

Inferring Electrocardiography From Optical Sensing Using Lightweight Neural Network

Yuenan Li, *Senior Member, IEEE*, Xin Tian, *Member, IEEE*, Qiang Zhu, *Member, IEEE*,
and Min Wu, *Fellow, IEEE*

Abstract—This paper presents a computational solution that enables continuous cardiac monitoring through cross-modality inference of electrocardiogram (ECG). While some smartwatches now allow users to obtain a 30-second ECG test by tapping a built-in bio-sensor, these short-term ECG tests often miss intermittent and asymptomatic abnormalities of cardiac functions. It is also infeasible to expect persistently active user participation for long-term continuous cardiac monitoring in order to capture these and other types of cardiac abnormalities. To alleviate the need for continuous user attention and active participation, we design a lightweight neural network that infers ECG from the photoplethysmogram (PPG) signal sensed at the skin surface by a wearable optical sensor. We also develop a diagnosis-oriented training strategy to enable the neural network to capture the pathological features of ECG, aiming to increase the utility of reconstructed ECG signals for screening cardiovascular diseases (CVDs). We also leverage model interpretation to obtain insights from data-driven models, for example, to reveal some associations between CVDs and ECG/PPG and to demonstrate how the neural network copes with motion artifacts in the ambulatory application. The experimental results on three datasets demonstrate the feasibility of inferring ECG from PPG, achieving a high fidelity of ECG reconstruction with only about 40K parameters.

Impact Statement—Existing wearable ECG monitors require users to wear adhesive patches or tap a designated area of a smartwatch during sensing, which is uncomfortable or infeasible for long-term cardiac monitoring. As optical sensors have become ubiquitous in wearable devices, this paper proposes a lightweight neural network for inferring ECG from the optical measure of the pulse wave induced by heartbeat, enabling a wearable device to function as an ECG monitor. This work can support long-term ECG monitoring without having users suffer from adhesive-related skin irritation associated with multi-day recorder patch, or constantly hold on a sensor. The learned model has the potential to build the physiological digital twin that facilitates personalized cardiovascular healthcare. The exploration of model interpretation reveals the influence of heart diseases on physiological signals, and the insights obtained through the proposed mechanism help identify diagnostic markers.

Index Terms—Cross-modality inference, physiological digital twin, electrocardiogram (ECG), neural network, photoplethysmogram (PPG), tele-health.

This work was supported in part by NSF under Grant 2124291. (*Corresponding author: Min Wu.*)

Yuenan Li is with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (email: ynli@tju.edu.cn). This work was initiated when Yuenan Li was a visiting researcher at the University of Maryland, College Park.

Xin Tian and Qiang Zhu were with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA, where the work was carried out, and are now with Meta Inc., Menlo Park, CA 94025 USA (e-mail: xtian17@terpmail.umd.edu, zhuqiang@terpmail.umd.edu).

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

I. INTRODUCTION

CARDIOVASCULAR diseases (CVDs) are the most prevalent causes of mortality. According to the statistics in a study conducted in 2019 [1], one person dies from CVDs every 37 seconds in the United States. Early treatment can effectively reduce the risk of sudden cardiac death. However, some CVDs, such as heart muscle dysfunction, show no obvious symptoms in the early stage. The presence of symptoms usually indicates the onset of heart failure. A study conducted in the aged population shows that around one-third to one-half of heart attacks are clinically unrecognized [2]. The unawareness of diseases makes some patients lose the opportunity to receive early medical intervention.

Electrocardiogram (ECG) is an essential tool for non-invasive diagnosis of CVDs. The patients at higher risk, such as the aging population, can benefit from ECG monitoring. Among the currently available options for continuous ECG monitoring, the Holter monitor is bulky to wear; newer devices attached to the chest with adhesives, such as the Zio Patch, are lightweight, but the prolonged use of adhesives with multi-day monitoring may increase the risk for skin irritations, especially for persons with sensitive skins. These patch-type sensors may slide or fall off under excessive sweating, or restrict patterns from certain activities. Recent technical advances have integrated bio-sensors into smart wearable devices. For example, using the crown and back crystal as electrodes, an Apple Watch allows its user to take ECG tests for up to 30 seconds at a time from the wrist by tapping the crown. Still, asymptomatic and intermittent events could be missed, and continuous user participation by keeping his/her hand on the sensor is impractical. It is desirable that smart wearable devices can continuously monitor cardiac conditions for extended periods without requiring users' active participation.

