PulseGAN: Learning to generate realistic pulse waveforms in remote photoplethysmography

Rencheng Song, *Member, IEEE,* Huan Chen, Juan Cheng, *Member, IEEE,* Chang Li, *Member, IEEE,* Yu Liu, *Member, IEEE,* and Xun Chen, *Senior Member, IEEE*

***Abstract*—Remote photoplethysmography (rPPG) is a non-contact technique for measuring cardiac signals from facial videos. High-quality rPPG pulse signals are urgently demanded in many fields, such as health monitoring and emotion recognition. However, most of the existing rPPG methods can only be used to get average heart rate (HR) values due to the limitation of inaccurate pulse signals. In this paper, a new framework based on generative adversarial network, called PulseGAN, is introduced to generate realistic rPPG pulse signals through denoising the chrominance (CHROM) signals. Considering that the cardiac signal is quasi-periodic and has apparent time-frequency characteristics, the error losses defined in time and spectrum domains are both employed with the adversarial loss to enforce the model generating accurate pulse waveforms as its reference. The proposed framework is tested on three public databases. The results show that the PulseGAN framework can effectively improve the waveform quality, thereby enhancing the accuracy of HR, the interbeat interval (IBI) and the related heart rate variability (HRV) features. The proposed method significantly improves the quality of waveforms compared to the input CHROM signals, with the mean absolute error of AVNN (the average of all normal-to-normal intervals) reduced by 41.19%, 40.45%, 41.63%, and the mean absolute error of SDNN (the standard deviation of all NN intervals) reduced by 37.53%, 44.29%, 58.41%, in the cross-database test on the UBFC-RPPG, PURE, and MAHNOB-HCI databases, respectively. This framework can be easily integrated with other existing rPPG methods to further improve the quality of waveforms, thereby obtaining more reliable IBI features and extending the application scope of rPPG techniques.**

***Index Terms*—Heart rate estimation, remote photoplethysmography, generative adversarial network, pulse waveform, heart rate variability**

I. Introduction

CARDIAC signal is an important physiological signal to monitor the human body's health and emotional status. The common ways for obtaining cardiac signals include electrocardiogram (ECG) and photoplethysmography (PPG). Both of them rely on specific sensors to contact with skins of subjects, which may be uncomfortable or unsuitable for people with sensitive skins [1]. In recent years, there is a trend to develop non-contact heart rate measurements through the microwave Doppler or computer vision techniques. The remote photoplethysmography (rPPG) [2] is a kind of computer vision based technique to record color changes of facial skins caused by corresponding heartbeats using consumer-level cameras.

After years of development, a variety of rPPG methods have been introduced according to different assumptions and mechanisms [3], [4]. For example, blind source separation (BSS) [5] based methods are proposed under some specific statistical assumption, while the model-based rPPG methods [6], [7] are derived from a skin optical reflection model. In addition, there are some other methods to achieve heart rate extraction through signal filtering [8], [9]. These conventional methods usually perform well when their model assumptions are met. However, in the actual environment with diverse types of noise, it is likely that the underlying assumptions of the original method are not fully met, which greatly decreases the performance of the method and leads to low quality of extracted waveforms. Therefore, conventional methods usually only aim to obtain average heart rate (HR) values by calculating the dominate frequency of the rPPG pulses [10], [11].

However, there is a growing demand to calculate more diverse cardiac features in rPPG applications, such as stress detection, emotional classification, and health monitoring, etc, where high-quality waveforms are critical. For example, HRV is the variation of HR cycles. It is a valuable predictor of sudden cardiac death and arrhythmic events. The spectral component of HRV can also reflect the activities of the parasympathetic and sympathetic nervous systems. Currently, these diverse cardiac features can usually be obtained from high-quality pulse waveforms measured by contact ECG or PPG. They usually require electrodes or sensors to be in contact with the human body, and thereby limiting the application scopes. Therefore, it is urgent to develop new rPPG technology which can extract accurate pulse waveform for calculating more physiological characteristics.

On the other hand, inspired by the rapid development of deep learning (DL) techniques, DL-based rPPG algorithms have also been proposed in recent years. The rPPG approaches based on DL can be generally divided into two types, the end-to-end type and the feature-decoder type. The former ones directly establish the mapping from video frames to the target HR values or pulse signals, while the latter ones get the HR targets through decoding the latent information preprocessed from video frames. Since DL is data-driven and

This work was supported in part by the National Natural Science Foundation of China (Grants 61922075 and 41901350), in part by the Provincial Natural Science Foundation of Anhui (Grant 2008085QF285), and in part by the Fundamental Research Funds for the Central Universities (Grant JZ2019HGBZ0151). (Corresponding author: Xun Chen).

R. Song, H. Chen, J. Cheng, C. Li and Y. Liu are with the Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China (e-mail: rcsong@hfut.edu.cn; 2018110057@mail.hfut.edu.cn; chengjuan@hfut.edu.cn; changli@hfut.edu.cn; yuliu@hfut.edu.cn).

Xun Chen is with Epilepsy Center, Department of Neurosurgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, and also with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, Anhui, 230001, China (e-mail: xunchen@ustc.edu.cn).

neural networks have strong fitting capabilities, the results of DL-based rPPG methods often outperform the conventional ones as demonstrated in [12], which inspires us to extract rPPG pulse waveforms under a DL framework.

The extraction of rPPG pulse waveforms can be considered as a generative problem from the perspective of generative models. Since firstly proposed by Ian in 2014, generative adversarial networks (GAN) [13] has become the mainstream generative method due to its state-of-the-art performance, especially in image processing and computer vision areas. The GAN is consisted of two neural networks, the generator G and the discriminator D. The two networks are trained in an adversarial way, where G generates a fake target signal to confuse the discriminator, and D makes judgments on the generated signals from the real ones, thereby prompting the results of G to be closer to the references. With the rapid development of GAN, it has also been applied to denoise one-dimensional signals, such as speech signals [14], [15], and ECG signals [16]. These studies enlighten us to acquire reliable rPPG waveforms using GAN models.

In this paper, we propose a new framework, named as PulseGAN, to extract rPPG pulse signal with a conditional GAN (cGAN) [17]. The rough pulse signal derived from CHROM method [6] is taken as the input of generator G, and the PPG signal synchronously recorded by a pulse oximeter is used as a reference. The discriminator D judges the generated signal from the reference one, where the rough input of G is taken as a conditioning. Considering the apparent characteristics of pulse signal, besides the adversarial loss, we also combine the waveform error loss in the time domain and the spectrum error loss in the frequency domain to enforce a multi-level match between the generated waveform and its reference. Through the adversarial training between G and D, the generator learns to construct a rPPG pulse as close as its ground truth. The proposed method is tested on public databases in two scenarios, including both within- and cross-database cases. The test results reveal that the PulseGAN effectively improves the quality of input waveforms like the signal-to-noise ratio (SNR), so that more cardiac features including the interbeat interval (IBI) indexes, and the HRV can be calculated more reliably.

In summary, the main contribution of this paper is that we introduce a PulseGAN framework to extract realistic rPPG pulse waveforms from rough input signals derived by some conventional method. The high-quality waveform makes it possible to further calculate reliable cardiac features like HRV, which can potentially extend the application scopes of rPPG techniques. The framework effectively combines the benefits of conventional methods and GAN. The generator is enforced to learn features of reference PPG signals through error losses defined in both time and spectrum domains in addition to the adversarial loss. The PulseGAN framework can also be easily integrated with some existing rPPG methods to achieve high-quality waveform reconstruction, thereby extending the application scope of rPPG techniques.

II. Related Work

In 2008, Verkruysse *et al.* [2] first verified the validity of rPPG for HR estimation from facial videos. They demonstrated that the green channel signal extracted from skin pixels contained strong pulsating information. Since then, a variety of rPPG methods have been proposed. Among them, the typical ones include those methods based on blind source separation (BSS) or the skin optical reflection model. The BSS method assumes that the pulse signal is linearly mixed with other noise signals, and all those signals satisfy some statistical property. For example, Poh *et al.* [5] applied independent component analysis (ICA) to separate the pulse signals from the color RGB signals. Wei *et al.* [18] employed the second-order blind source separation to extract the target signal from six RGB channels obtained in two facial regions of interest (ROIs). On the other hand, the methods based on the optical reflection model extract pulse signal explicitly through a combination of individual color channels are combined with specific ratios. This is considered to eliminate the common interference sources from the RGB channels. For example, De Haan *et al.* [6] proposed a chrominance method (CHROM) to calculate the pulse signal. The CHROM method eliminates the specular reflection component with a projection and then obtains the pulse through an "alpha tuning". In [7], Wang *et al.* used a different projection plane orthogonal to skin color (POS) for rPPG signal extraction. These conventional methods have achieved excellent results in calculating the average HR values of rPPG, during both laboratory and realistic scenarios. However, the quality of the waveforms remains poor due to noise interference and model limitations, which still has large room for improvement.

