DeepFakesON-Phys: DeepFakes Detection based on Heart Rate Estimation

Javier Hernandez-Ortega, Ruben Tolosana, Julian Fierrez and Aythami Morales

Biometrics and Data Pattern Analytics - BiDA Lab, Universidad Autonoma de Madrid

{javier.hernandezo, ruben.tolosana, julian.fierrez, aythami.morales}@uam.es

Abstract—This work introduces a novel DeepFake detection framework based on physiological measurement. In particular, we consider information related to the heart rate using remote photoplethysmography (rPPG). rPPG methods analyze video sequences looking for subtle color changes in the human skin, revealing the presence of human blood under the tissues. In this work we investigate to what extent rPPG is useful for the detection of DeepFake videos.

The proposed fake detector named DeepFakesON-Phys uses a Convolutional Attention Network (CAN), which extracts spatial and temporal information from video frames, analyzing and combining both sources to better detect fake videos. This detection approach has been experimentally evaluated using the latest public databases in the field: Celeb-DF and DFDC. The results achieved, above 98% AUC (Area Under the Curve) on both databases, outperform the state of the art and prove the success of fake detectors based on physiological measurement to detect the latest DeepFake videos.

Index Terms—Fake News, DeepFakes, Media Forensics, Face Manipulation, Fake Detection, Heart Rate, rPPG

I. INTRODUCTION

DeepFakes have become a great public concern recently [1], [2]. The very popular term "DeepFake" is usually 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. This type of digital manipulation is also known in the literature as Identity Swap, and it is moving forward very fast [3], [4].

Currently, most face manipulations are based on popular machine learning techniques such as AutoEncoders (AE) [5] and Generative Adversarial Networks (GAN) [6], achieving in general very realistic visual results, specially in the latest generation of public DeepFakes [7], and the present trends [8], [9]. However, and despite the impressive visual results, are current face manipulations also considering the physiological aspects of the human being in the synthesis process?

Physiological measurement has provided very valuable information to many different tasks such as e-learning [10], health care [11], human-computer interaction [12], and security [13], among many other tasks.

In physical face attacks, a.k.a. Presentation Attacks (PAs), real subjects are often impersonated using artifacts such as photographs, videos, and masks [13], [14]. Face recognition systems are known to be vulnerable against these attacks unless proper detection methods are implemented [15], [16]. Some of these detection methods are based on liveness detection by using information such as eye blinking [17] or natural facial micro-expressions [18]. Specifically for detecting 3D mask impersonation, which is one of the most challenging type of attacks, detecting pulse from face videos using remote photoplethysmography (rPPG) has shown to be an effective countermeasure [19]. When applying this technique to a video sequence with a fake face, the estimated heart rate signal is significantly different to the heart rate extracted from a real face [20].

Seeing the good results achieved by rPPG techniques when dealing with physical 3D face mask attacks, and since Deep-Fakes are digital manipulations somehow similar to them, in this work we hypothesize that fake detectors based on physiological measurement can also be used against DeepFakes after adapting them properly.

The present work proposes a novel DeepFake detector based on physiological measurement named DeepFakesON-Phys. In particular, the information related to the heart rate is considered to decide whether a video is real or fake. Our physiological detector intends to be a robust solution to the weaknesses of most state-of-the-art DeepFake detectors based on the visual features existing in fake videos [21], [22] and also on the artifacts/fingerprints inserted during the synthesis process [23], [24], which are highly dependent on a specific fake manipulation technique.

In this context, the main contributions of our work are:

- An approach based on physiological measurement to detect DeepFake videos: DeepFakesON-Phys[1]. Fig. 1 graphically summarizes the proposed fake detection approach based on a Convolutional Attention Network (CAN), composed of two parallel Convolutional Neural Networks (CNN) able to extract spatial and temporal information from video frames.
- An in-depth literature review of DeepFake detection approaches with special emphasis to physiological techniques, including the key aspects of the detection systems, the databases used, and the main results achieved.
- A thorough experimental assessment of the proposed DeepFake detector, considering the latest public databases of the 2nd DeepFake generation [3] such as Celeb-DF v2 and DFDC Preview.
- DeepFakesON-Phys achieves high-accuracy results, outperforming the state of the art on both Celeb-DF and DFDC databases.

