BiDA Lab
Biometrics & Data Pattern Analytics Lab

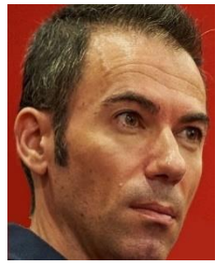
UAM
Universidad Autónoma de Madrid

DeepFakesON-Phys:
DeepFakes Detection based on Heart Rate Estimation



Msc. Javier HERNANDEZ-ORTEGA

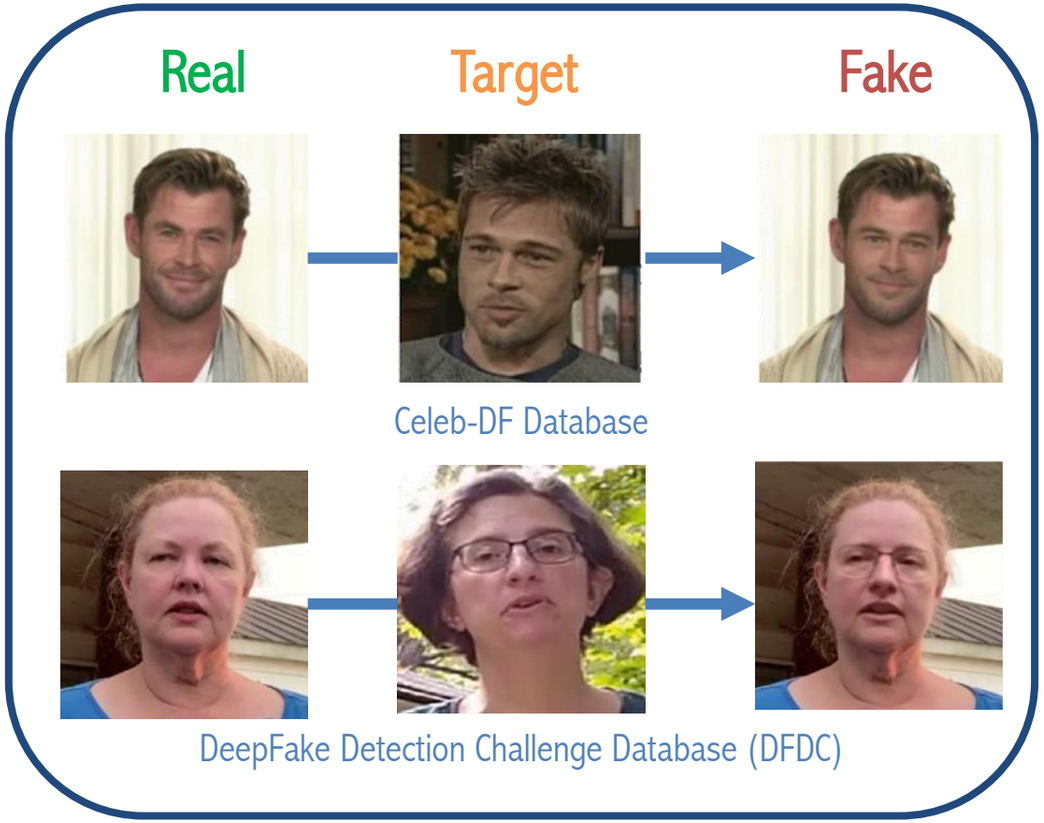

Dr. Ruben TOLOSANA


Prof. Julian FIERREZ


Prof. Aythami MORALES

Introduction

- **DeepFake (Identity Swap)** is referred to a deep learning based technique able to create fake videos by **swapping** the face of a person by the face of another person [1].



[1] Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; and Ortega-Garcia, J. 2020. **"DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection"**. *Information Fusion* 64: 131–148.

Introduction

- **Face manipulation techniques:** mostly based on AutoEncoders (AE) [2] and Generative Adversarial Networks (GAN) [3].
- **Very realistic visual results**: specially in the latest generation of public DeepFakes [4].

Real Video
(Robert de Niro)

DeepFake Video
(Al Pacino)

[2] Kingma, D. P.; and Welling, M. 2013. **"Auto-Encoding Variational Bayes"**. In *Proc. Int. Conf. on Learning Represent.*

[3] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.;Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. **"Generative Adversarial Nets"**. In *Proc. Advances in Neural Information Processing Systems.*

[4] Tolosana, R.; Romero-Tapiador, S.; Fierrez, J.; and Vera-Rodriguez, R. 2020. **"DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance"**. In *Proc. International Conference on Pattern Recognition Workshops* .

