11TH INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION SYSTEMS

17-19 March 2021, Universidad de Talca, Curicó - Chile
(Virtually via Whova and Zoom)

Keynote Lecture

Securing our Identity: from Biometric Anti-Spoofing to DeepFakes Detection

Prof. Julian FIERREZ
http://biometrics.eps.uam.es

Universidad Autónoma de Madrid

With contributions from: Javier GALBALLY, Ruben TOLOSANA, Sergio ROMERO-TAPIADOR, and Ruben VERA-RODRIGUEZ

1

Attacks to Biometric Systems: Introduction

A. Hadid, et al., "Biometrics systems under spoofing attack: an evaluation methodology and lessons learned", *IEEE Signal Processing Magazine*, Sept. 2015.

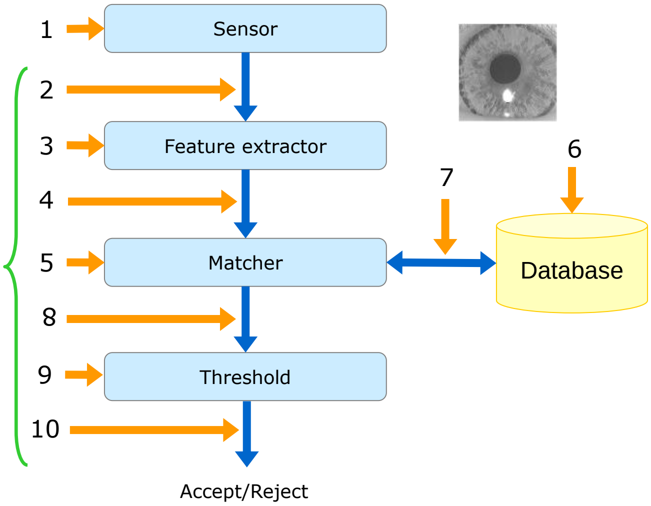
J. Galbally, J. Fierrez, J. Ortega-Garcia, "Vulnerabilities in biometric systems: attacks and recent advances in liveness detection", in *Proc. Spanish Workshop on Biometrics*, SWB, Girona, Spain, June 2007. [PDF]

2

Attack Points in Biometric Systems

DIRECT ATTACKS (aka PRESENTATION ATTACKS)
(Spoofing, mimicry)

INDIRECT ATTACKS
(Trojan Horse, Hill Climbing, Brute Force, channel interception, replay attacks, …)



A. Hadid, et al., "Biometrics systems under spoofing attack: an evaluation methodology and lessons learned", *IEEE Signal Processing Magazine*, Sept. 2015.

J. Galbally, J. Fierrez, J. Ortega-Garcia, "Vulnerabilities in biometric systems: attacks and recent advances in liveness detection", in *Proc. Spanish Workshop on Biometrics*, SWB, Girona, Spain, June 2007. [**PDF**]

Security Evaluation in Biometric Systems

- Steps for security evaluation of biometric systems:
  1) Description of the attack
  2) Description of the biometric systems being evaluated
  3) Description of the information required to be known by the attacker
  4) Description of the database
  5) Description of the tests that will be performed
  6) Compute the performance (FAR and FRR curves) of the systems being evaluated → determine the operating points where they will be tested
  7) Execution of the vulnerability evaluation in the defined operating points: Success Rate (SR), and Efficiency (E_{ff})
- Reporting the results
  - SR: percentage of accounts broken out of the total attacked
  - E_{ff}: average number of attempts needed to break an account

A. Hadid, et al., "Biometrics systems under spoofing attack: an evaluation methodology and lessons learned", *IEEE Signal Processing Magazine*, Sept. 2015.

J. Galbally, J. Fierrez, J. Ortega-Garcia, "Vulnerabilities in biometric systems: attacks and recent advances in liveness detection", in *Proc. Spanish Workshop on Biometrics*, SWB, Girona, Spain, June 2007. [**PDF**]

Security Evaluation in Biometric Systems: Standards

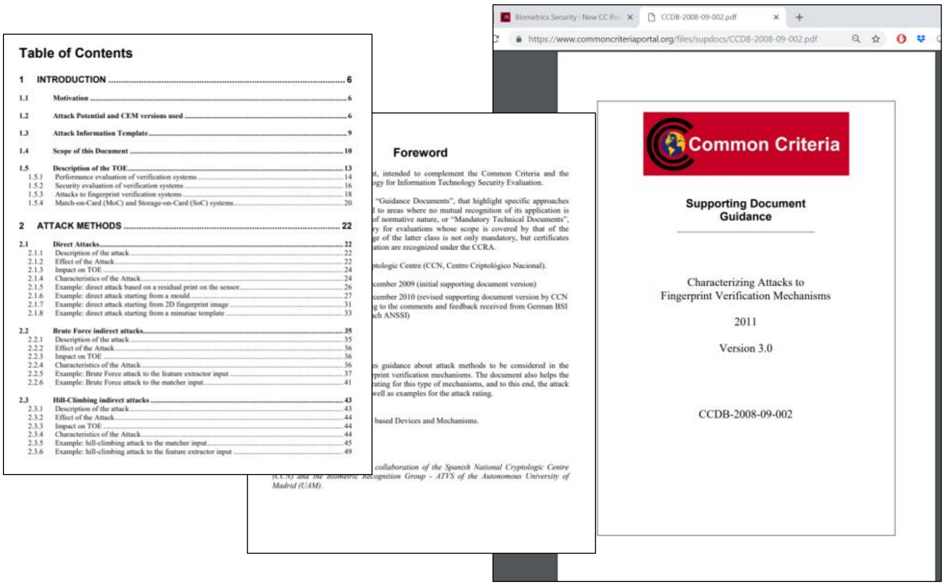


Table of Contents

Foreword

t, intended to complement the Common Criteria and the
ogy for Information Technology Security Evaluation.

"Guidance Documents", that highlight specific approaches
d to areas where no mutual recognition of its application is
of normative nature, or "Mandatory Technical Documents",
ry for evaluations whose scope is covered by that of the
ge of the latter class is not only mandatory, but certificates
ation are recognized under the CCRA.

ptologic Centre (CCN, Centro Criptológico Nacional).

cember 2009 (initial supporting document version)

cember 2010 (revised supporting document version by CCN
g to the comments and feedback received from German BSI
ch ANSSI)

s guidance about attack methods to be considered in the
rprint verification mechanisms. The document also helps the
rating for this type of mechanisms, and to this end, the attack
well as examples for the attack rating.

based Devices and Mechanisms.

collaboration of the Spanish National Cryptologic Centre (CCN) and the Biometric Recognition Group - ATVS of the Autonomous University of Madrid (UAM).

Biometrics Security : New CC Po
CCDB-2008-09-002.pdf
https://www.commoncriteriaportal.org/files/supdocs/CCDB-2008-09-002.pdf

Common Criteria

Supporting Document
Guidance

Characterizing Attacks to
Fingerprint Verification Mechanisms

2011

Version 3.0

CCDB-2008-09-002

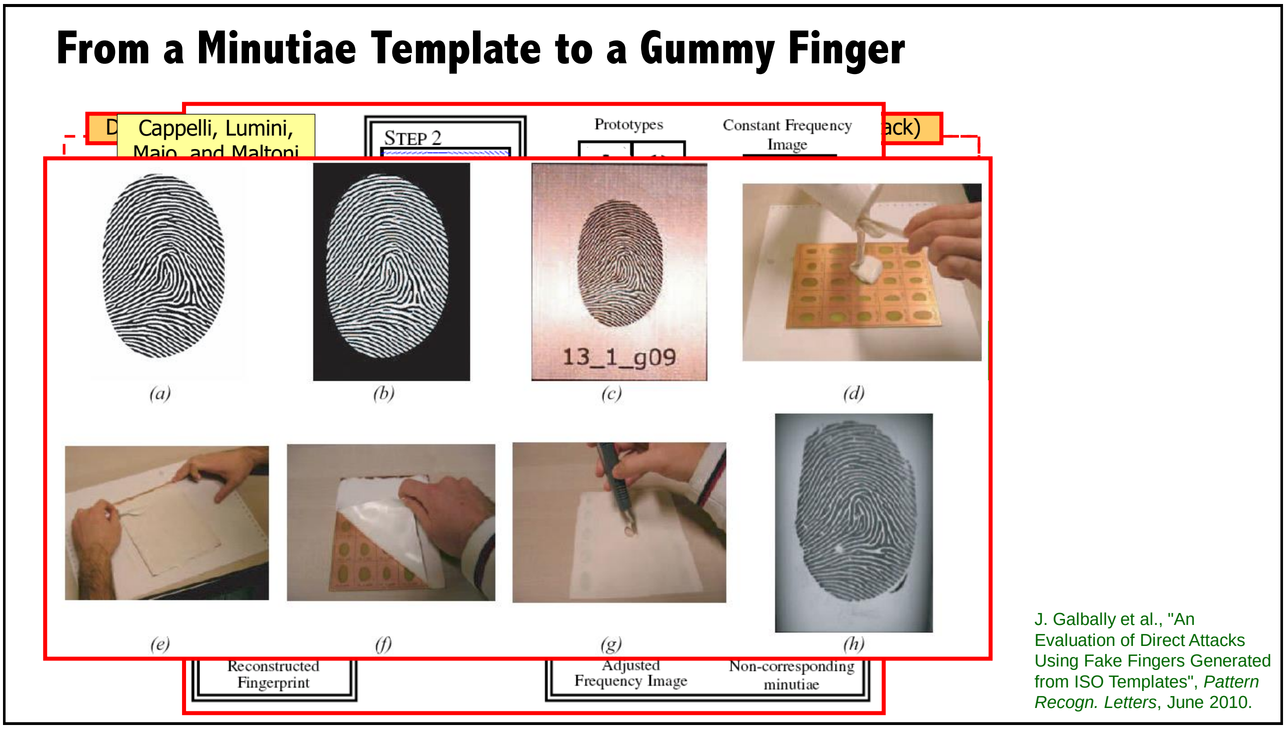
A. Merle, J. Bringer, J. Fierrez and N. Tekampe, "BEAT: A Methodology for Common Criteria Evaluations of Biometrics Systems", in *Proc. Intl. Common Criteria Conf.*, ICCC, London, UK, September 2015.

