

# Can We Obtain Fine-grained Heartbeat Waveform via Contact-free RF-sensing?

Shujie Zhang<sup>1</sup>, Tianyue Zheng<sup>1</sup>, Zhe Chen<sup>2</sup>, Jun Luo<sup>1</sup>

<sup>1</sup> School of Computer Science and Engineering, Nanyang Technological University, Singapore

<sup>2</sup> China-Singapore International Joint Research Institute, Guangzhou, China

Email: {shujie002, tianyue002, junluo}@ntu.edu.sg, chenz@ssijri.com

**Abstract**—*Contact-free* vital-signs monitoring enabled by *radio frequency* (RF) sensing is gaining increasing attention, thanks to its non-intrusiveness, noise-resistance, and low cost. Whereas most of these systems only perform respiration monitoring or retrieve heart rate, few can recover fine-grained heartbeat waveform. The major reason is that, though both respiration and heartbeat cause detectable micro-motions on human bodies, the former is so strong that it overwhelms the latter. In this paper, we aim to answer the question in the paper title, by demystifying how heartbeat waveform can be extracted from RF-sensing signal. Applying several mainstream methods to recover heartbeat waveform from raw RF signal, our results reveal that these methods may not achieve what they have claimed, mainly because they assume linear signal mixing whereas the composition between respiration and heartbeat can be highly nonlinear. To overcome the difficulty of decomposing nonlinear signal mixing, we leverage the power of a novel deep generative model termed *variational encoder-decoder* (VED). Exploiting the universal approximation ability of deep neural networks and the generative potential of variational inference, VED demonstrates a promising capability in recovering fine-grained heartbeat waveform from RF-sensing signal; this is firmly validated by our experiments with 12 subjects and 48-hour data.

**Index Terms**—Vital signs monitoring, contact-free sensing, RF-sensing, deep learning, variational encoder-decoder.

## I. INTRODUCTION

In the past decade, the need for a ubiquitous solution to continuous vital signs monitoring has driven both academic and industry to invest in contact-free sensing. Leveraging reflected signals to capture body micro-motions, both respiration and heartbeat can be acquired by exploiting typically visible light [1]–[3], acoustic wave [4]–[8], and *radio-frequency* (RF) signal [9]–[14]. Among these methods, contact-free RF-sensing appears to be the most promising, because it is well supported by various commercial-grade devices [15]–[17] and it offers sufficient resolution without being affected by light or acoustic interference pertaining to an environment. Therefore, we have witnessed a wide range of developments concerning RF-enabled vital signs monitoring, where Wi-Fi [18]–[20], impulse radio ultra-wideband (IR-UWB) [14], [21], [22], and Frequency-Modulated Continuous Wave (FMCW) radars [10], [23]–[25] are usually adopted.

Although contact-free RF-sensing has demonstrated impressive performance in respiration monitoring and heart rate estimation [9]–[14], [18]–[24], few systems are capable of extracting *fine-grained heartbeat waveform*. Normally, waveform

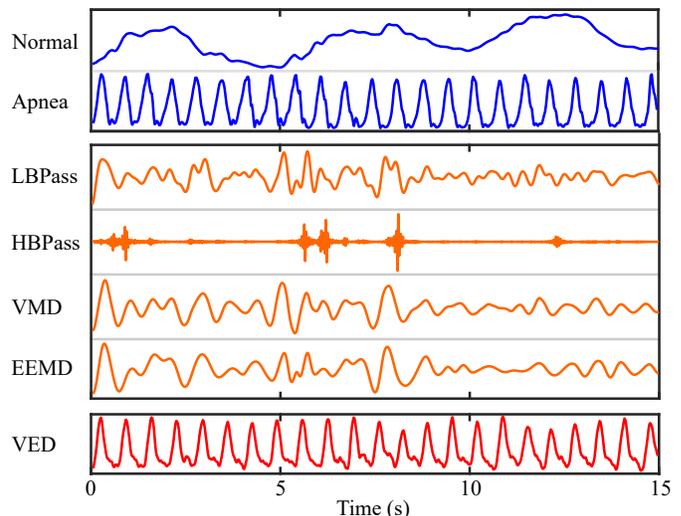


Fig. 1: Raw RF (phase) signal w/o and w/ breath holding (top two), heartbeat waveforms obtained by four 1-D signal processing methods, and that recovered by VED (the last one).

is deemed as much more useful than rate, as it depicts whole cardiac cycles [25]; yet recovering waveform is challenging because *heartbeat is often drowned out by respiration, which is (almost) co-located but incurs a much higher body-motion strength*. As shown by the first two curves in Fig. 1, whereas respiratory waveform can be readily identified from raw RF signal, heartbeat waveform becomes discernible only after the breath is held. Several methods have been developed to mitigate respiration interference [12], [26]–[28]. For example, bandpass and adaptive filters have been utilized in [12] to extract heartbeat waveform, while filters with a higher passband are adopted by [27], [28] to retrieve phonocardiograph (PCG) or seismocardiogram (SCG) waveform. Other proposals [14], [26] apply more sophisticated methods such as Variational Mode Decomposition (VMD) and Ensemble Empirical Mode Decomposition (EEMD) to decompose raw RF signal and obtain the heartbeat waveform component therein.

Consequently, one may naturally ask: *can fine-grained heartbeat waveform really be obtained via contact-free RF-sensing?* Aiming to answer this question, we use both modeling and experiments to demystify how to recover heartbeat waveform from raw RF signal. As a teaser, we briefly summarize our experience with the state of art methods in

Fig. 1, where raw RF signals (with breathing) are respectively processed by four methods: low bandpass filter (LBPass, 0.8-3.6 Hz), high bandpass filter (HBPass, 16-35 Hz), VMD, and EEMD. It can be observed that, under the breath interference, none of these methods could effectively recover fine-grained heartbeat waveform. In particular, LBPass roughly obtains the fundamental frequency component of heartbeat waveform (albeit highly distorted), while HBPass hardly achieves the morphology similarity claimed in [27], [28]. VMD delivers a smoother waveform compared to that of LBPass (yet losing more details), and EEMD performs similarly to VMD.

Given the inadequacy of existing proposals, we first analyze the reason behind the unsatisfactory results, and we figure out that these methods all operate under the wrong assumption that heartbeat and respiratory waveforms are linearly composed. To handle the nonlinear decomposition problem, we innovatively design a deep generative model termed *variational encoder-decoder* (VED). In order to better leverage the signal representation in complex in-phase/quadrature (I/Q) domain (rather than only amplitude or phase adopted by earlier proposals), our VED model takes both components of raw complex RF signals as inputs. More importantly, VED leverages i) the universal approximation ability of its encoder-decoder network, and ii) the generative potential of the variational inference in its latent space, in order to perform both nonlinear decomposition and waveform recovery. To briefly demonstrate the effectiveness of our method, a VED-recovered fine-grained heartbeat waveform is shown at the bottom of Fig. 1. In summary, we make the following major contributions in this paper:

- We reveal the true challenges of recovering fine-grained heartbeat waveform from contact-free RF-sensing data.
- By both modeling and experiments, we draw lessons from existing proposals and derive guidelines for developing contact-free cardiac monitoring system effectively.
- We propose VED as a novel deep generative model to approximate a nonlinear decomposition, so as to recover fine-grained heartbeat waveform from raw RF signal.
- We conduct extensive evaluations on Wi-Fi, IR-UWB, and FMCW devices with a 48-hour dataset.

The rest of this paper is organized as follows. Section II aims to demystify heartbeat waveform recovery by contact-free RF sensing, via both modeling and experiments. We then propose our RF-sensing solution to fine-grained heartbeat waveform recovery in Sec. III. Sec. IV reports the evaluation results of our method. Finally, Sec. V concludes our paper.

## II. CONTACT-FREE CARDIAC MONITORING

In this section, we start with the rationale of contact-free cardiac monitoring and of adopting RF as the sensing media. Then we introduce the working principles of RF cardiac monitoring and perform a comparative study among three mainstream schemes (i.e., Wi-Fi, IR-UWB, and FMCW) to verify their respective capabilities. We finally clarify the inefficacy of existing proposals and uncover the true challenges, using FMCW as the sensing technology.

