; Leibenluft et al., 2004; Ramon et al., 2015; Sugiura et al., 2015; Taylor et al., 2009), there are no studies

* Corresponding authors.
E-mail addresses: hamid.karimi-rouzbahani@mrc-cbu.cam.ac.uk (H. Karimi-Rouzbahani), ghodrati.masoud@gmail.com (M. Ghodrati).

https://doi.org/10.1016/j.neuroimage.2021.117896
Received 22 October 2020; Received in revised form 10 February 2021; Accepted 17 February 2021
Available online 3 March 2021
1053-8119/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)


ELSEVIER

Cognition

Volume 172, March 2018, Pages 46-58

COGNITION

Original Articles

Understanding face familiarity

Robin S.S. Kramer [a, b], Andrew W. Young [a], A. Mike Burton [a]

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Registered report

How robust is familiar face recognition? A repeat detection study of more than 1000 faces

Angus F. Chapman[1,2,†], Hannah Hawkins-Elder[1,†] and Tirta Susilo[1,3]

[1]School of Psychology, Victoria University of Wellington, Wellington 6040, New Zealand
[2]Department of Psychology, University of California San Diego, La Jolla, CA, USA
[3]ARC Centre of Excellence in Cognition and its Disorders, Sydney, New South Wales, Australia

AFC, 0000-0002-0354-8536

Cite this article: Chapman AF, Hawkins-Elder H, Susilo T. 2018 How robust is familiar face recognition? A repeat detection study of more than 1000 faces. *R. Soc. open sci.* **5**: 170634. http://dx.doi.org/10.1098/rsos.170634

Received: 8 June 2017
Accepted: 24 April 2018

Subject Category:
Psychology and cognitive neuroscience

Subject Areas:
psychology/cognition/behaviour

Keywords:
face recognition, face memory, familiarity, repeat detection

Author for correspondence:
Tirta Susilo
e-mail: tirta.susilo@vuw.ac.nz

†Co-first authors.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4106426.

THE ROYAL SOCIETY PUBLISHING

Recent theories suggest that familiar faces have a robust representation in memory because they have been encountered over a wide variety of contexts and image changes (e.g. lighting, viewpoint and expression). By contrast, unfamiliar faces are encountered only once, and so they do not benefit from such richness of experience and are represented based on image-specific details. In this registered report, we used a repeat detection task to test whether familiar faces are recognized better than unfamiliar faces across image changes. Participants viewed a stream of more than 1000 celebrity face images for 0.5 s each, any of which might be repeated at a later point and has to be detected. Some participants saw the same image at repeats, while others saw a different image of the same face. A post-experimental familiarity check allowed us to determine which celebrities were and were not familiar to each participant. We had three predictions: (i) detection would be better for familiar than unfamiliar faces, (ii) detection would be better across same rather than different images, and (iii) detection of familiar faces would be comparable across same and different images, but detection of unfamiliar faces would be poorer across different images. We obtained support for the first two predictions but not the last. Instead, we found that repeat detection of faces, regardless of familiarity, was poorer across different images. Our study suggests that the robustness of familiar face recognition may have limits, and that under some conditions, familiar face recognition can be just as influenced by image changes as unfamiliar face recognition.

ELSEVIER

Neuropsychologia

Volume 61, August 2014, Pages 65-79

Reviews and perspectives

Beyond the FFA: The role of the ventral anterior temporal lobes in face processing

Jessica A. Collins, Ingrid R. Olson

Show more

Share Cite

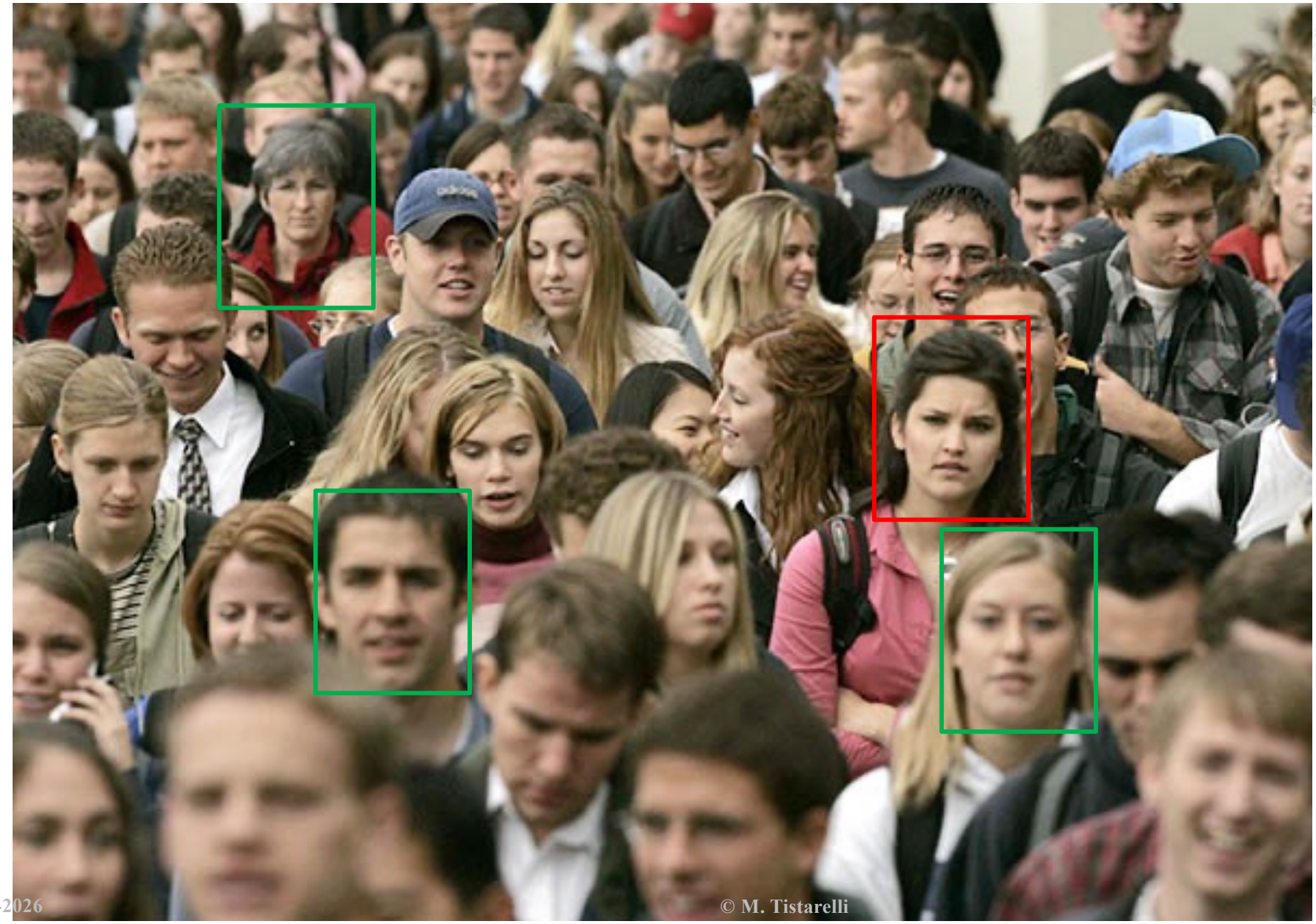
https://doi.org/10.1016/j.neuropsychologia.2014.06.005 Get rights and content

