PulseEdit: Editing Physiological Signal in Facial Videos for Privacy Protection

Mingliang Chen [1,1,1,1,1], Xin Liao [2], Min Wu [2], and Min Wu [2]

[1]ECE
[2]Affiliation not available

November 8, 2023

Abstract

Recent studies have shown that physiological signals can be remotely captured from human faces using a portable color camera under ambient light. This technology, namely remote photoplethysmography (rPPG), can be used to collect users' physiological status who are sitting in front of a camera, which may raise physiological privacy issues. To avoid the privacy abuse of the rPPG technology, this paper develops PulseEdit, a novel and efficient algorithm that can edit the physiological signals in facial videos without affecting visual appearance to protect the user's physiological signal from disclosure. PulseEdit can either remove the trace of the physiological signal in a video or transform the video to contain a target physiological signal chosen by a user. Experimental results show that PulseEdit can effectively edit physiological signals in facial videos and prevent heart rate measurement based on rPPG. It is possible to utilize PulseEdit in adversarial scenarios against some rPPG-based visual security algorithms. We present analyses on the performance of PulseEdit against rPPG-based liveness detection and rPPG-based deepfake detection, and demonstrate its ability to circumvent these visual security algorithms.

Hosted file

`demo_rem.mp4` available at https://authorea.com/users/680724/articles/677272-pulseedit-editing-physiological-signal-in-facial-videos-for-privacy-protection

Hosted file

`demo_mod.mp4` available at https://authorea.com/users/680724/articles/677272-pulseedit-editing-physiological-signal-in-facial-videos-for-privacy-protection

PulseEdit: Editing Physiological Signals in Facial Videos for Privacy Protection

Mingliang Chen, *Member, IEEE,* Xin Liao, *Senior Member, IEEE,* and Min Wu, *Fellow, IEEE*

***Abstract*—Recent studies have shown that physiological signals such as heart beat and breathing can be remotely captured from human faces using a regular color camera under ambient light. This technology, referred to as remote photoplethysmography (rPPG), can be used to collect the physiological status of users who are in front of a camera, which may raise privacy concerns. To avoid the privacy abuse of the rPPG technology, this paper develops PulseEdit, a novel and efficient algorithm that can edit the physiological signals in facial videos without affecting visual appearance and thus protect the user's physiological signal from disclosure. PulseEdit can either remove the trace of the physiological signal in a video or transform the video to contain a target physiological signal chosen by a user. Experimental results show that PulseEdit can effectively edit physiological signals in facial videos and prevent heart rate measurement based on rPPG. It is possible to utilize PulseEdit in adversarial scenarios against rPPG-based visual security algorithms. We present analyses on the performance of PulseEdit against rPPG-based liveness detection and rPPG-based deepfake detection, and demonstrate its ability to circumvent these visual security algorithms and its important role in supporting the design of attack-resilient systems.**

***Index Terms*—Remote photoplethysmography (rPPG), privacy protection, visual security, video editing, video forgery.**

I. Introduction

VIDEO-CAPTURING devices are ubiquitous in our daily life. These devices help us share our experiences with friends and communicate online with others. Yet have we realized whenever a person appears in front of a camera, not only can people recognize his/her identity based on the facial appearance, but also monitor some aspects of his/her physiological status such as cardiac activities?

Recent research has shown that contact-free measurement of human physiological signals from facial videos is feasible through computer vision algorithms [1]–[4]. For instance, remote photoplethysmography (rPPG) technology has attracted a growing amount of interests in capturing the subtle color changes of the skin caused by heartbeats in facial videos under ambient light. We can further infer heart rate (HR) [5]–[10], respiration rate (RR) [11], [12], and heart rate variability (HRV) [13] from the extracted rPPG signals. This promising technology can facilitate remote monitoring stress and fatigue during computer tasks [14] and sports training [15].

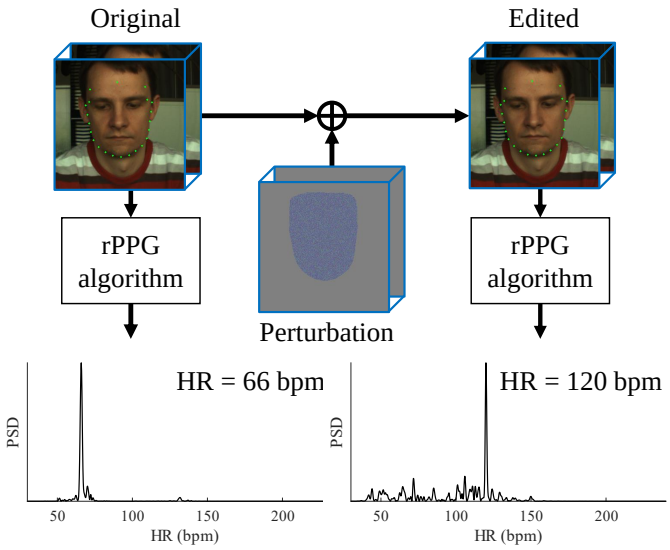


Fig. 1. PulseEdit can edit the rPPG signal in a facial video to conceal a person's true physiological status, without visual distortion of his/her appearance. We introduce negligible additive perturbations onto the facial region in the video, and successfully modify the HR extracted by the rPPG algorithm. In this example, HR is edited from 66 to 120 beats per minute (bpm) to avoid the disclosure of the user's true heart rate in the video.

Recalling the question raised at the very beginning of the paper, we recognize that this emerging technology may cause concerns about physiological privacy. With such a technology, video-capturing devices can record both a person's appearance and his/her physiological status such as cardiac activities simultaneously. This kind of physiological information that is intrinsically present in facial videos may subject to abuse, such as secretly collecting and analyzing a person's physiological features with ulterior motives. For example, opponents can read one's physiological status and analyze his/her conditions to gain an advantage in mission-critical negotiations. In daily life, a person's certain health conditions may be revealed without his/her explicit consent from a video taken by a party.

To address these physiological privacy issues, it is important to investigate how to effectively protect the physiological signals from disclosure in facial videos. To this end, we propose *PulseEdit* illustrated in Fig. 1, a novel method that edits rPPG signals in facial videos by superimposing specifically designed perturbation of small amplitude onto the input videos. Our method outputs a video that is visually the same but has its

Mingliang Chen was with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA when he performed the work, and now he is with Meta Platform Inc., Menlo Park, CA 94025 USA (e-mail: mchen126@terpmail.umd.edu).

Xin Liao is with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan 410082, China. (e-mail: xinliao@hnu.edu.cn).

Min Wu is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: minwu@umd.edu).

rPPG signal either removed completely or transformed to a target HR based on the user's choice. Processed by PulseEdit, the users' rPPG signals are protected from disclosure in the facial videos.

To make PulseEdit effective in practical use, we consider the following requirements when designing and evaluating the algorithm:

- **Invisibility:** the editing on the face should be negligible without obvious distortion in appearance.
- **Universality:** the protection should be valid on the face globally and locally. The processed video should no longer contain the user's true rPPG signal, and the edited rPPG signal can be detected from the whole face as well as local skin regions.
- **Generality:** the protection should be able to conceal a person's true rPPG signal against various rPPG algorithms in the literature. In other words, the edited rPPG signal can be measured by various rPPG algorithms.
- **Resistance:** an advanced capability is to make the editing on the face not detectable under visual forensic analysis.

In addition to privacy protection, PulseEdit can impact other applications where rPPG is employed. More specifically, rPPG signal has been demonstrated as a useful and discriminative feature in various visual security tasks, such as liveness detection [16]–[18] and deepfake detection [19], because real/live videos and fake/synthetic videos have different representations in rPPG signals extracted from the facial regions. Empowered by PulseEdit, we can edit the rPPG signals in facial videos and circumvent the above rPPG-based visual security algorithms. It is not difficult to see that PulseEdit is a potential threat to invalidate these algorithms, providing a direction to revise them and improve the confidence of their output decisions.

Our main contributions are summarized as follows:

- We develop *PulseEdit*, a novel algorithm that can edit rPPG signals in a facial video to conceal a person's true cardiac activity and physiological status, without introducing noticeable visual distortion in the video.
- We demonstrate that PulseEdit can provide effective privacy protection under various rPPG extraction algorithms in the literature and robustly edit rPPG signals in global and local facial regions. We further investigate the forensic detectability of PulseEdit against forensic steganalysis.
- We analyze the effectiveness of PulseEdit in circumventing rPPG-based liveness detection and rPPG-based deepfake detection. We show that PulseEdit is promising in circumventing these rPPG-based algorithms, which suggests that more research efforts are needed to improve these rPPG-based visual security algorithms from this adversarial perspective.

In the rest of the paper, we first introduce the prior work related to rPPG technology and its application in visual security tasks in Section II. Section III describes the proposed *PulseEdit* to edit rPPG signals in facial videos. We carry out comprehensive performance analysis on the PulseEdit algorithm for removing/modifying rPPG signals in facial videos in Section IV and explore its feasibility as a potential adversary against rPPG-based liveness detection and deepfake detection algorithms in Section V. Finally, Section VI discusses several related issues and Section VII concludes the paper.

II. Related Work

A. rPPG Technology

Monitoring cardiac activity is essential for understanding a person's health status and is actively used in clinical practices and home care. Conventional methods require contact-based sensors attached to the human skin, such as electrocardiogram leads, a pulse oximeter, or a fitness tracker.

Recently, rPPG enables contact-free HR measurement using color cameras. The principle of rPPG is that the blood volume changes under the skin influence the intensity and color of the reflected light from the skin, whose pattern is consistent with heartbeat cycles. Although such subtle momentary changes in the reflected light from the facial skin are not detectable by the human eyes, they can be captured by a color camera [1]. Eulerian video magnification [20] can amplify and visualize the subtle color changes in a facial video caused by the blood flow. Independent component analysis (ICA) [13], chrominance mapping (CHROM) [2], and plane-orthogonal-to-skin (POS) [4] were proposed to extract robust rPPG features from three color channels. Li *et al.* [5] applied adaptive filtering to handle environmental illumination and voluntary motion issues in remote HR measurement. Tulyakov *et al.* [6] proposed self-adaptive matrix completion to denoise rPPG features and offer robust HR estimation. The challenging fitness scenario [21], [22] has also been studied to improve the robustness of the rPPG technology. End-to-end models [7], [9] employing deep learning were also introduced to estimate HR from videos.