Case Study:

Security Evaluation of Direct Attacks from Stolen Fingerprint Templates

J. Galbally, J. Fierrez and R. Cappelli, "An Introduction to Fingerprint Presentation Attack Detection", in *Handbook of Biometric Anti-Spoofing*, S. Marcel and M. Nixon and J. Fierrez and N. Evans (Eds.), Springer, 2019, pp. 3-31.

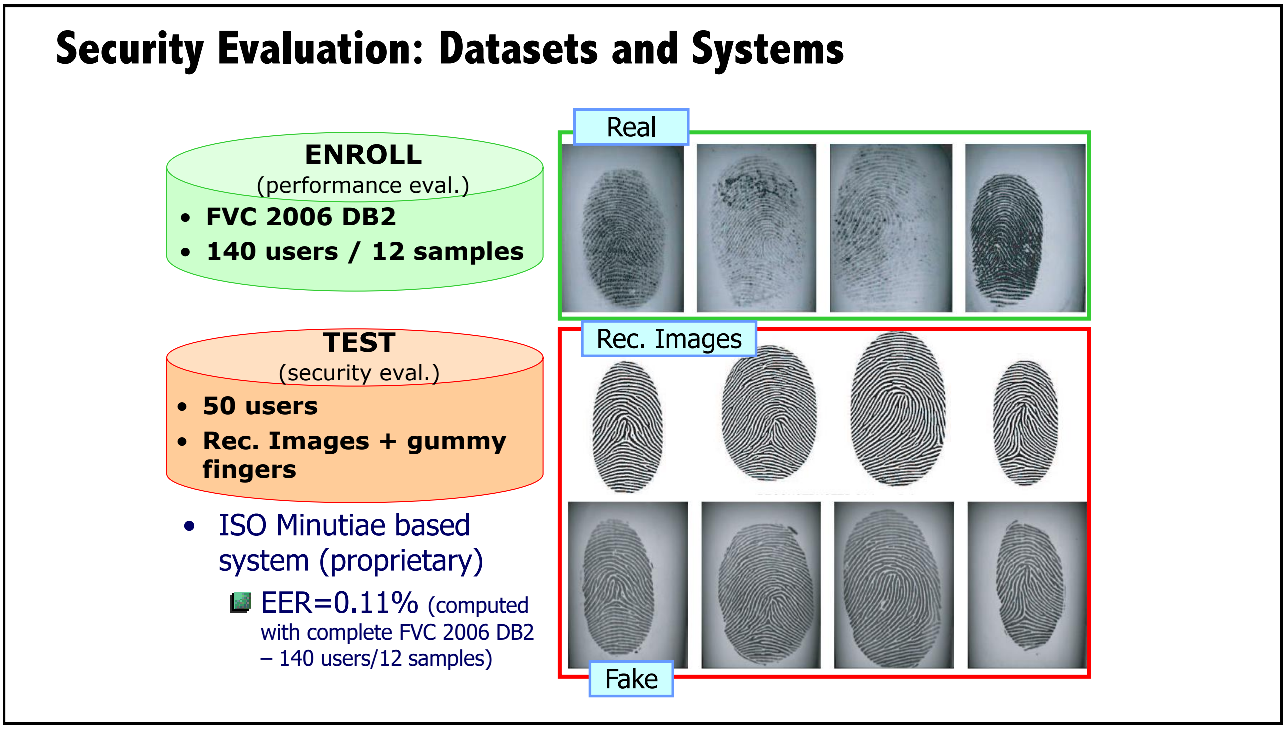
J. Galbally, R. Cappelli, A. Lumini, G. Gonzalez-de-Rivera, D. Maltoni, J. Fierrez, J. Ortega-Garcia and D. Maio, "An Evaluation of Direct Attacks Using Fake Fingers Generated from ISO Templates", *Pattern Recognition Letters*, June 2010.



From a Minutiae Template to a Gummy Finger

J. Galbally et al., "An Evaluation of Direct Attacks Using Fake Fingers Generated from ISO Templates", *Pattern Recogn. Letters*, June 2010.

7



Security Evaluation: Datasets and Systems

ENROLL
(performance eval.)

- **FVC 2006 DB2**
- **140 users / 12 samples**

TEST
(security eval.)

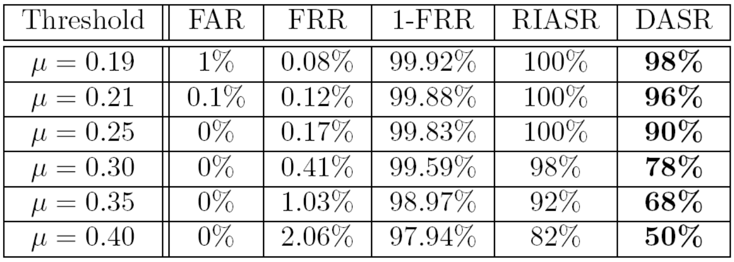
- **50 users**
- **Rec. Images + gummy fingers**

- ISO Minutiae based system (proprietary)
  - EER=0.11% (computed with complete FVC 2006 DB2 – 140 users/12 samples)

8

Security Evaluation: Results

- **RIASR** → Reconstructed Images Attack Success Rate
- **DASR** → Direct Attack Success Rate



Threshold	FAR	FRR	1-FRR	RIASR	DASR
$\mu = 0.19$	1%	0.08%	99.92%	100%	**98%**
$\mu = 0.21$	0.1%	0.12%	99.88%	100%	**96%**
$\mu = 0.25$	0%	0.17%	99.83%	100%	**90%**
$\mu = 0.30$	0%	0.41%	99.59%	98%	**78%**
$\mu = 0.35$	0%	1.03%	98.97%	92%	**68%**
$\mu = 0.40$	0%	2.06%	97.94%	82%	**50%**

- Loss of performance between the indirect and direct attack → related to quality loss
- The system is still highly vulnerable to the direct attack: SR=50% for very high security point, SR=78% for more realistic op. point
- Standards are positive, BUT provide information to attackers → solutions should be found

J. Galbally, R. Cappelli, A. Lumini, G. Gonzalez-de-Rivera, D. Maltoni, J. Fierrez, J. Ortega-Garcia and D. Maio, "An Evaluation of Direct Attacks Using Fake Fingers Generated from ISO Templates", *Pattern Recognition Letters*, June 2010.

9

Countermeasuring Direct Attacks to Biometric Systems

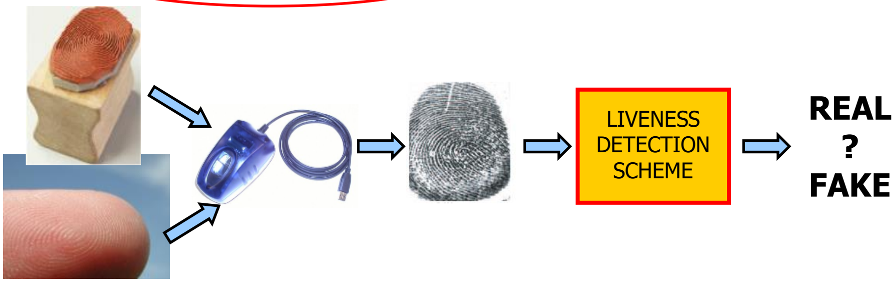
J. Galbally, J. Fierrez and R. Cappelli, "An Introduction to Fingerprint Presentation Attack Detection", in *Handbook of Biometric Anti-Spoofing*, S. Marcel and M. Nixon and J. Fierrez and N. Evans (Eds.), Springer, 2019, pp. 3-31.

J. Galbally, S. Marcel and J. Fierrez, "Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint and Face Recognition", *IEEE Trans. on Image Processing*, February 2014.

10

Introduction to Liveness Detection (aka Anti-Spoofing, aka Presentation Attack Detection)

- Countermeasures to direct attacks:
  - Multibiometrics (accurate trait + trait difficult to spoof)
  - Liveness Detection (use physiological property)
- Liveness Detection → Two class problem: REAL / FAKE
  - Hardware-based solutions: odour, heart beat, electric properties...
  - Software-based solutions: elastic properties, ridge pattern...