Abstract

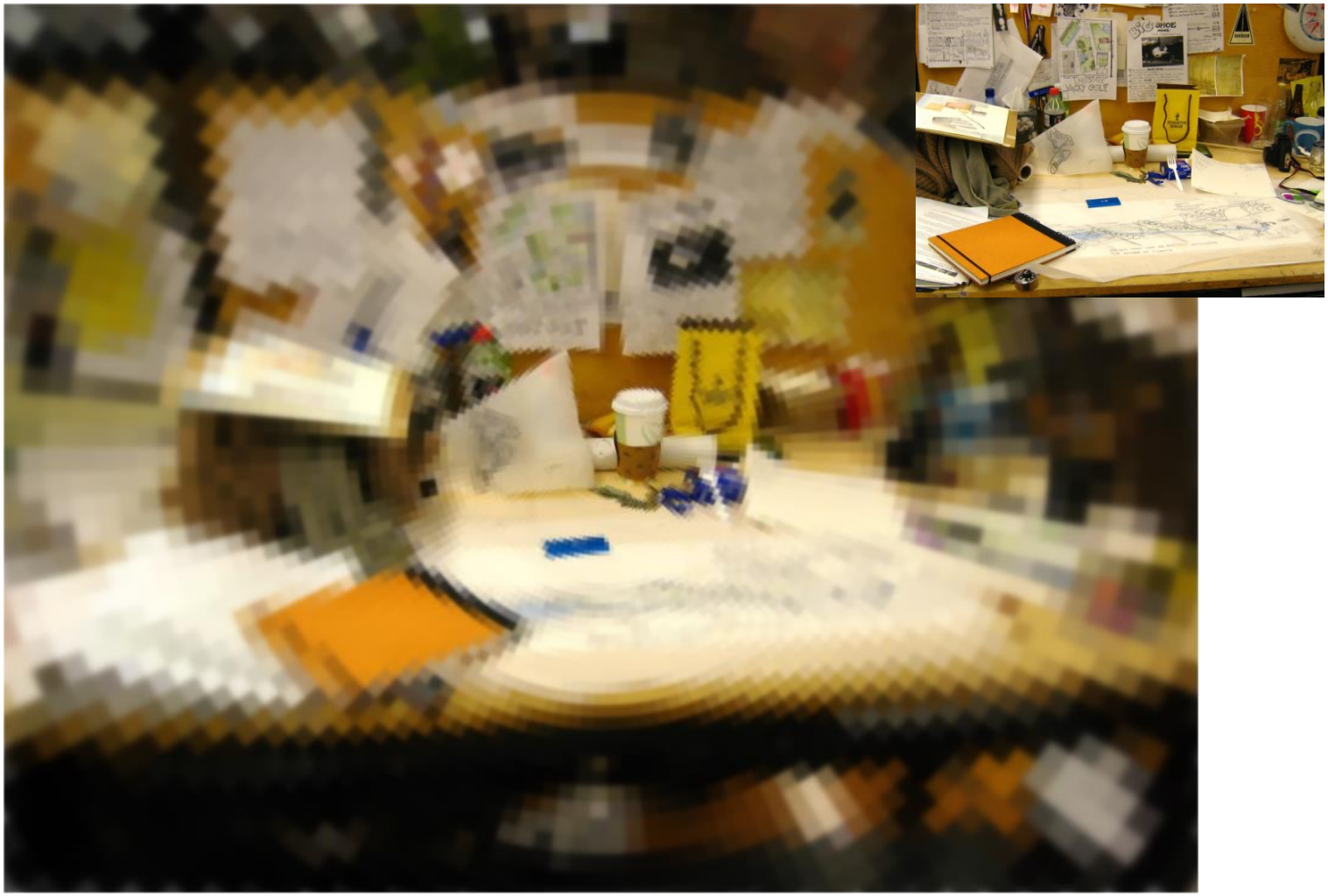
Extensive research has supported the existence of a specialized face-processing network that is distinct from the visual processing areas used for general object recognition. The majority of this work has been aimed at characterizing the response properties of the fusiform face area (FFA) and the occipital face area (OFA), which together are thought to constitute the core network of brain areas responsible for facial identification. Although accruing evidence has shown that face-selective patches in the ventral anterior temporal lobes (vATLs) are interconnected with the FFA and OFA, and that they play a role in facial identification, the relative contribution of these brain areas to the core face-processing network has remained unarticulated. Here we review recent research critically implicating the vATLs in face perception and memory. We propose that current models of face processing should be revised such that the ventral anterior temporal lobes serve a centralized role in the visual face-processing network. We speculate that a hierarchically organized system of face processing areas extends bilaterally from the inferior occipital gyri to the vATLs, with facial representations becoming increasingly complex and abstracted from low-level perceptual features as they move forward along this network. The anterior temporal face areas may serve as the apex of this hierarchy, instantiating the final stages of face recognition. We further argue that the anterior temporal face areas are ideally suited to serve as an interface between face perception and face memory, linking perceptual representations of individual identity with person-specific semantic knowledge.