B. Biometric Privacy Protection

Biometric privacy protection [23], [24] aims to conceal a person's privacy in biometric data and prevent possible thefts and misuses of this information. Traditional biometric privacy protection algorithms were proposed to de-identify a person's identity from these biometric features, including face [25], [26], iris [27], and fingerprint [28]. Deep learning has been introduced to protect privacy in multimodal biometrics [29].

In spite of privacy protection at the image perception level, several researches studied the privacy protection approaches at the feature representation level. Several facial representation methods were proposed to eliminate facial expressions [30] or selected biometric attributes (*e.g.*, age and gender) [31] in facial feature level. SensitiveNets [32] generates a learned embedding space that eliminates specific sensitive biometric information from the existing representation subspace. Terhorst *et al.* [33] proposed a privacy-preserving solution to suppress biometric attributes in an unsupervised manner. The privacy-preserving feature representations can improve the robustness of training models and benefit the fairness in model inference across biometric attributes.

As many methods have been proposed in the recent decade to extract physiological signals from facial videos, concerns are raised concurrently on the privacy issues of physiological information in videos. This information may be misused to

collect and analyze a person's physiological features with ulterior motives. Chen *et al.* [34] applied motion elimination in facial videos to subtract the pulse-induced pixel intensity variation on the subjects' faces to avoid the disclosure of the rPPG signal. The experimental results show that the rPPG signals are successfully removed without appearance distortion. As the work only studied the steady case in the research, it is unclear whether Chen's method can deal with the subjects' voluntary motion (*e.g.,* talking, head translation, and rotation) in video recording.

In this paper, we propose to edit the rPPG signals that are intrinsically presented in facial videos by perturbing the skin pixels on the face and conduct experiments on motion cases as well as steady cases. Compared with the prior art, not only is our work capable of removing the rPPG signal in a facial video, but also transforming it to a target rPPG signal of the user's choice.

C. rPPG Feature in Visual Security Tasks

rPPG signal has been employed as a discriminative feature to tackle several visual security tasks involving face videos, such as liveness detection against spoofing and deepfake detection. Liveness detection is crucial to protect face recognition systems from spoofing attacks, including printing a face on paper, replaying a facial video on a digital device, wearing a 3D face mask, and other approaches by adversaries. Liu *et al.* [16], [17] used the cross-correlation of rPPG features in multiple facial regions to classify live faces vs. spoofed faces. Hernandez *et al.* [18] proposed to analyze the signal quality of rPPG extracted from faces to discriminate live faces and spoofed faces.

"Deepfake" refers to a family of computer technologies to transform a person's face to another's in images or videos. Since deepfake videos circulated in social media have brought serious concerns such as through celebrity pornographic videos, fake news, hoaxes, and financial fraud, which largely impairs the integrity of social media, deepfake detection has attracted a lot of attention in the recent computer vision research. For example, FakeCatcher [19] explored the discriminative features of rPPG signals extracted from facial videos and utilized them for deepfake detection.

III. Proposed Method

PulseEdit has three main steps as shown in Fig. 2. We first detect the facial region in the video and extract skin intensity signals from multiple subregions on the face. We then obtain the perturbation signal via an optimization problem that transforms the rPPG signal in the video to a target signal. Finally, we manipulate the skin pixels in the video according to the perturbation signal, so that the PulseEdit video successfully removes the rPPG signal or transforms the rPPG signal to a target rPPG signal of the user's choice. We refer to the two modes as "removal" and "modification", respectively, in short. In the removal mode, the target signal can be white Gaussian noise; and in the modification mode, the target signal can be a simulated sinusoid with the frequency of a target HR or the rPPG signal extracted from a reference video of the user's choice.

A. rPPG Extraction

Similar to the prior art in the rPPG research, we first track the subject's face in the video to extract rPPG signal. We apply the facial landmark detector by Dlib [35] to locate and track 68 facial landmarks, from which we define the facial region of interest (ROI) shown with the green dots in the video frame in Fig. 2. To facilitate rPPG extraction from multiple subregions [6], the ROI is normalized to a rectangle using piecewise linear geometric transformation, and skin color pixels are masked by a Gaussian skin color model in the chrominance space [36]:

$$p(\mathbf{x}) = \exp\big(-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\Sigma^{-1}(\mathbf{x}-\mathbf{m})\big) \underset{\text{non-skin}}{\overset{\text{skin}}{\gtrless}} p_t, \quad (1)$$

where $\mathbf{x} = [cb, cr]^T$, and $\mathbf{m}$ and Σ are the mean and covariance matrix of the Gaussian skin color model. Within the masked rectangle facial ROI, we use a rectangle of a quarter size to uniformly select M subregions (subregions can have overlap with their neighbors). We compute the spatial average of the skin pixels in each subregion to form the skin intensity signal $R \in \mathbb{R}^{M\times 3\times N}$, for M subregions, 3 color channels, and N frames in the video. In the subsequent discussions, we refer to the subscripts i and c as subregion i and color channel c, respectively. For example, $R_{i,c}$ denotes the skin intensity signal in subregion i and color channel c.

B. rPPG Editing

In this module, our goal is to find a suitable perturbation on the skin intensity signals to change the rPPG in videos to the target signal given by users. We first detrend the skin intensity signal $R_{i,c}, \forall i, c$, to eliminate the illumination interference in the environment. In the detrending process, we use l_1 trend filtering [37] to obtain the signal trend and subtract it from the skin intensity signal. The detrending process can be solved by the optimization problem as

$$\min_{S_{i,c}} \ \frac{1}{2}||S_{i,c}||_2^2 + \mu||D(R_{i,c}-S_{i,c})||_1, \forall i, c, \quad (2)$$

where $S \in \mathbb{R}^{M\times 3\times N}$ denotes the corresponding detrended signal, the subscripts i and c denote the subregion and the color channel, and $D \in \mathbb{R}^{(N-2)\times N}$ is the second-order difference matrix

$$D = \begin{bmatrix} -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & \mathbf{0} & \\ & & \ddots & \ddots & \ddots & & \\ & \mathbf{0} & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \end{bmatrix}. \quad (3)$$

We denote $\delta \in \mathbb{R}^{3\times N}$ as the additive RGB perturbation imposed onto the detrended signal S, which gives rise to the edited signal $\tilde{S} \in \mathbb{R}^{M\times 3\times N}$, *i.e.* $\tilde{S}_{i,c} = S_{i,c} + \delta_c, \forall i, c$, where δ_c denotes the perturbation in the color channel c.

Next, we generate the target rPPG signal $T \in \mathbb{R}^{3\times N}$. To ensure the output video contains the target rPPG signal T, we maximize the similarity between the edited signals $\tilde{S}$ and the target signal T using the Pearson correlation coefficient:

$$\mathrm{P} = \frac{1}{M}\sum_{i,c}\rho(\tilde{S}_{i,c}, T_c) = \frac{1}{M}\sum_{i,c}\rho(S_{i,c}+\delta_c, T_c). \quad (4)$$

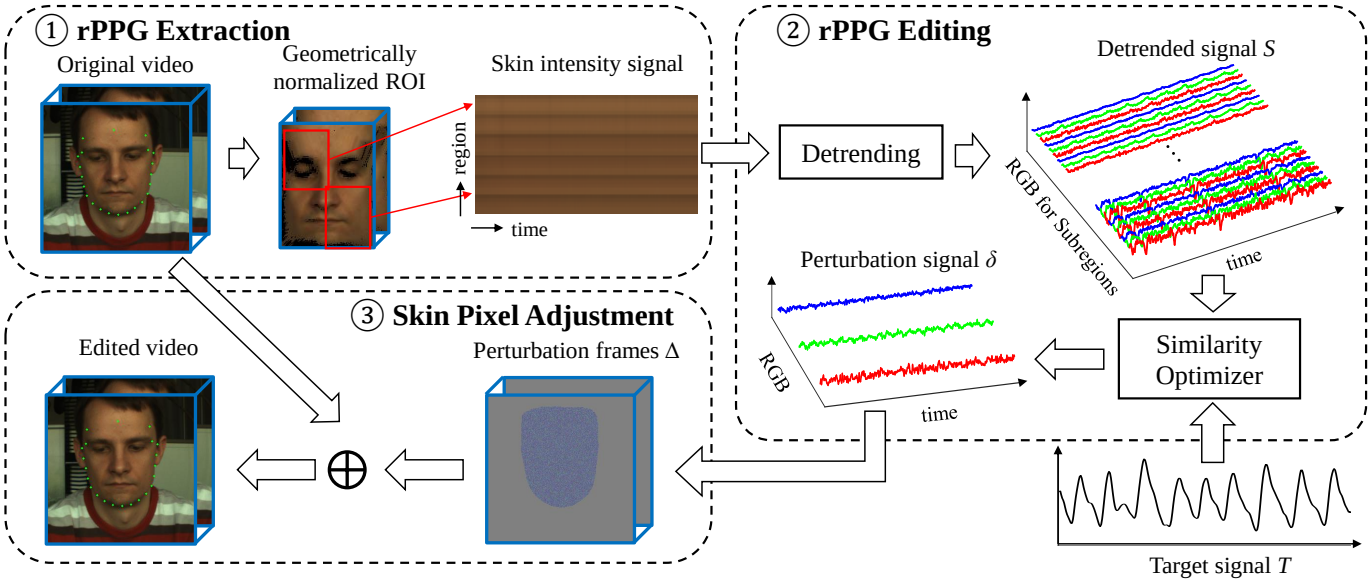


Fig. 2. Pipeline of PulseEdit system. We first extract skin intensity signals from multiple facial subregions in the video. Then, we compute the perturbation signal that can change the rPPG signals in the video to the target rPPG signal. Finally, we edit the skin pixels in the video, and the rPPG signal extracted from the video processed by PulseEdit is successfully transformed to the target signal.