Researchers have been working toward the above objective through cross-modality signal inference. The pilot study of Zhu *et al.* explored the possibility of inferring ECG from photoplethysmography (PPG) [3], [4]. PPG manifests the variation of blood volume caused by the movements of the heart muscle. The signal can be sensed by an optical sensor attached to the wrist or finger, without a user consciously participating all the time. Since PPG carries useful vital signs, the miniaturized PPG sensor has become an integral part of smart wearable devices. Using artificial intelligence (AI) and machine learning (ML) algorithms to augment PPG sensing data for ECG monitoring can alleviate the need to re-design bio-sensor hardware and mitigate continuous user attention to

carry out the sensing. The software Apps can be seamlessly integrated into existing devices. In light of the advantages of optical sensing in convenience, availability, and cost, some novel PPG-based biomedical schemes have been developed to complement the conventional ECG-based ones, aiming to utilize cellphones or smartwatches to monitor cardiac health at a larger scale. For example, the Apple Heart Study uses smartwatches to detect atrial fibrillation from intermittent wrist PPG [5].

Exploiting PPG sensors for long-term ECG monitoring also enables in-home health management and expands the geographic reach of cardiologists. Effective techniques for reconstructing ECG from PPG have the potential to provide general users with the opportunities of long-term ECG monitoring (e.g., for weeks or months long), and those with chronic cardiovascular problems and lifetime risk-assessment needs are among the primary beneficiaries. This class of techniques can enable the creation of the cardio-physiological digital twin, helping cardiologists dynamically monitor the status of a patient's heart over time and provide timely and targeted medical assistance [6].

More specifically, ECG and PPG are two sensing modalities of the same physiological process recorded by electrical and optical measures, respectively. Their correlation establishes the feasibility of estimating one from the other. Previous studies validate that the vital signs derived from PPG and ECG show strong agreement [7]. As will be seen in this paper, our analytical results show that the causal influence of the heart on blood circulation can be clearly observed from PPG and ECG waveforms. In this work, we leverage deep learning to infer ECG from PPG, aiming to achieve low-cost, user-friendly, and interpretable continuous cardiac monitoring. The main contributions of this work are summarized as follows:

- 1) We propose a lightweight neural network for deriving ECG from PPG. The network captures the correlations between ECG and PPG at multiple scales by taking advantage of the interactions among the convolutional layers with different receptive fields. For faithful representations of pathological ECG patterns, we use a diagnosis-oriented training algorithm to regularize the distribution of reconstructed ECG signals in a feature space that is discriminative of CVDs.

- 2) Considering the resource-constrained nature of wearable devices, we propose a model compression algorithm to further lower the memory consumption and computational complexity of ECG reconstruction. The knowledge learned by the original network is transferred to a compressed one via attention-based knowledge distillation. To the best of our knowledge, this is the first attempt at using a lightweight neural network to enable the cross-modality inference of ECG.

- 3) Beyond algorithm design, we focus on explainability. We use gradient-based interpretation to check if the CVD-related features of ECG learned from data for regularizing ECG reconstruction are clinically plausible. This work also addresses the ambulatory application, and the influence of motion on PPG-based ECG inference is examined using causal analysis. Based on that, we take advantage of the motion information sensed by wearable devices to enhance the robustness of ECG reconstruction during exercise, and the effects of the auxiliary

information are analyzed via model interpretation.

The rest of this paper is organized as follows. Section II briefly reviews related work. Section III elaborates on the neural network architecture, training, and model compression algorithms. Experimental results and discussions are presented in Section IV. Finally, Section V presents conclusions.

