Video-based heart rate estimation from challenging scenarios using synthetic video generation

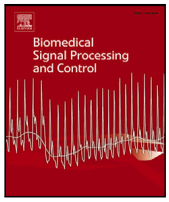
Yannick Benezeth, Deepak Krishnamoorthy, Deivid Johan Botina Monsalve, Keisuke Nakamura, Randy Gomez, Johel Mitéran

Biomedical Signal Processing and Control 96 (2024) 106598

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Video-based heart rate estimation from challenging scenarios using synthetic video generation

Yannick Benezeth [a,*], Deepak Krishnamoorthy [b], Deivid Johan Botina Monsalve [a], Keisuke Nakamura [c], Randy Gomez [c], Johel Mitéran [a]

[a] ImViA EA7535, Université de Bourgogne, Dijon, France
[b] Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, 601103, India
[c] Honda Research Institute Japan Co., Honcho, Wako-shi, Saitama, Japan

ARTICLE INFO

Keywords:
rPPG estimation
Data augmentation
Near-infrared
Fitness scenarios

ABSTRACT

Remote photoplethysmography (rPPG) is an emerging technology that allows for non-invasive monitoring of physiological signals such as heart rate, blood oxygen saturation, and respiration rate using a camera. This technology has the potential to revolutionize healthcare, sports science, and affective computing by enabling continuous monitoring in real-world environments without the need for cumbersome sensors. However, rPPG technology is still in its early stages. It faces challenges such as motion artifacts, low signal-to-noise ratio, and the challenge of conducting near-infrared measurements in low-light or nighttime conditions. The performance of existing rPPG techniques has been significantly improved by deep learning approaches, primarily due to the availability of large public datasets. However, most of these datasets are limited to the regular RGB color modality, with only a few available in near-infrared. Additionally, training deep neural networks for specific applications with distinctive movements, such as sports and fitness, would require extensive amounts of video data to achieve optimal specialization and efficiency, which can be prohibitively expensive. Therefore, exploring alternative methods to augment datasets for specific applications is crucial to improve the performance of deep neural networks in rPPG. In response to these challenges, this paper presents a novel methodology to generate synthetic videos for pre-training deep neural networks to estimate heart rates from videos captured under challenging conditions accurately. We have evaluated this approach using two near-infrared publicly available datasets, *i.e.* MERL (Nowara et al., 2020) and Tokyotech (Maki et al., 2019), and one challenging fitness dataset, *i.e.* ECG-Fitness (Špetlík et al., 2018). Furthermore, we have collected and made publicly available a novel collection of near-infrared videos named IMVIA-NIR. Our data augmentation strategy involves generating video sequences that animate a person in a source image based on the motion captured in a driving video. Furthermore, we integrate a synthetic rPPG signal into the faces, considering various important aspects such as the temporal shape of the signal, its spatial and spectral distribution, as well as the distribution of heart rates. This comprehensive integration process ensures a realistic incorporation of the rPPG signals into the synthetic videos. Experimental results demonstrated a significant reduction in the mean absolute error (MAE) score on all datasets. Overall, this approach provides a promising solution for improving the performance of deep neural networks in rPPG under challenging conditions.

* Corresponding author.
E-mail address: yannick.benezeth@u-bourgogne.fr (Y. Benezeth).

https://doi.org/10.1016/j.bspc.2024.106598
Received 20 November 2023; Received in revised form 27 May 2024; Accepted 25 June 2024
Available online 9 July 2024

1. Introduction

In recent years, there have been significant advancements in contactless technologies for measuring physiological signals (*e.g.* [1–3]). Remote photoplethysmography (or rPPG) is one such method that remotely measures subtle skin color fluctuations, which reflect complex light-tissue interactions [4]. rPPG signals can be reliably measured using webcams or more advanced professional cameras. Due to its non-invasive and contactless nature, rPPG technology has numerous potential applications. One of the most promising applications is in healthcare, where rPPG can be used for continuous monitoring of cardiovascular function in patients including newborns [5] or patients with skin pathologies that prevent the use of contact sensors [6]. In addition, rPPG can be used in sports and fitness applications for monitoring athletes' heart rates [7], emotion detection and stress monitoring, security and surveillance applications such as monitoring drivers for fatigue, or even for human–computer interaction [8]. With its diverse range of potential applications, rPPG technology has the potential to revolutionize various fields.

The initial approach for rPPG signal estimation relied solely on the green channel [9]. Later, techniques using blind source separation, such as PCA [10], ICA [11], constrained ICA [12,13], PVM [14], and those based on light-tissue interaction models, such as PbV [15], POS [16], and CHROM [17], were proposed. Recently, some methods have adopted deep neural networks for physiological measurements from video sequences, leveraging their strong and non-linear modeling abilities. These methods produce good results without requiring in-depth problem analysis from the designer [18], which can be critical for rPPG technology.

In 2018, the pioneering architecture featuring a series of 2D convolution layers and an attention mechanism was introduced [19]. Subsequently, numerous more sophisticated architectures have emerged, enhancing performance. These models typically combine spatial and temporal modules; the former employs face tracking to extract facial frames from videos [20–22], sometimes using predefined regions of interest [23–26]. Certain methods, like Deep-HR [27] and DeepPhys [19], utilize CNNs to analyze the spatial rPPG signal distribution. In addition to spatial information, most deep learning architectures utilize a temporal module to process features along the temporal dimension. Several researchers have introduced prior knowledge to the neural networks by computing spatial–temporal maps [23,25,26,28,29]. These maps are generated by manipulating the spatial regions of interest and concatenating them temporarily. End-to-end solutions for rPPG signal extraction from videos have also been developed, often using 3D-CNNs for simultaneous spatial–temporal analysis, although real-time processing remains a challenge due to high computational costs [20, 30–34].

As mentioned earlier, recent rPPG methods have shown high performance in some scenarios due to the availability of numerous datasets. In 2017, UBFC-rPPG dataset [35] containing approximately 50 relatively simple videos was released, and since then, several other databases dedicated to rPPG applications, with several hundred or even thousands of videos, have been proposed (*e.g.* in [25]). Although these datasets have significantly contributed to the success of recent rPPG methods, they still have some limitations. For instance, a notable limitation is the lack of sufficient representation of individuals with dark skin tones. For example, the UBFC-rPPG datasets primarily consist of participants with white skin tones, whereas the majority of participants in the VIPL-HR dataset are Asian. Furthermore, there need to be more datasets with challenging movements like ECG-Fitness [22]. There are numerous applications of rPPG technology in a fitness context. Still, they require improvements in the robustness of current models to movement, especially periodic movements whose frequency may be close to the heart rate, making separating the two signals challenging. Utilizing large databases that include these periodic movements to train the deep learning models could enable them to be trained to remove the movement component effectively. This training approach would significantly enhance the models' ability to capture heart rate signals despite confounding movement artifacts accurately. Moreover, there also need to be more datasets with near-infrared (NIR) videos. Estimating heart rate from NIR videos presents a particularly challenging task. This is primarily due to the lower sensitivity of cameras in these wavelengths and weaker absorption of hemoglobin, resulting in weaker temporal variations due to blood perfusion. Consequently, the signal-to-noise ratio is much less favorable than in color videos, making it more challenging to estimate heart rate from NIR videos accurately. Despite these challenges, enhancing rPPG model performance in the NIR is critical and would be beneficial to multiple applications. First, NIR cameras with invisible NIR illumination can robustly operate under various lighting conditions. This is particularly advantageous in unstable lighting environments, such as in driving scenarios [36]. Another important application of rPPG in NIR imaging is for remote monitoring of vital signs in darkness, such as for unobtrusive sleep monitoring. The availability of new massive datasets specifically tailored to address these particular challenges would undoubtedly enhance the performance of existing models. However, it is essential to note that acquiring a substantial number of videos for such datasets incurs significant costs in a project.