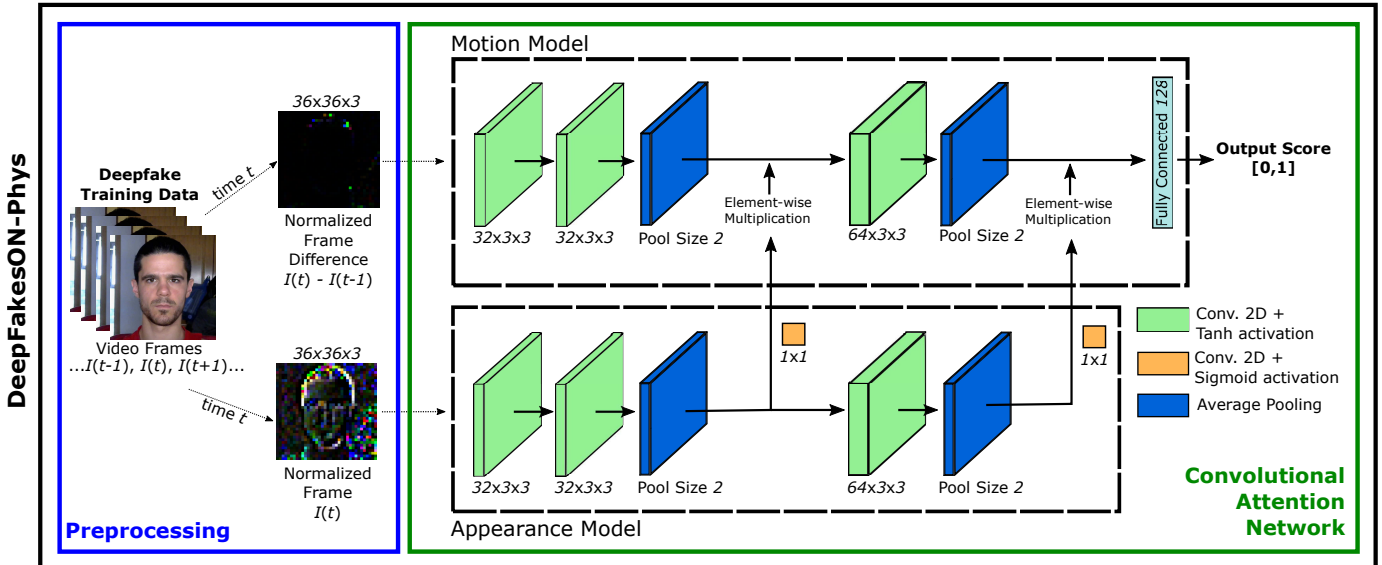[1]The code will be available in GitHub soon.

Fig. 1. **DeepFakesON-Phys architecture**. It comprises two stages: *i)* a preprocessing step to normalize the video frames, and *ii)* a Convolutional Attention Network composed of Motion and Appearance Models to better detect fake videos.

The remainder of the paper is organized as follows. Sec. II summarizes previous studies focused on the detection of Deep-Fakes. Sec. III describes the proposed DeepFakesON-Phys fake detection approach. Sec. IV summarizes all databases considered in the experimental framework of this study. Sec. V describes the experimental protocol and the results achieved in comparison with the state of the art. Finally, Sec. VI draws the final conclusions and points out future research lines.

II. RELATED WORKS

Different approaches have been proposed in the literature to detect DeepFake videos. Table I shows a comparison of the most relevant approaches in the area, paying special attention to the fake detectors based on physiological measurement. For each study we include information related to the method, classifiers, best performance, and databases for research. It is important to remark that in some cases, different evaluation metrics are considered, e.g., Area Under the Curve (AUC) and Equal Error Rate (EER), which complicate the comparison among studies. Finally, the results highlighted in *italics* indicate the generalization ability of the detectors against unseen databases, i.e., those databases were not considered for training. Most of these results are extracted from [25].

The first studies in the area focused on the visual artifacts existed in the 1st generation of fake videos. Matern *et al.* proposed in [21] fake detectors based on simple visual artifacts such as eye colour, missing reflections, and missing details in the teeth areas, achieving a final 85.1% AUC.

Approaches based on the detection of the face warping artifacts have also been studied in the literature. Li *et al.* proposed in [25], [26] detection systems based on CNN in order to detect the presence of such artifacts from the face and the surrounding areas, being one of the most robust detection approaches against unseen face manipulations.

Undoubtedly, fake detectors based on pure deep learning features are the most popular ones: feeding the networks with as many real/fake videos as possible and letting the networks to automatically extract the discriminative features. In general, these fake detectors have achieved very good results using popular network architectures such as Xception [27], [30], novel ones such as Capsule Networks [28], and novel training techniques based on attention mechanisms [29].

