Generative Adversarial Network for Photoplethysmography Reconstruction

Master of Science Thesis
University of Turku
Department of Computing
Health Technology
2022
Yuning Wang

Supervisors:
Ph.D. Iman Azimi
Ph.D. Pasi Liljeberg

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

UNIVERSITY OF TURKU
Department of Computing

YUNING WANG: Generative Adversarial Network for Photoplethysmography Reconstruction

Master of Science Thesis, 53 p.
Health Technology
June 2022

Photoplethysmography (PPG) is an optical measurement method for blood pulse wave monitoring. The method has been widely applied in both clinical and wearable devices to collect physiological parameters, such as heart rate (HR) and heart rate variability (HRV). Unfortunately, the PPG signals are very vulnerable to motion artifacts, caused by inevitable movements of human users. To obtain reliable results from PPG-based monitoring, methods to denoise the PPG signals are necessary.
Methods proposed in the literature, including signal decomposition, time-series analysis, and deep-learning based methods, reduce the effect of noise in PPG signals. However, their performance is insufficient for low signal-to-noise ratio PPG signals, or limited to noise from certain types of activities. Therefore, the aim of this study is to develop a method to remove the motion artifacts and reconstruct noisy PPG signals without any prior knowledge about the noise.
In this thesis, a deep convolutional generative adversarial network (DC-GAN) based method is proposed to reconstruct the PPG signals corrupted by real-world motion artifacts. The proposed method leverages the temporal information from the distorted signal and its preceding data points to obtain the clean PPG signal. A GAN-based model is trained to generate succeeding clean PPG signals by previous data points. A sliding window moving at a fixed step on the noisy signal is used to select and update the input for the trained model by the information within the noisy signal. A PPG dataset collected by smartwatches in a health monitoring study is used to train, validate, and test the method in this study. A noisy dataset generated with real-world motion artifacts of different noise levels and lengths is used to evaluate the proposed and baseline methods. Three state-of-the-art PPG reconstruction methods are compared with our method. Two metrics, including maximum peak-to-peak error and RMSSD error, are extracted from the original and reconstructed signals to estimate the reconstruction error for HR and HRV.
Our method outperforms state-of-the-art methods with the lowest values of the two evaluation matrices at all noise levels and lengths. The proposed method achieves 0.689, 1.352 and 1.821 seconds of maximum peak-to-peak errors for 5-second, 10-second, and 15-second noise at the highest noise level, respectively, and achieves 0.021, 0.048 and 0.067 seconds of RMSSD errors for the same noise cases.
Consequently, our method performs the best in reconstructing distorted PPG signals and provides reliable estimation for both HR and HRV.

Keywords: Motion artifacts removal, Photoplethysmography reconstruction, Generative Adversarial Network

Contents

List of Figures

List of Tables

List of acronyms

ADC Analog-to-Digital Conversion

ANS Autonomic Nervous System

AR Auto-regressive

BRDAE Bidirectional Recurrent Denoising Auto-encoder

CNN Convolutional Neural Network

CycleGAN Cycle Generative Adversarial Network

DC-GAN Deep Convolutional Generative Adversarial Network

DWT Discrete Wavelet Transform

ECG Electrocardiogram

EMD Empirical Mode Decomposition

HR Heart Rate

HRV Heart Rate Variability

IMF Intrinsic Mode Function

LED Light emitting diode

LSTM Long Short-Term Memory

MA Motion Artifact

PPG Photoplethysmography

PSD Power Spectral Density

RAM Random-Access Memory

ReLU Rectified Linear Units

RMSSD Root Mean Square of Successive Differences between Heart Beat Interval

RNN Recurrent Neural Network

ROM Read-Only Memory

SNR Signal-to-Noise Ratio

SVM Support Vector Machine

1 Introduction

Photoplethysmography (PPG) is a non-invasive measuring method to reflect the blood volume change of the pulse wave. The method uses a light emitting diode (LED) to send light to a tissue and a light detector to receive the changing amount of light transmission from tissue such as the surface of wrist, finger, forehead and earlobe caused by blood flow [1]. Smartwatches apply the technique to collect PPG signals from the wrist of the watch users. Various healthcare applications using PPG signals have been proposed, focusing on sleep monitoring, blood pressure monitoring, and cardiovascular diseases. [2]–[4].

However, the PPG signals can be easily affected by noise during data collection. Such noises caused by the external environment and user's motion artifacts are commonly seen in PPG signals. The external environment may distort the collected PPG. For example, the ambient light from a room light with variable frequency causes saturation noise to PPG signals [5]. The motion artifacts caused by the daily activities of the subject may also be collected together with PPG by a wearable device. PPG signals with these noises may result in unreliable decision making.

The quality of the PPG signal is indicated by the signal-to-noise ratio (SNR), which is the ratio between signal power and noise power. The larger the SNR value is, the less the PPG signal is distorted. Although spectral analysis and different filtering methods can reduce the effect of most of the noises in PPG signals, the motion artifacts, which have overlapped frequency-domain features with clean PPG signals,

are especially difficult to remove. Therefore, an accurate PPG signal reconstruction method to eliminate the effect of motion artifacts is necessarily required for studies working with PPG signals.

Many studies have been proposed to detect the motion artifacts in PPG signals. Various time and frequency-domain features are extracted to classify whether the signal is clean or not. The noisy segment is then excluded from further analysis [6]. However, these methods can not produce continuous PPG signals or provide reliable estimation for certain health-related parameters, such as heart rate variability.

Signal components analysis methods are proposed to improve the PPG signal quality. These methods decompose the noisy PPG signals into noise components and clean components based on mathematical algorithms. Then, the noise components are removed, and the methods reconstruct the PPG with only clean components provided by the algorithms to improve the signal quality [7]–[11].

Time-series based methods are also applied for PPG signal reconstruction. These methods regard the PPG signal as time-series data, and train the model with the clean signal before the distorted PPG signal. Then, the trained model predicts the succeeding data points based on the previous clean PPG signals to perform the reconstruction for PPG signal [12]. However, these methods fail to reconstruct signals with low SNR values. The reconstruction results from these methods are also highly dependent on the quality of proceeding signals.

In addition, deep-learning based methods are recently applied to reconstruct the signals. These models are trained to produce the clean PPG signals as output [13], [14]. However, the models are only developed for certain types of noises.

As a result, we believe that a robust PPG signal reconstruction method is required to regenerate the distorted PPG signals without prior knowledge about the noise such as type, noise level, or length of noise.

1.1 Research questions

This study, therefore, intends to focus on the following two research questions in motion artifacts removal in PPG signals.

1. Can the distorted PPG signals be reconstructed accurately, especially when SNRs are low?

2. Is the information in the distorted signal and the proceeding points enough to reconstruct the PPG signal?

One of the important issues in reconstructing noisy PPG signals is to deal with signals with low SNR. The waveform of such signals at high noise levels is largely corrupted. Hence, no useful information can be extracted from the fiducial points such as peaks and valleys to reconstruct clean PPG.

Another issue is that real-world motion artifacts can be caused by various types of activities in daily life. Without any other signal as a reference for motion artifacts, the motion artifacts in the noisy PPG signals do not have a certain representative feature, which makes it hard to be distinguished from clean signals. The only available information is the proceeding PPG data points and the distorted segment itself.

Therefore, we focus on the solutions to these two research questions in this study.

1.2 Contributions

In this thesis, we propose a deep convolutional GAN-based method to reconstruct noisy PPG signals distorted by motion artifacts. The proposed method leverages the information from both the noisy segment itself and its proceeding signals. The PPG data used in the study is collected by smartwatches. The model is trained with clean PPG signals and tested with PPG signals at different noise levels. Our method is evaluated by the noisy dataset generated by real-world motion artifacts

with different noise duration and SNRs. Finally, we compare the proposed method with the other three state-of-the-art methods by extracting types of errors that represent the reconstruction error of heart rate and heart rate variability. The major contributions of the thesis can be summarized as follows:

1. A GAN-based method is proposed to reconstruct noisy PPG signals distorted by real-world motion artifacts using the information within the signal and proceeding points.

2. A test dataset containing distorted PPG signals with different noise lengths at noise levels with SNR values ranging from -25 dB to 30 dB is generated to evaluate the performance of the proposed method, compared to the existing three state-of-the-art methods.

3. Two metrics representing the reconstruction error of HR and HRV are extracted to evaluate the performance of proposed method, compared to three existing state-of-the-art methods.

This thesis is related to the accepted publication [15].

1.3 Thesis content summary

The rest of the thesis is organized as follows. Chapter 3 reviews the related work on PPG noise cancellation. Chapter 2 introduces the background of PPG and the used model. Chapter 4 provides the details of the collection, participant recruitment, and the wearable device for the dataset used in this study. The data processing, model architecture, and the reconstruction method is presented in Chapter 5. Chapter 6 presents the implementation of the other three state-of-art methods to compare with our proposed method. The results and discussion of the signal reconstruction from all the methods are introduced in 7. Chapter 8 is the conclusion.

2 Background

In this chapter, we introduce the PPG and GAN.

2.1 Photoplethysmography

Photoplethysmography (PPG) is a non-invasive optical technique that reflects the change of blood flow by measuring the amount of light absorption or reflection [16]. When the heart beats, the blood flow in the vessels changes correspondingly. The tissue on the blood vessels has different light absorption or reflectivity rate for the same wavelength with or without blood flows. The technique uses a light emitter and a light sensor to detect the light change caused by cardiovascular activity on the skin.

A typical PPG waveform is shown in Figure 2.1. One cycle of a typical PPG waveform includes systolic peak, dicrotic norch, and diastolic peak. The rising edge of the waveform reflects the systole, and the amplitude of the systolic peak is related to stroke volume [17]. The falling edge of the waveform is the diastole and wave reflections. The dicrotic notch is caused by the closure of heart valves and the reverse flow of blood, which is commonly seen in subjects with healthy heart valves [18].

The PPG signals can be easily collected using various wearable devices, including smart rings and smartwatches. As it is easy to implement with low cost, PPG has been frequently used in different health monitoring tasks. The cyclic change of the PPG waveform is able to measure various physiological parameters, such as

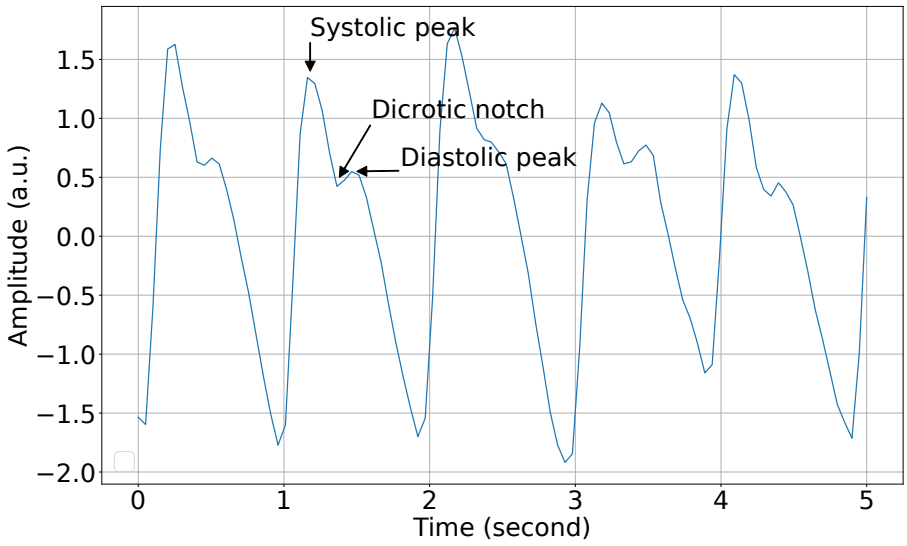

Figure 2.1: Typical waveform of PPG signals.

heart rate (HR), respiration rate [19], heart rate variability (HRV)[20], blood oxygen saturation [21], blood pressure [22] and ankle pressure [23].

2.1.1 Noise in Photoplethysmography

The raw PPG signals are usually noisy. The noises are mainly caused by the environment, measuring devices, and body movements of participants. We describe three types of noise, including white noise, baseline wander and motion artifacts, which are commonly seen in PPG signals.

White noise White noise is the type of noise with a uniform distribution. In other words, the white noise has a zero mean and standard deviation of 1 [24].

Baseline wander Baseline wander is a low-frequency artifact usually caused by subjects' movements, such as respiration. The baseline wander drifts PPG signals baseline up or down over time [25].

Motion artifact The motion artifact (MA) is the noise caused by the movement of subjects. Although the PPG signals are widely used in daily monitoring devices, the signal is vulnerable to noises caused by participants' daily activities such as running, waving hands, and finger tapping. Such artifacts corrupt the PPG signals, thus leading to inaccurate estimation for the health-related parameters.

Figure 2.2 shows an example of a 10-second clean PPG signal and a PPG signal corrupted by motion artifact. Figure 2.2 (a) and (c) are the waveforms, and Figure 2.2 (b) and (d) illustrate the corresponding power spectral density (PSD). The PSD is the power of the signal at the corresponding frequency. As indicated in Figure 2.2(b), the frequency with the largest power in the clean PPG signal is 1.5 Hz, which corresponds to the HR. However, the HR frequency in Figure 2.2(d) is not clear due to the MAs. Therefore, the MAs cannot be removed by frequency-based filters, because the frequencies of clean signals and artifacts overlap.

2.1.2 Heart rate variability

Heart rate variability (HRV) refers to the change between the successive heartbeat [26]. It is caused by the interaction between the brain and heart, which is regulated by the non-linear activity of the autonomic nervous system (ANS) [27]. The HRV is able to reflect neurocardiac health. Previous studies have proved that HRV is correlated with stress, panic, and worry [28].

The root mean square of successive differences between normal heartbeats (RMSSD) is one of the important HRV parameters, which reflects the differences between successive heartbeat interval. It can estimate HRV mediated by vagal activities [27]. The reduced RMSSD level is highly correlated with the risk of sudden unexplained death in epilepsy [29].