For all these reasons, researchers have proposed methods to augment existing datasets. One of the main advantages of using data augmentation (DA) is that it can increase the amount and diversity of training data [37], which can help to improve the performance of the deep-learning models. For example, by applying various transformations to the original videos, such as rotation, scaling, and flipping, it is possible to generate many new synthetic videos that can be used to train a model to estimate heart rate from videos. Additionally, data augmentation can help to reduce overfitting, which occurs when a model is too closely fit to the training data and performs poorly on new, unseen data. Recently, more advanced methods dedicated to the rPPG application have been proposed for augmenting datasets through generative models [38] or computer graphics simulations [39]. Synthetic approaches offer the flexibility to simulate various appearances and precisely control all synthetic video parameters. However, it is hard to generate completely realistic videos with computer-graphic tools, and a significant "sim-to-real" performance gap can be expected, given that models trained exclusively on computer-graphics-generated data frequently struggle to generalize successfully to real-world videos [40]. Other approaches focus on improving the training process; for example, Zije et al. [41] utilized a self-supervised learning approach, employing data augmentation to generate positive and negative samples that match the signal frequencies of a given video sample. Tsou et al. [42] proposed a multi-task learning-based video augmentation technique, simultaneously augmenting the training data while learning the rPPG estimation model. Birla et al. [43] introduced contrastive learning, utilizing temporal scaling-based data augmentation to overcome the skewed training data distribution. Additionally, Song et al. [29] utilized data augmentation through transfer learning with synthetic rPPG signals to train their CNN model. Each approach showcases different strategies for augmenting data to enhance the robustness of their models.

This paper introduces a novel approach for augmenting small datasets, specifically aimed at training deep learning models for heart rate estimation in challenging scenarios. The main contributions of this study are summarized as follows:

- We propose a new methodology dedicated to augmenting small rPPG datasets. Our method is based on generating animation from a single image and a video of another person and integrating the synthetic rPPG signal into the faces, considering various essential aspects such as the temporal shape of the signal, its spatial and spectral distribution, and the distribution of heart rates. This comprehensive integration process ensures a realistic incorporation of the rPPG signals into the augmented videos.

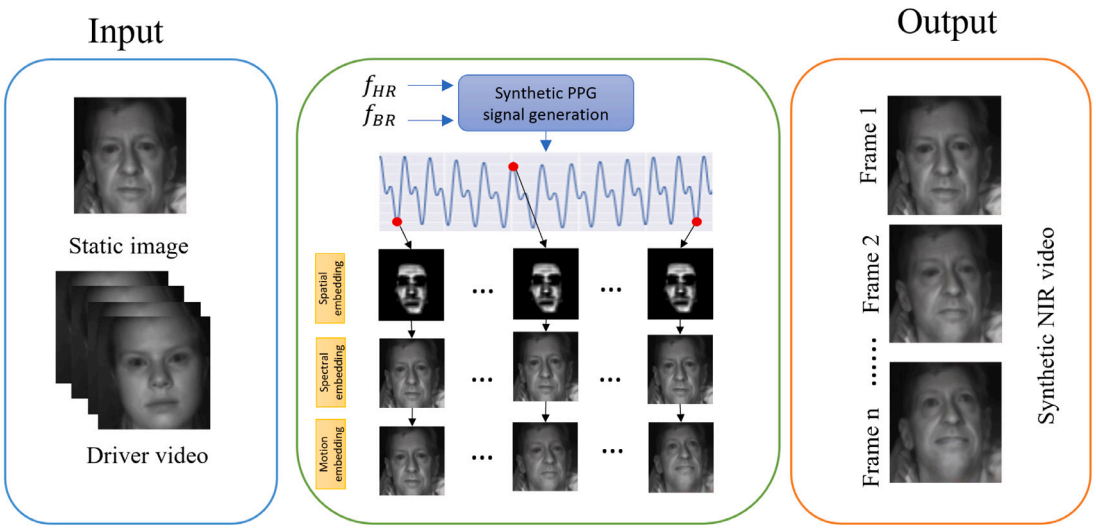


Fig. 1. Generating synthetic videos process.

- This methodology has been validated for augmenting small databases for two distinct applications. Therefore, we propose an evaluation of the use of synthetic videos in training rPPG models in the context of NIR imaging and fitness activities. These two examples are especially relevant due to the scarcity of available NIR datasets and the repetitive movements of fitness scenarios.
- We also collected and made publicly available a novel collection of near-infrared videos named IMVIA-NIR.

The remainder of the paper is organized as follows: Section 2 elaborates the proposed data augmentation methodology, Section 3 presents the evaluation protocol, Section 4 the results and Section 5 the conclusion.

2. Proposed data augmentation methodology

The proposed methodology is outlined in Fig. 1. Our method leverages an image animation technique to generate synthetic video sequences that animates a person in a source image, harnessing the captured motion from a driving video. Through this process, we can generate many synthetic videos by initially utilizing only a few facial images and motion videos. We integrate a synthetic PPG signal into the augmented videos to ensure the synthetic videos are suitable for training neural networks in heart rate estimation applications. Several important factors are considered during this integration process. Firstly, we account for factors that influence the shape of the PPG signal, including breathing rate and cardiac variability. Additionally, we carefully consider the spatial distribution of the signal on the face, the distribution of the signal across the RGB channels, and the distribution of heart rate values among the synthetic videos.

In the subsequent sections, we provide an in-depth, step-by-step explanation of our data augmentation approach, offering detailed insights into the implementation and integration of each component to generate synthetic videos.

2.1. Synthetic PPG signals generation

It is important to clarify that we use the term "PPG signal" to refer to the analytically created synthetic signal. On the other hand, we use the term "rPPG signal" when referring to the integrated signal within the video sequence depicting a person's face.

