



OPEN ACCESS

EDITED BY

Matteo Ferrara,
University of Bologna, Italy

REVIEWED BY

Deepayan Bhowmik,
Newcastle University, United Kingdom
Giuseppe Boccignone,
University of Milan, Italy

*CORRESPONDENCE

Peter Eisert
✉ peter.eisert@hhi.fraunhofer.de

[†]These authors have contributed equally to this work

RECEIVED 30 September 2024

ACCEPTED 25 February 2025

PUBLISHED 30 April 2025

CITATION

Seibold C, Wisotzky EL, Beckmann A, Kossack B, Hilsmann A and Eisert P (2025) High-quality deepfakes have a heart!. *Front. Imaging* 4:1504551. doi: 10.3389/fimag.2025.1504551

COPYRIGHT

© 2025 Seibold, Wisotzky, Beckmann, Kossack, Hilsmann and Eisert. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

High-quality deepfakes have a heart!

Clemens Seibold^{1,2†}, Eric L. Wisotzky^{1,2†}, Arian Beckmann¹, Benjamin Kossack¹, Anna Hilsmann¹ and Peter Eisert^{1,2*}

¹Computer Vision & Graphics, Vision & Imaging Technologies, Fraunhofer Heinrich-Hertz-Institute HHI, Berlin, Germany, ²Visual Computing, Department of Computer Science, Humboldt University, Berlin, Germany

Introduction: Deepfakes have become ubiquitous in our modern society, with both their quantity and quality increasing. The current evolution of image generation techniques makes the detection of manipulated content through visual inspection increasingly difficult. This challenge has motivated researchers to analyze heart-beat-related signal to distinguish deep fakes from genuine videos.

Methods: In this study, we analyze deepfake videos of faces generated with novel methods regarding their heart-beat-related signals using remote photoplethysmography (rPPG). The rPPG signal describes the blood flow based, or rather local blood volume changes, and thus reflects the pulse signal. For our analysis, we present a pipeline that extracts rPPG signals and investigate the origin of the extracted signals in deepfake videos using correlation analyses. To validate our rPPG extraction pipeline and analyze rPPG signals of deepfakes, we captured a dataset of facial videos synchronized with an electrocardiogram (ECG) as a ground-truth pulse signal. Additionally, we generated high-quality deepfakes and incorporated publicly available datasets into our evaluation.

Results: We prove that our heart rate extraction pipeline produces valid estimates for genuine videos by comparing the estimated results with ECG reference data. Our high-quality deepfakes exhibit valid heart rates and their rPPG signals show a significant correlation with the corresponding driver video that was used to generate them. Furthermore, we show that this also holds for deepfakes from a publicly available dataset.

Discussion: Previous research assumed that the subtle heart-beat-related signals get lost during the deepfake generation process, making them useful for deepfake detection. However, this paper shows that this assumption is no longer valid for current deepfake methods. Nevertheless, preliminary experiments indicate that analyzing spatial distribution of bloodflow regarding its plausibility can still help to detect high quality deepfakes.

KEYWORDS

deepfakes, video forensics, remote photoplethysmography (rPPG), biological signals, remote heart rate estimation, imaging photoplethysmography (IPPG)

1 Introduction

In recent years, deepfakes have emerged as a prominent and concerning phenomenon. Notably, political figures such as Barack Obama, Donald Trump, and Wladimir Klitschko have become targets, drawing significant public attention. The societal and ethical implications of deepfake technology have become increasingly evident. Initial examples were characterized by visible artifacts, particularly when static images were synthesized into video sequences (DeepFakes, 2019). However, advancements in image generation techniques have significantly improved the realism of these manipulations, making it increasingly difficult to detect alterations through visual inspection alone (Ramesh et al., 2021; Karras et al., 2020).

Modern state-of-the-art deepfake detection approaches rely on features learned by convolutional filters sensitive to inconsistencies in both the spatial and the temporal domain (Wang et al., 2023; Haliassos et al., 2022). Despite receiving outstanding performance results on benchmark datasets, these techniques suffer from a lack of explainability. This weakness becomes critical when human supervisors of video-identification systems face potential misclassifications by these detectors, leading to challenges due to their opaque decision-making processes.

However, contemporary deepfake generation techniques, while increasingly sophisticated in their ability to visually mimic real individuals, do not explicitly model physiological signals present in genuine videos. The cardiovascular pulse, inducing individual pulsating blood flow in human skin, causes subtle color variations that are assumed to be plausible in genuine videos only. This inadequacy has been employed by several researchers for leveraging locally resolved signals, such as those captured by techniques like remote photoplethysmography (rPPG), which capture these subtle variations (Yu et al., 2021a). For example, rPPG can extract physiological information, such as pulse rate, from a recorded video, providing valuable data for deepfake detection (Kossack et al., 2019a). Traditional approaches have primarily focused on extracting a global pulse signal from an entire video sequence (Yu et al., 2021a). Detectors leveraging this global rPPG signal have demonstrated promising results concluding that deepfakes do not include such physiologically induced signals. However, contrary findings indicate that deepfakes can indeed exhibit a one-dimensional signal resembling a heart rate (HR), further complicating the detection process (Fernandes et al., 2019). Additionally, recent advancements in synthetic face generation explicitly incorporate pulsation signals (Ciftci and Yin, 2019) or enable the manipulation of physiological signals in facial videos (Chen et al., 2022), thus blurring the distinction between real and fake rPPG signals. It is also important to note that rPPG-based deepfake detectors may inadvertently rely on non-physiological cues, such as background artifacts, noise, or comparisons between image pairs, rather than purely detecting pulse-related color changes in the skin (Ciftci et al., 2024; Qi et al., 2020; Ciftci et al., 2020b,a). For instance, Ciftci et al. (2020a) demonstrated that filtering rPPG signals with a bandpass filter between 4.68 Hz to 15 Hz (i.e., 180 bpm to 900 bpm), can more effectively distinguish real videos from deepfakes compared to filtering signals based on human heart rate frequencies. This highlights a critical limitation in current deepfake detection approaches that rely on rPPG signals, as they often fail to account for the fact that deepfakes can still produce realistic HR signals.

In this article, we demonstrate that HR signals can indeed be derived from deepfake videos, and, more importantly, these signals closely match those of the original driving video, which define the head motion and facial expressions. This finding challenges the assumption that deepfakes inherently lack valid physiological signals and emphasizes the need for detection methods that go beyond simple pulse detection. Our contribution provides new insights into the physiological consistency of deepfakes, raising the bar for future detection techniques.

To validate our findings, we propose a pipeline that extracts the pulse rate from videos while incorporating motion compensation and background noise reduction for enhanced robustness. To further substantiate our approach, we collected a dataset consisting

of video recordings synchronized with electrocardiogram (ECG) data. Our experiments demonstrate that the HRs extracted from the videos using our pipeline closely align with those from the ECG signal, confirming the accuracy of the rPPG-based extraction process. To explore the origin of the heart beats detected in the rPPG signals of the deepfake videos, we generated a set of deepfakes based on these original video recordings. In our experiments, we show that the HRs derived from the deepfakes significantly overlap with those of the source (or “driver”) videos, highlighting that deepfake HR signals are not random but rather reflect the physiological information present in the driving video. Furthermore, we extend our analysis to older generations of deepfakes by utilizing the publicly available KoDF dataset (Kwon et al., 2021), where we similarly demonstrate the presence of valid HR signals. These results emphasize that even older deepfake methods can carry realistic physiological signals, further complicating traditional detection methods.

The remainder of this paper is organized as follows: In Section 2, we provide an overview of existing work on deepfake generation, deepfake detection, and rPPG. Our proposed method is presented in Section 3. Section 4 outlines the experiments conducted, along with the presentation of our used dataset and results. Thereafter, we discuss our method’s limitations and conclude our paper with a summary of our results and findings in Section 6.

2 Related work

2.1 Deepfakes

Deepfakes represent a category of manipulated videos and audio files created through deep learning techniques. These manipulations involve altering faces, modifying gestures and facial expressions, and adjusting physical appearances and mouth movements to align with manipulated audio content. The widespread popularity of deepfakes is evident in various applications, with common usage found in AI-based face swapping techniques. Notably, there is a surge in popularity with smartphone apps that facilitate seamless face swapping, demonstrating the accessibility and user-friendly nature of these technologies. These apps leverage advanced voice synthesis, facial synthesis, and video generation methods to produce convincing and often deceptive content.

The development of GANs (Goodfellow et al., 2014), VAEs (Kingma and Welling, 2014) and, lately, diffusion models (Ho et al., 2020) enabled various possibilities for the forgery of digital content. The seminal deepfake generation method utilizes a dual-decoder autoencoder, with each decoder dedicated to one of the set target identities for swapping (DeepFakes, 2019). Subsequently, this foundational method has been enhanced by the integration of adversarial training, application of more sophisticated convolutional neural networks or advanced blending techniques (Perov et al., 2020; Beckmann et al., 2023). Numerous methods have been developed for manipulating face expressions and appearances, with modern approaches capable of synthesizing a face with a given appearance and an expression of choice in the one-shot scenario (Drobyshev et al., 2022; Nirkin et al., 2022; Wang et al., 2021b,a). Recently, several approaches leverage denoising

diffusion models for the generation and manipulation of high-quality face images (Ho et al., 2020; Zhao et al., 2023; Ding et al., 2023; Huang et al., 2023). This continuous evolution of deepfake technologies poses challenges for content authentication and necessitates the development of robust detection mechanisms.

Early approaches to detect deepfakes exploit physical inconsistencies in the behaviour and appearance of the head. Li et al. (2018) exploit the fact that early deepfake generation approaches merely use training images with opened eyes, by utilizing facial landmarks to identify the eye-blinking behaviour in videos. In Yang et al. (2019), the authors take advantage of the fact that the process of cropping, aligning and inserting a face onto another head leads to a misalignment of the attributes in the inner face and the head pose. Other approaches aim to manually generate fake training images by simulating the artifacts introduced by warping or blending operations in genuine images (Li and Lyu, 2018; Li et al., 2020). With the rapid increase of quality in visual fake content, research focus shifted from more obvious and explainable artifacts to high dimensional complex convolutional feature maps. In the foundational FaceForensics++ (FF++) paper (Rössler et al., 2019), the authors propose a benchmark dataset for the evaluation of deepfake detectors and analyze the detection performance of several CNN based detectors.

