On using rPPG signals for DeepFake detection: a cautionary note

Alessandro D'Amelio[1][0000−0002−8210−4457], Raffaella Lanzarotti[1][0000−0002−8534−4413], Sabrina Patania[1][0000−0001−7279−723X], Giuliano Grossi[1][0000−0001−9274−4047], Vittorio Cuculo[2][0000−0002−8479−9950], Andrea Valota[1], and Giuseppe Boccignone[1][0000−0002−5572−0924]

[1] PHuSe Lab - Università degli Studi di Milano
[2] AImageLab - Università degli Studi di Modena e Reggio Emilia
{alessandro.damelio, raffaella.lanzarotti, sabrina.patania,giuliano.grossi,andrea.valota,giuseppe.boccignone}@unimi.it
vittorio.cuculo@unimore.it

Abstract. An experimental analysis is proposed concerning the use of physiological signals, specifically remote Photoplethysmography (rPPG), as a potential means for detecting Deepfakes (DF). The study investigates the effects of different variables, such as video compression and face swap quality, on rPPG information extracted from both original and forged videos. The experiments aim to understand the impact of face forgery procedures on remotely-estimated cardiac information, how this effect interacts with other variables, and how rPPG-based DF detection accuracy is affected by these quantities. Preliminary results suggest that cardiac information in some cases (e.g. uncompressed videos) may have a limited role in discriminating real videos from forged ones, but the effects of other physiological signals cannot be discounted. Surprisingly, heart rate related frequencies appear to deliver a significant contribution to the DF detection task in compressed videos.

Keywords: Deepfake detection · rPPG · Video forensics · Physiological signals

1 Introduction

Fake videos generated through deep learning techniques (Deepfakes [1], DF) blur the line between truth and deception (surmising, for simplicity, that such a line can be drawn). This very fact might lead to a responsible use in the service of many realms (entertainment, education [2], advertising, or privacy protection via de-identification [3]). However, the major concern lies in that DFs might pave the way to the murky realm of fake identity creation for unethical and malicious applications, posing a variety of threats to individuals (e.g. fake porn), organizations (e.g. blackmail to managers to stop sharing their compromising DFs), and politicians (e.g. fake news to sabotage government leaders) [4].

It is no surprise that, since the paradigmatic generation of a synthesized version of Obama [5] in 2017, increasing efforts have been devoted to develop

DF detection (DFD) methods that can differentiate between real and forged videos [6, 7]. Yet, DFD raises subtle issues that, in spite of the flourishing of published methods achieving good metrics performance on public datasets, have been hitherto most often neglected, an attitude that can negatively impinge on DFD generalization from the lab to real-world contexts. A chief concern of this note is to make a step forward in unveiling some of such issues.

DFD methods might be coarsely grouped according to the kind of artifacts that a DF technique can introduce into the forged video (but see [4]), from spatial artifacts and disarranged temporal coherence to anomalies in human behaviours and semantic inconsistencies (such as those between visemes and phonemes).

A class of methods that have recently gained currency is that relying on virtual measurements of physiological data. The idea is simple and straightforward: DF techniques are likely to disrupt physiological signals (such as heart rate, blood flow, and breathing) that can be detected in a contactless way from the RGB video stream. One such case is represented by measurements (e.g., heart rate, HR, and respiration rhythm, RR) derived from the Blood Volume Pulse (BVP) signal estimated via remote photoplethysmography (rPPG, [8]).

In this note we investigate rPPG-based DFD in order to eventually gauge the effects of different variables (e.g. video compression, rPPG method, etc.) on the rPPG information extracted from both original and forged videos in a controlled experimental setting. Specifically, we address the following research questions:

1. (*RQ1*) How a face forgery procedure impinges the remotely-estimated cardiac information?
2. (*RQ2*) How such effect interacts with other factors, such as video compression and DF quality?
3. (*RQ3*) How rPPG-based DF detection accuracy is overall affected by the above factors?

Preliminary results so far achieved suggest that, in discriminating real videos from forged ones, cardiac information displays a nuanced role but depending on a number of factors, while effects of physiological signals other than the cardiac one embedded within the BVP cannot be completely ruled out. Under such circumstances, the application of rPPG-based DFD in real-world contexts should be carefully designed and weighed up.