The approach employed for generating synthetic PPG signals is inspired by the work of Perepelkina et al. [44], which utilized sinusoidal signals to generate synthetic PPG signals. In addition to this, our technique models various components of the PPG signal [45], including the distinctive shape of a PPG signal with the presence of the dicrotic notch, respiratory rate with its associated baseline wander and amplitude modulation, Gaussian noise to simulate instrument measurement noise and frequency modulation reflecting cardiac variability. By incorporating these factors into the generation of synthetic PPG signals, our technique ensures a realistic representation of the PPG signal's characteristics. This enhances the suitability of the synthetic data for training and evaluating neural networks in the domain of heart rate estimation. We define the synthetic PPG signal $s(t)$ as:

$$s(t) = p(t) + d(t) + C_1 b(t) + n(t) \tag{1}$$

where $s(t)$ is composed of four main components: the pulse signal and its dicrotic notch $p(t)$ and $d(t)$, the breathing rate $b(t)$ and the Gaussian noise $n(t)$. $p(t)$ and $d(t)$ are defined in Eqs. (2) and (3), respectively:

$$p(t) = \left(A_1 + C_2 b(t)\right) \cdot \sin\left(2\pi \cdot \left(f_{hr}(t) + C_3 b(t)\right) \cdot t + \phi_p\right) \tag{2}$$

$$d(t) = \left(A_2 + C_2 b(t)\right) \cdot \sin\left(4\pi \cdot \left(f_{hr}(t) + C_3 b(t)\right) \cdot t + 2\phi_p\right) \tag{3}$$

where A_1 and A_2 are the pulse signal and its dicrotic notch amplitudes, respectively. $f_{hr}(t)$ is the instantaneous heart rate at time t, and ϕ_p is the pulse signal phase at origin. $f_{hr}(t)$ is sampled from a uniform distribution $f_{hr} \pm f_{hr}\delta_{hr}$, where f_{hr} is a reference heart rate for the signal and δ_{hr} refers heart rate variability (we use $\delta_{hr} = 0.05$). Similarly, the breathing baseline wander is given by:

$$b(t) = A_3 \cdot \sin\left(2\pi \cdot f_{br}(t) \cdot t + \phi_b\right) \tag{4}$$

with A_3 the baseline wander amplitude, and ϕ_b the breathing signal phase at origin. C_1, C_2, and C_3 are constants and were found empirically, being $C_1 = 0.05$, $C_2 = 0.01$, and $C_3 = 0.15$. A_1, A_2 and A_3 are randomly sampled from $[0.2, 0.7]$, $[0, 0.3]$ and $[0.3, 2]$ respectively. The breathing rate $f_{br}(t)$ at time t is sampled from $f_{br} \pm f_{br}\delta_{br}$ with $\delta_{br} = 0.1$. The reference breathing rate of the signal f_{br} is set between 0.2 and 0.4 Hz. The frequency range for the reference heart rate f_{hr} is set between 0.7 and 3 Hz; however, we do not sample reference heart rates from a uniform distribution but a non-parametric distribution that allows us to best match the data distribution of a set of heart rates estimated on the dataset we want to augment. Indeed, as observed in [46], the heart rate distribution in a training database has a major impact on the performance of a neural network architecture for an rPPG application.

In order to make as few assumptions as possible about this distribution, we use a non-parametric estimate based on kernel smoothing,

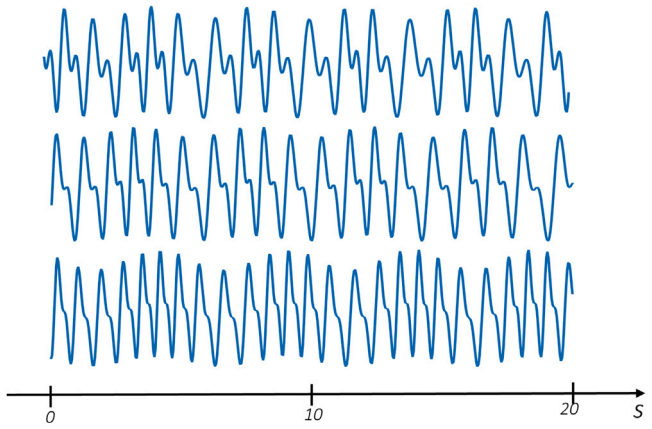


Fig. 2. Examples of three synthetic PPG signals.

or kernel density estimation (KDE). Let f_{hr}^{i} be the average heart rate estimated from the ith video of a training dataset, where $i = 1, 2, \ldots, n$ and n the number of videos. The KDE model estimates the underlying probability density function of the heart rates with the following:

$$g(f_{hr}) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{f_{hr} - f_{hr}^{i}}{h}\right). \tag{5}$$

In this expression, $g(f_{hr})$ represents the KDE distribution at point f_{hr}. The kernel Gaussian function $K(\cdot)$ is applied to each individual sample $(f_{hr} - f_{hr}^{i})$, scaled by the bandwidth h. The bandwidth h controls the smoothness of the estimated density. By sampling the instantaneous heart rates from this KDE distribution, we ensure that the distribution of heart rates in the synthetic video matches the existing dataset. Fig. 2 depicts some examples of the final synthetic PPG signals.

2.2. Synthetic videos generation

Spatial embedding. In order to integrate the PPG signal in the facial region of a person in a synthetic video, it is essential to consider that the signal is not uniformly distributed over the skin. Some areas of the skin are more vascularized than others. Notably, studies have demonstrated that the amplitude of the signal is higher in regions such as the cheeks and forehead [47]. Therefore, we employ the following procedure to determine the spatial distribution of the signal on the face. Let I be an image of a subject. We use a deep learning-based model specialized in skin detection, denoted as Ψ_s, to acquire a binary mask of skin pixels $M_s = \Psi_s(I, \theta_s)$ with θ_s the model weights. In this work, the BlazeFace [48] approach is used to detect and extract faces and a semantic segmentation network [49] is used for skin detection in both NIR and RGB videos. The edges of the binary mask M_s are smoothed using a median filter. In order to consider the distribution of the amplitude of the signal on various areas of the skin, we use a neural network specialized in generating an attention mask on the skin (such as the *appearance* branch of DeepPhys [19]), namely Ψ_a, having $M_a = \Psi_a(I, \theta_a)$ with θ_a the weights of the network and M_a the attention mask. Finally, the spatial distribution of the rPPG signal M is obtained with the element-wise multiplication between the attention mask and the binary skin mask: $M = M_a \odot M_s$.

Channel embedding. At the basis of most rPPG methods, it is well known that the proportion of the rPPG signal is different in the 3 RGB channels. The relative amplitude of the rPPG signal in the light reflected from the skin varies as a function of wavelength. The signal is strongest in the green channel due to the absorption peak of hemoglobin around these wavelengths. Fig. 3 shows the relative amplitude of the PPG signal (AC/DC) in light reflected from the skin. The amplitude varies as a function of wavelength, peaking at around 550 nm. Consequently, we use a weight vector w to distribute the rPPG signal across the 3 RGB channels [50]:

$$[w_R, w_G, w_B] = [0.23, 0.41, 0.36]. \tag{6}$$

It is important to note that, in this work, we consider that NIR videos are monochrome, which is often the case. Consequently, the question of channel distribution does not arise in that case.