While recent and ongoing works on generating better deepfakes focus mostly on making them look more realistic and appealing, the coherence of biological rPPG signals is not considered. This motivated several researches to work on the promising line of fake detection methods, analyzing the coherence of biological rPPG signals in the spatial and temporal domain and thereby increasing the explainability of the detection process (Ciftci et al., 2020a; Hernandez-Ortega et al., 2020). FakeCatcher (Ciftci et al., 2020a) extracts rPPG signals from three face regions which are subject to various signal transformations. Moreover, the extracted signals are consolidated into image-like PPG maps, which represent the temporal and spatial distribution of biological signals across the analyzed facial regions. Those signal maps are then fed to a CNN for classification. DeepFakeON-Phys (Hernandez-Ortega et al., 2020) adapts the heart rate estimation method proposed in Chen and McDuff (2018) and modifies it through the usage of a two branch convolutional attention network to assess both appearance and motion related information for deepfake video detection. In Wu et al. (2023), the authors propose the usage of a temporal transformer in combination with a mask-guided local attention module in order to capture spatial and temporal inconsistencies over long distances in the used PPG maps. Detection methods that specifically pay attention to the heart rate (HR) information extracted from rPPG were proposed in Ciftci et al. (2020b) and Boccignone et al. (2022).

2.2 rPPG

The extraction of human vital signs from face videos is a rapidly growing and emerging field with numerous recent publications (Poh et al., 2010; De Haan and Jeanne, 2013; Wang et al., 2017; Tulyakov et al., 2016). The medical measurement of the HR typically relies on the optical measuring technique known as photoplethysmography (PPG) (Zaunseder et al., 2018). This

technique capitalizes on human blood circulation, where blood's light absorption exceeds that of surrounding tissue. Consequently, variations in blood volume influence light transmission or reflectance accordingly (Tamura et al., 2014). A PPG sensor, commonly used for measuring the human pulse rate, is placed directly on the skin to optically detect changes in blood volume (Tamura et al., 2014). Remote photoplethysmography employs the same principle, allowing for contactless HR measurements using a standard RGB camera (Zaunseder et al., 2018). In this technique, the continuous change in skin color, resulting from blood flow through the circulatory system, is analyzed by rPPG methods to determine HR (Poh et al., 2010; De Haan and Jeanne, 2013; Wang et al., 2017; Tulyakov et al., 2016).

To robustly extract an rPPG signal, irrespective of the subject's skin tone and non-white illumination conditions, the Plane-Orthogonal-to-Skin Transformation (POS) (Wang et al., 2017) of the rPPG signal has been developed for pre-processing the input video sequence.

Given that global model-based methods may be susceptible to noise, compression artifacts, or masking, recent rPPG-related publications leverage deep neural networks for HR extraction from video data (Chen and McDuff, 2018; Yu et al., 2019, 2020). Yang et al. (2021) conducted a comparative study of three neural networks [Deepphys (Chen and McDuff, 2018), rPPGNet (Yu et al., 2019), and Physnet (Yu et al., 2020)] against model-based approaches [independent component analysis (ICA) (Poh et al., 2010), CHROM (De Haan and Jeanne, 2013), and POS (Wang et al., 2017)] using the publicly available UBFC-rPPG dataset (Bobbia et al., 2019). In these experiments, under constant lighting conditions, deep-learning-based approaches outperformed model-based ones. However, model-based approaches (ICA, CHROM, and POS) exhibited more accurate and robust results in varying lighting conditions (Yang et al., 2021).

The locally analyzed rPPG signal extracted from videos is visualized based on amplitude, velocity, or signal-to-noise ratio (SNR) maps (Kossack et al., 2019b; Yang Jun, Guthier B, 2015; Zaunseder et al., 2018). Particularly, blood flow in facial videos has been scrutinized (Yang Jun, Guthier B, 2015; Kossack et al., 2019b,a), where blood flow velocity is calculated from the relative phase shift of the frequency component corresponding to HR in the frequency domain. These methods assume that the difference between neighboring phase values directly corresponds to the velocity at that point.

Beyond medical applications (Schraven et al., 2023; Kossack et al., 2023), rPPG analysis has also been employed to detect presentation attacks on authentication systems (Kossack et al., 2022). In multiple studies, rPPG methods are applied to facial videos to discern whether the face is covered by a mask (Li et al., 2017; Kossack et al., 2019a; Yu et al., 2021b). However, deepfake detection proposes another challenge, and as described in Section 2.1, discrepancies between images resulting from deepfake generation disrupt the natural color variations in the skin induced by the heartbeat.

3 Methods and data

We propose a pipeline for extracting and analyzing physiologically related signals, specifically focusing on those

associated with the cardiovascular cycle, which typically occur in the frequency range of 0.7 Hz to 3 Hz. To ensure the accurate detection of these signals, the pipeline requires an input video showing the face of a single person for at least 10 s. The proposed pipeline incorporates motion compensation techniques and accounts for frequencies introduced by external factors, such as compression or camera properties, to ensure a robust extraction of physiologically related signals. The details of these components are discussed in the following two sections.

Following the components of our pipeline, we describe the data used for the experiments. This includes the dataset of videos and ECG data that we captured, the method used to generate deepfakes and finally an external dataset that was used for evaluation.

3.1 Reference face and temporal alignment

We focus on global rPPG signals over time, specifically the averaged color changes across various spatial positions on the facial surface. To ensure accurate signal extraction, it is essential to compensate for any movements made by the person in the video. To achieve this, each frame of the input video is aligned with a reference face by detecting facial landmarks using MediaPipe (Google, 2022). These landmarks form the basis for Delaunay triangulation (de Berg et al., 2008), generating a 2D mesh over the facial region. The reference 2D mesh consists of 918 triangles, and serves as a foundation for tracking and stabilizing the facial movements across the video. While this approach is easy to implement, it does not consider motion blur, and the accuracy of the registration is constrained by the facial landmarks. To further enhance this important process, we will extend our method in the future by using the approach proposed by Seibold et al. (2017) for removing motion blur and Seibold et al. (2024) for a pixel-wise registration.

In each input frame, we track the detected facial landmarks and use the 2D mesh to warp each triangle to its corresponding reference position. This warping process aligns the facial features from the input video to the reference face, as illustrated in Figure 1. The outcome is a motion-compensated image sequence that serves as the foundation for our subsequent analysis, ensuring that the extracted rPPG signals are not affected by facial movements.

3.2 Encoding of heart rate related features

To extract heart rates, we perform a global analysis of the entire video to obtain a single robust reference rPPG signal, which is associated with the subject's pulse signal (Kossack et al., 2021). For rPPG calculation, we apply the Plane-to-Orthogonal skin (POS) transformation (Wang et al., 2017) on a 10 s window orange, i.e., including 300 frames. A preliminary analysis about the optimal window length showed a standard deviation of the differences between extracted HR and ground truth for the 10 s window of 1.39 bpm, increasing to 3.38 bpm, 3.76 bpm, and 4.15 bpm for 8 s, 6 s, and 5 s windows, respectively. The entire video is processed by sliding this window over the video duration with a step size of one frame, ensuring continuous analysis and accurate pulse signal extraction.

After processing the entire video sequence, the output signal is normalized and filtered with a fifth-order Butterworth digital band-pass filter with a frequency range between 0.7 Hz and 3.0 Hz (corresponding to a HR of 42 bpm to 180 bpm). This filtering produces the final rPPG signals. For each time step, we then transform rPPG signal from the time domain to the frequency domain using a fast Fourier transform (FFT), mapping the signal magnitudes across all time steps for further analysis.

This processing is applied to the entire face region to capture physiologically relevant signals, as well as to two homogeneous square regions in the background to collect image noise information, see Figure 2. The two background regions are averaged during transformation, resulting in a single FFT map for the background. The two FFT maps - one from the face and one from the background - are then subtracted based on their intensities to generate a background-free FFT map that focuses on the physiologically induced signals.

To determine the HR signal over the entire video duration, the highest magnitude in the subtraction FFT map is identified, and based on this peak, the HR for each time instant is extracted through a single optimization step.

3.3 Captured dataset

Given that many of the most popular datasets for deepfake analysis are several years old and deepfake generation techniques have advanced significantly, we created a fully controlled, high-quality dataset to ensure optimal compression and realism. To validate the functionality of our method, we collected recordings of twelve individuals, representing diverse genders, ages and ethnic background in a controlled studio environment. The recordings were captured with participants positioned in front of a white background, under uniform lighting provided by white LED illumination. For each participant, 10-20 frontal view recordings were taken, with the head centered throughout the video. During each recording, participants were asked to perform a range of activities, including talking, reading, and interacting with the recording supervisor. All participants provided written consent for the use of their recordings in this experiment and its subsequent publication.

We used an industry RGB camera¹ to capture the video recordings. The recordings vary in length, ranging from 10 s up to several minutes, with a frame-rate of 25 fps and a resolution of 2448×2048 pixels. In addition to the RGB video, we measured the ECG and PPG of selected subjects. These physiological signals were used to calculate the heart rate (HR) as ground truth for our analysis. Selected frames from our dataset are shown in Figure 3A.

3.4 Creation of high-quality deepfakes

Publicly available datasets have not kept pace with the rapid development in deepfake technology as new techniques and architectures continuously emerge, leading to increasingly realistic and higher-quality deepfakes. This progress likely impacts previous

¹ ace acA2440-75uc, Basler AG, Germany.