In the last few years, DL techniques have been increasingly used in rPPG extraction. Here we list some typical methods. In 2018, Chen *et al.* [19] introduced an end-to-end system to obtain HR and respiration rate. A convolutional neural network (CNN) combined with an attention mechanism was designed to establish the mapping between video frames and the desired physiological information. In the same year, Špetlłk *et al.* [20] put forward a two-step CNN composed by a feature extractor and an HR estimator to estimate the HR from a series of facial images. Niu *et al.* [21] proposed a spatiotemporal representation of HR information and designed a general-to-special transfer learning strategy to estimate HR from the representation. Later, the authors also applied a channel and spatial-temporal attention mechanism to further improve the HR estimation from face videos [22]. Yu *et al.* [23] proposed an end-to-end deep learning method to retrieve rPPG pulse signals from videos in highly compressed formats. They also explored the benefit of employing neural architecture search to enhance the performance of end-to-end rPPG methods [24]. Song *et al.* [12] designed a feature-decoder framework to map a novel spatiotemporal map to the corresponding HR value through a CNN. They also took a transfer learning to reduce the demand of training data and accelerate the convergence of model.

The goal of above DL-based rPPG methods is to determine accurate HR values. There are also some DL methods that

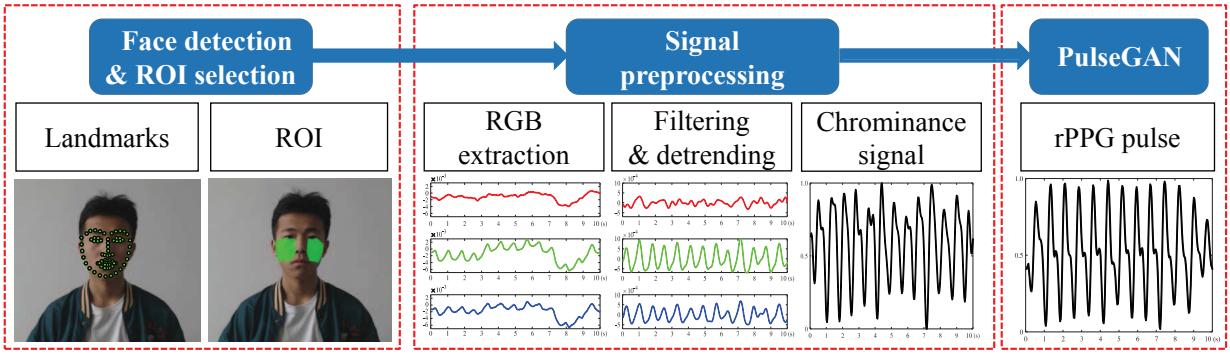


Fig. 1. The framework of the proposed PulseGAN method. First, 68-point facial landmarks [26] are detected and a region of interest (ROI) is defined. Subsequently, the RGB signals of ROI are extracted from the videos, and after filtering, etc., the CHROM algorithm is used to obtain rough pulse signals. Finally, a high-quality pulse waveform is obtained through denoising the rough CHROM signal with the PulseGAN, thereby calculating more accurate physiological parameters.

can directly generate pulse waveforms. For example, Bian *et al.* [25] proposed a new regression model that used a two-layer long short-term memory (LSTM) to filter the noisy rPPG signals. Slapničar *et al.* [26] also employed a LSTM model to enhance the rough rPPG signals obtained by the POS algorithm. In [27], Yu *et al.* introduced an end-to-end way to extract pulse signal with deep spatial-temporal convolutional networks from the original face sequences. Particularly, the authors also calculate the HRV features to evaluate the quality of pulse waveforms. Although these articles have made significant progresses in extracting waveforms, we still need to consider various factors that affect the generation of waveforms, such as loss functions, the network structures, and the design of input and output etc., to further improve the generated waveforms.

This paper aims to introduce a new framework for enhancing pulse waveform quality with cGAN. We will verify that the proposed PulseGAN framework employing a combination of the waveform loss, the spectrum loss, and the adversarial loss outperforms the one with only a waveform loss on the quality of waveform enhancement from coarse inputs.

III. Method

In this section, we introduce the details of the proposed PulseGAN framework for cardiac pulse extraction. The overall framework of PulseGAN is shown in Fig.1. First, 68-point facial landmarks [28] are detected and a region of interest (ROI) is defined according to those landmarks covering the left and right cheeks. Second, the pixels within the selected ROI are averaged to get the RGB channels, and the CHROM algorithm is used to obtain a rough pulse signal that will be taken as the input of PulseGAN. Finally, a high-quality pulse waveform is obtained through denoising the rough CHROM signal with the PulseGAN.

A. Acquisition of rough rPPG pulses

A rough rPPG pulse signal is obtained with some conventional method before feeding into the PulseGAN. It can significantly simplify the training difficulty of PulseGAN if the rough rPPG pulse is close enough to its reference one. In this paper, the CHROM [6] proposed by De Haan *et al.* as the candidate to extract the rough pulse signal. Theoretically, other conventional methods can also be used. We choose the CHROM method because it is fast and stable against motion artifacts.

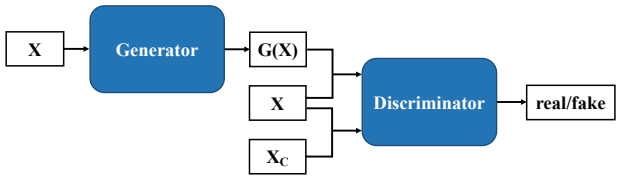


Fig. 2. The conditional GAN structure used in PulseGAN.

The principle of CHROM is based on the skin optical reflection model [7]. The chrominance signals S_1 and S_2 are defined based on a projection of standardized RGB signals to remove the specular reflection terms. The rough pulse signal X is then calculated through an alpha tuning technique as $X = S_{1,f} - \alpha S_{2,f}$, where $\alpha = \sigma\left(S_{1,f}\right)/\sigma\left(S_{2,f}\right)$, σ indicates the standard deviation operation, and the $S_{1,f}$ and $S_{2,f}$ are band-pass version of S_1 and S_2. To standardize all input signals, the obtained CHROM signal is de-trended and then normalized to a range of $[0, 1]$.

B. The PulseGAN framework

The overall structure of the PulseGAN is as shown in Fig.2. The PulseGAN is composed of a generator G and a discriminator D. The generator G is taken to map the rough CHROM signal X to a target rPPG signal $G(X)$ that is close to the reference PPG signal X_c. The discriminator D is used to distinguish the ground truth X_c from the signals $G(X)$. To better pair the inputs and outputs, we refer to the conditional GAN [17] approach, where the input X is set as a condition in the discriminator. Therefore, the input of the discriminator is composed of two channels as $(G(X), X)$ or (X_c, X). The discriminator D outputs a lower score for the input $(G(X), X)$, while it gives a higher score for the input (X_c, X). The characteristics of the PPG signal are continuously learned through an adversarial learning between

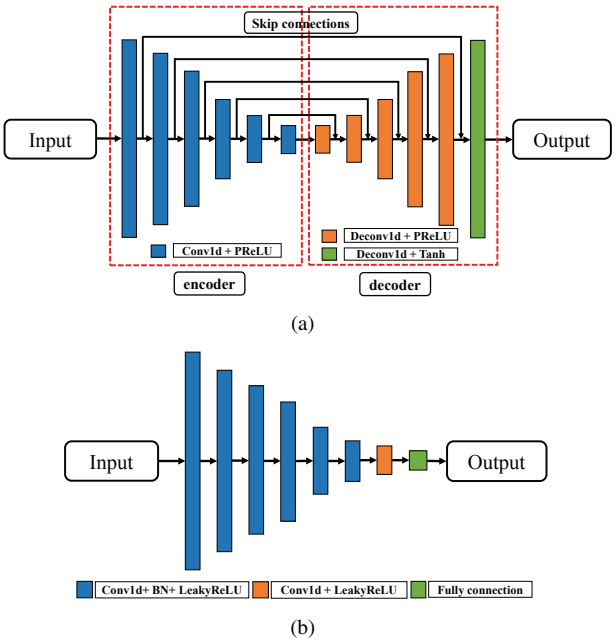


Fig. 3. The network structure of PulseGAN. (a) The generator network. (b) The discriminator network.

the generator and the discriminator, so that the output signal has a distribution as close as that of the reference PPG signal.