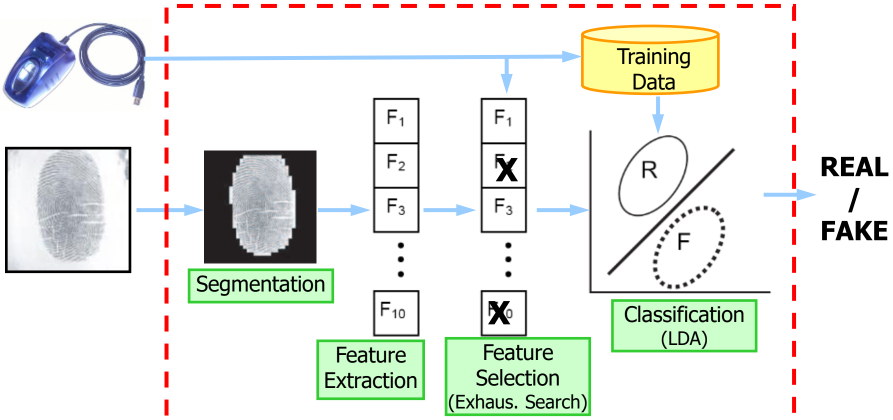


LIVENESS DETECTION SCHEME

REAL
?
FAKE

11

Liveness Detection based on Quality Features

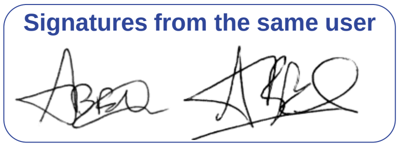


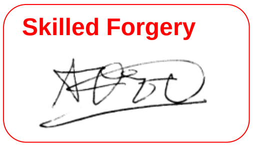
J. Galbally, "A High Performance Fingerprint Liveness Detection Method Based on Quality Related Features", *Future Generation Computer Sys.*, Jan 2012.

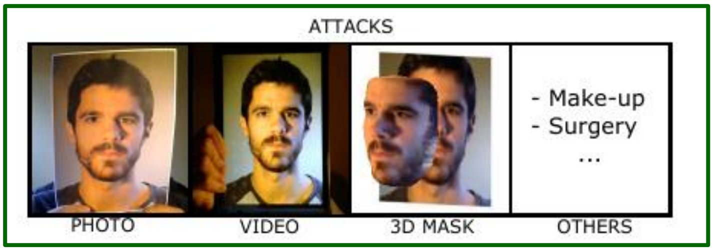
F. Alonso-Fernandez, "A comparative study of fingerprint image-quality estimation methods", *IEEE Trans. on Information Forensics and Security*, Dec. 2007.

12

Other Biometrics







A. Morales, J. Fierrez and J. Galbally and Marta Gomez-Barrero, "Introduction to Iris Presentation Attack Detection", in *Handbook of Biometric Anti-Spoofing*, S. Marcel and M. Nixon and J. Fierrez and N. Evans (Eds.), Springer, 2019, pp. 135-150. [PDF]

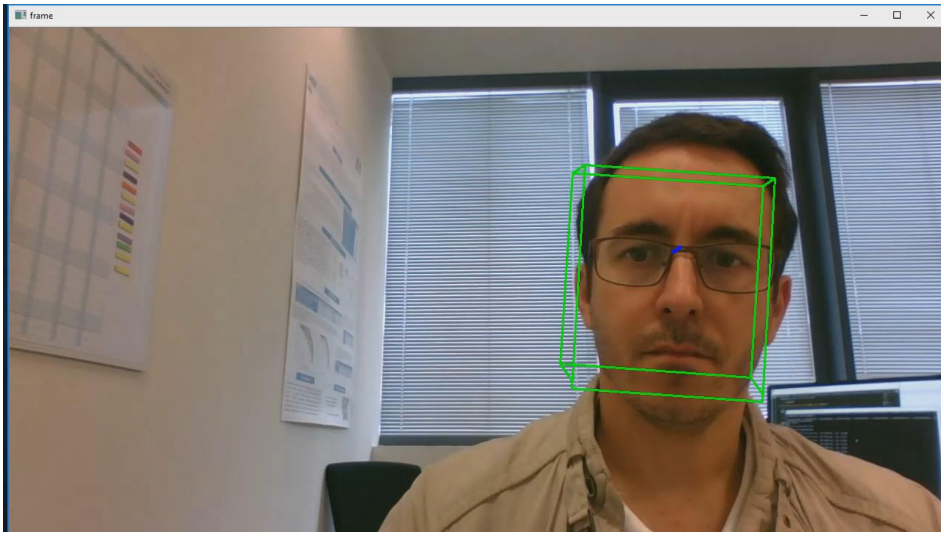
R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Presentation Attacks in Signature Biometrics: Types and Introduction to Attack Detection", in *Handbook of Biometric Anti-Spoofing*, S. Marcel and M. Nixon and J. Fierrez and N. Evans (Eds.), Springer, 2019, pp. 439-453. [PDF]

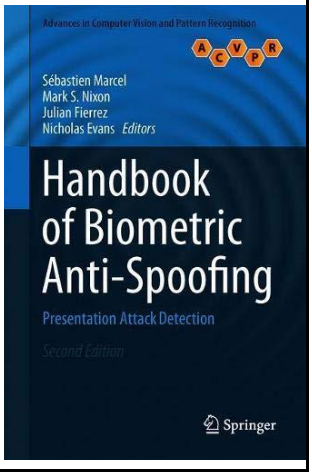
J. Hernandez-Ortega, J. Fierrez, A. Morales and J. Galbally, "Introduction to Face Presentation Attack Detection", in *Handbook of Biometric Anti-Spoofing*, S. Marcel and M. Nixon and J. Fierrez and N. Evans (Eds.), Springer, 2019, pp. 187-206. [PDF]

J. Galbally, S. Marcel and J. Fierrez, "Biometric Anti-spoofing Methods: A Survey in Face Recognition", *IEEE Access*, December 2014.

13

Attacks to other Biometrics: Face




Advances in Computer Vision and Pattern Recognition

Sébastien Marcel
Mark S. Nixon
Julian Fierrez
Nicholas Evans *Editors*

Handbook of Biometric Anti-Spoofing

Presentation Attack Detection

Second Edition

Springer

J. Galbally, S. Marcel and J. Fierrez, "Biometric Anti-spoofing Methods: A Survey in Face Recognition", *IEEE Access*, December 2014.

14

DeepFakes: old Problem, new Threat

R. Tolosana, R. Vera-Rodriguez, et al., “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” *Information Fusion*, December 2020.

J. Galbally, J. Fierrez, J. Ortega-Garcia, "Vulnerabilities in biometric systems: attacks and recent advances in liveness detection", in *Proc. Spanish Workshop on Biometrics*, SWB, Girona, Spain, June 2007. [**PDF**]

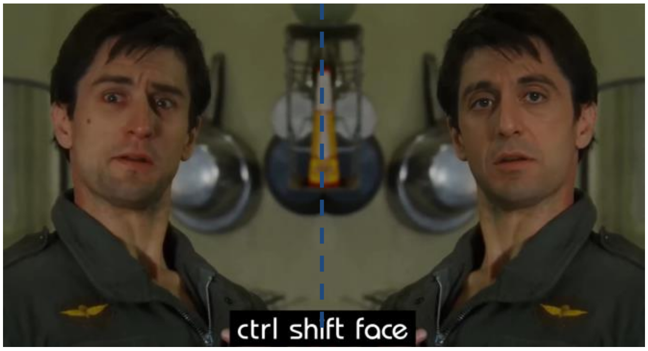
15

What are DeepFakes?

In general, the popular term DeepFakes is referred to all digital fake content created by means of deep learning techniques.

Real Video
(Robert de Niro)

DeepFake Video
(Al Pacino)

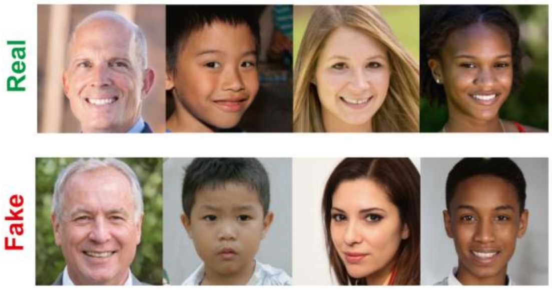


R. Tolosana, R. Vera-Rodriguez, et al., “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” *Information Fusion*, December 2020.

16

Types of DeepFakes

Image Level



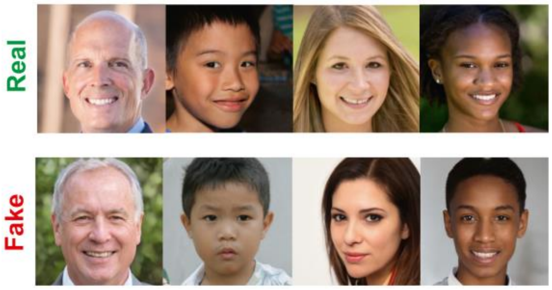
Entire Face Synthesis

R. Tolosana, R. Vera-Rodriguez, et al., “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” *Information Fusion*, December 2020.

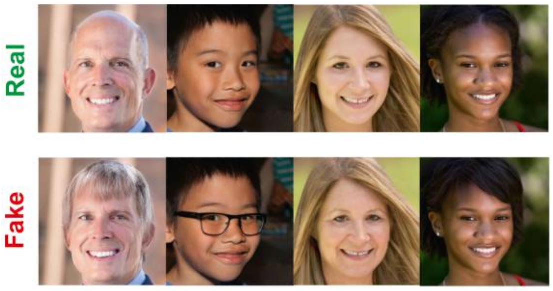
17

Types of DeepFakes

Image Level



Entire Face Synthesis



Attribute Manipulation
(aka Face Retouching)

R. Tolosana, R. Vera-Rodriguez, et al., “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” *Information Fusion*, December 2020.