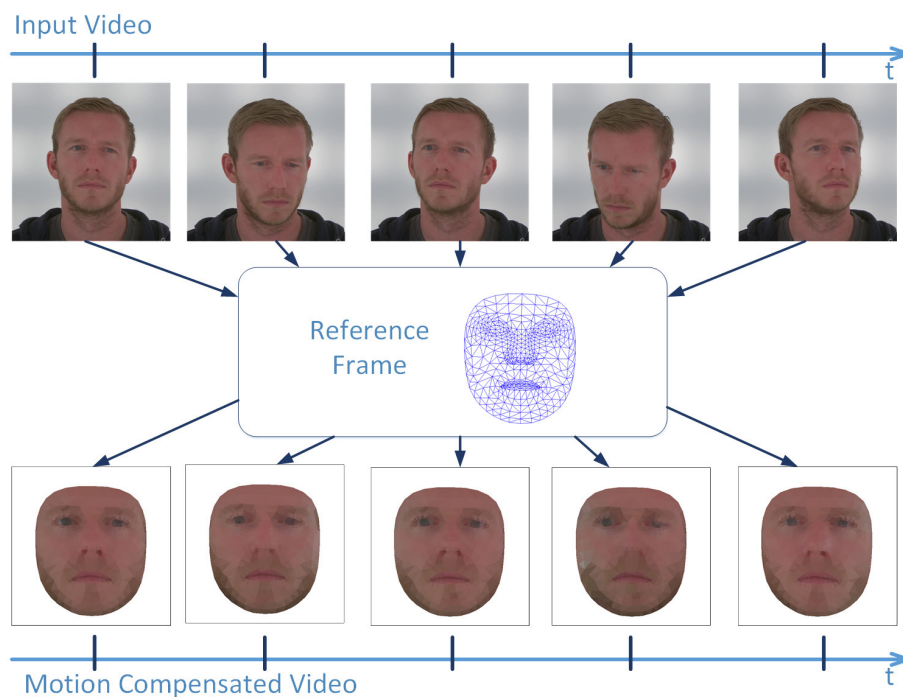


FIGURE 1

Illustration of the temporal alignment process. A reference mesh, composed of 918 triangles formed from MediaPipe facial landmarks, (**center**) is used to spatially warp each frame from the video sequence (**top**) to a reference position (**bottom**).

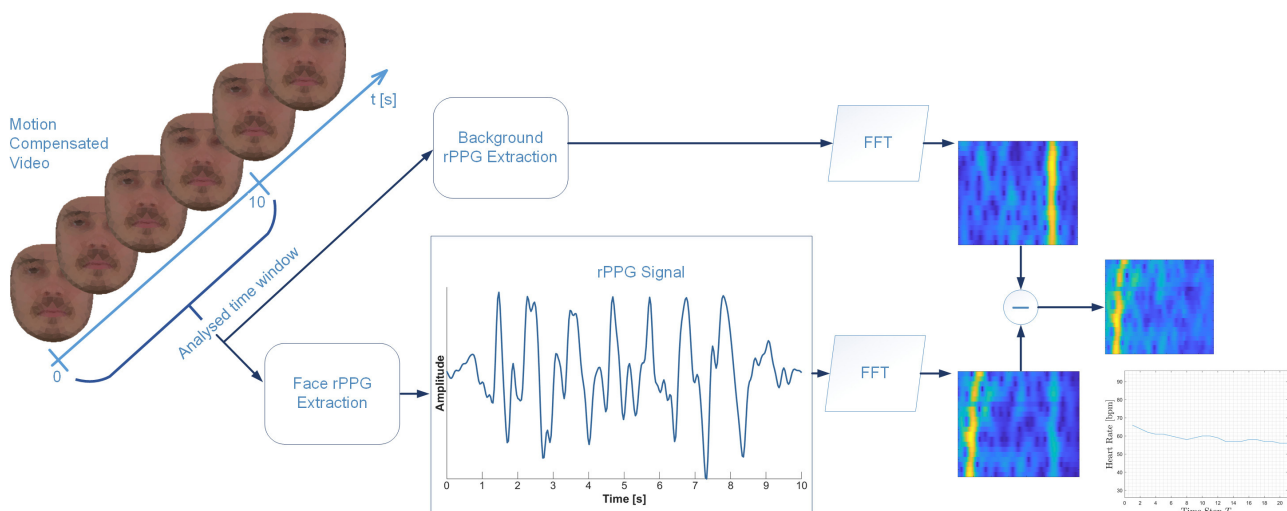


FIGURE 2

Heart rate extraction pipeline. From the registered video sequence, we calculate a global rPPG signal of the face as well as the background. Following, we determine the magnitudes in frequency space for each signal over time. To robustly extract the heart rate both signals are "subtracted".

assumptions about deepfakes, particularly the notion that they do not contain HR-related signals. As deepfake generation methods improve, it becomes necessary to reassess these conclusions in light of more sophisticated and physiologically accurate manipulations.

To generate our own set of high-quality deepfakes, we employed a dual-decoder autoencoder architecture along with an advanced blending procedure, as described in Beckmann et al. (2023). Unlike a standard autoencoder with a single decoder to

reconstruct the input image, this model utilizes two decoders. Each decoder is trained to reconstruct the input image but with the identity of a specific person respectively, the source and target person. During training, the autoencoder is fed with pairs of images of the source and target person. Once trained, the model can be used to swap faces in images and, accordingly, in videos. The advanced blending procedure enhances quality of the deepfakes by modifying the mask used for blending. Specifically, the mask is

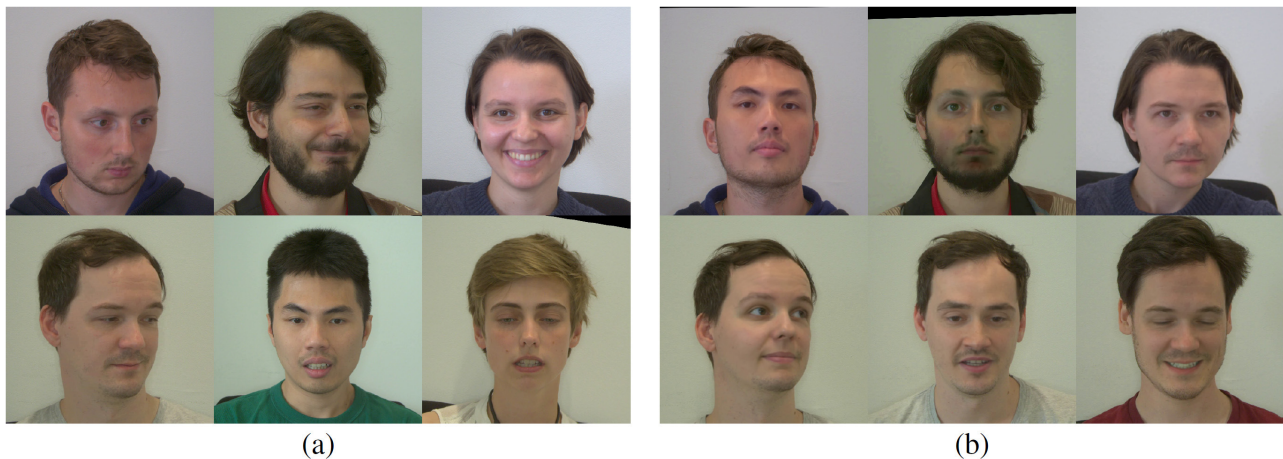


FIGURE 3

A subset of our recorded data representing six participants (A) and a correlating subset of generated deepfakes (B).

adjusted to create a greater distance between the edges of the face and the boundaries of the mask by “squeezing” it by approximately 15 pixels on each side. This adjustment excludes non-facial regions from the blending process, thereby reducing blending artifacts at the boundaries and improving the overall realism of the generated deepfakes.

Following data collection, we created various identity pairs and trained a separate deepfake autoencoder for each pair. Using these autoencoders, we performed face swaps between all videos for each identity pair, generating a total of 858 identity-specific deepfake videos and 156 unaltered counterparts. Figure 3B shows examples of our deepfakes. For more details on the used method see Beckmann et al. (2023).

In addition to these deep fakes, we generated additional ones using the open-source tool DeepFaceLive (DFL) (Petrov, 2023). This tool was developed for real-time face swapping. It requires a driver video and swaps the face in the video with that of a target face model, while maintaining the driver’s expression and head pose. A set of target face models is provided by the tool. We used four of these provided face models to generate 32 deep fake videos. These videos are used in our experiments to show that the rPPG signal of a deep fake is similar to that of its driver video and also to our fakes generated using the same driver video.

3.5 External data

In addition to our own dataset, which includes videos with ECG data and corresponding deepfakes, we also utilized publicly available datasets to enhance the scope of our analysis. First, we used the deepfakes generated in Beckmann et al. (2023) based on the “actors” subset of the deepfake detection dataset (Dufour et al., 2019), its corresponding originals as well as the fakes from that dataset based on the same originals.

Recognizing that many existing deepfake datasets may have limitations in terms of size and diversity, we selected the KoDF dataset (Kwon et al., 2021), which is designed generalize more

effectively to real-world deepfakes compared to other public datasets like FF++ (Rössler et al., 2019) or Celeb-DF (Dang-Nguyen et al., 2020). KoDF contains 403 Korean subjects and a few ten-thousands of real and fake videos. In addition, KoDF includes six synthesis models for deepfake creation, which brings a large diversity of fakes to the set; in our study, we utilized four of these six methods due to fake quality.

Finally, we selected 45 videos from the KoDF dataset and generated an additional 45 Deepfake videos using the PicSi.AI platform², leveraging its available synthesis methods.

4 Results

4.1 Signal analysis

In the initial phase of our analysis, we focused on our own dataset, where we successfully extracted meaningful heart rate (HR) signals from both genuine and deepfake videos. In all cases, the detected HR corresponded to the face of the subject in the video, regardless of whether the video was real or a deepfake.

The average signal-to-noise ratio (SNR) of the extracted HR was significantly higher in the original videos compared to the deepfake videos with values of -1.97 dB for genuine and -3.35 dB for deepfakes. This difference in SNR highlights the lower quality of rPPG signals in deepfakes, likely due to artifacts introduced during the generation process. As all participants were seated during the recordings and made only slight movements, it is reasonable to assume that the resting heart rate (normally between 60 bpm to 90 bpm) was detected for all participants. A higher HR was only measured for two participants, but this was consistent across all recordings and verified by the ECG measurements, suggesting the reliability of our extraction process. The results of our analysis on four videos, two fake and two genuine, are exemplary depicted in Figure 4.