Fake detectors based on the image and temporal discrepancies across frames have also been proposed in the literature. Sabir *et al.* proposed in [31] a Recurrent Convolutional Network similar to [37], trained end-to-end instead of using a pre-trained model. Their proposed detection approach was tested using FaceForensics++ database [27], achieving AUC results above 96%.

Although most approaches are based on the detection of fake videos using the whole face, Tolosana *et al.* evaluated in [7] the discriminative power of each facial region using state-of-the-art network architectures, achieving interesting results on DeepFake databases of the 1st and 2nd generations.

Finally, we pay special attention to the fake detectors based on physiological information. The eye blinking rate was studied in [33], [35]. In [33], Li *et al.* proposed Long-Term Recurrent Convolutional Networks (LRCN) to capture the temporal dependencies existed in human eye blinking. Their method was evaluated on the UADFV database, achieving a final 99.0% AUC. More recently, Jung *et al.* proposed a different approach named DeepVision. Their proposed approach fused the Fast-HyperFace [38] and EAR [39] algorithms to track the blinking, achieving an accuracy of 87.5% over an in-house database.

Fake detectors based on the analysis of the way we speak were studied by Agarwal *et al.* in [22], focusing on the distinct facial expressions and movements. These features

TABLE I

COMPARISON OF DIFFERENT STATE-OF-THE-ART FAKE DETECTORS. RESULTS IN *italics* INDICATE THE GENERALIZATION CAPACITY OF THE DETECTORS AGAINST UNSEEN DATABASES. FF++ = FACEFORENSICS++, AUC = AREA UNDER THE CURVE, ACC. = ACCURACY, EER = EQUAL ERROR RATE.

Study	Method	Classifiers	Best Performance	Databases
Matern *et al.* (2019) [21]	Visual Features	Logistic Regression MLP	AUC = 85.1%	Own
			AUC = 78.0%	*FF++ / DFD*
			AUC = 66.2%	*DFDC Preview*
			AUC = 55.1%	*Celeb-DF*
Li *et al.* (2019) [25], [26]	Face Warping Features	CNN	AUC = 97.7%	UADFV
			AUC = 93.0%	*FF++ / DFD*
			AUC = 75.5%	*DFDC Preview*
			AUC = 64.6%	*Celeb-DF*
Rössler *et al.* (2019) [27]	Mesoscopic Features Steganalysis Features Deep Learning Features	CNN	Acc. $\simeq$ 94.0%	FF++ (DeepFake, LQ)
			Acc. $\simeq$ 98.0%	FF++ (DeepFake, HQ)
			Acc. $\simeq$ 100.0%	FF++ (DeepFake, RAW)
			Acc. $\simeq$ 93.0%	FF++ (FaceSwap, LQ)
			Acc. $\simeq$ 97.0%	FF++ (FaceSwap, HQ)
			Acc. $\simeq$ 99.0%	FF++ (FaceSwap, RAW)
Nguyen *et al.* (2019) [28]	Deep Learning Features	Capsule Networks	*AUC = 61.3%*	*UADFV*
			AUC = 96.6%	*FF++ / DFD*
			AUC = 53.3%	*DFDC Preview*
			AUC = 57.5%	*Celeb-DF*
Dang *et al.* (2019) [29]	Deep Learning Features	CNN + Attention Mechanism	AUC = 99.4% EER = 3.1%	DFFD
Dolhansky *et al.* (2019) [30]	Deep Learning Features	CNN	Precision = 93.0% Recall = 8.4%	DFDC Preview
Sabir *et al.* (2019) [31]	Image + Temporal Features	CNN + RNN	AUC = 96.9%	FF++ (DeepFake, LQ)
			AUC = 96.3%	FF++ (FaceSwap, LQ)
Tolosana *et al.* (2020) [7]	Facial Regions Features	CNN	AUC = 100.0%	UADFV
			AUC = 99.5%	FF++ (FaceSwap, HQ)
			AUC = 91.1%	DFDC Preview
			AUC = 83.6%	Celeb-DF
Conotter *et al.* (2014) [32]	Physiological Features	-	Acc. = 100%	Own
Li *et al.* (2018) [33]	Physiological Features	LRCN	AUC = 99.0%	UADFV
Agarwal and Farid (2019) [22]	Physiological Features	SVM	AUC = 96.3%	Own (FaceSwap, HQ)
Ciftci *et al.* (2020) [34]	Physiological Features	SVM/CNN	Acc. = 94.9% Acc. = 91.5%	FF++ (DeepFakes) Celeb-DF
Jung *et al.* (2020) [35]	Physiological Features	Distance	Acc. = 87.5%	Own
Qi *et al.* (2020) [36]	Physiological Features	CNN + Attention Mechanism	Acc. = 100.0%	FF++ (FaceSwap)
			Acc. = 100.0%	FF++ (DeepFake)
			Acc. = 64.1%	*DFDC Preview*
DeepFakesON-Phys [Proposed Approach]	**Physiological Features**	**CAN**	**AUC = 99.9%**	**Celeb-DF v2**
			AUC = 98.2%	**DFDC Preview**