Video generation and motion embedding. Using an image of a face I, along with a synthetic PPG signal $s(t)$ and its corresponding spatial distribution M and spectral distribution w, we employ an image animation technique to generate a synthetic video sequence. This process involves integrating the motion captured from a driving video into the synthetic sequences, resulting in realistic and dynamic animations. We begin by duplicating T times the frame I, generating the T-length video F. Then, the PPG signal is embedded into F with:

$$F'_c(i, j, t) = F_c(i, j, t) + \gamma w_c M(i, j) s(t) \tag{7}$$

with $c \in \{R, G, B\}$ the red, green and blue channel, γ a parameter to adjust the signal amplitude and F' is the video with the embedded rPPG component. In the case of monochrome video, the rPPG signal embedding is given by $F'(i, j, t) = F(i, j, t) + \gamma M(i, j) s(t)$ omitting w. i, j and t are the location and temporal indices. This integration is based on Shafer's dichromatic reflection model [16], which provides a framework for modeling both color changes and motions used in multiple previous works (*e.g.* [15] or [17]). The dichromatic reflection model explains skin color in images as a sum of two components: a static term representing the constant skin tone and a dynamic, pulsatile term that varies with blood flow. Now that the PPG signals are embedded in F', we can refer to this signal as rPPG. Although F' contains an rPPG signal, the subject within the scene remains static. Therefore, it is necessary to add motion to the video frame to make it more realistic, using an image animation technique. In this work, we employ a neural network specialized in transmitting motion (Ψ_m), based on the first-order motion model proposed by Siarohin et al.et al. [52]. We eventually obtain $F'' = \Psi_m(F', \theta_m)$ where θ_m are the network weights, and F'' is the final synthetic video that embeds an rPPG component and realistic motion. These videos can be used in the neural network training procedure to estimate an rPPG signal. Examples of synthetic videos are given in Section 4.

3. Evaluation protocol

3.1. Datasets

Our particular focus lies on scenarios involving NIR imaging and fitness activities because these two distinct applications are especially relevant due to the scarcity of available NIR datasets and the repetitive movements of fitness scenarios. As a consequence, we have selected the following three public databases and we also collected and made publicly available a novel collection of near-infrared videos, named IMVIA-NIR.[1]

MERL-RICE [36] indoor dataset is the first dataset of face videos for remote photoplethysmography (rPPG) that were collected simultaneously in broadband RGB and narrow-band NIR, with pulse oximeter recordings as ground truth of the vital signs. The dataset contains eight subjects with *still* and *motion* experiments. We only use the *still* experiment in this work. The dataset totally contains 15 videos, and the length of each video is 3 min. TokyoTech-NIR [48] contains NIR facial videos of nine subjects and corresponding reference finger-attached PPG sensor data. The nine 3-minute long videos are split into nine 20-second segments. Each 3-minute video contains three parts (about

[1] IMVIA-NIR dataset webpage: https://sites.google.com/view/ybenezeth/imvia-nir.

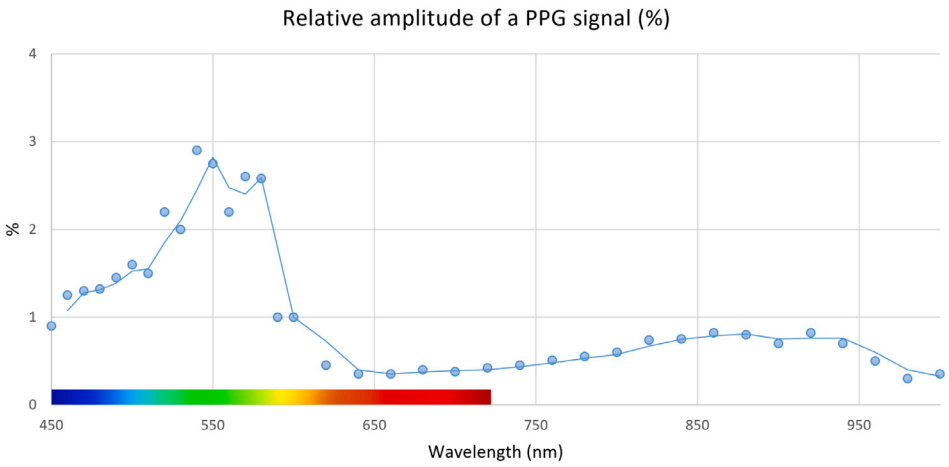


Fig. 3. Relative amplitude of an rPPG signal with respect to the wavelengths. Plot generated from data in [51].

60 s each): *relax*, *exercise*, and *relax* again. In the exercise session, the subjects were asked to perform hand-grip exercises. ECG-Fitness [22] database is challenging due to its versatile nature and realistic lighting conditions. It consists of RGB videos of subjects performing physical activities on fitness equipments. There are a total of 17 subjects. The subjects perform activities such as rowing, speaking, running in an elliptical trainer, and riding a stationary bike. There are a total of 204 videos lasting one minute. The dataset addresses challenges such as non-uniform lighting, rapid motions with blur, strong facial expressions etc.

The applications of rPPG in NIR vision are very numerous, but unfortunately, the existing public databases are very few. Therefore, we decided to collect a new database of NIR videos and make this database publicly accessible to researchers working on the subject. IMVIA-NIR dataset contains a set of 45-second-long NIR videos acquired from 10 subjects in an indoor environment. The dataset has both male and female subjects. We have also chosen subjects from varying ethnicities to improve the generality of the database. A few sample frames from the NIR datasets and ECG-Fitness databases used in our experiments are shown in Figs. 4 and 5, respectively.

3.2. Metrics

In this work, we use a set of four metrics to evaluate our data augmentation strategy, with two metrics evaluating the heart rate estimation, namely Mean Absolute Error (MAE) and Pearson's correlation coefficient (r), and two metrics evaluating the quality of the estimated rPPG signals namely, Template Match Correlation (TMC) and Signal-to-Noise-Ratio (SNR). Where the value of MAE should be minimized, and the values of SNR, r, and TMC should be maximized. The analysis was carried over sequential temporal windows of 15 s and a step size of 0.5 s.

Mean Absolute Error. MAE corresponds to the average absolute error between $\mathbf{HR_r}$ and $\mathbf{HR_c}$ in *bpm* calculated over all the windows for all videos.

$$MAE = \frac{1}{n}\sum_{n}^{k=1} \left|\mathbf{HR_r}(k) - \mathbf{HR_c}(k)\right| \tag{8}$$

where $\mathbf{HR_r}(k)$ and $\mathbf{HR_c}(k)$ denote the heart rates estimated from the kth window of the rPPG and the contact-based or synthetic PPG signals, respectively. These window-wise heart rates are derived from the highest peak in the FFT of both the rPPG and contact-based or synthetic PPG signals, within the standard heart rate frequency range of $f \in [0.7, 3]$ Hz.

Pearson correlation coefficient. r coefficient measures the linear correlation between vectors $\mathbf{HR_r}$ and $\mathbf{HR_c}$. A value of $r = 1$ means a positive total linear correlation, while $r = -1$ implies a negative linear correlation. Finally, $r = 0$ indicates that there is no linear correlation between the estimations and the reference values. r is given by:

$$r = \frac{\sum_{k=1}^{n}\left(\mathbf{HR_r}(k) - \overline{HR_r}\right)\left(\mathbf{HR_c}(k) - \overline{HR_c}\right)}{\sqrt{\sum_{k=1}^{n}\left(\mathbf{HR_r}(k) - \overline{HR_r}\right)^2}\sqrt{\sum_{k=1}^{n}\left(\mathbf{HR_c}(k) - \overline{HR_c}\right)^2}} \tag{9}$$

where $\overline{HR_r}$ and $\overline{HR_c}$ represents the average of $\mathbf{HR_r}$ and $\mathbf{HR_c}$ respectively.