For the videos with a captured PPG reference signal and deepfakes based on these videos, we further analysed the rPPG in

² <https://www.picsi.ai>

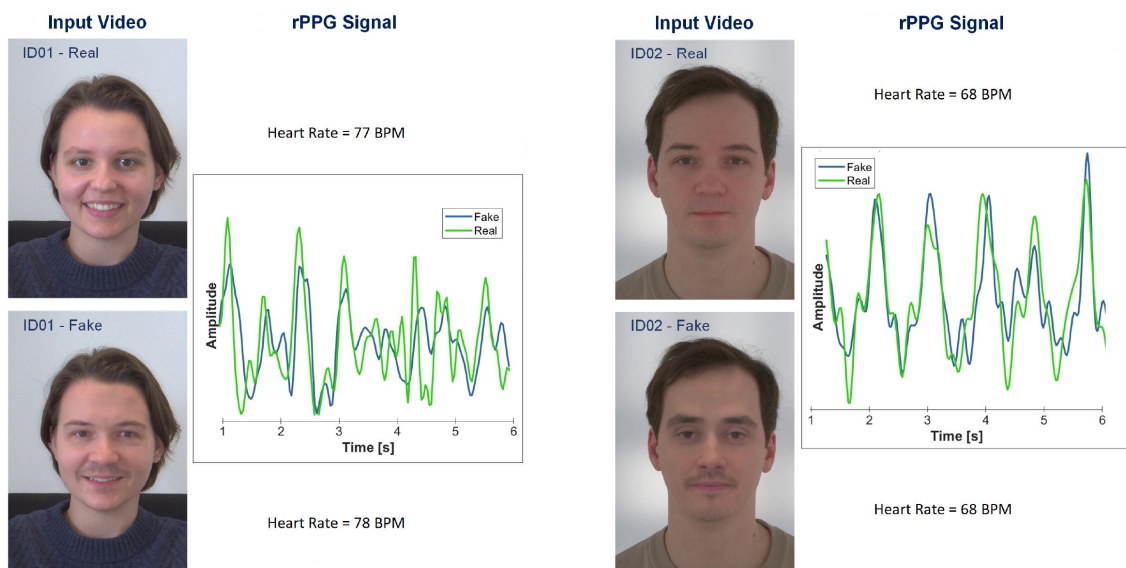


FIGURE 4

This illustration presents two pairs of genuine and fake videos. On the **left** of each example, frames from each video sequence are displayed. On the **right**, the extracted reference rPPG signal is plotted for each paired fake and original video. Additionally, the measured heart rate of the person recorded is displayed.

time domain. We calculated the Pearson correlation coefficients for the PPG and rPPG signals for the genuine videos. Since no ground truth PPG signal can be measured for the deepfakes, the signal from the underlying driver video is used instead. In addition to the correlation, we calculated the Mean Squared Error (MSE) for the rPPG signals, using the PPG signals as ground truth. Before calculating the MSE, the mean and variance of both signals were normalized to zero and one, respectively. The results are shown in [Figure 5](#).

For all types of videos, there is a moderate to strong correlation in most samples. The correlation between the PPG and rPPG signals for the genuine videos and the DeepFaceLive (DFL) fakes shows a similar distribution, while the correlation for deepfakes generated with the method of [Beckmann et al. \(2023\)](#) is slightly lower. These high correlations between the rPPG signals of the deepfakes and the ground truth PPG signal of the driver videos show that these fakes replicate the rPPG signal of the driver videos. This point is further supported by the HR gained from the rPPG signal. The absolute difference to the ground truth across all videos and time periods is on average 1.80 bpm, 1.85 bpm and 3.24 bpm for the genuine videos, the [Beckmann et al. \(2023\)](#) and DFL fakes, respectively. It should be noted that rPPG signal extraction from videos includes, alongside the PPG-related signal, additional components induced by body motion and other noise sources, and thus cannot perfectly reflect a true PPG signal.

In addition to comparisons with ground truth PPG signals, we calculated the Pearson correlation coefficients between these deepfakes and their underlying driver videos. The results are shown in [Figure 6](#). While the correlation for videos generated with DFL is strong in most cases, the correlation for those generated with the method of [Beckmann et al. \(2023\)](#) is moderate for most videos. This

provides further evidence that deepfakes mimic the rPPG signal of the driver video.

Building on these results, we further analyzed the generated deepfakes to investigate the origin of their rPPG signals. In the majority of cases, the rPPG signals in the deepfakes closely mirrored those of the original source videos, with only minor variations observed. When comparing the HR measured in the genuine videos with those from their deepfake counterparts, we found that the global HR in the deepfake videos was remarkably similar to the HR of the original source recordings, as well as to the measured ECG ground truth, see examples in [Figure 7](#). For all fakes in our dataset, we found a high correlation to the HR of the original driving video, see [Figure 8](#). The average correlation between the HR of the fakes created by using the method of [Beckmann et al. \(2023\)](#) is $\bar{r} = 0.57$ (median $r = 0.55$) and for the fakes generated with DFL $\bar{r} = 0.82$ (median $r = 0.89$). For the other fakes (KoDF dataset, both methods on actor subset), the correlation is above $r > 0.4$. However, for the public available FF++ fakes, the deviation is remarkably high (min $r = -0.23$ to max $r = 0.91$). These findings confirm that the heart rate signals in high-quality deepfakes are often inherited from the source video, further complicating the task of distinguishing between real and fake content based solely on global HR analysis.

The FFT maps visualize that the rPPG signal, which can be traced back to physiological properties, clearly originates from the source video. [Figures 9, 10](#) show two examples with a set of six FFT maps, the background, face and subtraction FFT maps of an original video and an deepfake, where the original served as source. The extracted HRs for both examples can be found in [Figure 7](#).

Example *ID_04_0100* demonstrates the influence of our proposed background analysis on signal detection. In this instance, a strong noise signal around 150 bpm is detectable in the

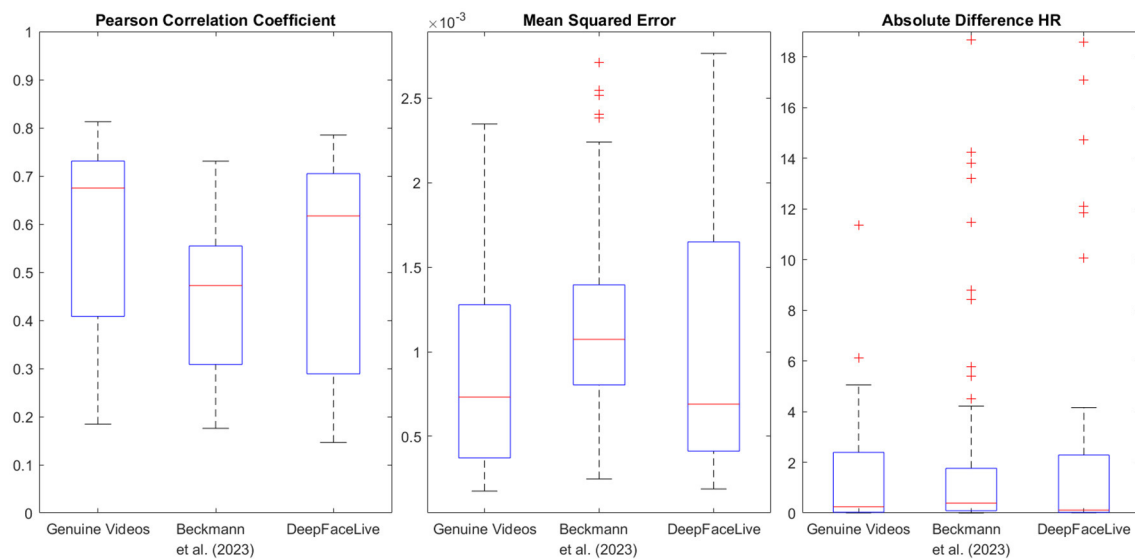


FIGURE 5

Correlation and deviation of rPPG to PPG signal as well as the absolute difference of the heart rate (HR) between detected and ground truth. The rPPG signal of the deep fake videos generated with DeepFaceLive (DFL) shows a similarly strong correlation to the measured PPG signal as the rPPG signal for the genuine videos. The correlation for the rPPG signal of the deep fake videos generated with the methods of (Beckmann et al., 2023) is slightly weaker but still moderate. The MSE is in a similar range for all types of videos.