Introduction

- **Face Recognition Presentation Attack:** using photographs, videos, and masks [5].

- **3D Masks :** somehow similar to DeepFake digital manipulations.
  - Physical vs digital mask over the real face.
- **Texture and shape**-based techniques **not efficient** against hyperrealistic 3D Masks [6].
  - Same with realistic DeepFake methods.
  - Other approaches are necessary → Physiology.

[5] Hernandez-Ortega, J.; Fierrez, J.; Morales, A.; and Galbally, J. 2019. **"Introduction to Face Presentation Attack Detection"**. In *Handbook of Biometric Anti-Spoofing*, 187–206. Springer.

[6] Erdogmus, N.; and Marcel, S. 2014. **"Spoofing Face Recognition with 3D Masks"**. *IEEE Transactions on Information Forensics and Security* 9(7): 1084–1097.

Introduction

- 3D Masks do not emulate the physiology of human beings [6], i.e. HR, blood oxygen, breath rate.
  - **Estimating them** is a powerful tool for 3D Masks detection.

- Do DeepFake manipulations consider the physiological aspects in the synthesis process?

- Detection based on pulse detection → Remote Photoplethysmography [7], used in:
  - E-learning [Hernandez-Ortega *et al.* 2020].
  - Health Care [Mc-Duff *et al.* 2015].
  - Human-Computer Interaction [Tan and Nijholt 2010].
  - Security [Marcel *et al.* 2019].

[7] Hernandez-Ortega, J.; Fierrez, J.; Morales, A.; and Tome, P. 2018. **"Time Analysis of Pulse-Based Face Anti-Spoofing in Visible and NIR"**. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition Workshops*.