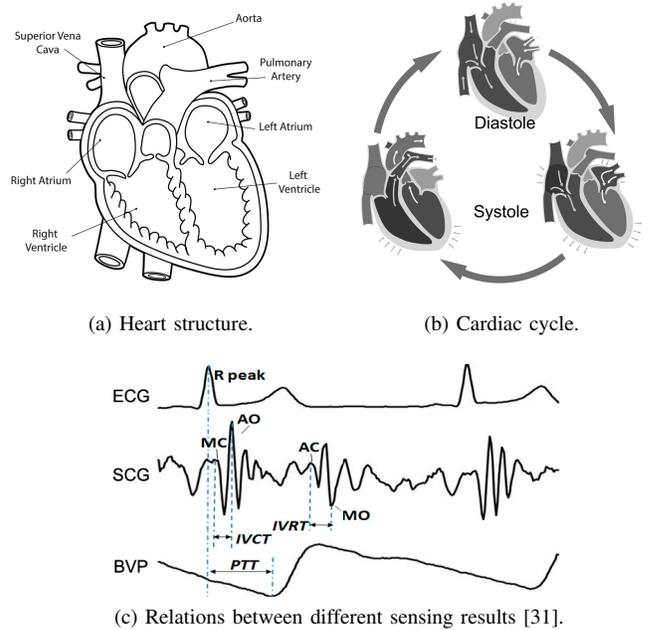


Fig. 2: Heart structure, dynamics, and various sensing results.

### A. Understanding Heartbeat

We first expose a cardiac cycle from both its intrinsic operation details and the resulting extrinsic manifestation. As shown in Fig. 2a, a heart is separated into left and right parts; each part contains an atrium and a ventricle. The heart's electrical system generates rhythmical impulses to regulate muscular contraction of the heart, hence producing heartbeat. Each cardiac cycle (or beat), as shown in Fig. 2b, consists of a sequence of cardiac events commonly separated into two periods: relaxation (diastole) and contraction (systole). The cycle starts with heart relaxing and filled with blood, then ends with heart contracting and pumping blood through lungs and peripheral circulation [29]. As each beat forces a new surge of blood to fill the arteries, these vessels dilate and recoil to maintain pressure while accommodating blood volume changes. Such dilation and constriction lead to the so-called *blood volume pulse* (BVP), creating an extrinsic representation of heartbeat [25], [30].

Some of these physiological events can only be observed via wearable sensing, such as electrocardiogram (ECG) that records the rhythmical electrical impulses. However, our following discussions focus only on contact-free cardiac monitoring. In terms of intrinsic operations, previous studies claim that phonocardiogram (PCG, sounds generated by heartbeat) and seismocardiogram (SCG, chest wall micro-vibrations induced by heartbeat) can be captured by RF sensing [27], [28]. However, our experiments in Sec. II-C shall reveal their sensing distance being too short to be practically usable for contact-free monitoring. To reach a longer sensing distance, extrinsic BVP has been leveraged in two ways. On one hand, remote photoplethysmography (rPPG) adopts a camera to capture the subtle color changes of a subject's facial regions caused

by BVP under skin.<sup>1</sup> On the other hand, acoustic [4] and RF [12]–[14], [19]–[24], [26] signals are used to sense the micro-vibrations induced by BVP, where the most conspicuous vibrations are produced by the common carotid artery on neck. As a vision-based method, rPPG is highly vulnerable to illumination changes and other light interference; a similar situation happens to acoustic sensing as sound interference is inherent to our living environments. Therefore, RF sensing appears to be the most reasonable solution for measuring BVP. We illustrate ECG, SCG (similar to PCG), and BVP along with corresponding events in Fig. 2c, but we refer readers to [31] for detailed explanations on the marked events.

### B. Capturing Heartbeat with RF signals

Existing proposals have employed various RF schemes (Wi-Fi, IR-UWB, and FMCW) for capturing heartbeat. Although they apply different signal frequencies and modulation techniques, the received signal model can be unified [34], [35] to facilitate exposition and further processing. Generally speaking, they all explore the reflected signal  $y(t)$  off human bodies:

$$y(t) = \sum_{p=1}^P \alpha_p(t) e^{-j2\pi f_c \frac{2d_p(t)}{c}} + n(t), \quad (1)$$

where  $P$  is the number of reflectors,  $\alpha_p(t)$  is the baseband signal after a real channel gain of the  $p$ -th reflected signal,  $f_c$  is the carrier frequency,  $c$  is the speed of light,  $d_p(t)$  is the distance from sensor to  $p$ -th reflector, and  $n(t)$  is Gaussian noise. Here we omit the detailed representation of  $\alpha_p(t)$  for brevity. After I/Q downconversion, the demodulated baseband signal can be represented as a complex sequence  $r(t) = r_I(t) + jr_Q(t)$ , with the phase of  $r(t)$  representing that of the carrier in Eqn. (1). Plotting the amplitude of  $r(t)$  in Fig. 3a, we can observe that the *subjects* (human bodies in our case) indicated by peaks mark the strongest reflectors within sensing range; they can be clearly differentiated via distinct distances.<sup>2</sup>

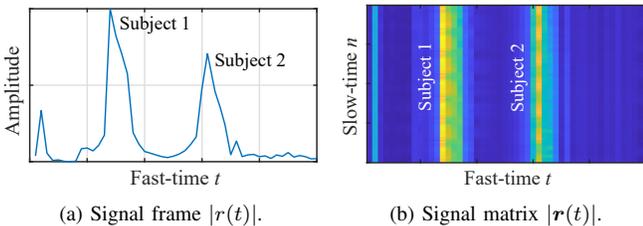


Fig. 3: Reflections from two human subjects and their corresponding heartbeats can be observed in a single frame and a signal matrix stacking multiple frames, respectively.

However, a single *frame*  $r(t)$  only allows one to detect the distance of a subject. To monitor heartbeat, one has to transmit

<sup>1</sup>Similarly, photoplethysmography (PPG) is a wearable sensing method commonly adopted to obtain ground truth for experimentally validating contact-free sensing [32], [33].

<sup>2</sup>The ability of differentiating reflectors at different distances relies on the bandwidth of the baseband signal. Therefore, Wi-Fi, with a bandwidth of only several tens of MHz (in other words,  $\alpha_p(t)$  remains almost constant with a frame), is insufficient to achieve such a high distance resolution.

multiple frames at a regular interval to capture the micro-motion of a subject. Stacking a sequence of such frames forms a *signal matrix*  $\mathbf{r}(t) = [r_1(t), \dots, r_n(t), \dots, r_N(t)]^T$ , where  $t$  and  $n$  are respectively the *fast-time* and *slow-time* indices, and  $N$  denotes the number of slow-time frames. As the fast-time index  $t$  indicates distance, we also term it *distance bin*, or simply *bin*. We illustrate a signal matrix of two subjects holding their breath in Fig. 3b, in which one may readily observe the heartbeats. Since this paper is meant to reveal how heartbeat waveform can be recovered from a single subject, we hereby focus only on a single column  $r(n)$ , whose I/Q components are respectively represented as:

$$r_I(n) = \alpha(n) \cos\left(\frac{4\pi d_0}{\lambda} + \frac{4\pi d_h(n)}{\lambda}\right) + o_I^B(n), \quad (2)$$

$$r_Q(n) = \alpha(n) \sin\left(\frac{4\pi d_0}{\lambda} + \frac{4\pi d_h(n)}{\lambda}\right) + o_Q^B(n), \quad (3)$$

where  $\alpha(n)$  describes the magnitude of the reflected signal off a human body,  $d_0$  is the mean distance from the radar to the body,  $\lambda$  is the wavelength of the carrier, and  $d_h(n)$  is the micro-displacement caused by heartbeat. We use  $o^B(n) = o_I^B(n) + jo_Q^B(n)$  to describe the total effect of interference from respiration and body background reflection.