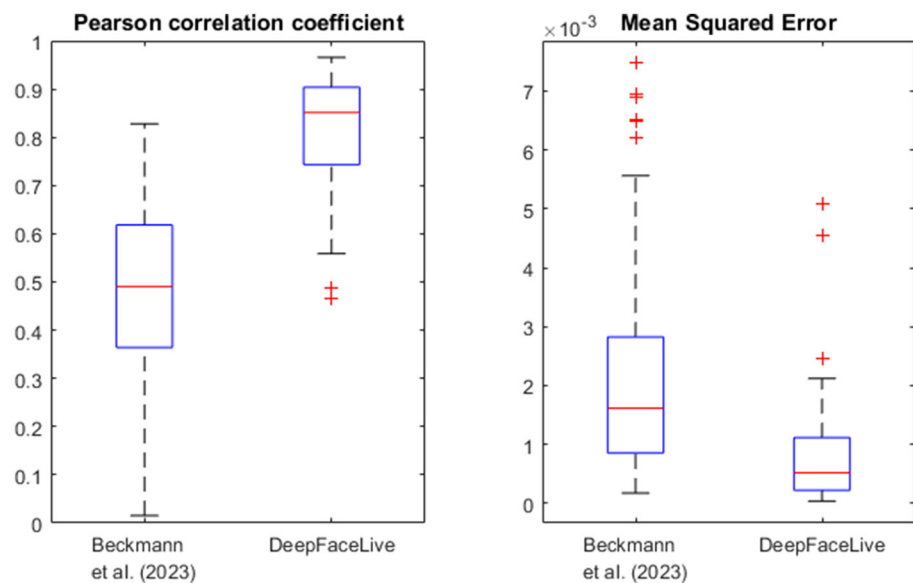


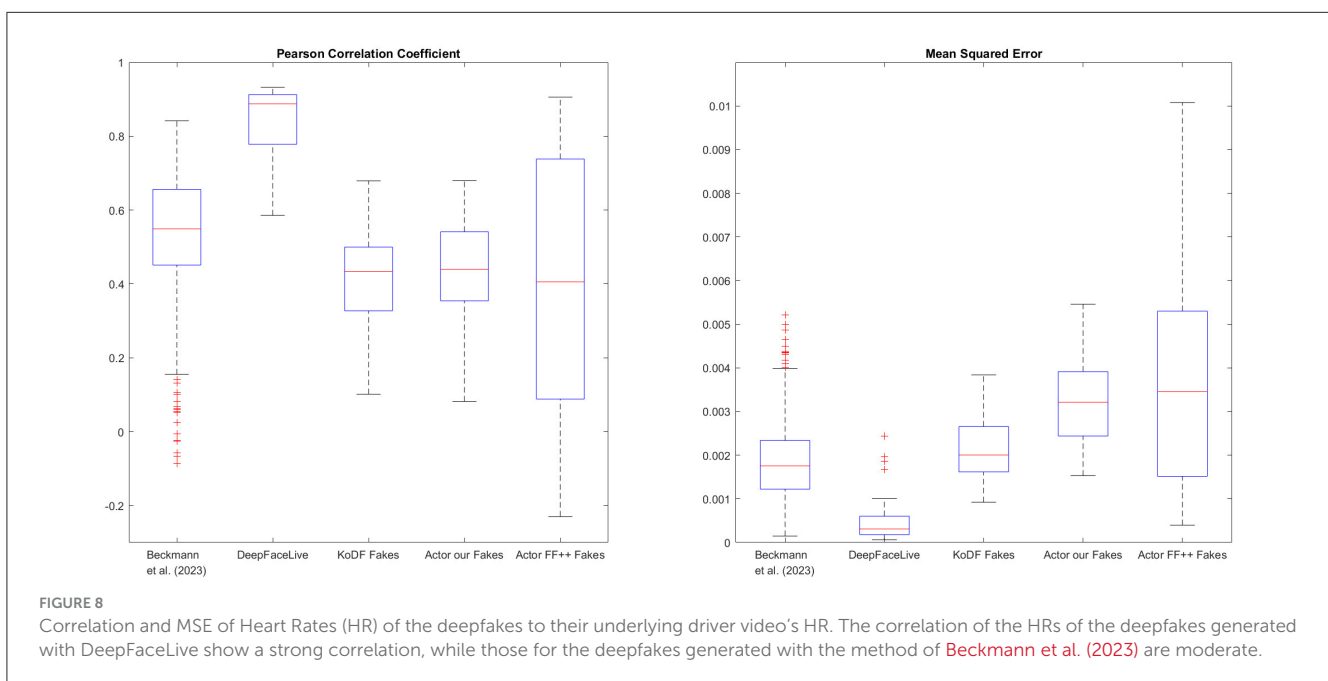
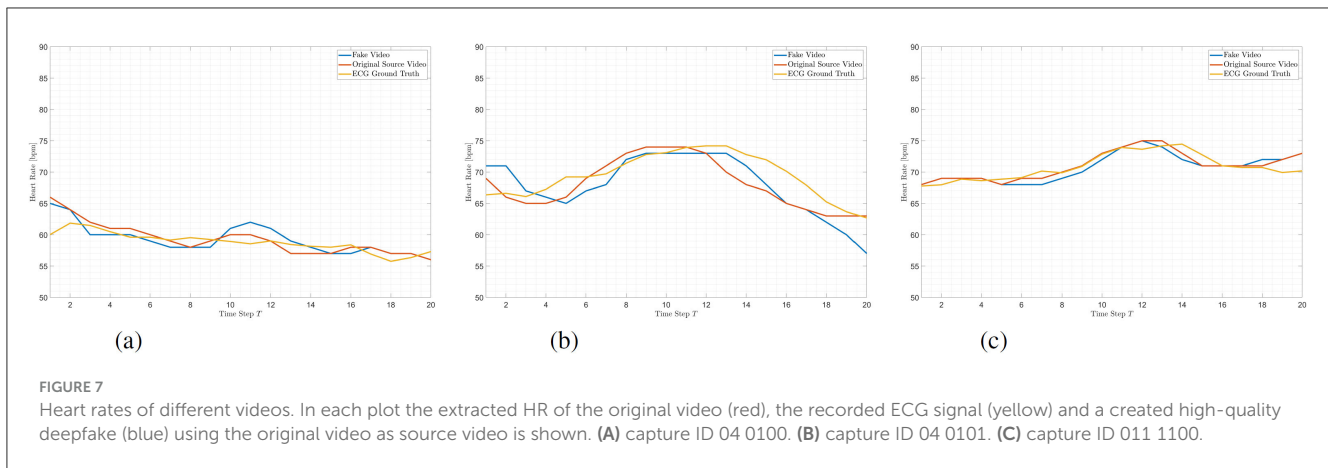
FIGURE 6

Correlation and deviation of rPPG signals of deepfakes to their underlying driver video's rPPG signal. The rPPG signals of the deepfakes generated with DeepFaceLive (DFL) show a strong correlation to the rPPG signal of the underlying driver videos in most cases, while those for the deepfakes generated with the method of (Beckmann et al., 2023) are weaker but, on average, still moderate. The DFL deepfakes outperform also in terms of MSE those of (Beckmann et al., 2023).

background. Due to the nature of deepfake generation, this noise signal is also present in all fakes where that capture served as source, resulting in a high correlation between the FFT maps of original and deepfakes (with a correlation of 0.96). In the original face video, the physiological signal (at about 59 bpm) is twice as strong as the background noise signal (at 150 bpm), making it easy to extract. However, in the deepfake face, the HR and noise signals are of

comparable magnitudes, complicating clear pulse extraction. This issue is resolved by incorporating background analysis, as shown in Figure 9.

The correlation between the original and deepfake FFT maps increases slightly, 0.7667 for the face FFT maps to 0.7826 for the subtraction maps, further emphasizing that the rPPG signal in the deepfake originates from the source video. This strong relationship



between the original and deepfake signals extends to cases where the background signals in the deepfakes differ more significantly from the originals, reinforcing the notion that deepfakes inherit their rPPG signals from the driver video (cf. Figure 10).

Both examples clearly demonstrate that, in the analysis of the face region in deepfakes, the background signal (induced by noise, compression, etc.) plays a significantly stronger role, as the transferred HR signal is reproduced with less intensity compared to the original video. This is also reflected in the corresponding SNRs (see above). Due to the weaker transmission and artificial replication of the pulse signal, a strong correlation between the original and deepfake signal is not always observed, see Figure 11 as example for a moderate relationship between original source and deepfake subtraction maps with a correlation of 0.53. However, upon closer examination, a trace of the original video's HR signal can still be detected in the faked face. This 'signal trace' underscores that, despite noise and degradation, elements of the physiological signal from the source video remain present in the deepfake.

4.2 Analysis on external data

Given the limited size of our dataset, we extended of HR analysis to the KoDF and FF++ dataset. Despite varying compression rates and relatively high image noise, we were able to consistently extract HR signals from all genuine videos. Although some deepfake videos presented challenges due to noise and compression artifacts, we were still able to extract signals in most cases that could be associated with HR (cf. Figure 8). However, as the datasets do not include the participants' actual HR data, we were unable to validate these extracted HR signals against ground truth measurements.

A closer look to quality parameter (Table 1) shows a extremely low signal-to-noise ratio (SNR) of the extracted HRs for the external datasets, especially for FF++, while the deviation of the HR over time is high although all videos involve individuals who are at rest and should therefore have a stable pulse. This indicates that for a certain amount of videos, the detected HR is not plausible,

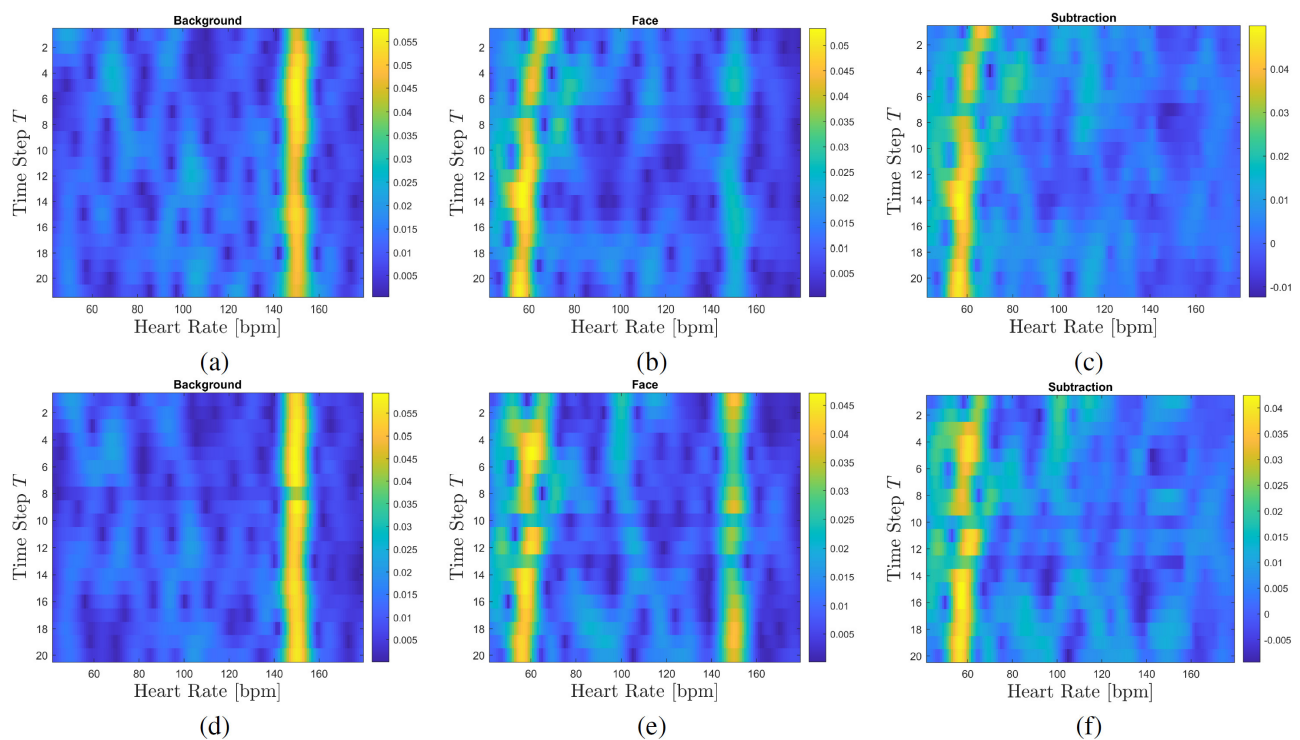


FIGURE 9

FFT maps of capturing *ID_04_0100*. A similar HR can be detected in both cases, original source (A–C) and deepfake (D–F) video with about 59 bpm. The correlations between the original and deepfake FFT maps shows a strong relationship for all three map pairs: 0.96 for background, 0.77 for face, and 0.78 for subtraction map.