The network structures of PulseGAN are designed with reference to SEGAN [14]. The generator, as shown in Fig.3(a), is similar as a denoising autoencoder with several skip connections. As seen, both the encoder and the decoder have six hidden layers, which are less than the ones in SEGAN. Besides, we also remove the latent vector z in SEGAN. These modifications can reduce the risk of overfitting in generating the rPPG waveforms. In detail, the encoder is composed of six one-dimensional convolution layers, while the decoder has six deconvolution layers. The parametric rectified linear units (PReLUs) and Tanh are taken as the nonlinear activation functions. The skip connections are taken to transfer fine-grained features from the encoder to its counterpart in the decoder. This is important for the generator to construct high-quality waveforms.

The discriminator is also a stack of several 1D convolutional layers together with a fully connected layer in the last layer as shown in Fig.3(b). The LeakyReLU is chosen as the nonlinear activation function and batch normalization is employed to accelerate the convergence. The input of D has two channels, where the CHROM signal X is used as a condition. The discriminator makes judgments on the generated waveform $(G(X), X)$ and its reference one (X_c, X), respectively. The output value of D represents the probability that the discriminator considers the input to be real data.

C. Loss function

The purpose of PulseGAN is to generate a waveform $G(X)$ from its input X. $G(X)$ is expected to be as close as its reference signal X_c. This is achieved through training the PulseGAN with a lot of paired data. Since the pulse signal has a clear time-domain and frequency-domain characteristics, we define error losses in both domains to better guide the generator to learn the features of the reference signal. Therefore, we define the loss function of the generator and discriminator as follows:

$$L_G = \frac{1}{2}(D(G(X), X) - 1)^2 + \lambda \parallel X_c - G(X) \parallel_1 + \beta \parallel X_{cf} - G_f(X) \parallel_1 \quad (1)$$

and

$$L_D = \frac{1}{2}(D(G(X), X))^2 + \frac{1}{2}(D(X_c, X) - 1)^2. \quad (2)$$

The first term of L_G is an adversarial loss similar as the least square GAN (LSGAN) [29], the second and third ones are the waveform loss and the spectrum loss defined in time domain and frequency domain, respectively. The loss function of discriminator remains the same as the LSGAN. It enforces D to distinguish the generated and the reference signals. Here the $G_f(X)$ and X_{cf} in the spectrum loss are calculated as the spectrums by a 1024-point fast Fourier transform (FFT) on $G(X)$ and X_c, respectively. And $\parallel \cdot \parallel_1$ indicates the L_1 norm. The λ and β are the weights of the waveform loss and the spectrum loss, respectively. The generator is enforced to learn the time-frequency characteristics through minimizing the error losses. Therefore, the quality of generated waveforms can be effectively improved.

IV. Experiments

In this section, we will evaluate the proposed PulseGAN on several public databases to illustrate its effectiveness. In detail, the following experiments are conducted: 1) test on the UBFC-RPPG database [30] in a within-database way; 2) test on the UBFC-RPPG, PURE [10], and MAHNOB-HCI [31] databases in a cross-database way; 3) evaluate the influence of different loss functions through ablation study. The proposed method is compared with several other methods using quality metrics defined on the generated waveforms, such as the averaged HR, HRV, IBI, and the SNR.

A. Experimental setup

The following five databases are involved in the experiment including the UBFC-RPPG, the PURE, the VIPL-HR [32], the MAHNOB-HCI, and the in-house BSIPL-RPPG databases. All databases are comprised of facial videos and physiological signals. Particularly, the reference physiological signals in MAHNOB-HCI database are ECG signals, while the BVP signals are provided in the other four databases. The HR distribution of each database is shown in Fig.4. As can be seen, the in-house BSIPL-RPPG, UBFC-RPPG and MAHNOB-HCI databases have wider ranges of HR distributions compared to that of the other two. The PURE database has a HR distribution mainly concentrating at both ends, whereas most of the HR values of the VIPL-HR database fall into the range from 50 to 80 bpm. That is because we have removed reference PPG signals with poor quality in the VIPL-HR database to avoid affecting the training of the network.

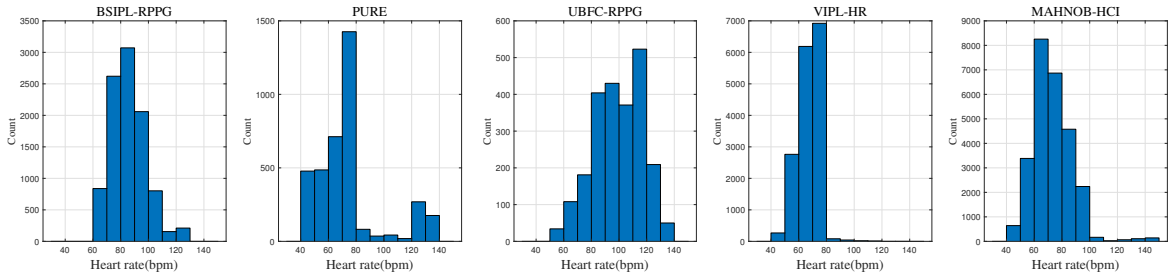


Fig. 4. The HR distributions of reference PPG pulses in BSIPL-RPPG, PURE, UBFC-RPPG, VIPL-HR, and MAHNOB-HCI databases, respectively.

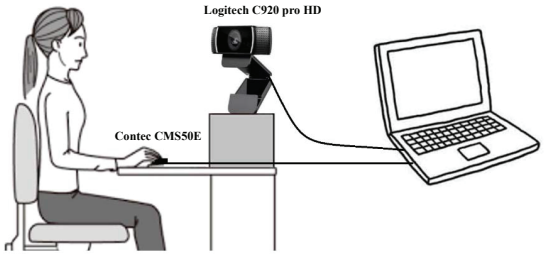


Fig. 5. Setup of the in-house BSIPL-RPPG database.

The proposed method is tested in two scenarios, the within-database case and cross-database case. For the within-database scenario, the UBFC-RPPG database is divided into training set and testing set according to the principle of subject-independent. For the cross-database scenario, the UBFC-RPPG, PURE and MAHNOB-HCI databases are used as the testing sets, respectively. According to characteristics of testing data sets, different training sets are set up in the cross-database scenario. Particularly, the in-house BSIPL-RPPG and PURE databases are taken as the training set to test the UBFC-RPPG database. The in-house BSIPL-RPPG and UBFC-RPPG databases are taken as the training set to test the PURE database. The VIPL-HR database is taken as the training set to test the MAHNOB-HCI database.

We use a 10-second sliding window to process all videos and physiological signals for both scenarios. However, the sliding step in the within-database case is taken as 0.5 seconds, while in the cross-database case, 0.5 seconds are set for generating training samples, and 1 second is used for generating the testing samples. A smaller sliding step can help to increase the number of training samples for the within-database case. All reference physiological signals are resampled to be aligned with the video frame rate.

We train the proposed PulseGAN for 30 epochs using the Adam optimizer. The initial learning rate is set to 0.001, and it is adaptively adjusted through a dynamic learning rate scheduler, the 'ReduceLROnPlateau' with the factor to 0.1 and patience to 3.0. The weight parameters α and β in Eq. (1) are both taken as 10.0 to balance the waveform and spectrum losses.

B. Databases

The UBFC-RPPG database [30] includes 42 videos under a realistic situation. The subjects were asked to play a time-sensitive mathematical game in order to keep the HR varied. The videos were recorded by a webcam (Logitech C920 HD Pro) with a spatial resolution of 640×480 pixels and a frame rate of 30 fps. Each video is about 2 minutes long, and the PPG pulse signals are collected simultaneously by the pulse oximeter (Contec Medical CMS50E) with a 60 Hz sampling rate.

The PURE database [10] contains 60 videos from 10 subjects (8 male and 2 female). Each subject performed six different kinds of head motions, including steady, talking, slow translation, fast translation, small rotation, and medium rotation. Each video is about 1 minute long and recorded by an ECO274CVGE camera with a resolution of 640×480 pixels and a frame rate of 30 fps. The PPG pulse signals are also collected by the Contec CMS50E pulse oximeter while recording each video.