II. RELATED WORK

The research on PPG-based ECG inference is still in its infancy. A few prior studies have been dedicated to this problem [3], [4], [8]–[14]. Going beyond the previous capability of mainly estimating the parameters of ECG from PPG [15], Zhu *et al.*'s work first presents an interpretable framework based on signals and systems theories and demonstrates the feasibility of generating ECG waveforms from PPG sensor via computational approach [3]. This work translates PPG to ECG in the Discrete Cosine Transform (DCT) domain using linear regression. Tian *et al.* cast PPG-to-ECG mapping as a cross-domain sparse coding problem [8]. The algorithm learns PPG and ECG dictionaries and a linear transform for domain translation simultaneously. Reference [9] presents a more discriminative variant of the sparse-coding-based algorithm, where dictionary learning is regularized by CVD labels to enhance the sensitivity of the sparse codes to pathological ECG patterns. A heartbeat comprises a sequence of short-term actions that determine the local morphologies of PPG and ECG. Hence, hand-crafted models have limited flexibility in capturing the multi-scale correlation between PPG and ECG due to using linear mapping or global bases in waveform analysis and synthesis.

Some recent studies leverage the strong expressive power of neural networks to enhance the accuracy of ECG reconstruction. Chiu *et al.* proposed a neural-network-based end-to-end algorithm [10]. The algorithm first uses transformation-and-attention networks to modulate raw PPG and then uses a pair of encoder and decoder to synthesize ECG. The training scheme emphasizes the QRS complex using a region-of-interest (ROI) enhanced loss. An adversarial learning framework was designed to learn the PPG-to-ECG mapping [11], where the ECG generator attempts to deceive a discriminator by making synthesized waveforms indistinguishable from the realistic ones. Tang *et al.* constructed a W-Net to infer ECG from PPG by cascading double U-Nets [12]. Bi-directional long short-term memory networks (LSTMs) were also adopted to analyze the temporal features of PPG and extract the cues for reconstructing ECG [13]. Lan's work uses a Transformer-based architecture to analyze the long-range dependencies between PPG and ECG [14], and the reconstructed ECG is combined with PPG for CVD diagnosis. Recently, we have also seen growing research interest in inferring ECG from other sensing modalities. For instance, it has been shown that ECG can be estimated from ballistocardiogram (BCG), which is a micro-signal capturing the slight body motion resulting from the abrupt ejection of blood during heartbeats. Deep learning algorithms were developed to reconstruct ECG from the body surface displacement measured by a wearable gyroscope or continuous-wave radar [16]–[18].

Neural networks are more effective than hand-crafted models in detecting the traces of short-term heart activities from waveforms, enabling the ECG reconstruction at a higher fidelity. However, deep neural networks often have high demands on storage and computation resources, which hinders mobile applications. Most prior efforts prioritized the fidelity of reconstructed ECG and have not given sufficient attention to the utility of ECG inference for CVD diagnosis. In particular, the explainability of deep-learning-based PPG-to-ECG inference remains unexplored. Existing studies do not elucidate the relationships between PPG and CVDs, while such knowledge is crucial for understanding the impact of CVDs on heart function. The above-mentioned issues motivate us to enhance the model compactness, efficiency, fidelity, and explainability of ECG inference while benefiting downstream clinical applications.

In addition to PPG-based ECG inference, deep learning has been extensively adopted in cardiac signal processing, such as automated PPG and ECG interpretation [19]–[22], artifacts removal [23], waveform synthesis [24], vital sign measurement [25], [26], heart function assessment [27], and biometric recognition [28]. Hannun *et al.* trained a deep neural network for detecting rhythm-based disorder [19]. The network can classify 12 kinds of arrhythmia from single-lead ECG at cardiologist-level accuracy. Deep learning was also applied to monitor the aggravation of heart diseases from ECG [21]. To improve the accuracy of patient-specific CVD diagnosis, Golany *et al.* developed a generative adversarial network (GAN) for synthesizing the ECG waveforms of a patient [24]. Deep learning also facilitates the measurement of vital signs. It has been demonstrated that blood pressure may be inferred from PPG using a deep belief network [25], making it promising to monitor continuous blood pressure in a cuffless way.

III. METHODS

A. Physiological Background and Problem Formulation

ECG measures the electrical signal generated by the depolarization and re-polarization of heart muscle cells. These activities are triggered by an electrical stimulus originating from the SA node (i.e., the pacemaker of the heart). The stimulus coordinates the contracting and expanding movements of the heart, which are the driving force of blood circulation.