18

Types of DeepFakes

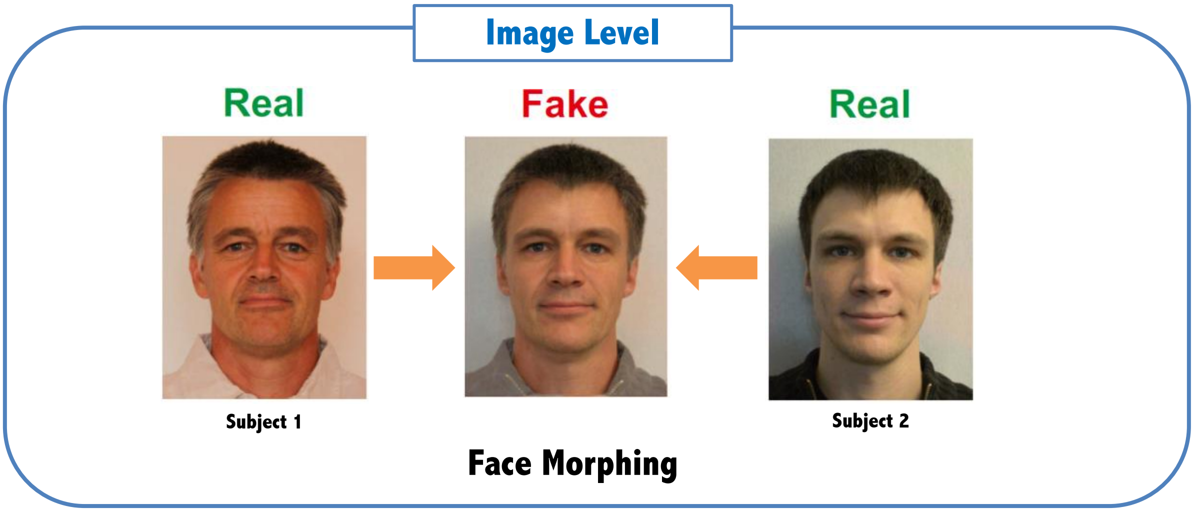


Image Level

Real

Fake

Real

Subject 1

Subject 2

Face Morphing

R. Tolosana, R. Vera-Rodriguez, et al., “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” *Information Fusion*, December 2020.

19

Types of DeepFakes

Video Level

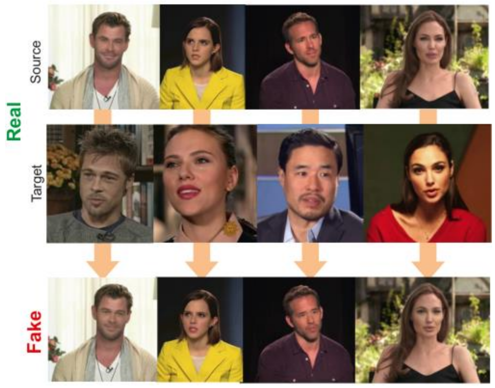
R. Tolosana, R. Vera-Rodriguez, et al., “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” *Information Fusion*, December 2020.

20

Types of DeepFakes

Video Level

Identity Swap



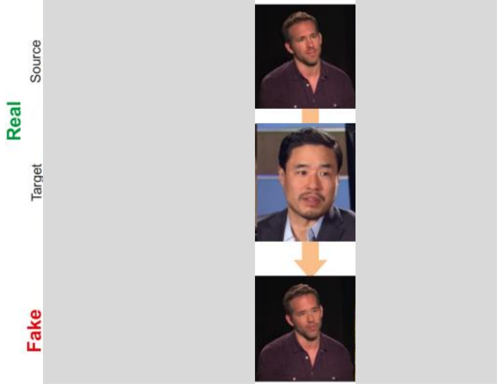
R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, December 2020.

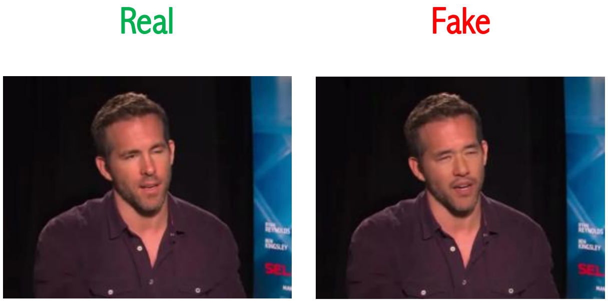
21

Types of DeepFakes

Video Level

Identity Swap





Examples from Celeb-DF database

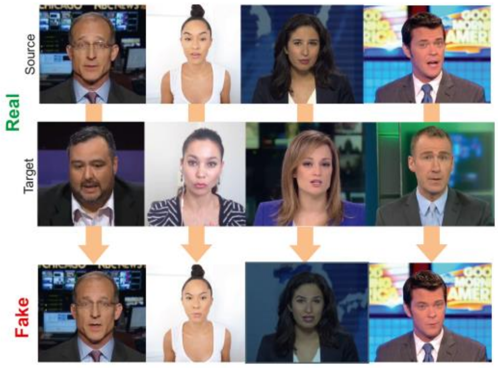
R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, December 2020.

22

Types of DeepFakes

Video Level

Expression Swap



R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, December 2020.

23

Types of DeepFakes

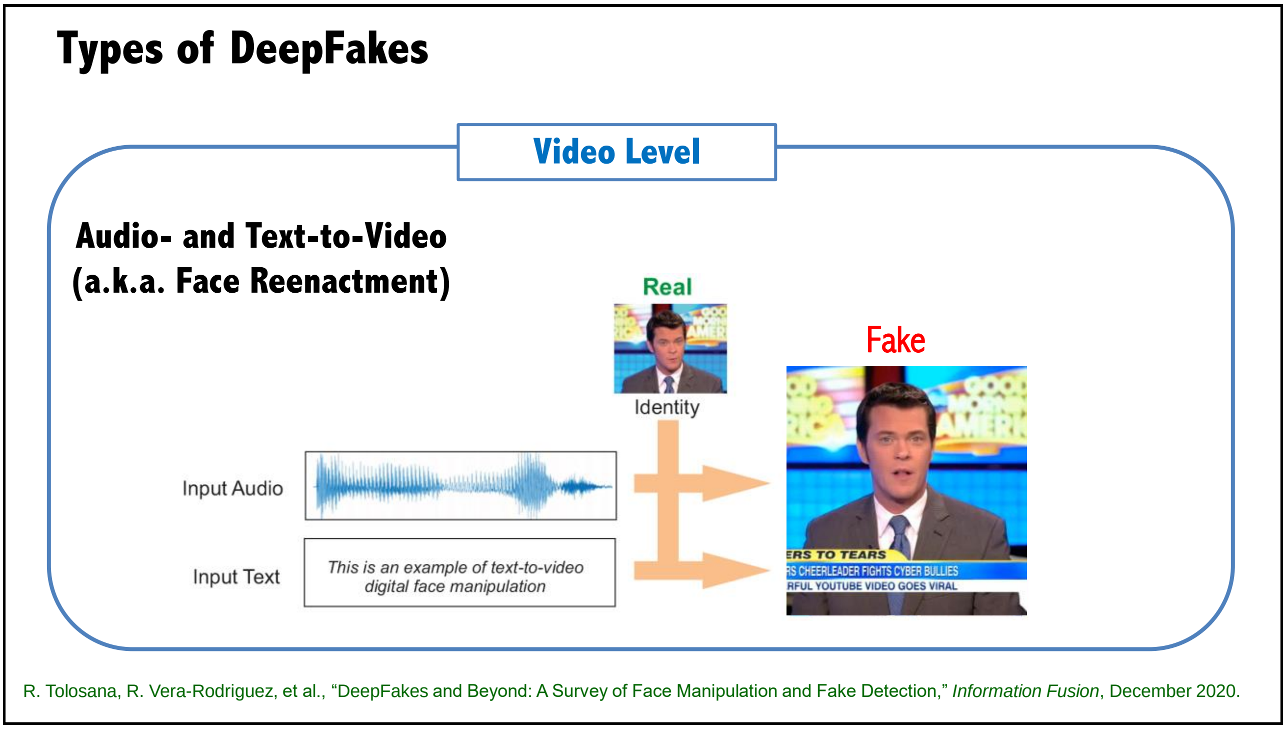
Video Level

Expression Swap

Examples from FaceForensics++ database

R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, December 2020.