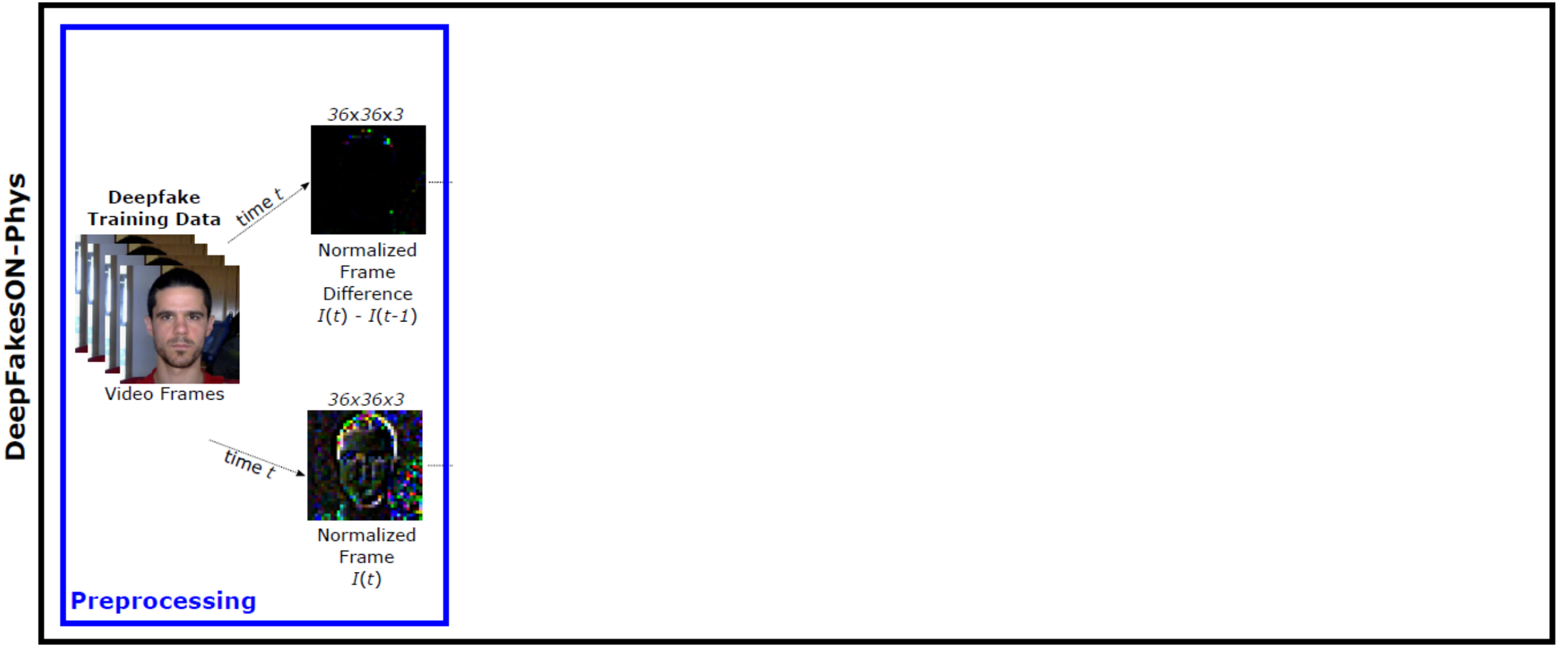
Contributions

- DeepFake detector based on physiological measurement: DeepFakesON-Phys.
  - Based on Deep Learning.
  - rPPG features pretrained for heart rate estimation.
  - Adapted using knowledge transfer.
  - Information related to the heart rate → Real or Fake.

- Trained and tested with 2nd generation DeepFake DBs:
  - **DFDC Preview.**
  - **Celeb-DF v2.**

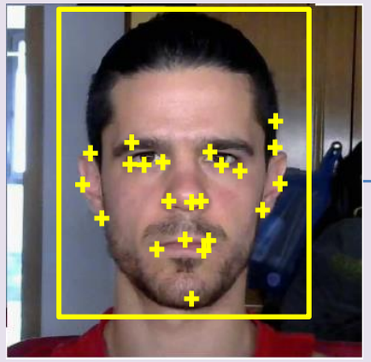
DeepFakesON-Phys → solution to the weaknesses of detectors based on the visual artifacts and fingerprints inserted during the synthesis process.

Proposed Framework



Preprocessing

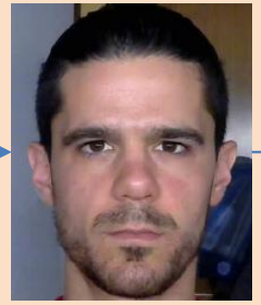
1. Face Detection & Tracking



MTCNN Face Detector
&
KLT Feature Tracker

2.1 Normalized Frame

Face Frame
F(t)



Normalized Frame
***I(t)*:**

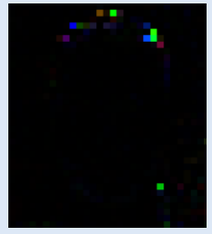
$$I(t)=(F(t+1) - F(t)) \ / \ (F(t+1) + F(t))$$

2.2 Normalized Frame Difference

Normalized Frames

I(t)

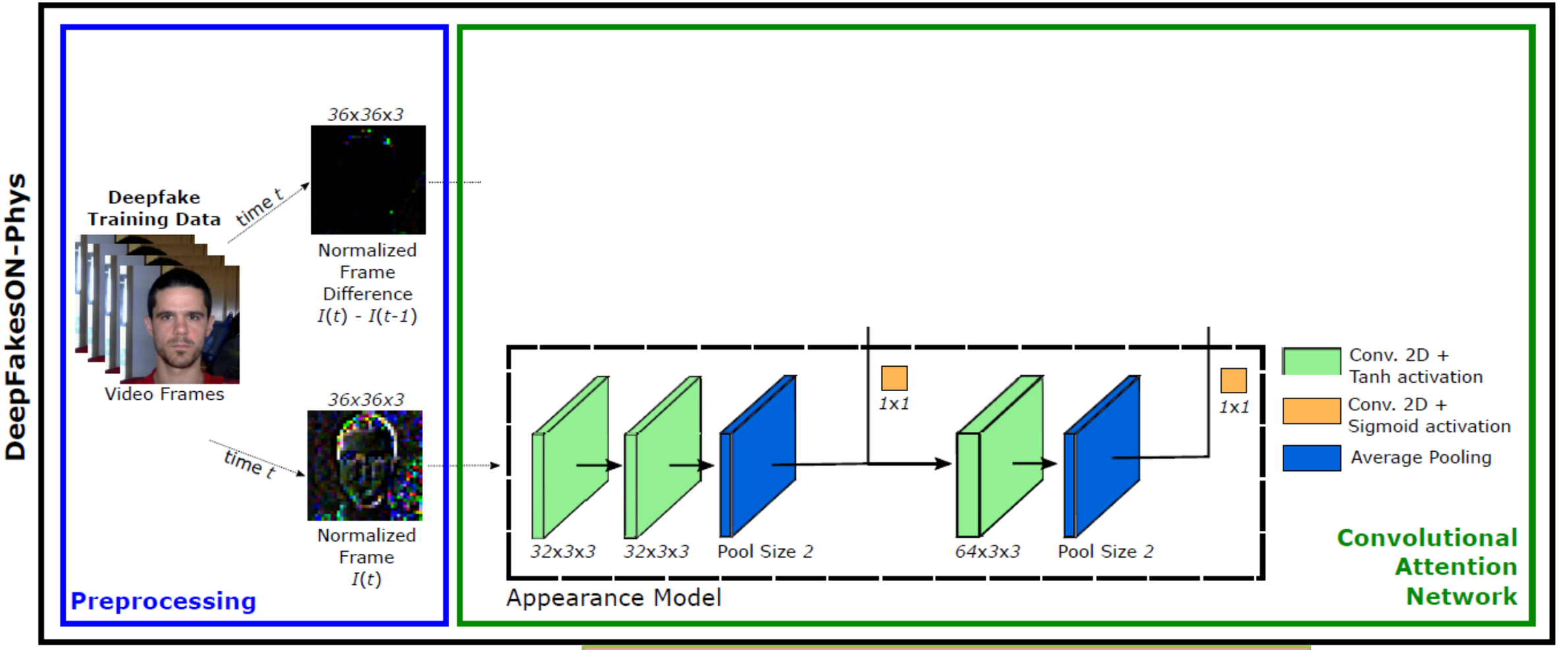
I(t-1)