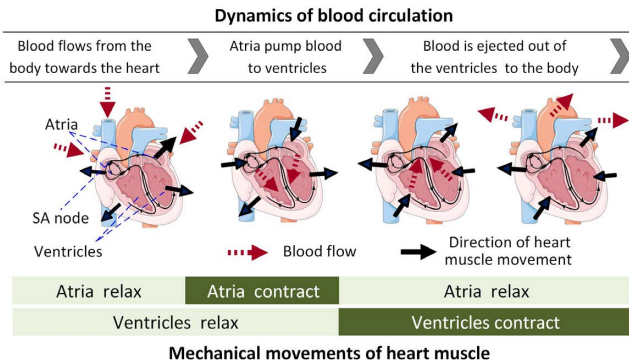


Fig. 1. Phases of a cardiac cycle.

Fig. 1 shows the major events during a cardiac cycle¹. A cardiac cycle begins when the four chambers of the heart relax. The two upper chambers (i.e., atria) expand to receive blood from the body. The stimulus first triggers the depolarization of the atria, resulting in the P-wave on ECG. The depolarization causes the atria muscle to contract and pump blood into the two bottom chambers (i.e., ventricles). The electrical stimulus then transmits to the ventricles through the conducting pathway and generates the QRS complex on ECG. As the ventricles contract, blood is ejected out of the heart. More specifically, the left ventricle pumps blood to vessels of the body. The increase of blood volume in the vessels gives rise to an ascending slope on PPG. After that, the ventricles start to relax, and the T-wave on ECG depicts this phase. Finally, both the atria and ventricles relax, so the pressure within the heart drops rapidly, and a new cycle is about to start. As a result, blood flows back toward the atria, leaving a descending slope on PPG. Fig. 2 depicts the traces of several key cardiac events on ECG and the corresponding blood circulation reflected by PPG.

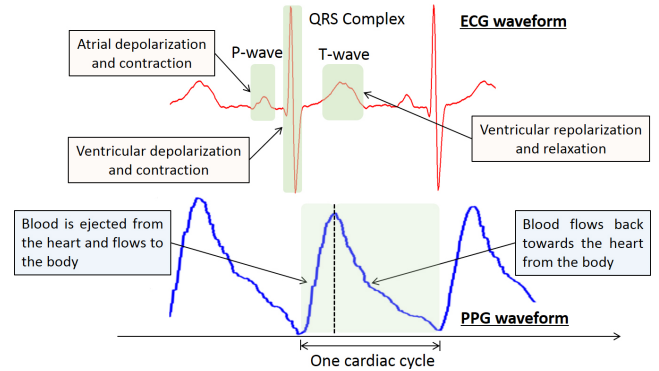


Fig. 2. Association between the electrical activities of the heart and the blood flow dynamics represented by ECG and PPG.

Consider a simplified conceptual model of the ECG (E) and PPG (P) signals. Denoting by δ the electrical stimulus that initiates a heartbeat, we have:

$$E = H_E(\delta) + N_E, \quad (1)$$

$$P = H_P(\delta) + N_P, \quad (2)$$

where $H_E(\cdot)$ and $H_P(\cdot)$ are the impulse response functions describing the electrical and mechanical activations of the heart, respectively, and N_E and N_P are sensing noise. Without direct access to $H_E(\cdot)$ and $H_P(\cdot)$, inferring ECG from PPG is an under-determined inverse problem. In this work, we solve this problem by training a lightweight neural network $G_{P \rightarrow E}(\cdot)$. Instead of resorting to a highly simplified model with a universal basis of DCT and linear mapping in the prior art [3], we harness data in capturing potentially complex relations while striving to keep the model compact and explainable.

¹The heart images in Fig. 1 and Fig. 3 are adopted from Servier Medical Art at <https://smart.servier.com>, licensed under a Creative Commons Attribution 3.0 unported license.