2 Background and Related Works

A variety of techniques have been developed to efficiently and effectively estimate vital signs relying solely on standard cameras [8–10]. DF detection techniques exploiting physiological information rely on the assumption that face manipulation would produce a (partial) corruption of such information, thus introducing physiological artifacts. Several methods relying on remotely estimated cardiac information have proven effective in DFD, yet, to the best of our knowledge, an analysis of the determinants of such effectiveness is still missing. Notoriously, rPPG approaches are fraught with hurdles, requiring specific conditions to be

met, one among many, the availability of uncompressed (or lightly compressed) videos [11]. Interestingly enough, remarkable DFD results have been achieved even on heavily compressed datasets using rPPG information.

DeepRythm [12], employs motion-magnified spatial-temporal maps to highlight chrominance spatio-temporal signals and a dual-spatial-temporal attention network to reduce interference. Another approach, described in [13], uses a two-stage network to identify rhythmic patterns in PPG signals that persist in deepfakes. DeepFakesonPhys [14] relies on a convolutional attention network composed of two parallel CNNs to extract and combine spatio-temporal information from videos. FakeCatcher method [15] exploits an SVM to train on 126-dimensional feature vectors computed from rPPG-derived signals extracted from three facial regions of interest by two rPPG methods. To improve performance, a CNN classifier is trained on PPG maps. Similarly, in [16] the authors use PPG maps and show that various manipulation techniques produce distinct patterns of heartbeats. In [17] a simple and explainable approach is proposed, which relies on gauging both intra-patch complexity measures and inter-patch coherence of rPPG signals. In [18], Spatial-Temporal Filtering Network (STFNet) is considered for rPPG filtering together with a Spatial-Temporal Interaction Network (STINet) accounting for the interaction of PPG signals. More recently, a Multi-scale Spatial–Temporal PPG map has been used to exploit cardiac signals extracted from multiple facial regions [19]. In order to capture both spatial and temporal inconsistencies, the authors laid down a two-stage network consisting of a Mask-Guided Local Attention module together with a Temporal Transformer.

3 Material and Methods

Over the last few years a number of benchmark datasets have been proposed with the aim of scoring different DF detection approaches (e.g. [20–23]). Typically, these benchmarks are built by crawling YouTube videos and subsequently adopting DF techniques to swap identities, reenact faces or manipulate attributes. Clearly, these datasets have been conceived for direct methods comparison and not to understand the effects of face forgery techniques on rPPG signals, which is the problem we are addressing here. To this end, a physiological ground truth is mandatory to perform appropriate comparisons. Moreover, albeit some benchmarks provide a distinction between high quality and low quality videos, uncompressed video is not in general available for the task of DF detection. Under such circumstances, to lay down a controlled setup, we opted to build our own artificially manipulated dataset (UBFC1-forged Dataset, UBFC1-F) as described in the following.

3.1 The UBFC1-forged Dataset

UBFC1-F is constructed from the publicly available UBFC1 dataset [24]. The latter is typically employed to assess the quality of rPPG methods. It is composed of 8 videos (about 16500 frames) recorded using a low-cost webcam (Logitech

C920 HD Pro) at a resolution of 640x480 pxl in uncompressed 8-bit RGB format, with a frame rate of 30fps. The ground truth PPG data, including the PPG waveform and heart rates, were obtained using a CMS50E transmissive pulse oximeter. The recordings were made while the subjects sat in front of the camera, positioned approximately 1 meter away, with their face visible. The participants were instructed to remain motionless; even so, some of the recorded videos exhibit noticeable movement.

UBFC1-F dataset was generated from UBFC1 as follows. The freely available *FaceSwap* tool[3] has been employed in order to perform face forgery. This method replaces a face in a target sequence with one observed in a source sequence using a simple method based on neural image style transfer [25].

Each subject identity was swapped with every other individual and for each video a variant with different compression rates: 1) *No Compression*, 2) *High Compression*. Compressed videos were obtained via H.264 encoding with a Constant Rate Factor (CRF) set to 30 (High Compression).

Moreover, in order to simulate 1) *Low Quality* and 2) *High Quality* forged videos, training iterations of the *FaceSwap* model were stopped to 25000 in the former case and brought to convergence in the latter.