Given these theoretical backgrounds, we are now ready to investigate the ability of capturing heartbeat waveform by the three major RF schemes. We use three RF devices (Wi-Fi [15], IR-UWB [16], and FMCW [17]) to simultaneously record reflected signals when a subject is either holding the breath or breathing naturally, and some typical results are plotted in Fig. 4, leading to the following observations:

- All three RF schemes are able to capture heartbeat waveform to some extent with subject holding the breath; otherwise, heartbeat is totally overwhelmed by breath and further signal processing to separate them is necessary.
- Given the constellation of  $r(n)$  in a complex plane illustrated in the first two rows, a trajectory caused by (body) micro-displacements always forms an elliptic arc, whose central angle is proportional to the magnitude of micro-displacements, hence allowing the breath (if not held) to drown out the heartbeat.
- Another key factor affecting the central angle of an elliptic arc is the carrier frequency  $f_c$ : Wi-Fi at 5.8 GHz, IR-UWB at 7.29 GHz, and FMCW at 77 GHz. Essentially, the higher the  $f_c$ , the more sensitive the carrier phase, hence that of  $r(t)$ , is to the changes in reflected signals induced by micro-displacements.
- The arc may not be circular and centered at the origin, because  $o^B(n)$  induces a varying offset to it. Compared to Wi-Fi with 40 MHz bandwidth, IR-UWB and FMCW (with much wider 1.5 GHz and 3 GHz bandwidths, respectively), yield narrower bins with less interference from  $o^B(n)$ , hence leading to much cleaner traces.
- When studying the 1-D representation of  $r(t)$  (i.e., its amplitude and phase) in Fig 4, the same phenomenon that respiration overwhelms heartbeat can again be observed. The different distortions existing in amplitude and phase

representations are likely to be caused by the distinct constructions of these devices: we consistently observe that  $r(t)$  has a better projection on its amplitude for IR-UWB, as opposed to phase for FMCW.

In conclusion, heartbeat waveform could be captured rather accurately (by either IR-UWB and FMCW, but not certainly not Wi-Fi), when there is no interference from respiration. And 1-D representation of  $r(t)$  (amplitude or phase) is able to describe the periodic changes of body micro-displacement despite the device-dependent variations. However, *if 1-D representation is enough for signal separation* is still a question that we would like to find out in the following section.

### C. What Went Wrong with Previous Methods

Given the aforementioned background, we hereby study several state-of-the-art methods and analyze why they fail to obtain fine-grained heartbeat waveform through experiments. We use FMCW as the sensing technology considering its high distance resolution and motion sensitivity. Moreover, as existing methods mainly take 1-D signal as their input and FMCW performs better in phase (see Fig. 4), we shall take phase signal as the input in the following experiments.

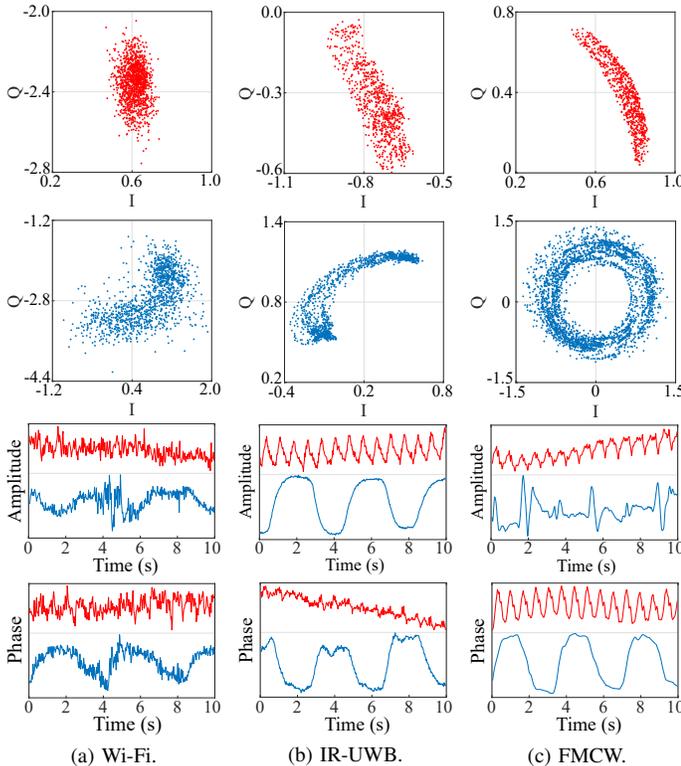


Fig. 4: Comparison of three RF schemes in terms of capturing heartbeat waveform. We use the first two rows to depict  $r(n)$  (in I/Q domain) with (red) or without (blue) holding the breath. Then the last two rows respectively demonstrate what happens if only the amplitude or phase of  $r(n)$  is considered, with (red) or without (blue) holding the breath.

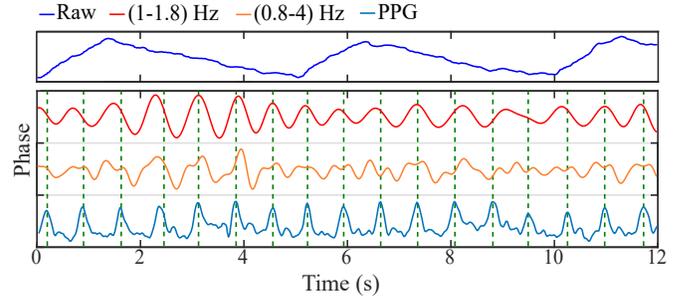


Fig. 5: Low-range bandpass filtering suffers from losing details, as well as distortion caused by respiration harmonics.

1) *Low-Range Bandpass Filtering*: Applying a bandpass filter may be the most intuitive way, because the heartbeat and respiration have very different fundamental frequencies. For healthy adults at rest, the normal respiratory rate often falls between 12 and 16 rpm (respirations per minute), while the heart rate ranges from 60 to 100 bpm (beats per minute) [36]. Therefore, applying a bandpass filter of (1-1.8) Hz to extract heartbeat waveform appears to be a plausible solution [20]. In practice, range of (0.8-4) Hz is usually adopted to include abnormal cases [10], [13], [21], and also to enrich the waveform with more high frequency details.

Fig. 5 shows the raw and filtered phase of the RF signal, as well as the BVP ground truth provided by a PPG wearable sensor [33]. It is clear that the waveforms yielded by bandpass filters largely fail to characterize cardiac cycles: while the narrower range filter leads to only the fundamental frequency without details, the wider range filter results in waveform distorted by harmonic frequency components of respiration. In fact, their peaks cannot be well aligned with those of the ground truth, rendering wrong estimation even to heart rate. The above results were obtained with a subject facing the radar at a distance of 1 m. Bounded by the requirement on signal-to-noise ratio (SNR), the maximum distance at which BVP waveform could be detected is roughly 5 m.

2) *High-Range Bandpass Filtering*: According to Sec. II-A, PCG and SCG represent intrinsic cardiac operations correlated with ECG signals, as the mechanical vibrations are induced by the electrical signals [30], [37]. Therefore, as opposed to BVP extrinsically representing heartbeat, PCG/SCG occupy much higher frequencies, making researchers believe that they would evade the interference from respiration. Consequently, bandpass filters at a higher frequency range are employed to obtain such waveforms, e.g., (16-80) Hz filter for PCG [27], and (18-35) Hz one for SCG [28].

The results of applying two high-range bandpass filters are presented in Fig. 6, with ECG waveform collected by another wearable sensor [38] as ground truth, and *R wave peak* and *T wave end* labeled as important reference points [30], [37]. These results evidently demonstrate that it is incorrect to believe that PCG/SCG are not prone to the interference from respiration. On one hand, the correlation between the existence of PCG/SCG and the respiration states is very strong. On the other hand, even during the period when PCG/SCG waveforms

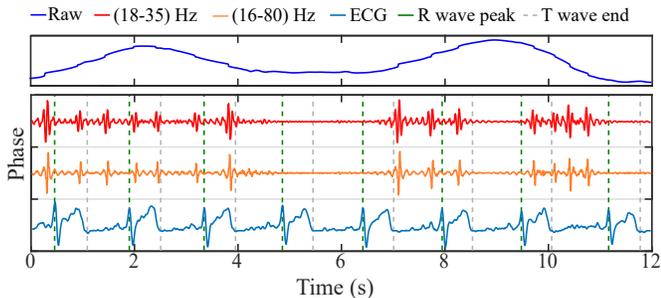


Fig. 6: High-range bandpass filtering may avoid respiration interference, yet it still fails to recover meaningful waveform.

are observable, these waveforms fail to represent heartbeat patterns as they are not well aligned with the ECG events. In addition, the above results were obtained at a distance of 20cm. Due to the high SNR requirement to preserve the details in waveforms for indicating every cardiac event, PCG/SCG are only obtainable at close proximity. As shown in Fig. 7, a slight increase in distance makes it more difficult to observe heartbeat pattern from PCG/SCG waveforms. Compared with the meter-level sensing distance of BVP, the applicability of PCG/SCG in real-life monitoring is highly questionable.