Signal-to-noise ratio. SNR measures the quality of rPPG signals as the ratio of the power of the main pulsatile component and the power of background noise, computed in dB due to the wide dynamic range of the signals:

$$SNR = 10\log_{10}\left(\frac{\int_{f_1}^{f_2} h_{signal}(f)|\mathcal{F}\{\tilde{s}(t)\}|^2 \mathrm{d}f}{\int_{f_1}^{f_2} h_{noise}(f)|\mathcal{F}\{\tilde{s}(t)\}|^2 \mathrm{d}f}\right) \tag{10}$$

where $\mathcal{F}\{\tilde{s}(t)\}$ is the FFT transform of the estimated rPPG signal $\tilde{s}(t)$, f_1 and f_2 the lower and upper limit of the integral defined by the possible physiological range of the heart rate (40 to 240 bpm in our case), and a double-step function h for the first and second harmonics, defined by the convolutions:

$$\begin{aligned} h_{signal}(f) &= [\delta(f - f_0) + \delta(f - 2f_0)] * \prod(\pm f_r) \\ h_{noise}(f) &= 1 - h_{signal}(f) \end{aligned} \tag{11}$$

with δ the Dirac delta function, f_0 the fundamental frequency (*i.e.* peak of the periodogram), convoluted with the *rect* function, noted as $\prod$ of half-width f_r.

Template match correlation. TMC is another signal quality metric in the temporal domain [53]. The following steps are performed:

1. The signal peaks are detected.
2. The median beat-to-beat interval is calculated.
3. With a window width equal to the median beat-to-beat interval, all pulses are extracted individually centered on their respective peak.
4. The template is calculated as the average of all the pulses.
5. The TMC coefficient is calculated as the average of the correlation of all the pulses with the template.

A value close to $TMC = 1$ means that the pulse shape of the evaluated signal is uniform, and, therefore close to the expected signal, while a value close to $TMC = 0$ indicates the contrary.

Fig. 4. Sample frames from NIR databases - Row 1: TokyoTech, Row 2:MERL-RICE, Row 3:IMVIA-NIR.

Fig. 5. Sample frames from ECG-Fitness database. Row 1: Rowing exercise, Row 2: Running on an Elliptic trainer, Row 3: Stationary bike.

3.3. Models and implementation details

We evaluate our proposed data augmentation strategy with two different 3D-CNN architectures, namely Physnet [20] and the real-time rPPG model named RTrPPG [33] networks. These networks consist of an encoder E and decoder D. The encoder E is used to transform the given input into a compressed form. The latent representation of the input is then fed as an input to the decoder D to generate the rPPG signal $\tilde{s} = [\tilde{s}(1), \tilde{s}(2), \ldots \tilde{s}(T)]$. The rPPG estimation is given by the equation:

$$[\tilde{s}(1), \tilde{s}(2), \ldots \tilde{s}(T)] = \Psi_{rPPG}([F(1), F(2) \ldots F(T)]); \theta_{rPPG} \quad (12)$$

where $F(1), F(2) \dots F(T)$ represents the input frames given to network and θ_{rPPG} represents the parameters of the network. We use all the default parameters suggested in the original papers of Physnet [20] and RTrPPG [33]. The BlazeFace approach [54] is specifically designed for the rapid and efficient detection of faces, ideal for real-time image processing applications. This model operates using a streamlined, lightweight convolutional network architecture that ensures both high speed and accuracy. In addition to face detection, we utilize a color-agnostic semantic segmentation network [49] for skin detection across NIR and RGB videos. For more precise physiological signal embedding, we apply the appearance branch of the DeepPhys model [19], which generates soft-attention probability maps. These maps focus on areas of the skin showing stronger physiological signals. Furthermore, image animation in our work is driven by the first order motion model proposed by Siarohin et al. [52]. This model separates appearance from motion using a self-supervised learning framework, resulting in realistic animations through learned keypoints and local affine transformations that respond to the dynamics of a driving video. It also features an occlusion-aware generator, which maintains the visibility and continuity of animated figures.

Once the rPPG signal is estimated, the window-wise heart rate was calculated from the highest peak of the FFT within the limits of normal heart rate $f \in [0.7, 3]$ Hz over a temporal moving window of 15 s using a step size of 0.5 s. Subject-independent 5-fold cross-validation is performed for all the databases used for evaluation. A learning rate of 0.0001 is chosen based on fold-wise hyper-parameter tuning. The batch size for training and evaluation is set as 8, and all the folds are trained for 20 epochs. Synthetic videos are used to pre-train the rPPG models, *i.e.* Physnet and RTrPPG, and models are then fine-tuned on the target dataset. For this, a set of 1000 synthetic videos were created to augment the NIR datasets and a total of 2000 synthetic videos were created to augment the ECG-Fitness dataset. Data augmentation process is applied only to the videos of the training sets.

4. Result analysis

In this section, we present the evaluation of our data augmentation method for training deep learning models in rPPG applications. First, we present in Fig. 6 examples of synthetic videos generated by the proposed method. The first row presents frames extracted from a driving video, whereas the following rows (2 to 4) showcase the synthetic NIR videos. In each of rows 2 to 4, the first image presents the original NIR image that underwent the animation process. It is possible to observe that the motion of the driving video is correctly applied to NIR videos, resulting in realistic movements in our synthetic videos. However, complex movements can cause unrealistic deformations, especially during large head rotations (e.g. in penultimate column) and mouth movements like smiling (e.g. in final column). Since photorealism is not our goal, we can argue that these artifacts introduce diversity among synthetic videos generated from the same driving video.

Quantitative results of heart rate estimations are displayed in Tables 1 and 2, illustrating the performance achieved with and without data augmentation on both the NIR datasets and the ECG-Fitness dataset. We compare the results obtained with Physnet and RTrPPG. Additionally, we compare the reference heart rate sampling from the conventional uniform distribution to the proposed non-parametric distribution implemented using KDE.

Performance over NIR datasets. It is possible to draw some conclusions from Table 1. First and foremost, we observed notable performance variations across different datasets, models and data augmentation strategies. While the MERL and TokyoTech-NIR datasets initially showed acceptable results, the performance on the IMVIA-NIR dataset was notably poor, likely due to its relatively small size, highlighting the importance of pre-training. Second, the application of data augmentation consistently improved the performance of the models, with only a few exceptions where the initial results were already excellent, such as with the TokyoTech-NIR dataset and the Physnet model. In these cases, the pre-training had a negligible effect. Nonetheless, for most cases, the performance was consistently enhanced. For instance, the IMVIA-NIR dataset saw a substantial improvement in mean absolute error (MAE) from 25.92 bpm without data augmentation to 11.20 bpm with uniform sampling and further to an impressive 3.25 bpm using the proposed non-parametric distribution with KDE. Our experiments clearly indicated that the KDE sampling method outperformed the uniform distribution sampling. This difference was particularly notable for the IMVIA-NIR dataset as noted previously. While the Physnet model consistently outperformed the RTrPPG model, the overall trends remained similar between the two models with respect to data augmentation. Additionally, the RTrPPG model exhibited a notable advantage in inference speed compared to Physnet, showing an approximately 90% improvement in inference speed according to Botina et al. [33] with, as can be seen, very similar performance in most cases.