The VIPL-HR database [32] contains 2378 visible light videos and 752 near-infrared videos from 107 subjects (79 males and 28 females, the ages are between 22 and 41 years old). Only visible light videos are used in the experiments. The database contains 9 scenarios recorded by 3 different devices (Logitech C310 web-camera, the front camera of HUAWEI P9 smartphone, and RealSense F200 camera). The frame rates of the videos in VIPL-HR database from 25 fps to 30 fps, and the resolution is 960×720 and 1920×1080. The ground-truth physical signals were recorded using a pulse oximeter (CONTEC CMS60C BVP sensor).

The MAHNOB-HCI database [31] consists of 527 videos in total. 15 female and 12 male participants are involved with ages varying between 19 to 40 years old. All videos are recorded at 61 fps with 780×580. Only the ECG signals are recorded but not the BVP signals.

The BSIPL-RPPG is an in-house rPPG database including 37 healthy student subjects (24 male and 13 female with age ranging from 18 to 25 years old). The experimental setup is illustrated in Fig.5. The subjects were asked to sit in front of the camera (Logitech C920 pro HD) at a distance of 1.0 meter. A Contec CMS50E pulse oximeter was clamped on the subject's finger to acquire the PPG signal synchronously. Both the camera and the pulse oximeter were connected to a computer to transfer the acquired data in real time. The videos were recorded with a resolution of 640×480 pixels under a frame rate of 30 fps. Meanwhile, the PPG signal was collected by the pulse oximeter at a 60 Hz sampling rate. Each video and its counterpart PPG signal last about 4.5 minutes long.

TABLE I
THE RESULTS ON UBFC-RPPG DATABASE: A WITHIN-DATABASE CASE.

Method	HR(bpm)				HRV(ms)		IBI(ms)	Pulse(dB)
	MAE	RMSE	MER	R	AVNN_{mae}	SDNN_{mae}	IBI_{mae}	SNR
GREEN [2]	7.50	14.41	7.82%	0.62	—	—	—	—
ICA [5]	5.17	11.76	5.30%	0.65	—	—	—	—
POS [7]	4.05	8.75	4.21%	0.78	—	—	—	—
CHROM [6]	2.37	4.91	2.46%	0.89	16.54	40.90	63.20	6.63
Bobbia *et al*. [30]	—	2.388	—	0.961	—	—	—	—
Benezeth *et al*. [33]	1.21	2.41	—	0.82	—	—	—	—
Tsou *et al*. [34]	**0.48**	**0.97**	—	—	—	—	—	—
DAE	1.48	2.49	1.55%	0.97	9.52	19.25	41.27	3.58
PulseGAN	1.19	2.10	**1.24%**	**0.98**	**7.52**	**18.36**	**39.60**	**7.90**

The subjects were requested to sit still for the first 2 minutes, and perform some apparent head movements for the last 2.5 minutes.

C. Metrics

We define several metrics to evaluate the quality of the generated pulse waveform. First, the IBI sequences are calculated separately for the generated and reference pulse signals. A series of cardiac features can then be defined according to the calculated IBI. For example, the average HR can be calculated from IBI as [35]

$$\text{HR} = 60/\overline{\text{IBI}}. \tag{3}$$

where $\overline{\text{IBI}}$ is the average value of the IBI sequence for the current processing window. Similarly, we can also get HRV features [36] of AVNN and SDNN as follows,

$$\text{AVNN} = \frac{1}{T}\sum_{i=1}^{T}\text{RR}_i \tag{4}$$

and

$$\text{SDNN} = \sqrt{\frac{1}{T-1}\sum_{i=1}^{T}(\text{RR}_i - \text{AVNN})}, \tag{5}$$

where AVNN indicates the average of all normal-to-normal (NN) intervals, SDNN is the standard deviation of all NN intervals, RR_i represents the i-th R-R interval, and T is the total number of R-R intervals.

Finally, we define the following error metrics to compare the HR, HRV (AVNN and SDNN), IBI, and SNR calculated from the PulseGAN and the reference signals.

1) **HR**: The metrics of HR values include the mean absolute error (MAE), the root mean square error (RMSE), the mean error rate percentage (MER), and the Pearsons correlation coefficient (R). The formulas of these metrics refer to [12].
2) **HRV**: The mean absolute error of AVNN (or SDNN) is calculated as below:

$$Y_{mae} = \frac{1}{N}\sum_{n=1}^{N}\left|Y_n^{'} - Y_n\right|, \tag{6}$$

where Y_n indicates the AVNN (or SDNN) for the nth window calculated from PulseGAN, $Y_n^{'}$ is the AVNN (or SDNN) from its reference PPG signal, and N is the total number of time windows.

3) **IBI**: We also define metrics to evaluate the quality of IBI directly. Since the length of the IBI vectors may be different, we refer to a similar way in [37] to solve this issue. Namely, each IBI vector is expanded to the same length as the PPG signal. We pad the i-th RR interval of the IBI sequence with values all equal to RR_i. After the padding operation, we define the absolute error $\text{IBI}_{ae}^{(n)}$ for the nth window as below

$$\text{IBI}_{ae}^{(n)} = \mathbb{E}(\mid \text{IBI}_{predict}^{(n)} - \text{IBI}_{label}^{(n)} \mid), \tag{7}$$

where $\mathbb{E}$ refers to the mathematical expectation, $\text{IBI}_{predict}^{(n)}$ is the padded IBI vector of rPPG pulse, and $\text{IBI}_{label}^{(n)}$ is the padded IBI vector of the ground truth. Finally, a mean absolute error for IBI vectors from all samples is calculated by

$$\text{IBI}_{mae} = \frac{1}{N}\sum_{n=1}^{N}\text{IBI}_{ae}^{(n)}, \tag{8}$$

where N is the total number of time windows.
4) **SNR**: In order to directly compare the quality of generated waveforms, we also calculate the SNR of the pulses referring to the definition in [6].

D. Experimental results

The experimental results are introduced following a sequence of within-database and cross-database configurations, respectively.

Within-database: We first perform the within-database testing on the UBFC-RPPG database. According to the time window and the sliding step configuration, we totally get 4234 samples, where we take the 3192 samples from the first 30 subjects as the training set, and the remaining 1042 samples from the last 12 subjects as the testing set.

The estimation results are summarized in Table I, which are also compared with some existing methods. Among them, GREEN [2], ICA [5], POS [7], and CHROM [6] have been implemented with an open source toolbox [38]. The DAE (denoising autoencoder) here refers to the method of using the generator G of PulseGAN with only a waveform error loss. The results of the other three methods were directly taken from corresponding papers due to complexity of implementation. From the results, we observe that the proposed method outperforms the other comparison methods except for the average HR results in [34]. We need to note that the testing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3051176, IEEE Journal of Biomedical and Health Informatics

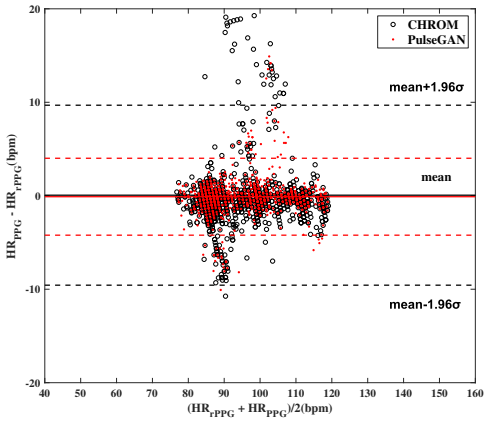


Fig. 6. Bland-Altman plots between the predicted HR ($\mathrm{HR_{rPPG}}$) and the reference HR ($\mathrm{HR_{PPG}}$) on UBFC-RPPG database for a within-database case: CHROM vs PulseGAN.