were considered in combination with Support Vector Machines (SVM), achieving a 96.3% AUC over their own database.

Finally, fake detection methods based on the heart rate have been also studied in the literature. One of the first studies in this regard was [32]. In that study, Conotter *et al.* preliminary evaluated the potential of blood flow changes in the face to distinguish between computer generated and real videos. Their proposed approach was evaluated using 12 videos (6 real and fake videos each), concluding that it is possible to use this metric to detect computer generated videos.

Changes in the blood flow have also been studied in [34], [36] using DeepFake videos. In [34], the authors considered rPPG techniques to extract robust biological features. Classifiers based on SVM and CNN were analyzed, achieving final accuracies of 94.9% and 91.5% for the DeepFakes videos of FaceForensics++ and Celeb-DF, respectively.

Recently, Qi *et al.* developed in [36] a more sophisticated fake detector named DeepRhythm. Their approach was also based on features extracted using rPPG techniques. DeepRhythm was enhanced through two modules: *i)* motion-magnified spatial-temporal representation, and *ii)* dual-spatial-temporal attention. These modules were incorporated in order to provide a better adaptation to dynamically changing faces and various fake types. In general, good results with accuracies of 100% were achieved on FaceForensics++ database. However, this method suffers from a demanding preprocessing stage, needing a precise detection of 81 facial landmarks and the use of a color magnification algorithm prior to fake detection. Also, poor results were achieved on databases of the 2nd generation such as the DFDC Preview (Acc. = 64.1%).

In the present work, in addition to the proposal of a different DeepFake detection architecture, we enhance previous

approaches keeping the preprocessing stage as light and robust as possible, only composed of a face detector and frame normalization. To provide an overall picture, we include in Table I the results achieved with our proposed DeepFakesON-Phys in comparison with key related works, which shows that we outperform the state of the art on Celeb-DF v2 and DFDC Preview databases.

III. PROPOSED METHOD: DEEPFAKESON-PHYS

Fig. 1 graphically summarizes the architecture of DeepFakesON-Phys, the proposed fake detector based on heart rate estimation. We hypothesize that rPPG methods should obtain significantly different results when trying to estimate the subjacent heart rate from a video containing a real face, compared with a fake face. Since the changes in color and illumination due to oxygen concentration are subtle and invisible to the human eye, we think that most of the existing DeepFake manipulation methods do not consider the physiological aspects of the human being yet.

The initial architecture of the proposed fake detector is based on the DeepPhys model described in [40], whose objective was to estimate the human heart rate using facial video sequences. The model is based on deep learning and was designed to extract spatio-temporal information from videos mimicking the behavior of traditional handcrafted rPPG techniques. Features are extracted through the color changes in users' faces that are caused by the variation of oxygen concentration in the blood. Signal processing methods are also used for isolating the color changes caused by blood from other changes that may be caused by factors such as external illumination, noise, etc.

As can be seen in Fig. 1, after the first preprocessing stage, the Convolutional Attention Network (CAN) is composed of two different CNN branches:

- **Motion Model**: it is designed to detect changes between consecutive frames, i.e., performing a short-time analysis of the video for detecting fakes. To accomplish this task, the input at a time t consists of a frame computed as the normalized difference of the current frame $I(t)$ and the previous one $I(t-1)$.
- **Appearance Model**: it focuses on the analysis of the static information on each video frame. It has the target of providing the Motion Model with information about which points of the current frame may contain the most relevant information for detecting DeepFakes, i.e., a batch of attention masks that are shared at different layers of the CNN. The input of this branch at time t is the raw frame of the video $I(t)$, normalized to zero mean and unitary standard deviation.

