DeepFakes Have No Heart: A Simple rPPG-Based Method to Reveal Fake Videos

Giuseppe Boccignone, Sathya Bursic, Vittorio Cuculo, Alessandro D'Amelio(✉), Giuliano Grossi, Raffaella Lanzarotti, and Sabrina Patania

PHuSe Lab - Dipartimento di Informatica, Università degli Studi di Milano, 20133 Milan, Italy
{giuseppe.boccignone,sathya.bursic,vittorio.cuculo,alessandro.damelio, giuliano.grossi,raffaella.lanzarotti,sabrina.patania}@unimi.it

Abstract. We present a simple, yet general method to detect fake videos displaying human subjects, generated via Deep Learning techniques. The method relies on gauging the complexity of heart rate dynamics as derived from the facial video streams through remote photoplethysmography (rPPG). Features analyzed have a clear semantics as to such physiological behaviour. The approach is thus explainable both in terms of the underlying context model and the entailed computational steps. Most important, when compared to more complex state-of-the-art detection methods, results so far achieved give evidence of its capability to cope with datasets produced by different deep fake models.

Keywords: DeepFake detection · rPPG · Image forensics · Biological signals

1 Introduction

The term "DeepFake" (DF, a portmanteau of "deep learning" and "fake") refers to videos created by deep learning techniques, especially generative models such as variational auto-encoders and generative adversarial networks, aiming at producing a believable media [1]. Concerning human faces, four DF categories can be identified: re-enactment [2–4], swapping [5], editing [6,7], and synthesis [8].

DF techniques date back to 2017 (a celebrated synthesized version of Obama [9]). Since then, impressive improvements have been achieved, in terms both of realism [10], accessibility and reduction of source data required to realize DF [11]. Yet, very realistic DFs have fostered unethical and malicious applications, posing a series of threats for individuals (e.g. fake porn), organizations (blackmail to managers to stop sharing their compromising DFs), politicians (e.g. fake news to sabotage government leaders) [12]. Thus, efforts have been devoted to DF detection methods allowing to discriminate between real and forged videos [13,14]. Based on the artifacts such methods look for, [12] seven categories can be drawn up: *Blending*, spatial artifacts that appear when the generated content is

S. Sclaroff et al. (Eds.): ICIAP 2022, LNCS 13232, pp. 186–195, 2022.
https://doi.org/10.1007/978-3-031-06430-2_16

blended back to the frame; *Environment*, spatial artifacts that emerge comparing the fake face with the context of the rest of the frame; *Forensics*, artifacts corresponding to subtle features or patterns introduced in the video by the model; *Behaviour*, anomalies in mannerisms or other human behaviours introduced by the model; *Physiology*, physiological signals (e.g. heart beat, blood flow, breathing) disrupted by DF methods; *Synchronization*, temporal inconsistencies (e.g., between visemes and phonemes) introduced in fake videos; *Coherence*, disrupted temporal coherence.

Markedly, behavioural and physiological signals allow to characterize the original videos, resulting in general detectors, independent of the DF generative model, or the dataset adopted for training [15]. Behavioural artifacts involve inconsistencies in physical attributes. The main features investigated are facial attributes [16], head pose [17], facial expression [18], emotions [19,20], gaze tracking [21,22], and blink detection [23].

Physiological artifacts relate to the corruption of physiological signals, such as respiratory pattern [24,25] and heart rate variability [26,27], latently conveyed by original human videos. In particular, several methods have proven their effectiveness as fake detectors using heart rate estimation. In [28] the virtual heart rate is not explicitly computed, while motion-magnified spatial-temporal maps are derived to highlight the chrominance spatio-temporal signal. A dual-spatial-temporal attention network is adopted to alleviate the influence of various interferences such as head movement or illumination variations. In [29] a two stage network is proposed based on the conjecture that in DFs the real PPG signal is lost, while a rhythmic pattern persists, which is a mixture of PPG signals, depending on the adopted generative method. DeepFakesonPhys [30] consists of a convolutional attention network composed of two parallel CNNs to extract and combine spatio-temporal information from the video.