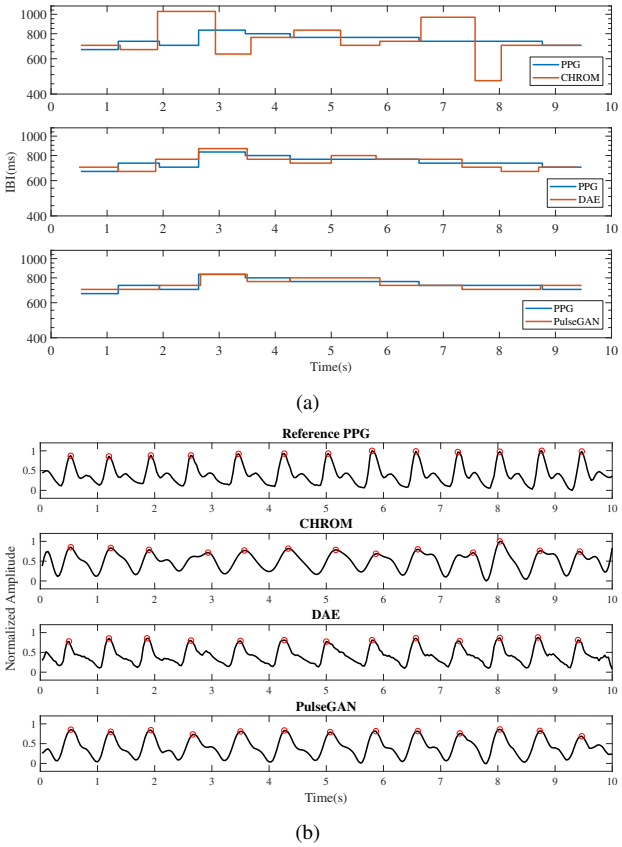


Fig. 7. A comparison example of IBI sequence in (a) and rPPG pulse signals in (b) on UBFC-RPPG database: a within-database case.

data used in [34] was lack of details and it may not be the same as what we used here. However, we can still see that the PulseGAN significantly improves the quality of generated pulses compared to the input CHROM signal, especially for the IBI and HRV related metrics.

To further evaluate the proposed method, the Bland-Altman plots are shown in Fig.6. We can observe that the PulseGAN has much better consistency with the ground truth compared to

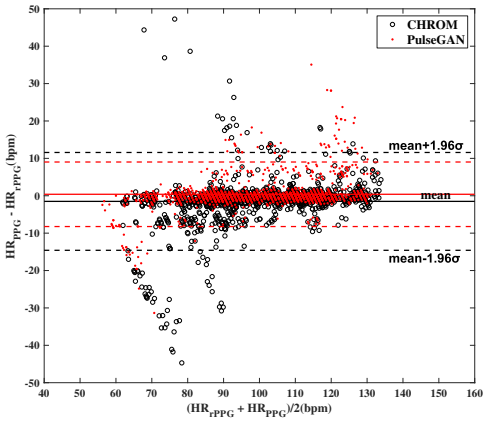


Fig. 8. Bland-Altman plots between the predicted HR ($\mathrm{HR_{rPPG}}$) and the reference HR ($\mathrm{HR_{PPG}}$) on UBFC-RPPG database for a cross-database case: CHROM vs PulseGAN.

CHROM. To demonstrate the improvement of the waveform quality more intuitively, in Fig.7, we show a sample of the pulse signal and corresponding IBI sequence. It can be seen that the waveform and IBI sequence of the example pulse signal are both significantly improved by DAE and PulseGAN compared to CHROM. In Fig. 7(a), the three sub-figures respectively show the comparison between the IBI of the reference signal PPG and the IBI obtained by CHROM, DAE and PulseGAN. The IBI_{ae} errors of the example in Fig. 7(a) are 112.44, 42.50, and 24.67 ms for CHROM, DAE, and PulseGAN, respectively.

***Cross-database**:* In the case of cross-database, we test three databases, including the UBFC-RPPG, PURE, and MAHNOB-HCI, respectively.

***1) Test with UBFC-RPPG**:* We take the PURE and BSIPL-RPPG databases as the training set. This combination can effectively balance the number of samples in different HR ranges to achieve a more consistent HR distribution with the testing set. According to the configuration of cross-database scenario, there are total 13484 training samples obtained from the PURE (3727 samples) and BSIPL-RPPG (9757 samples) databases. We note that the sliding step of 0.5 seconds was applied to generate training samples and only part of the samples were kept for training. Next, we get total 1470 samples from the UBFC-RPPG database as the testing set. The number of testing samples is less than we used for the within-database case since a 1-second sliding step is taken for the cross-database testing case instead of the 0.5 seconds used for the within-database case.

The average HR measurements are summarized in Table II. From the results, we can see that our method achieves the best performance among all the comparison methods except for the MAE of HR in [34]. Particularly, the PulseGAN improves 20.85% (41.19%) for the AVNN_{mae}, and improves 20.28% (37.53%) for the SDNN_{mae}, compared to the DAE (CHROM). Similarly, the Bland-Altman plots are illustrated in Fig.8 to show the consistency of the predicted HR values with the reference ones. We can see that the PulseGAN clearly outperform the CHROM method.

TABLE II
THE RESULTS ON UBFC-RPPG DATABASE: A CROSS-DATABASE CASE.

Method	HR(bpm)				HRV(ms)		IBI(ms)	Pulse(dB)
	MAE	RMSE	MER	R	AVNN_{mae}	SDNN_{mae}	IBI_{mae}	SNR
GREEN [2]	8.29	15.82	7.81%	0.68	—	—	—	—
ICA [5]	4.39	11.60	4.30%	0.82	—	—	—	—
POS [7]	3.52	8.38	3.36%	0.90	—	—	—	—
CHROM [6]	3.10	6.84	3.83%	0.93	25.30	38.96	60.16	6.681
Bousefsaf *et al*. [39]	5.45	8.64	—	—	—	—	—	—
Tsou *et al*. [34]	**1.29**	8.73	—	—	—	—	—	—
Lee *et al*. [40]	5.97	7.42	—	0.53	—	—	—	—
DAE	2.70	5.17	2.85%	0.96	18.80	30.53	49.65	4.847
PulseGAN	2.09	**4.42**	**2.23%**	**0.97**	**14.88**	**24.34**	**42.27**	**7.633**

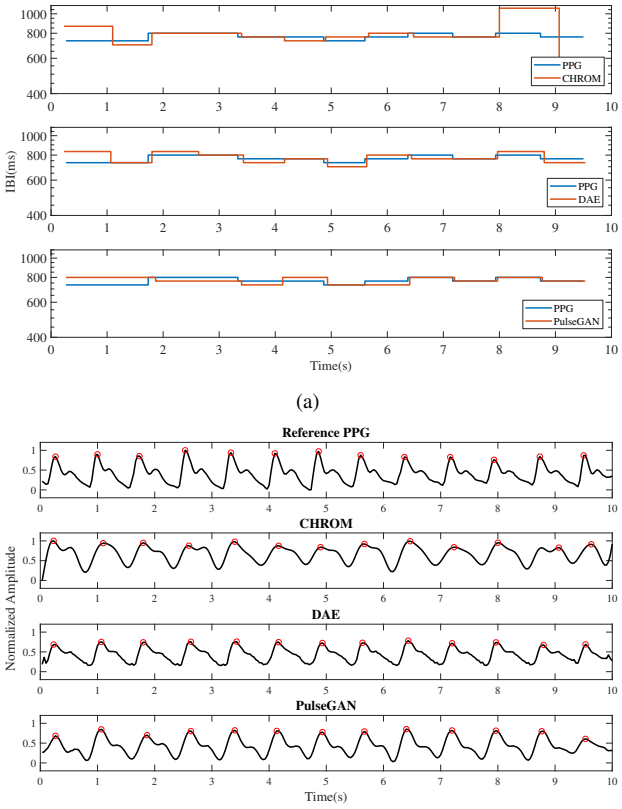


Fig. 9. A comparison example of IBI sequence in (a) and rPPG pulse signals in (b) on UBFC-RPPG database: a cross-database case.

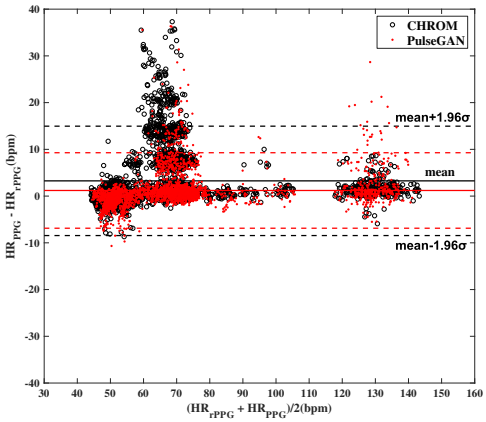


Fig. 10. Bland-Altman plots between the predicted HR (HR_{rPPG}) and the reference HR (HR_{PPG}) on PURE database for a cross-database case: CHROM vs PulseGAN.