Just look once…

Just look once…

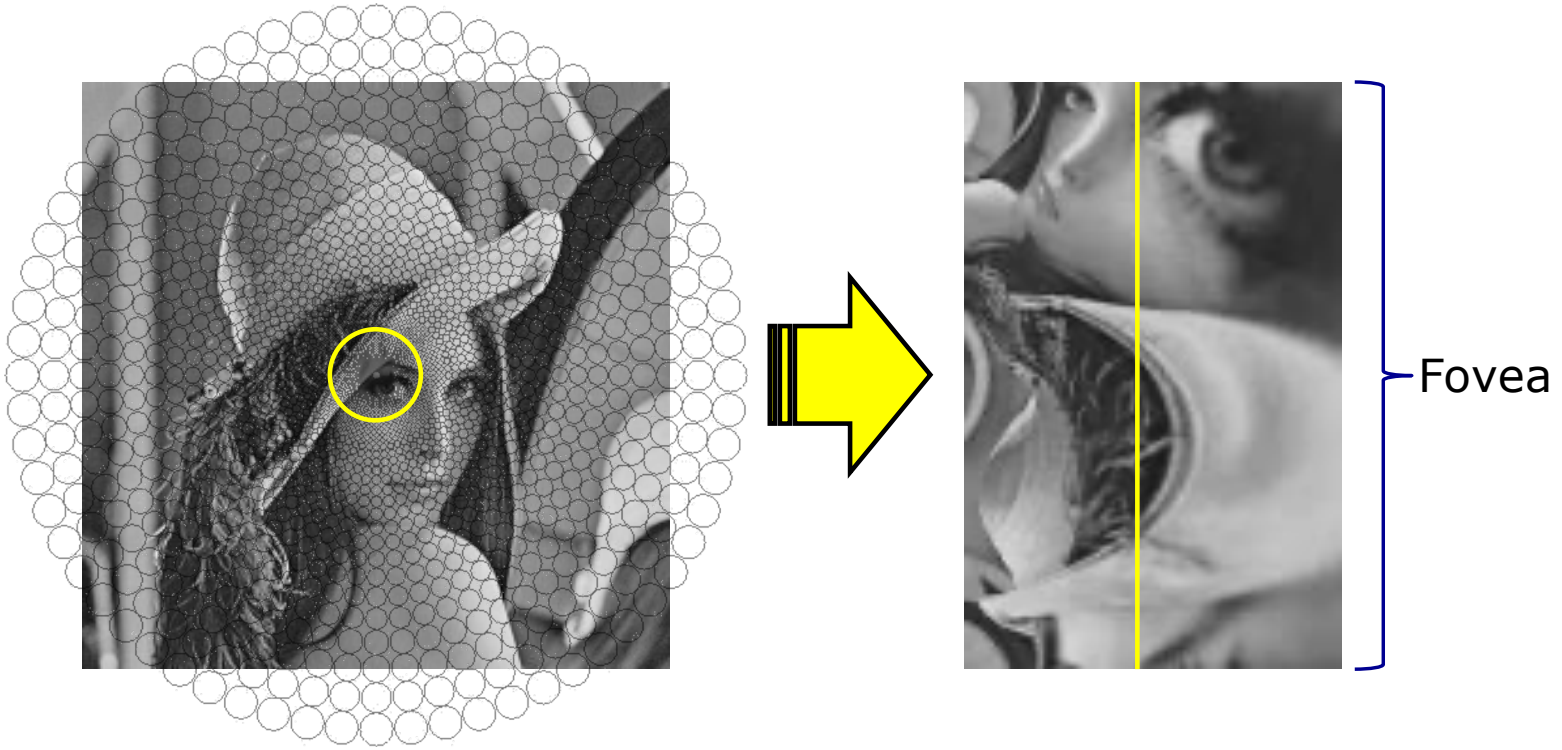


Visual attention



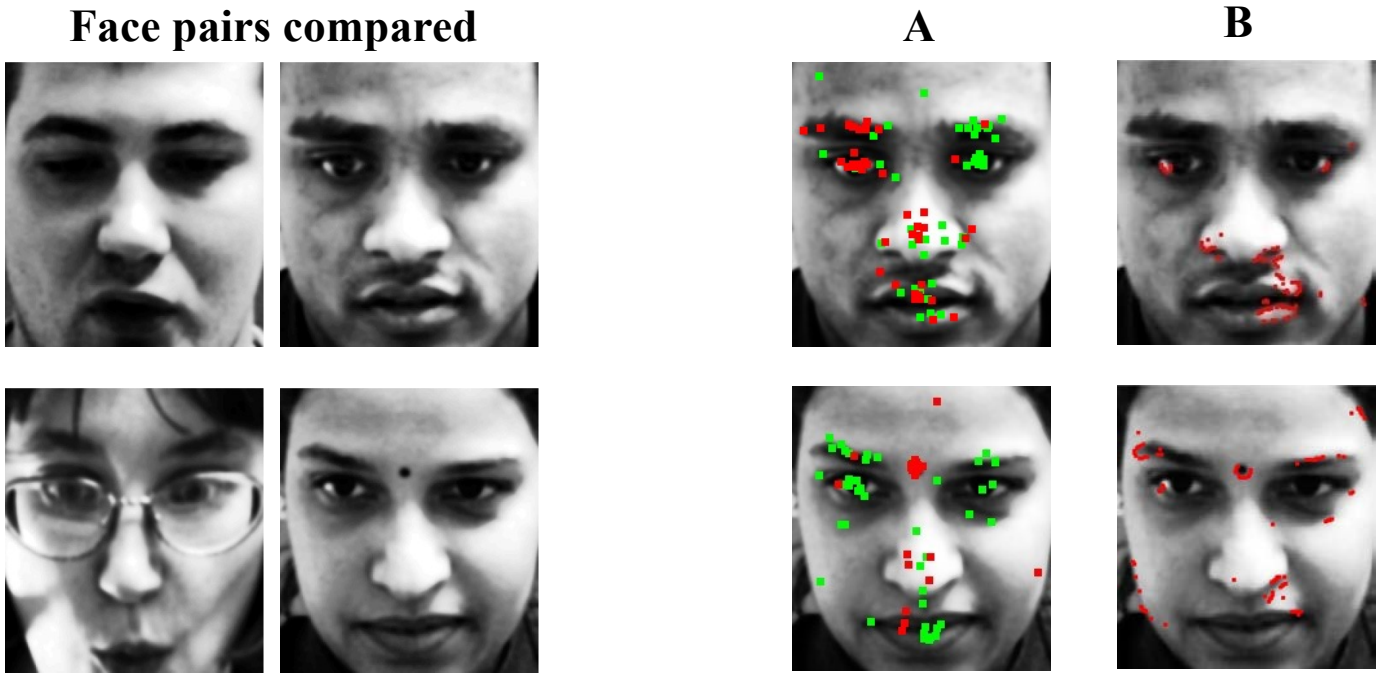
Log-Polar/Retino-Cortical mapping

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 in face comparison



(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.

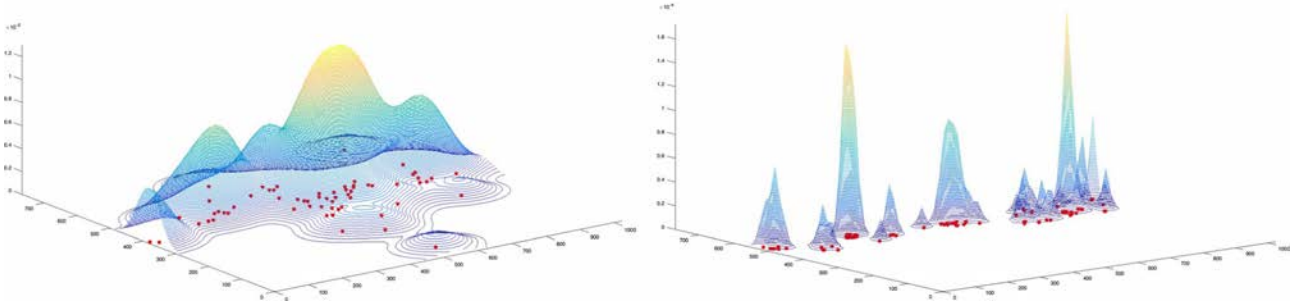
CNNs and visual attention

M. Cadoni, A. Lagorio, S. Khellat-Kihel, E. Grosso "**On the correlation between human fixations, handcrafted and CNN features**", Neural Computing and Applications, 2021.

CNNs and visual attention

Fixation points

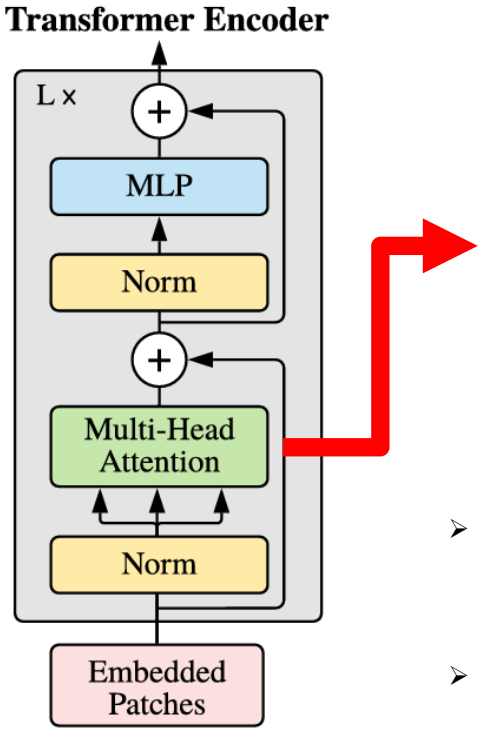
AlexNet interest points.

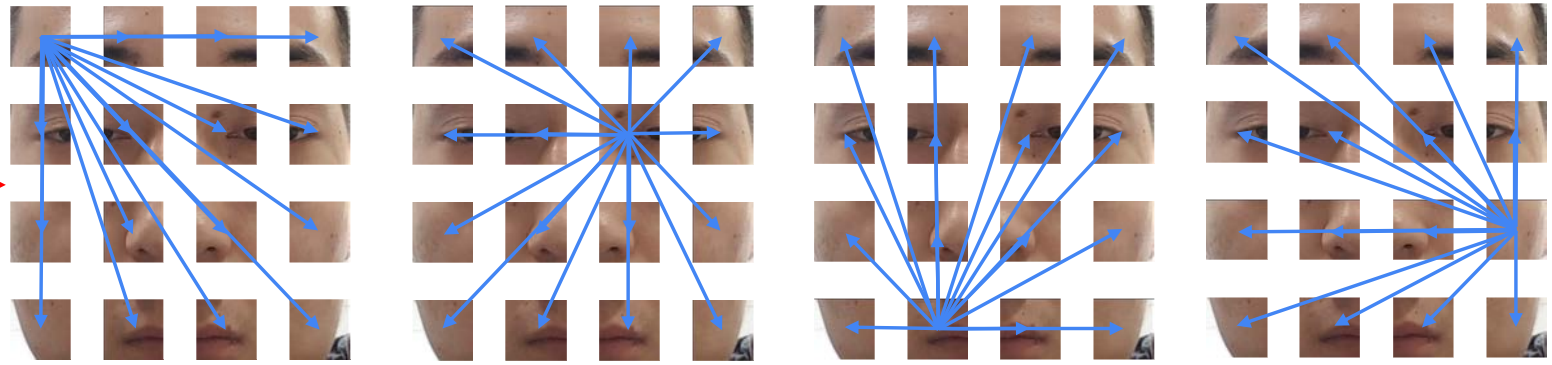


Interest regions are modeled via **K**erne**l** **D**ensity **E**stimation.