For the edited facial video, we require that the person in the video has negligible perceptual distortion. Thus, we regularize the perturbation signal δ with L_2 loss to control the perturbation budget in the facial video:

$$E = \frac{1}{N}||\delta||_2^2. \tag{5}$$

Combining the above two terms, we obtain the perturbation signal δ by solving the optimization problem:

$$\min_{\delta} \quad -\frac{1}{M}\sum_{i,c}\rho(S_{i,c}+\delta_c, T_c) + \lambda\frac{1}{N}||\delta||_2^2. \tag{6}$$

We can use a gradient-based solver, such as the Adam solver [38], to solve the optimization problem in (6).

C. Skin Pixel Adjustment

The goal of this module is to map the perturbation signal $\delta \in \mathbb{R}^{3\times N}$ in time series to the spatial-temporal perturbation frames $\Delta \in \mathbb{R}^{h\times w\times 3\times N}$, where h and w refers to the height and width of the frames in pixel count. We denote $\delta_c(n)$ as the perturbation of the color channel c in the n-th frame. One simple and intuitive approach to edit the pixels on the face is to directly add $\delta(n)$ to every skin pixel on the facial region in the n-th frame of the input video. Due to the integer quantization of pixel values in video frames, the decimal part of $\delta(n)$ needs special consideration in order to ensure the pixel values are collectively changed by the expected amount.

We adopt randomized dithering to skin pixels to achieve decimal perturbation in a statistical sense. Specifically, for the color channel c in the n-th frame, we adjust the skin pixels in an amount of either $\lfloor\delta_c(n)\rfloor$ with probability p or $\lceil\delta_c(n)\rceil$ with probability $1-p$, where p should be chosen so that

$$\delta_c(n) = \lfloor\delta_c(n)\rfloor p + \lceil\delta_c(n)\rceil(1-p). \tag{7}$$

Algorithm 1 Skin Pixel Adjustment

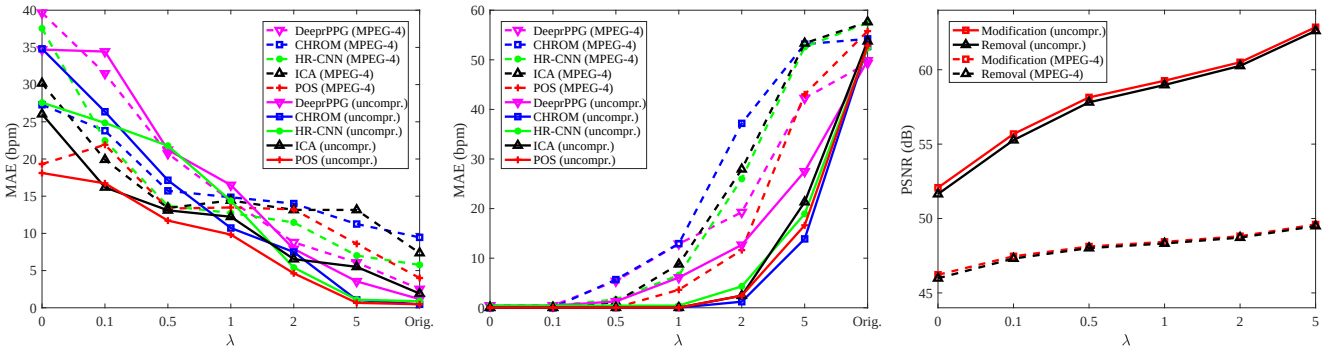
Input: Original video $\mathcal{I}$ containing N frames, $\mathcal{I} = [\mathcal{I}_1, \mathcal{I}_2, ..., \mathcal{I}_N]$, (frame dimension $h \times w \times 3$); perturbation signal δ (dimension $3 \times N$).
Output: PulseEdit video $\tilde{\mathcal{I}}$.

```
Δ ← ZeroLike(I)                                  ▷ memory allocation
for n = 1 → N do
    R_face ← FaceSkinPixel(I_n)                  ▷ detect skin pixels
    for c = 1 → 3 do                             ▷ each color channel
        for all (x, y) ∈ R_face do               ▷ each skin pixel
            p ← Rand(0, 1)
            if p < ⌈δ_c(n)⌉ − δ_c(n) then
                Δ_n(x, y, c) ← ⌊δ_c(n)⌋
            else
                Δ_n(x, y, c) ← ⌈δ_c(n)⌉
            end if
        end for
    end for
end for
Ĩ ← I + Δ
```

Equation (7) yields $p = \lceil\delta_c(n)\rceil - \delta_c(n)$. Algorithm 1 presents the detailed procedure of skin pixel adjustment to generate the final PulseEdit video.

IV. Experimental Results

In this section, we present experimental results on the PURE dataset [39] to demonstrate the effectiveness and robustness of PulseEdit in editing rPPG signals in facial videos. To further validate the forensic undetectability of PulseEdit when being used as a potential attack, we test the PulseEdit videos against digital forensic analysis. Lastly, we compare PulseEdit with the prior art of rPPG removal method [34] and study



(a) rPPG removal (b) rPPG modification (c) Frame-level PSNR

Fig. 3. Performance of PulseEdit on the PURE dataset with different λ: (a) HR estimation error in the rPPG removal mode with respect to the reference HR from pulse oximeter, (b) HR estimation error in the rPPG modification mode with respect to the target HR, and (c) average frame-level PSNR.

the influence of different subject motion settings in video recordings on the performance of rPPG removal. In the paper, we set $M = 6 \times 6 = 36$ and use Adam [38] to solve (6) with the learning rate 0.1 and the number of iterations 200.

A. Performance on PURE dataset

The PURE dataset [39] contains 60 facial video recordings of 640×480 pixel resolution and 30 frames per second (fps) in well-lit rooms from 10 subjects. Each subject was recorded in 6 different setups: steady, talking, slow translation, fast translation, small rotation, and medium rotation. The videos were stored without lossy compression. To validate the effectiveness of PulseEdit in editing rPPG signals in facial videos, we analyzed the PulseEdit outputs of the PURE videos with five representative rPPG algorithms: ICA [13], CHROM [2], POS [4], HR-CNN [7], and DeeprPPG [9]. The first three methods are classical signal processing methods and the last two are deep learning methods. For the HR-CNN method, we used the model provided by the authors[1], which was trained on the PURE dataset. For the DeeprPPG method, we re-implemented the rPPG extraction model and trained on the PURE dataset with an $80/20$ split for training and testing.

We extracted rPPG signal from the whole facial region in this part of the experiments to estimate HR, and evaluated the performance using mean absolute error (MAE). For the rPPG removal mode, we computed the error between the estimated HR from the video and the reference HR from pulse oximeter provided by the dataset. For the rPPG modification mode, we computed the error between the estimated HR from the video and the target HR.

We applied PulseEdit on the PURE videos for both the removal and modification modes. In the removal mode, we generated white Gaussian noise as the target rPPG signal T to remove the intrinsic rPPG signal in the original video. In the modification mode, we aimed at changing the rPPG signal to HR $= 120$ bpm as an example. We generated a sinusoid of frequency 120 bpm as the target rPPG signal T for all the color channels. To simulate the noise condition of rPPG signals, we added white Gaussian noise with -10 dB, 0 dB, and -10 dB in red, green, and blue channels, respectively, since the green channel generally contains the strongest level of pulse signal among all three channels [1]. We used the whole face region in rPPG analysis to estimate HR from facial videos.

We study the effect of different $\lambda = \{0, 0.1, 0.5, 1, 2, 5\}$ on the performance of PulseEdit, which governs the perturbation budget in the facial video. To investigate the robustness of PulseEdit against video lossy compression, we compressed the edited frames by MPEG-4 format at the average bitrate of around 500 kbps. Fig. 4 shows the qualitative comparison of the video frames and the corresponding rPPG spectrograms with different λ. Fig. 3(a) and (b) present the performance of HR estimation before and after PulseEdit in the removal and modification modes, respectively.

In the removal mode, we aim to increase the error of HR estimation with respect to the reference HR, and Fig. 3(a) shows that the error increases as λ decreases. When λ is less than 0.5, the rPPG-removed videos have a very large estimation error (*i.e.*, $>$ 10 bpm), indicating the successful removal of the intrinsic rPPG signal by PulseEdit. In the modification mode, our goal is to reduce the error of HR estimation with respect to the target HR, and Fig. 3(b) shows that the error is reduced as λ decreases. When λ is less than 0.5, the rPPG-modified videos have HR estimations very close to the target HR, with an error no more than 1 bpm for uncompressed videos and 10 bpm for MPEG-4 videos. This suggests that PulseEdit can effectively transform the rPPG signal in a video to a target HR. From Fig. 4, we observe that when λ increases from 0 to 5, the original rPPG signals gradually appear in the spectrograms of the edited videos. This indicates that we need to spend enough editing expense (smaller λ) in the video to successfully conceal the original rPPG signal.