I(t) – I(t-1)

Normalized Frame Difference

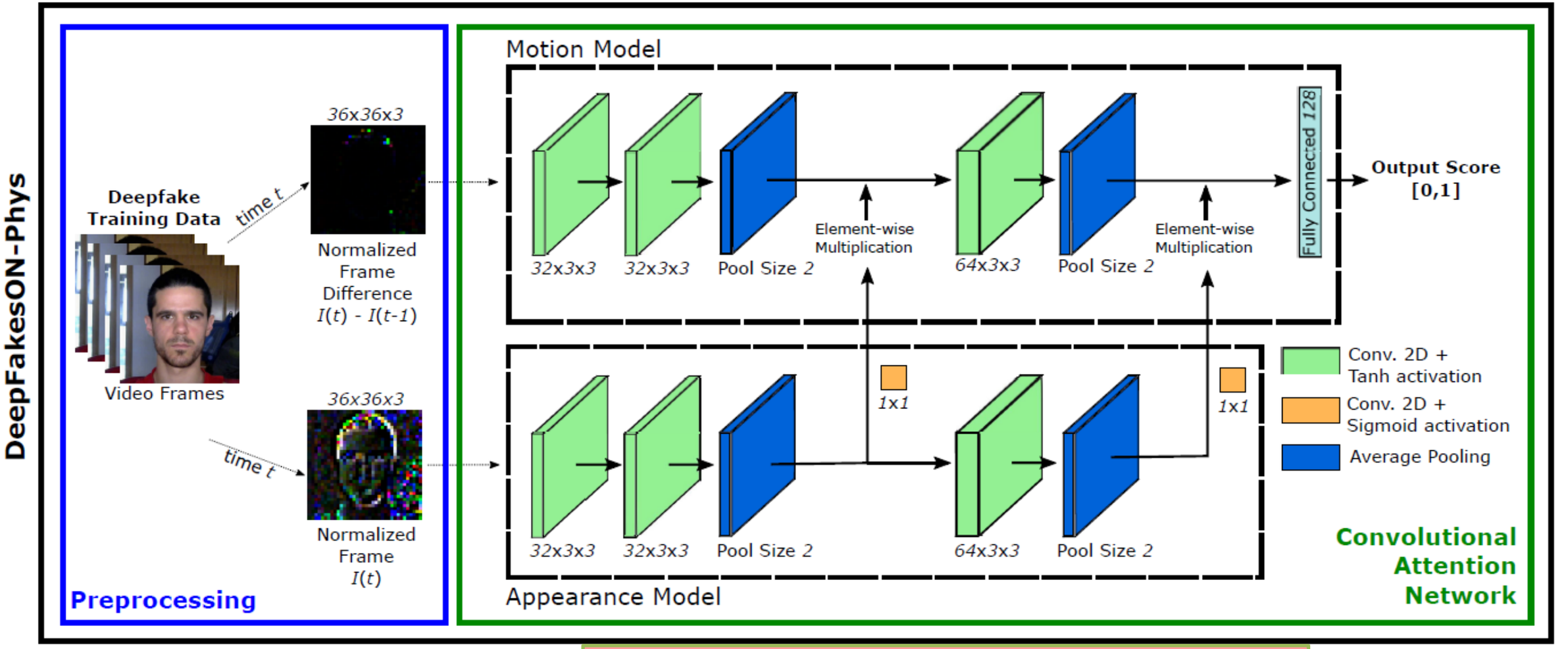
Proposed Framework



Appearance model: static information → Attention

Proposed Framework

Motion model: temporal information + attention



Appearance model: static information → Attention

Databases – 2nd Generation

Celeb-DF v2 [9]