3.2 Physiological estimation and Analysis

For each video in the UBFC1-F, either real of forged, the *Blood Volume Pulse* (BVP) signal is estimated from displayed faces via rPPG by exploiting the `pyVHR` framework[4] [9, 10]. `pyVHR` implements well established pipelines allowing to derive cardiac information using either classic signal processing or learning-based approaches. Here, the classic pipeline is adopted due to its simplicity and higher level of explainability. Though Deep Learning based approaches exhibit superior performance on many benchmark datasets, their generalization abilities are still subject to some controversy [9, 10, 26].

The RGB videos are used to estimate hidden physiological information, starting with the detection of the face of a possibly manipulated subject, and the automatic tracking of a set of patches on the face around the cheeks area (which typically contains pixels belonging to the swapping area). For each of the P patches, the color intensities of the pixels within it at time t, denoted by $\{p_i^j(t)\}_{j=1}^{N_i}$ $(i = 1, \ldots, P)$, are averaged to create P RGB traces. Let $q_i(t)$ be the RGB trace obtained from the i-th patch of length N_i:

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

These traces are then split into K overlapping time-windows, $q_i^k(t) = q_i(t)w\left(t - k\tau F_s\right)$, $k = 0, \ldots, K-1$, F_s being the video frame rate, $\tau < 1$ the fraction of overlap, and w a rectangular window. For each patch i and at each time frame k, the

[3] `https://github.com/deepfakes/faceswap`
[4] `https://github.com/phuselab/pyVHR`

BVP signal is estimated using either the GREEN [27] or the POS [28] method. Denote $\mathcal{M} := \{\text{GREEN}, \text{POS}\}$; the BVP estimate for the i-th patch and the k-th time frame is obtained as:

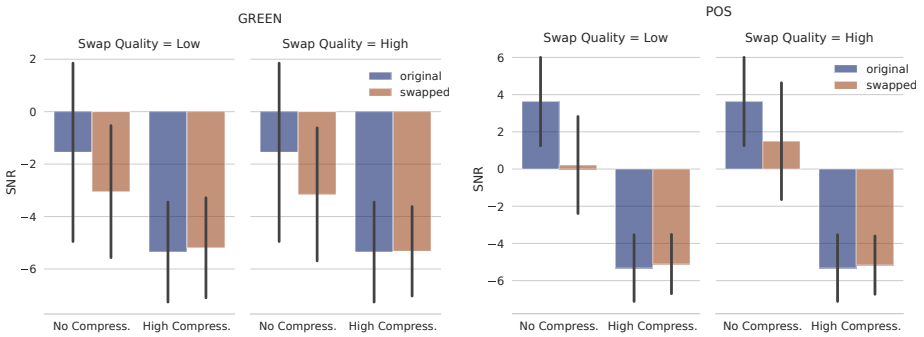
$$x_i^k(t) = \mathcal{M}\left[q_i^k(t)\right]. \tag{2}$$

SNR Analysis In order to address (*RQ1*) and (*RQ2*), a *Signal-to-Noise Ratio* (SNR) analysis is performed. The goal is to assess to what extent physiological information in forged videos is either disrupted or maintained. At the same time, we gauge the effects of both video compression and forgery method's quality.

The SNR can be operationalised according to [29]; namely, the ratio of the power around the reference HR frequency (i.e., the ground truth HR frequency) plus the first harmonic of the estimated pulse-signal and the remaining power contained in the spectrum of the estimated BVP:

$$\text{SNR} = \frac{1}{K}\sum_k 10\log_{10}\left(\frac{\sum_v \left(U^k(v)S^k(v)\right)^2}{\sum_v \left(1-U^k(v)\right)S^k(v)\right)^2}\right), \tag{3}$$