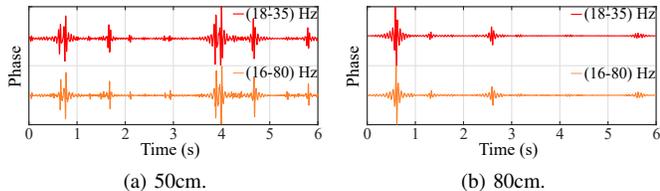


Fig. 7: Heartbeat waveform obtained by high-range bandpass filter at different distances: a minor distance increase can totally ruin heartbeat patterns already barely observable.

3) *Decomposition Approaches*: Compared with the aforementioned classical filtering methods focusing on spectral analysis, mode decomposition methods [39], [40] targeting time-domain analysis are supposed to have better adaptability in signal separation. Among them, EEMD [26] and VMD [14] are recently adopted for heartbeat waveform extraction. These methods decompose signals into components called *intrinsic membership functions* (IMFs). For heartbeat waveform extraction, these methods normally break down raw radar signals into the BVP signal, respiration signal, body movements, as well as noises. We apply both VMD and EEMD to decompose raw phase signals measured at 1m; the extracted heartbeat waveforms and the PPG ground truth are illustrated in Fig. 8. Although decomposition methods recover the fundamental frequency components of heartbeat better than the low-range bandpass filtering, they again fail to obtain fine-grained heartbeat waveform, because the resulting waveforms are totally void of necessary details.

4) *Summary*: According to our experiment analysis, all existing methods fail to distill fine-grained heartbeat waveform from respiration interference. The reason for this failure is twofold: i) these methods all involve linear operations, capable

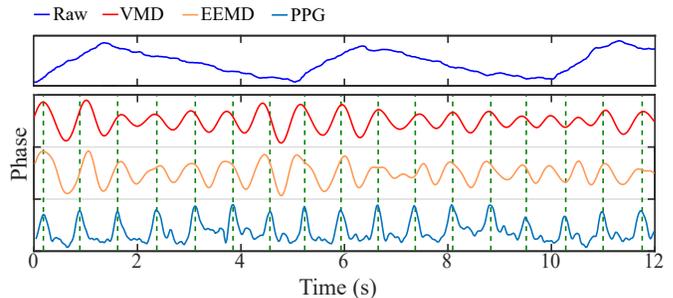


Fig. 8: Mode decomposition methods again yield fundamental frequency components but fail to recover necessary details.

of handling only linearly composed signal mixing, and ii) they all rely on only 1D (phase or amplitude) data input. As indicated in Fig. 4, I/Q data provide a much more comprehensive characterization of body motions, so it appears to be imperative to design a nonlinear decomposition method taking I/Q data as its input.

### III. AN INTEGRATED DEEP LEARNING APPROACH

Given the inefficacy of linear approaches, we intend to work directly on I/Q data and propose a deep learning module to disentangle the nonlinear mixing. We firstly briefly motivate our approach and introduce the design rationale of our VED model. We then give an overview of the VED architecture, followed by further elaborations on the working principles of individual VED components, as well as how VED gets trained.

#### A. Design Rationale

Though nonlinear filtering methods exist, they may not help recovering heartbeat waveform for two reasons. On one hand, recursive approximation methods (e.g., those similar to EEMD [39]) cannot achieve satisfying performance, as discussed in Sec. II-C3. On the other hand, designing task-specific nonlinear filters (e.g., median filter [41] and linear-in-the-parameters nonlinear filter [42]) requires extensive professional experience and domain knowledge, yet they can hardly be generalized. To overcome these challenges to deterministic filtering, it may help to view the problem from the probabilistic estimation perspective. Essentially, extracting heartbeat waveform  $\mathbf{x}$  from the raw RF signal  $\mathbf{r}$  can be viewed as modelling the underlying distribution  $p(\mathbf{x}|\mathbf{r})$  of  $\mathbf{x}$  conditioned on  $\mathbf{r}$ . Driven by raw RF signals and ground truth waveforms, these approaches can potentially discover the nonlinear relationship between the input and output automatically; a recent avatar of such an approach is learning via deep neural networks [43].

Whereas basic deep neural networks may not adapt well to model the specific distribution  $p(\mathbf{x}|\mathbf{r})$  in our case, the deep learning community has been employing Encoder-Decoder (ED) models for handling signal separation [44], [45]. In an ED model, an encoder learns from raw input  $\mathbf{r}$  and encodes it into a compact representation  $\mathbf{h}$  in a latent space. Then a decoder uses this latent representation as input to reconstruct the desired output  $\mathbf{x}$ . This “bottleneck” structure is especially efficient for separating signal mixing, as the structure forces

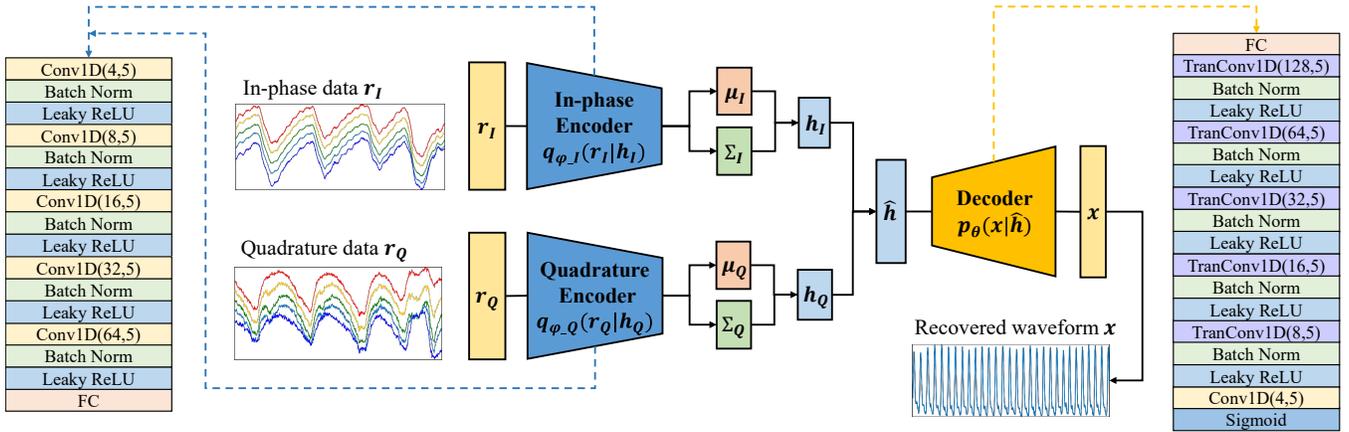


Fig. 9: The VED Architecture. I/Q encoders take raw I/Q data and encode them to respective latent probability distributions. Latent vectors are then sampled from the distributions and fed to the decoder for recovering the heartbeat waveform.

$\mathbf{h}$  to “forget” irrelevant signals while keeping compressed knowledge of the desired one.

Unfortunately, a regular ED model is often prone to overfitting, as it may simply memorize the computation leading to a discontinuous encoding in the latent space. To improve the generalizability of the model, we regularize the latent space by encoding latent vector  $\mathbf{h}$  to probability distributions instead of discrete values. As such, the encoder  $q_\phi(\mathbf{h}|\mathbf{r})$  with parameters  $\phi$  can produce more meaningful representations better generalizable to unseen data, and the decoder  $p_\theta(\mathbf{x}|\mathbf{h})$  with parameters  $\theta$  can fully leverage these regularized representations to reconstruct heartbeat waveforms without overfitting to training samples. Because the model uses *variational inference* [46] to model the latent distributions, we name our model Variational Encoder-Decoder (VED). Nevertheless, different from Variational AutoEncoder (VAE) models [47] generating new samples similar to training data from random vectors, VED is designed to decompose the input data and reconstruct specific component. To provide an overview before diving into details, we present the overall VED architecture along with example inputs and outputs in Fig. 9.