Overall, these approaches prove their effectiveness in the intra-method and inter-method experiments, yet we argue that they suffer from two cogent issues. The first is related to the lack of explicit assessment of the physiological information; indeed, adopting end-to-end approaches brings to consider a broader set of factors (e.g. appearance, texture, behaviour) besides the physiology itself. Clearly, this would potentially allow to reach higher levels of accuracy at the cost of hiding the actual contribution of the information coming from physiological signals alone, if not framed in a principled framework. Second, in this concerning realm, explainability of the method and results achieved should be a serious issue. To the best of our knowledge, the FakeCatcher method [31] is the only proposal in such direction. To characterize interconsistency, such method trains an SVM on 126-dimensional feature vectors computed from the rPPG-derived signal, extracted from three face regions of interest by two rPPG methods. To improve performance, a CNN classifier is trained on PPG maps. However, this way, the explainability is lost, though the system is still based on biological signals.

2 Proposed Method

Background and Rationale. Consider a video of a talking agent involved in some kind of interaction (dyadic, small group, giving a speech). This can be conceived as the observable measurement of the behaviour of the agent according to an agent-in-context (AIC) model [32]. In brief (cfr. Fig. 1), over time, the agent is influenced at the conceptual level by the social and environmental context, beliefs, memories and learning. At the perceptual level the agent takes into account, both exteroceptive sensations from the world and interoceptive sensations from the body. Accordingly, the agent regulates his/her body's visceral physiology and behavioural outflow.

It is out of the scope of this note to discuss in detail the AIC model. It will suffice to remark that level coordination over time stands on a principled generative/predictive framework [32,33], where forward (bottom-up, from periphery to cortex) and backward (top-down, feedback) signalling is synchronized and enforced to support embodied simulation [33]. Indeed, the measurable behavioural/physiological outflow - e.g., facial behaviour or heart rate variability (HRV) that are cogent for the work presented here - is the result of such simulation loop [32–34].

Clearly, our research hypothesis here is that out-of-context manipulation of (faking) one of the observable variables (markedly, the face) the overall coordination is disrupted.

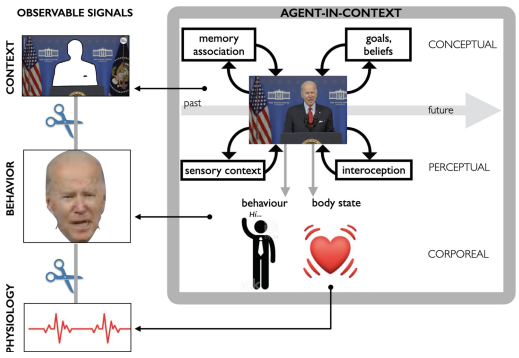

Fig. 1. Agent-in-context model. Right: over time, the agent exhibits a dynamics which involves coordination at different levels: conceptual, perceptual, and corporeal (behavioural/physiological) responses. Left: faking interventions on the observable face dynamics might disrupt (iconized by scissors) observable context/behaviour and behaviour/physiology coordination.

Gauging context/behaviour disruption (cfr. left-most side of Fig. 1), e.g., such as addressed in *Synchronization* and *Behaviour* approaches, is intriguing though complex. Thus, in this work we address behaviour/physiology disruption.

To such end, rPPG [35–38] analysis is a promising approach since HR behaviour and related vascular control can be straightforwardly assessed from the video signal itself. Indeed, rPPG is a suitable mean to sense capillary dilation and constriction related to heart beats. Dilation and constriction modulate the transmission or reflection of visible (or infra-red) light emitted to and detected from the skin. The amount of reflected light changes according to the blood volume and cardiac-synchronous variations (but see [39]).

Fake Detection via rPPG. The hidden physiological information is firstly estimated from RGB videos. To this end, the face of the (possibly) manipulated subject is detected and a set of 100 patches is automatically tracked on it.

The pixels color intensities within the i-th patch at time t $\{p_i^j(t)\}_{j=1}^{N_i}$ are averaged, thus resulting in 100 RGB traces. Denote $q_i(t)$ the RGB trace obtained from the i-th patch:

$$q_i(t) = \frac{1}{N_i} \sum_{j=1}^{N_i} p_i^j(t), \qquad i = 1, \ldots, 100. \tag{1}$$

The trace is then split into K overlapping time-windows, $q_i^k(t) = q_i(t)w\left(t - k\tau F_s\right)$, $k = 0, \ldots, K-1$, where F_s represents the video frame rate, τ is the amount of overlap and w is the rectangular window. The Blood Volume Pulse (BVP) signal is then estimated for each patch, at each time frame, using the POS rPPG method [39],

$$x_i^k(t) = POS\left[q_i^k(t)\right]. \tag{2}$$

This procedure is carried out by resorting to the `pyVHR` Python framework [26], which is suitable to scrutinize and interpret the processing steps outlined above.