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.

Transformers and visual attention

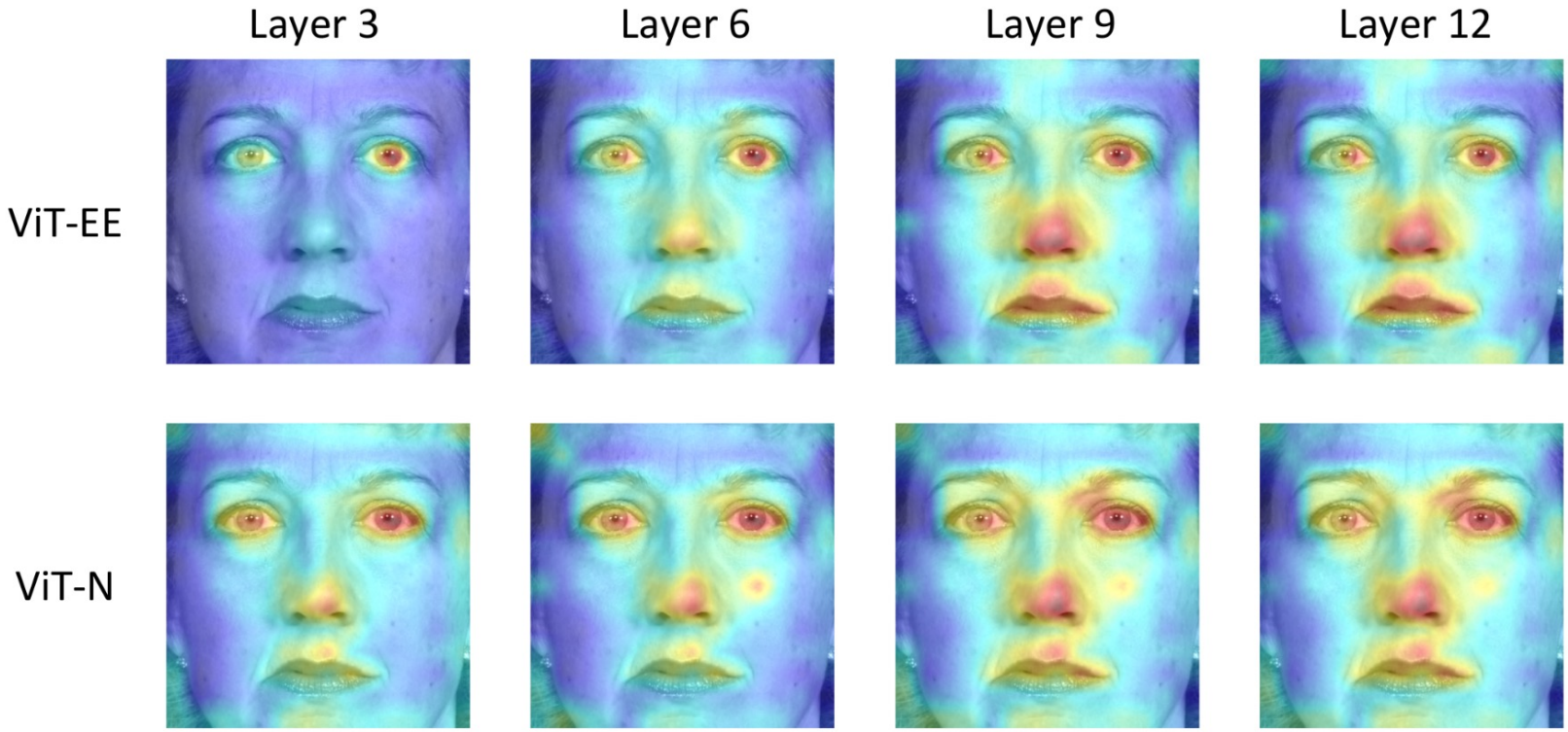




- **Multi Head Self Attention** is computed as the normalized dot product between input patch embeddings.
- Data- and task-dependant up-weighting of the most informative image regions
- Through repeated, parallel comparison and feed-forward networks a robust representation is obtained.

S. Abnar and W. Zuidema. **Quantifying attention flow in transformers**. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 4190–4197, 2020.

ViT attention



M. Cadoni, S. Nixon, A. Lagorio and M. Fadda, **Exploring attention on faces: similarities between humans and Transformers**, 2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Madrid, Spain, 2022, pp. 1-8.

S. Nixon, P. Ruiu, M. Cadoni, A. Lagorio and M. Tistarelli, **Exploiting Face Recognizability with Early Exit Vision Transformers**, 2023 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2023, pp. 1-7

Driving Attention in ViTs

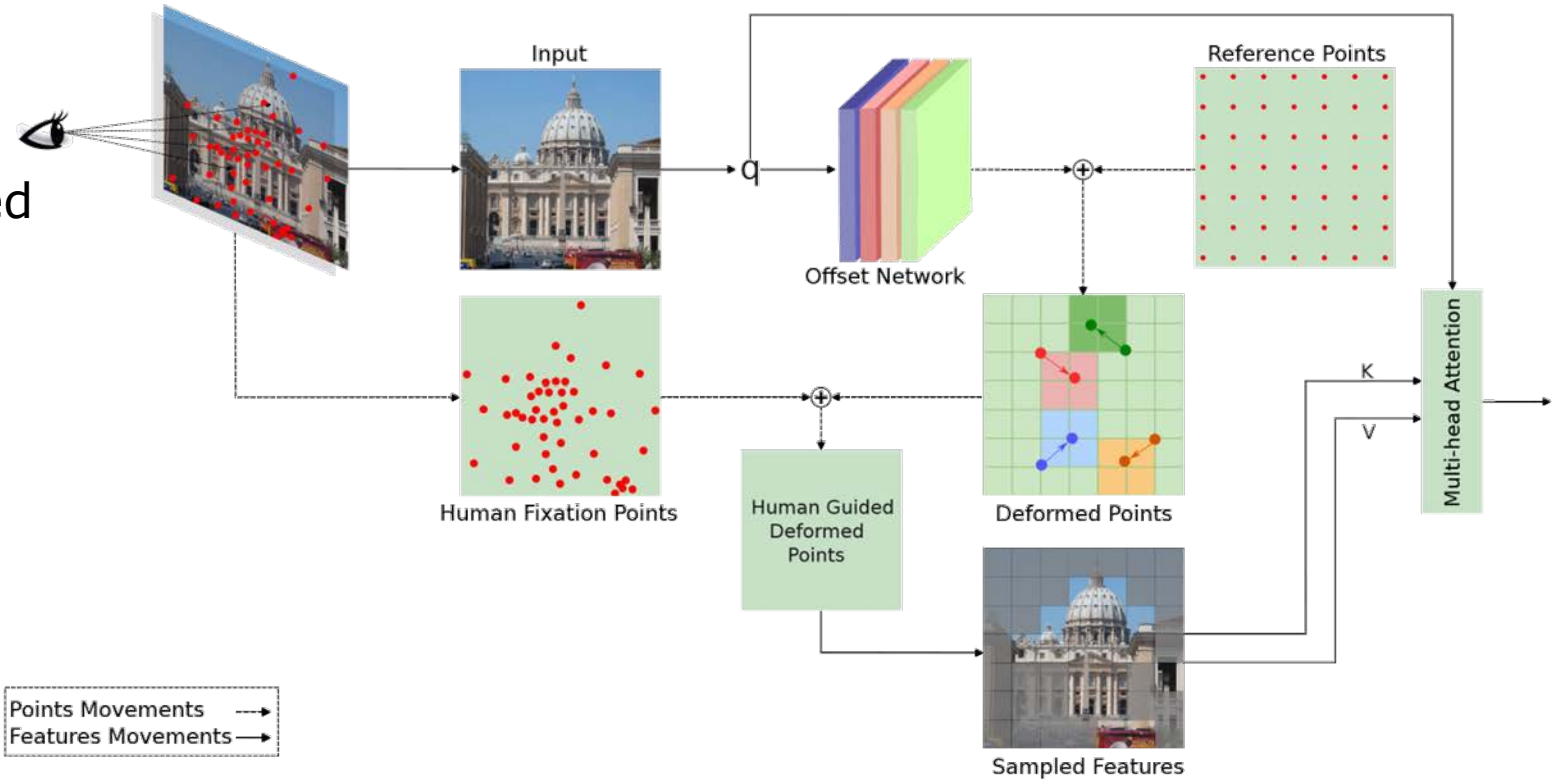
- Vision Transformers often attend to irrelevant image regions.
- Human vision selects salient regions through eye fixations.
- **Approach**: injecting human eye fixation points directly into transformer self-attention.
- **Goal**: align model attention with human attention and improve classification accuracy.