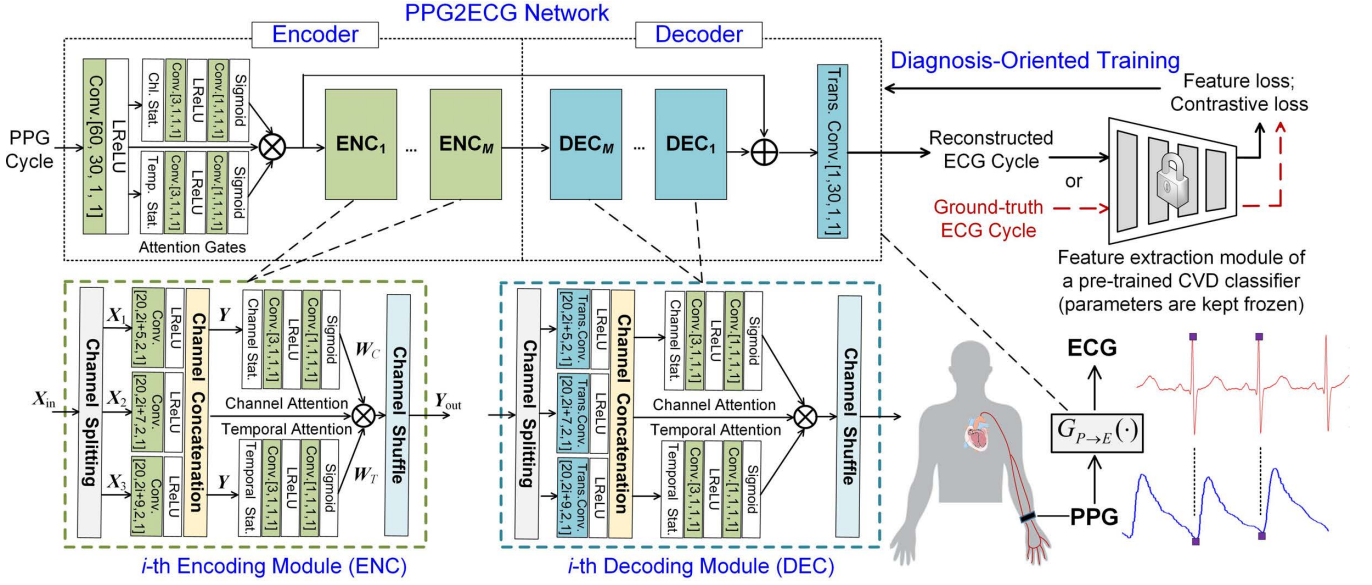


Fig. 3. Architecture of the PPG2ECG network and the diagnosis-oriented training scheme. Parameter settings of the network are shown in the figure. The dashed boxes below the PPG2ECG network show the architectures of ENCs and DECs. The numbers of ENCs and DECs were set to $M = 4$. In each ENC or DEC, input feature maps are split into $G = 3$ groups. We use the notation $[N, K, D, S]$ to represent the parameter setting of a 1D convolutional (or transposed convolutional) layer, where N is the kernel number, K is the kernel length, D is the dilation, and S is the stride.

B. Signal Preprocessing

The training ECG and PPG sequences are preprocessed using the same procedures as the prior work of Zhu *et al.* [3], [4] and Tian *et al.* [8], [9]. We take the moment when the ventricles contract as the anchor point for PPG-ECG synchronization, where the onset points of PPG are aligned to the R-peaks of ECG. The detrending algorithm adopted by Zhu *et al.* [3] is then applied on aligned sequences to eliminate the slow-varying trends introduced by breathing, motion, etc. The detrended sequences are partitioned into cycles. Each cycle starts at an onset point of PPG or an R-peak of ECG. The PPG and ECG cycles are then interpolated to fixed length as $P \in \mathbb{R}^L$ and $E \in \mathbb{R}^L$, respectively.

C. Neural Network Architecture

At the core of PPG-based ECG reconstruction (a.k.a. PPG2ECG) is mining the intrinsic correlation between PPG and ECG. The cardiac events within a heartbeat are of different durations, and the electro-mechanical activities of the heart are multi-scale in nature [29]. For example, the contraction of the atria lasts about 0.1s, while the blood ejection phase of the ventricles has a much longer duration of about 0.4s. To effectively capture their representations on PPG and ECG, the neural network needs to explore the signal space at various scales. Hence, a key challenge in designing a lightweight PPG2ECG model is maintaining a strong capability of multi-scale feature learning/fusion under a strict constraint on parameter budget.