Many studies have been carried out to monitor the daily HRV of participants. Due to the simplicity and low cost, PPG-based wearable devices are used in such

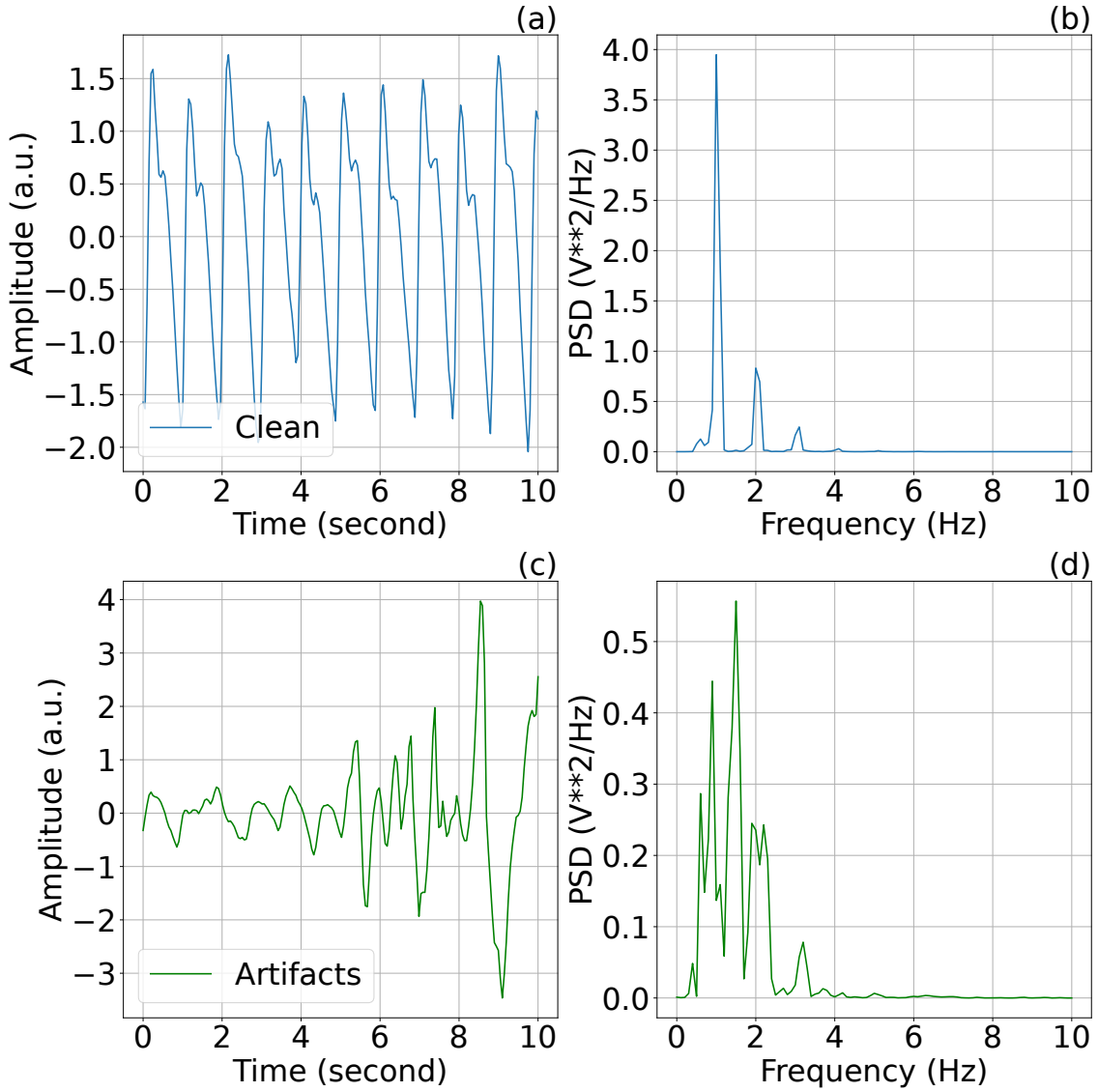

Figure 2.2: Waveform and power spectral density from 10-second PPG segment and artifact. The clean PPG and artifacts are in blue and green. (a) Clean PPG waveform, (b) clean PPG PSD, (c) motion artifact waveform, and (d) motion artifact PSD.

health-monitoring studies [30]–[32]. It is also easy to extract different HRV parameters from the PPG waveform. Therefore, the quality of PPG signals has a large effect on HRV parameters.

2.2 Generative adversarial network

The Generative adversarial network (GAN) is a network structure proposed by Goodfellow et al. to generate synthetic data [33]. Figure 2.3 illustrates the standard structure of GAN. The model consists of two sub networks, i.e., a generator and a discriminator. The generator learns the distribution of the input data and reconstructs the input. The discriminator classifies whether the data is original or generated by the generator. The two sub networks compete with each other during training until the difference between the original and generated data is minimized. Therefore, GAN can be used to generate and reconstruct the data. Different studies have applied such feature to de-noise images in computer vision tasks. [34], [35].

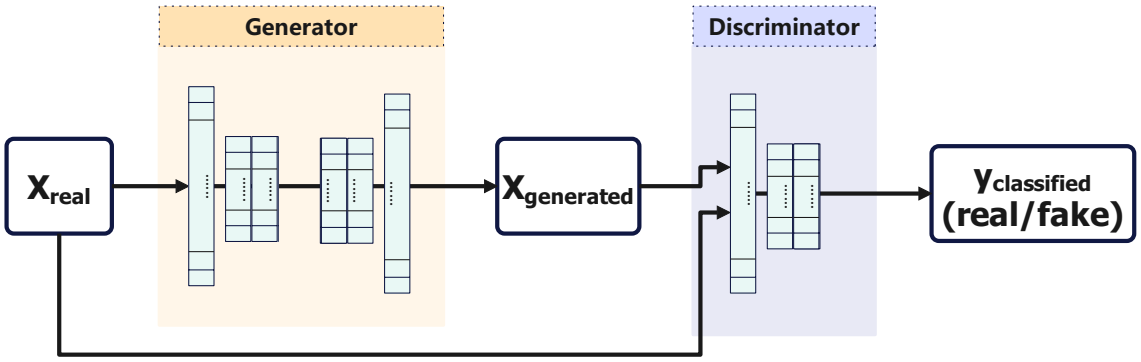

Figure 2.3: The architecture of Generative Adversarial Network

2.2.1 Deep Convolutional Generative Adversarial Network

Although GAN can maximize the likelihood by the design of competition between generator and discriminator, the training of GAN is often unstable, which results in nonsensical synthesized data [36]–[38]. The deep convolutional generative adversarial network (DC-GAN) is proposed to solve the problem [39]. DC-GAN is a GAN-based neural network with modified convolutional structure. Compared to

the initial GAN, the DC-GAN adopted three latest changes in convolutional neural networks:

- Using stride convolutional layers to downsample the data instead of max pooling.
- Excluding fully-connected hidden layers in deep network achitecture.
- Using Batch Normalization between layers.

The DC-GAN architecture has been proved to be more stable and effective in learning the representation of data [39]. Compared to GAN, DC-GAN is more used to infer data based on contextual information, such as image colorization [40] or inpainting corrupted images [41].

3 Related work

Noises distort the waveform of PPG signals, making it difficult to extract accurate health-related parameters for variable-based analysis. Methods to reduce the effect of noise are needed. However, PPG de-noising is a challenging task, especially with high noise level and long noise duration. In this chapter, we describe previous studies dealing with the MAs in PPG signals. To tackle the problem, previous studies can be divided into 2 categories: (1) signal quality assessment methods, which give a classification result of whether a PPG segment is clean or not, and (2) noise removal methods, which remove the noise component and produce de-noised signals.

3.1 Quality assessment methods

The quality assessment methods classify the PPG signals as noisy or clean by extracting various features in the time or frequency domain. For instance, in the study from [6], the authors calculated the statistic values, including kurtosis, skewness, and standard deviation of each segment, and detected the motion artifacts by the statistic threshold set from clean signals. Likewise, in [42], other features such as variation in approximate entropy, Shannon entropy, and spectral entropy are extracted to further classify the signal as clean or not by the elliptical envelope algorithm.

Recently, deep-learning-based anomaly detection methods are also applied to assess the PPG signal quality. In deep learning-based methods, the model itself acts

as the feature extractor and consequently relieves the work of feature engineering. The study conducted by [43] proposed a 1-dimensional convolution neural network to detect whether each 5-second PPG signal segment was distorted by motion artifacts. The PPG signals were blindly sectioned to 5-second segments regardless of any peak or valley information. Then the segments were normalized and fed into the 13-layer 1-D convolutional neural network (CNN) model to output the "clean" or "artifact" labels. Similarly, in [44], the authors implemented a bi-directional recurrent neural network-based auto-encoder to detect and remove motion artifacts in an unsupervised manner. The methods mentioned above can accurately evaluate the PPG signal quality. However, such methods only differentiated the PPG signals as clean or noisy, but did not improve the signal quality.

3.2 Noise removal methods

Classifying the signal as good or bad is insufficient to obtain the noise-free PPG signal, as excluding the noisy part in the signal would cause loss of collected data. Therefore, in addition to the signal quality assessment methods, many studies are proposed to eliminate the noise components in signals by analyzing different time and frequency features in reference and PPG signals.

3.2.1 Reference-signal-based method

Reference-signal-based methods use another reference signal which is recorded simultaneously as the PPG signals to provide information about either noise or clean PPG. Based on different design of data collection, the collected simultaneous signal can either be the reference of noise or clean PPG signals. The authors in [7][8] designed adaptive filters with accelerometers and gyroscopes signals as references of noise to filter out the motion artifacts in noisy PPG signals. Masinelli et al. used

the Electrocardiogram (ECG) signals as a reference of ground true HR to remove motion artifacts by spectral analysis [45]. Despite the improved HR estimation, these reference signal-based methods still suffer from the limitations of the inaccurate detection of reference sensors. Moreover, extra sensors such as accelerator and ECG sensors are needed for the research, which makes the setup more complicated.

3.2.2 Signal-component-analysis methods

Signal components analysis methods are applied to reduce the noise in PPG signals. These methods decompose the noisy signal into noise and clean components, and then remove the noise component in the signal to obtain the reconstructed clean signal. In [9], the noisy PPG signals were first decomposed into a sum of Intrinsic Mode Function (IMF) by Empirical Mode Decomposition(EMD) to improve the signal quality and then reduce the noise by Discrete Wavelet Transform (DWT). In [10] , the authors set the upper and lower envelope boundaries to minimize the MAs in the signals. However, these methods fail to reconstruct the signals when the noise level is high.

3.2.3 Time-series methods

Time-series methods are applied to estimate clean PPG signals while the signal is noisy. These methods take the predecidual samples into account during reconstruction since the physiological activities are time-causal processes. In [12], the combination of a Kalman filter and Auto-regressive (AR) model was introduced to predict the value on the time $t + 1$ using the values on and before time t, hence avoiding the motion artifacts caused by sudden subjects' movements. In [46], the authors used the fiducial points such as peak, foot, diastolic peak, and diastolic notch from a previous clean PPG segment to reconstruct a new one by cubic spline interpolation. However, these methods highly rely on the amount and quality of

proceeding data to reconstruct the signal.

3.2.4 Deep-learning-based methods

Deep learning models are also developed to reconstruct PPG signals. The study from [13] first transform 1-D PPG signals to its representative 2-D images. Then, a Cycle Generative Adversarial Network (CycleGAN) was trained to de-noise and reconstruct images. Finally, the de-noised image was transformed back to clean PPG signals. Similarly, in [14], a bidirectional recurrent auto-encoder was trained by minimizing the difference between synthesized noisy PPG signals and their original clean signals. Then, the trained model was able to produce clean PPG signals when the input signals were noisy. Although deep learning models have achieved good performance on large datasets, these deep-learning-based methods only considered the noise caused by certain types of movement, such as waving and shaking hands. Moreover, these methods were tested by manually generated noisy PPG dataset. The noisy signals in the test set were synthesized by adding the noises to clean signals, which guaranteed that the noisy segment must contain clean PPG information. Hence, these methods may fail when clean PPG signals are faded in the corrupted signals in real life.

The methods using either quality assessment or noise removal methods are insufficient to eliminate the effect of motion artifacts on PPG signals. The mentioned methods are limited to either short-term analysis of signals or certain types of noises caused by specified activities. Consequently, a method to reconstruct the PPG waveform distorted by real-world motion artifacts without any prior knowledge of the noise is needed.

4 Dataset

In this chapter, we introduce the dataset used to train, validate and test the model in this thesis.

The PPG data used in this thesis is a part of a health monitoring study conducted by the Health Technology Group at the University of Turku [47]. The PPG data was collected in Southwest Finland for one month, from July to August 2019. The participants were monitored without any limitation of activities, which allows to collect real-world signals in daily life.

There are 46 participants included in the study. The participants were asked to wear the wearable device on their non-dominant hands 24 hours for 7 days. The recruited participants meet the following requirements:

- Being able to use the wearable device for 24 hours.
- No cardiovascular diseases.
- No symptoms of other diseases during monitoring.
- No restricted physical activities.

Table 4.1 summarizes the background information of all the participants in the study [48].

The wearable device used in the study is a Samsung Gear Sport smartwatch [49]. The size of each smartwatch is $44.6 \times 42.9 \times 11.6mm$ in height, width, and depth with a 30.2 mm main display. The watch is embedded with Wi-Fi at 802.11 b/g/n

Table 4.1: Background information of participants

Parameters Values		**Values**
Age, years, mean (SD)		32.5 (6.6)
BMI, mean (SD)		25.4 (5.2)
Exercise	Almost daily	8
	A few times a week	13
	Once a week or fewer	6
Education level	Primary school	1
	High school	6
	College	6
	University	14
Employment status	Working	22
	Student	1
	Unemployed	3
	Other	1

2.4GHz and Bluetooth v4.2. It is water resistant with the pressure of 5 atmospheres or 50 meters under water. The internal memory of the smartwatch is 0.75 GB of random-access memory (RAM), and 4 GB of read-only memory (ROM). The watch is programmed by the Tizen operating system and utilizes sensors, including accelerator, gyroscope, and PPG sensor.

With our customized data collection application, 16-minute continuous PPG and accelerator/gyroscope signals are collected every half an hour at sampling frequency of 20 Hz. The first minute of the collected signals is removed because the time is for device calibration. The setup enables the battery life of the smartwatches to be longer than 24 hours.

The internal memory of the smartwatches is sufficient for storing the collected

data. The collected data is first stored in the watch's internal storage. After the monitoring, the data is sent to our cloud service for storage.