A. Waseem, P. Ruiu, S. Nixon, A. Lagorio and M. Tistarelli, **Attention Mechanisms in Vision Transformers: A Survey**, Int.l J. of Computer Vision, 2026

A. Waseem, P. Ruiu, S. Nixon, A. Lagorio and M. Tistarelli, **HuGDAT: Guiding transformer attention with human vision**, ICPR 2026.

Driving Attention in ViTs

- Fixation points are injected into deformable attention sampling.
- Fixations are combined with deformable attention offsets.
- Sampling positions = reference points + learned offsets + fixations.
- The backbone architecture remains unchanged.

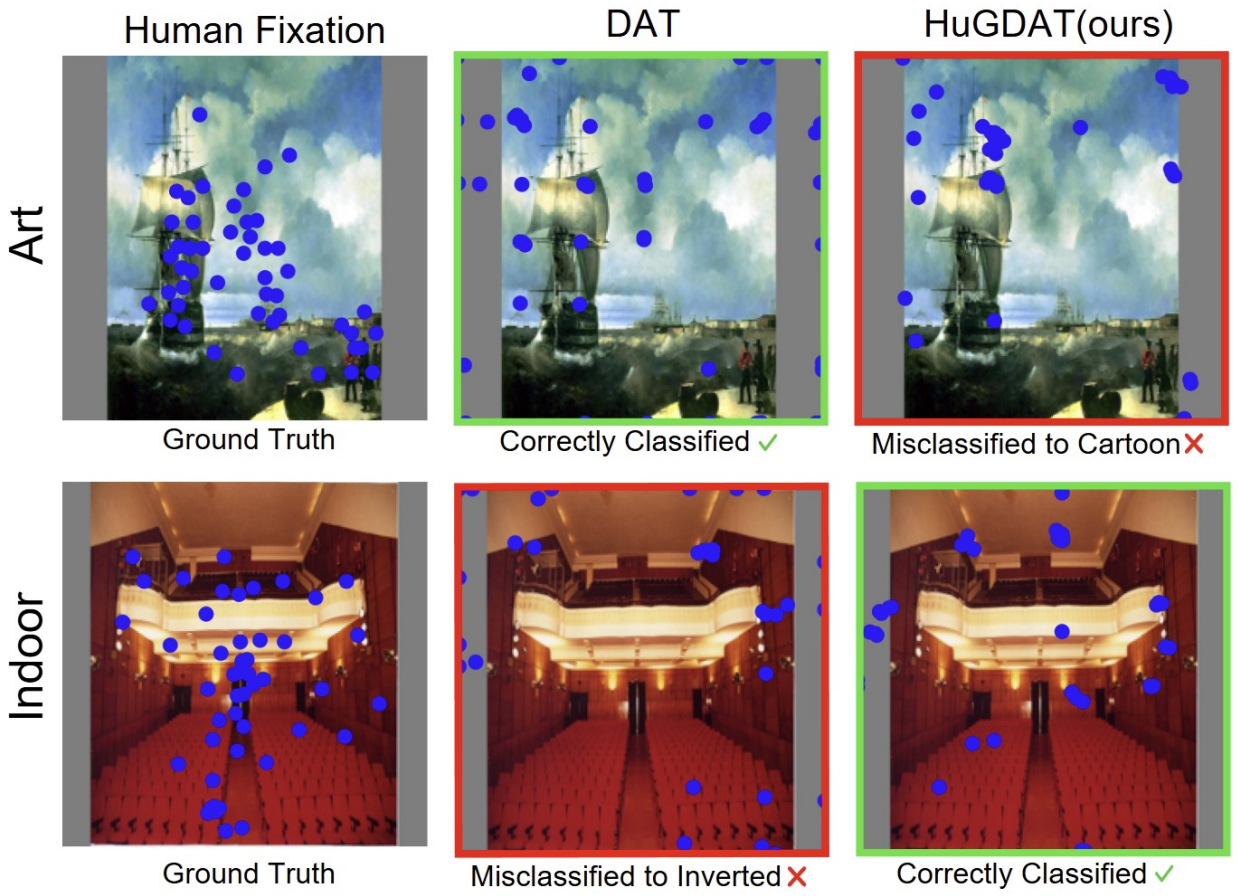


A. Waseem, P. Ruiu, S. Nixon, A. Lagorio and M. Tistarelli, **Attention Mechanisms in Vision Transformers: A Survey**, Int.l J. of Computer Vision, 2026

A. Waseem, P. Ruiu, S. Nixon, A. Lagorio and M. Tistarelli, **HuGDAT: Guiding transformer attention with human vision**, ICPR 2026.

Driving Attention in ViTs

- Largest gains appear in **object-centric categories** (faces, objects, indoor scenes benefit most).
- Human fixations are effective **attention priors**.
- Hybrid human-guided + learned attention performs best.
- Improved interpretability and stability.



A. Waseem, P. Ruiu, S. Nixon, A. Lagorio and M. Tistarelli, **Attention Mechanisms in Vision Transformers: A Survey**, Int.l J. of Computer Vision, 2026

A. Waseem, P. Ruiu, S. Nixon, A. Lagorio and M. Tistarelli, **HuGDAT: Guiding transformer attention with human vision**, ICPR 2026.