### B. Encoder

The encoder  $q_\phi(\mathbf{h}|\mathbf{r})$  takes in the I/Q signal  $\mathbf{r}$  and encodes it to a latent representation  $\mathbf{h}$ . Since  $\mathbf{r}$  consists of complex values as explained in Sec. II-B, we choose to design two parallel encoders of the same structure to respectively handle the in-phase component  $r_I$  and quadrature component  $r_Q$  of  $\mathbf{r}$  (see Fig. 9). Although existing proposals have explored deep complex networks [48] to handle complex inputs, they are usually difficult to train (with extremely slow convergence) as they require redefining calculus operations (e.g., differentiation) crucial to backpropagation. Therefore, applying two encoders is certainly a more efficient solution in our case.

The encoder network consists of multiple stacks, with each stack containing a one-dimensional convolutional layer [49], a batch normalization layer [50], and a leaky ReLU layer [51] in sequential order. As shown in Fig. 9, each convolutional layer

$\text{Conv1D}(m, n)$  takes  $m$  input channels and adopts a kernel of size  $n$ . In the meantime, the feature map length is reduced as each convolutional layer downsamples its input by a factor of 4 (i.e., stride 4), funneling input data into the latent space. Essentially, the encoder applies i) 1D-CNN to extract temporal features related to heartbeat, ii) the batch normalization layer to make the training process more efficient and stabilized, and iii) the Leaky ReLU layer as activation function for adding nonlinearity to the network. Finally, the encoded features are mapped into the latent distributions by fully connected layers.

### C. Latent Space

The outputs of the encoders are two Gaussian distributions  $\mathbf{h}_I \sim \mathcal{N}(\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I^2)$  and  $\mathbf{h}_Q \sim \mathcal{N}(\boldsymbol{\mu}_Q, \boldsymbol{\Sigma}_Q^2)$  parameterized by respective means and variances. During training,  $\mathbf{h}_I$  and  $\mathbf{h}_Q$  should be sampled from respective distributions. However, since we cannot obtain derivatives of random variables, backpropagation does not work if  $\mathbf{h}_I$  and  $\mathbf{h}_Q$  are directly sampled. To this end, we apply the *reparameterization trick* to restructure how we take derivatives of the latent representations. We randomly sample  $\boldsymbol{\epsilon}$  from a unit Gaussian distribution  $\mathcal{N}(\mathbf{0}, \mathbf{I})$ , then shift the sample by respective means and scaled by respective covariances to obtain  $\mathbf{h}_I$  and  $\mathbf{h}_Q$ :

$$\mathbf{h}_I = \boldsymbol{\mu}_I + \boldsymbol{\Sigma}_I \odot \boldsymbol{\epsilon}, \quad (4)$$

$$\mathbf{h}_Q = \boldsymbol{\mu}_Q + \boldsymbol{\Sigma}_Q \odot \boldsymbol{\epsilon}, \quad (5)$$

with  $\odot$  denoting an element-wise product. Consequently, the parameters of the distributions can be leveraged to obtain derivatives during backpropagation while maintaining the stochastic nature of the latent representations. After each I/Q encoded distribution is sampled as  $\mathbf{h}_I$  and  $\mathbf{h}_Q$ , we concatenate them to form  $\hat{\mathbf{h}} = [\mathbf{h}_I, \mathbf{h}_Q]$ ; this  $\hat{\mathbf{h}}$  becomes the input to the first fully connected layer of the decoder for further heartbeat waveform reconstruction.

### D. Decoder

The decoder  $p_\theta(\mathbf{x}|\hat{\mathbf{h}})$  serves as an expansion model to reconstruct heartbeat waveform. As shown in Fig. 9, the

decoder network is virtually a reverse of the encoder. We use stacks of 1-D transposed convolutional layers [52], batch normalization layers and leaky ReLU layers. Specifically, each transposed convolutional layer  $\text{TranConv1D}(m, n)$  takes in  $m$  input channels and adopts a kernel size of  $n$ . In the meantime, the feature map length is increased as each transposed convolutional layer upsamples input by a factor of 4 (i.e., stride 4), and expands the latent representation to a longer sequence. Finally, the decoded features are transformed by a 1-D convolutional layer to match the output dimension, and sigmoid is used as the activation function to finally produce  $\mathbf{x}$ .

### E. Training Strategy

Theoretically, the best performance of heartbeat recovery can be reached by maximizing the likelihood  $p_\theta(\mathbf{x})$ , which can be derived as follows:

$$\begin{aligned} \log p_\theta(\mathbf{x}) &= \mathbb{E}_{q_\phi(\mathbf{h}|\mathbf{r})} [\log p_\theta(\mathbf{x})] \\ &= \mathbb{E}_{q_\phi(\mathbf{h}|\mathbf{r})} [\log p_\theta(\mathbf{x}|\mathbf{h})] - D_{\text{KL}}(q_\phi(\mathbf{h}|\mathbf{r})\|p_\theta(\mathbf{h})) \\ &\quad + D_{\text{KL}}(q_\phi(\mathbf{h}|\mathbf{r})\|p_\theta(\mathbf{h}|\mathbf{r})), \end{aligned} \quad (6)$$

where  $D_{\text{KL}}(\cdot)$  denotes the Kullback-Leibler (KL) divergence [53]. Although only the first two terms are tractable, together they form the so called *variational lower bound* (VLB), which can be maximized instead.

$$\begin{aligned} \text{VLB}_{\text{VED}}(\mathbf{x}; \phi, \theta) &= \mathbb{E}_{q_\phi(\mathbf{h}|\mathbf{r})} [\log p_\theta(\mathbf{x}|\mathbf{h})] \\ &\quad - D_{\text{KL}}(q_\phi(\mathbf{h}|\mathbf{r})\|p_\theta(\mathbf{h})) \\ &= \log p_\theta(\mathbf{x}) - D_{\text{KL}}(q_\phi(\mathbf{h}|\mathbf{r})\|p_\theta(\mathbf{h}|\mathbf{r})) \\ &\leq \log p_\theta(\mathbf{x}), \end{aligned} \quad (7)$$

Intuitively,  $\mathbb{E}_{q_\phi(\mathbf{h}|\mathbf{r})} [\log p_\theta(\mathbf{x}|\mathbf{h})]$  describes the reconstruction of heartbeat waveform, and  $D_{\text{KL}}(q_\phi(\mathbf{h}|\mathbf{r})\|p_\theta(\mathbf{h}))$  measures KL divergence between approximate posterior distribution to the true prior  $p(\mathbf{h})$ . These two terms can be practically implemented by the reconstruction loss and two I/Q regularization losses:

*a) Reconstruction Loss:* As a practical measure of  $\mathbb{E}_{q_\phi(\mathbf{h}|\mathbf{r})} [\log p_\theta(\mathbf{x}|\mathbf{h})]$ , the reconstruction loss  $\mathcal{L}_{RC}$  is defined in  $L_2$  that measures the sum of point-wise differences between the recovered waveform  $\mathbf{x}$  and the ground truth  $\mathbf{x}_{\text{gt}}$ .

$$\mathcal{L}_{RC} = \|\mathbf{x}_{\text{gt}} - \mathbf{x}\|_2, \quad (8)$$

*b) I/Q Regularization Loss:* Corresponding to the term  $D_{\text{KL}}(q_\phi(\mathbf{h}|\mathbf{r})\|p_\theta(\mathbf{h}))$  in Eqn. (7), we have two losses for regularizing the in-phase and quadrature distributions. Since we assume both true prior distributions as unit Gaussian, the I/Q regularizing losses  $\mathcal{L}_{IR}$  and  $\mathcal{L}_{QR}$  are defined as:

$$\mathcal{L}_{IR} = \mathbf{D}_{\text{KL}}(\mathcal{N}(\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I^2), \mathcal{N}(\mathbf{0}, \mathbf{I})), \quad (9)$$

$$\mathcal{L}_{QR} = \mathbf{D}_{\text{KL}}(\mathcal{N}(\boldsymbol{\mu}_Q, \boldsymbol{\Sigma}_Q^2), \mathcal{N}(\mathbf{0}, \mathbf{I})), \quad (10)$$

Combining these loss functions, the overall loss function for training VED becomes:

$$\mathcal{L}_{\text{VED}} = \mathcal{L}_{RC} + \gamma(\mathcal{L}_{IR} + \mathcal{L}_{QR}), \quad (11)$$

where  $\gamma$  is the weight for regularizing losses.  $\mathcal{L}_{IR}$  and  $\mathcal{L}_{QR}$  share the same weight  $\gamma$ , because the I/Q data are of equal importance and I/Q encoders are symmetric.