Since lossy compression may attenuate the rPPG signal on the face, it is expected that the HR error is larger in MPEG-4 videos than in uncompressed videos. Specifically, in the rPPG modification mode, the HR error with respect to the target HR is larger in the MPEG-4 video than in the uncompressed video. Nevertheless, the modified rPPG signal of the target HR can still be detected by the rPPG methods within an acceptable error range, when we choose $\lambda < 0.5$. In comparison, lossy compression has less impact on the rPPG removal mode. Overall, these results indicate that although

[1]Model is available at https://cmp.felk.cvut.cz/%7espetlrad/ecg-fitness/

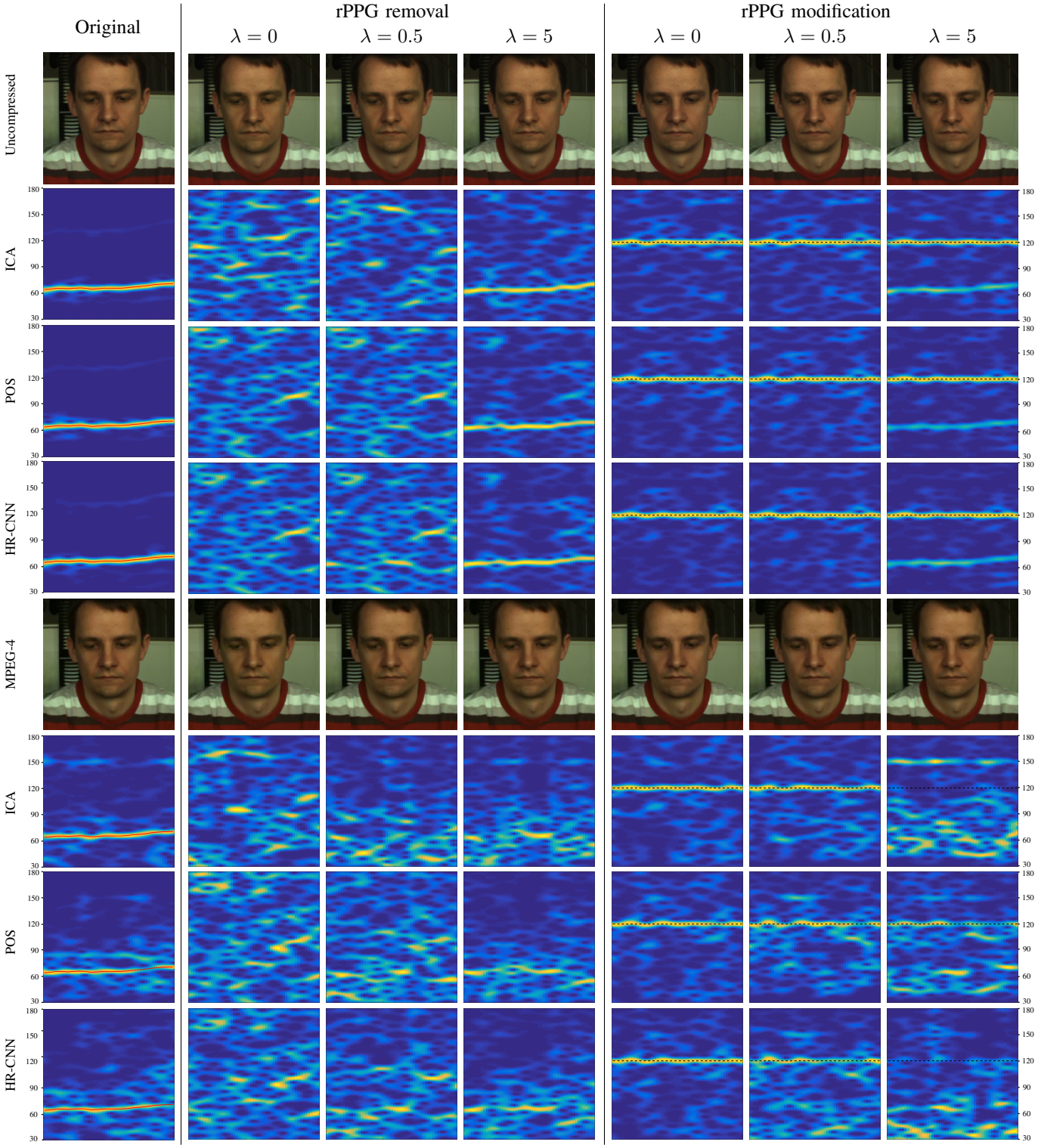
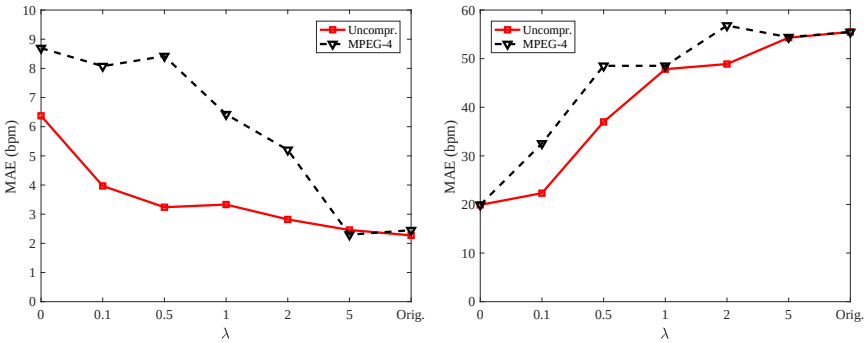


Fig. 4. Exemplary face crop from the videos and spectrograms of the rPPG extracted from the videos with two classical rPPG methods, ICA and POS, and one deep learning method, HR-CNN. We set the target HR $= 120$ bpm in the rPPG modification mode. The x-axis and y-axis denote the time and heart rate (30 bpm to 180 bpm), respectively. The red lines in the spectrograms of the original video indicate the reference HR from pulse oximeter and the black dash lines in the spectrograms of the rPPG modified videos indicate the target HR $= 120$ bpm. The figure is best viewed in color.

lossy compression can weaken the manipulations introduced by PulseEdit, the privacy protection of the intrinsic rPPG signal remains effective when choosing a proper λ.

An important observation is that the five rPPG methods have similar HR estimation performance on the PulseEdit videos, indicating that PulseEdit is effective to various rPPG algorithms, including the classical signal processing methods and the deep learning methods as well. This satisfies the



(a) rPPG removal (b) rPPG modification

Fig. 5. HR estimation of PulseEdit videos with different λ via a motion-based method: (a) rPPG removal mode and (b) rPPG modificaton mode.

Fig. 6. Illustration of the three facial subregions: forehead (red), left cheek (green), and right cheek (blue). The figure is best viewed in color.

"generality" requirement.

Fig. 3(c) shows the objective image quality assessment for the PulseEdit videos within the facial ROI with a size of 300×300. Since λ governs the editing strength in the video, frame-level PSNR increases when λ increases. By vision examination, we can hardly notice the distortion on the person's appearance shown in Fig. 4.

Motion-based physiological signal extraction. The prior art has demonstrated that physiological signals can also be extracted from facial videos via subtle head motions caused by ballistocardiogram (BCG). We evaluate how effective PulseEdit can remove heart rate information extracted using the motion-based method [3]. Since voluntary head motions can easily sabotage the subtle involuntary head motions induced by BCG, we analyze the steady cases for fair evaluations in Fig. 5.

From our intuition, PulseEdit may not perform well against motion-based methods, because it focuses on altering skin color and does not deliberately modify the subtle head motions in the steady facial videos. Nevertheless, we can observe that PulseEdit can still amplify the HR error estimated by the motion-based method though it can more effectively remove heart rate infromation obtained via rPPG extraction methods. One possible reason is that the imposed perturbation on pixels influences the estimated optical flow of the facial pixel points in tracking, degrading the pixel-level trajectory analysis of the involuntary subtle head motion.

Running time. Overall, PulseEdit runs efficiently. On average, the step of rPPG extraction runs at around 10 fps, the step of rPPG editing reaches 170 fps (the detrending runs at 200 fps and the optimization runs at 1000 fps, respectively), and the step of skin pixel adjustment runs at around 100 fps. These running times were measured using a single-core Python implementation on a PC with an Intel Core i5-4440 processor.

B. rPPG Analysis on Multiple Facial Subregions

To examine the universality of PulseEdit, we analyze the presence of rPPG signals in three facial subregions: forehead, and left and right cheek, shown in Fig. 6. The regions are detected automatically via the facial landmarks. Fig. 7 presents the performance of HR estimation from the three facial subregions using the five rPPG algorithms. For classical non-deep learning methods, we apply the algorithms within the selected subregions; for deep learning methods, we first warp the polygon regions to regular rectangles with the fitting input size, and then feed them into the models. Since a larger size of ROI generally gives a better average quality of rPPG extraction [40], we expect a reduced accuracy of HR estimation from facial subregions alone, compared with using the whole face region.

From Fig. 7, we observe that HR error from the three facial subregions has a similar trend as that from the whole face region under different λ values. For the rPPG-removed videos, the error is much larger than the original videos, when λ is less than 0.5. This suggests that the intrinsic rPPG signals are completely erased in all three facial subregions. For the rPPG-modified videos, the HR error with respect to the target HR is in an acceptable range, when λ is less than 0.5. We can see that the rPPG signals in all three facial subregions are successfully transformed to the target HR. In summary, these results indicate that PulseEdit can effectively edit the rPPG signals not only in the global facial region but also in local facial subregions, which satisfies the "universality" requirement.

PulseEdit computes the original skin intensity variations $R \in \mathbb{R}^{M \times 3 \times N}$ (M denotes the number of subregions) from multiple facial subregions and finds the optimal perturbation in (6) that can change the heartbeat information in the extracted local regions. This design can help the perturbation universally change the heartbeat information in the global face and the local facial regions.