The attention masks coming from the Appearance Model are shared with the Motion Model at two different points of the CAN. Finally, the output layer of the Motion Model is also the final output of the entire CAN.

In the original architecture [40], the output stage consisted of a regression layer for estimating the time derivative of the subject's heart rate. In our case, as we do not aim to estimate

TABLE II
IDENTITY SWAP PUBLICLY AVAILABLE DATABASES OF THE 2ND GENERATION CONSIDERED IN OUR EXPERIMENTAL FRAMEWORK.

2nd Generation		
Database	Real Videos	Fake Videos
Celeb-DF v2 (2020) [25]	590 (Youtube)	5,639 (DeepFake)
DFDC Preview (2019) [30]	1,131 (Actors)	4,119 (Unknown)

the pulse of the subject, but the presence of a fake face, so we change the final regression layer with a classification layer, using a sigmoid activation function for obtaining a final score in the $[0,1]$ range for each instant t of the video, related to the probability of the face being real.

We initialize the CAN parameters with the weights from the model pretrained for heart rate estimation in [41], instead of training a new CAN from scratch. Then, we freeze the weights of all the layers of the original CAN model apart from the new classification layer and the last fully-connected layer, and we retrain the model. Due to this transfer-learning process we take benefit of the weights learned for heart rate estimation, just adapting them for the DeepFake detection task. This way, we make sure that the weights of the convolutional layers remain looking for information relative to heart rate and the last layers learn how to use that information for detecting the existence of DeepFakes.

IV. DATABASES

Two different public databases are considered in the experimental framework of this study. In particular, Celeb-DF v2 and DFDC Preview, the two most challenging DeepFake databases up to date. Table II summarizes their main features.

A. Celeb-DF v2

The aim of the Celeb-DF v2 database [25] was to generate fake videos of better visual quality compared with the previous UADFV database. This database consists of 590 real videos extracted from Youtube, corresponding to celebrities with a diverse distribution in terms of gender, age, and ethnic group. In addition, these videos exhibit a large range of variations in aspects such as the face sizes (in pixels), orientations, lighting conditions, and backgrounds. Regarding fake videos, a total of 5,639 videos were created swapping faces using DeepFake technology. The final videos are in MPEG4.0 format.

B. DFDC Preview

The DFDC database [30] is one of the latest public databases, released by Facebook in collaboration with other companies and academic institutions such as Microsoft, Amazon, and the MIT. In the present study we consider the DFDC Preview dataset consisting of 1,131 real videos from 66 paid actors, ensuring realistic variability in gender, skin tone, and
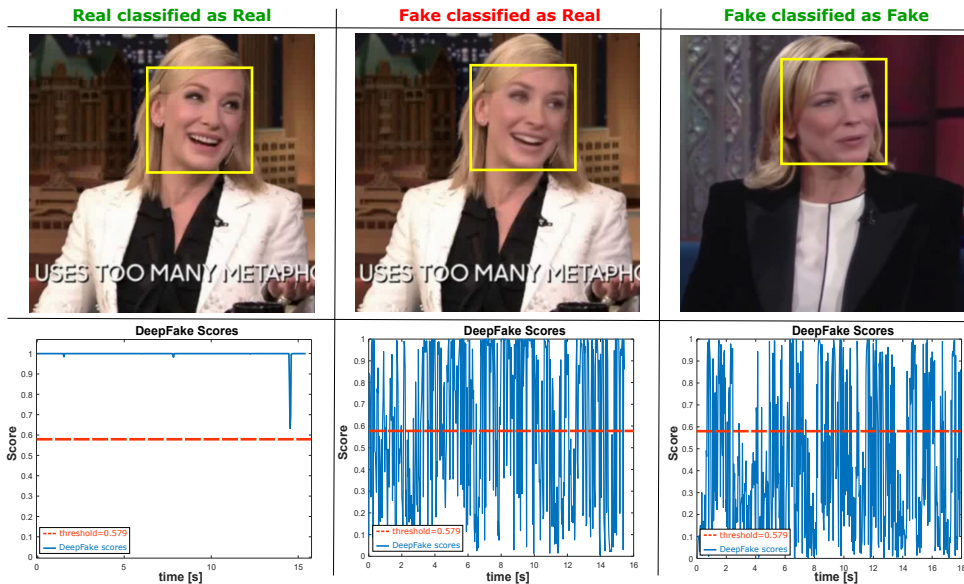
Fig. 2. **Examples of successful and failed DeepFake detections**. Top: sample frames of evaluated videos. Bottom: score distribution for each sample video. For the fake video misclassified as containing a real face, the DeepFake detection scores present a higher mean compared to the case of the fake video correctly classified as a fake.

age. It is important to remark that no publicly available data or data from social media sites were used to create this dataset, unlike other popular databases. Regarding fake videos, a total of 4,119 videos were created using two different unknown approaches for fakes generation. Fake videos were generated by swapping subjects with similar appearances, i.e., similar facial attributes such as skin tone, facial hair, glasses, etc. After a given pairwise model was trained on two identities, the identities were swapped onto the other's videos.