## IV. EVALUATION AND DISCUSSION

In this section, we evaluate VED's capabilities of recovering fine-grained heartbeat waveform, given several real-life scenarios and under various parameter settings.

### A. Evaluation Setup

We recruit 12 healthy subjects (6 females and 6 males), aged from 18 to 60 and weighted from 50 to 80 kg, and we conduct experiments with them under the IRB approval of our institute. During data collection, the subjects are asked to sit (quasi-)statically in real-life environments (e.g., offices and classrooms), and we collect RF-sensing data under their natural states (except very few breath-holding cases). To demonstrate the effectiveness of VED in recovering heartbeat waveform, we use both IR-UWB [16] and FMCW [17] radars to capture 100 frames per second during data collection, considering their competent performance briefly demonstrated in Sec. II-B. To better understand the performance of VED under various conditions, we conduct measurements when subjects sit at different distances and orientations from radar, and we even consider cases where subjects exhibit minor body movements. In total, 48 hours of IR-UWB and FMCW radar data (equivalent to approximately 180,000 heartbeat cycles) are collected along with the ground truth recordings. We divide the data into 20-second segments, leading to a total of 8,640 data samples. We use 30% of the samples (only 2 female and 2 male subjects) to train the VED, then we test it upon the remaining 70% samples.

We choose EEMD [26] as the *baseline* method for comparison, as it is a state-of-art method for signal decomposition, and it appears to have arguably the best performance among other existing methods. For data collected by IR-UWB and FMCW radars, we respectively use the amplitude and phase of the signal as input to EEMD. Apart from the waveform shape, we consider another two features to describe the performance of heartbeat waveform recovery: *interbeat interval* (IBI) and *heart rate* (HR). To quantitatively evaluate the recovered waveform quality, we adopt the following metrics:

- *Cosine Similarity.* Defined as the cosine between two vectors, cosine similarity is employed to measure the morphology similarity between the recovered heartbeat waveform  $\mathbf{x}(n)$  and ground truth waveform  $\mathbf{x}_{\text{gt}}(n)$ . A higher cosine similarity between the two waveforms indicates a better recovery performance.
- *Relative Error.* Defined as the ratio between the absolute error and the ground truth value, relative error is applied to measure how much (in percentage) an estimated quantity deviates from its ground truth value. A lower relative error indicates a better recovery performance.

### B. Overall Performance Results

We first present two examples of heartbeat waveform recovery by IR-UWB and FMCW radars in Fig. 10a and 10b, respectively, where the raw RF signal, VED recovered waveform, ground truth waveform, and baseline recovered waveform are illustrated from top to bottom. It is evident that

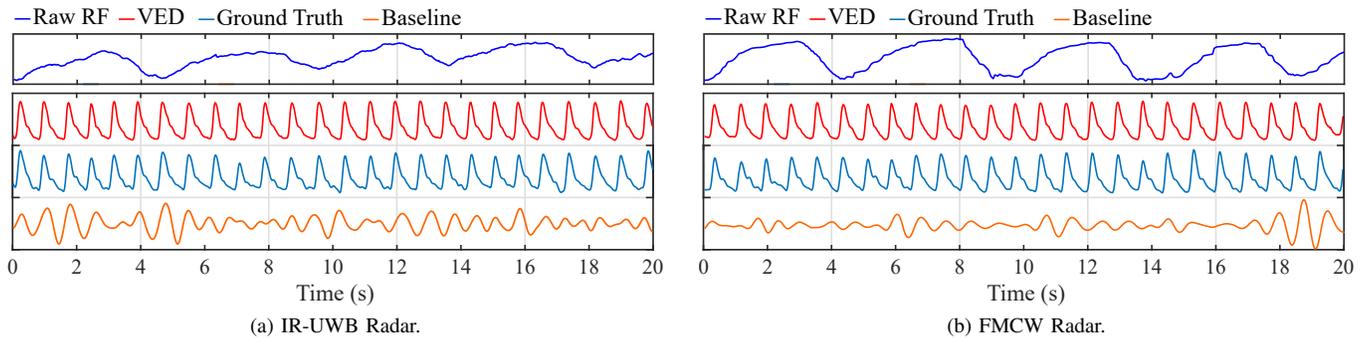


Fig. 10: Examples of heartbeat waveform recovery from two different devices.

VED captures not only the accurate HR and IBI, but also the exact shape of the heartbeat waveform, whereas the baseline generally fails to provide meaningful waveform with correct details. Moreover, the performance of VED waveform recovery appears to be rather independent of the underlying hardware, indicating VED’s cross-technology ability.

We further report the qualitative evaluation results obtained by applying the metrics defined earlier. In Fig. 11a and Fig. 11c, it can be seen that VED achieves a median cosine similarity higher than 0.9 and 0.92 respectively for IR-UWB and FMCW radars, while the same quantity of the baseline method is always below 0.85. For heart rate estimations, VED achieves median errors less than 2.4% for both IR-UWB and FMCW radars, while the baseline method has median errors around 5% and maximum errors reaching up to 20%. For IBI estimations, VED achieves median errors less than 3.6%, as opposed to those of the baseline method getting beyond 6.9% (even worse for FMCW radar). Although the performance of the baseline method can be acceptable for certain applications, VED consistently outperforms the baseline in all metrics and is hence suitable for more advanced and broad application scenarios demanding high-accuracy heartbeat waveform. In summary, all these results have evidently demonstrated the su-

perior performance of VED in recovering heartbeat waveform from the interference of respiration.

### C. Impact of Practical Factors

In the following, we study the impact of different practical factors. Because all features can be inferred from waveform, we most focus on evaluating the cosine similarity that measures the overall performance of waveform recovery.

1) *Human Subject*: The performance of VED for all 12 subjects (with the first 4 involved in training) are presented in Fig. 12, showing a median higher than 0.9 for 10 subjects, and that for the rest still higher than 0.88. These results have confirmed that the performance of VED is virtually insensitive to the discrepancies among the human subjects, and is effective across all subjects who are not involved in training.

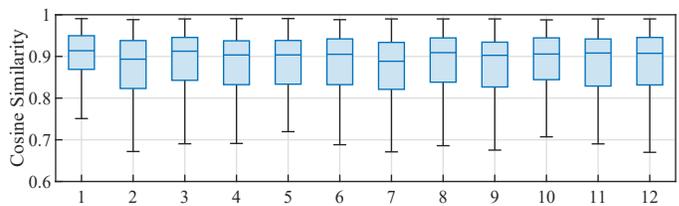


Fig. 12: Impact of different subjects.

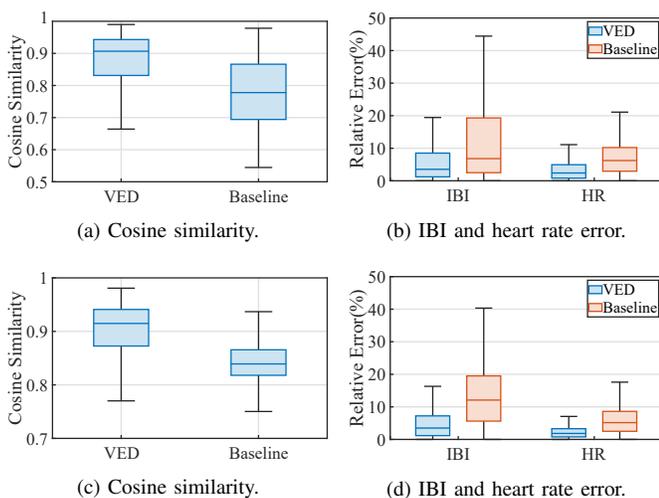


Fig. 11: Comparing the performance of IR-UWB (a), (b) and FMCW (c), (d) with the baseline.