5 Proposed method

In this chapter, we introduce our GAN-based proposed method, including the model architecture, loss functions, training process, and reconstruction method with trained model. We also describe the preparation of training, validation, and test PPG datasets, which are used for model training, hyperparameter selection, and evaluation of the method performance, respectively.

5.1 Proposed model architecture

In our study, we develop a deep convolutional GAN-based model to reconstruct the noisy signals distorted by MAs. Previous studies showed that GAN is a powerful approach to de-noise images[34], [50]. We apply such a feature to our study to train a model by clean PPG signals and then use the trained model to denoise PPG signals distorted by MAs.

Our GAN-based model consists of two sub networks: a generator and a discriminator. The generator learns to generate synthesized clean PPG signals from their previous data points, while the discriminator learns to distinguish whether the signals are original input or synthesized by the generator. The two sub-networks compete with each other during training, and consequently the model learns how to synthesize clean PPG signals by the information from previous data.

Inspired by [51], we adopt a similar structure to GANomaly, which uses an Auto-encoder architecture for the generator and an encoder architecture for the discrim-

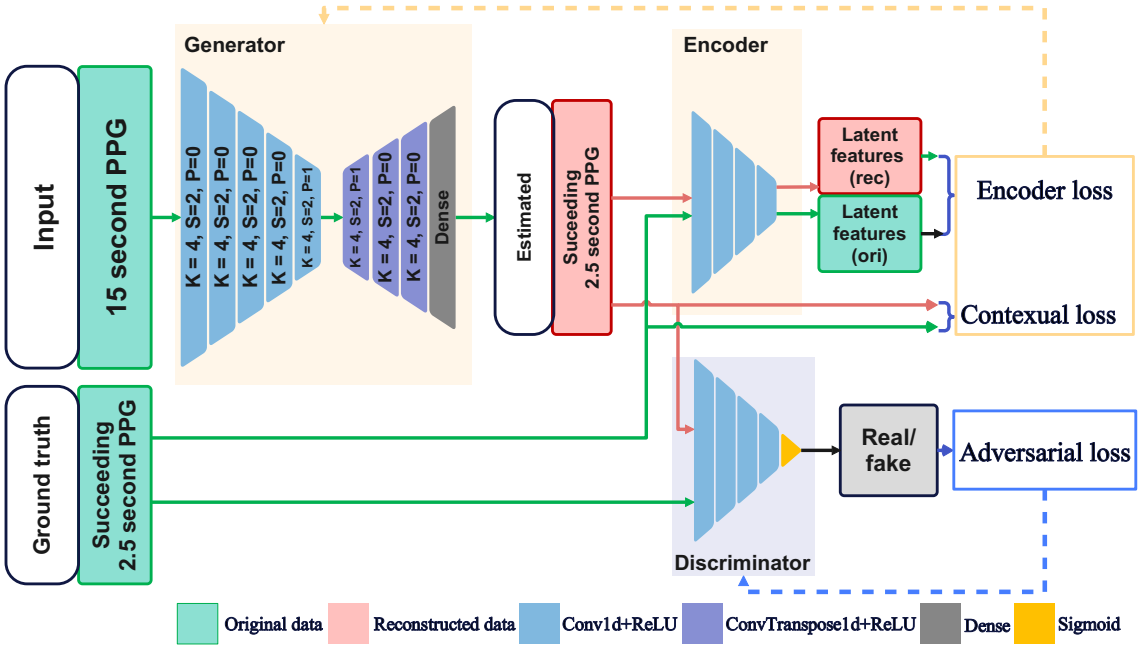

Figure 5.1: The architecture of proposed model. K, S, and P represent the kernel size, stride and padding parameters in convolutional layers respectively

inator. The deep convolutional layers are modified and applied to our model. The architecture of our proposed model is illustrated by Figure 5.1. In each epoch, the model takes 15-second clean PPG signal as input, and then the generator generates its succeeding 2.5 second PPG signal as output. The original succeeding 2.5-second PPG signal is the ground truth. The generator is optimized by minimizing the difference between the estimated succeeding PPG and its ground truth. The discriminator is updated by minimizing the classification error of whether the signal is the original ground truth or estimated by the generator. The K, S, and P in Figure 5.1 represent the kernel size, stride, and padding parameters in convolutional layers, respectively.

Generator The generator follows an Auto-encoder structure, i.e., an encoder and a decoder. The generator is trained to generate synthetic succeeding clean signals.

The encoder stacks 5 1D-convolutional layers to compress the PPG input to its representative features in latent space. The decoder stacks 3 transposed 1D-convolution layers and flattens its output by one dense layer to re-generate the PPG signal by the bottle-neck features.

Encoder The generator is then followed by another encoder. The encoder acts as a feature extraction part of the generator, which stacks 4 1D-convolution layers to downscale both the estimated signal from the generator and the real PPG signal to find their representative features in latent space.

Discriminator The discriminator, aiming to distinguish whether the signal is real or synthesized by the generator, uses the same encoder architecture to compress both reconstructed signals and original input to their corresponding representative features. Then, the Sigmoid layer classifies the features to values "1" or "0", representing the true and fake label.

5.1.1 Loss functions

To optimize each sub-network in the pipeline, we use the same three loss functions i.e., the encoder loss, contextual loss, and adversarial loss, as in [51] to optimize the model during training.

Encoder loss The encoder loss L_{enc} is the difference between the representative features compressed from synthesized and real PPG signals by the generator encoder mentioned in Section 5.1. The loss is calculated as:

$$J(fg, fo) = \frac{1}{256} \sum_{i=1}^{256} |fg_i - fo_i|^2 \tag{5.1}$$

where fg and fo are the latent features of generated and original PPG signal, respectively, and i is the index in a batch of samples with a size of 256. The value is backpropagated to update the generator in each iteration.

Contextual loss The contextual loss L_{con} is the difference between the original and the estimated succeeding signal from the generator in 5.1. The loss is calculated as:

$$J(sg, so) = \frac{1}{256} \sum_{i=1}^{256} |sg_i - so_i| \tag{5.2}$$

where sg and so are the generated signal and its original ground truth, respectively. The value is also backpropagated to update the generator each iteration.

Adversarial loss The adversarial loss is the standard discriminator loss. Binary cross-entropy is used as the loss function to calculate the difference between the actual and classification labels (i.e., real/fake) obtained by the discriminator. The value is backpropagated to optimize the discriminator each iteration.

In this model, except for the last layer of the discriminator which uses Sigmoid to give a binary output, all the other activation functions are Rectified Linear Units (ReLU). The batch normalization is performed between layers with dropout rate of 0.5 to avoid overfitting. The weights for networks are randomly initialized, and Adam is chosen as the optimizer.

5.1.2 Reconstruction method

In this thesis, we propose a sliding window method to reconstruct the noisy PPG signals with the trained model. As the real-world distorted PPG signals are not guaranteed to contain clean PPG signals, our proposed method uses a sliding window reconstruction method to leverage the information from both noisy signals and its proceeding clean signals.

Figure 5.2 indicates the reconstruction process of proposed method. A sliding window with a fixed length is used to select input and feed the signals to the generator. The first iteration starts at only clean signal before the distorted part. The generator takes the window of input and gives a prediction of succeeding data points.

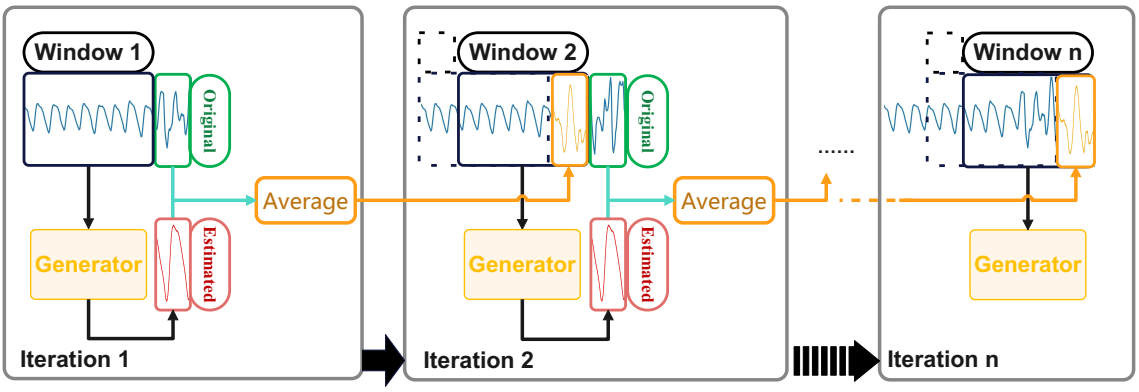

Figure 5.2: The reconstruction process of proposed method

Then the signal is updated by averaging the prediction and the original noise. In the next iteration, the window moves forward with a step, and repeats the process of making a prediction, updating the signal by average value, and moving the window to feed updated signals as input to the trained model. The iterations continues until the window sweeps to the end of the segment.

In our case, the sliding window length is 15 seconds, and the shift size of each step is 2.5 seconds to ensure each prediction at least contains one cardiovascular cycle.

5.2 Model training

In this section, we represent the training process of our proposed GAN-based model. The model is only trained with clean PPG signals. Thus the model only learns the distribution of clean PPG signals during training, and the trained parameters are not suitable for generating other signals than clean PPG signals. For example, if the input contains clean PPG and other noise at the same time, and the generator encoder compresses both clean PPG and noise to their representative bottleneck features, the trained generator decoder with its parametrization, can not reconstruct

the noise from the bottleneck feature of noise. Therefore, training with only clean signals enables the model to produce clean PPG signals from the noisy segment.

The model calculates the three types of loss mentioned in Section 5.1.1 and back-propagates them to optimize the generator and discriminator of the model during training. The model performing best on the validation set is selected to reconstruct noisy signal in the testing phase. Note that for the reconstruction method in this thesis, z-score standardization is applied to each PPG segment before feeding it into models.

5.2.1 Training and validation data preparation

We preprocess the raw PPG signals before feeding the data into our proposed model. Figure 5.3 is the block diagram, which illustrates all the steps of the preparation for the training and validation dataset. We introduce each preprocessing step shown in Figure 5.3 in the remainder of this section.

5.2.2 Segmentation and filtering

We use the PPG data collected in the study described in Chapter 4 to train and validate the model. The dataset contains 15-minute raw PPG signals from 46 participants, and the data from 37 of them form the training and validation dataset.

We first section the 15-minute raw PPG signals into 30-second segments to keep consistent with the input size of our signal quality assessment method. Then, we implement a 4-order high-pass Butterworth filter with a cutoff frequency of 0.5 Hz to remove baseline wander. Figure 5.4 indicates one of the filtering results after the high-pass Butterworth filter. Figure 5.4 (a) is one of the 30-second segment before filtering, while Figure 5.4 (b) is the segment after filtering. As indicated in Figure 5.4, the baseline of the raw PPG signal drifts up at 16 seconds, while after applying the high-pass filter, the baseline becomes stable throughout the 30-second segment.