i.e., not related to the real HR, although we have selected the videos with best signal quality and analyzed the corresponding deepfakes. It is important to note, that it is not possible for the KoDF dataset to identify the exact driver video used for each fake. Therefore, we only looked at whether it is possible to identify a physiologically meaningful HR in both the original and the fake videos. Here, similar results as with our dataset could be achieved (cf. Figure 12). For the deepfakes, a signal which can be related to a HR is in most cases detectable. For further examples of FFT maps of deepfakes from the KoDF dataset are shown in Appendix.

5 Discussion

As discussed in Section 3.5, numerous datasets have been developed to support deepfake research, such as the DeepFake Detection Challenge Dataset (Dolhansky et al., 2020), FF++ (Rössler et al., 2019), and Celeb-DF (Dang-Nguyen et al., 2020). These datasets have significantly advanced deepfake detection techniques. However, few authors have explored deepfake detection through the analysis of physiologically related signals, such as rPPG. Despite their importance, public datasets present several challenges when used for analyzing rPPG signals in the context of deepfake detection as rPPG is sensitive to video quality.

Many deepfake datasets suffer from compression artifacts, low resolution, inconsistent frame rates, high background noise, and challenging illumination settings (D'Amelio et al., 2023; Kwon et al., 2021). These factors can substantially degrade the quality of

rPPG signals, making it difficult to reliably extract physiological features (Wang et al., 2024; Zaunseder et al., 2018; McDuff et al., 2017). Consequently, the utility of rPPG analysis in deepfake detection has been limited, particularly in datasets where video quality is compromised.

Previous studies (Çiftçi et al., 2024; Qi et al., 2020; Ciftci et al., 2020b,a; Hernandez-Ortega et al., 2020) concluded that deepfakes do not exhibit a detectable heartbeat (Boccignone et al., 2022), suggesting that this could be used as a reliable marker for deepfake detection. However, much of this research was conducted on datasets of low image quality. In contrast, our study reveals that for recent and high-quality deepfakes, such as those generated using the method described in Beckmann et al. (2023), DeepFaceLive or present in the KoDF dataset, it is possible to robustly detect a HR signal that originates from the source (driver) video. Our experiments demonstrated that deepfakes can exhibit realistic heart rates, contradicting previous findings. Specifically, in all fake videos from our dataset and most videos from the KoDF dataset, valid HR signals were successfully extracted. This indicates that solely relying on the analysis of global HR signals is no longer sufficient to detect deepfakes.

Another significant challenge in existing deepfake datasets is the lack of reference measurements, such as concurrent ECG or PPG sensor readings, which are crucial for validating the accuracy of extracted rPPG signals. Without these ground truth data, it becomes difficult to assess the reliability of physiological signal extraction and, consequently, the conclusions drawn from them.

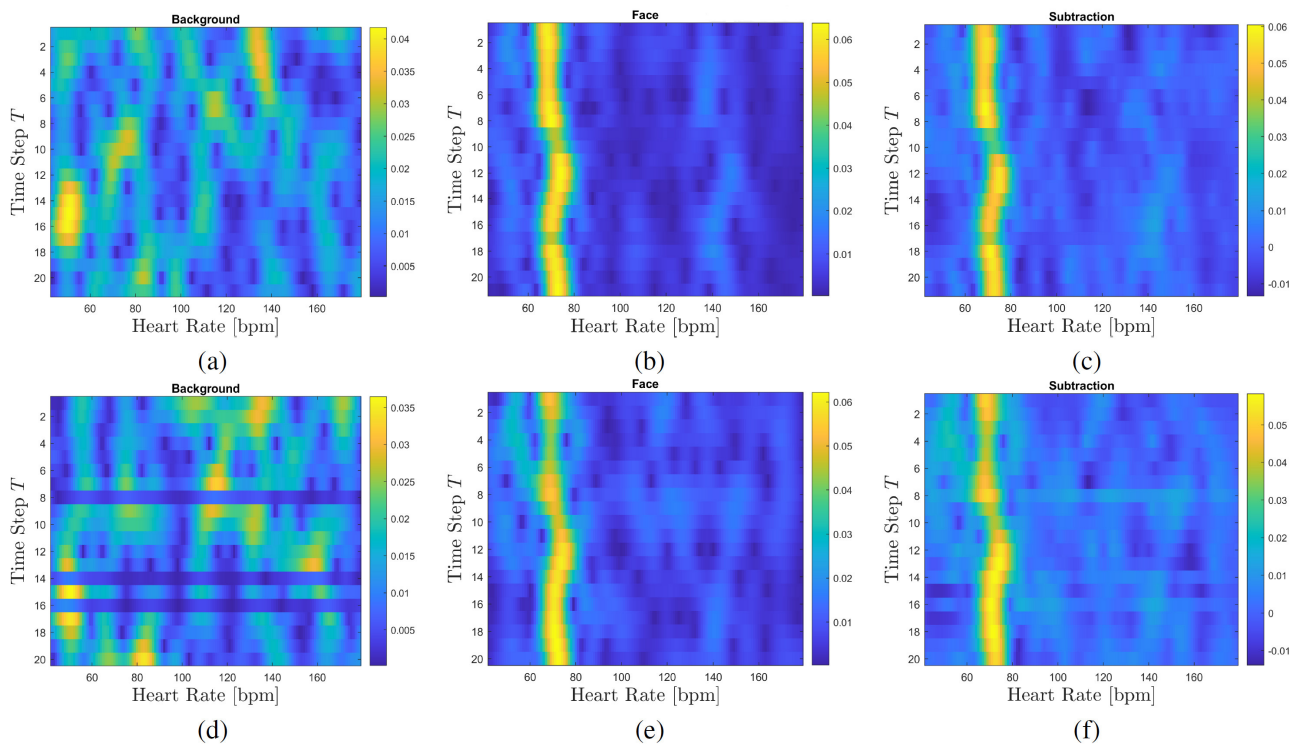


FIGURE 10

FFT maps of capturing ID_11_1100. A similar HR can be detected in both cases, original source (A–C) and deepfake (D–F) video with about 71 bpm. The correlations between the original and deepfake FFT maps shows a moderate to strong relationship for all three map pairs: 0.50 for background (moderate relationship), 0.91 for face (strong), and 0.89 for subtraction map (strong).