24



Types of DeepFakes

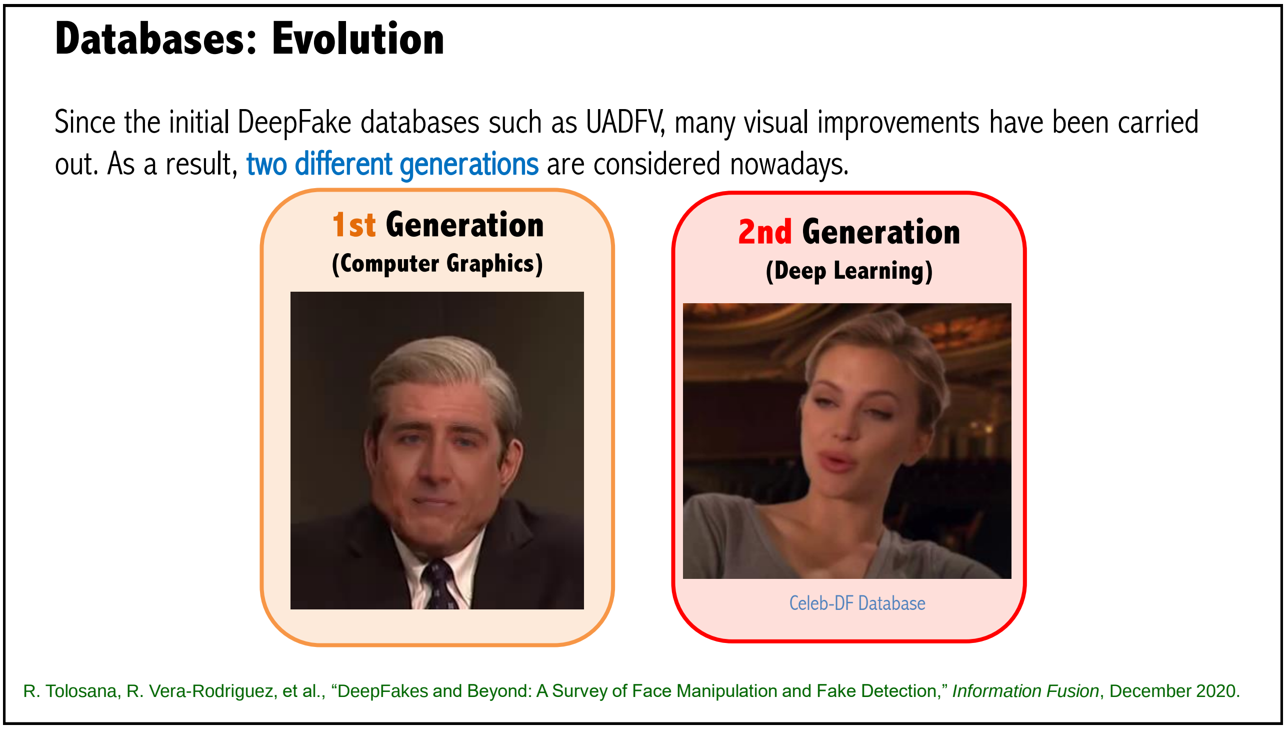
Video Level

Audio- and Text-to-Video (a.k.a. Face Reenactment)

Real
Identity
Fake
Input Audio
Input Text
This is an example of text-to-video digital face manipulation

R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, December 2020.

25

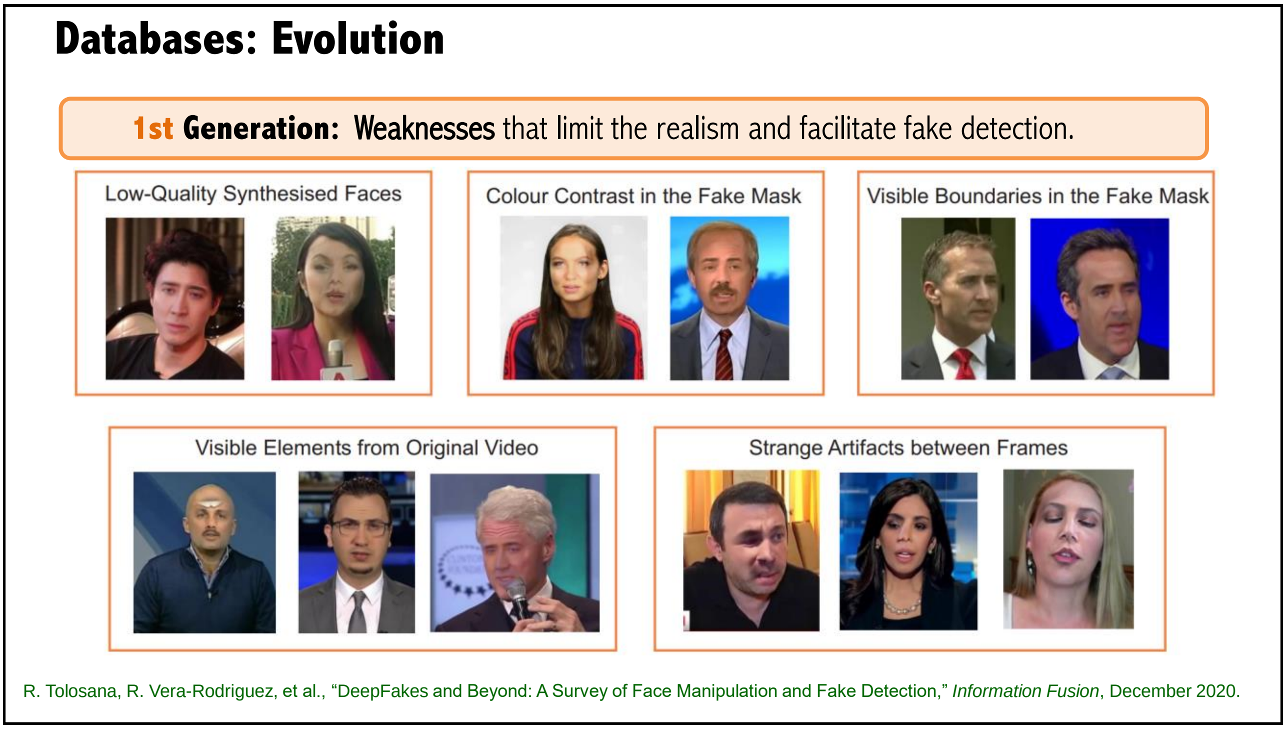


Databases: Evolution

Since the initial DeepFake databases such as UADFV, many visual improvements have been carried out. As a result, two different generations are considered nowadays.

R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, December 2020.

26



Databases: Evolution

1st Generation: Weaknesses that limit the realism and facilitate fake detection.

Low-Quality Synthesised Faces

Colour Contrast in the Fake Mask

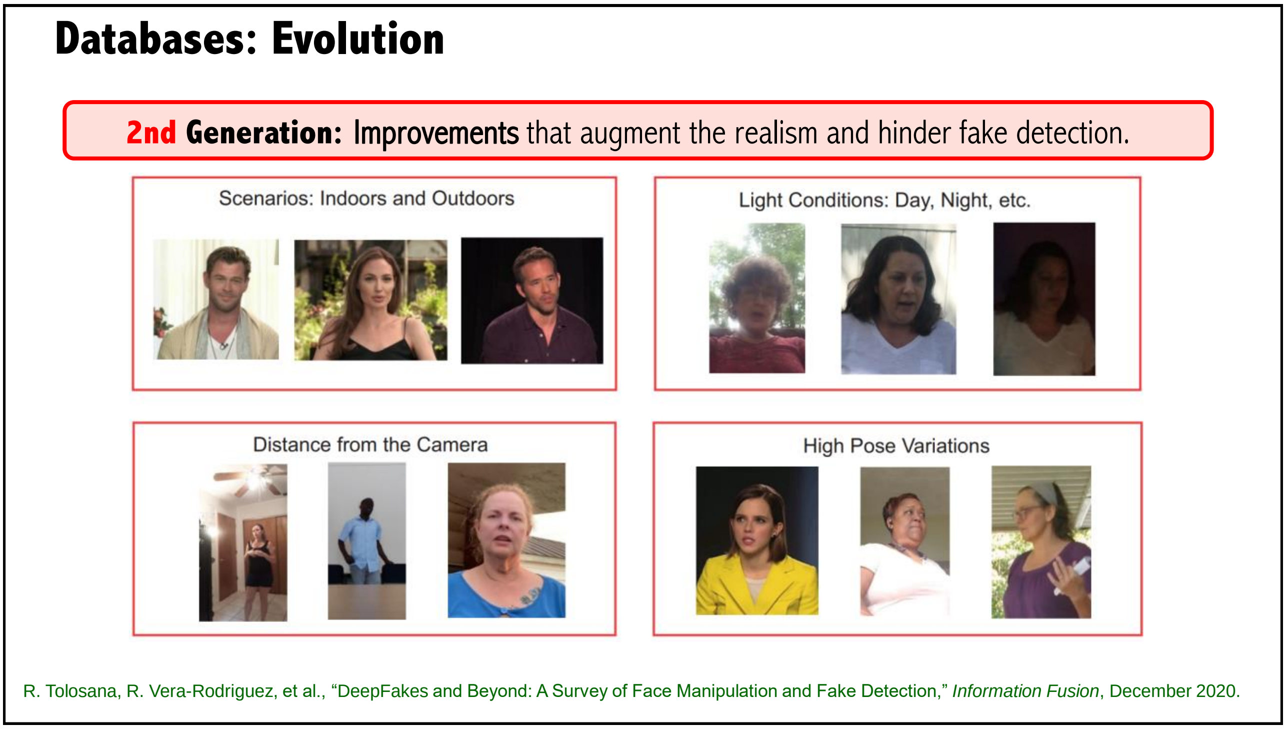
Visible Boundaries in the Fake Mask

Visible Elements from Original Video

Strange Artifacts between Frames

R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, December 2020.

27



Databases: Evolution

2nd Generation: Improvements that augment the realism and hinder fake detection.