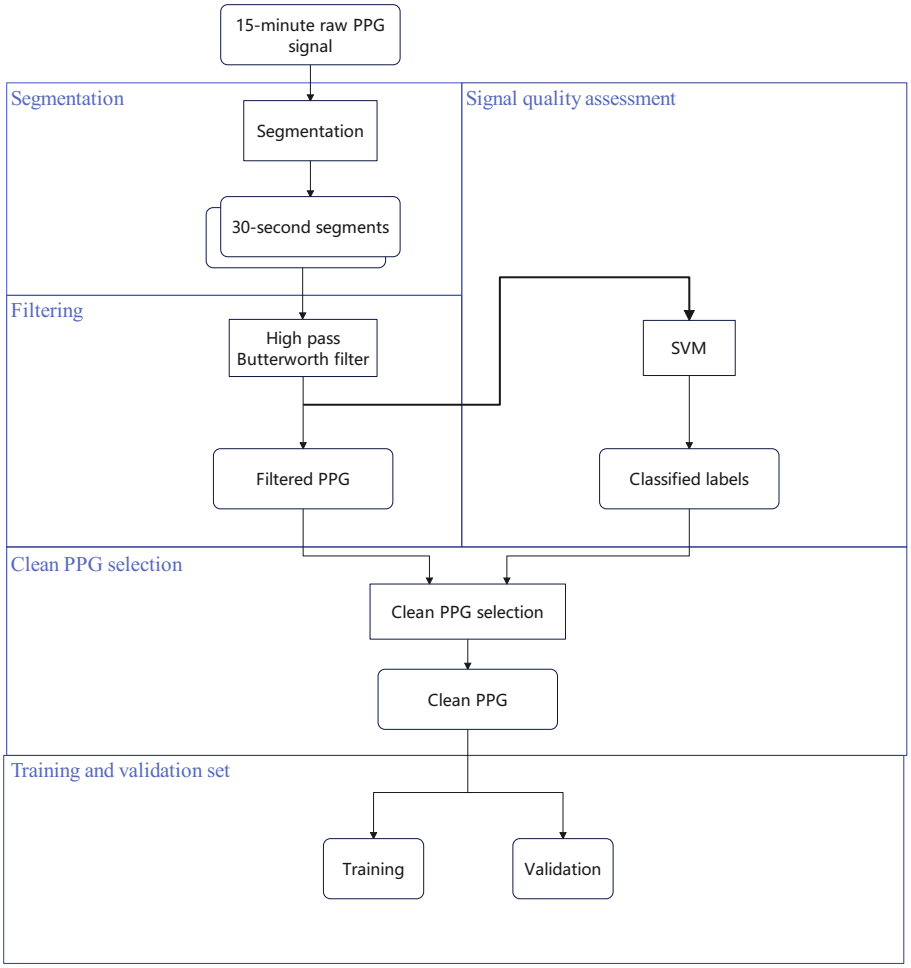

Figure 5.3: Block diagram of training data preparation

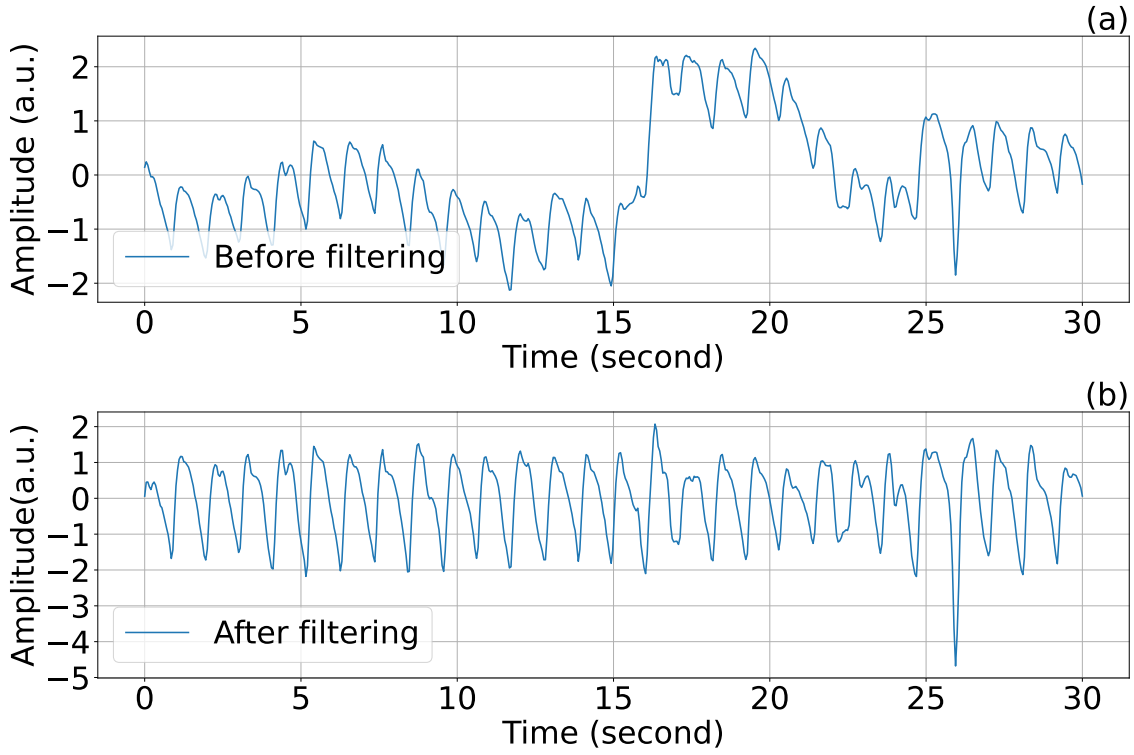

Figure 5.4: PPG signal (a) before filtering, and (b) after filtering

The filter can remove the unrelated low-frequency components and baseline wander.

5.2.3 Signal quality assessment

We used a Support Vector Machine (SVM) to assess the PPG signal quality in this thesis. The assessment technique takes 30-second PPG signal as input, and produces a classification result of whether the segment is clean or noisy. The assessment method is from [42], which extracts five features from the PPG waveform, including spectral entropy, Shannon entropy, approximate entropy, kurtosis, and skewness from each 30-second PPG. Then these features are fed into an SVM to classify the clean signals and noise.

Figure 5.5 indicates one of the classification results from SVM. (a) is one of the 30-second segments classified as clean, while (b) as noisy by the SVM. The SVM

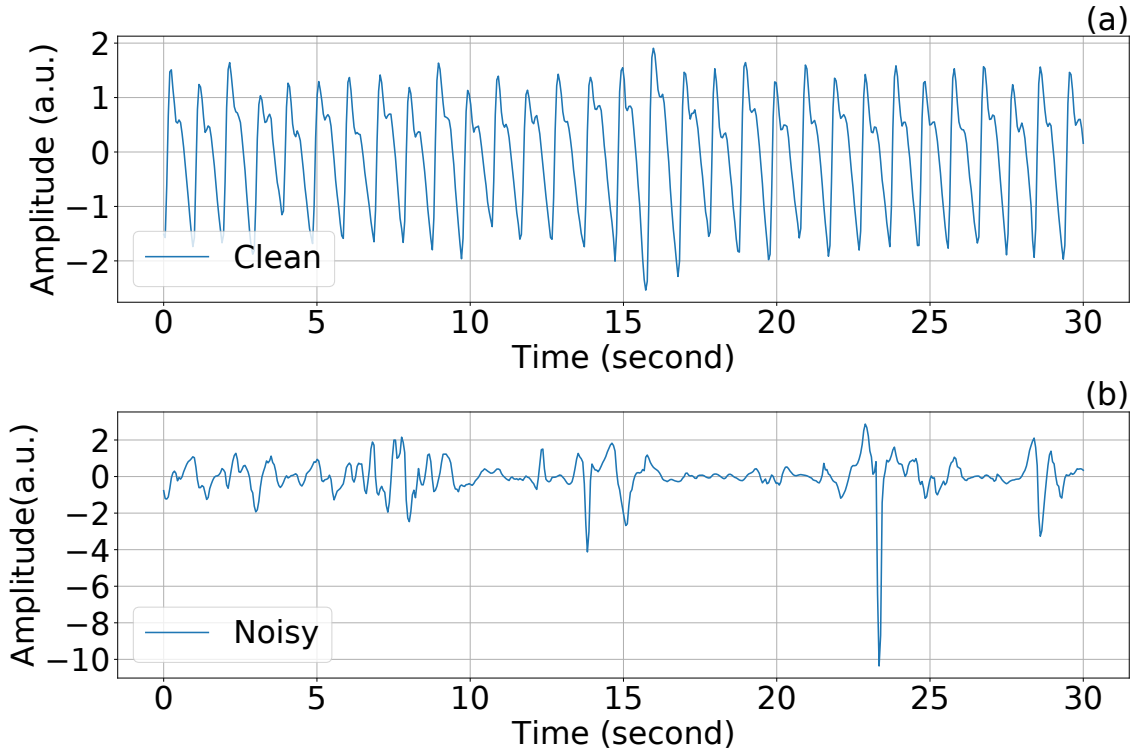

Figure 5.5: PPG signal classified as (a) clean segment, and (b) noisy segment by SVM.

can effectively assess the quality of PPG signals.

Finally, we used the classified labels to select only clean PPG signals from all the filtered 30-second signals. The obtained clean PPG signals are used as the data for the training and validation set.

The clean PPG data from 27 participants of the study is used for training, and the clean PPG from the other 10 participants forms the validation dataset. In total, 61117 30-second clean PPG segments are used to train the model, and 19661 segments for the validation set to select hyperparameters for the model.

5.3 Test data preparation

We synthesize a test dataset containing noisy PPG signals distorted by real-world MAs. The test dataset is generated to evaluate the method's performance on different noise levels and noise lengths. In this section, we introduce the preparation of the test dataset used in the study.

The steps of preparing the test dataset are shown in Figure 5.6. The preprocessing steps are similar to the preparation of the training and validation dataset. The raw PPG signals are first sectioned into 30-second segments. Then, the sectioned segments are filtered by a high-pass filter with cut-off frequency of 0.5 Hz to remove baseline wander. The same signal quality assessment method is used to produce the clean or noisy label for 30-second PPG segments. The classified clean and noisy segments are put into our noisy signal generation function to synthesize noisy PPG signals with required noise levels.

5.3.1 Noisy signal generation

The PPG data from the remaining 9 participants is used to synthesize the noisy PPG test set at different noise levels and duration. A noisy signal generation method is used in this thesis to synthesize the noisy test dataset. The noisy signals with different SNR values are needed for evaluating the reconstruction results from methods. The noisy signals are generated by the clean data and MAs mentioned in Section 5.2.3

In the thesis, we used a generate function to integrate clean signals and MAs at different noise levels. The noisy generation process follows the Algorithm 5.3.1. The explicit steps are introduced as follows.

Clean PPG signal selection: Randomly select a clean 30-second segment from the 9 individuals (S).

Noise selection: Randomly select a noise segment (N) from the MA dataset from

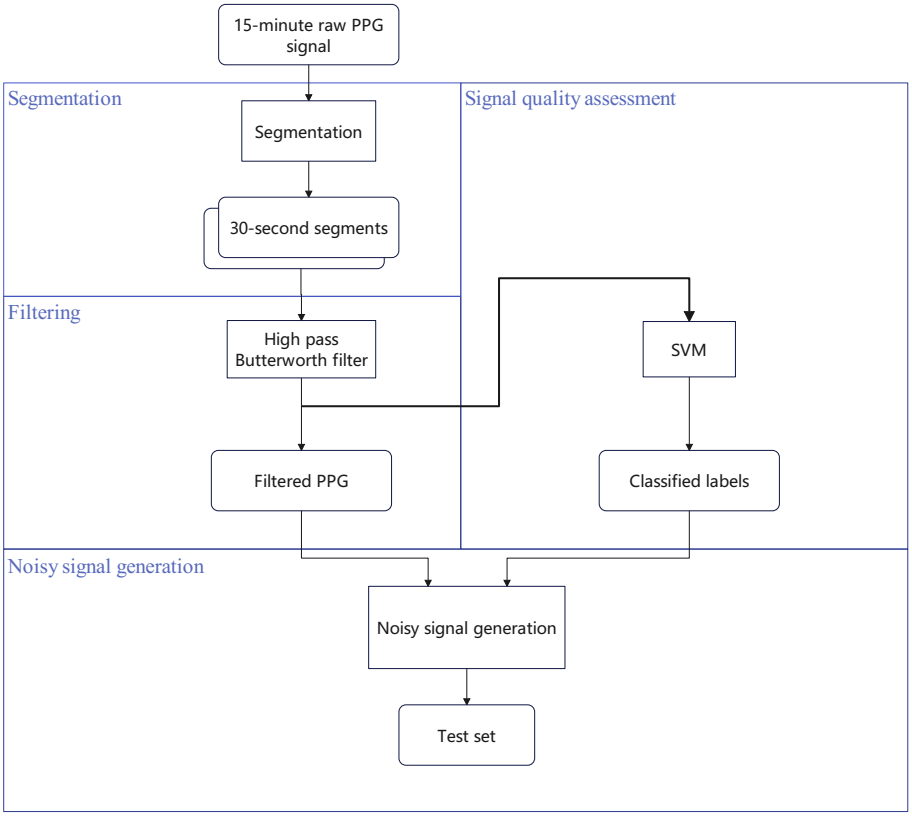

Figure 5.6: Block diagram of test data preparation

Listing 1 The generator function

Initialize:

win_size ← window size

batch_size ← number of batch size

while $j <$ batch_size **do**

$S \leftarrow$ a window of the clean PPG signal is chosen randomly

$N \leftarrow$ select a noise frame randomly

$w_N \leftarrow$ a random number with uniform distribution (0,16)

$X \leftarrow S + w_N N$

SNR ← calculate the SNR

$j += 1$

end while

the 9 individuals. The MA dataset contains MAs with three noise lengths, including 5-second, 10-second, and 15-second motion artifacts. The noise segments are selected accordingly with these three different lengths.

Noisy PPG generation: A noisy PPG signal is calculated by a weighted arithmetic mean:

$$X = w_S S + w_N N \tag{5.3}$$

where w_S and w_N are the weights of the clean signals and noise, respectively. In this thesis, w_S equals to 1, while w_N is a random generated number which follows a uniform distribution ranging from 0 to 16. Note that we select a window corresponding to the noise length (i.e. 5-second, 10-second, and 15-second) at the end of each segment to add noises.

SNR extraction: We then calcuate the SNR values for each generated noisy PPG signal using the following Equation 5.4:

$$SNR = 10 \log \frac{P_{Signal}}{P_{Noise}} \tag{5.4}$$

where P_{Signal} is the power of clean signal, and P_{Noise} is the power of the selected motion artifacts.

The generated noisy 5-second, 10-second, and 15-second PPG signals with SNRs ranging from -25 dB to 30 dB are used as the data for the test set. In total, the generate function in Section 5.3.1 generates 16384 30-second segments for each noise length.

6 State-of-the-art methods

This chapter introduces the state-of-the-art PPG reconstruction methods that we use as baseline. Three PPG noise removal methods are selected in this thesis. All the baseline methods do not need any reference signals to keep consistent with our proposed method, and each of them is selected to represent the corresponding method category, including a time time-series, a signal decomposition, and another deep-learning based method mentioned in Chapter 3. The time-series method uses the proceeding samples to predict the distorted part of the signal, where we select the combination of Kalman filter and AR model proposed in [12] to represent the method of this category. The signal decomposition methods analyze the noise and clean PPG components in the noisy signal, and only exclude the noise component to reduce the effect of motion artifacts in corrupted signals. In this regard, we select the EMD and DWT method proposed in [9] as a representation of reconstruction methods in this category. Deep-learning based method develop different models to reconstruct PPG signals. Other than our proposed GAN-based method, we select the Bidirectional Long Short-Term Memory (LSTM) method proposed in [14] to represent the performance of the deep-learning models with other network architecture, which only focus on certain types of noises.

6.1 Kalman filter and Auto regression model

We select the Kalman filter and AR model from [12]. The method uses the signals before noise as training data to predict the succeeding data. The Kalman filter and AR model is selected as the state-of-the-art method to represent the time-series method.

The Kalman filter is a filter that leverages the observation from the past to estimate unknown variables. The estimation is calculated by providing solutions with the least-squares method [52]. The Kalman filter in the selected method is used to update the coefficients for the AR model.

The Auto regressive (AR) model is a prediction method on time-series data, which holds the assumption that the variable of interest is linearly correlated with its past values. It predicts the value of the variable on time t by values of the variable in the past from time 0 to time $t-1$ [53]. The AR model in the selected method is implemented to predict the succeeding PPG data points. The combination of Kalman filter and AR model works together to optimize the prediction over time by all the data before the current time.

In our case, the parameters for the Kalman filter follow the same values in [12], that is, the noise covariance matrix of is a normal distribution with the values of mean and standard deviation equaling to 0 and 0.2, respectively. The order p for the AR model is obtained by AIC criterion.

6.2 Empirical Mode Decomposition and Discrete Wavelet Transform

The method that combines EMD and DWT to reconstruct PPG signals is from [9]. The EMD and DWT method is chosen to represent the signal decomposition method for PPG signal reconstruction.

The EMD is a signal decomposition method for non-stationary signals. The method decomposes the signal into a set of components, namely Intrinsic Mode Function (IMF) [54], which only have one zero between successive extrema and a zero local mean. In the selected state-of-the-art method, the EMD first decomposes the signal to different levels of IMF components. Then the most relevant component is chosen to improve the signal quality.

The DWT is also a signal decomposition method that maps the signal in time domain to frequency domain [55]. A mother wavelet is used to decompose the signal into a set of high-frequency and low-frequency wavelets. At each level, the signals are filtered by high-pass and low-pass filters to obtain the detail coefficients (i.e. the high-frequency coefficients) and the approximation coefficients (i.e. the low-frequency coefficients). The signals are then reconstructed by a linear combination of the wavelet functions weighted by this set of wavelet coefficients. In the selected method, the DWT takes the output component from the EMD, and then uses the output IMF as input of DWT to further reduce the noise.