In summary, our findings highlight the significance of data augmentation for improving performance, particularly in cases where the initial results are poor. Moreover, the superiority of the KDE sampling approach over uniform distribution sampling was evident, especially on the IMVIA-NIR dataset. Although the RTrPPG model consistently trailed behind Physnet, its advantage in terms of computational efficiency makes it a favorable choice in certain scenarios, especially when coupled with the pre-training strategy.

To give a better idea of the performance improvement with and without data augmentation, we present in Figs. 7, 8, and 9, the correlation plots obtained with RTrPPG, with and without data augmentation (with KDE sampling) on the 3 NIR datasets. It is possible to observe that the fitting line is consistently closer to the 45° line using the proposed pre-training technique.

Performance over fitness database. ECG Fitness is a particularly difficult dataset, so it is interesting to observe in Table 2 that the data augmentation strategy proposed in this paper increases performance regarding heart rate estimation and signal quality. More specifically, we can observe that the conclusions are the same as for NIR. However, here the ECG Fitness database is quite large, so the improvement is less marked than with IMVIA-NIR, for example. With RTrPPG, signal quality metrics were improved with an increase from 0.56 to 0.76 for TMC, and HR estimation metrics were also improved with a significant decrease in MAE from 25.89 to 9.32 bpm. The trends are similar for PhysNet.

5. Conclusion

Deep learning approaches have significantly enhanced the performance of existing rPPG techniques, largely attributed to the availability of extensive public datasets. Consequently, it becomes crucial to explore methods for augmenting datasets rapidly and efficiently, tailored to specific applications, in order to improve further the accuracy of deep neural networks in rPPG. This research paper introduces a novel methodology that generates synthetic videos to train deep neural networks accurately in estimating heart rates from videos captured under challenging conditions. Our method involves generating video sequences that animate a person in a source image by leveraging motion information extracted from a driving video. We integrate the synthetic rPPG signal into the facial regions, considering key aspects such as the temporal shape of the signal, its spatial and spectral distribution, and the distribution of heart rates. This methodology has been successfully evaluated to expand small databases for two specific applications: training rPPG models to estimate heart rates in NIR imaging and fitness activities. These scenarios are particularly pertinent due to the limited availability of NIR datasets and the repetitive nature of movements in fitness environments. To assess the proposed approach, we utilized two publicly available near-infrared

Fig. 6. Examples of synthetic videos. The top row illustrates frames from one driving video, while the subsequent rows (2 to 4) depict the synthetic videos created by the animation method. The driving video is extracted from the *speaking* scenario of ECG-Fitness, where subjects are speaking in front of the camera, not engaging in physical activities. The first image in each of rows 2–4 shows the original NIR image that was animated.

Table 1
Synthesis of results obtained on the NIR datasets.

Model	Data augmentation	MERL				TokyoTech-NIR				IMVIA-NIR			
		MAE	r	SNR	TMC	MAE	r	SNR	TMC	MAE	r	SNR	TMC
Physnet	No DA	2.00	0.80	5.2	0.91	1.30	0.98	7.2	0.94	25.92	−0.50	−4.6	0.75
	DA w/ uniform	3.32	0.71	3.6	0.87	1.11	0.98	7.3	0.94	11.20	0.19	−1.7	0.80
	DA w/ KDE	1.63	0.88	7.7	0.93	1.09	0.99	9.0	0.93	3.45	0.92	6.5	0.94
RTrPPG	No DA	3.05	0.78	3.3	0.83	4.23	0.79	3.0	0.80	44.60	0.18	−10.9	0.52
	DA w/ uniform	2.23	0.90	3.5	0.84	4.26	0.79	3.3	0.81	14.97	0.52	−5.1	0.64
	DA w/ KDE	1.84	0.92	5.5	0.87	4.08	0.78	4.9	0.84	3.25	0.92	4.7	0.85

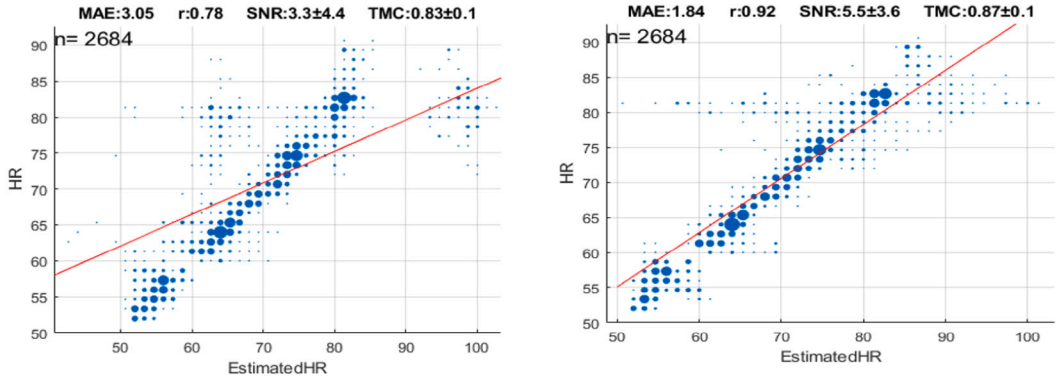


Fig. 7. Correlation plots comparing results obtained without data augmentation (left) vs. using proposed data augmentation technique (right) on MERL-RICE dataset. (A few outliers are ignored on the left plot to maintain the homogeneity of the X-axis in both plots).

datasets, namely MERL-RICE [36] and Tokyotech-NIR [48], along with a challenging RGB fitness dataset called ECG-Fitness [22]. Additionally, we collected and made publicly available a novel collection of near-infrared videos named IMVIA-NIR. The experimental results demonstrate significant improvements, particularly when the initial datasets are relatively small. These findings suggest that our approach holds great promise in enhancing the performance of deep neural networks in rPPG under challenging conditions.

Looking ahead, it would be interesting to conduct a more systematic and in-depth study of the various factors that contribute to the system's final performance. Furthermore, exploring the benefits of this strategy in addressing the limitations of rPPG technology for individuals with dark skin tones would be a valuable direction for future research. Ultimately, we aimed for realism in generating synthetic videos by using an analytical formula to create PPG signals, which incorporates multiple factors such as respiration, heart rate variability, and also considering the spatial and spectral distribution of signals on the skin. The primary goal of these synthetic data is to aid in training deep learning models by providing data tailored to specific applications. However, it would be interesting to investigate in future work to what extent the realism

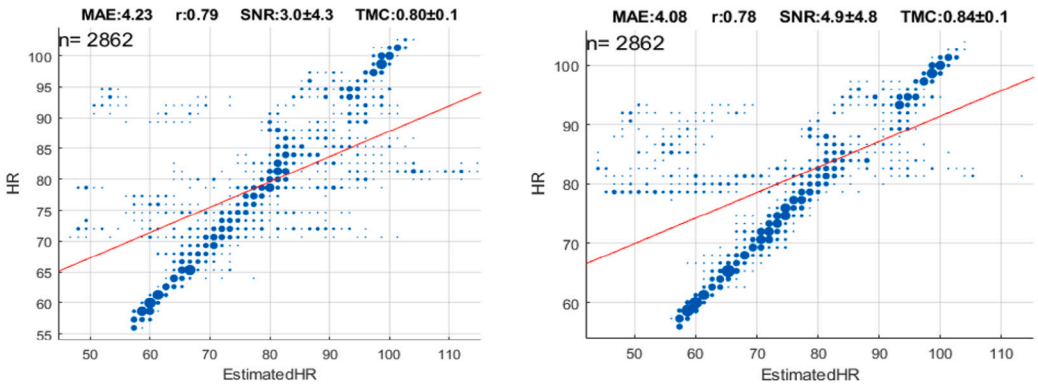


Fig. 8. Correlation plots comparing results obtained without data augmentation (left) vs. using proposed data augmentation technique (right) on TokyoTech-NIR dataset. (A few outliers are ignored on the left plot to maintain the homogeneity of the X-axis in both plots).