C. User Study on Perceptual Distortion

We conducted a user study to investigate whether a human viewer can notice the perceptual distortion introduced by PulseEdit under different λ. Each question shows two videos, the original video and the edited video by PulseEdit, and provides three options: video one, video two, and "cannot determine". The respondents were asked to choose the original video from the two given videos. If they could not distinguish the two videos, they might select the "cannot determine" option. Thus, one respondent has three conditions for each question: select the correct video, cannot determine, or select

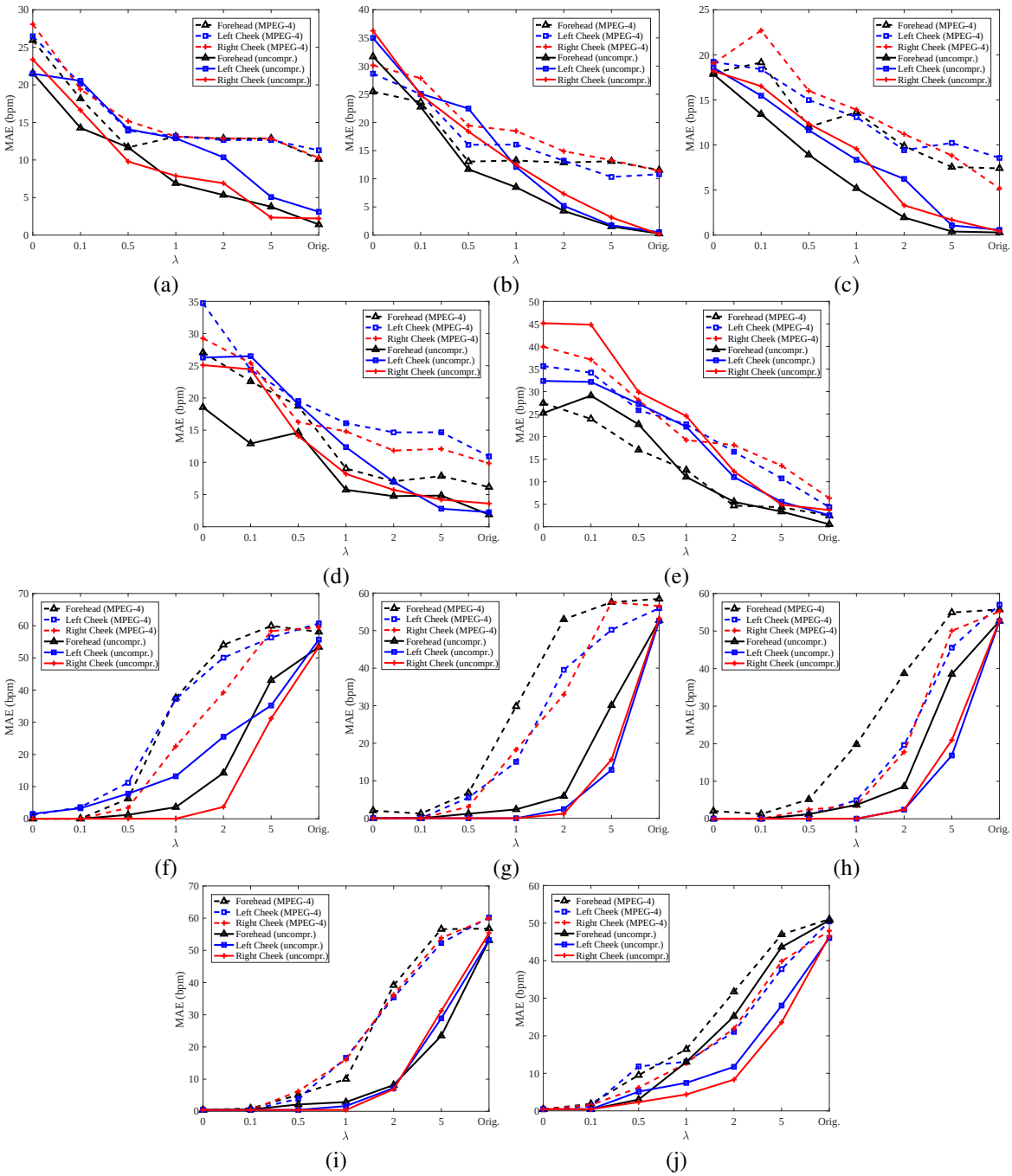


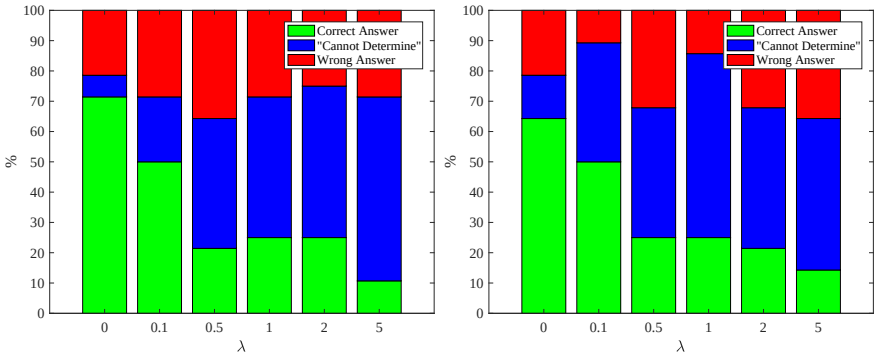
Fig. 7. HR estimation of PulseEdit videos on multiple face subregions with different λ using five rPPG methods: (a) ICA, (b) CHROM, (c) POS, (d) HR-CNN, (e) DeeprPPG in the rPPG removal mode, and (f) ICA, (g) CHROM, (h) POS, (i) HR-CNN, (j) DeeprPPG in the rPPG modification mode.

the wrong video. We collected 28 responses and present the survey result in Fig. 8 which illustrates the numerical proportion of the correct answer, the "cannot determine" option, and the wrong answer under different λ. The user study shows that more people could not distinguish between the original video and the edited video and fewer people could select the original video correctly as λ increases. This indicates that the large λ can reduce the perceptual distortion in human vision. There is an abrupt drop of correct answer rate at $\lambda = 0.5$, suggesting that $\lambda = 0.5$ is a good choice to balance the editing performance and the perceptual distortion.

D. PulseEdit against Forensic Analysis

From the previous performance analysis on PulseEdit, we can see that PulseEdit is effective in editing the intrinsic rPPG signals in facial videos for privacy protection. As motivated in Section I, it is possible to utilize PulseEdit in adversarial scenarios by forgers. In this subsection, we examine the resistance of PulseEdit against forensic analysis tools to help us understand its strengths and limitations.

PulseEdit perturbs the skin pixels by a small amount in the video frames to edit rPPG signals, which is similar to how steganography [43] manipulates the images. Based on



(a) rPPG removal (b) rPPG modification

Fig. 8. Results of user study on perceptual distortion in (a) the rPPG removal mode and (b) the rPPG modification mode.

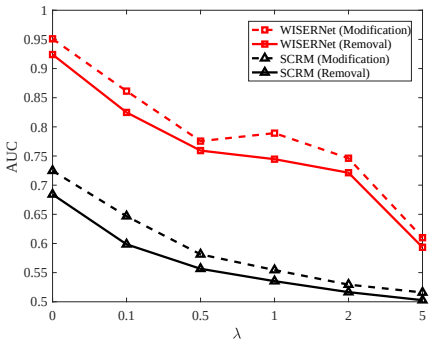


Fig. 9. Performance of PulseEdit against forensic analysis in MPEG-4 videos.

TABLE I
RESULT OF PULSEEDIT WITH $\lambda = 0.5$ ON HR ERROR, PERCEPTUAL DISTORTION, AND FORENSIC ANALYSIS

			rPPG removal				rPPG modification			
			Uncompressed		MPEG-4		Uncompressed		MPEG-4	
			Original	Edited	Original	Edited	Original	Edited	Original	Edited
MAE of HR estimation (bpm)	ICA	Face	1.91	13.70	7.36	13.41	53.75	0.03	57.63	1.12
		Forehead	1.41	11.68	10.12	11.67	53.36	1.23	58.09	6.23
		Left cheek	3.11	14.09	11.28	13.91	55.67	7.81	60.70	11.21
		Right cheek	2.21	9.80	10.28	15.19	53.47	0.03	59.50	3.41
	CHROM	Face	0.47	17.14	9.47	15.72	52.68	0.03	54.22	5.68
		Forehead	0.26	11.69	11.55	13.05	52.71	1.22	58.44	6.75
		Left cheek	0.52	22.46	10.80	16.03	52.57	0.03	55.95	5.51
		Right cheek	0.18	18.40	11.39	19.44	53.24	0.03	56.53	3.09
	POS	Face	0.47	12.33	4.00	13.34	52.84	0.03	55.79	0.03
		Forehead	0.29	8.91	7.41	12.01	52.67	1.22	55.73	5.18
		Left cheek	0.58	11.64	8.56	14.98	52.64	0.03	57.02	1.17
		Right cheek	0.42	12.35	5.18	16.01	52.84	0.03	55.23	2.44
	HR-CNN	Face	0.88	21.78	5.77	13.70	57.46	0.44	57.46	1.68
		Forehead	1.89	14.65	6.17	18.76	53.11	5.09	56.76	6.75
		Left cheek	2.25	19.02	10.93	19.53	60.22	0.44	55.95	3.91
		Right cheek	3.60	14.09	9.84	16.25	55.38	0.43	59.99	6.25
	DeeprPPG	Face	1.16	21.23	2.52	20.73	49.86	1.28	49.33	5.38
		Forehead	0.55	22.70	2.44	17.06	50.74	3.01	51.06	9.52
		Left cheek	2.55	27.28	4.33	25.87	46.06	5.13	50.45	11.84
		Right cheek	3.70	29.88	6.33	28.09	46.57	2.33	47.96	6.19
PSNR (dB) (Orig. as ref.)			ref.	57.81	ref.	48.01	ref.	58.14	ref.	48.12
PSNR (dB) (Uncompr. orig. as ref.)			ref.	57.81	33.02	33.04	ref.	58.14	33.02	33.05
SCRM [41] (AUC)			n/a	1.00	n/a	0.58	n/a	1.00	n/a	0.56
WISERNet [42] (AUC)			n/a	1.00	n/a	0.78	n/a	1.00	n/a	0.76

this point of view, we examine the forensic detectability of PulseEdit against two representative steganalysis methods: spatio–color rich model (SCRM) [41] with ensemble training [44], and WISERNet [42] based on deep learning. Since PulseEdit only edits the facial regions, we cropped facial ROI with a size of 300×300. We set the original video frames as negative and the PulseEdit video frames as positive, and used 5-fold cross-validation to evaluate the performance. For deep models, we changed the size of feature maps in the intermediate layers accordingly to cater to the input size of 300×300.