Scenarios: Indoors and Outdoors

Light Conditions: Day, Night, etc.

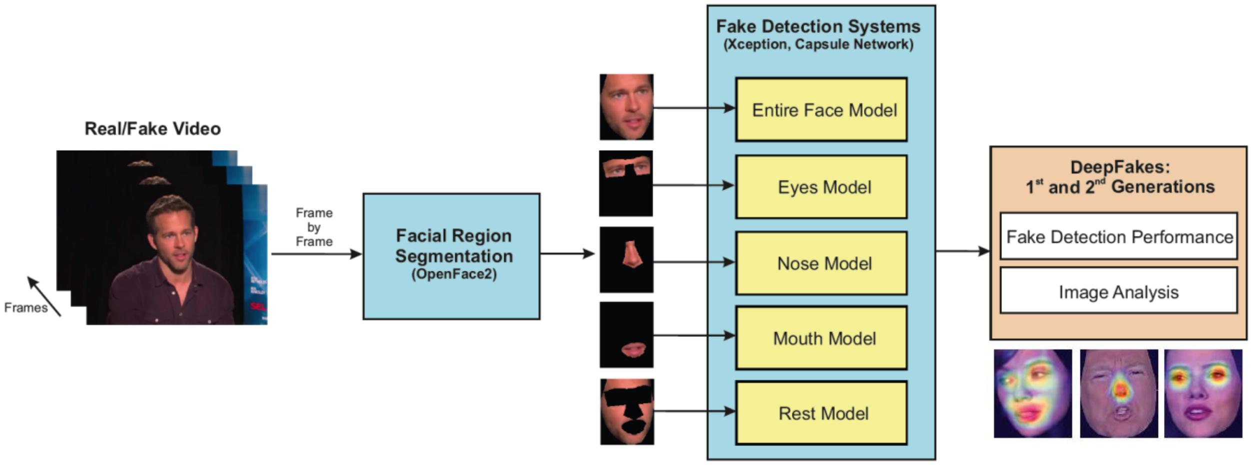
Distance from the Camera

High Pose Variations

R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, December 2020.

28

DeepFakes Evolution: Analysis of Facial Regions and Performance



Real/Fake Video

Frames

Frame by Frame

Facial Region Segmentation (OpenFace2)

Fake Detection Systems (Xception, Capsule Network)

Entire Face Model

Eyes Model

Nose Model

Mouth Model

Rest Model

DeepFakes: 1st and 2nd Generations

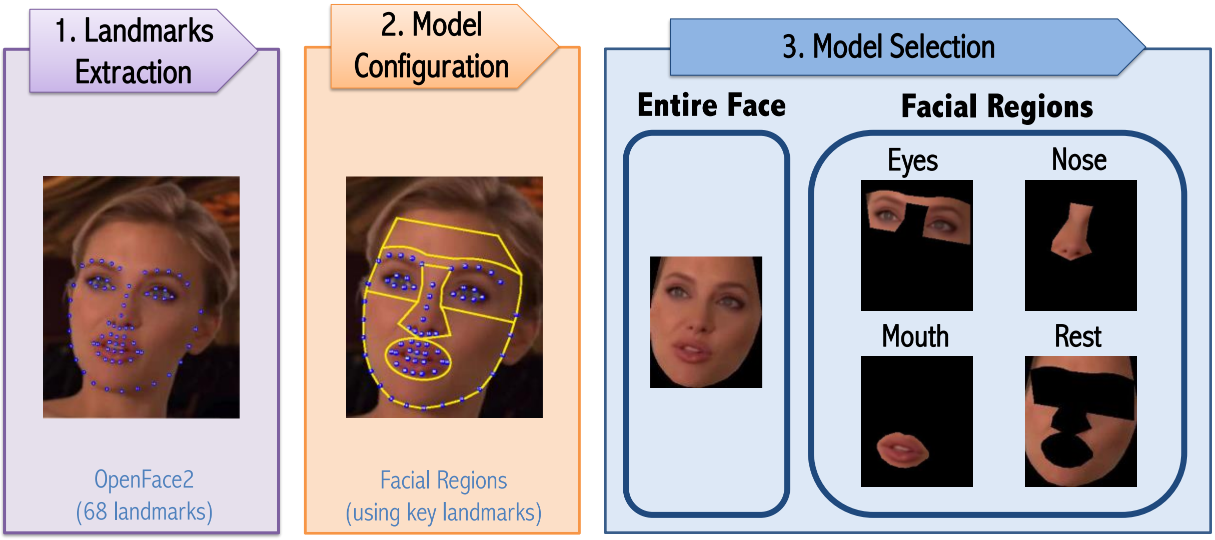
Fake Detection Performance

Image Analysis

R. Tolosana, S. Romero-Tapiador, J. Fierrez and R. Vera-Rodriguez, "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance", in *Proc. International Conference on Pattern Recognition Workshops*, ICPRw, Milan, Italy, 2021.

29

Facial Region Segmentation



1. Landmarks Extraction

OpenFace2 (68 landmarks)

2. Model Configuration

Facial Regions (using key landmarks)

3. Model Selection

Entire Face

Facial Regions

Eyes

Nose

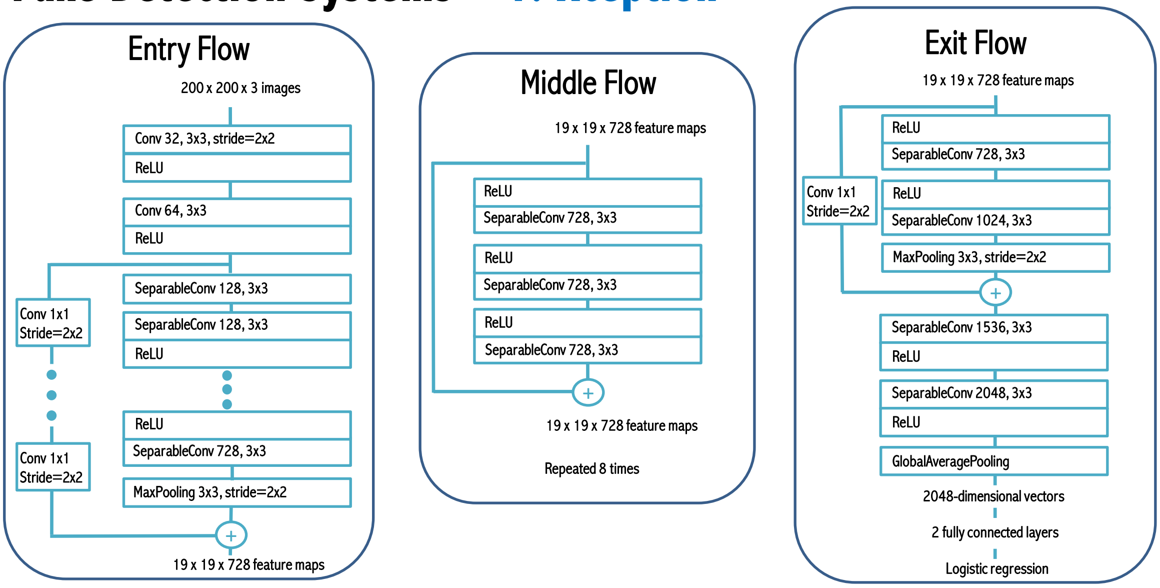
Mouth

Rest

R. Tolosana, S. Romero-Tapiador, J. Fierrez and R. Vera-Rodriguez, "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance", in *Proc. International Conference on Pattern Recognition Workshops*, ICPRw, Milan, Italy, 2021.

30

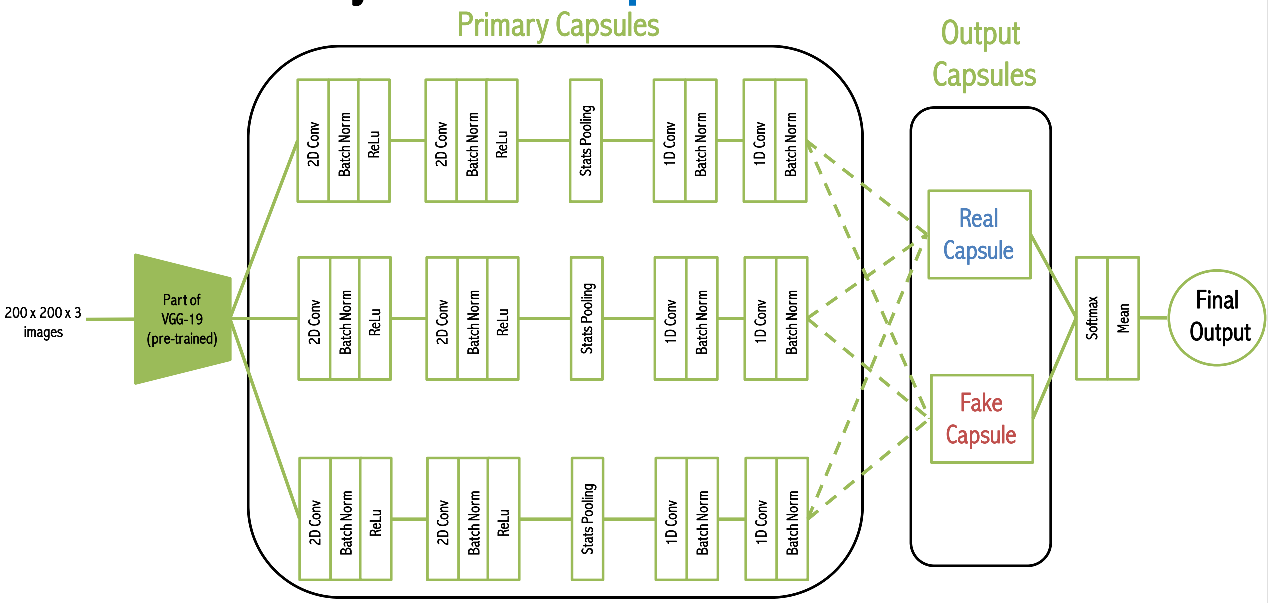
Fake Detection Systems – 1. Xception



F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *CVPR* 2017.