2) *Latent Space Dimension*: As the latent space is a key to VED’s generalization ability, its dimension directly affects the performance of VED. A small number of the dimension can degrade the performance due to the limited capacity of a latent representation, but a large number can render VED harder to converge in training. Consequently, this dimension should be set to strike a balance between capacity and training efficiency. We train the VED with different latent dimensions and report the results in Fig. 13. According to the results, the performance improves with the dimension increasing from 16 to 64 and then starts to decrease gradually thereafter. Therefore, we choose 64 as the latent dimension in our implementation.

3) *Weight  $\gamma$  of the Loss Function*: The weight  $\gamma$  for the regularization loss in Eqn. (11) is a crucial parameter to be tuned for VED: increasing  $\gamma$  stresses on the latent space regularization, thus improving robustness; but doing so also sacrifices the accuracy of reconstruction. To determine the optimal weight to balancing the reconstruction and regularization

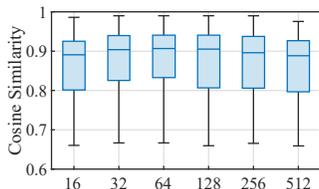


Fig. 13: Impact of the latent space dimension.

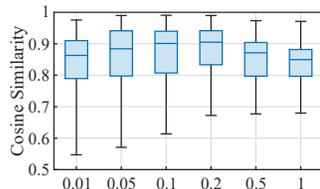


Fig. 14: Impact of the weight  $\gamma$  in the loss function.

losses, we evaluate the impact of different weight values. As shown in Fig. 14, increasing  $\gamma$  from 0.01 to 0.2 has positive effect on the cosine similarity, but the effect becomes negative afterward. Therefore we choose  $\gamma = 0.2$ , which works best in practice, for our implementation.

4) *Impact of Distance:* In real-life scenarios, the distance between a radar and subject often affects the sensing performance. To evaluate the distance impact, we let our subjects vary their distance from 1 m to 3 m and collect measurements accordingly. The results shown in Fig. 15 clearly indicate that distance has a negative effect on performance, because the reflected signal becomes weaker when the distance increases. However, even at the farthest distance of 3 m, VED achieves a median cosine similarity of 0.82, high enough for most applications, thus firmly proving the effectiveness of VED.

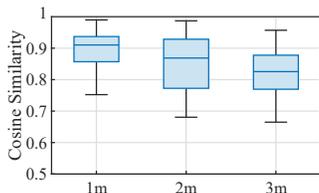


Fig. 15: Impact of distance.

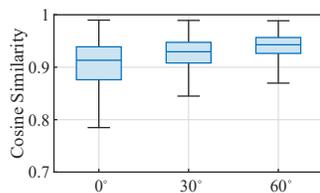


Fig. 16: Impact of orientation.

5) *Impact of Subject Orientation:* Apart from distance, the subject orientation could also critically impact the sensing ability. Therefore, we change the position of the radar so that its normal and that of the subject (sitting at 1 m distance) from three different angles:  $0^\circ$ ,  $30^\circ$ ,  $60^\circ$ . Given the results reported in Fig. 16, one may readily observe that the medians of cosine similarity at all three angles are above 0.91, indicating a successful heartbeat waveform recovery at all orientations. However, it might be a bit counter-intuitive that the recovery performance becomes better as the angle increases. We conjecture that this phenomenon can be attributed to the decreasing signal strength (chest motion) caused by respiration and relatively intact signal strength (neck artery micro-motion) incited by heartbeat: basically, heartbeat is less interfered with by breath when sensing at a larger angle.

6) *Impact of Minor Body Movements:* By far, we have conducted our evaluations under the condition that subjects all sit statically. However, minor body movements may be present in real-life settings, for example, subjects may shake their legs under natural states. Existing proposals simply suspend heartbeat monitoring (e.g., [10]), or rely on the same methods discussed in Sec. II-C to remove the interference caused

by body movements, which inevitably fail to recover fine-grained heartbeat waveform. Since VED can remove non-linear interference by design, we believe that VED can also be robust to minor body movements.

In order to verify this point, we try to recover heartbeat waveforms when subjects' bodies exhibit minor movements, such as swaying limb slowly (SW), playing phone (PP), and shaking leg (SL). We report the results obtained at 1 m distance in Fig. 17; these results indicate that SW has the least impact on heartbeat waveform recovery, yielding a median cosine similarity, IBI error, and HR error respectively of 0.89, 2.4%, and 1.3%, while PP and SL degrade the recovery performance slightly more due to their larger strength and randomness. Overall, all median cosine similarities are above 0.85, and all relative errors are under 15%, proving the robustness of VED to minor body movements.

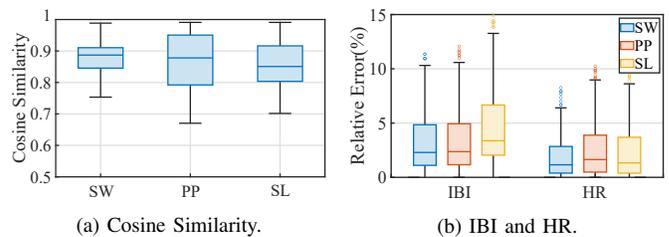


Fig. 17: Impact of minor body movements.

## V. CONCLUSION

In this paper, we have demystified and firmly addressed the question stated in the title: *whether contact-free RF sensing can capture fine-grained heartbeat waveform sensing?* We have first studied the principle of heartbeat sensing and analyzed the inadequacy of existing proposals. We have then revealed the true challenge as the heartbeat signal being mixed nonlinearly with respiration and thus overwhelmed by the latter. To solve this problem, we have proposed a deep learning model VED; it leverages both an encoder-decoder network and variational inference to perform nonlinear decomposition and robust waveform recovery. Our extensive evaluations have clearly demonstrated the promising potential of employing VED in fine-grained heartbeat monitoring under practical scenarios. Meanwhile, there are still a few extensions that we plan to implement in the future. On one hand, we would like to cooperate with medical institutions to further evaluate the performance of VED with real clinical datasets. On the other hand, we are on the way to develop multi-subject heartbeat monitoring leveraging the spatial diversity of multi-antenna and even multi-radar systems.

## REFERENCES

- [1] W. Chen and D. McDuff, "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks," in *Proc. of the 15th ECCV*, September 2018, pp. 349–365.
- [2] 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 *Proc. of the IEEE/CVF ICCV*, 2019, pp. 151–160.