Finally, we take an example to demonstrate the intuitive enhancement on waveforms and the IBI sequence. As can be seen in Fig.9, the IBI sequence and the related pulse waveform obtained by PulseGAN are more close to their ground truths compared to that of DAE and CHROM. The IBI_{ae} errors of the example in Fig. 9(a) are 65.01, 27.44, and 23.11 ms for the CHROM, DAE, and PulseGAN, respectively. The experimental results of PulseGAN for the cross-database case indicates the good generalization capability of the proposed model.

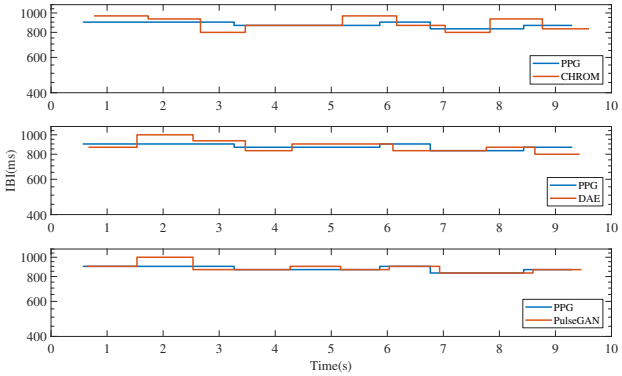
2) Test with PURE: To test the PURE database, the UBFC-RPPG together with the in-house BSIPL-RPPG databases are used to prepare training set. Particularly, data augmentation similar as [22] was done to balance the reference HR distributions of training and testing databases as shown in Fig.4. After the above preprocessing, we get a total of 32505 samples for the training data set. The number of all testing samples is 3198 under a 1-second sliding step.

The comparison results of different methods are summarized in Table III, where the results of the first three methods are directly taken from the corresponding papers. We observe that the proposed PulseGAN clearly improves the results compare to the input CHROM, which verifies the effectiveness of the proposed method. A comparison example of IBI sequence and rPPG pulse signal is shown in Fig.11. The comparison results also show that the proposed method has demonstrated the improvement of waveform and IBI vector. The IBI_{ae} errors of this example are 42.11, 36.67, and 17.67 ms for CHROM, DAE, and PulseGAN, respectively. Finally, the Bland-Altman plots are illustrated in Fig.10. We can see that the PulseGAN has much better consistency with the ground truth compared to CHROM.

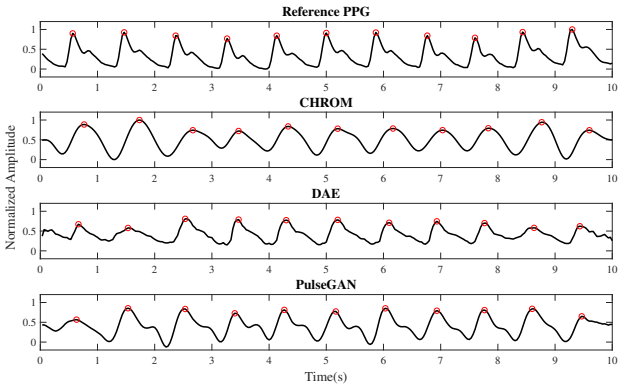
3) Test with MAHNOB-HCI: To test the proposed method with the MAHNOB-HCI database, we prepare the training set with the VIPL-HR database. The CHROM signals are also extracted from these two databases for the input signals of the network. As reported in many studies [19], [43], [44], the CHROM method does not perform well in these two databases because the videos are compressed and the acquisition scenarios are complicated. We observe that some

TABLE III
THE RESULTS ON PURE DATABASE: A CROSS-DATABASE CASE.

Method	HR(bpm)				HRV(ms)		IBI(ms)	Pulse(dB)
	MAE	RMSE	MER	R	AVNN_{mae}	SDNN_{mae}	IBI_{mae}	SNR
NMD-HR [41]	8.68	—	—	—	—	—	—	—
Zhao *et al*. [42]	3.09	4.26	—	—	—	—	—	—
Tsou *et al*. [34]	**0.63**	**2.51**	—	—	—	—	—	—
CHROM [6]	3.82	6.8	5.30%	0.97	49.63	89.3	107.4	5.499
DAE	3.24	5.97	4.33%	0.97	39.09	74.28	90.97	5.219
PulseGAN	2.28	4.29	**3.33%**	**0.99**	**28.92**	**49.39**	**65.26**	**6.56**



(a)



(b)

Fig. 11. A comparison example of IBI sequence in (a) and rPPG pulse signals in (b) on PURE database: a cross-database case.

of the extracted CHROM signals are totally distorted and the quality of some reference PPG signals in the VIPL-HR database are also quite poor.

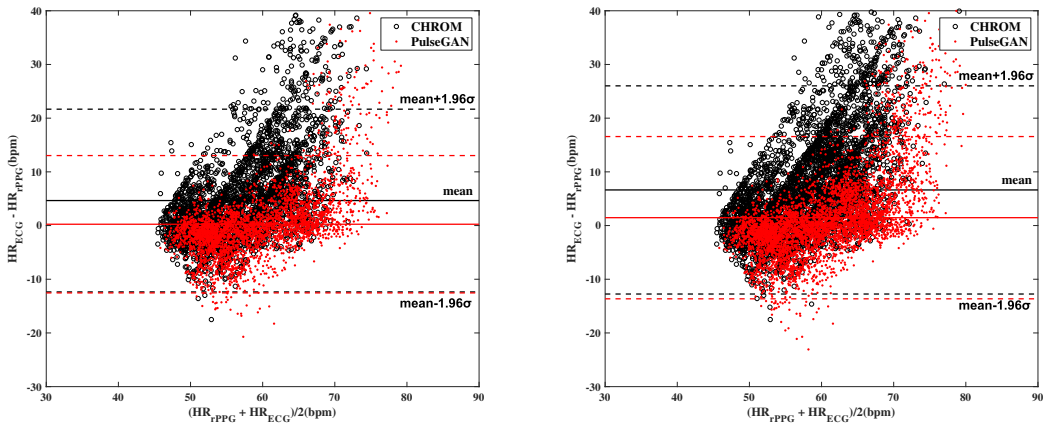
To ensure the convergence of training, we clean the training data to discard low quality signals. In detail, we first remove training samples with bad reference PPG signals through a check based on template matching. For the input CHROM signals in the remaining training data set, it is hard to directly remove the bad samples because most of the samples are occupied by noise. In order to balance the quality and the size of training data set, we implement the quality enhancement on the input CHROM signals for the training data as described below. Finally, we take the same data augmentation as done above for testing the PURE database to balance the HR distributions. After the above preprocessing, we get 31575 samples for the training data set.

The enhancement details on the CHROM signals for training data are shown here. Suppose the ground truth PPG signal is y and the CHROM signal is x. Let d denote the difference signal between x and y as $d = y - x$. The purpose is to reconstruct the input as $x' = x + f(d)$, which should have better quality compared to the original x. We implement a multi-scale reconstruction of the input signal using the discrete cosine transform (DCT). Suppose the signal d is decomposed into DCT basis vectors X, of which the expansion coefficients measure the energy stored in each of the components. We retain those largest expansion coefficients to achieve the multi-scale reconstruction. Let $\tau (0 \leq \tau \leq 1)$ indicate the ratio of the energy kept for the remaining DCT coefficients. If $\tau = 0$, it represents $f(d) = 0$ and x' is equivalent to x. If $\tau = 1$, it represents $f(d) = d$ and x' is the same as the reference y. In the experimental results listed below, we choose the ratio $\tau = 0.7$ for enhancing CHROM signals in the training data. We have also tried to train the model using enhanced CHROM signals with $\tau = 0.5$. We observe that the testing performance is similar as that of $\tau = 0.7$, and thereby omitting the results here.