Same parameters in [9] are used in this thesis, that is, to choose the first IMF from EMD as the input for DWT and use Daubechies 4 with decomposition level of 5 as the mother wavelet for DWT.

6.3 Bidirectional Long Short-Term Memory Auto-encoder

The bidirectional recurrent denoising auto-encoder (BRDAE) is selected from [14]. It represents the state of the art of other deep-learning based methods.

The bidirectional recurrent neural network (RNN) is a deep-learning based method commonly used in tasks dealing with sequential data. Compared with the feed-forward network, RNN considers the time dependencies between the previous and

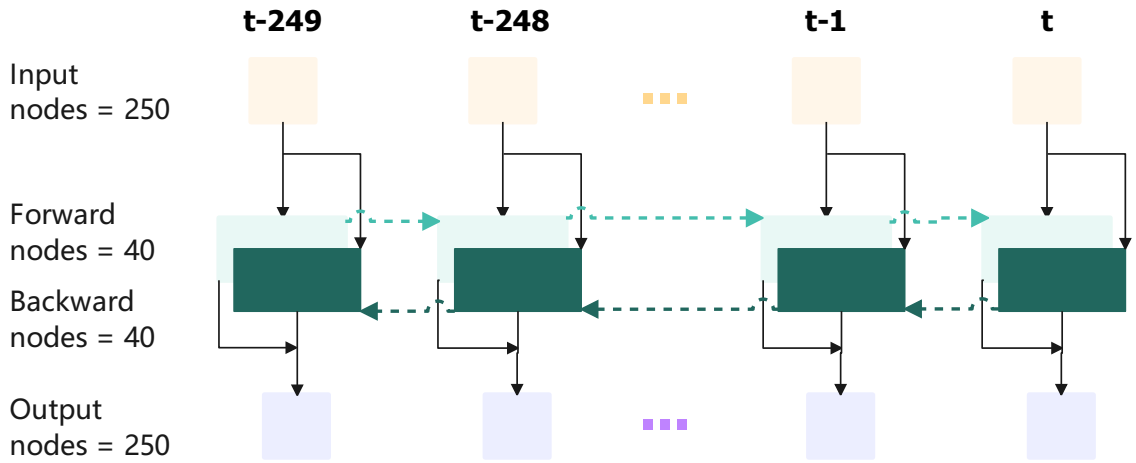

Figure 6.1: The architecture of Bidirectional LSTM method

current hidden state. Therefore, the RNN is able to store the history information of data. In the selected method, a bidirectional Long Short-Term Memory (LSTM) layer is applied to the Auto-Encoder model to extract time dependencies in PPG signals.

We follow the same BRDAE architecture in [14] (see Figure 6.1). The model takes 250 nodes as input and uses a 40-node bidirectional LSTM layer as hidden layer. The bottleneck feature is then reconstructed to an output size of 250 nodes again. The root mean square error between reconstructed signal and true values is calculated and backpropagated to the model as the loss of each training epoch. Adam is selected as the optimizer with a learning rate of 0.001.

The training data for the model is augmented by three types of noise, including white noise, slope noise, and saturation noise. They are randomly added to the clean PPG signal to train the model.

The white noise is the random noise with zero mean and normal distribution described in Chapter 2.1.1. The noisy signal with white noise X^{white} can be synthesized using the following equation:

$$X_i^{white} = S_i + \frac{1}{3} \cdot N(0,1) \tag{6.1}$$

where, S_i is the i^{th} sample of the original clean signal, and $N(0,1)$ indicates the normally distributed white noise with mean and standard deviation equaling to 0 and 1, respectively. Figure 6.2 shows an example of the synthesized noisy PPG signal with white noise (in orange), and its clean ground truth (in blue).

The slope noise is the noise that causes sloping shift to the baseline of signals. The noisy signal with slope noise X^{slope} can be synthesized using the following equation:

$$X_i^{slope} = S_i + \frac{4}{N} \cdot i \cdot U(-1,1) \tag{6.2}$$

where S_i is the i^{th} sample of the original clean signal, and i is the index of the sample. N is the total sample number in the signal segment. U(-1, 1) represents a random number that follows a uniform distribution between -1 and 1. 4 is considered as the highest slope for the synthesized signal, which is the same value from [14].

The saturation noise represents the distorted PPG signals below or above the analog-to-digital conversion (ADC) range during measurement. The noisy signal with saturation noise $X^{saturation}$ can be synthesized using the following equation:

$$\begin{aligned} X_i^{saturation} &= 0 or 1, \quad if x_1 \leq i \leq x_2 \\ X_i^{saturation} &= S_i, \quad otherwise \end{aligned} \tag{6.3}$$

where x_1 is the starting index of the saturation noise, which is a random number selected from a uniform distribution between 0 and total sample number N. x_2 is the end index of the saturation noise, which is dependent on the value of x_1. The x_1 and x_2 should meet the requirement that $X_2 - X_1 \leq 20\% \cdot N$ because the maximum duration of saturation noise should be less than 20% of the total signal noise. The 20% value is chosen to keep consistent with the training details in [14].

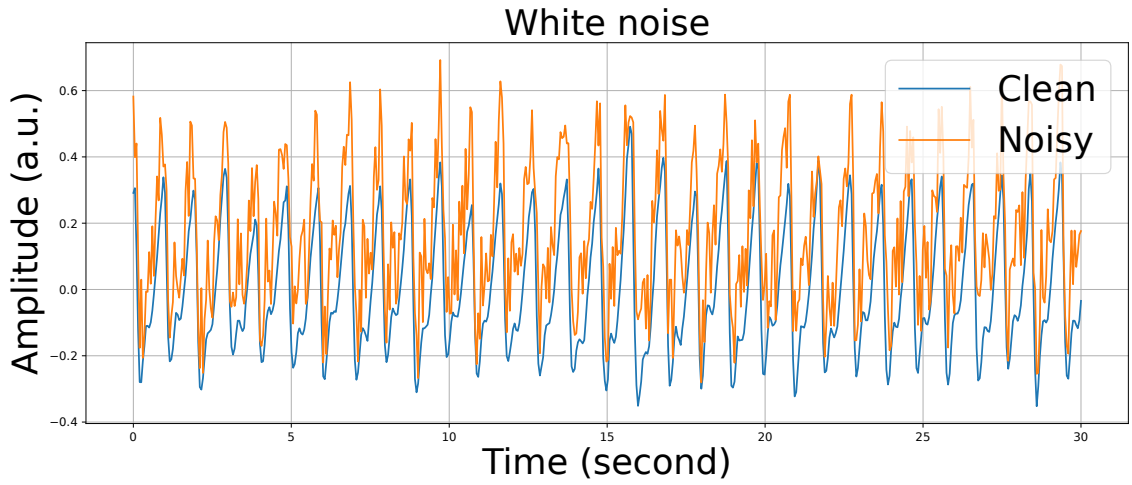

Figure 6.2: An example of synthesized PPG signal with white noise (orange) and ground truth (blue)

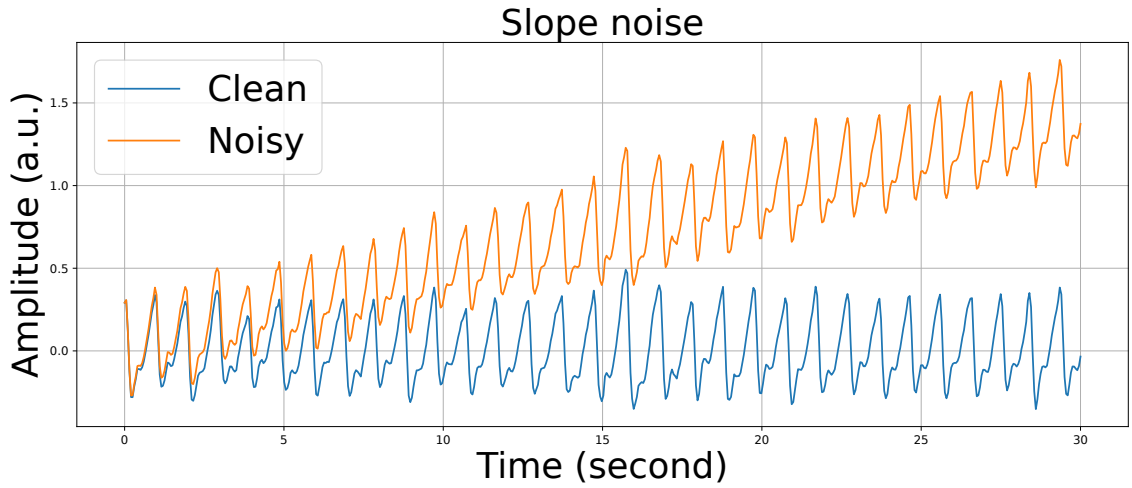

Figure 6.3: An example of synthesized PPG signal with slope noise (orange) and ground truth (blue)

The model is trained by minimizing the differences between the synthesized noisy signal and its original clean signal. In testing, the final trained model takes the noisy signals as input and then outputs the reconstructed clean signal.

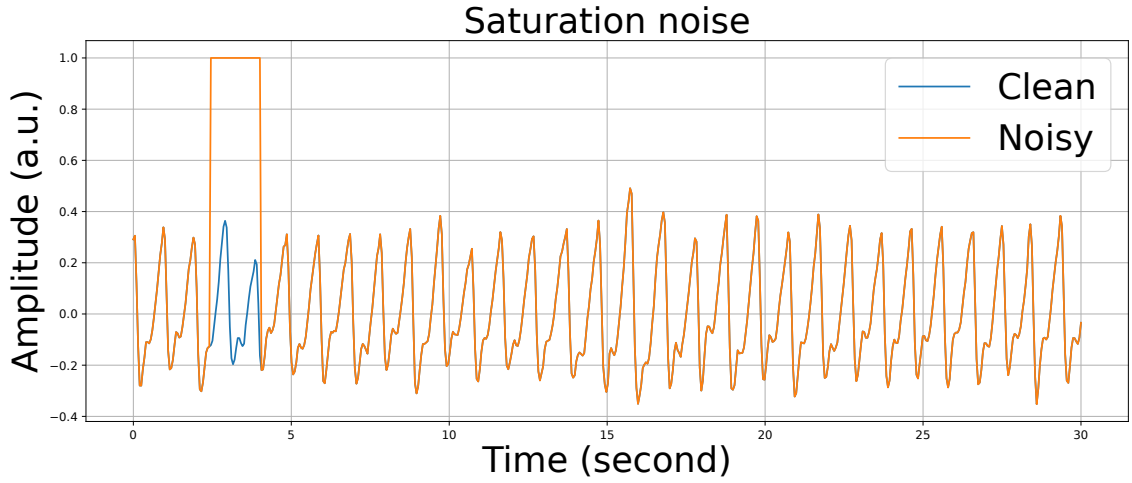

Figure 6.4: An example of synthesized PPG signal with saturation noise (orange) and ground truth (blue)

7 Results and discussion

In this chapter, we present the reconstruction results obtained by the proposed and the selected three state-of-the-art methods. All the methods are evaluated by the test dataset obtained in Section 5.3.1, which contains noisy PPG signals distorted by 5-second, 10-second, and 15-second real-world motion artifacts. In each noise length group, the SNR values of the signals range from -25 dB to 30 dB.

We introduce the two metrics, i.e., the maximum peak-to-peak error and the RMSSD error, which are used to evaluate the method performance.

7.1 Setup

The proposed method and the Bidirectional LSTM method are implemented by PyTorch [56]. The Kalman and AR methods are implemented by pykalman [57], statsmodel [58] package. The EMD and DWT method uses pyEMD [59], and PyWavelets [60] package in Python.

In the thesis, the training process of the proposed method runs 1000 epochs in total, and a batch size of 256 is used for each epoch. The Bidirectional LSTM method runs 2000 epochs for training with a batch size of 250 samples, which is the same as the process proposed in [14]. For both the proposed and Bidirectional LSTM method, the model that performs best on the validation set is chosen as the final model for testing.

For experiments from all the methods, we used a Linux machine with AMD

Ryzen Threadripper 2920x 12-Core processor, NVIDIA TITAN RTX GPU (24 GB memory), and 126 GB RAM.

7.2 Performance evaluation

In this section, we introduce the two metrics used for the evaluation of method performance. Two types of errors: maximum peak-to-peak error and RMSSD error, are extracted from original and reconstructed signals, which represent the HR and HRV errors respectively, to evaluate the reconstruction results.

7.2.1 Maximum peak-to-peak error

The peaks in the original signal and the reconstructed signals are first paired in time order. The pair with the maximum distance from each other is the maximum peak-to-peak error. The maximum peak-to-peak error is used to estimate the heart rate error between the reconstructed and original signals. The error is calculated as:

$$Error_{peak-to-peak} = max(|Peaks_{reconstructed} - Peaks_{original}|) \tag{7.1}$$

7.2.2 RMSSD error

The root mean square of the successive peak interval difference (RMSSD) is an important HRV parameter[61]. The RMSSD error is used to estimate the HRV error between the reconstructed and original signals. The RMSSD error between reconstructed signal and ground truth is calculated as:

$$Error_{RMSSD} = |RMSSD_{reconstructed} - RMSSD_{original}| \tag{7.2}$$

7.3 Reconstruction results

The model performance is evaluated by three lengths of noise, including 5-second, 10-second, and 15-second. The three types of noise duration represent PPG signal distorted by short to long MAs. The reconstruction errors, including maximum peak-to-peak and RMSSD error of signals at different noise levels and lengths of noise, are presented in this section.

An example of the reconstructed PPG waveform distorted by 5-second noise at a high noise level (i.e. SNR = -8.093 dB) is illustrated in Figure 7.1. The blue, orange and red lines represent the ground truth, synthesized noisy signal, and the reconstructed signal respectively. As shown in Figure 7.1 (b), our proposed method performs the best at reconstructing PPG peaks and waveforms compared to the ground truth in blue.