We observe that steganalysis models are effective on uncompressed video frames as their detection performance has an area under curve (AUC) of $0.99+$ for every λ value. They can almost perfectly differentiate the original video frames and the edited video frames by PulseEdit. Without incorporating additional constraints, the randomized pixel adjustment in Section III-C perturbs the skin pixels independently in the frame, introducing artificial changes among local neighboring pixels that are not presented in the direct output of video cameras. This kind of unconstrained distortion can be easily extracted by various image forensic models and discriminative

to natural images and edited images [45]–[47]. Fig. 9 presents the steganalysis results on the lossily compressed videos. Compared with the uncompressed videos, the steganalysis result of the MPEG-4 videos degrades in a noticeable amount. For the two steganalysis models, the deep model has a better ability to detect the manipulation trace in the lossily compressed videos than the classic model. We also find that the steganalysis performance on the lossily compressed videos decreases significantly in both forensic methods as λ increases. This suggests that lossy compression can alleviate the detectability of the manipulation traces in videos introduced by PulseEdit.

In the current form, PulseEdit focuses on altering the rPPG information for privacy protection and has not explicitly concealed the traces of manipulation. As such, the presence of perturbation can be detected from the uncompressed frames by such forensic tools as steganalysis. Because of the limitation of such forensic analysis for lossy compressed frames and the small and random perturbation of PulseEdit by design, a lossy compression on PulseEdit videos can evade forensic steganalysis and remain effective in concealing/modifying the intrinsic rPPG information. It is possible to further include various forensic undetectability into the algorithm, to gain insights on the ability of PulseEdit as an antiforensic tool and the competing direction of detecting the manipulations made by PulseEdit.

TABLE II
PERFORMANCE COMPARISON OF RPPG REMOVAL METHODS ON UBFC-RPPG DATASET

			Original	Chen's [34]	PulseEdit
MAE of HR estimation (bpm)	ICA	Face	0.90	16.78	19.84
		Forehead	0.43	15.86	21.48
		Left cheek	0.66	14.92	22.79
		Right cheek	1.50	16.00	18.51
	CHROM	Face	0.96	16.28	20.42
		Forehead	0.64	13.63	18.96
		Left cheek	1.30	14.84	13.89
		Right cheek	0.64	10.70	20.95
	POS	Face	0.87	18.17	17.31
		Forehead	0.64	16.79	17.53
		Left cheek	0.67	16.95	16.10
		Right cheek	0.82	15.70	20.84
	HR-CNN	Face	1.72	15.44	19.85
		Forehead	2.01	17.67	20.09
		Left cheek	1.23	14.62	19.25
		Right cheek	0.99	17.29	21.62
	DeeprPPG	Face	1.44	16.34	19.23
		Forehead	3.16	18.76	24.02
		Left cheek	1.59	17.11	23.32
		Right cheek	1.66	19.85	22.24
PSNR (dB)			ref.	43.64	52.83

E. Performance Summary of PulseEdit

Taking into consideration HR estimation error, perceptual distortion, and resistance of PulseEdit videos against forensics, we choose $\lambda = 0.5$ in PulseEdit and use it for the following experiments. We summarize the experimental results of PulseEdit with $\lambda = 0.5$ in Table I. The first five macro-rows show MAE of HR estimation (unit: bpm), using different rPPG algorithms. Note that, we compute MAE between the estimated HR and the reference HR from pulse oximeter in the rPPG removal mode, and compute MAE between the estimated HR and the target HR$= 120$ bpm in the rPPG modification mode. The next row shows frame-level perceptual distortion analysis within the facial ROI between original videos and edited videos. The last two rows present the forensic analysis on PulseEdit.

Table I shows that the error of HR estimation with respect to the reference HR from the facial video increases after PulseEdit in the rPPG removal mode; the error of HR estimation with respect to the target HR decreases significantly after PulseEdit in the rPPG modification mode. This indicates that the proposed PulseEdit can effectively remove/modify rPPG information both in the whole face sense and in the local subregion sense, tested by various rPPG methods. High PSNR index suggests that PulseEdit hardly introduces perceptual distortion on the subject's appearance. Comparing the HR estimation error between the uncompressed and MPEG-4 videos, we can see that lossy compression can weaken the manipulation applied in the facial videos, but PulseEdit can still edit the rPPG signals to some extent. From the perspective of antiforensics, the AUC index reduces more than 0.4 in the SCRM and more than 0.2 in the WISERNet. This indicates that lossy compression can greatly help PulseEdit videos defend forensic analysis.

F. Comparison with Prior Art

We compare the proposed PulseEdit in the rPPG removal mode with the prior art Chen's method [34]. We re-implemented Chen's method and tuned the hyperparameter to obtain the best performance. We report the HR error from the facial videos using the five rPPG methods: ICA, CHROM, POS, HR-CNN, and DeeprPPG. The performance is evaluated on the uncompressed videos.

Steady case. We compare the rPPG-removing methods under steady cases using the UBFC-RPPG dataset [48]. From Table II, we can see that the five rPPG methods can accurately estimate HR from the original videos. Given the fact that we can extract rPPG signals accurately from the original videos, Chen's methods and PulseEdit have the comparable capability to amplify the HR estimation error in steady cases. The PSNR index indicates that the proposed PulseEdit has less distortion than Chen's method on the video frames.

Realistic case. We compare the rPPG-removing methods under realistic cases using the PURE dataset [39]. Realistic cases reflect the facial conditions in practical applications, including steady cases and motion cases. Similarly, we can see that the five rPPG methods can accurately estimate HR from the original videos in Table III. Given the fact that we can extract rPPG signals accurately from the original videos, PulseEdit has larger amplification of HR error than Chen's method in realistic cases, which indicates that PulseEdit has better editing performance to remove the intrinsic rPPG signal

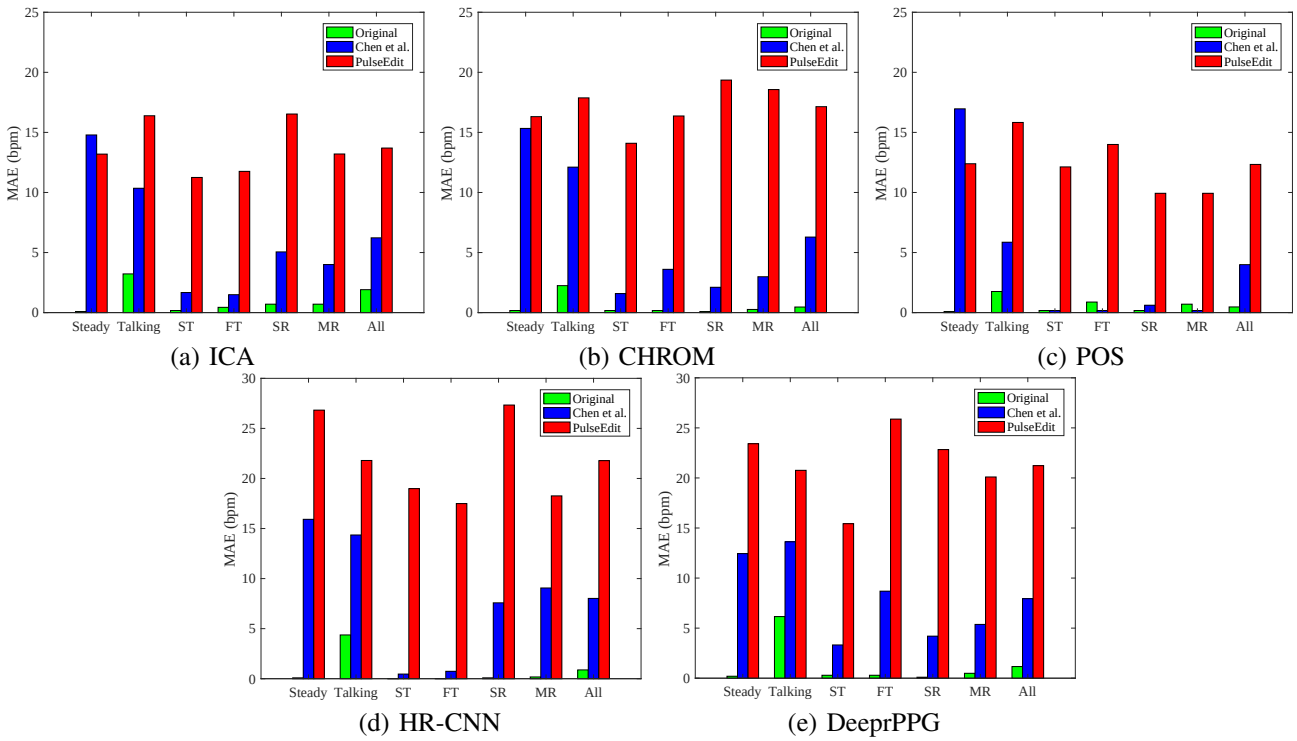


Fig. 10. Performance comparison between PulseEdit and Chen's method in different subject motion settings with five rPPG methods: (a) ICA, (b) CHROM, (c) POS, (d) HR-CNN, and (e) DeeprPPG. The motion settings are steady, talking, slow translation (ST), fast translation (FT), small rotation (SR), and medium rotation (MR).