Cogently, our goal here is to address signal features that have a clear semantics as to the physiological behavior. The proposed approach employs two sets of features as predictors of the presence of faking interventions. According to the AIC model, we would expect the BVP signals estimated from each patch to be disrupted. Disruption is here gauged in terms of complexity of the estimated BVP signals. Signal complexity can be addressed in many different ways; we refer to the ensemble of predictors as to *Intra-Patch BVP Complexity Measures*.

Further, it is reasonable to assume that a genuine face video would yield BVP estimates with a certain amount of coherence across the tracked patches, while exhibiting manifold behaviours in forged ones. Hence, besides individual patch-based BVP complexity, we quantify the amount of coherence between patches by means of *Inter-Patch BVP Coherence Measures*.

Intra-Patch BVP Complexity Measures. A simple and widely used measure of complexity of a time series is represented by the number of Zero-Crossings of the signal; that is the rate at which a time-series changes from positive to negative, or vice-versa. Additionally, the Hjorth Mobility and Complexity parameters

[40] are computed as indicators of the statistical properties of the signal in the time domain. In the frequency domain, the Shannon entropy of the Power Spectral Density of the estimated BVP is computed. Another common approach is to measure the entropy of a time series; more precisely, a BVP signal x can be defined as a family of time-indexed random variables $\{x(t)\}_{t\in\mathcal{T}}$. The *Entropy Rate* of a sequence of N random variables,

$$H(x) = \lim_{N\to\infty} \frac{1}{N} H\left(x(1), x(2), \ldots x(N)\right) \tag{3}$$

describes the rate of growth of the entropy of the sequence with N. The analytical computation of the Entropy Rate of a time-series requires knowing the joint distribution of the random variables composing it; however, many efficient algorithms for its approximation from a data sample have been proposed, e.g., Approximate Entropy [41], Sample Entropy, Permutation Entropy [42], SVD Entropy [43].

Another powerful statistical index of complexity, is the fractal dimension (FD) of the waveform. A variety of algorithms are available for the computation of the FD. Here we consider four approaches that have been widely adopted for the analysis of biological signals [44], namely Detrended Fluctuation Analysis (DFA), Katz FD, Petrosian FD and Higuchi FD. All such quantities are averaged across the 100 patches in each time frame in order to yield a 12-D feature vector of Intra-Patch Complexity Measures.

Inter-Patch BVP Coherence Measures. The degree of consistency between the BVP estimates across the patches is quantified for a given time window both in the time and frequency domain.

The level of spectral concordance between the BVP signals estimated from the i-th and j-th patches in the k-th time frame, can be computed as their *magnitude-squared Coherence* function,

$$C(x_i^k, x_j^k) = \frac{|G(x_i^k, x_j^k)|^2}{G(x_i^k, x_i^k)G(x_j^k, x_j^k)},$$

where $G(x_i^k, x_j^k)$ is the *cross-spectral density* between the BVP estimates from the i-th and j-th patches, respectively; $G(x_i^k, x_i^k)$ and $G(x_j^k, x_j^k)$ are the *Power Spectral Densities* of the signals. The magnitude-squared Coherence function denotes the amount of similarity between the two signals at each frequency. Here we consider the average of the magnitude-squared Coherence function over frequencies as a scalar measure of spectral similarity between BVP estimates.

We also evaluate the similarity of the signals in the time domain by measuring the Mutual Information between the BVPs:

$$MI(x_i^k, x_j^k) = \int_{x_i^k}\int_{x_j^k} P(x_i^k, x_j^k) log\left(\frac{P(x_i^k, x_j^k)}{P(x_i^k)P(x_j^k)}\right) d_{x_i^k} d_{x_j^k}$$

Such quantities are computed for each possible pair (i, j) of patches and for each time frame k. The 8-D Inter-Patch BVP Coherence feature vector for the k-th time window is then obtained by computing the first four moments of the resulting distributions (mean, standard deviation, skewness, kurtosis) for both spectral similarity and Mutual Information.