7.3.1 5-second noise

The noisy signals distorted by 5-second noise reflect the impact of short-term MA on PPG signals. The performance of all the methods on PPG signals distorted by 5-second noise in different SNR groups is indicated in Figure 7.2. Figure 7.2 (a) shows the maximum peak-to-peak error, and Figure 7.2 (b) is the RMSSD error. The Kalman and AR, EMD and DWT, bidirectional LSTM, and the proposed method are indicated by blue, yellow, green and red lines, respectively. The quantitative values are shown in Table 7.1.

As indicated in Figure 7.2 (a), our method outperforms other baseline methods in all SNR groups with the lowest peak-to-peak error. For example, when SNR is at -25 dB, the maximum peak-to-peak error from the proposed method is 0.689 seconds, while the error from Kalman and AR method is 2.948 seconds, the EMD and DWT is 1.131 seconds, and the Bidirectional LSTM is 1.053 seconds. This indicates that the proposed method is able to estimate reliable HR with PPG distorted by 5-second

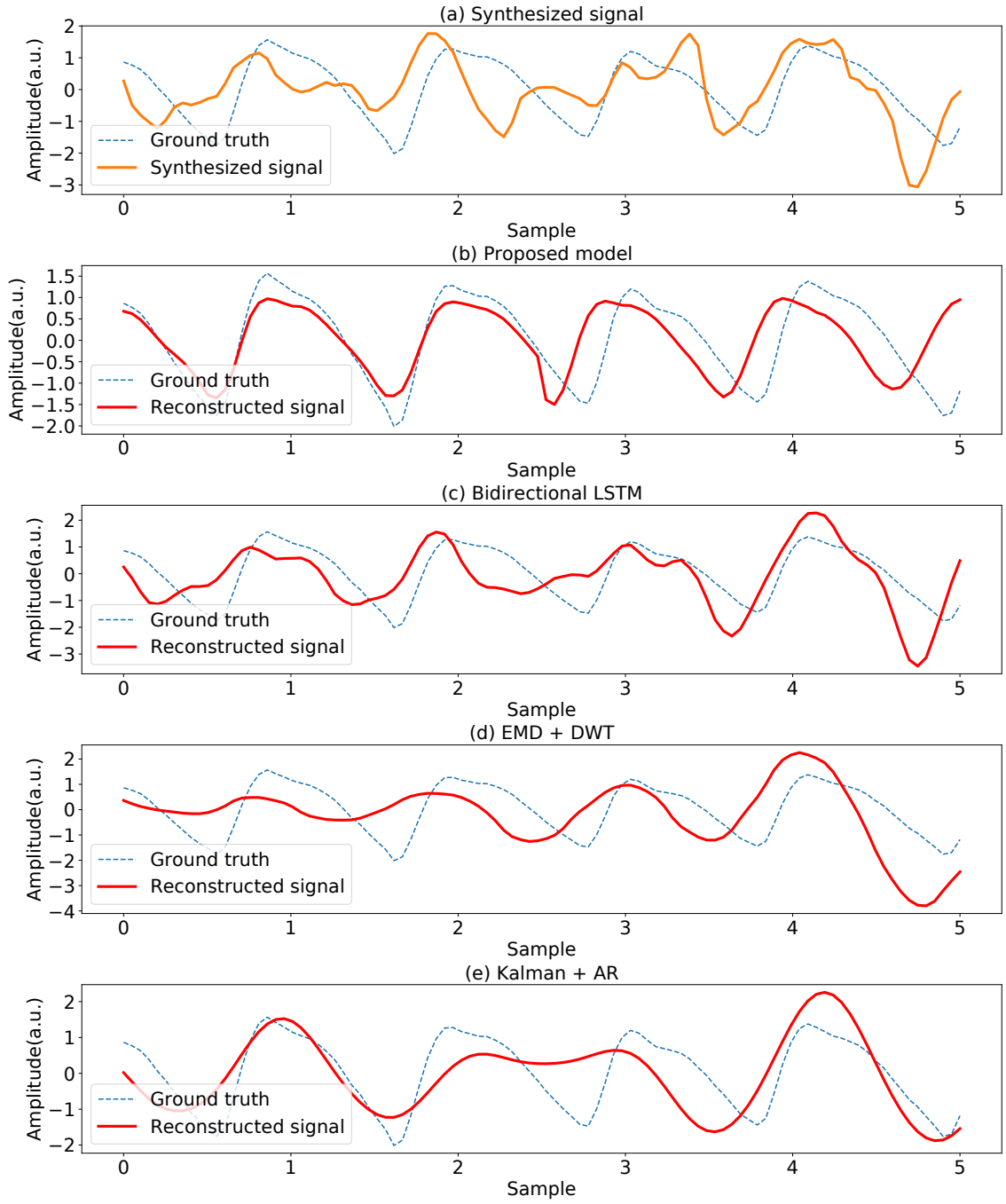

Figure 7.1: Reconstruction results from proposed method and baseline methods (SNR=-8.09 dB). Blue dotted line is the clean ground truth.(a) Synthesized signal, (b) Proposed method, (c) Bidirectional LSTM, (d) EMD and DWT, and (e) Kalman and AR.

Table 7.1: Performance of all methods on 5-second noise

SNR (dB)	Proposed method		Kalman + AR		EMD + DWT		Bidirectional LSTM	
	P-to-P	RMSSD	P-to-P	RMSSD	P-to-P	RMSSD	P-to-P	RMSSD
-25	0.689	0.021	2.948	0.109	1.131	1.526	1.053	0.045
-20	0.599	0.018	2.045	0.080	1.090	1.095	1.052	0.045
-15	0.533	0.014	1.672	0.061	1.083	0.687	0.997	0.043
-10	0.407	0.011	1.342	0.041	1.046	0.442	0.926	0.038
-5	0.312	0.009	1.200	0.033	0.939	0.256	0.828	0.034
0	0.264	0.007	1.155	0.030	0.835	0.200	0.691	0.028
5	0.220	0.006	1.142	0.029	0.691	0.147	0.595	0.023
10	0.211	0.006	1.094	0.028	0.613	0.129	0.548	0.020
15	0.224	0.006	1.066	0.026	0.633	0.143	0.515	0.019
20	0.215	0.006	1.098	0.028	0.562	0.134	0.504	0.019
25	0.222	0.006	1.066	0.027	0.599	0.146	0.52	0.019
30	0.212	0.006	1.120	0.027	0.591	0.138	0.489	0.018

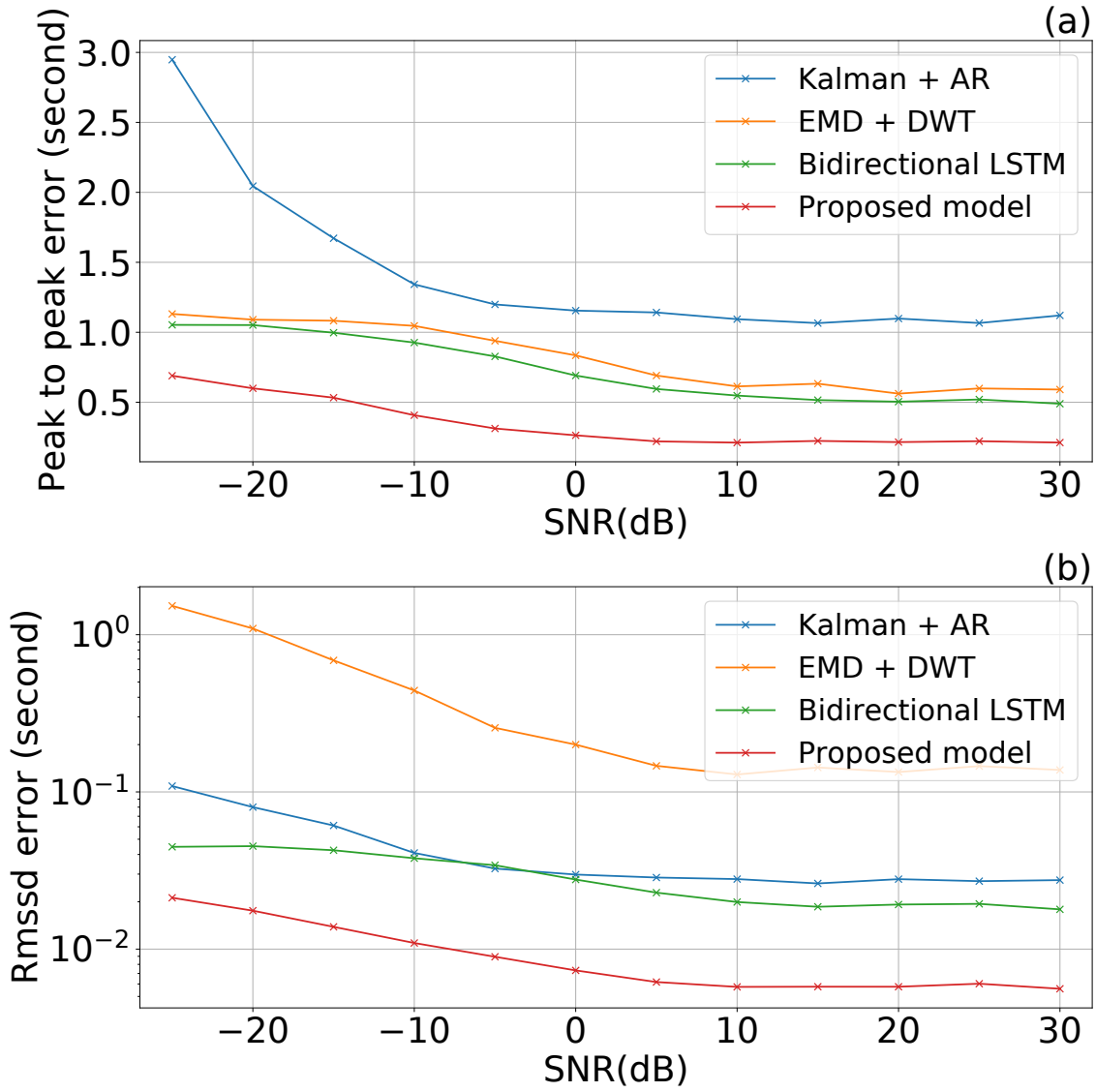

Figure 7.2: Reconstruction errors from all methods with 5-second noise. (a) Maximum peak-to-peak error, and (b) RMSSD error.

noise.

Our method also obtains the lowest RMSSD error at all noise levels. As indicated in Figure 7.2 (b), when SNR is at -25 dB, the RMSSD error from proposed method is 0.021 seconds. However, the error from Kalman and AR method is 0.109 seconds, the EMD and DWT is 1.526 seconds, and the Bidirectional LSTM is 0.045 seconds.

The results show that the proposed method can also produce reliable estimation for HRV with PPG distorted by 5-second noise.

Moreover, the proposed method is more robust to low SNR signals. As shown in Figure 7.2, all methods except Kalman and AR obtain a maximum peak-to-peak error less than 1.0 seconds when the SNR is at 30 dB. However, the errors increase when the SNR values decrease. For instance, when the SNR values drop from 30dB to -25dB, the peak-to-peak error from the proposed method only increases from 0.212 to 0.684 seconds, while such increase is higher in Kalman and AR, EMD and DWT, and Bidirectional LSTM method, with values of 1.120 to 2.948 seconds, 0.591 to 1.131 seconds, and 0.489 to 1.053 seconds, respectively.

Similarly, for RMSSD errors, such increase from the proposed method is still the smallest. When SNR decreases from 30 dB to -25 dB, the RMSSD error from proposed method only increases from 0.006 to 0.021 seconds. However, the EMD and DWT, which has the highest RMSSD error, increases from 0.138 to 1.526 seconds. The RMSSD error increases of Kalman and AR method is 0.027 to 0.109 seconds, and Bidirectional LSTM method is 0.018 to 0.045 seconds.

For signals distorted by short-term MAs, the HRV parameters are more affected than the HR parameters. However, Our method obtains the lowest RMSSD errors in all SNR groups. At the same time, when SNR decreases, the increase of RMSSD error from the proposed method is the smallest. In addition, the proposed method also obtains the lowest peak-to-peak error in all SNR groups. Therefore, our method is robust to high noise level, and can produce reliable reconstruction estimation for both HR and HRV for PPG distorted by short-term MAs.

7.3.2 10-second noise

The noisy signals distorted by 10-second noise evaluate the model performance of increased length of MA to PPG signals. The performance of all the methods on PPG

Table 7.2: Performance of all methods on 10-second noise

SNR (dB)	Proposed method		Kalman + AR		EMD + DWT		Bidirectional LSTM	
	P-to-P	RMSSD	P-to-P	RMSSD	P-to-P	RMSSD	P-to-P	RMSSD
-25	1.352	0.048	4.965	0.282	2.385	2.540	2.041	0.082
-20	1.187	0.039	4.368	0.228	2.340	2.214	1.986	0.080
-15	0.986	0.030	3.413	0.150	2.368	1.687	1.879	0.075
-10	0.790	0.023	2.796	0.098	2.164	0.994	1.816	0.070
-5	0.610	0.017	2.358	0.071	1.897	0.510	1.558	0.059
0	0.448	0.012	2.193	0.060	1.595	0.280	1.239	0.046
5	0.365	0.009	2.159	0.058	1.357	0.186	1.011	0.035
10	0.355	0.009	2.224	0.059	1.209	0.166	0.878	0.030
15	0.319	0.008	2.079	0.055	1.084	0.160	0.804	0.027
20	0.332	0.008	2.150	0.058	1.091	0.162	0.813	0.028
25	0.305	0.008	2.022	0.054	1.070	0.150	0.812	0.027
30	0.350	0.009	2.164	0.059	1.094	0.168	0.796	0.026

signals distorted by 10-second noise in different SNR groups is indicated in Figure 7.3. Figure 7.3 (a) shows the maximum peak-to-peak errors, and Figure 7.3 (b) is the RMSSD errors from all SNR groups. The Kalman and AR, EMD and DWT, bidirectional LSTM and proposed method are indicated by blue, yellow, green and red lines, respectively. The quantitative values are shown in Table 7.2

As indicated in Figure 7.3 (a), the proposed method still performs the best in all SNR groups compared to other baseline methods when PPG signals are distorted by 10-second noise. The peak-to-peak errors from the proposed method are the lowest at all noise levels. For example, when SNR is at -25 dB, the maximum peak-to-peak error from the proposed method is 1.352 seconds, while the error from Kalman and AR method is 4.965 seconds, the EMD and DWT is 2.385 seconds, and the Bidirectional LSTM is 2.041 seconds. This indicates that the proposed method is able to estimate reliable HR with PPG distorted by 10-second noise.