- [3] A. Pai, A. Veeraraghavan, and A. Sabharwal, "HRVCam: Robust Camera-based Measurement of Heart Rate Variability," *J. Biomed Opt.*, vol. 26, pp. 1–23, 2021.
- [4] K. Qian, C. Wu, F. Xiao, Y. Zheng, Y. Zhang, Z. Yang, and Y. Liu, "Acousticcardiogram: Monitoring Heartbeats using Acoustic Signals on Smart Devices," in *Proc. of the 37th IEEE INFOCOM*, 2018, pp. 1574–1582.
- [5] T. Wang, D. Zhang, Y. Zheng, T. Gu, X. Zhou, and B. Dorizzi, "C-FMCW Based Contactless Respiration Detection Using Acoustic Signal," in *Proc. of the 20th ACM UbiComp*, 2018, pp. 170:1–20.
- [6] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, and M. Li, "BreathListener: Fine-Grained Breathing Monitoring in Driving Environments Utilizing Acoustic Signals," in *Proc. of the 17th ACM MobiSys*, 2019, p. 54–66.
- [7] A. Wang, J. E. Sunshine, and S. Gollakota, "Contactless Infant Monitoring Using White Noise," in *Proc. of the 26th ACM MobiCom*, 2019, pp. 52:1–16.
- [8] X. Xu, J. Yu, and Y. Chen, "Leveraging Acoustic Signals for Fine-grained Breathing Monitoring in Driving Environments," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2020.
- [9] T.-Y. J. Kao, Y. Yan, T.-M. Shen, A. Y.-K. Chen, and J. Lin, "Design and Analysis of a 60-GHz CMOS Doppler Micro-Radar System-in-Package for Vital-Sign and Vibration Detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 4, pp. 1649–1659, 2013.
- [10] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart Homes that Monitor Breathing and Heart Rate," in *Proc. of the 33rd ACM CHI*, 2015, pp. 837–846.
- [11] P. Nguyen, X. Zhang, A. Halbower, and T. Vu, "Continuous and Fine-Grained Breathing Volume Monitoring from Afar Using Wireless Signals," in *Proc. of the 35th IEEE INFOCOM*, 2016, pp. 1–9.
- [12] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren, "Cardiac Scan: A Non-Contact and Continuous Heart-Based User Authentication System," in *Proc. of the 23rd ACM MobiCom*, 2017, p. 315–328.
- [13] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Vital Sign and Sleep Monitoring using Millimeter Wave," *ACM Transactions on Sensor Networks*, vol. 13, no. 2, pp. 1–32, 2017.
- [14] T. Zheng, Z. Chen, C. Cai, J. Luo, and X. Zhang, "V<sup>2</sup>iFi: in-Vehicle Vital Sign Monitoring via Compact RF Sensing," in *Proc. of the 22nd ACM UbiComp*, 2020, pp. 70:1–27.
- [15] Mango Communications, "WARP v3," <https://warpproject.org/trac/>, 2013, accessed: 2021-05-30.
- [16] Novelda AS, "The World Leader in Ultra Wideband (UWB) Sensing," <https://novelda.com/technology/>, 2017, accessed: 2021-05-30.
- [17] Texas Instruments, "IWR1843BOOST," <https://www.ti.com/store/ti/en/p/product/?p=IWR1843BOOST>, 2020, accessed: 2021-05-30.
- [18] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-Invasive Detection of Moving and Stationary Human with WiFi," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2329–2342, 2015.
- [19] K. Niu, F. Zhang, Z. Chang, and D. Zhang, "A Fresnel Diffraction Model based Human Respiration Detection System using COTS Wi-Fi Devices," in *Proc. of the 20th UbiComp*, 2018, pp. 416–419.
- [20] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking Vital Signs During Sleep Leveraging Off-the-Shelf WiFi," in *Proc. of the 16th ACM MobiHoc*, 2015, p. 267–276.
- [21] A. Lazaro, D. Girbau, and R. Villarino, "Analysis of Vital Signs Monitoring Using an IR-UWB Radar," *Progress In Electromagnetics Research*, vol. 100, pp. 265–284, 2010.
- [22] S. K. Leem, F. Khan, and S. H. Cho, "Vital Sign Monitoring and Mobile Phone Usage Detection Using IR-UWB Radar for Intended Use in Car Crash Prevention," *MDPI Sensors*, vol. 17, no. 6, pp. 1240:1–25, 2017.
- [23] J. Muñoz-Ferreras, Z. Peng, R. Gómez-García, and C. Li, "Random Body Movement Mitigation for FMCW-radar-based Vital-sign Monitoring," in *Proc. of IEEE BioWireless*, 2016, pp. 22–24.
- [24] M. Zhao, F. Adib, and D. Katabi, "Emotion Recognition Using Wireless Signals," *Commun. ACM*, vol. 61, no. 9, p. 91–100, 2018.
- [25] Z. Chen, T. Zheng, C. Cai, and J. Luo, "MoVi-Fi: Motion-robust Vital Signs Waveform Recovery via Deep Interpreted RF Sensing," in *Proc. of the 27th ACM MobiCom*, 2021, pp. 392–405.
- [26] W. Hu, Z. Zhao, Y. Wang, H. Zhang, and F. Lin, "Noncontact Accurate Measurement of Cardiopulmonary Activity Using a Compact Quadrature Doppler Radar Sensor," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 725–735, 2014.
- [27] C. Will, K. Shi, S. Schellenberger, T. Steigleder, F. Michler, J. Fuchs, R. Weigel, C. Ostgathe, and A. Koelplin, "Radar-based Heart Sound Detection," *Nature Scientific Reports*, vol. 8, no. 1, pp. 1–14, 2018.
- [28] Z. Xia, M. M. H. Shandhi, O. T. Inan, and Y. Zhang, "Non-Contact Sensing of Seismocardiogram Signals Using Microwave Doppler Radar," *IEEE Sensors Journal*, vol. 18, no. 14, pp. 5956–5964, 2018.
- [29] J. E. Hall and M. E. Hall, *Guyton and Hall Textbook of Medical Physiology Hardcover*. Elsevier Health Sciences, 2020.
- [30] K. E. Barrett, S. Boitano, S. M. Barman, and H. L. Brooks, *Ganong's Review of Medical Physiology*. McGraw Hill, 2010.
- [31] M. D. Rienzo, G. Rizzo, Z. M. Islay, and P. Lombardi, "SeisMote: A Multi-SensorWireless Platform for Cardiovascular Monitoring in Laboratory, Daily Life, and Telemedicine," *MDPI Sensors*, vol. 20, no. 3, pp. 680:1–15, 2020.
- [32] J. Allen, "Photoplethysmography and its Application in Clinical Physiological Measurement," *Physiological Measurement*, vol. 28, no. 3, pp. 1–39, feb 2007.
- [33] NeuLog, "Heart Rate & Pulse Logger Sensor NUL-208," <https://neulog.com/heart-rate-pulse/>, 2017, accessed: 2021-05-30.
- [34] S. Ding, Z. Chen, T. Zheng, and J. Luo, "RF-Net: A Unified Meta-Learning Framework for RF-Enabled One-Shot Human Activity Recognition," in *Proc. of the 18th ACM SenSys*, 2020, pp. 517–530.
- [35] T. Zheng, Z. Chen, S. Ding, and J. Luo, "Enhancing RF Sensing with Deep Learning: A Layered Approach," *IEEE Communications Magazine*, vol. 59, no. 2, pp. 70–76, 2021.
- [36] Johns Hopkins Medicine, "Vital Signs (Body Temperature, Pulse Rate, Respiration Rate, Blood Pressure)," <https://www.hopkinsmedicine.org/health/conditions-and-diseases/vital-signs-body-temperature-pulse-rate-respiration-rate-blood-pressure>, 2021, accessed: 2021-05-30.
- [37] R. S. Crow, P. Hannan, D. Jacobs, L. Hedquist, and D. M. Salerno, "Relationship Between Seismocardiogram and Echocardiogram for Events in the Cardiac Cycle," *American journal of noninvasive cardiology*, vol. 8, pp. 39–46, 1994.
- [38] SparkFun, "SparkFun Single Lead Heart Rate Monitor - AD8232," <https://www.sparkfun.com/products/12650>, 2014, accessed: 2021-05-30.
- [39] Z. Wu and N. E. Huang, "Ensemble Empirical Mode Decomposition: A Noise-assisted Data Analysis Method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [40] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," *IEEE Trans. on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [41] D. R. K. Brownrigg, "The Weighted Median Filter," *Communications of the ACM*, vol. 27, no. 8, p. 807–818, 1984.
- [42] A. Carini, S. Cecchi, M. Gasparini, and G. L. Sicuranza, "Introducing Legendre Nonlinear Filters," in *Proc. of IEEE ICASSP*, 2014, pp. 7939–7943.
- [43] S. Aslan, L. Hocke, N. Schwarz, and B. Frederick, "Extraction of the Cardiac Waveform from Simultaneous Multislice fMRI Data Using Slice Sorted Averaging and a Deep Learning Reconstruction Filter," *NeuroImage*, vol. 198, pp. 303–316, 2019.
- [44] Y. Luo and N. Mesgarani, "TasNet: Time-Domain Audio Separation Network for Real-Time, Single-Channel Speech Separation," in *Proc. of IEEE ICASSP*, 2018, pp. 696–700.
- [45] S. I. Mimitakis, K. Drossos, T. Virtanen, and G. Schuller, "A Recurrent Encoder-Decoder Approach with Skip-Filtering Connections for Monaural Singing Voice Separation," in *Proc. of the 27th IEEE MLSP*, 2017, pp. 1–6.
- [46] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational Inference: A Review for Statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.
- [47] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *Proc. of ICLR*, 2013, pp. 1–14.
- [48] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep Complex Networks," in *Proc. of ICLR*, 2018, pp. 1–19.
- [49] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object Recognition with Gradient-Based Learning," in *Shape, Contour and Grouping in Computer Vision*. Springer, 1999, pp. 319–345.
- [50] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proc. of the 32nd ICML*, 2015, pp. 448–456.
- [51] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in *Proc. of the 14th AISTATS*, 2011, pp. 315–323.
- [52] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional Networks," in *Proc. of the 23rd IEEE CVPR*, 2010, pp. 2528–2535.
- [53] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.