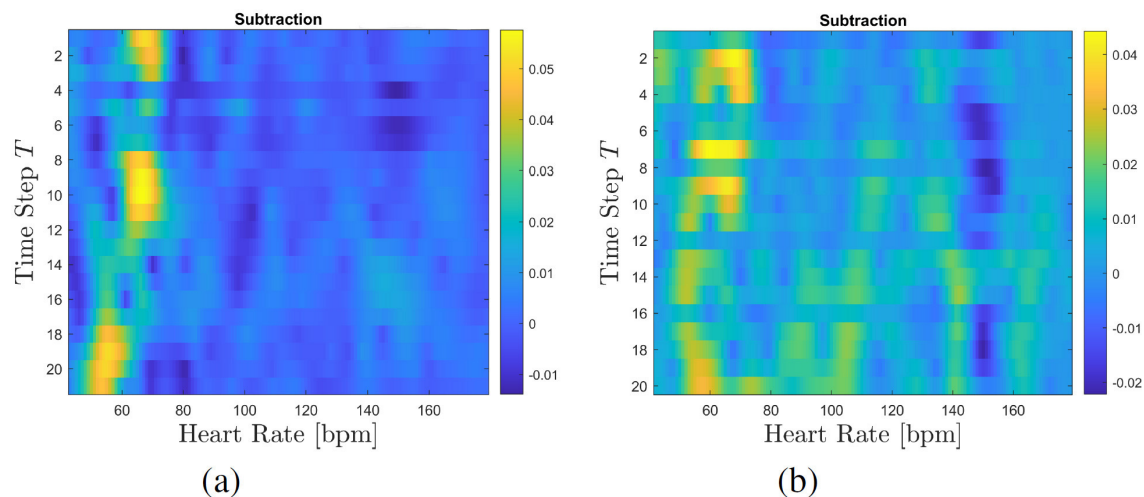


FIGURE 11

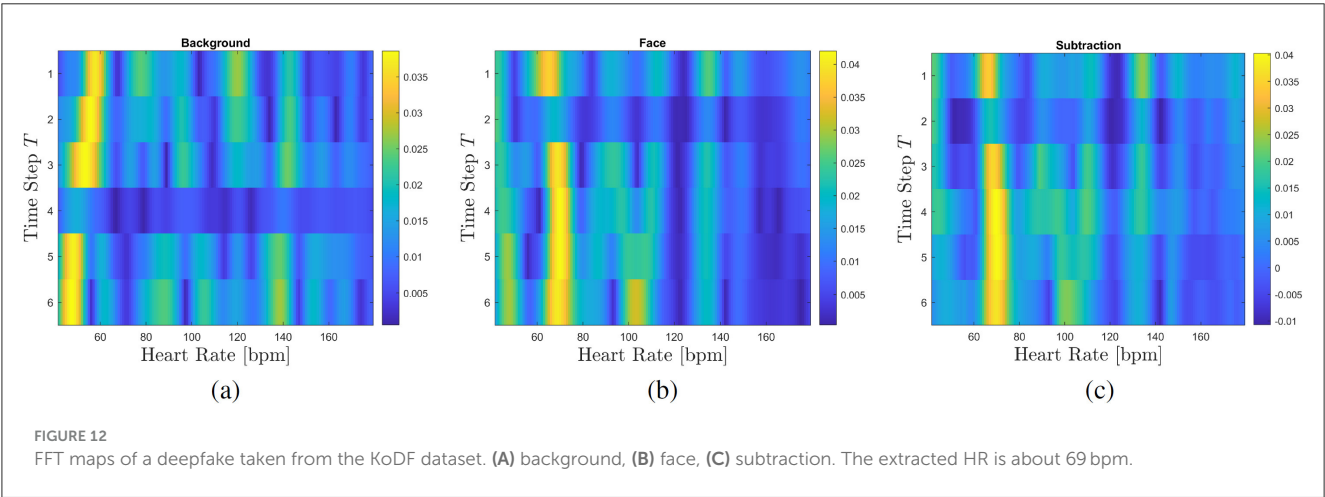
FFT subtraction maps of (A) capturing ID_004_0110 and (B) a related deepfake. In the original FFT map, a HR can be identified clearly at around 65 bpm. The FFT subtraction map of the deepfake is more noisy but the HR of the underlying original is detectable as well. The correlation between both maps is moderate with 0.53. (A) Original. (B) Deepfake.

To improve the utility of physiological signals for deepfake detection, we propose shifting from global HR analysis to locally resolved signals within the face. Recent advances in video-based

vital sign analysis have moved toward capturing local pulse signals from specific facial regions (Kossack et al., 2019b, 2021), which better reflect the anatomical blood flow patterns of the

TABLE 1 Signal-to-noise ratio (SNR) and standard deviation (STD) of the extracted HR signal for all included data.

Method	Beckmann et al. (2023)	DFL	KoDF	Actor our fakes	Actor FF++ fakes
SNR originals [dB]	5.595	8.947	4.470	2.951	1.852
SNR fakes [dB]	3.566	7.838	3.027	2.730	2.345
STD originals [bpm]	5.033	2.779	7.892	6.929	8.291
STD fakes [bpm]	7.300	2.245	8.609	8.931	8.833



human face. By leveraging these localized physiological patterns, we aim to enhance both the robustness and interpretability of deepfake detection. Building on this idea, we performed initial experiments where we extracted rPPG-related feature maps from a subset of our dataset following the approach described in [Schraven et al. \(2023\)](#) and trained an EfficientNet-B4 model as a convolutional deepfake detector ([Tan and Le, 2019](#)). Our preliminary results (AUROC score of 87.4%) show promising results that these local maps can be used for deepfake detection. Using the rPPG-based features improves interpretability by providing more understandable features, but the detector itself lacks transparency by design. To overcome this issue, we adapt in the future the concept proposed by [Seibold et al. \(2021\)](#), which leads to detectors that accurately determines which part of the input contributes to the prediction that an input is a forgery.

6 Conclusion

In conclusion, our study demonstrates that high-quality deepfakes exhibit rPPG signals that correspond to the HR of the source (driver) video. By comparing the different PPG signals and analyzing the FFT maps as well as the extracted HRs and its correlations, we confirmed that the globally derived rPPG signal originates from the driving video, rather than being artificially generated. This finding challenges previous assumptions that deepfakes inherently lack valid physiological signals, revealing the

limitations of using simple HR analysis for detecting high-quality deepfakes.

One of the key contributions of our study is the demonstration that HR signals in deepfakes can closely match those of the source video, making traditional global HR-based detection methods insufficient for distinguishing between real and fake content. By performing our analysis not only on our own dataset but also on fakes created with DeepFaceLive and from the KoDF dataset, we confirmed the generalization of our findings, showing that even older deepfake datasets contain valid HR signals.

To address this limitation, we propose leveraging local blood flow information for deepfake detection. Preliminary experiments indicate that this localized analysis holds significant promise for improving detection accuracy. As part of ongoing work, we are further refining this approach, which also offers the added benefit of enhanced explainability. Visualizing local blood flow patterns could provide clearer insight into the decision-making process of detection algorithms. Another important factor in ensuring robust detection is the availability of good and diverse training data. An attacker may attempt to mimic blood flow patterns to evade detection; therefore, we plan to enhance our deepfake dataset using style-transfer with a temporal component by extending the work on improved image forgeries of [Seibold et al. \(2019\)](#).

In summary, our contributions include: (1) providing evidence that deepfakes can exhibit realistic heart rate signals, (2) highlighting the insufficiency of global HR analysis for detecting high-quality deepfakes, and (3) proposing the use

of localized rPPG signals to enhance both the robustness and explainability of deepfake detection. Our approach could serve as a valuable complement to existing techniques, with the potential to improve the security and integrity of multimedia content across platforms.

Data availability statement

The datasets presented in this article are not readily available because no commercial use. Requests to access the datasets should be directed to Peter Eisert, peter.eisert@hhi.fraunhofer.de.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

CS: Conceptualization, Formal analysis, Validation, Investigation, Writing – original draft. EW: Conceptualization, Formal analysis, Validation, Investigation, Writing – original draft. AB: Methodology, Investigation. BK: Investigation, Writing – original draft. AH: Supervision, Writing – review & editing. PE: Funding acquisition, Supervision, Writing – review & editing.

References

- Beckmann, A., Hilsmann, A., and Eisert, P. (2023). "Fooling state-of-the-art deepfake detection with high-quality deepfakes," in *Proceedings of the 2023 ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec '23* (New York, NY: Association for Computing Machinery), 175–180.
- Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., and Dubois, J. (2019). Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognit. Lett.* 124, 82–90. doi: 10.1016/j.patrec.2017.10.017
- Boccignone, G., Bursic, S., Cuculo, V., D'Amelio, A., Grossi, G., Lanzarotti, R., et al. (2022). "Deepfakes have no heart: A simple rppg-based method to reveal fake videos," in *Image Analysis and Processing - ICIAP 2022: 21st International Conference* (Berlin, Heidelberg: Springer-Verlag), 186–195.
- Chen, M., Liao, X., and Wu, M. (2022). Pulseedit: Editing physiological signals in facial videos for privacy protection. *IEEE Trans. Inform. Forens. Secur.* 17, 457–471. doi: 10.1109/TIFS.2022.3142993
- Chen, W., and McDuff, D. (2018). "Deepphys: Video-based physiological measurement using convolutional attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)* (The Computer Vision Foundation (CVF)).
- Ciftci, U. A., Demir, İ., and Yin, L. (2024). Deepfake source detection in a heart beat. *Vis. Comput.* 40, 2733–2750. doi: 10.1007/s00371-023-02981-0
- Ciftci, U. A., Demir, İ., and Yin, L. (2020a). "Fakecatcher: Detection of synthetic portrait videos using biological signals," in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (IEEE).
- Ciftci, U. A., Demir, İ., and Yin, L. (2020b). "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)* (Houston, TX: IEEE), 1–10.
- Ciftci, U. A., and Yin, L. (2019). "Heart rate based face synthesis for pulse estimation," in *Advances in Visual Computing: 14th International Symposium on Visual Computing, ISVC 2019* (Lake Tahoe: Springer), 540–551.
- D'Amelio, A., Lanzarotti, R., Patania, S., Grossi, G., Cuculo, V., Valota, A., et al. (2023). "On using rppg signals for deepfake detection: a cautionary note," in *International Conference on Image Analysis and Processing* (Cham: Springer), 235–246.
- Dang-Nguyen, D.-T., Dang-Nguyen, D.-S., Piras, L., Giacinto, G., and Boato, G. (2020). *CelebDF: A Large-Scale Challenging Dataset for Deepfake Forensics*. Available online at: <https://github.com/yuezunli/celeb-deepfakeforensics> (accessed March 7, 2025).
- de Berg, M., Cheong, O., van Kreveld, M., and Overmars, M. (2008). *Computational Geometry: Algorithms and Applications*. Cham: Springer Science & Business Media.
- De Haan, G., and Jeanne, V. (2013). Robust pulse rate from chrominance-based rPPG. *IEEE Trans. Biomed. Eng.* 60, 2878–2886. doi: 10.1109/TBME.2013.2266196
- DeepFakes (2019). *Faceswap*.
- Ding, Z., Zhang, C., Xia, Z., Jebe, L., Tu, Z., and Zhang, X. (2023). "Diffusionrig: Learning personalized priors for facial appearance editing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE).
- Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., et al. (2020). The deepfake detection challenge (DFDC) dataset. *arXiv [preprint]* arXiv:2006.07397. doi: 10.48550/arXiv.2006.07397
- Drobyshev, N., Chelishchev, J., Khakhulin, T., Ivakhnenko, A., Lempitsky, V., and Zakharov, E. (2022). "Megaportraits: One-shot megapixel neural head avatars," in *Proc. of the 30th ACM International Conference on Multimedia* (New York, NY: Association for Computing Machinery).
- Dufour, N., Gully, A., Karlsson, P., Vorbyov, A., Leung, T., Childs, J., et al. (2019). *Deepfakes Detection Dataset*. New York: Google and Jigsaw.
- Fernandes, S., Raj, S., Ortiz, E., Vintila, I., Salter, M., Urosevic, G., et al. (2019). "Predicting heart rate variations of deepfake videos using neural ode," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (Seoul: IEEE).