To verify the generalization capability, the PulseGAN model (ratio $\tau = 0.7$) is tested on input signals of MAHNOB-HCI database with SNR above different thresholds. We indicate that the CHROM signals obtained in MAHNOB-HCI database are still the original ones expect a selection based on SNR. The input selection ensures that the quality of the input data is not too bad, and otherwise the method will fail as discussed later. The testing results for inputs with different SNRs are summarized in Table IV. The numbers of testing samples are 3512, 4983, 6783 and 8885 for SNR above 0 dB to -3 dB, respectively. It can be seen that the proposed PulseGAN can significantly improve the quality of input signals with different SNRs. We also observe that the average errors in Table IV of CHROM with SNR> 0 dB are slightly larger than the results obtained by Song *et al*. [12]. After mapping by the PulseGAN model, the quality metrics of generated pulses consistently outperform the ones from [12]. This verifies the benefit of the proposed method to combine with some existing methods to further improve their results. This is especially important when pulse waveforms are required for calculation of HRV related features. Similarly, the Bland-Altman plots of SNR> 0 dB and SNR> -1 dB are illustrated in Fig.12 to show the consistency of the predicted HR values with the reference ones. We can see that the results of CHROM method can be significantly improved by the PulseGAN.

TABLE IV
THE RESULTS ON MAHNOB-HCI DATABASE: A CROSS-DATABASE CASE.

Method	HR(bpm)				HRV(ms)		IBI(ms)	Pulse(dB)
	MAE	RMSE	MER	R	AVNN_{mae}	SDNN_{mae}	IBI_{mae}	SNR
DeepPhys [19]	4.57	—	—	—	—	—	—	—
RhythmNet [43]	—	8.28	8.00%	0.64	—	—	—	—
PhysNet128 [27]	6.85	8.76	—	0.69	—	—	—	—
Song *et al.* [12]	5.98	7.45	7.97%	**0.75**	—	—	—	—
CHROM [6] (SNR>-3dB)	11.35	15.39	15.91%	0.08	155.28	196.17	266.68	-0.16
PulseGAN	7.03	10.46	10.07%	0.46	84.77	94.94	149.40	1.47
CHROM [6] (SNR>-2dB)	9.70	13.69	14.00%	0.15	138.47	192.97	251.2	0.56
PulseGAN	5.98	9.11	9.00%	0.55	75.33	88.53	136.94	2.59
CHROM [6] (SNR>-1dB)	8.08	11.91	11.85%	0.26	121.66	186.62	235.45	1.31
PulseGAN	5.03	7.85	7.77%	0.64	67.52	81.91	125.85	3.66
CHROM [6] (SNR>0dB)	6.38	9.85	9.72%	0.36	103.48	178.88	218.00	2.08
PulseGAN	**4.15**	**6.53**	**6.77%**	0.71	**60.40**	**74.39**	**114.74**	**4.67**



(a) (b)

Fig. 12. Bland-Altman plots between the predicted HR (HR_{rPPG}) and the reference HR (HR_{ECG}) on MAHNOB-HCI database for a cross-database case: CHROM vs PulseGAN. (a): SNR>0 dB, (b): SNR>-1 dB.

Ablation study: We further take an ablation study to evaluate different factors that affect the performance of the proposed method. The ablation study is also done on the MAHNOB-HCI database using the same training model obtained from VIPL-HR database with ratio $\tau = 0.7$. In the ablation study, we consider two aspects that may affect the performance of the PulseGAN. To simplify the analysis, we only conduct ablation study with input testing data satisfying SNR>0 dB in the MAHNOB-HCI database.

First, we evaluate the influence of using adversarial and spectrum losses in the proposed method. Different combinations of loss functions will be tested with details listed as below:
1) Use three losses simultaneously *(wf, sp, adv)*;
2) Remove the adversarial loss *(wf, sp)*;
3) Remove the spectrum loss *(wf, adv)*;
4) Remove both the adversarial loss and spectrum loss *(wf)*, also denoted as DAE,
where *adv* represents the adversarial loss, *wf* is the waveform loss in the time domain, and *sp* is the spectrum loss. We add the case 4) here to understand the effect of using simultaneously the adversarial loss and spectrum loss.

The results of ablation study on loss functions are summarized in Table V. It can be seen that the performance of PulseGAN degrades compared to the case of using full loss functions if either the adversarial loss or the spectrum loss is removed. However, the gap is not very significant. However, the improvement on the last four columns in Table V between the full PulseGAN *(wf, sp, adv)* and the DAE is clear. It can be seen that the SDNN_{mae}, IBI_{mae}, and the SNR improves 13.62%, 13.13%, and 15.22%, respectively. Similar results were also observed in Table II and Table III for the UBFC-RPPG and PURE database, respectively. This indicates that the simultaneous use of both spectrum and adversarial losses can significantly enhance the waveform quality, which is helpful to calculate reliable waveform-related features. Theoretically, the waveform and spectrum losses enforce the generated pulse to match with the reference one from the spatial-temporal feature features. The use of adversarial loss can further restrict the generation to be more realistic as the reference one from high-level features.

Second, we evaluate the performance of the PulseGAN model to test input signals obtained in different ways. The purpose is to verify the robustness and generalization of the PulseGAN model. Three types of input signals, including the CHROM signals obtained from only the left cheek region and only the right cheek region (as illustrated in Fig. 1), respectively, and also the green signal obtained from the whole

TABLE V
THE RESULTS OF THE ABLATION STUDY FOR LOSS FUNCTIONS ON MAHNOB-HCI DATABASE: A CROSS-DATABASE CASE.

Method	HR(bpm)				HRV(ms)		IBI(ms)	Pulse(dB)
	MAE	RMSE	MER	R	AVNN_{mae}	SDNN_{mae}	IBI_{mae}	SNR
CHROM [6]	6.38	9.85	9.72%	0.36	103.48	178.88	218	2.08
DAE	4.54	7.13	7.40%	0.63	64.67	86.12	132.09	4.05
(wf, adv)	4.46	6.77	7.35%	0.68	65.22	79.03	123.76	4.28
(wf, sp)	4.28	6.63	7.00%	0.70	62.72	78.48	119.82	4.27
(wf, sp, adv)	**4.15**	**6.53**	**6.77%**	**0.71**	**60.40**	**74.39**	**114.74**	**4.67**

TABLE VI
THE RESULTS OF THE ABLATION STUDY FOR DIFFERENT TESTING INPUTS ON MAHNOB-HCI DATABASE: A CROSS-DATABASE CASE.

Method	HR(bpm)				HRV(ms)		IBI(ms)	Pulse(dB)
	MAE	RMSE	MER	R	AVNN_{mae}	SDNN_{mae}	IBI_{mae}	SNR
CHROM [6] (left cheek)	5.93	9.13	9.30%	0.38	100.71	178.09	218.41	1.90
PulseGAN	4.35	6.43	7.40%	0.69	65.55	76.89	121.22	4.11
CHROM [6] (right cheek)	5.60	8.50	8.94%	0.40	100.23	176.61	214.36	1.88
PulseGAN	4.16	5.99	7.27%	0.71	65.84	76.9	121.38	4.18
GREEN [2]	9.07	14.29	12.31%	0.17	108.5	149.75	200.22	3.05
PulseGAN	5.53	9.91	7.54%	0.58	58.41	68.92	113.03	5.04

ROI. The testing PulseGAN model is still the one trained by VIPL-HR database with ratio $\tau = 0.7$. The results are summarized in Table VI. All the testing inputs are obtained with SNR> 0 dB. Particularly, the errors of GREEN [2] inputs are larger than that of CHROM signals. We can clearly observe that the PulseGAN significantly improves the quality metrics compared to that of the input signals for all cases. This proves good generalization capability of the proposed method.

E. Discussion

The above experimental results verify the effectiveness of the proposed method. Particularly, the testing results of cross-database case on the three databases all reveal the benefit of simultaneous usage of adversarial and spectrum losses. It indicates that the proposed method can significantly improve the quality of coarse CHROM inputs, especially on the cardiac features like IBI and HRV. However, we still need to emphasize that the convergence of training will be affected if the qualities of training samples are too low. The usage of data augmentation and other data enhancement techniques are helpful to achieve reliable training. From the testing aspect, the proposed method usually works well when the quality of inputs is above some criterion.

To demonstrate the restriction on the quality of inputs in an intuitive way, we show a failure case of applying the PulseGAN method on the PURE database as shown in Fig.13. As seen, the quality of input CHROM signal is very low with the SNR as -7.80 dB. It indicates that the CHROM signal is still occupied by noise, of which the discrepancy with the corresponding reference PPG signals is huge. After mapping by the PulesGAN model, the SNR of the waveform is improved to -5.53 dB. Although the PulseGAN model improves the quality metrics compared to the input, the generated results are still degraded due to influence of noise in the inputs. The IBI_{ae} for the CHROM signal in this sample is 123.22 ms, which is reduced to 103.11 ms by the PulseGAN. However, it is still obviously larger than the average IBI_{mae} (65.26 ms) in Table III. This failure case shows that the quality of the testing input data is critical to the success of the PulseGAN.