TABLE III
PERFORMANCE COMPARISON OF RPPG REMOVAL METHODS ON PURE DATASET

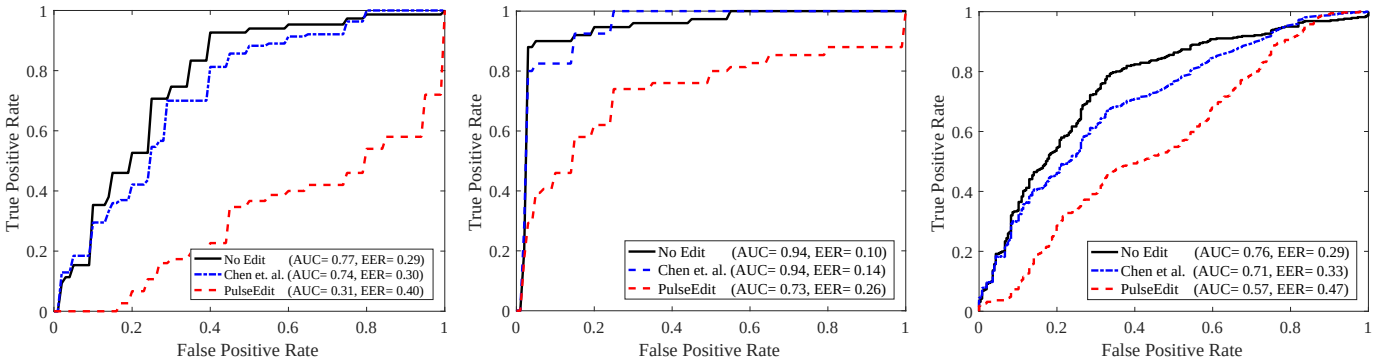
			Original	Chen's [34]	PulseEdit
MAE of HR estimation (bpm)	ICA	Face	1.91	6.23	13.70
		Forehead	1.41	6.32	12.68
		Left cheek	3.11	6.38	14.09
		Right cheek	2.21	6.04	15.80
	CHROM	Face	0.47	6.29	17.14
		Forehead	0.26	6.73	16.69
		Left cheek	0.52	4.75	22.46
		Right cheek	0.18	13.34	18.40
	POS	Face	0.47	3.99	11.72
		Forehead	0.29	4.57	14.91
		Left cheek	0.58	5.36	11.64
		Right cheek	0.42	8.34	12.35
	HR-CNN	Face	0.88	8.02	17.78
		Forehead	1.89	6.76	14.65
		Left cheek	2.25	11.31	19.02
		Right cheek	3.60	13.50	14.09
	DeeprPPG	Face	1.16	7.94	21.23
		Forehead	0.55	11.56	22.70
		Left cheek	2.55	10.18	22.28
		Right cheek	3.70	7.40	21.88
PSNR (dB)			ref.	47.01	57.81

in facial videos. The PSNR index indicates that the proposed PulseEdit has less distortion than Chen's method on the video frames.

Fig. 10 presents barplots of performance comparison between the proposed PulseEdit and Chen's method regarding 6 motion settings: steady, talking, slow translation, fast translation, small rotation, and medium rotation. We can observe that Chen's method has similar performance to our method in the steady case but does not perform well in the talking, head translation, and head rotation cases.

Overall, the two methods have the similar performance of rPPG removal in steady cases, but Chen's method is not effective when dealing with head motions. In comparison, PulseEdit has little performance variation in steady cases and motion cases, indicating that our proposed method is effective in a variety of motion settings.

Chen's method first estimates the color intensity variations from pixel level, and then subtracts the pixel-wise intensity variations from the original video to remove the physiological signals. The color intensity variation in each facial pixel is a combined consequence of pulse-induced color variation and voluntary motion. For head motion cases, the color intensity variation caused by voluntary motions can easily overwhelm the pulse-induced color variation in the video. This could explain the reason why Chen's method has ineffective editing performance for motion cases. In contrast, PulseEdit tracks multiple facial subregions and then extracts the pulse-induced skin color variations from them. The tracking of facial subregions can alleviate the interference of color variations caused by voluntary motions. Also, PulseEdit finds the optimal perturbation such that the original rPPG signals in multiple facial subregions can directly transfer to the target rPPG signal. Hence, PulseEdit can deal with both steady and motion cases.



(a) Liveness by Hernandez *et al.* (b) Liveness by Liu *et al.* (c) FakeCatcher

Fig. 11. ROC curves of (a) Hernandez's and (b) Liu's methods of rPPG-based liveness detection, and (c) FakeCatcher of rPPG-based deepfake detection before and after PulseEdit

V. Analysis of Adversarial Scenarios

Since PulseEdit can edit rPPG signals in facial videos, we expect that PulseEdit, as an adversarial operation, can circumvent rPPG-based liveness detection [16], [18] and rPPG-based deepfake detection [19]. Thus, we conducted experiments on the HKBUMARsV1+ dataset [17] for liveness detection and the Celeb-DFv1 dataset [49] for deepfake detection to evaluate the effectiveness of PulseEdit on above two aspects, respectively.

A. Analysis against rPPG-based Liveness Detection

Liveness detection aims at detecting whether a person seen by a camera is in his/her true live appearance or wearing a camouflaging mask with different facial appearances, a profile photo, or a video replay, to prevent face spoofing in identity authentication. Since live faces and many spoofed faces often have different characteristics in rPPG features extracted from the facial area, several prior publications have presented classifiers based on rPPG features. We test two rPPG-based liveness detection methods, namely, Hernandez's method [18] and Liu's method [16], as a proof-of-concept, to analyze the performance of PulseEdit on circumventing the rPPG-based methods.

We conducted experiments on the HKBUMARsV1+ dataset [17], which consists of video recordings from 12 subjects in flesh and wearing 3D face masks of different appearances. We set live facial videos as negative and 3D mask videos as positive. The classifier settings are the same as stated in [16], [18]. PulseEdit was applied to the 3D mask videos, with the target rPPG signals generated using the same procedure as in the rPPG modification mode in Section IV-A. We used subject-based 5-fold cross-validation to evaluate the performance of the detector on the videos before and after PulseEdit.

We report the equal error rate (EER) and AUC in Fig. 11 to show the impact of PulseEdit on the rPPG-based liveness detection algorithms. EER refers to the point where false positive rate and false negative rate are equal. AUC refers to the area under the receiver operating characteristic (ROC) curve. Smaller EER and larger AUC indicate better detection ability. We can see that PulseEdit increases the EER from 0.29 to 0.40 and decreases the AUC from 0.77 to 0.31 for Hernandez's method [18], and increases the EER from 0.10 to 0.26 and decreases the AUC from 0.94 to 0.73 for Liu's method [16]. These results suggest that the current form of PulseEdit can already circumvent the rPPG-based liveness detection to some extent and additional optimization may enhance such evasion by incorporating information from the existing research of liveness detection.

B. Analysis against rPPG-based Deepfake Detection

The fast development of deep learning enables computers to transform a person's face to another's in images and videos. These "deepfake" videos can spread misinformation and fake news and impair the integrity of social media, prompting a strong and urgent need of developing the detection algorithms for deepfake videos [50]. Recently, FakeCatcher [19] was proposed to utilize rPPG signals from the video as features to detect whether the video is real or deepfake. To analyze the effectiveness of PulseEdit, we tested PulseEdit videos using the FakeCatcher CNN model.

We conducted experiments on the Celeb-DFv1 dataset [49], which consists of 370 real videos and 733 deepfake videos in the training set, and 38 real videos and 62 deepfake videos in the test set. We considered real videos as negative and fake videos as positive, and trained the FakeCatcher CNN model in the training set. The CNN architecture is the same as stated in [19]. As shown in Fig. 11(c), FakeCatcher achieves an EER of 0.29 and an AUC of 0.76 in the test set.

We applied PulseEdit on the deepfake videos in the test set, with the rPPG signals extracted from the corresponding real videos as the target rPPG signals. In other words, we tried to restore the original rPPG signal in the deepfake videos. From the classification performance of the FakeCatcher on the test set with PulseEdit, we observe that the EER increases to 0.47 and the AUC reduces to 0.57, indicating that the rPPG signals inserted by PulseEdit can circumvent FakeCatcher, making it consider the deepfake videos as trustworthy. The above observations show that PulseEdit can degrade the reliability of the FakeCatcher classifier and fool it to make wrong decisions on the deepfake videos.

C. Comparison with Prior Art

We analyze the prior art, Chen's method [34], in the above adversarial scenarios. Based on EER and AUC indices from Fig. 11, we can see that there is no significant performance degradation on rPPG-based visual security algorithms when we process the fake/synthesized videos using Chen's method [34].

The rPPG-based visual security algorithms utilize the feature discrepancy between the extracted rPPG signals from real videos and fake/synthesized videos to do the classification. Typically, the real videos contain meaningful rPPG signals while the fake/synthesized videos may mainly contain noise. PulseEdit in the modification mode can synthesize designed physiological signals for fake/synthesized videos to fool the rPPG-based visual security algorithms. In constrast, Chen's method [34] was mainly designed removing the physiological signals from the facial videos. Since fake/synthesized videos do not contain rPPG signals already, removing the rPPG signals does not change the feature of the extracted rPPG signal in fake/synthesized videos. This explains why Chen's method [34] is not an effective adversarial tool. Overall, our proposed PulseEdit provides a better adversarial capability to circumvent rPPG-based visual security algorithms.

VI. Discussions

In terms of the running time and the HR estimation error of PulseEdit, the proposed PulseEdit is an effective algorithm to edit rPPG signal in facial videos. Compared with the prior art [34] that only focuses on eliminating the rPPG information, we have designed PulseEdit with two modes: rPPG removal and rPPG modification. The former mode can remove the rPPG information and the latter mode can change the rPPG information to a target HR of user's choice. The proposed algorithm offers the users more options of editing operations on the physiological signal in facial videos regarding physiological privacy protection. PulseEdit also provides a better capability to remove the physiological signal from videos with head motions (*i.e.*, talking, translation, and rotation), more robust to deal with practical recording scenarios.

Considering PulseEdit as an adversarial operation to the rPPG technology, we have studied to what extent PulseEdit can circumvent rPPG-based visual security algorithms. As a proof-of-concept, we considered the rPPG-based liveness detection and deepfake detection algorithms. The experimental results demonstrate noticeable performance drops between the spoofed videos before and after PulseEdit processes them, indicating that PulseEdit can successfully mitigate the rPPG-based visual security algorithms. From the perspective of threat modeling for these visual security algorithms, our PulseEdit research suggests that it is important to investigate this and other similar vulnerabilities and improve the rPPG-based visual security algorithms against adversarial operations.

Over the past decade, rPPG technology has been prospering and it is becoming feasible to monitor vital signs, such as HR, using commercial digital cameras in daily life. One common bottleneck in the R&D of rPPG technology is the lack of sufficient facial videos with known HR of a wide range [51]. PulseEdit in the rPPG modification mode may be used to synthesize facial videos with controllable HR to enlarge the dataset and facilitate the R&D of rPPG technology.