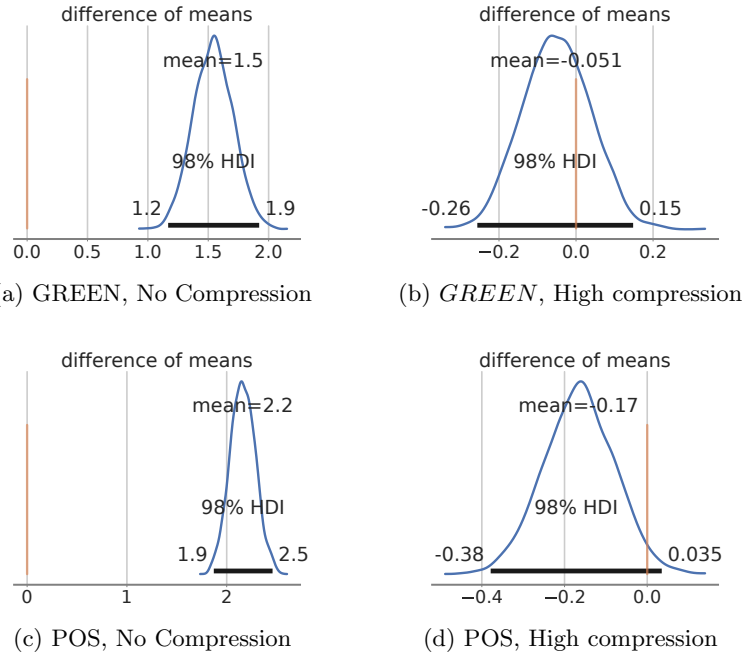
where $S^k(v)$ is the power spectral density of the estimated BVP in the k-th time window and $U^k(v)$ is a binary mask that selects the power contained within ± 12 BPM around the reference HR and its first harmonic. The SNR is computed for each BVP signal estimated in each video of the UBFC1-F dataset in order to highlight differences in *real vs. forged* identities, *uncompressed vs. compressed* videos and *Low Quality vs. High Quality* scenarios. Results are visualised in Figure 1.



(a) (b)

Fig. 1: SNR values obtained on the rPPG estimates on original and swapped videos using the GREEN **(a)** and POS **(b)** methods with different iterations of the *FaceSwap* DF approach.

Quantitatively, statistical significance is gauged via the BEST statistical analysis (a Bayesian version of the t-Test) [30]; results are presented in Figure 2, where the posterior probability distributions of the difference of means between Signal-to-Noise Ratio (SNR) values obtained from original and swapped videos of the UBFC1-F dataset are shown. The results are demonstrated for both the GREEN and POS rPPG methods and for scenarios where video compression is applied or not.



(a) GREEN, No Compression (b) *GREEN*, High compression

(c) POS, No Compression (d) POS, High compression

Fig. 2: Result of the BEST test for statistical comparison of two groups. Posterior probability distributions of the difference of means between SNR values obtained from original and swapped videos of the UBFC1-F dataset in case video compression is applied or not (for both GREEN and POS rPPG methods). If the 0 difference value falls outside the highest density interval (98% HDI) of the posterior, it can be deemed an implausible value (the two distributions can be considered distinct). The test reveals significant differences in uncompressed videos (**(a)** and **(c)**) and non-significant differences in compressed ones (**(a)** and **(d)**). Figure shows results for the *High Quality* swap case; similar results are obtained for the *Low Quality* case.

3.3 DeepFake Detection analysis

As to (*RQ3*), the BVP signals estimated on the UBFC1-F dataset are employed to gauge the effect of the same variables considered above on the DF classification accuracy. To this end, the DF detection method proposed in [17] and depicted at a glance in Figure 3 is exploited. Two sets of features are used as predictors of the presence of faking interventions:

- *Intra-Patch BVP Complexity Measures*: A set of features quantifying the entropy rate of BVP signals
- *Inter-Patch BVP Coherence Measures*: A set of features measuring the degree of consistency between the BVP estimates across the patches.

Feature extraction is followed by an SVM binary classifier for the final DFD step.