Feature Selection and Classification. The procedure described above, yields two sets of measures that when joined produce a 20-D vector of features for every frame k of a given video. Sequential floating forward selection (SFFS) [45] is thus employed in order to select the best subset of features without sensible loss of information. According to SFFS, the following 6 features result to be the most informative for DF classification via rPPG signals: Zero Crossing Rate, Petrosian FD, SVD Entropy, Higuchi FD, Average Spectral similarity and Standard Deviation of Mutual Information. Remarkably, 4 features are Complexity related quantities on BVP signals, while 2 are between-patches coherence measures.

A Support Vector Machine (SVM) is eventually adopted for classification (real vs. fake). Training and classification are carried out at the time-frame level; video level predictions are obtained by picking the most frequently predicted label.

3 Results

We test the effectiveness of the proposed approach on the FaceForensics++ dataset [46] consisting of 1000 original video sequences that have been manipulated with five methods, namely DeepFakes [47], Face2Face [48], FaceShifter [49], FaceSwap [50] and NeuralTextures [51]. Figure 2 shows one example (via FaceShifter) of swapping faces from a source to a target subject. The proposed approach is tested on each DF algorithm outcome separately. Table 1 reports the 5-fold Nested Cross-Validation Accuracy results. As can be observed, regardless of the swapping method, we are able to discriminate real from fake videos with remarkable accuracy. The table shows the results obtained by a recent approach [31]; reported accuracy scores are comparable. Most important, we test the ability of our method to generalize across different face swapping algorithms. To

Fig. 2. Face swapping. Source (left), target (center) and swapped (right) example from the FaceForensics++ dataset

Table 1. 5-Fold nested cross validation accuracy for the video-level classification of Real *vs.* Fake videos for each face swapping algorithm

	DeepFakes	Face2Face	FaceShifter	FaceSwap	NeuralText.	Avg.
Our	90.68 ± 0.61	94.46 ± 0.03	98.88 ± 0.13	95.39 ± 0.07	87.57 ± 0.35	93.40%
[31]	**94.87**%	**96.37**%	-	**95.75**%	**89.12**%	**94.02**%

this end, the original videos belonging to the "Real" class are partitioned into train/test sets with 80/20%, while the "Fake" class is built by joining all the data from every swap algorithm except one. We train our SVM on this dataset and test the model on the left out data. As a consequence, the "Fake" class contains about 5 times the number of examples if compared to the "Real" one. In Table 2, the results of the cross-method evaluation are reported. Due to the class imbalance, besides accuracy scores we also report the F1 and AUC scores, too. Notably, the proposed approach outperforms the results obtained by [31], thus exhibiting better generalization abilities despite representing a much simpler and straightforward method.

Table 2. Cross-method results: testing on a swap algorithm while training on the others

	Metric	DeepFakes	Face2Face	FaceSh.	FaceSwap	NeuralText.	Avg.
Our	AUC	89.82%	88.97%	91.52%	88.59%	89.76%	89.75%
	F1	96.72%	96.70%	98.21%	97.44%	94.14%	96.64%
	Acc.	**94.53**%	94.48%	**96.98**%	95.66%	**90.61**%	**94.45**%
[31]	Acc.	93.75%	**95.25**%	-	**96.25**%	81.25%	91.62%

4 Conclusions

We have presented a simple method for DF detection. The features extracted from the rPPG-based signal have a clear semantics as to the heart rate physiological behaviour. The approach is thus explainable both in terms of the underlying context model and the entailed computational steps. Most important, when compared to more complex state-of-the-art detection methods, present results give evidence of its capability to cope with datasets produced by different DF models. Beyond the reported results, the agent-in-context model, which frames our approach, is suitable for paving the way to a seamless integration of analyses addressing different levels of the faked agent's behavior. The present investigation is to be intended as a *proof of concept*, hence further experiments on larger and more challenging datasets (e.g. [52,53]) are planned to be carried out in future works.