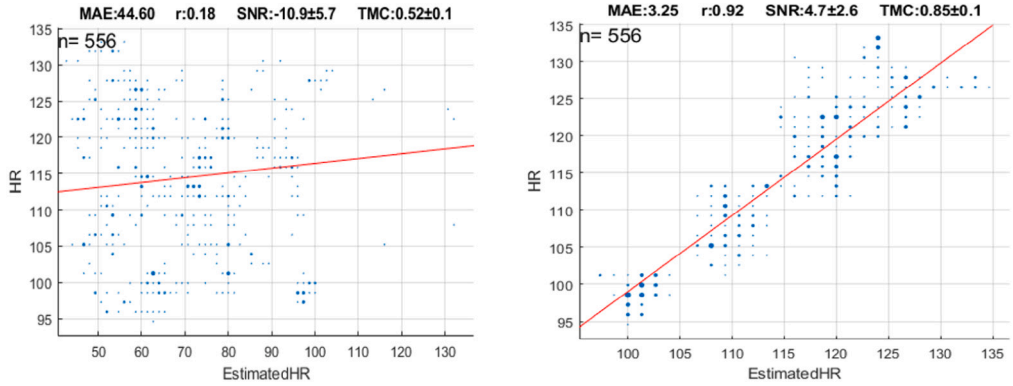


Fig. 9. Correlation plots comparing results obtained without data augmentation (left) vs. using the proposed data augmentation technique (right) on IMVIA-NIR dataset. (A few outliers are ignored on the left plot to maintain the homogeneity of the X-axis in both plots).

Table 2
Synthesis of results obtained on the RGB ECG-Fitness dataset.

Model	Data augmentation	ECG-Fitness			
		MAE	r	SNR	TMC
Physnet	No DA	18.08	0.32	−1.0	0.75
	DA w/ uniform	9.50	0.64	−0.0	0.73
	DA w/ KDE	9.08	0.68	2.0	0.50
RTRPPG	No DA	25.89	0.05	−6.0	0.56
	DA w/ uniform	9.93	0.71	0.0	0.74
	DA w/ KDE	9.32	0.72	0.1	0.76

of these data influences the performance improvements of the deep learning models. This exploration could provide valuable insights into the practical applications and efficacy of synthetic training datasets in machine learning contexts.

CRediT authorship contribution statement

Yannick Benezeth: Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Methodology, Funding acquisition, Conceptualization. **Deepak Krishnamoorthy:** Software. **Deivid Johan Botina Monsalve:** Software. **Keisuke Nakamura:** Supervision, Funding acquisition. **Randy Gomez:** Supervision, Funding acquisition. **Johel Mitéran:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Lalit Maurya, Pavleen Kaur, Deepak Chawla, Prasant Mahapatra, Non-contact breathing rate monitoring in newborns: A review, Comput. Biol. Med. 132 (2021) 104321.
[2] Chi Pham, Khashayar Poorzargar, Mahesh Nagappa, Aparna Saripella, Matteo Parotto, Marina Englesakis, Kang Lee, Frances Chung, Effectiveness of consumer-grade contactless vital signs monitors: a systematic review and meta-analysis, J. Clin. Monit. Comput. (2022) 1–14.
[3] William Taylor, Qammer H Abbasi, Kia Dashtipour, Shuja Ansari, Syed Aziz Shah, Arslan Khalid, Muhammad Ali Imran, A review of the state of the art in non-contact sensing for COVID-19, Sensors 20 (19) (2020) 5665.
[4] Smera Premkumar, Duraisamy Jude Hemanth, Intelligent remote photoplethysmography-based methods for heart rate estimation from face videos: A survey, in: Informatics, Vol. 9, MDPI, 2022, p. 57.
[5] Fatema-Tuz-Zohra Khanam, Asanka G Perera, Ali Al-Naji, Kim Gibson, Javaan Chahl, Non-contact automatic vital signs monitoring of infants in a neonatal intensive care unit based on neural networks, J. Imag. 7 (8) (2021) 122.
[6] Won Kyu Lee, Heenam Yoon, Chungmin Han, Kwang Min Joo, Kwang Suk Park, Physiological signal monitoring bed for infants based on load-cell sensors, Sensors 16 (3) (2016) 409.
[7] Justin Amadeus Albert, Bert Arnrich, A computer vision approach to continuously monitor fatigue during resistance training, Biomed. Signal Process. Control 89 (2024) 105701.

[8] Elena Magán, M Paz Sesmero, Juan Manuel Alonso-Weber, Araceli Sanchis, Driver drowsiness detection by applying deep learning techniques to sequences of images, Appl. Sci. 12 (3) (2022) 1145.

[9] Wim Verkruysse, Lars O. Svaasand, J. Stuart Nelson, Remote plethysmographic imaging using ambient light, Opt. Express 16 (26) (2008) 21434–21445.

[10] Magdalena Lewandowska, Jacek Ruminski, Tomasz Kocejko, Jedrzej Nowak, Measuring pulse rate with a webcam — A non-contact method for evaluating cardiac activity, in: 2011 Federated Conference on Computer Science and Information Systems, FedCSIS, 2011, pp. 405–410.

[11] Ming-Zher Poh, Daniel J. McDuff, Rosalind W. Picard, Advancements in noncontact, multiparameter physiological measurements using a webcam, IEEE Trans. Biomed. Eng. 58 (1) (2011) 7–11.

[12] Richard Macwan, Yannick Benezeth, Alamin Mansouri, Keisuke Nakamura, Randy Gomez, Remote photoplethysmography measurement using constrained ica, in: 2017 E-Health and Bioengineering Conference, EHB, IEEE, 2017, pp. 430–433.

[13] Richard Macwan, Yannick Benezeth, Alamin Mansouri, Heart rate estimation using remote photoplethysmography with multi-objective optimization, Biomed. Signal Process. Control 49 (2019) 24–33.

[14] Richard Macwan, Serge Bobbia, Yannick Benezeth, Julien Dubois, Alamin Mansouri, Periodic variance maximization using generalized eigenvalue decomposition applied to remote photoplethysmography estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 1332–1340.

[15] Gerard De Haan, Arno Van Leest, Improved motion robustness of remote-PPG by using the blood volume pulse signature, Physiol. Meas. 35 (9) (2014) 1913.

[16] Wenjin Wang, Albertus C Den Brinker, Sander Stuijk, Gerard De Haan, Algorithmic principles of remote PPG, IEEE Trans. Biomed. Eng. 64 (7) (2016) 1479–1491.

[17] Gerard de Haan, Vincent Jeanne, Robust pulse rate from chrominance-based rPPG, IEEE Trans. Biomed. Eng. 60 (10) (2013) 2878–2886.