31

Fake Detection Systems – 2. Capsule Network



H. Nguyen, J. Yamagishi, I. Echizen, "Use of a Capsule Network to Detect Fake Images and Videos," arXiv:1910.12467, 2019.

32

Databases – 1st Generation

UADFV

- 49 real/fake videos
- FakeApp

FaceForensics++

- 1000 real/fake videos
- FaceSwap

Y. Li, M. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking,” in *WIFS* 2018.

A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and ¨ M. Nießner, “FaceForensics++: Learning to Detect Manipulated Facial Images,” in *ICCV* 2019.

33

Databases – 2nd Generation

Celeb-DF v1

- 408 real and 795 fake videos
- Deep Learning

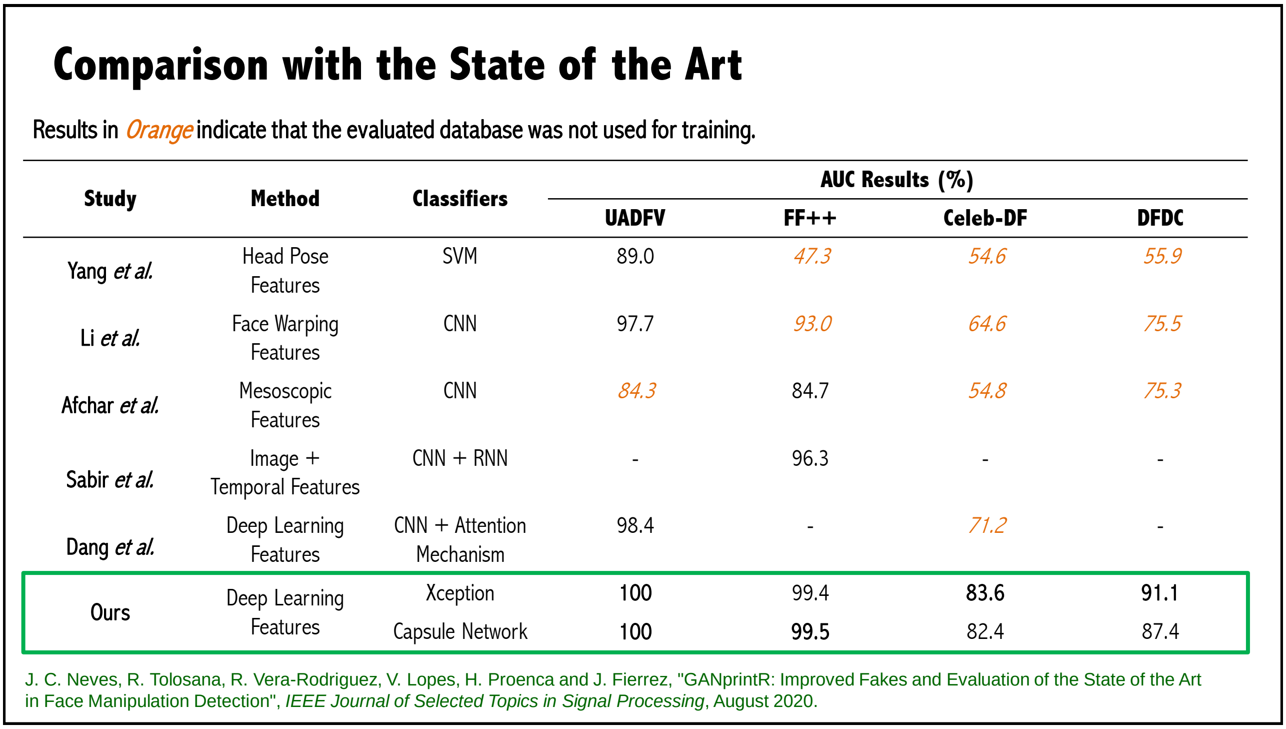
DFDC

- ≃ 1000 real and 5000 fake videos
- Two different approaches

Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, “Celeb-DF: A LargeScale Challenging Dataset for DeepFake Forensics,” in *CVPR* 2020.

B. Dolhansky, R. Howes, et al., “The Deepfake Detection Challenge (DFDC) Preview Dataset,” arXiv preprint arXiv:1910.08854, 2019.

34



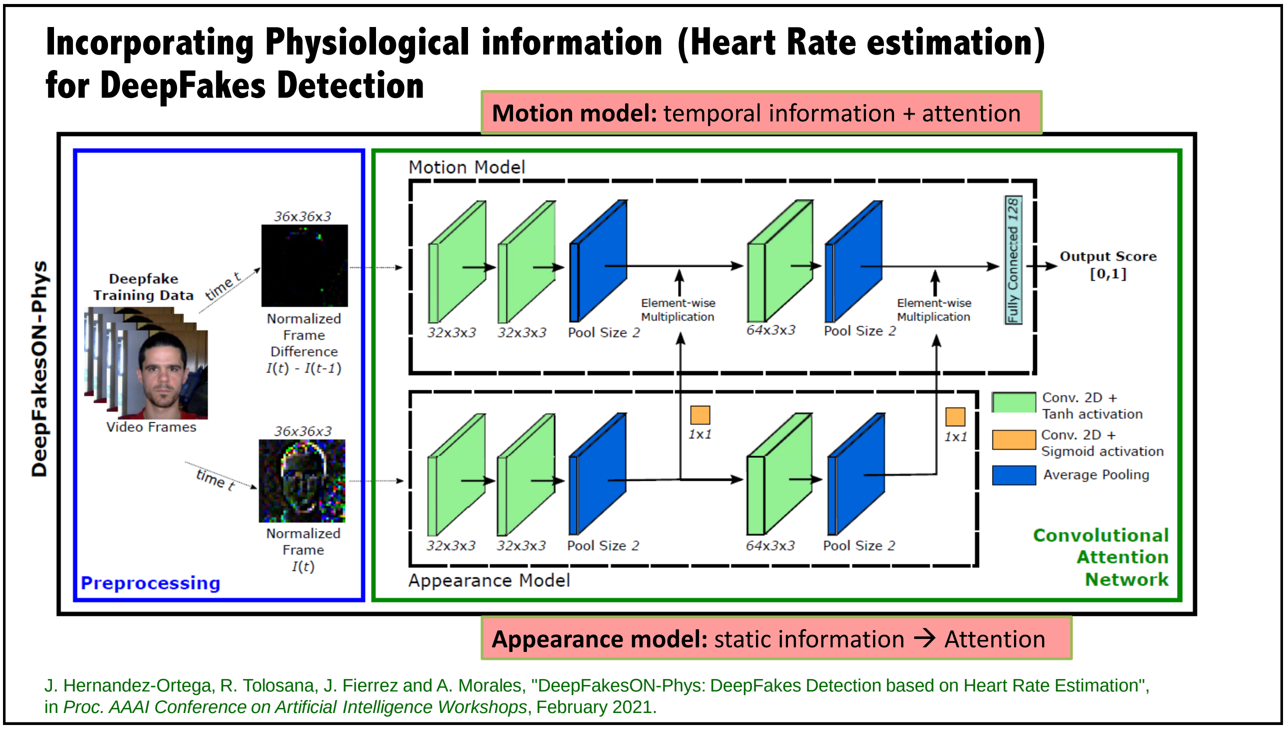
Comparison with the State of the Art

Results in *Orange* indicate that the evaluated database was not used for training.

Study	Method	Classifiers	AUC Results (%)			
			UADFV	**FF++**	**Celeb-DF**	**DFDC**
Yang *et al.*	Head Pose Features	SVM	89.0	*47.3*	*54.6*	*55.9*
Li *et al.*	Face Warping Features	CNN	97.7	*93.0*	*64.6*	*75.5*
Afchar *et al.*	Mesoscopic Features	CNN	*84.3*	84.7	*54.8*	*75.3*
Sabir *et al.*	Image + Temporal Features	CNN + RNN	-	96.3	-	-
Dang *et al.*	Deep Learning Features	CNN + Attention Mechanism	98.4	-	*71.2*	-
Ours	Deep Learning Features	Xception	**100**	99.4	**83.6**	**91.1**
		Capsule Network	**100**	**99.5**	82.4	87.4

J. C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proenca and J. Fierrez, "GANprintR: Improved Fakes and Evaluation of the State of the Art in Face Manipulation Detection", *IEEE Journal of Selected Topics in Signal Processing*, August 2020.