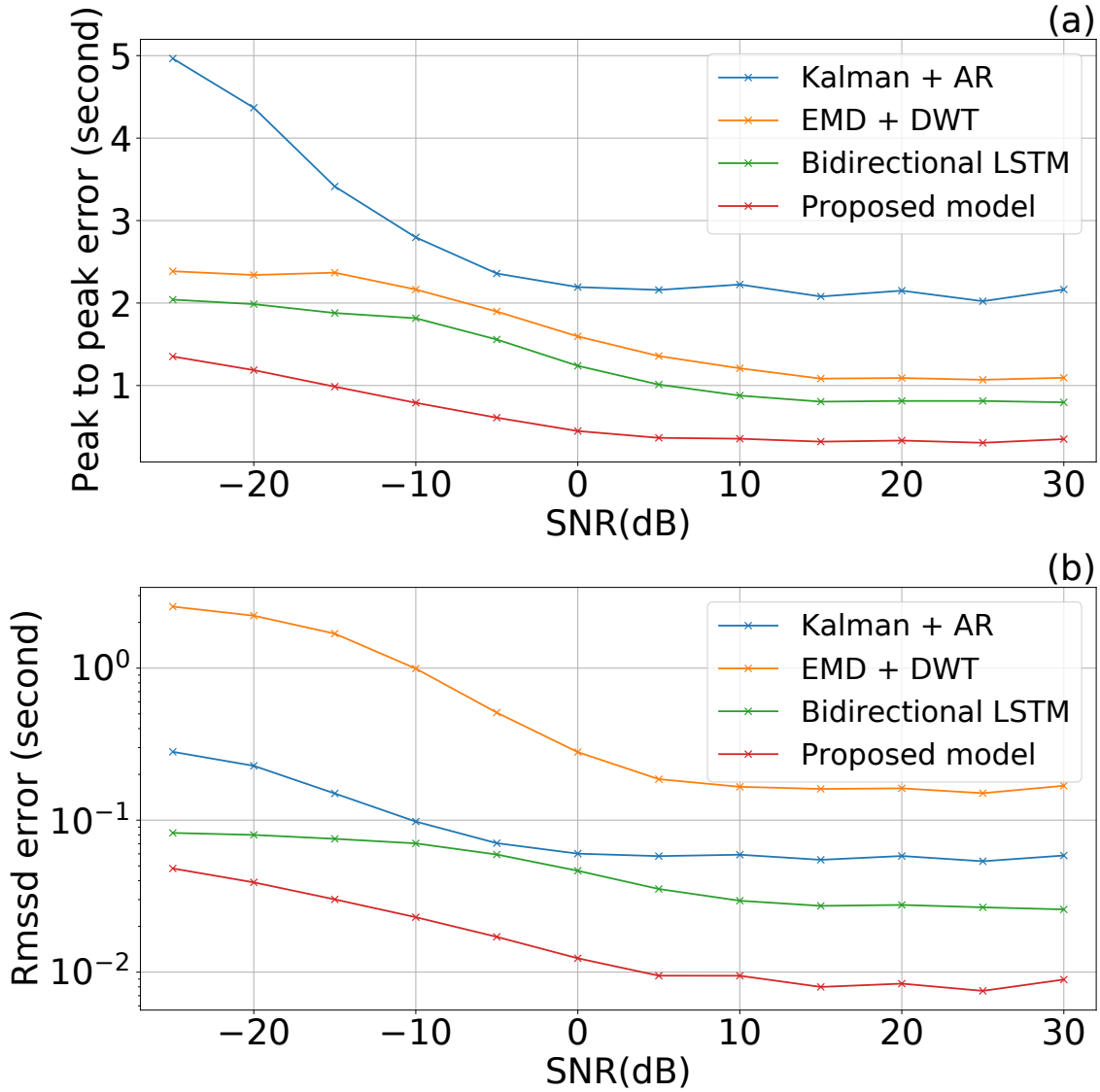

Figure 7.3: Reconstruction errors from all methods with 10-second noise. (a) Maximum peak-to-peak error, and (b) RMSSD error.

The RMSSD errors from our proposed method in 10-second noise case are also the lowest at all noise levels. As indicated in Figure 7.3 (b), when SNR is at -25 dB, the RMSSD error from proposed method is 0.048 seconds. However, the error from Kalman and AR method is 0.282 seconds, the EMD and DWT is 2.540 seconds, and the Bidirectional LSTM is 0.082 seconds. The results show that the proposed method can also produce reliable estimation for HRV with PPG distorted by 10-second noise.

Similar to the 5-second case, the proposed method shows robustness to signals with low SNR values. As shown in Figure 7.3, all methods except Kalman and AR obtain a maximum peak-to-peak error less than 2.0 seconds when the SNR is at 30 dB. However, the errors increase when the SNR values decrease. For instance, when the SNR values drop from 30dB to -25dB, the peak-to-peak error from the proposed method only increases from 0.350 to 1.352 seconds, while such increase is higher in Kalman and AR, EMD and DWT, and Bidirectional LSTM method, with values of 2.164 to 4.965 seconds, 1.094 to 2.385 seconds, and 0.796 to 2.041 seconds, respectively.

For RMSSD errors in 10-second noise length case, such increase from the proposed method is still the smallest. When SNR decreases from 30 dB to -25 dB, the RMSSD error from the proposed method only increases from 0.009 to 0.048 seconds. However, the EMD and DWT increases from 0.168 to 2.540 seconds. The RMSSD error increases of Kalman and AR method is 0.059 to 0.282 seconds, and Bidirectional LSTM method is 0.026 to 0.082 seconds.

When the noise length increases, the effect of noise on both HR and HRV also increases. However, our method still outperforms all state-of-the-art methods with the lowest peak-to-peak and RMSSD errors in all SNR groups. Moreover, when the reconstruction errors increase with the noise level, our method has the smallest increase in the reconstruction error for HR and HRV. Consequently, our method

Table 7.3: Performance of all methods on 15-second noise

SNR (dB)	Proposed method		Kalman + AR		EMD + DWT		Bidirectional LSTM	
	P-to-P	RMSSD	P-to-P	RMSSD	P-to-P	RMSSD	P-to-P	RMSSD
-25	1.821	0.067	7.772	0.567	3.647	1.476	2.848	0.112
-20	1.647	0.059	6.479	0.420	3.655	1.441	2.818	0.109
-15	1.374	0.047	5.303	0.290	3.565	1.202	2.774	0.104
-10	1.019	0.031	4.298	0.173	3.320	0.778	2.513	0.094
-5	0.723	0.021	3.502	0.113	2.975	0.480	2.255	0.082
0	0.499	0.013	3.263	0.094	2.425	0.285	1.727	0.063
5	0.389	0.009	3.367	0.094	1.821	0.170	1.400	0.048
10	0.350	0.008	3.190	0.088	1.693	0.163	1.124	0.036
15	0.324	0.007	3.077	0.086	1.437	0.135	1.076	0.035
20	0.328	0.007	3.086	0.086	1.469	0.158	1.042	0.034
25	0.328	0.007	3.298	0.093	1.530	0.153	0.996	0.033
30	0.312	0.007	3.134	0.088	1.543	0.165	0.998	0.031

is still robust to low SNRs when noise length increases, and can produce reliable estimation for HR and HRV in signals distorted by 10-second MAs.

7.3.3 15-second noise

The noisy signals distorted by 15-second noise are used to evaluate the model performance of PPG distorted by longer MAs (compared to the MAs evaluated earlier). The performance of all the methods on PPG signals distorted by 15-second noise in different SNR groups is indicated in Figure 7.4. Figure 7.4 (a) shows the maximum peak-to-peak, Figure 7.4 (b) is the RMSSD error. The Kalman and AR, EMD and DWT, bidirectional LSTM and the proposed method are indicated by blue, yellow, green and red lines, respectively. The quantitative values are shown in Table 7.3

As indicated in Figure 7.4 (a), the proposed method outperforms other state-of-the-art methods in all SNR groups when PPG signals are distorted by 15-second

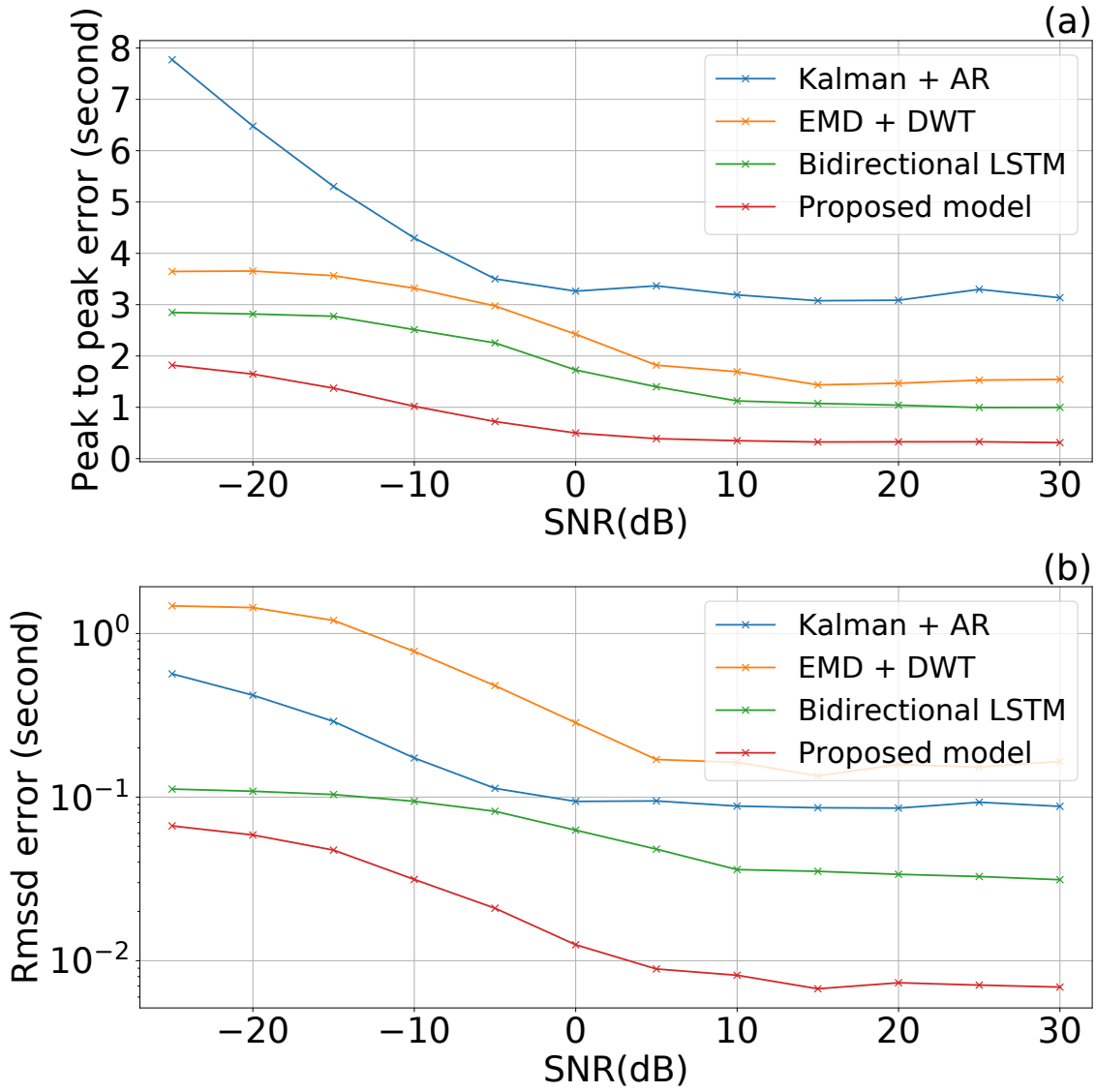

Figure 7.4: Reconstruction errors from all methods with 15-second noise. (a) Maximum peak-to-peak error, and (b) RMSSD error.

noise. Our proposed method achieves the lowest value of peak-to-peak error at all noise levels. For example, when SNR is at -25 dB, the maximum peak-to-peak error from the proposed method is 1.821 seconds, while the error from Kalman and AR method is 7.772 seconds, the EMD and DWT is 3.647 seconds, and the Bidirectional LSTM is 2.848 seconds. This indicates that the proposed method is able to estimate

reliable HR with PPG distorted by 15-second noise.

The proposed method also achieves the lowest value of RMSSD error in all SNR groups. As indicated in Figure 7.4 (b), when SNR is at - 25 dB, the RMSSD error from the proposed method is 0.067 seconds, while the error from Kalman and AR method is 0.567 seconds, the EMD and DWT is 1.476 seconds, and the Bidirectional LSTM is 0.112 seconds. This indicates that the proposed method is able to estimate reliable HRV with PPG distorted by 15-second noise.

Similar to previous noise length cases, the proposed method shows robustness to signals with low SNR values. As shown in Figure 7.3, all methods except Kalman and AR obtain a maximum peak-to-peak error less than 2.0 seconds when the SNR is at 30 dB. However, the errors increase when the SNR values decrease. For instance, when the SNR values drop from 30dB to -25dB, the peak-to-peak error from the proposed method only increases from 0.312 to 1.821 seconds, while such increase is higher in Kalman and AR, EMD and DWT, and Bidirectional LSTM method, with values of 3.134 to 7.772 seconds, 1.543 to 3.647 seconds, and 0.998 to 2.848 seconds, respectively.

For RMSSD errors in signals with 15-second noise case, such increase from the proposed method is still the smallest. When SNR decreases from 30 dB to -25 dB, the RMSSD error from the proposed method only increases from 0.007 to 0.067 seconds. However, the EMD and DWT, increases from 0.165 to 1.476 seconds. The RMSSD error increases of Kalman and AR method is 0.088 to 0.567 seconds, and Bidirectional LSTM method is 0.031 to 0.112 seconds.

When the PPG signals are distorted by long MAs, the MAs have a significant impact on both HR and HRV. However, our method still outperforms the baseline methods with the lowest peak-to-peak and RMSSD errors. In addition, when the peak-to-peak and RMSSD errors increase as the SNR values decrease, our method obtains the smallest value for such increase of error in signals distorted by 15-second

noise. Therefore, our methods show robustness to high noise level and can produce reliable estimation for both HR and HRV in signals distorted by long MAs.