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was funded by the German Federal Ministry of Education and Research (BMBF) under Grant No. 13N15735 (FakeID) and by Horizon Europe under Grant No. 101121280 (Einstein).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, eds. Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger (New York: Curran Associates, Inc).
- Google (2022). *MediaPipe: A Framework for Building Multimodal Applied Machine Learning Pipelines*. Available online at: <https://mediapipe.dev/> (accessed: March 14, 2024).
- Haliassos, A., Mira, R., Petridis, S., and Pantic, M. (2022). "Leveraging real talking faces via self-supervision for robust forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (New Orleans, LA: IEEE), 14950–14962.
- Hernandez-Ortega, J., Tolosana, R., Fierrez, J., and Morales, A. (2020). Deepfakeson-phys: Deepfakes detection based on heart rate estimation. *arXiv [preprint]* arXiv:2010.00400. doi: 10.48550/arXiv.2010.00400
- Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *arXiv [preprint]* arxiv:2006.11239. doi: 10.48550/arXiv.2006.11239
- Huang, Z., Chan, K. C., Jiang, Y., and Liu, Z. (2023). "Collaborative diffusion for multi-modal face generation and editing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Vancouver, BC: IEEE).
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). "Analyzing and improving the image quality of StyleGAN," in *Proceedings of CVPR (IEEE/CVF)*.
- Kingma, D. P., and Welling, M. (2014). "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations* (Banff, AB: Conference Track Proceedings).
- Kossack, B., Wisotzky, E., Eisert, P., Schraven, S. P., Globke, B., and Hilsmann, A. (2022). "Perfusion assessment via local remote photoplethysmography (RPPG)," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (New Orleans, LA: IEEE), 2192–2201.
- Kossack, B., Wisotzky, E. L., Hilsmann, A., and Eisert, P. (2019a). "Local remote photoplethysmography signal analysis for application in presentation attack detection," in *Vision, Modeling and Visualization-VMV* (London: The Eurographics Association), 135–142.
- Kossack, B., Wisotzky, E. L., Hilsmann, A., and Eisert, P. (2021). "Automatic region-based heart rate measurement using remote photoplethysmography," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (Montreal, BC: IEEE), 2755–2759.
- Kossack, B., Wisotzky, E. L., Hilsmann, A., Eisert, P., and Hänsch, R. (2019b). Local blood flow analysis and visualization from RGB-video sequences. *Curr. Direct. Biomed. Eng.* 5:1. doi: 10.1515/cdbme-2019-0094
- Kossack, B., Wisotzky, E. L., Schraven, S., Skopnik, L., Hilsmann, A., and Eisert, P. (2023). Modified allen test assessment via imaging photoplethysmography. *Curr. Direct. Biomed. Eng.* 9, 571–574. doi: 10.1515/cdbme-2023-1143
- Kwon, P., You, J., Nam, G., Park, S., and Chae, G. (2021). "Kodf: A large-scale korean deepfake detection dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (Montreal, QC: IEEE), 10744–10753.
- Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F., et al. (2020). "Face x-ray for more general face forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA: IEEE).
- Li, X., Komulainen, J., Zhao, G., Yuen, P. C., and Pietikainen, M. (2017). "Generalized face anti-spoofing by detecting pulse from face videos," in *Proceedings - International Conference on Pattern Recognition* (Cancun: IEEE), 4244–4249.
- Li, Y., Chang, M.-C., and Lyu, S. (2018). "In Ictu Oculi: Exposing ai generated fake face videos by detecting eye blinking," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)* (Hong Kong: IEEE).
- Li, Y., and Lyu, S. (2018). Exposing deepfake videos by detecting face warping artifacts. *arXiv [preprint]* arXiv:1811.00656. doi: 10.48550/arXiv.1811.00656
- McDuff, D. J., Blackford, E. B., and Estep, J. R. (2017). "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)* (Washington, DC: IEEE), 63–70.
- Nirkin, Y., Keller, Y., and Hassner, T. (2022). *Fsganv2: Improved Subject Agnostic Face Swapping and Reenactment* (IEEE).
- Perov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Umé, C., et al. (2020). Deepfacelab: A simple, flexible and extensible face swapping framework. *arXiv [preprint]* arXiv:2005.05535. doi: 10.48550/arXiv.2005.05535
- Petrov, I. (2023). *DeepFaceLive: Real-Time Face Swap for PC Streaming or Video Calls*. Available online at: <https://github.com/iperov/DeepFaceLive> (accessed: October 25, 2024).
- Poh, M.-Z., McDuff, D. J., and Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* 18:10762. doi: 10.1364/OE.18.010762
- Qi, H., Guo, Q., Juefei-Xu, F., Xie, X., Ma, L., Feng, W., et al. (2020). "Deephythm: exposing deepfakes with attentional visual heartbeat rhythms," in *Proceedings of the 28th ACM International Conference on Multimedia* (New York: ACM), 4318–4327.
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., et al. (2021). "Zero-shot text-to-image generation," in *Proceedings of the 38th International Conference on Machine Learning*, eds. M. Meila, and T. Zhang (New York: PMLR), 8821–8831.
- Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Niefßner, M. (2019). "FaceForensics++: Learning to detect manipulated facial images," in *International Conference on Computer Vision (ICCV)*.
- Schraven, S. P., Kossack, B., Strüder, D., Jung, M., Skopnik, L., Gross, J., et al. (2023). Continuous intraoperative perfusion monitoring of free microvascular anastomosed fasciocutaneous flaps using remote photoplethysmography. *Sci. Rep.* 13:1532. doi: 10.1038/s41598-023-28277-w
- Seibold, C., Hilsmann, A., and Eisert, P. (2017). Model-based motion blur estimation for the improvement of motion tracking. *Comp. Vision Image Understand.* 160, 1077–1142. doi: 10.1016/j.cviu.2017.03.005
- Seibold, C., Hilsmann, A., and Eisert, P. (2019). "Style your face morph and improve your face morphing attack detector," in *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)* (IEEE), 1–6.
- Seibold, C., Hilsmann, A., and Eisert, P. (2021). Feature focus: towards explainable and transparent deep face morphing attack detectors. *Computers* 10:117. doi: 10.3390/computers10090117
- Seibold, C., Hilsmann, A., and Eisert, P. (2024). "Towards better morphed face images without ghosting artifacts," in *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (Rome: SCITEPRESS). doi: 10.5220/0012302800003660
- Tamura, T., Maeda, Y., Sekine, M., and Yoshida, M. (2014). Wearable Photoplethysmographic sensors—past and present. *Electronics* 3, 282–302. doi: 10.3390/electronics3020282
- Tan, M., and Le, Q. (2019). "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, eds. K. Chaudhuri, and R. Salakhutdinov (New York: PMLR), 6105–6114.
- Tulyakov, S., Alameda-Pineda, X., Ricci, E., Yin, L., Cohn, J. F., and Sebe, N. (2016). "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE).
- Wang, J., Shan, C., Liu, L., and Hou, Z. (2024). Camera-based physiological measurement: Recent advances and future prospects. *Neurocomputing* 2024:127282. doi: 10.1016/j.neucom.2024.127282
- Wang, T.-C., Mallya, A., and Liu, M.-Y. (2021a). "One-shot free-view neural talking-head synthesis for video conferencing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Nashville, TN: IEEE).
- Wang, W., Den Brinker, A. C., Stuijk, S., and De Haan, G. (2017). Algorithmic Principles of Remote PPG. *IEEE Trans. Biomed. Eng.* 64, 1479–1491. doi: 10.1109/TBME.2016.2609282
- Wang, Y., Chen, X., Zhu, J., Chu, W., Tai, Y., Wang, C., et al. (2021b). "Hiface: 3D shape and semantic prior guided high fidelity face swapping," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence* (Montreal: IJCAI-21).
- Wang, Z., Bao, J., Zhou, W., Wang, W., and Li, H. (2023). "Altfreezing for more general video face forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Vancouver, BC: IEEE), 4129–4138.
- Wu, J., Zhu, Y., Jiang, X., Liu, Y., and Lin, J. (2023). Local attention and long-distance interaction of rppg for deepfake detection. *Vis. Comput.* 40, 1083–1094. doi: 10.1007/s00371-023-02833-x
- Yang, J., Guthrie, B., and Saddik, E. A. (2015). "Estimating two-dimensional blood flow velocities from videos," in *International Conference on Image Processing (ICIP)* (Quebec City, QC: IEEE), 3768–3772.
- Yang, X., Li, Y., and Lyu, S. (2019). "Exposing deep fakes using inconsistent head poses," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Brighton: IEEE), 8261–8265.
- Yang, Z., Wang, H., and Lu, F. (2021). Assessment of deep learning-based heart rate estimation using remote photoplethysmography under different illuminations. *arXiv [preprint]* arXiv:2107.13193. doi: 10.48550/arXiv.2107.13193
- Yu, P., Xia, Z., Fei, J., and Lu, Y. (2021a). A survey on deepfake video detection. *IET Biomet.* 10, 607–624. doi: 10.1049/bme2.12031
- Yu, Z., Li, X., Wang, P., and Zhao, G. (2021b). Transrppg: Remote photoplethysmography transformer for 3d mask face presentation attack detection. *IEEE Signal Process. Lett.* 28, 1290–1294. doi: 10.1109/LSP.2021.3089908

Yu, Z., Li, X., and Zhao, G. (2020). "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," in *30th British Machine Vision Conference 2019* (Glasgow: BMVC).

Yu, Z., Peng, W., Li, X., Hong, X., and Zhao, G. (2019). "Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul: IEEE).

Zaunseder, S., Trumpp, A., Wedekind, D., and Malberg, H. (2018). Cardiovascular assessment by imaging photoplethysmography - a review. *Biomedizinische Technik* 2018, 1–18. doi: 10.1515/bmt-2017-0119

Zhao, W., Rao, Y., Shi, W., Liu, Z., Zhou, J., and Lu, J. (2023). "Diffswap: high-fidelity and controllable face swapping via 3D-aware masked diffusion," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Vancouver, BC: IEEE), 8568–8577.