- 590 real (Youtube)
- 5,639 fake videos (Deep Learning)

DFDC Preview [10]

- 1,131 real (Actors)
- 4,139 fake videos (Various)

[9] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu. 2020. "Celeb-DF: A LargeScale Challenging Dataset for DeepFake Forensics". In *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR).*

[10] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer. 2019. "The Deepfake Detection Challenge (DFDC) Preview Dataset". *arXiv preprint.:1910.08854.*

DeepFakesON-Phys: Development

1) Model based on DeepPhys [11] (Heart rate from facial video) → Not public.

[11] W. Chen, and D. McDuff. 2018. **"Deepphys: Video-based Physiological Measurement using Convolutional Attention Networks".** In *Procs. of the European Conf. on Computer Vision (ECCV).*

DeepFakesON-Phys: Development

1) Model based on DeepPhys [11] (Heart rate from facial video) → Not public.
2) Own implementation trained with COHFACE DB [12].

[11] W. Chen, and D. McDuff. 2018. "**Deepphys: Video-based Physiological Measurement using Convolutional Attention Networks**". In *Procs. of the European Conf. on Computer Vision (ECCV).*
[12] J. Hernandez-Ortega, *et al.* 2020. "**A Comparative Evaluation of Heart Rate Estimation Methods using Face Videos**". In *Procs. of the Computers, Software, and Applications Conf. (COMPSAC).*

DeepFakesON-Phys: Development

1) Model based on DeepPhys [11] (Heart rate from facial video) → Not public.
2) Own implementation trained with COHFACE DB [12].
3) Celeb-DF v2 and DFDC Preview split into 2 non-overlapping datasets: dev. and eval.

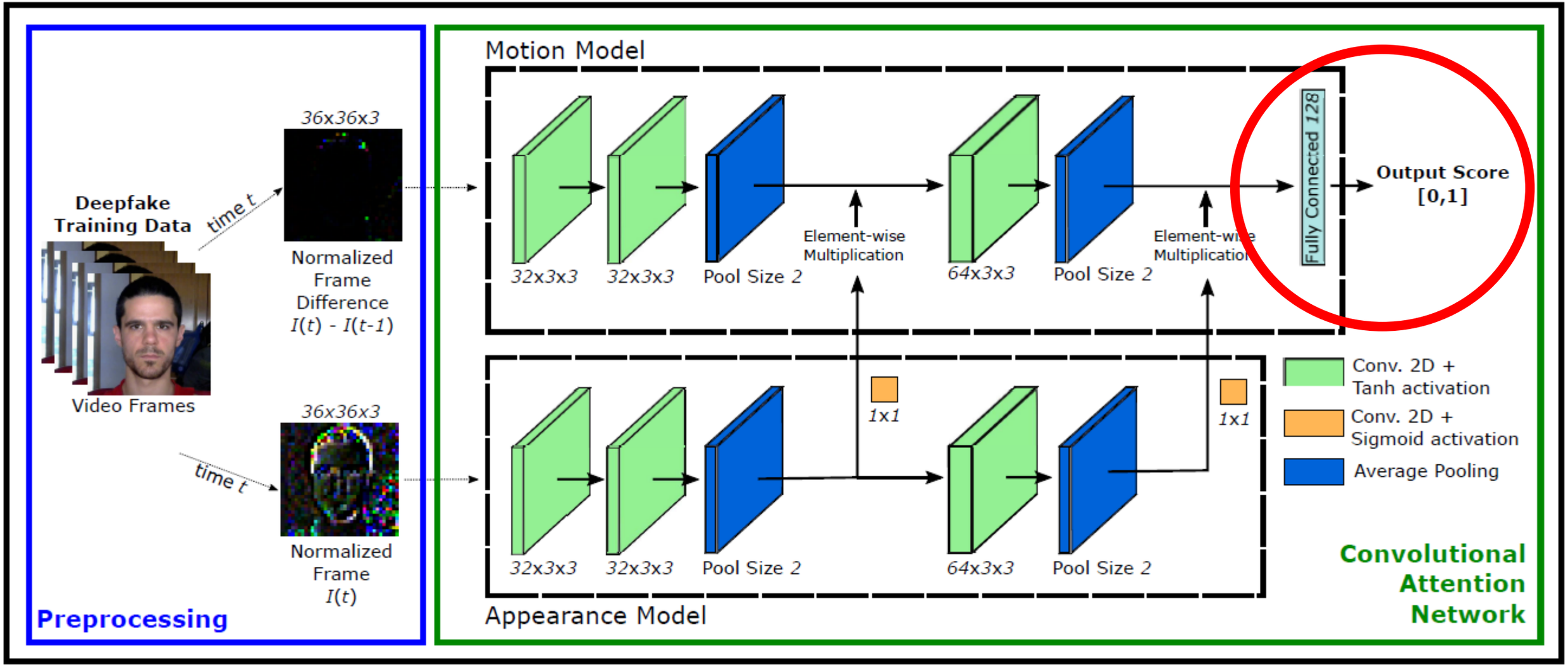
[11] W. Chen, and D. McDuff. 2018. **"Deepphys: Video-based Physiological Measurement using Convolutional Attention Networks".** In *Procs. of the European Conf. on Computer Vision (ECCV).*
[12] J. Hernandez-Ortega, *et al.* 2020. **"A Comparative Evaluation of Heart Rate Estimation Methods using Face Videos".** In *Procs. of the Computers, Software, and Applications Conf. (COMPSAC).*