There are some potential directions to improve the proposed algorithm. PulseEdit has a hyperparameter λ to balance the editing performance and the perceptual distortion. For facial videos with different skin tones and illumination conditions, the λ for the optimal performance is different. So far, to find the optimal editing performance for each video, we can heuristically try λ in ascending order until the optimal perturbation is less than a maximum intensity threshold or the perceptual distortion can not be discovered by human examination. In future research, the adaptive λ method is one direction to improve the overall performance of PulseEdit.

Our algorithm focuses on altering rPPG information for physiological privacy protection. Physiological information can also be extracted from facial videos via BCG or involuntary subtle head motions. It is interesting to develop BCG editing algorithm to synergically edit the physiological information from facial videos in parallel with PulseEdit.

In the current form, one limitation of PulseEdit as an adversarial tool for video forgery is that we have not explicitly conceal the manipulation traces introduced by itself. Forensic tools such as steganalysis can detect the presence of perturbation from the uncompressed frames if available. Nevertheless, we have found that lossy video compression is a feasible approach to improve the resistance of the edited frames against forensic analysis and retain the edited rPPG signal in the video. In future work, the inclusion of various forensic undetectability into the framework of PulseEdit and the development of new detectors to detect these manipulations may be two intertwining research directions. In addition, beyond the current form of PulseEdit perturbing the facial pixels independently, the future algorithm can take spatial and temporal correlations of facial pixels into consideration for the pixel perturbation to further minimize the perceptual distortion of facial videos.

VII. Conclusion

In this paper, we have proposed *PulseEdit*, a novel algorithm that can edit the rPPG signal in facial videos without visible distortion, to protect the physiological information from disclosure. We design a set of perturbation frames and impose them onto the input video frames to change a person's intrinsic rPPG signal present in the facial region. PulseEdit can either remove the rPPG signals on the face or change them to a target heart rate of a user's choice. Extensive experimental results demonstrate the effectiveness and robustness of PulseEdit in different facial subregions, and various rPPG algorithms can no longer detect heart rate accurately from facial videos after PulseEdit. We also show that PulseEdit can potentially circumvent rPPG-based liveness detection and deepfake detection, suggesting a direction for improvement in these areas.

Several improvements on PulseEdit can be explored in future research. Adaptive λ could be better than fixed λ to optimize the editing performance and the perceptual distortion in each video. Apart from rPPG editing, BCG editing algorithm can be developed to synergically edit the physiological

information from facial videos in parallel with PulseEdit. The inclusion of various forensic detectability criteria into the algorithm can help gain insights into the ability of PulseEdit as an antiforensic tool and the competing direction of detecting the manipulations made by PulseEdit.

REFERENCES

[1] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[2] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, 2013.

[3] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2013, pp. 3430–3437.

[4] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, 2016.

[5] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2014, pp. 4264–4271.

[6] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2016.

[7] R. Špetlík, V. Franc, and J. Matas, "Visual heart rate estimation with convolutional neural network," in *British Machine Vision Conference*, 2018.

[8] Q. Zhu, M. Chen, C.-W. Wong, and M. Wu, "Adaptive multi-trace carving based on dynamic programming," in *Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 1716–1720.

[9] S.-Q. Liu and P. C. Yuen, "A general remote photoplethysmography estimator with spatiotemporal convolutional network," in *International Conf. on Automatic Face and Gesture Recognition*, 2020, pp. 481–488.

[10] Q. Zhu, M. Chen, C.-W. Wong, and M. Wu, "Adaptive multi-trace carving for robust frequency tracking in forensic applications," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1174–1189, 2021.

[11] M. Chen, Q. Zhu, H. Zhang, M. Wu, and Q. Wang, "Respiratory rate estimation from face videos," in *IEEE EMBS Conf. on Biomedical and Health Informatics*, 2019, pp. 1–4.

[12] M. Chen, Q. Zhu, M. Wu, and Q. Wang, "Modulation model of the photoplethysmography signal for vital sign extraction," *IEEE J. Biomed. and Health Informatics*, 2020.

[13] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2010.

[14] D. J. McDuff, J. Hernandez, S. Gontarek, and R. W. Picard, "COGCAM: Contact-free measurement of cognitive stress during computer tasks with a digital camera," in *CHI Conf. on Human Factors in Computing Systems*, 2016, pp. 4000–4004.

[15] B.-F. Wu, C.-H. Lin, P.-W. Huang, T.-M. Lin, and M.-L. Chung, "A contactless sport training monitor based on facial expression and remote-PPG," in *IEEE Conf. on Systems, Man, and Cybernetics*, 2017, pp. 846–851.

[16] S. Liu, P. C. Yuen, S. Zhang, and G. Zhao, "3D mask face anti-spoofing with remote photoplethysmography," in *European Conf. on Computer Vision*, 2016, pp. 85–100.

[17] S. Liu, X. Lan, and P. C. Yuen, "Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection," in *European Conf. on Computer Vision*, 2018, pp. 558–573.

[18] J. Hernandez-Ortega, J. Fierrez, A. Morales, and P. Tome, "Time analysis of pulse-based face anti-spoofing in visible and NIR," in *IEEE Conf. on Computer Vision and Pattern Recognition Workshop*, 2018, pp. 544–552.

[19] U. A. Ciftci, I. Demir, and L. Yin, "FakeCatcher: Detection of synthetic portrait videos using biological signals," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.

[20] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graphics*, vol. 31, no. 4, 2012.

[21] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Robust heart rate from fitness videos," *Physiological Measurement*, vol. 38, no. 6, pp. 1023–1044, 2017.

[22] Q. Zhu, C.-W. Wong, C.-H. Fu, and M. Wu, "Fitness heart rate measurement using face videos," in *IEEE Conf. on Image Processing*, 2017, pp. 2000–2004.

[23] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security & Privacy*, vol. 1, no. 2, pp. 33–42, 2003.

[24] M. Barni, G. Droandi, and R. Lazzeretti, "Privacy protection in biometric-based recognition systems: A marriage between cryptography and signal processing," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 66–76, 2015.

[25] L. Du, M. Yi, E. Blasch, and H. Ling, "Garp-face: Balancing privacy protection and utility preservation in face de-identification," in *IEEE International Joint Conference on Biometrics*, 2014, pp. 1–8.

[26] M. Gomez-Barrero, C. Rathgeb, J. Galbally, J. Fierrez, and C. Busch, "Protected facial biometric templates based on local gabor patterns and adaptive bloom filters," in *International Conference on Pattern Recognition*, 2014, pp. 4483–4488.

[27] S. Cimato, M. Gamassi, V. Piuri, R. Sassi, and F. Scotti, "A multi-biometric verification system for the privacy protection of iris templates," in *International Workshop on Computational Intelligence in Security for Information Systems*, 2009, pp. 227–234.

[28] S. Li and A. C. Kot, "Fingerprint combination for privacy protection," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 2, pp. 350–360, 2012.

[29] V. Talreja, M. C. Valenti, and N. M. Nasrabadi, "Deep hashing for secure multimodal biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1306–1321, 2020.

[30] A. Peña, J. Fierrez, A. Morales, and A. Lapedriza, "Learning emotional-blinded face representations," in *International Conference on Pattern Recognition*, 2021, pp. 3566–3573.

[31] B. Bortolato, M. Ivanovska, P. Rot, J. Križaj, P. Terhörst, N. Damer, P. Peer, and V. Štruc, "Learning privacy-enhancing face representations through feature disentanglement," in *International Conf. on Automatic Face and Gesture Recognition*, 2020, pp. 45–52.

[32] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "SensitiveNets: Learning agnostic representations with application to face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 2158–2164, 2020.

[33] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unsupervised privacy-enhancement of face representations using similarity-sensitive noise transformations," *Applied Intelligence*, vol. 49, no. 8, pp. 3043–3060, 2019.

[34] W. Chen and R. W. Picard, "Eliminating physiological information from facial videos," in *IEEE Conf. on Automatic Face and Gesture Recogition*, 2017, pp. 48–55.

[35] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.

[36] J.-C. Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images," in *International Conf. on Automatic Face and Gesture Recogition*, 2000, pp. 54–61.

[37] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "L1 trend filtering," *SIAM review*, vol. 51, no. 2, pp. 339–360, 2009.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Internation Conf. on Learning Representation*, 2015.

[39] R. Stricker, S. Müller, and H.-M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 1056–1062.

[40] S. Kwon, J. Kim, D. Lee, and K. Park, "ROI analysis for remote photoplethysmography on facial video," in *Annual Conf. of the IEEE Engineering in Medicine and Biology Society*, 2015, pp. 4938–4941.

[41] M. Goljan, J. Fridrich, and R. Cogranne, "Rich model for steganalysis of color images," in *IEEE Workshop on Information Forensics and Security*, 2014, pp. 185–190.

[42] J. Zeng, S. Tan, G. Liu, B. Li, and J. Huang, "WISERNet: Wider separate-then-reunion network for steganalysis of color images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 10, pp. 2735–2748, 2019.

[43] X. Liao, J. Yin, M. Chen, and Z. Qin, "Adaptive payload distribution in multiple images steganography based on image texture features," *IEEE Trans. Dependable Secure Comput.*, 2020.

[44] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 432–444, 2011.

[45] A. Swaminathan, M. Wu, and K. R. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 101–117, 2008.
[46] M. C. Stamm, M. Wu, and K. R. Liu, "Information forensics: An overview of the first decade," *IEEE Access*, vol. 1, pp. 167–200, 2013.
[47] A. Swaminathan, M. Wu, and K. R. Liu, "Component forensics," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 38–48, 2009.
[48] 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.
[49] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A large-scale challenging dataset for deepfake forensics," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2020, pp. 3207–3216.
[50] J. Hu, X. Liao, W. Wang, and Z. Qin, "Detecting compressed deepfake videos in social networks using frame-temporality two-stream convolutional network," *IEEE Trans. Circuits Syst. Video Technol.*, 2021.
[51] U. A. Ciftci and L. Yin, "Heart rate based face synthesis for pulse estimation," in *International Symposium on Visual Computing*, 2019, pp. 540–551.