It is important to highlight that the DFDC database considers different acquisition scenarios (i.e., indoors and outdoors), light conditions (i.e., day, night, etc.), distances from the person to the camera, and pose variations, among others.

V. EXPERIMENTS

A. Experimental Protocol

Celeb-DF v2 and DFDC Preview databases have been divided into non-overlapping datasets, development and evaluation. It is important to remark that each dataset comprises videos from different identities (both real and fake), unlike some previous studies. This aspect is very important in order to perform a fair evaluation and predict the generalization ability of the fake detection systems against unseen identities.

For the Celeb-DF v2 database, we consider real/fake videos of 40 and 19 different identities for the development and evaluation datasets respectively, whereas for the DFDC Preview database, we follow the same experimental protocol proposed in [30] as the authors already considered this concern.

In addition, it is important to highlight that the evaluation is carried out at frame level as in most previous studies [3], not video level, using the popular AUC and accuracy metrics.

TABLE III
FAKE DETECTION PERFORMANCE RESULTS IN TERMS OF AUC AND ACCURACY OVER THE FINAL EVALUATION DATASETS.

Database	AUC Results (%)	Acc. Results (%)
Celeb-DF v2	99.9	98.7
DFDC Preview	98.2	94.4

B. Results: Fake Detection with DeepFakesON-Phys

This section evaluates the ability of DeepFakesON-Phys to detect the most challenging DeepFake videos of the 2nd generation. Table III shows the fake detection performance results achieved in terms of AUC and accuracy over the final evaluation datasets of Celeb-DF v2 and DFDC Preview. It is important to highlight that a separate fake detector is trained for each database.

In general, very good results are achieved in both DeepFake databases. For the Celeb-DF v2 database, DeepFakesON-Phys achieves an accuracy of 98.7% and an AUC of 99.9%. Regarding the DFDC Preview database, the results achieved are 94.4% accuracy and 98.2% AUC, similar ones to the obtained for the Celeb-DF database.

Observing the results, it seems clear that the fake detectors have learnt to distinguish the spatio-temporal differences between the real/fake faces of Celeb-DF v2 and DFDC Preview databases. Since all the convolutional layers of the proposed fake detector are frozen (the network was originally initialized with the weights from the model trained to predict the heart rate [41]), and we only train the last fully-connected layers, we can conclude that the proposed detection approach based on physiological measurement is succesfully using pulse-related features for distinguishing between real and fake faces. These results prove that current face manipulation techniques do not

TABLE IV
COMPARISON OF DIFFERENT STATE-OF-THE-ART FAKE DETECTORS WITH OUR PROPOSED DEEPFAKESON-PHYS. THE BEST RESULTS ACHIEVED FOR EACH DATABASE ARE REMARKED IN **BOLD**. RESULTS IN *ITALICS* INDICATE THAT THE EVALUATED DATABASE WAS NOT USED FOR TRAINING.

Study	Method	Classifiers	AUC Results (%)	
			Celeb-DF [25]	**DFDC** [30]
Yang *et al.* (2019) [42]	Head Pose Features	SVM	*54.6*	*55.9*
Li *et al.* (2020) [25]	Face Warping Features	CNN	*64.6*	*75.5*
Afchar *et al.* (2018) [43]	Mesoscopic Features	CNN	*54.8*	*75.3*
Dang *et al.* (2020) [29]	Deep Learning Features	CNN + Attention Mechanism	*71.2*	-
Tolosana *et al.* (2020) [7]	Deep Learning Features	CNN	83.6	91.1
Qi *et al.* (2020) [36]	Physiological Features	CNN + Attention Mechanism	-	*Acc. = 64.1*
Ciftci *et al.* (2020) [34]	Physiological Features	SVM/CNN	Acc. = 91.5	-
DeepFakesON-Phys [Present Paper]	**Physiological Features**	**CNN + Attention Mechanism**	**AUC = 99.9** **Acc. = 98.7**	**AUC = 98.2** **Acc. = 94.4**

pay attention to the physiological information of the human being when synthesizing fake videos.