Question 4

- **How can we build "ethical" systems which propery address current concerns?**

Question 4

- **How can we build "ethical" systems which propery address current concerns?**

1. **Only recently arised, because of the data greedy behavior of CNNs, and widesprad applications**

Question 4

- **How can we build "ethical" systems which propery address current privacy concerns?**

1. Only recently arised, because of the data greedy behavior of CNNs, and widesprad applications

2. Another dimension of the same ill-posed problem

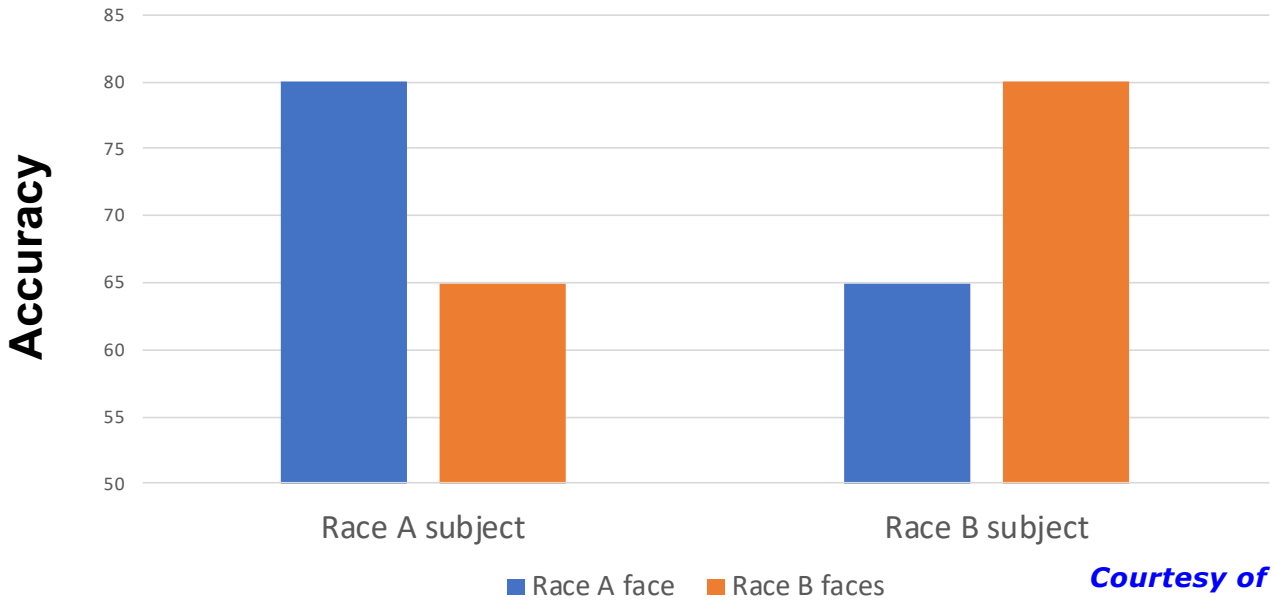
- **Feature-disentagle-based methods** learn unbiased face representations. Still race, age and gender convey useful information for recognition. [Y. Li, K. Swersky, R. Zemel "Learning unbiased features", 2014]
- **Adaptation-based methods** learn unbiased representation spaces where all faces are equally represented. [R. Ragonesi, R. Volpi, J. Cavazza, V. Murino, "Learning Unbiased Representations via Mutual Information Backpropagation", 2021]
- **Re-weighting-based methods** learn an adaptive weight or margin to be adopted for each specific group. [X. Xu, Y. Huang, P. Shen, S. Li, J. Li, F. Huang, Y. Li, Z. Cui, "Consistent Instance False Positive Improves Fairness in Face Recognition", 2021]

Question 4

- **How can we build "ethical" systems which propery address current privacy concerns?**

1. Only recently arised, because of the data greedy behavior of CNNs, and widesprad applications
2. Another dimension of the same ill-posed problem
3. **The bias mainly depends on our ecosystem**

The *other race* effect



Courtesy of J. Phillips et al. (2022)

K. S. Krishnapriya, V. Albiero, K. Vangara, M. C. King and K. W. Bowyer, "Issues Related to Face Recognition Accuracy Varying Based on Race and Skin Tone," in **IEEE Transactions on Technology and Society**, vol. 1, no. 1, pp. 8-20, March 2020.

Cavazos JG, Phillips PJ, Castillo CD, O'Toole AJ. "Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?," **IEEE Transactions on Biometrics, Behavior, and Identity Science** 3(1):101-111, 2021.

C.A. Meissenr and J.C. Brigham, "Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review," **Psychology Public Policy, and Law**, vol. 7, no. 1, pp. 3-35, 2001.

R.S. Malpass and J. Kravits, "Recognition for faces of own and other race," **J. Pers. Soc. Psychol**., vol. 13, no. 4, pp. 330-334, 1969.

Question 4

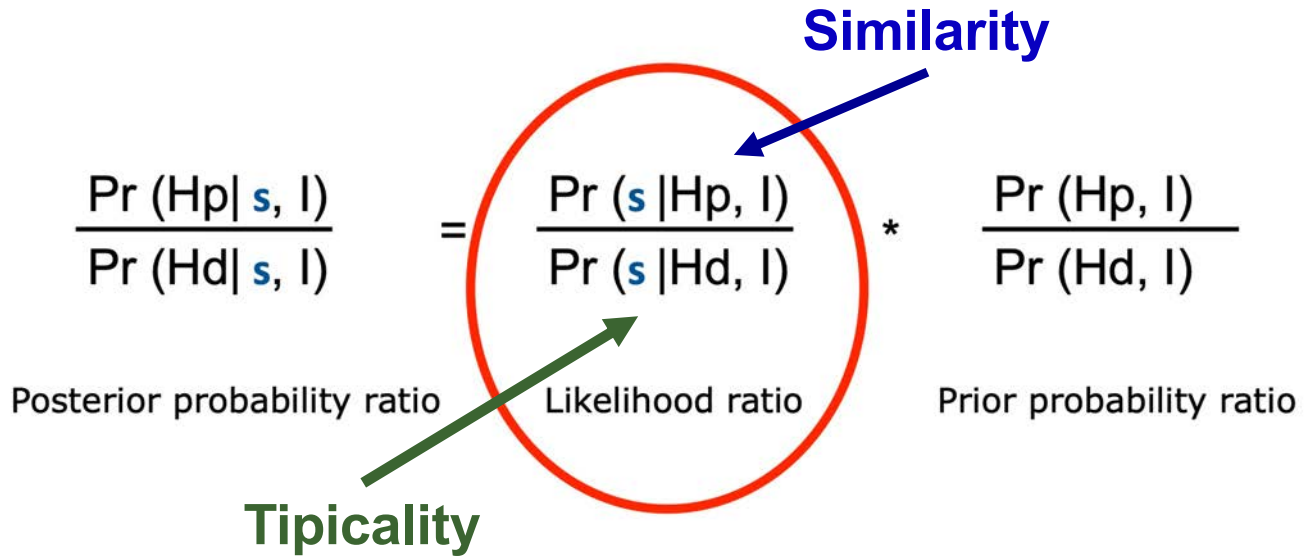
- **How can we build "ethical" systems which propery address current privacy concerns?**

1. Only recently arised, because of the data greedy behavior of CNNs, and widesprad applications
2. Another dimension of the same ill-posed problem
3. The bias mainly depends on our ecosystem
4. **Concept of universality… and the *likelihood ratio***

Question 4

- **How can we build "ethical" systems which propery address current concerns?**

• Assign the evidential value (Likelihood Ratio)



Question 4

- **How can we build "ethical" systems which propery address current privacy concerns?**

1. Only recently arised, because of the data greedy behavior of CNNs, and widesprad applications
2. Another dimension of the same ill-posed problem
3. The bias mainly depends on our ecosystem
4. Concept of universality … and the *likelihood ratio*
5. **How many bias sources should be considered?**