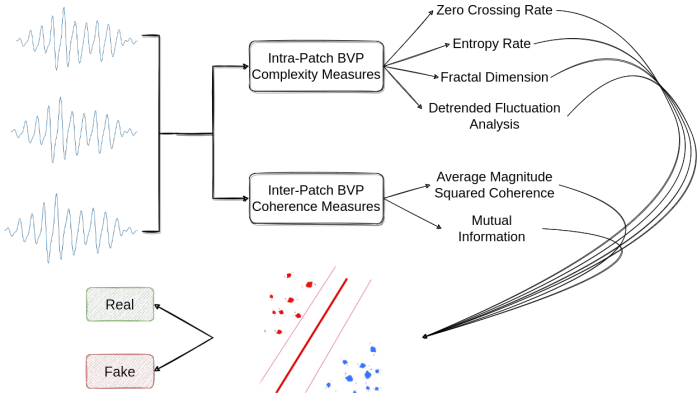


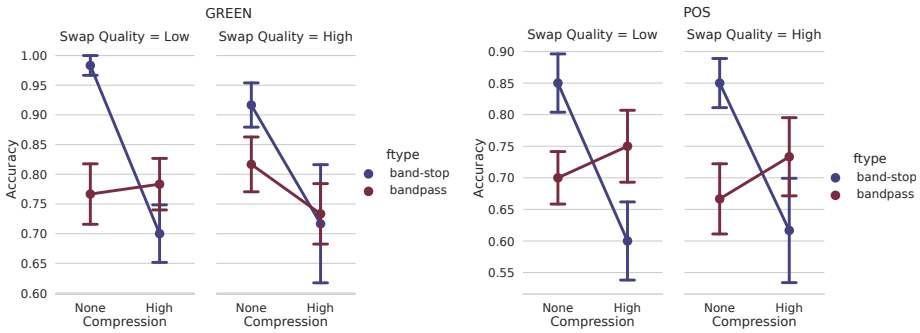
Fig. 3: Overview of the DF detection approach presented in [17] adapted for the analyses presented here

As with the SNR analysis we quantify the effect of video compression and *FaceSwap* level of convergence (method quality) on the DF detection task. In addition, to gauge the role played by heart-rate (HR) related information, we exploit two different filters (*ftype*) applied on the Power Spectral Density (PSD) of the BVP signal:

1. *Band-stop*: A band-stop filter removing the HR-related frequencies (0.75 - 4.0 Hz)
2. *Bandpass*: A band-stop filter keeping only the HR related frequencies (0.75 - 4.0 Hz)

The DFD accuracy levels achieved on the UBFC1-F dataset by varying the above factors are reported in Figure 4.

Further statistical hypothesis testing to quantify significant differences between accuracy levels is performed according to [31]. We evaluate the potential



(a) (b)

Fig. 4: Accuracy levels for DF detection using the GREEN **(a)** and POS **(b)** rPPG methods with varying compression rates, filter types and *FaceSwap*'s swap quality.

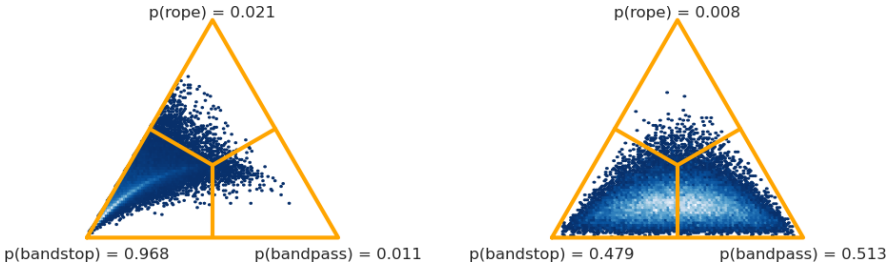
improvement of one scenario over another by utilizing the Bayesian Sign-Rank Test [32], a Bayesian non-parametric method that extends the Wilcoxon signed-rank test.

Figure 5 depicts the results of the analysis by reporting the posterior samples for the Bayesian Sign-Rank Test on the simplex. Each vertex of the triangle represents the case where a scenario is either more probable to yield higher DF detection accuracy w.r.t the other or equivalent (probability over the Region of Practical Equivalence, $P(\text{ROPE})$).

4 Discussion

SNR analysis. This analysis clearly shows that on uncompressed videos displaying real identities, physiological information is baldly present; this is particularly evident when using the POS rPPG method, which notoriously exhibits more robust performances if compared to the baseline GREEN method (cfr. Figure 1(b)). Interestingly enough, when *FaceSwap* is applied, the SNR drastically drops, albeit betraying some residual original cardiac information (especially when the swapping method is trained to convergence). Notably, on compressed videos the SNR reveals weak presence of cardiac information with non significant differences between real and forged videos either using the GREEN or POS rPPG method.

Indeed, from a statistical standpoint, BEST analysis (Fig. 2) shows that in the case of uncompressed videos, significant differences are observed between the SNR values of original and forged videos (Fig. 3, panels **(a)** and **(c)**), whereas non-significant differences are observed in compressed videos (panels **(b)** and **(d)**). Presented results refer to the "*High Quality* swap" case, but the same conclusion can be drawn for the "*Low Quality*" scenario.