Fig. 2 shows some examples of successful and failed detections when evaluating the proposed approach with real/fake faces of Celeb-DF v2. In particular, all the failures correspond to fake faces generated from a particular video, misclassifying them as real faces. Fig. 2 shows a frame from the original real video (top-left), one from a misclassified fake video generated using that scenario (top-middle), and another from a fake video correctly classified as fake and generated using the same real and fake identities but from other source videos (top-right).

Looking at the score distributions along time of the three examples (Fig. 2, bottom), it can be seen that for the real face video (left) the scores are 1 for most of the time and always over the detection threshold. However, for the fake videos considered (middle and right), the score changes constantly, making the score of some fake frames to cross the detection threshold and consequently misclassifying them as real. Nevertheless, it is important to remark that these mistakes only happen if we analyze the results at frame level (traditional approach followed in the literature [3]). In case we consider an evaluation at video level, DeepFakesON-Phys would be able to detect fake videos by integrating the temporal information available in short-time segments, e.g., in a similar way as described in [19] for continuous face anti-spoofing.

We believe that the failures produced in this particular case are propitiated by the interferences of external illumination. rPPG methods that use handcrafted features are usually fragile against external artificial illumination in the frequency and power ranges of normal human heart rate, making difficult to distinguish those illumination changes from the color changes caused by blood perfusion. Anyway, the proposed physiological approach presented in this work is more robust to this kind of illumination perturbations due to its CAN training process.

C. Comparison with the State of the Art

Finally, we compared in Table IV the results achieved in the present work with other state-of-the-art DeepFake detection approaches: head pose variations [42], face warping artifacts [25], mesoscopic features [43], pure deep learning features [7], [29], and physiological features [34], [36]. The best results achieved for each database are remarked in **bold**.

Results in *italics* indicate that the evaluated database was not used for training. Some of these results are extracted from [25].

Note that the comparison in Table IV is not always under the same datasets and protocols, therefore it must be interpreted with care. Despite of that, it is patent that the proposed DeepFakesON-Phys has achieved state-of-the-art results in both Celeb-DF and DFDC Preview databases. In particular, it has further outperformed popular fake detectors based on pure deep learning approaches such as Xception and Capsule Networks [7] and also other recent physiological approaches based on SVM/CNN [34].

VI. CONCLUSIONS

This work has evaluated the potential of physiological measurement to detect DeepFake videos. In particular, we have proposed a novel DeepFake detector named DeepFakesON-Phys based on a Convolutional Attention Network (CAN) originally trained for heart rate estimation using remote photoplethysmography (rPPG). The proposed CAN approach consists of two parallel CNN networks that extract and share temporal and spatial information from video frames.

DeepFakesON-Phys has been evaluated using Celeb-DF v2 and DFDC Preview databases, two of the latest and most challenging DeepFake video databases. Regarding the experimental protocol, each database was divided into development and evaluation datasets, considering different identities in each dataset in order to perform a fair evaluation of the technology.

The soundness and competitiveness of DeepFakesON-Phys has been proven by the very good results achieved, AUC values of 99.9% and 98.2% for the Celeb-DF and DFDC databases, respectively. These results have outperformed other state-of-the-art fake detectors based on face warping and pure deep learning features, among others. Finally, the experimental results of this study reveal that current face manipulation techniques do not pay attention to the physiological information of the human being when synthesizing fake videos.

Future work will be oriented to the analysis of the robustness of the proposed fake detection approach against face manipulations unseen during the training process [3], temporal integration of frame data [19], and the application of the proposed physiological approach to other face manipulation techniques such as face morphing [44].

REFERENCES

[1] D. Citron, "How DeepFake Undermine Truth and Threaten Democracy," 2019. [Online]. Available: https://www.ted.com

[2] R. Cellan-Jones, "Deepfake Videos Double in Nine Months," 2019. [Online]. Available: https://www.bbc.com/news/technology-49961089

[3] 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*, vol. 64, pp. 131–148, 2020.

[4] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, pp. 910–932, 2020.

[5] D.P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *Proc. International Conference on Learning Representations*, 2013.

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

[7] R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance," *arXiv preprint arXiv: 2004.07532*, 2020.

[8] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in *Proc. IEEE/CVF Conf. on Computer Vision and Patter Recognition*, 2020.