Question 5

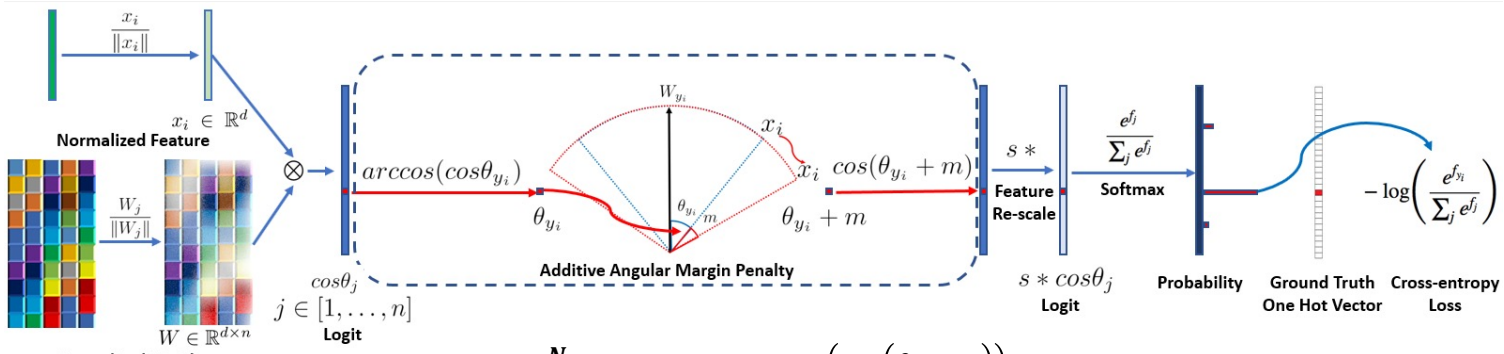
- **Will face recognition advance AI?**

Question 5

- **Will face recognition advance AI?**

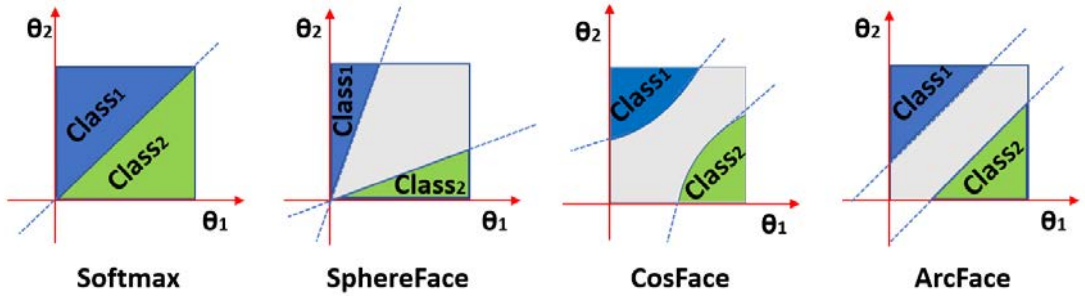
…Maybe

Question 5



NOT THIS WAY!

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

Natural vs Artificial Intelligence

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?

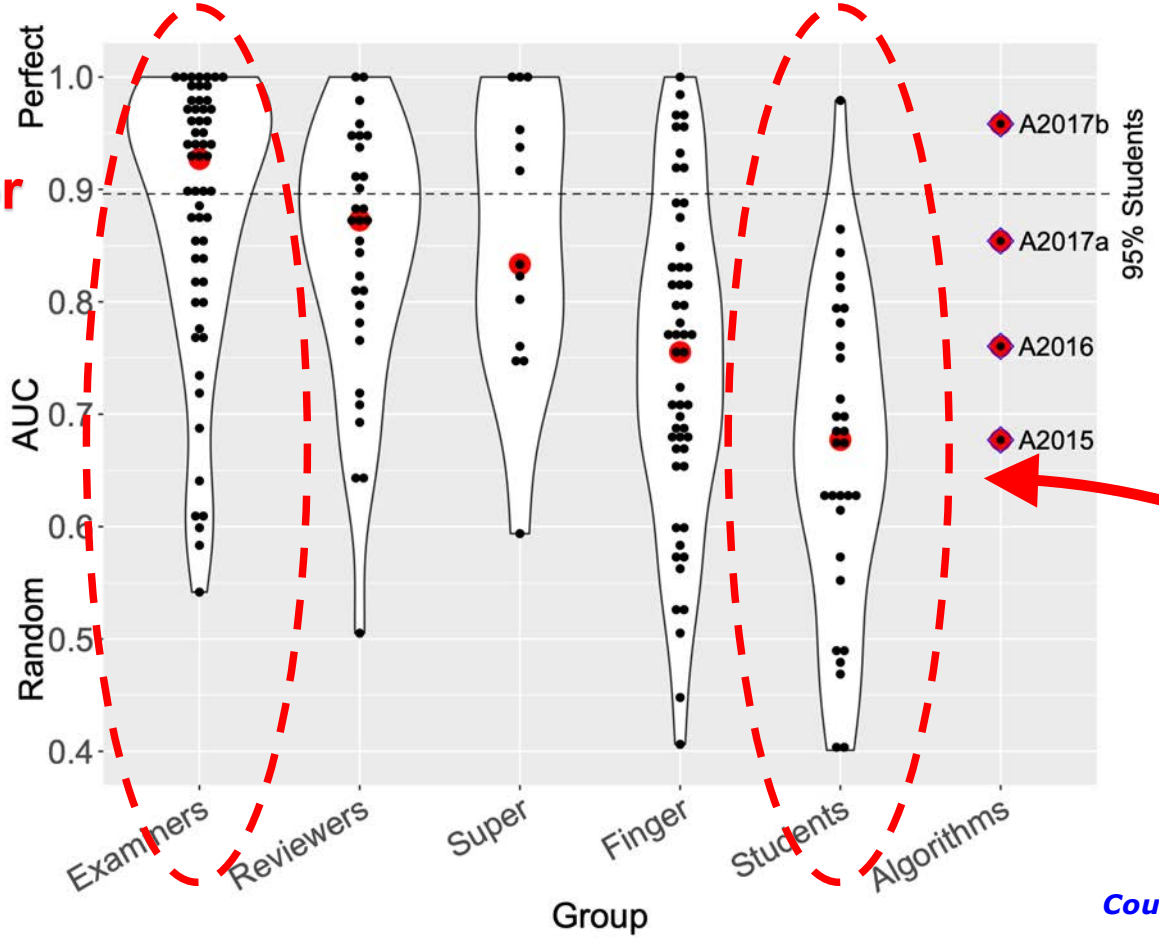
Alan L. Yuille[1] · **Chenxi Liu**[1]

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.