DeepFakesON-Phys: Development

1) Model based on DeepPhys [11] (Heart rate from facial video) → Not public.
2) Own implementation trained with COHFACE DB [12].
3) Celeb-DF v2 and DFDC Preview split into 2 non-overlapping datasets: dev. and eval.
4) Changed the last FC and the output layers of the former model (two classes, real or fake).

[11] W. Chen, and D. McDuff. 2018. "**Deepphys: Video-based Physiological Measurement using Convolutional Attention Networks**". In *Procs. of the European Conf. on Computer Vision (ECCV).*
[12] J. Hernandez-Ortega, *et al.* 2020. "**A Comparative Evaluation of Heart Rate Estimation Methods using Face Videos**". In *Procs. of the Computers, Software, and Applications Conf. (COMPSAC).*

DeepFakesON-Phys: Development and Evaluation



DeepFakesON-Phys: Development

1) Model based on DeepPhys [11] (Heart rate from facial video) → Not public.
2) Own implementation trained with COHFACE DB [12].
3) Celeb-DF v2 and DFDC Preview split into 2 non-overlapping datasets: dev. and eval.
4) Changed the last FC and the output layers of the former model (two classes, real or fake).
5) Fixed all weights up to the final fully-connected layer.

[11] W. Chen, and D. McDuff. 2018. "**Deepphys: Video-based Physiological Measurement using Convolutional Attention Networks**". In *Procs. of the European Conf. on Computer Vision (ECCV).*
[12] J. Hernandez-Ortega, et al. 2020. "**A Comparative Evaluation of Heart Rate Estimation Methods using Face Videos**". In *Procs. of the Computers, Software, and Applications Conf. (COMPSAC).*

DeepFakesON-Phys: Development

1) Model based on DeepPhys [11] (Heart rate from facial video) → Not public.
2) Own implementation trained with COHFACE DB [12].
3) Celeb-DF v2 and DFDC Preview split into 2 non-overlapping datasets: dev. and eval.
4) Changed the last FC and the output layers of the former model (two classes, real or fake).
5) Fixed all weights up to the final fully-connected layer.
6) Trained the network for 100 more epochs and choose the best performing model based on validation accuracy.
    - One model per training database.

[11] W. Chen, and D. McDuff. 2018. "**Deepphys: Video-based Physiological Measurement using Convolutional Attention Networks**". In *Procs. of the European Conf. on Computer Vision (ECCV).*
[12] J. Hernandez-Ortega, *et al.* 2020. "**A Comparative Evaluation of Heart Rate Estimation Methods using Face Videos**". In *Procs. of the Computers, Software, and Applications Conf. (COMPSAC).*

Experimental Results

Evaluation Metrics → Area Under the Curve (AUC) and Accuracy (Frame level).

Celeb-DF v2

Study	Method	Classifier	AUC (%)
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 [Ours]	Physiological	CNN + Attention	99.9 Acc. = 98.7

Experimental Results

Evaluation Metrics → Area Under the Curve (AUC) and Accuracy (Frame level).