[9] J. Naruniec, L. Helminger, C. Schroers and R.M. Weber, "High-Resolution Neural Face Swapping for Visual Effects," in *Proc. Eurographics Symposium on Rendering*, 2020.

[10] J. Hernandez-Ortega, R. Daza, A. Morales, J. Fierrez and R. Tolosana, "Heart Rate Estimation from Face Videos for Student Assessment: Experiments on edBB," in *Proc. IEEE Computer Software and Applications Conference*, 2020.

[11] D. J. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford, "A Survey of Remote Optical Photoplethysmographic Imaging Methods," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015.

[12] D. Tan and A. Nijholt, "Brain-Computer Interfaces and Human-Computer Interaction," in *Brain-Computer Interfaces*. Springer, 2010, pp. 3–19.

[13] S. Marcel, M. Nixon, J. Fierrez, and N. Evans, *Handbook of Biometric Anti-Spoofing (2nd Edition)*, 2019.

[14] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics Systems Under Spoofing Attack: An Evaluation Methodology and Lessons Learned," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 20–30, 2015.

[15] J. Galbally, S. Marcel, and J. Fierrez, "Biometric Anti-Spoofing Methods: A Survey in Face Recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.

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

[17] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face Liveness Detection with Component Dependent Descriptor," in *Proc. IAPR International Conference on Biometrics*, 2013.

[18] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally Efficient Face Spoofing Detection with Motion Magnification," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2013.

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

[20] N. Erdogmus and S. Marcel, "Spoofing Face Recognition with 3D Masks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1084–1097, 2014.

[21] F. Matern, C. Riess, and M. Stamminger, "Exploiting Visual Artifacts to Expose DeepFakes and Face Manipulations," in *Proc. IEEE Winter Applications of Computer Vision Workshops*, 2019.

[22] S. Agarwal and H. Farid, "Protecting World Leaders Against Deep Fakes," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[23] J. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proença, 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.

[24] F. Marra, D. Gragnaniello, L. Verdoliva, and G. Poggi, "Do GANs Leave Artificial Fingerprints?" in *Proc. IEEE Conference on Multimedia Information Processing and Retrieval*, 2019, pp. 506–511.

[25] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[26] Y. Li and S. Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[27] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in *Proc. IEEE/CVF Int. Conference on Computer Vision*, 2019.

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

[29] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. Jain, "On the Detection of Digital Face Manipulation," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

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

[31] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[32] V. Conotter, E. Bodnari, G. Boato, and H. Farid, "Physiologically-Based Detection of Computer Generated Faces in Video," in *Proc. IEEE International Conference on Image Processing*, 2014.

[33] Y. Li, M. Chang, and S. Lyu, "In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking," in *Proc. IEEE International Workshop on Information Forensics and Security*, 2018.

[34] U. A. Ciftci, I. Demir, and L. Yin, "FakeCatcher: Detection of Synthetic Portrait Videos Using Biological Signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[35] T. Jung, S. Kim, and K. Kim, "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," *IEEE Access*, vol. 8, pp. 83 144–83 154, 2020.

[36] H. Qi, Q. Guo, F. Juefei-Xu, X. Xie, L. Ma, W. Feng, Y. Liu, and J. Zhao, "DeepRhythm: Exposing DeepFakes with Attentional Visual Heartbeat Rhythms," *arXiv preprint arXiv:2006.07634*, 2020.

[37] D. Güera and E. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in *Proc. International Conference on Advanced Video and Signal Based Surveillance*, 2018.

[38] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 121–135, 2017.

[39] T. Soukupova and J. Cech, "Real-Time Eye Blink Detection Using Facial Landmarks," in *Proc. Computer Vision Winter Workshop*, 2016.

[40] W. Chen and D. McDuff, "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks," in *Proc. European Conference on Computer Vision*, 2018, pp. 349–365.

[41] J. Hernandez-Ortega, J. Fierrez, A. Morales, and D. Diaz, "A Comparative Evaluation of Heart Rate Estimation Methods using Face Videos," in *Proc. IEEE Intl. Workshop on Medical Computing*, 2020.

[42] X. Yang, Y. Li, and S. Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2019.

[43] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," in *Proc. IEEE International Workshop on Information Forensics and Security*, 2018.

[44] K. Raja *et al.*, "Morphing Attack Detection - Database, Evaluation Platform and Benchmarking," *arXiv preprint arXiv:2006.06458*, 2020.