References

1. Lee, S.-H., Yun, G.-E., Lim, M.Y., Lee, Y.K.: A study on effective use of bpm information in deepfake detection. In: 2021 International Conference on Information and Communication Technology Convergence (ICTC), pp. 425–427. IEEE (2021)
2. Bansal, A., Ma, S., Ramanan, D., Sheikh, Y.: Recycle-GAN: unsupervised video retargeting. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 119–135 (2018)
3. Tran, L., Yin, X., Liu, X.: Representation learning by rotating your faces. IEEE Trans. Pattern Anal. Mach. Intell. **41**(12), 3007–3021 (2018)
4. Bursic, S., D'Amelio, A., Granato, M., Grossi, G., Lanzarotti, R.: A quantitative evaluation framework of video de-identification methods. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 6089–6095. IEEE (2021)
5. Peng, B., Fan, H., Wang, W., Dong, J., Lyu, S.: A unified framework for high fidelity face swap and expression reenactment. In: IEEE Transactions on Circuits and Systems for Video Technology (2021)
6. Gupta, A., Khan, F.F., Mukhopadhyay, R., Namboodiri, V.P., Jawahar, C.: Intelligent video editing: incorporating modern talking face generation algorithms in a video editor. In: Proceedings of the Twelfth Indian Conference on Computer Vision, Graphics and Image Processing, pp. 1–9 (2021)
7. Ding, H., Sricharan, K., Chellappa, R.: Exprgan: facial expression editing with controllable expression intensity. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
8. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8110–8119 (2020)
9. Suwajanakorn, S., Seitz, S.M., Kemelmacher-Shlizerman, I.: Synthesizing Obama: learning lip sync from audio. ACM Trans. Graph. **36**(4), 1–13 (2017)
10. Nirkin, Y., Keller, Y., Hassner, T.: FSGAN: subject agnostic face swapping and reenactment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7184–7193 (2019)
11. Lattas, A., et al.: Avatarme: realistically renderable 3D facial reconstruction in-the-wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 760–769 (2020)
12. Mirsky, Y., Lee, W.: The creation and detection of deepfakes: a survey. ACM Comput. Surv. **54**(1), 1–41 (2021)
13. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J.: Deepfakes and beyond: a survey of face manipulation and fake detection. Inf. Fusion **64**, 131–148 (2020)
14. Nguyen, T.T., Nguyen, C.M., Nguyen, D.T., Nguyen, D.T., Nahavandi, S.: Deep learning for deepfakes creation and detection: a survey. arXiv preprint arXiv:1909.11573 (2019)
15. Ciftci, U.A., Demir, I., Yin, L.: How do the hearts of deep fakes beat? deep fake source detection via interpreting residuals with biological signals. In: 2020 IEEE International Joint Conference on Biometrics (IJCB), pp. 1–10. IEEE (2020)
16. Matern, F., Riess, C., Stamminger, M.: Exploiting visual artifacts to expose deepfakes and face manipulations. In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 83–92. IEEE (2019)

17. Yang, X., Li, Y., Lyu, S.: Exposing deep fakes using inconsistent head poses. In: ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8261–8265. IEEE (2019)
18. Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., Li, H.: Protecting world leaders against deep fakes. In: CVPR Workshops, vol. 1 (2019)
19. Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., Manocha, D.: Emotions don't lie: an audio-visual deepfake detection method using affective cues. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 2823–2832 (2020)
20. Hosler, B., et al.: Do deepfakes feel emotions? a semantic approach to detecting deepfakes via emotional inconsistencies. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1013–1022 (2021)
21. Demir, I., Ciftci, U.A.: Where do deep fakes look? synthetic face detection via gaze tracking. In: ACM Symposium on Eye Tracking Research and Applications, pp. 1–11 (2021)
22. Cuculo, V., D'Amelio, A., Lanzarotti, R., Boccignone, G.: Personality gaze patterns unveiled via automatic relevance determination. In: Mazzara, M., Ober, I., Salaün, G. (eds.) STAF 2018. LNCS, vol. 11176, pp. 171–184. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04771-9_14
23. Jung, T., Kim, S., Kim, K.: Deepvision: deepfakes detection using human eye blinking pattern. IEEE Access **8**, 83144–83154 (2020)
24. Prathosh, A., Praveena, P., Mestha, L.K., Bharadwaj, S.: Estimation of respiratory pattern from video using selective ensemble aggregation. IEEE Trans. Signal Process. **65**(11), 2902–2916 (2017)
25. Chen, M., Zhu, Q., Zhang, H., Wu, M., Wang, Q.: Respiratory rate estimation from face videos. In: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 1–4. IEEE (2019)
26. Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., Lanzarotti, R.: An open framework for remote-PPG methods and their assessment. IEEE Access **8**, 216083–216103 (2020)
27. Rouast, P.V., Adam, M.T., Chiong, R., Cornforth, D., Lux, E.: Remote heart rate measurement using low-cost RGB face video: a technical literature review. Front. Comput. Sci. **12**(5), 858–872 (2018)
28. Qi, H., et al.: DeepRhythm: exposing deepfakes with attentional visual heartbeat rhythms. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 4318–4327 (2020)
29. Liang, J., Deng, W.: Identifying rhythmic patterns for face forgery detection and categorization. In: 2021 IEEE International Joint Conference on Biometrics (IJCB), pp. 1–8 (2021)
30. Hernandez-Ortega, J., Tolosana, R., Fierrez, J., Morales, A.: DeepFakesON-Phys: Deepfakes detection based on heart rate estimation. arXiv preprint arXiv:2010.00400 (2020)
31. Ciftci, U.A., Demir, I., Yin, L.: FakeCatcher: detection of synthetic portrait videos using biological signals. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
32. Koban, L., Gianaros, P.J., Kober, H., Wager, T.D.: The self in context: brain systems linking mental and physical health. Nat. Rev. Neurosci. **22**(5), 309–322 (2021)
33. Hutchinson, J.B., Barrett, L.F.: The power of predictions: an emerging paradigm for psychological research. Curr. Direct. Psychol. Sci. **28**(3), 280–291 (2019)

34. Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., Lanzarotti, R.: Deep construction of an affective latent space via multimodal enactment. IEEE Trans. Cognit. Develop. Syst. **10**, 865–880 (2018)
35. Wieringa, F.P., Mastik, F., Steen, A.F.W.v.d.: Contactless multiple wavelength photoplethysmographic imaging: a first step toward "spo2 camera"technology". Ann. Biomed. Eng. **33**(8), 1034–1041 (2005)
36. Humphreys, K., Ward, T., Markham, C.: Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry. Rev. Sci. Instrum. **78**(4), 044304 (2007)
37. Verkruysse, W., Svaasand, L.O., Nelson, J.S.: Remote plethysmographic imaging using ambient light. Opt. Express **16**(26), 21434–21445 (2008)
38. McDuff, D.J., Estepp, J.R., Piasecki, A.M., Blackford, E.B.: A survey of remote optical photoplethysmographic imaging methods. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6398–6404. IEEE (2015)
39. Wang, W., den Brinker, A.C., Stuijk, S., De Haan, G.: Algorithmic principles of remote PPG. IEEE Trans. Biomed. Eng. **64**(7), 1479–1491 (2016)
40. Hjorth, B.: Eeg analysis based on time domain properties. Electroencephalogr. Clin. Neurophysiol. **29**(3), 306–310 (1970)
41. Pincus, S.M., Gladstone, I.M., Ehrenkranz, R.A.: A regularity statistic for medical data analysis. J. Clin. Monitor. **7**(4), 335–345 (1991)
42. Bandt, C., Pompe, B.: Permutation entropy: a natural complexity measure for time series. Phys. Rev. Lett. **88**(17), 174102 (2002)
43. Roberts, S.J., Penny, W., Rezek, I.: Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing. Med. Biol. Eng. Comput. **37**(1), 93–98 (1999)
44. Esteller, R., Vachtsevanos, G., Echauz, J., Litt, B.: A comparison of waveform fractal dimension algorithms. IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. **48**(2), 177–183 (2001)
45. Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature selection. Pattern Recogn. Lett. **15**(11), 1119–1125 (1994)
46. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M.: FaceForensics++: learning to detect manipulated facial images. In: International Conference on Computer Vision (ICCV) (2019)
47. Deepfakes. https://github.com/deepfakes/faceswap
48. Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., Nießner, M.: "Face2face: real-time face capture and reenactment of RGB videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2387–2395 (2016)
49. Li, L., Bao, J., Yang, H., Chen, D., Wen, F.: Faceshifter: towards high fidelity and occlusion aware face swapping. arXiv preprint arXiv:1912.13457 (2019)
50. Faceswap. https://github.com/MarekKowalski/FaceSwap/
51. Thies, J., Zollhöfer, M., Nießner, M.: Deferred neural rendering: image synthesis using neural textures. ACM Trans. Graph. **38**(4), 1–12 (2019)
52. Li, Y., Yang, X., Sun, P., Qi, H., Lyu, S.: Celeb-df: a large-scale challenging dataset for deepfake forensics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3207–3216 (2020)
53. Dolhansky, B., et al.: The deepfake detection challenge (dfdc) dataset. arXiv preprint arXiv:2006.07397 (2020)