The proposed PPG2ECG network transforms PPG to ECG using an encoder-decoder architecture. As Fig. 3 shows, the network consists of a stack of multi-scale feature extraction and waveform reconstruction modules, and the building blocks are referred to as encoding and decoding modules. It is worth noting that such an encoding-and-decoding framework, also known as an analysis-and-synthesis framework, is common to many representation and inference tasks in learning as well

as signal processing. Using neural networks to accomplish the encoding/analysis and decoding/synthesis offers flexibility beyond simple models (such as linear and/or parametric ones). The encoder progressively aggregates input PPG to latent codes. Each module characterizes the local morphologies of PPG at multiple temporal resolutions, aiming to detect the short-term and long-term influence of heart activities on blood circulation. The decoder synthesizes ECG using the latent codes from coarse to fine scales. The encoder first uses a 1D convolutional layer and attention gates to extract the primary features of PPG. A stack of encoding modules (ENC) then extracts from PPG the information that is useful for synthesizing ECG. In each ENC, the input feature map $X_{in} \in \mathbb{R}^{C \times L_{in}}$ is first split along the channel direction into G non-overlapping groups, as $\{X_i \in \mathbb{R}^{C/G \times L_{in}} | i = 1, \dots, G\}$. The groups are fed to G 1D convolutional layers (see the detailed view of ENC in Fig. 3) whose kernel lengths increase with a step size of 2 from the top-down direction. The outputs of the convolutional layers, which are of the same length, are concatenated along the channel direction:

$$Y = [C_1(X_1), \dots, C_G(X_G)] \in \mathbb{R}^{C \times L_{out}}, \quad (3)$$

where $C_i(\cdot)$ is the i -th convolutional layer, and $[\dots]$ denotes the concatenation operation. Y encodes the temporal characteristics of the input extracted at G different scales.

A major benefit of grouped convolutions is reducing parameters. Take an ENC without channel splitting (i.e., $G = 1$) as a reference. Assume that both the input and output feature maps have C channels and the lengths of 1D convolutional kernels are K . Then the number of parameters is $P_1 = C^2 K$, where we omit the bias terms for simplicity. For an ENC with G groups ($G > 1$), each group takes C/G channels as input and outputs the same number of channels. If we increase the kernel length by two at a time starting at K , the total number of parameters becomes

$$P_G = \sum_{i=1}^G [K + 2(i-1)] \left(\frac{C}{G}\right)^2 = \frac{(K+G-1)C^2}{G} \approx \frac{P_1}{G}. \quad (4)$$

We set $G = 3$ and $K \geq 7$, so the grouped convolution has about $1/3$ the parameter amount compared to the one without channel splitting.

Two attention gates are placed after the convolutional layers in each ENC to re-weight the feature map $\mathbf{Y} \in \mathbb{R}^{C \times L_{\text{out}}}$ along the temporal and channel directions. Since blood circulation is driven by the instantaneous activation of the heart muscle, the cues for inferring the activities of the heart do not spread uniformly over the PPG cycle. The PPG2ECG network uses a statistics-based temporal attention gate to highlight the parts of PPG that are informative for synthesizing ECG. We compute the statistics of each column of $\mathbf{Y}$, including the mean value, max value, and standard deviation. Two 1D convolutional layers are applied to the column-wise statistics to generate weights $\mathbf{W}_T \in [0, 1]^{L_{\text{out}}}$. Similarly, the channel attention gate learns to re-weight $\mathbf{Y}$ based on row-wise statistics. The statistics reflect the distribution of the features extracted by each convolutional kernel, according to which the attention gate assigns a weight to each channel, giving $\mathbf{W}_C \in [0, 1]^C$. The feature map is modulated as follows:

$$\mathbf{Y}_{\text{out}} = \mathbf{Y} \otimes \mathbf{W}_T \otimes \mathbf{W}_C, \quad (5)$$

where $\otimes$ is the element-wise multiplication in the corresponding dimension. The attention gates are also appended at the end of the first convolutional layer.