[18] Qi Zhan, Wenjin Wang, Gerard de Haan, Analysis of CNN-based remote-PPG to understand limitations and sensitivities, 2020.

[19] Weixuan Chen, Daniel McDuff, Deepphys: Video-based physiological measurement using convolutional attention networks, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 349–365.

[20] Zitong Yu, Xiaobai Li, Guoying Zhao, Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks, 2019, arXiv preprint arXiv:1905.02419.

[21] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, Guoying Zhao, 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, 2019, pp. 151–160.

[22] Radim Špetlík, Vojtech Franc, Jirí Matas, Visual heart rate estimation with convolutional neural network, in: Proceedings of the British Machine Vision Conference, Newcastle, UK, 2018, pp. 3–6.

[23] Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, Guoying Zhao, Video-based remote physiological measurement via cross-verified feature disentangling, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, Springer, 2020, pp. 295–310.

[24] Xuesong Niu, S. Shan, Hu Han, Xilin Chen, RhythmNet: End-to-end heart rate estimation from face via spatial-temporal representation, IEEE Trans. Image Process. 29 (2019) 2409–2423.

[25] Xuesong Niu, Hu Han, Shiguang Shan, Xilin Chen, VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video, in: Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part V 14, Springer, 2019, pp. 562–576.

[26] Xuesong Niu, Hu Han, Shiguang Shan, Xilin Chen, Synrhythm: Learning a deep heart rate estimator from general to specific, in: 2018 24th International Conference on Pattern Recognition, ICPR, IEEE, 2018, pp. 3580–3585.

[27] Mohammad Sabokrou, Masoud Pourreza, Xiaobai Li, Mahmood Fathy, Guoying Zhao, Deep-HR: Fast heart rate estimation from face video under realistic conditions, Expert Syst. Appl. 186 (2021) 115596.

[28] Xuesong Niu, Xingyuan Zhao, Hu Han, Abhijit Das, Antitza Dantcheva, Shiguang Shan, Xilin Chen, Robust remote heart rate estimation from face utilizing spatial-temporal attention, in: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), IEEE, 2019, pp. 1–8.

[29] Rencheng Song, Senle Zhang, Chang Li, Yunfei Zhang, Juan Cheng, Xun Chen, Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks, IEEE Trans. Instrum. Meas. 69 (10) (2020) 7411–7421.

[30] J. Gideon, S. Stent, The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, IEEE Computer Society, Los Alamitos, CA, USA, 2021, pp. 3975–3984.

[31] Jeremy Speth, Nathan Vance, Patrick Flynn, Kevin Bowyer, Adam Czajka, Unifying frame rate and temporal dilations for improved remote pulse detection, Comput. Vis. Image Underst. 210 (2021) 103246.

[32] Zitong Yu, Xiaobai Li, Xuesong Niu, Jingang Shi, Guoying Zhao, Autohr: A strong end-to-end baseline for remote heart rate measurement with neural searching, IEEE Signal Process. Lett. 27 (2020) 1245–1249.

[33] Deivid Botina-Monsalve, Yannick Benezeth, Johel Miteran, RTrPPG: An ultra light 3DCNN for real-time remote photoplethysmography, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2146–2154.

[34] Hanguang Xiao, Tianqi Liu, Yisha Sun, Yulin Li, Shiyi Zhao, Alberto Avolio, Remote photoplethysmography for heart rate measurement: A review, Biomed. Signal Process. Control 88 (2024) 105608.

[35] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, Julien Dubois, Unsupervised skin tissue segmentation for remote photoplethysmography, Pattern Recognit. Lett. 124 (2019) 82–90.

[36] E.M. Nowara, T.K. Marks, H. Mansour, A. Veeraraghavan, Near-infrared imaging photoplethysmography during driving, IEEE Trans. Intell. Transp. Syst. (2020) 1–12.

[37] Alex Hernández-García, Peter König, Further advantages of data augmentation on convolutional neural networks, in: Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part I 27, Springer, 2018, pp. 95–103.

[38] Rencheng Song, Huan Chen, Juan Cheng, Chang Li, Yu Liu, Xun Chen, PulseGAN: Learning to generate realistic pulse waveforms in remote photoplethysmography, IEEE J. Biomed. Health Inf. 25 (5) (2021) 1373–1384.

[39] Daniel McDuff, Xin Liu, Javier Hernandez, Erroll Wood, Tadas Baltrusaitis, Synthetic data for multi-parameter camera-based physiological sensing, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 3742–3748.

[40] Daniel McDuff, Camera measurement of physiological vital signs, ACM Comput. Surv. 55 (9) (2023) 1–40.

[41] Zijie Yue, Miaojing Shi, Shuai Ding, Video-based remote physiological measurement via self-supervised learning, 2022, arXiv preprint arXiv:2210.15401.

[42] Yun-Yun Tsou, Yi-An Lee, Chiou-Ting Hsu, Multi-task learning for simultaneous video generation and remote photoplethysmography estimation, in: Proceedings of the Asian Conference on Computer Vision, 2020.

[43] Lokendra Birla, Sneha Shukla, Anup Kumar Gupta, Puneet Gupta, ALPINE: Improving remote heart rate estimation using contrastive learning, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 5029–5038.

[44] Olga Perepelkina, Mikhail Artemyev, Marina Churikova, Mikhail Grinenko, HeartTrack: Convolutional neural network for remote video-based heart rate monitoring, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 288–289.

[45] John Allen, Photoplethysmography and its application in clinical physiological measurement, Physiol. Meas. 28 (3) (2007) R1.

[46] Deivid Botina-Monsalve, Yannick Benezeth, Johel Miteran, Performance analysis of remote photoplethysmography deep filtering using long short-term memory neural network, BioMed. Eng. OnLine 21 (1) (2022) 1–27.

[47] Peixi Li, Yannick Benezeth, Keisuke Nakamura, Randy Gomez, Fan Yang, Model-based region of interest segmentation for remote photoplethysmography, in: VISIGRAPP (4: VISAPP), 2019, pp. 383–388.

[48] Yuichiro Maki, Yusuke Monno, Kazunori Yoshizaki, Masayuki Tanaka, Masatoshi Okutomi, Inter-beat interval estimation from facial video based on reliability of BVP signals, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2019, pp. 6525–6528.

[49] https://github.com/WillBrennan/SemanticSegmentation.

[50] Xueming Lin, Gerard de Haan, Using blood volume pulse vector to extract rppg signal in infrared spectrum, 2014.

[51] Damianos Damianou, The Wavelength Dependence of the Photoplethysmogram and Its Implication to Pulse Oximetry (Ph.D. thesis), University of Nottingham, 1995.

[52] Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, Nicu Sebe, First order motion model for image animation, in: Conference on Neural Information Processing Systems (NeurIPS), 2019.

[53] Christina Orphanidou, Christina Orphanidou, Quality assessment for the photoplethysmogram (PPG), in: Signal Quality Assessment in Physiological Monitoring: State of the Art and Practical Considerations, Springer, 2018, pp. 41–63.

[54] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, Matthias Grundmann, Blazeface: Sub-millisecond neural face detection on mobile gpus, 2019, arXiv preprint arXiv:1907.05047.