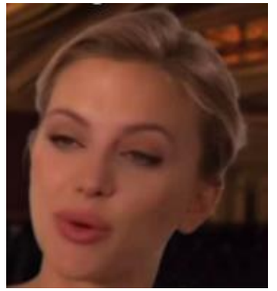
DFDC Preview

Study	Method	Classifier	AUC (%)
Yang, Li, and Lyu 2019	Head Pose	SVM	55.9
Li *et al*. 2020	Face Warping	CNN	75.5
Afchar *et al*. 2018	Mesoscopic	CNN	75.3
Dang *et al*. 2020	Deep Learning	CNN + Attention	-
Tolosana *et al*. 2020a	Deep Learning	CNN	91.1
Qi *et al*. 2020	Physiological	CNN + Attention	Acc. = 64.1
Ciftci, Demir, and Yin 2020	Physiological	SVM/CNN	-
DeepFakesON-Phys [Ours]	Physiological	CNN + Attention	98.2 Acc. = 94.4

Conclusions

Two 2nd generation DeepFake databases

- Celeb-DF v2



- DFDC Preview

Two of the latest and most challenging DeepFake video databases.

DeepFakesON-Phys:

Outperformed other **state-of-the-art fake detectors** based on face warping and pure deep learning features, among others.

Revealed that **current DeepFake techniques do not pay attention to** the heart-rate-related or blood-related **physiological information**.

Future Work

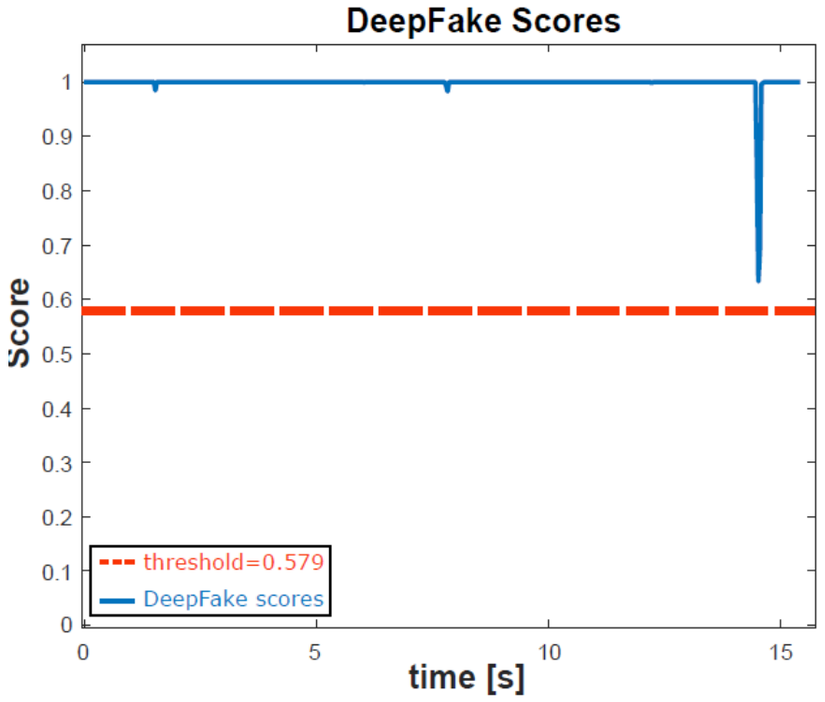
1. Analysis of the **robustness against unseen face manipulations** [13].

[13] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., and Ortega-Garcia, J. 2020. "**DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection**". *Information Fusion* 64: 131–148.