7.4 Impact of Noise Duration

In addition to the robustness of low SNR level discussed above, our method also shows a better performance to longer noise duration for estimating HR and HRV. We take the examples from the highest (30 dB) and lowest (-25 dB) SNR groups to clarify the problem. The maximum peak-to-peak error increases when the noise length increases. With our proposed method, such increase is the smallest compared to all baseline methods.

In -25 dB SNR, the maximum peak-to-peak error from the proposed method increases from 0.689 to 1.821 seconds, when the noise duration increases from 5 seconds to 15 seconds. However, the maximum peak-to-peak error from Kalman and AR method increases from 2.948 to 7.772 seconds, the EMD and DWT increases from 1.131 to 3.647 seconds, and the Bidirectional LSTM increases from 1.053 to 2.848 seconds.

In 30 dB SNR, the maximum peak-to-peak error from the proposed method increases from 0.212 to 0.312 seconds, when the noise duration increases from 5 seconds to 15 seconds. However, the maximum peak-to-peak error from Kalman and AR method increases from 1.120 to 3.134 seconds, the EMD and DWT increases from 0.591 to 1.543 seconds, and the Bidirectional LSTM increases from 0.489 to 0.998 seconds. Therefore, the proposed method is also robust to longer noise length.

8 Conclusion

In this thesis, we proposed a GAN-based method to reconstruct PPG signals distorted by real-world motion artifacts.

The proposed method was trained with clean PPG signals and reconstructed noisy signals using both current noisy signal, and the prediction based on previous clean PPG samples. The proposed method thus leveraged the temporal information in both current and proceeding data points, and enabled the reconstruction of PPG signals at high noise level. For preparing the training, validation and test dataset, the raw PPG signals were filtered and sectioned into 30-second segments. Then the quality of the signals is assessed to obtain the clean and distorted PPG signals.

A noisy test dataset was generated to evaluate the performance of our proposed and state-of-the-art methods in the literature. The test dataset contained noisy PPG signals corrupted by motion artifacts collected in daily life. The noisy segments in the test dataset were generated at different noise levels and noise duration.

The proposed method was compared with three other state-of-the-art PPG reconstruction methods in the literature, including a time-series, a signal decomposition and a deep-learning based method. All methods were evaluated by extracting two types of errors: maximum peak-to-peak error and RMSSD error, which represent the HR and HRV errors, respectively.

Our method performed the best at all noise levels with different noise lengths and produced the lowest values of the two types of error. The results indicated

that our method was effective to reconstruct noisy PPG signals distorted by real-world motion artifacts with low SNR values and long noise duration. The proposed method could consequently provide reliable estimation for both HR and HRV.

References

[1] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable photoplethysmography sensors and their potential future applications in health care", *International journal of biosensors & bioelectronics*, vol. 4, no. 4, p. 195, 2018.

[2] P. Renevey, R. Delgado-Gonzalo, A. Lemkaddem, *et al.*, "Optical wrist-worn device for sleep monitoring", in *EMBEC & NBC 2017*, Springer, 2017, pp. 615–618.

[3] C. El-Hajj and P. Kyriacou, "Cuffless blood pressure estimation from ppg signals and its derivatives using deep learning models", *Biomedical Signal Processing and Control*, vol. 70, p. 102 984, 2021.

[4] L. M. Eerikäinen, A. G. Bonomi, F. Schipper, *et al.*, "Comparison between electrocardiogram-and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions", *Physiological measurement*, vol. 39, no. 8, p. 084 001, 2018.

[5] J. Kim, T. Lee, J. Kim, and H. Ko, "Ambient light cancellation in photoplethysmogram application using alternating sampling and charge redistribution technique", in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 6441–6444.

[6] S. Hanyu and C. Xiaohui, "Motion artifact detection and reduction in ppg signals based on statistics analysis", in *2017 29th Chinese control and decision conference (CCDC)*, IEEE, 2017, pp. 3114–3119.

[7] D. Jarchi and A. J. Casson, "Estimation of heart rate from foot worn photoplethysmography sensors during fast bike exercise", in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2016, pp. 3155–2158.

[8] A. J. Casson, A. V. Galvez, and D. Jarchi, "Gyroscope vs. accelerometer measurements of motion from wrist ppg during physical exercise", *ICT Express*, vol. 2, no. 4, pp. 175–179, 2016.

[9] S. D. Tang, Y. S. Goh, M. D. Wong, and Y. E. Lew, "Ppg signal reconstruction using a combination of discrete wavelet transform and empirical mode decomposition", in *2016 6th International Conference on Intelligent and Advanced Systems (ICIAS)*, IEEE, 2016, pp. 1–4.

[10] H. Yuan, S. F. Memon, T. Newe, E. Lewis, and G. Leen, "Motion artefact minimization from photoplethysmography based non-invasive hemoglobin sensor based on an envelope filtering algorithm", *Measurement*, vol. 115, pp. 288–298, 2018.

[11] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, K. Sivani, and K. A. Reddy, "Ica-based improved dtcwt technique for ma reduction in ppg signals with restored respiratory information", *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 10, pp. 2639–2651, 2013.

[12] P. Nooralishahi, C. K. Loo, and L. W. Shiung, "Robust remote heart rate estimation from multiple asynchronous noisy channels using autoregressive model with kalman filter", *Biomedical Signal Processing and Control*, vol. 47, pp. 366–379, 2019.

[13] A. H. A. Zargari, S. A. H. Aqajari, H. Khodabandeh, A. M. Rahmani, and F. Kurdahi, "An accurate non-accelerometer-based ppg motion artifact removal technique using cyclegan", *arXiv preprint arXiv:2106.11512*, 2021.

[14] J. Lee, S. Sun, S. M. Yang, *et al.*, "Bidirectional recurrent auto-encoder for photoplethysmogram denoising", *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2375–2385, 2018.

[15] W. Yuning, A. Iman, K. Kianoosh, M. R. Amir, and L. Pasi, "Ppg signal reconstruction using deep convolutional generative adversarial network", in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2022.

[16] A. B. Hertzman, "The blood supply of various skin areas as estimated by the photoelectric plethysmograph", *American Journal of Physiology-Legacy Content*, vol. 124, no. 2, pp. 328–340, 1938.

[17] W. B. Murray and P. A. Foster, "The peripheral pulse wave: Information overlooked", *Journal of clinical monitoring*, vol. 12, no. 5, pp. 365–377, 1996.

[18] R. Blazek and C. Lee, "Multi-resolution linear model comparison for detection of dicrotic notch and peak in blood volume pulse signals", *Anal Biomed Signals Images*, vol. 20, no. 20, pp. 378–386, 2010.

[19] K. Nakajima, T. Tamura, and H. Miike, "Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique", *Medical engineering & physics*, vol. 18, no. 5, pp. 365–372, 1996.

[20] J. Allen, "Photoplethysmography and its application in clinical physiological measurement", *Physiological measurement*, vol. 28, no. 3, R1, 2007.

[21] A. D. Foster, C. Neumann, and E. Rovenstine, "Peripheral circulation during anesthesia, shock and hemorrhage; the digital plethysmograph as a clini-

cal guide", in *The Journal of the American Society of Anesthesiologists*, The American Society of Anesthesiologists, vol. 6, 1945, pp. 246–257.

[22] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A neural network-based method for continuous blood pressure estimation from a ppg signal", in *2013 IEEE International instrumentation and measurement technology conference (I2MTC)*, IEEE, 2013, pp. 280–283.

[23] B. Jönsson, C. Laurent, M. Eneling, T. Skau, and L.-G. Lindberg, "Automatic ankle pressure measurements using ppg in ankle-brachial pressure index determination", *European journal of vascular and endovascular surgery*, vol. 30, no. 4, pp. 395–401, 2005.

[24] H.-H. Kuo, *White noise distribution theory*. CRC press, 2018.

[25] M. Blanco-Velasco, B. Weng, and K. E. Barner, "Ecg signal denoising and baseline wander correction based on the empirical mode decomposition", *Computers in biology and medicine*, vol. 38, no. 1, pp. 1–13, 2008.

[26] R. McCraty and F. Shaffer, "Heart rate variability: New perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk", *Global advances in health and medicine*, vol. 4, no. 1, pp. 46–61, 2015.

[27] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms", *Frontiers in public health*, p. 258, 2017.

[28] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and heart rate variability: A meta-analysis and review of the literature", *Psychiatry investigation*, vol. 15, no. 3, p. 235, 2018.

[29] C. M. DeGiorgio, P. Miller, S. Meymandi, *et al.*, "Rmssd, a measure of vagus-mediated heart rate variability, is associated with risk factors for sudep: The sudep-7 inventory", *Epilepsy & behavior*, vol. 19, no. 1, pp. 78–81, 2010.

[30] H.-Y. Jan, M.-F. Chen, T.-C. Fu, W.-C. Lin, C.-L. Tsai, and K.-P. Lin, "Evaluation of coherence between ecg and ppg derived parameters on heart rate variability and respiration in healthy volunteers with/without controlled breathing", *Journal of Medical and Biological Engineering*, vol. 39, no. 5, pp. 783–795, 2019.

[31] C. Hoog Antink, Y. Mai, M. Peltokangas, S. Leonhardt, N. Oksala, and A. Vehkaoja, "Accuracy of heart rate variability estimated with reflective wrist-ppg in elderly vascular patients", *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[32] W.-H. Lin, D. Wu, C. Li, H. Zhang, and Y.-T. Zhang, "Comparison of heart rate variability from ppg with that from ecg", in *The international conference on health informatics*, Springer, 2014, pp. 213–215.

[33] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial nets", *Advances in neural information processing systems*, vol. 27, 2014.

[34] Z. Chen, Z. Zeng, H. Shen, X. Zheng, P. Dai, and P. Ouyang, "Dn-gan: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images", *Biomedical Signal Processing and Control*, vol. 55, p. 101 632, 2020.

[35] L. D. Tran, S. M. Nguyen, and M. Arai, "Gan-based noise model for denoising real images", in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, Nov. 2020.

[36] L. Weng, "From gan to wgan", *arXiv preprint arXiv:1904.08994*, 2019.

[37] J. Zhu, G. Yang, and P. Lio, "How can we make gan perform better in single medical image super-resolution? a lesion focused multi-scale approach", in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, 2019, pp. 1669–1673.

[38] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein, "Unrolled generative adversarial networks", *arXiv preprint arXiv:1611.02163*, 2016.

[39] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks", *arXiv preprint arXiv:1511.06434*, 2015.

[40] P. L. Suárez, A. D. Sappa, and B. X. Vintimilla, "Infrared image colorization based on a triplet dcgan architecture", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 18–23.

[41] R. Yeh, C. Chen, T. Y. Lim, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with perceptual and contextual losses", *arXiv preprint arXiv:1607.07539*, vol. 2, no. 3, 2016.

[42] A. Mahmoudzadeh, I. Azimi, A. M. Rahmani, and P. Liljeberg, "Lightweight photoplethysmography quality assessment for real-time iot-based health monitoring using unsupervised anomaly detection", *Procedia Computer Science*, vol. 184, pp. 140–147, 2021.

[43] C.-H. Goh, L. K. Tan, N. H. Lovell, S.-C. Ng, M. P. Tan, and E. Lim, "Robust ppg motion artifact detection using a 1-d convolution neural network", *Computer methods and programs in biomedicine*, vol. 196, p. 105 596, 2020.

[44] J. Azar, A. Makhoul, R. Couturier, and J. Demerjian, "Deep recurrent neural network-based autoencoder for photoplethysmogram artifacts filtering", *Computers & Electrical Engineering*, vol. 92, p. 107 065, 2021.

[45] G. Masinelli, F. Dell'Agnola, A. A. Valdés, and D. Atienza, "Spare: A spectral peak recovery algorithm for ppg signals pulsewave reconstruction in multimodal wearable devices", *Sensors*, vol. 21, no. 8, p. 2725, 2021.

[46] P. Ghosal, S. Himavathi, and E. Srinivasan, "Ppg motion artifact reduction using neural network and spline interpolation", in *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, IEEE, 2020, pp. 1–6.

[47] M. A. Mehrabadi, I. Azimi, F. Sarhaddi, *et al.*, "Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: Instrument validation study", *JMIR mHealth and uHealth*, vol. 8, no. 11, e20465, 2020.

[48] F. Sarhaddi, K. Kazemi, I. Azimi, *et al.*, "A comprehensive accuracy assessment of samsung smartwatch heart rate and heart rate variability", *medRxiv*, 2022. DOI: 10.1101/2022.04.29.22274461. eprint: https://www.medrxiv.org/content/early/2022/05/01/2022.04.29.22274461.full.pdf. [Online]. Available: https://www.medrxiv.org/content/early/2022/05/01/2022.04.29.22274461.

[49] *Samsung gear sport*, https://www.samsung.com/global/galaxy/gear-sport, Accessed: 2010-09-30.

[50] L. D. Tran, S. M. Nguyen, and M. Arai, "Gan-based noise model for denoising real images", in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[51] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training", in *Asian conference on computer vision*, Springer, 2018, pp. 622–637.

[52] G. Welch, G. Bishop, *et al.*, "An introduction to the kalman filter", 1995.

[53] R. Shibata, "Selection of the order of an autoregressive model by akaike's information criterion", *Biometrika*, vol. 63, no. 1, pp. 117–126, 1976.

[54] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method", *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2046, pp. 1597–1611, 2004.

[55] I. Daubechies, *Ten lectures on wavelets*. SIAM, 1992.

[56] A. Paszke, S. Gross, F. Massa, *et al.*, "Pytorch: An imperative style, high-performance deep learning library", *Advances in neural information processing systems*, vol. 32, 2019.

[57] D. D. et al., *Kalman filter, smoother, and em algorithm for python*, `https://github.com/pykalman/pykalman`, Github Repository, Sep. 2015.

[58] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python", in *Proceedings of the 9th Python in Science Conference*, Austin, TX, vol. 57, 2010, pp. 10–25 080.

[59] L. D. et al., *Python implementation of empirical mode decompoisition (emd) method*, `https://github.com/laszukdawid/PyEMD`, Github Repository, 2017.

[60] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "Pywavelets: A python package for wavelet analysis", *Journal of Open Source Software*, vol. 4, no. 36, p. 1237, 2019.

[61] F. Shaffer, Z. M. Meehan, and C. L. Zerr, "A critical review of ultra-short-term heart rate variability norms research", *Frontiers in Neuroscience*, vol. 14, 2020.