35



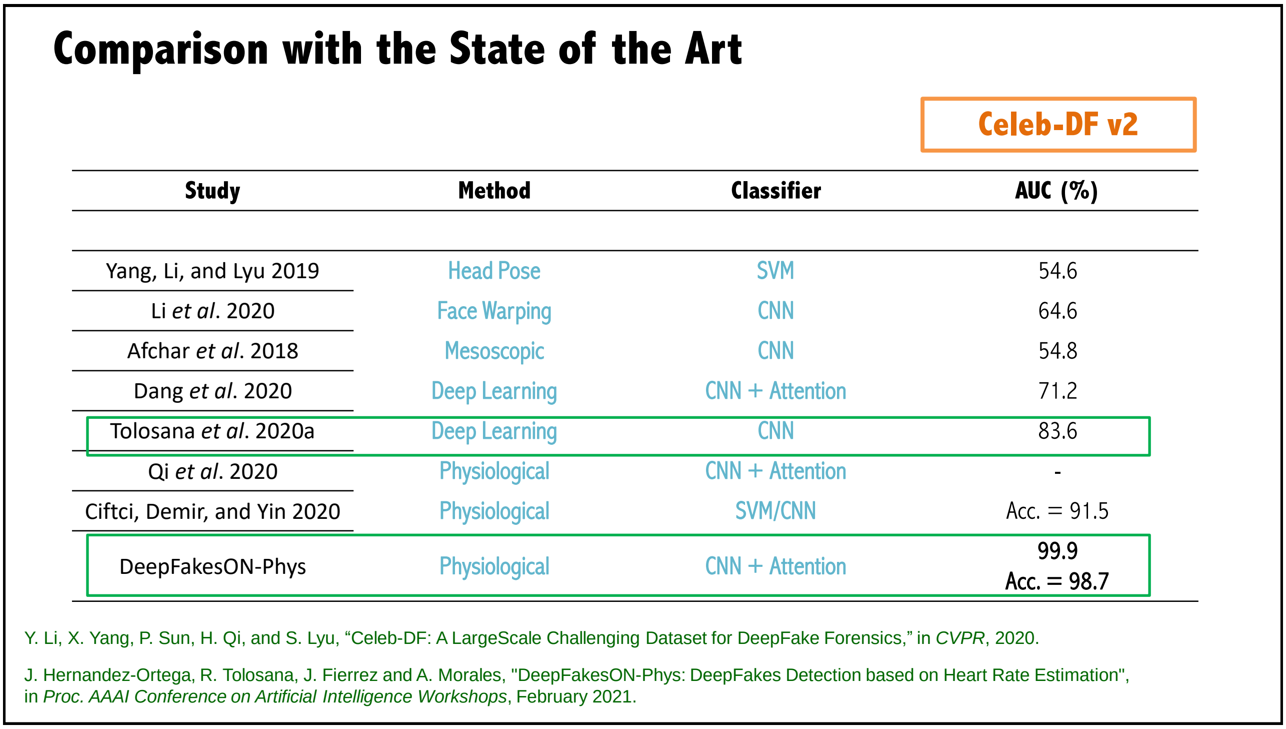
Incorporating Physiological information (Heart Rate estimation) for DeepFakes Detection

Motion model: temporal information + attention

Appearance model: static information → Attention

J. Hernandez-Ortega, R. Tolosana, J. Fierrez and A. Morales, "DeepFakesON-Phys: DeepFakes Detection based on Heart Rate Estimation", in *Proc. AAAI Conference on Artificial Intelligence Workshops*, February 2021.

36



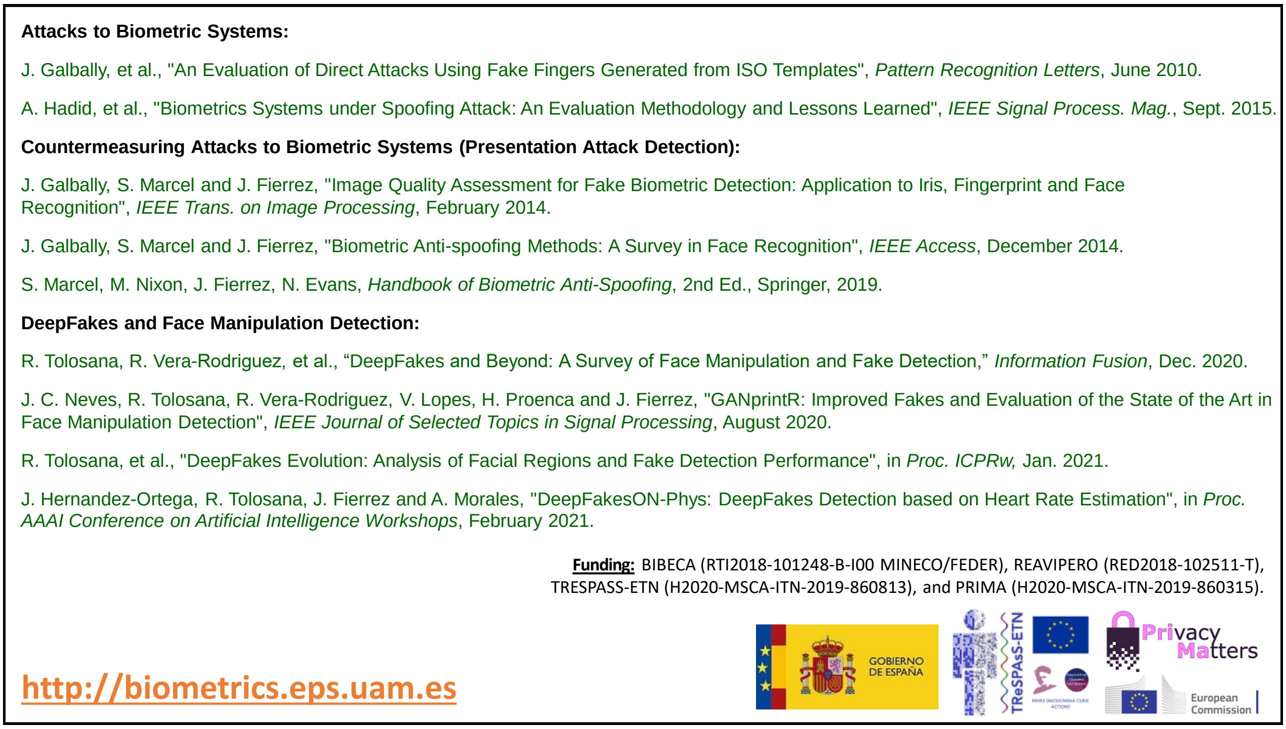
Comparison with the State of the Art

Study	Method	Classifier	AUC (%)
			Celeb-DF v2
Yang, Li, and Lyu 2019	Head Pose	SVM	54.6
Li *et al*. 2020	Face Warping	CNN	64.6
Afchar *et al*. 2018	Mesoscopic	CNN	54.8
Dang *et al*. 2020	Deep Learning	CNN + Attention	71.2
Tolosana *et al*. 2020a	Deep Learning	CNN	83.6
Qi *et al*. 2020	Physiological	CNN + Attention	-
Ciftci, Demir, and Yin 2020	Physiological	SVM/CNN	Acc. = 91.5
DeepFakesON-Phys	Physiological	CNN + Attention	99.9 Acc. = 98.7

Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A LargeScale Challenging Dataset for DeepFake Forensics," in *CVPR*, 2020.

J. Hernandez-Ortega, R. Tolosana, J. Fierrez and A. Morales, "DeepFakesON-Phys: DeepFakes Detection based on Heart Rate Estimation", in *Proc. AAAI Conference on Artificial Intelligence Workshops*, February 2021.

37



Attacks to Biometric Systems:

J. Galbally, et al., "An Evaluation of Direct Attacks Using Fake Fingers Generated from ISO Templates", *Pattern Recognition Letters*, June 2010.

A. Hadid, et al., "Biometrics Systems under Spoofing Attack: An Evaluation Methodology and Lessons Learned", *IEEE Signal Process. Mag.*, Sept. 2015.

Countermeasuring Attacks to Biometric Systems (Presentation Attack Detection):

J. Galbally, S. Marcel and J. Fierrez, "Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint and Face Recognition", *IEEE Trans. on Image Processing*, February 2014.

J. Galbally, S. Marcel and J. Fierrez, "Biometric Anti-spoofing Methods: A Survey in Face Recognition", *IEEE Access*, December 2014.

S. Marcel, M. Nixon, J. Fierrez, N. Evans, *Handbook of Biometric Anti-Spoofing*, 2nd Ed., Springer, 2019.

DeepFakes and Face Manipulation Detection:

R. Tolosana, R. Vera-Rodriguez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, Dec. 2020.

J. C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proenca and J. Fierrez, "GANprintR: Improved Fakes and Evaluation of the State of the Art in Face Manipulation Detection", *IEEE Journal of Selected Topics in Signal Processing*, August 2020.

R. Tolosana, et al., "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance", in *Proc. ICPRw,* Jan. 2021.

J. Hernandez-Ortega, R. Tolosana, J. Fierrez and A. Morales, "DeepFakesON-Phys: DeepFakes Detection based on Heart Rate Estimation", in *Proc. AAAI Conference on Artificial Intelligence Workshops*, February 2021.

Funding: BIBECA (RTI2018-101248-B-I00 MINECO/FEDER), REAVIPERO (RED2018-102511-T), TRESPASS-ETN (H2020-MSCA-ITN-2019-860813), and PRIMA (H2020-MSCA-ITN-2019-860315).

http://biometrics.eps.uam.es

38