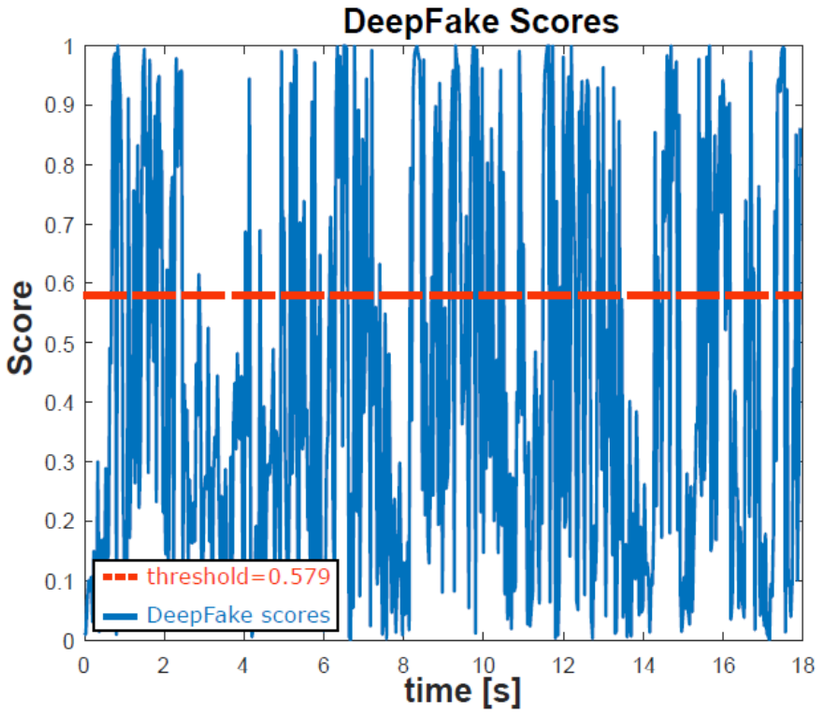
Future Work

1. Analysis of the **robustness against unseen face manipulations** [13].
2. **Applying temporal integration** to frame data [14].

Real Video



Fake Video



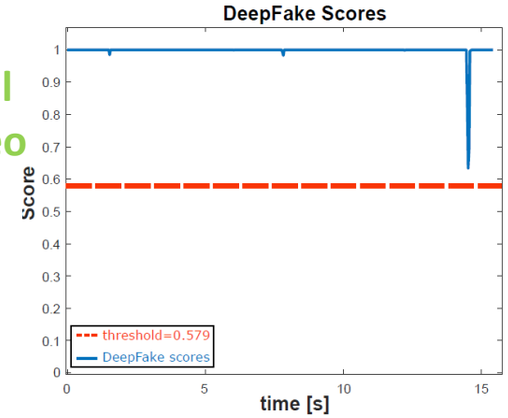
[13] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., and Ortega-Garcia, J. 2020. "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection". *Information Fusion* 64: 131–148.

[14] Hernandez-Ortega, J., Fierrez, J., Morales, A., and Tome, P. 2018. "Time Analysis of Pulse-Based Face Anti-Spoofing in Visible and NIR". In *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition Workshops (CVPRw).*

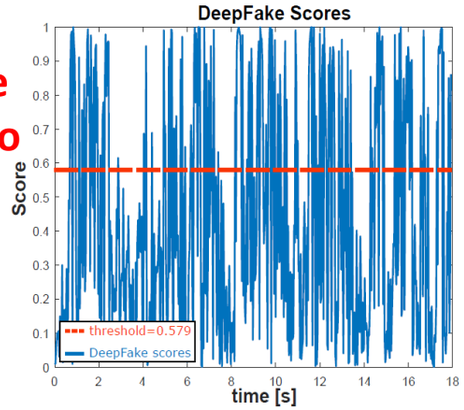
Future Work

1. Analysis of the **robustness against unseen face manipulations** [13].
2. **Applying temporal integration** to frame data [14].

Real Video



Fake Video



3. **Studying other face manipulation techniques**, e.g. face morphing [15].

[13] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., and Ortega-Garcia, J. 2020. "**DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection**". *Information Fusion* 64: 131–148.

[14] Hernandez-Ortega, J., Fierrez, J., Morales, A., and Tome, P. 2018. "**Time Analysis of Pulse-Based Face Anti-Spoofing in Visible and NIR**". In *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition Workshops (CVPRw).*

[15] Raja, K., *et al.* 2020. "**Morphing Attack Detection - Database, Evaluation Platform and Benchmarking**". *IEEE Transactions on Information Forensics and Security.*

BiDA Lab

Biometrics & Data Pattern Analytics Lab

UAM

Universidad Autónoma de Madrid

Know More:

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales and J. Ortega-Garcia, "**DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection**", *Information Fusion*, 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*, 2020.

J. Hernandez-Ortega, *et al*. "**Time Analysis of Pulse-based Face Anti-spoofing in Visible and NIR**". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018.

J. Hernandez-Ortega, *et al*. "**Introduction to Face Presentation Attack Detection**." *Handbook of Biometric Anti-Spoofing*. Springer. 2019.

http://biometrics.eps.uam.es

Funding: This work has been supported by projects: IDEA-FAST (IMI2-2018-15-two-stage-853981), PRIMA (ITN-2019-860315), TRESPASS-ETN (ITN-2019-860813), BIBECA (RTI2018-101248-B-I00 MINECO/FEDER), and edBB (Universidad Autonoma de Madrid, UAM). J. H.-O. is supported by a PhD fellowship from UAM. R. T. is supported by a Postdoctoral fellowship from CAM/FSE.

TReSPAsS-ETN

Privacy Matters

European Commission

GOBIERNO DE ESPAÑA