(a) GREEN (No compression) (b) GREEN (High compression)

(c) POS (No compression) (d) POS (High compression)

Fig. 5: Posterior samples for the Bayesian Sign-Rank Test on the simplex comparing the distributions of accuracy levels when choosing a *band-stop* vs. *bandpass* filtering of a BVP signal estimated using either GREEN or POS rPPG methods. Each plot shows the probability for a particular scenario to yield higher accuracy if compared with another.

DFD analysis. Results reported in Figure 4(a) show that the rPPG signals estimated by the GREEN method in general yield the highest accuracy levels on uncompressed videos. This result is rather surprising, considering that, on average, GREEN's signals exhibited negative SNR (cfr. Figure 1(a)), which in principle should result in poor capabilities of capturing BVP information. Video compression, significantly cuts down the DFD performance of the method. However, at high compression, cardiac information (selected via bandpass filtering) yields higher DFD accuracy than that obtain from other PSD bands (band-stop filtering) when *FaceSwap* simulates a *Low Quality* swapper, and comparable accuracy for *High Quality* swapping. Such results are statistically confirmed via the Bayesian Sign-Rank Test (Fig. 5, panels (a) and (b)). Further, when the POS method is considered, such trend becomes crystal clear, and the no compression/high compression factor swaps the roles played by information obtained after bandpass/band-stop filtering to achieve DFD accuracy (Fig. 3(b) and Fig. 5(c) and (d))

Overall, results can be summarized as follows: at zero compression, if we consider the classification obtained from *band-stop* filtered vs. *bandpass* filtered

rPPG signals, the former is significantly better for all methods. However, with compressed videos results are reversed: when considering the classification accuracy obtained from *band-stop* filtered vs. *bandpass* filtered rPPG signals, the latter is generally better for both rPPG methods (GREEN and POS). Therefore, *prima facie*, it can be surmised that starting from uncompressed videos, the DF method introduces artifacts that are better captured by the PSD bands of the BVP that do not strictly involve cardiac information. However, other physiological signals may be still present in the filtered BVPs and hence eventually be recruited for the classification. As a matter of fact, BVP signals carry a variety of vital signs (e.g. respiratory signals, blood oxygen levels, blood pressure etc.) that can be extracted (see, for instance, [8, 26]). Video compression, though, seems to smooth out these artifacts that no longer provide enough information, while the HR-related frequency band "withstand" compression as opposed to other frequency bands, in spite of the low SNR. This is an interesting result because, if we look at the SNR of the estimated BVPs (Figure 1), which is significantly different between the original (high SNR) and fake (low SNR) cases at zero compression, it becomes negative with no significant differences between original and fake videos in compressed videos (cfr. Figure 2). In other words, under conditions of high compression, the noise is higher than the signal, but the signal, markedly the HR-related component, is still present and is likely to be used.

5 Conclusions

The effective role played by HR information in DFD is the result of complex interactions between the original quality of the video stream and its compression, the power of the DF method, the quality of the estimated signal and the method adopted for BVP estimation. Such interactions can lead to counter-intuitive results. Although the current study is a simple simulation-based approach and has potential limitations as it solely relies on the *FaceSwap* technique to generate deepfakes, it provides insights into how different variables can impact remotely estimated BVP signals. It's worth remarking that, the reported classification results depend on the specific DF detection method [17] employed here; the dependence of the analysed quantities on other DF methods and classification approaches will be addressed in a future research. Preliminary results in such direction (not reported here) show that, in the outlined scenario, different BVP estimation methods can also lead to different outcomes over different datasets, thus impinging on DFD generalization when a cross-dataset evaluation is performed. To sum up, the obtained results suggest that the role of cardiac information in distinguishing genuine videos from manipulated ones may be limited in certain situations, such as with uncompressed videos. However, the impact of other physiological signals should not be disregarded. Interestingly, though, frequencies associated with heart rate seem to play a significant role in detecting manipulated videos, particularly in compressed formats.

All in all, in spite of the promising results reported "in-the-lab" condition, HR information for DFD should be handled with care when applications "in-the-wild" are targeted.