Natural vs Artificial Intelligence

Trained behavior

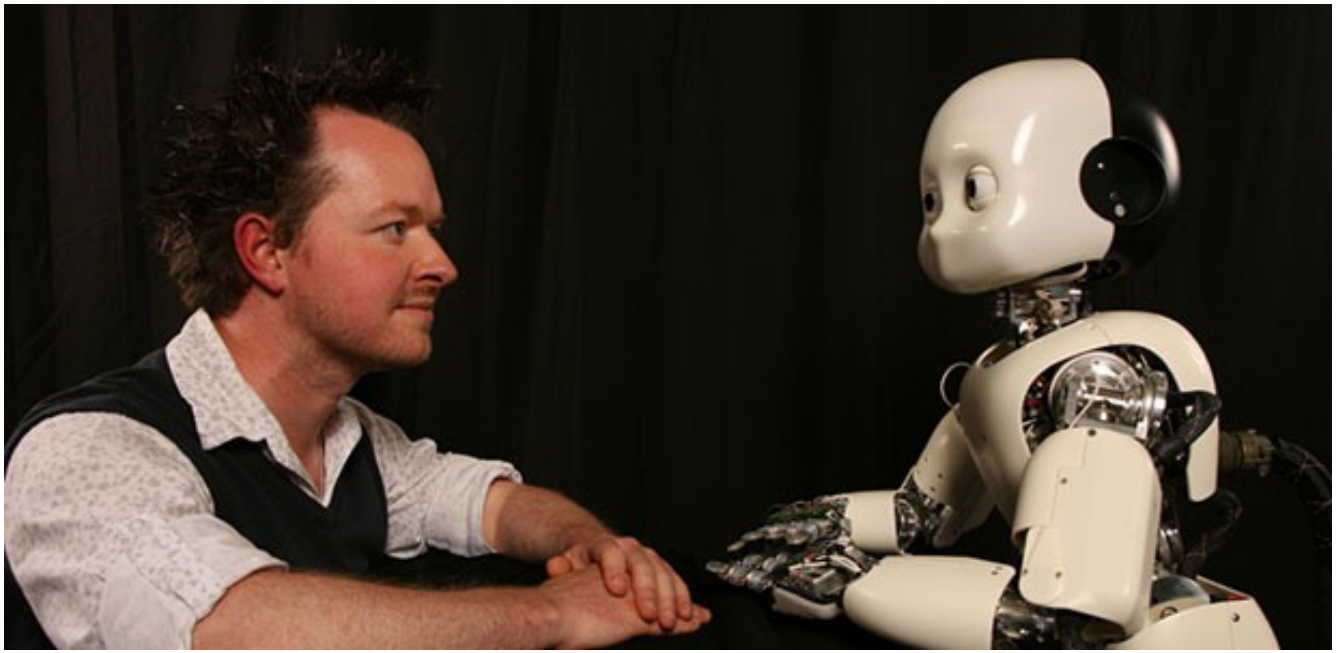


Normal/designed behavior

Courtesy of J. Phillips et al. (2022)

Phillips, Yates, Hu, Hahn, Noyes, Jackson, Jeckln, Ranjan, Sankaranarayanan, Chen, Castillo, Chellappa, White, O'Toole,
"**Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms**", Proc. *National Academy of Sciences*, 2018.

Natural vs Artificial Intelligence



- **PLASTICITY**

to **alter** over time, by creating new connections.

- **FLEXIBILITY**

to **adapt** to new, changing, or unplanned events.

- **ADAPTABILITY**

to **change** in response to external stimuli enhancing his own skills.

Natural vs Artificial Intelligence

- **Learn from human perceptual behaviors**
  - Exploit attention mechanisms; make models more *curious*
- **Change the learning paradigm**
  - Perform interactions; incremental and continuous learning
- **Add event-driven feedback to the system**

Understanding how faces **become** familiar appears to rely on ***both*** **bottom-up** statistical image descriptions, and **top-down** processes that cohere superficially different images of the same person...

Robin S.S. Kramer, Andrew W. Young, A. Mike Burton, "**Understanding face familiarity**", Cognition, Vol. 172, pp. 46-58, 2018, ISSN 0010-0277, https://doi.org/10.1016/j.cognition.2017.12.005.

Conclusion

Q1 If face recognition is so **simple**, why do we still need to pursue research on this topic?

Q2 What are the **drawbacks** and **limitations** of current deep learning models? How far can we go by exploiting increasing amounts of **face data**?

Q3 Is the **Human Visual System** still the best comparative face recognition model? If so, what can we learn from the way **humans** recognize faces?

Q4 How can we build "**ethical**" systems which propery address current privacy concerns?

Q5 Will face recognition **advance AI**?

Conclusions

Vision Lab

Any other questions?

THANK YOU FOR YOUR

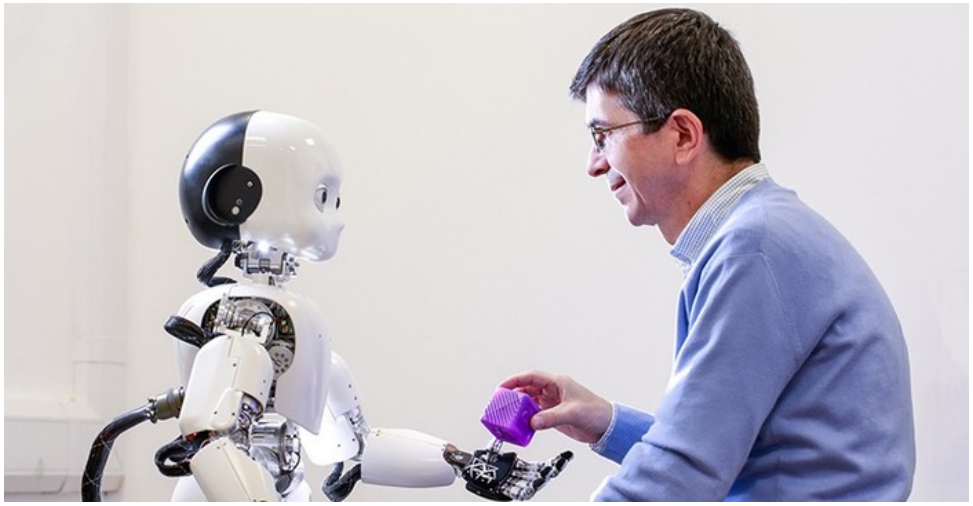
23rd Int.l Summer School for Advanced Studies on Biometrics, Behavior and Vision:

Human Interactions and Large Foundation Models

Alghero, Italy - June, 8-12 2026

https://biometrics.uniss.it

Contact: tista@uniss.it



- H. Wang, S. Wang, Z. Jin, Y. Wang, C. Chen, and M. Tistarelli, **Similarity-based gray-box adversarial attack against deep face recognition**, in IEEE International Conference on Automatic Face and Gesture Recognition 2021 (FG2021), 2021.
- Wang, H., Dong, X., Jin, Z., Teoh, A. B. J., & Tistarelli, M. (2021). **Interpretable Security Analysis of Cancellable Biometrics Using Constrained-Optimized Similarity-Based Attack**. In *Proceedings of the IEEE/CVF Winter Conf. on Appl.s of Computer Vision (WACV) Workshops,* pp. 70-77.
- Yen-Lung Lai, Xingbo Dong, Zhe Jin, Massimo Tistarelli, Wun-She Yap, Bok-Min Goi: **Breaking Free From Entropy's Shackles: Cosine Distance-Sensitive Error Correction for Reliable Biometric Cryptography**. IEEE Trans. Inf. Forensics Secur. 18: 3101-3115 (2023)
- Yen-Lung Lai, Zhe Jin, KokSheik Wong, Massimo Tistarelli: **Efficient Known-Sample Attack for Distance-Preserving Hashing Biometric Template Protection Schemes**. IEEE Trans. Inf. Forensics Secur. 16: 3170-3185 (2021).
- M. Ortega, L. Brodo, M. Bicego, M. Tistarelli (2009) **Measuring changes in face appearance through aging**, in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR 2009, pp.107-113, 2009.
- S. Nixon, P. Ruiu, M. Cadoni, A. Lagorio and M. Tistarelli, **Exploiting Face Recognizability with Early Exit Vision Transformers**, 2023 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2023, pp. 1-7.
- 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.
- M. Cadoni, A. Lagorio, S. Khellat-Kihel, E. Grosso (2021) **On the correlation between human fixations, handcrafted and CNN feature**s, Neural Computing and Applications.
- 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.
- M. Cadoni, S. Nixon, A. Lagorio and M. Fadda, **Exploring attention on faces: similarities between humans and Transformers**, 2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Madrid, Spain, 2022, pp. 1-8.
- Cadoni, M., Lagorio, A. & Grosso, E. **Face detection based on a human attention guided multi-scale model**. Biol Cybern 117, 453–466 (2023).
- Waseem, A., Ruiu, P., Lagorio, A., Nixon, S., & Tistarelli, M., **HuGDAT: Guiding transformer attention with human vision**, Under review ICPR 2026.
- S. Nixon, P. Ruiu, M. Cadoni, A. Lagorio and M. Tistarelli, **Assessing bias and computational efficiency in vision transformers using early exits**, EURASIP Journal on Image and Video Processing, 2025(1), 2. https://doi.org/10.1186/s13640-024-00658-9.