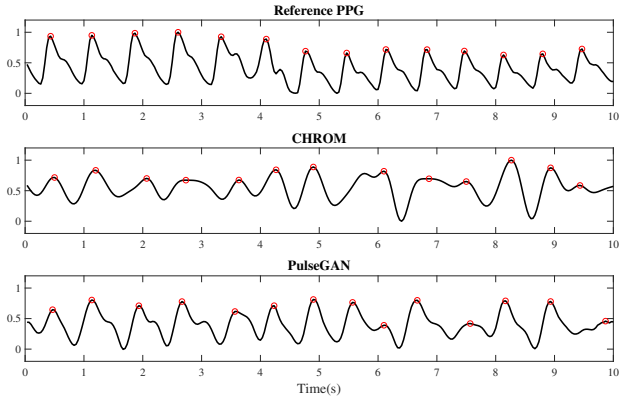


Fig. 13. A failure example from PURE database due to bad input quality: a cross-database case.

In summary, the proposed method provides a general framework to combine with existing methods to further enhance the quality of waveforms. This is useful when high-quality waveforms are required to calculate more cardiac features.

V. CONCLUSION

Cardiac signal is very important to evaluate the healthy and emotional status of human bodies. In this paper, we have proposed a PulseGAN method to extract high-quality pulse waveforms through remote photoplethysmogrpahy. The PulseGAN is designed based on a framework of generative adversarial network with error losses defined in both time and spectrum domains. It takes the rough CHROM signal as the input, and outputs an enhanced rPPG pulse through the deep generative model. It is also easy to integrate the PulseGAN framework with existing rPPG methods for further improving the quality of generated waveforms. The experimental results on three public databases demonstrate that the PulseGAN consistently enhances the quality of coarse input

waveforms for both within-database and cross-database cases. The comparison results with other typical rPPG methods such as the DAE verifies the superior performance of PulseGAN to generate high-quality waveforms. The proposed PulseGAN has demonstrated the feasibility of calculating more reliable cardiac features like the HRV characteristics through rPPG. Although the results in this paper are relatively preliminary, these attempts are meaningful to extend the application scope of rPPG techniques.

REFERENCES

[1] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 463–477, 2016.

[2] 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.

[3] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-based heart rate measurement: Recent advances and future prospects," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 3600–3615, 2018.

[4] Y. Deng and A. Kumar, "Standoff heart rate estimation from video: a review," in *Mobile Multimedia/Image Processing, Security, and Applications 2020*, vol. 11399, International Society for Optics and Photonics. SPIE, 2020, pp. 16 – 29.

[5] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." *Optics Express*, vol. 18, no. 10, pp. 10 762–10 774, 2010.

[6] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[7] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2017.

[8] A. M. Unakafov, "Pulse rate estimation using imaging photoplethysmography: generic framework and comparison of methods on a publicly available dataset," *Biomedical Physics & Engineering Express*, vol. 4, no. 4, p. 045001, apr 2018.

[9] X. Liu, X. Yang, D. Wang, and A. Wong, "Detecting pulse rates from facial videos recorded in unstable lighting conditions: An adaptive spatiotemporal homomorphic filtering algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2020.

[10] R. Stricker, S. Mller, and H. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 1056–1062.

[11] C. Zhao, W. Chen, C. Lin, and X. Wu, "Physiological signal preserving video compression for remote photoplethysmography," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4537–4548, 2019.

[12] R. Song, S. Zhang, C. Li, Y. Zhang, J. Cheng, and X. Chen, "Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7411 –7421, 2020.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems (NIPS)*. Curran Associates, Inc., 2014, pp. 2672–2680.

[14] S. Pascual, A. Bonafonte, and J. Serra, "SEGAN: Speech enhancement generative adversarial network," *arXiv preprint arXiv:1703.09452*, pp. 3642–3646, 2017.

[15] Y. Xiang and C. Bao, "Speech enhancement via generative adversarial LSTM networks," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2018, pp. 46–50.

[16] J. Wang, R. Li, R. Li, K. Li, H. Zeng, G. Xie, and L. Liu, "Adversarial de-noising of electrocardiogram," *Neurocomputing*, vol. 349, pp. 212–224, 2019.

[17] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[18] B. Wei, X. He, C. Zhang, and X. Wu, "Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions," *Biomedical Engineering Online*, vol. 16, no. 1, p. 17, 2017.

[19] W. Chen and D. 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] R. Špetlík, V. Franc, J. Čech, and J. Matas, "Visual heart rate estimation with convolutional neural network," in *Proceedings of British Machine Vision Conference 2018*. British Machine Vision Association (BMVA), 2018, pp. 1–12.

[21] X. Niu, H. Han, S. Shan, and X. 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.

[22] X. Niu, X. Zhao, H. Han, A. Das, A. Dantcheva, S. Shan, and X. 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.

[23] Z. Yu, W. Peng, X. Li, X. Hong, and G. 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 International Conference on Computer Vision*, 2019, pp. 151–160.

[24] Z. Yu, X. Li, X. Niu, J. Shi, and G. Zhao, "AutoHR: A strong end-to-end baseline for remote heart rate measurement with neural searching," *IEEE Signal Processing Letters*, vol. 27, pp. 1245–1249, 2020.

[25] M. Bian, B. Peng, W. Wang, and J. Dong, "An accurate LSTM based video heart rate estimation method," pp. 409–417, 2019.

[26] G. Slapničar, E. Dovgan, P. Čuk, and M. Luštrek, "Contact-free monitoring of physiological parameters in people with profound intellectual and multiple disabilities," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2019, pp. 1–9.

[27] Z. Yu, X. Li, and G. Zhao, "Recovering remote photoplethysmograph signal from facial videos using spatio-temporal convolutional networks," *arXiv preprint arXiv:1905.02419*, 2019.

[28] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 94–108.

[29] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 2794–2802.

[30] 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.

[31] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2011.

[32] X. Niu, H. Han, S. Shan, and X. Chen, "VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video," in *Asian Conference on Computer Vision*. Springer, 2018, pp. 562–576.

[33] Y. Benezeth, S. Bobbia, K. Nakamura, R. Gomez, and J. Dubois, "Probabilistic signal quality metric for reduced complexity unsupervised remote photoplethysmography," in *2019 13th International Symposium on Medical Information and Communication Technology (ISMICT)*, May 2019, pp. 1–5.

[34] Y.-Y. Tsou, Y.-A. Lee, C.-T. Hsu, and S.-H. Chang, "Siamese-rppg network: Remote photoplethysmography signal estimation from face videos," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. Association for Computing Machinery, 2020, pp. 2066–2073.

[35] M. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011.

[36] M. Malik, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology,," *Annals of Noninvasive Electrocardiology*, vol. 1, no. 2, pp. 151–181, 1996.

[37] X. Liu, X. Yang, J. Jin, and A. Wong, "Detecting pulse wave from unstable facial videos recorded from consumer-level cameras: a disturbance-adaptive orthogonal matching pursuit," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 12, pp. 3352 –3362, 2020.

[38] D. McDuff and E. Blackford, "iphys: An open non-contact imaging-based physiological measurement toolbox," *arXiv preprint arXiv:1901.04366*, pp. 1–4, 2019.

[39] F. Bousefsaf, A. Pruski, and C. Maaoui, "3D convolutional neural networks for remote pulse rate measurement and mapping from facial video," *Applied Sciences*, vol. 9, no. 20, 2019.

[40] E. Lee, E. Chen, and C.-Y. Lee, "Meta-rppg: Remote heart rate estimation using a transductive meta-learner," in *European Conference on Computer Vision*. Springer, 2020, pp. 392–409.

[41] H. Demirezen and C. E. Erdem, "Remote photoplethysmography using nonlinear mode decomposition," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1060–1064.

[42] C. Zhao, P. Mei, S. Xu, Y. Li, and Y. Feng, "Performance evaluation of visual object detection and tracking algorithms used in remote photoplethysmography," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019, pp. 1–10.

[43] X. Niu, S. Shan, H. Han, and X. Chen, "Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation," *IEEE Transactions on Image Processing*, vol. 29, pp. 2409–2423, 2020.

[44] X. Niu, Z. Yu, H. Han, X. Li, S. Shan, and G. Zhao, "Video-based remote physiological measurement via cross-verified feature disentangling," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 295–310.