References

1. S.-H. Lee, G.-E. Yun, M. Y. Lim, and Y. K. Lee, "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. D. Lee, "Deepfake Salvador Dalí takes selfies with museum visitors," *The Verge*, 2019.
3. S. Bursic, A. D'Amelio, M. Granato, G. Grossi, and R. Lanzarotti, "A quantitative evaluation framework of video de-identification methods," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 6089–6095, IEEE, 2021.
4. Y. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 1, pp. 1–41, 2021.
5. S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing obama: learning lip sync from audio," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 4, pp. 1–13, 2017.
6. 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.
7. T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep learning for deepfakes creation and detection: A survey," *arXiv preprint arXiv:1909.11573*, 2019.
8. D. McDuff, "Camera measurement of physiological vital signs," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–40, 2023.
9. G. Boccignone, D. Conte, V. Cuculo, A. D'Amelio, G. Grossi, and R. Lanzarotti, "An open framework for remote-PPG methods and their assessment," *IEEE Access*, vol. 8, pp. 216083–216103, 2020.
10. G. Boccignone, D. Conte, V. Cuculo, A. D'Amelio, G. Grossi, R. Lanzarotti, and E. Mortara, "pyvhr: a python framework for remote photoplethysmography," *PeerJ Computer Science*, vol. 8, p. e929, 2022.
11. D. J. McDuff, E. B. Blackford, and J. R. Estepp, "The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 63–70, IEEE, 2017.
12. 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," in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 4318–4327, 2020.
13. J. Liang and W. Deng, "Identifying rhythmic patterns for face forgery detection and categorization," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–8, 2021.
14. J. Hernandez-Ortega, R. Tolosana, J. Fierrez, and A. Morales, "DeepFakesON-Phys: Deepfakes detection based on heart rate estimation," *arXiv preprint arXiv:2010.00400*, 2020.

15. 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.
16. U. A. Ciftci, I. Demir, and L. Yin, "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.
17. G. Boccignone, S. Bursic, V. Cuculo, A. D'Amelio, G. Grossi, R. Lanzarotti, and S. Patania, "Deepfakes have no heart: A simple rppg-based method to reveal fake videos," in *Image Analysis and Processing–ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part II*, pp. 186–195, Springer, 2022.
18. J. Liang and W. Deng, "Identifying rhythmic patterns for face forgery detection and categorization," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–8, IEEE, 2021.
19. J. Wu, Y. Zhu, X. Jiang, Y. Liu, and J. Lin, "Local attention and long-distance interaction of rppg for deepfake detection," *The Visual Computer*, pp. 1–12, 2023.
20. A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to detect manipulated facial images," in *International Conference on Computer Vision (ICCV)*, 2019.
21. B. Dolhansky, J. B. B. Pflaum, R. H. Jikuo Lu, C. Menglin Wang, and C. Ferrer, "The deepfake detection challenge (DFDC) dataset," *arXiv preprint arXiv:2006.07397*, 2020.
22. P. S. H. Q. Yuezun Li, Xin Yang and S. Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *IEEE Conference on Computer Vision and Patten Recognition (CVPR)*, 2020.
23. L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection," in *CVPR*, 2020.
24. S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognition Letters*, vol. 124, pp. 82–90, 2019.
25. M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," *Advances in neural information processing systems*, vol. 30, 2017.
26. G. Boccignone, A. D'Amelio, O. Ghezzi, G. Grossi, and R. Lanzarotti, "An evaluation of non-contact photoplethysmography-based methods for remote respiratory rate estimation," *Sensors*, vol. 23, no. 7, p. 3387, 2023.
27. W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light.," *Opt. Express*, vol. 16, no. 26, pp. 21434–21445, 2008.
28. W. Wang, A. C. den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2016.
29. G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.
30. J. K. Kruschke, "Bayesian estimation supersedes the t test.," *Journal of Experimental Psychology: General*, vol. 142, no. 2, p. 573, 2013.
31. A. Benavoli, G. Corani, J. Demšar, and M. Zaffalon, "Time for a change: a tutorial for comparing multiple classifiers through bayesian analysis," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 2653–2688, 2017.
32. A. Benavoli, G. Corani, F. Mangili, M. Zaffalon, and F. Ruggeri, "A Bayesian Wilcoxon signed-rank test based on the Dirichlet process," in *International conference on machine learning*, pp. 1026–1034, PMLR, 2014.