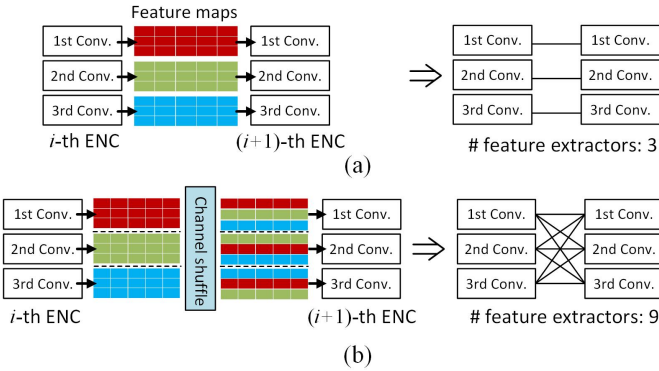


Fig. 4. Effect of channel shuffle. (a) and (b) illustrate the cases without and with channel shuffle, respectively, and the figures at the right illustrate the equivalent connection among the convolutional layers in neighboring ENCs.

The modulated feature map is shuffled along the channel direction before being fed to the next ENC (see the detailed view of ENC in Fig. 3), and we use the transpose-based shuffle operation [30]. The shuffle layer is parameter-free and can increase the richness of features by virtually connecting all the convolutional layers in different ENCs. Fig. 4 demonstrates the effect of channel shuffle. As Fig. 4(a) shows, without channel shuffle, the output of a convolutional layer in the first ENC only flows to the corresponding layer in the next ENC. Accordingly, stacking ENCs results in three parallel convolutional pipelines, which is equivalent to independently

applying three feature extractors over the input. By contrast, the shuffle operation gives the output of a convolutional layer equal opportunities to flow to all the convolutional layers in the next ENC, as illustrated in the right part of Fig. 4(b). By virtually connecting the convolutional layers in sequential ENCs, the channel splitting and shuffle operations increase the number of pathways for data flow within the neural network, and each pathway serves as a feature extractor. Since the convolutional layers in each ENC are of different kernel sizes, the neural network extracts features from PPG at various scales as the input flows through these pathways, and the characteristics of output ECG can be inferred at multiple granularities as well. As mentioned earlier, the movements of the atria and ventricles within a heartbeat are highly variable in time duration. Wider coverage of scales boosts the feature learning capability of the PPG2ECG neural network, which is beneficial for analyzing the responses of heartbeats on PPG.

The decoding modules (DECs) have the same architecture as ENCs, except that they use transposed convolutional layers to expand input feature maps. The output of the last DEC is fed to a transposed convolutional layer to reconstruct ECG. As the input PPG goes through cascaded ENCs and DECs, the high-resolution timing information about cardiac events carried by the original waveform is attenuated. To compensate, we follow the best practice of deep neural network design, as seen in the ResNet development [31], and link the first convolutional layer in the encoder and the final ECG reconstruction layer with a residual connection.

D. Diagnosis-Oriented Training Algorithm

Apart from fidelity, the PPG2ECG network also needs to guarantee the effectiveness of reconstructed ECG waveforms in screening CVDs. In automated and manual diagnosis, not all the ECG sample points contribute equally to the decision, and some CVDs cause local abnormalities. For example, myocardial infarction (a heart attack caused by the obstruction of the blood supply to the heart) sometimes manifests as an elevated ST segment. The reconstructed waveform should preserve such diagnostic features, but the norm between reconstructed ECG and the ground truth cannot emphasize the clinically significant parts of ECG. Therefore, it is necessary to use a task-driven loss to regularize ECG generation. In this work, we use the prior knowledge about CVDs learned from data to regularize the PPG2ECG network. We first train a classifier to detect CVDs from ECG, and the intermediate features, which are more sensitive to pathological ECG patterns, are exploited for training the PPG2ECG network. As validated by ablation experiments, the diagnosis-oriented training scheme makes reconstructed ECG waveforms show higher fidelity and more accurate diagnostic results (on both deep-learning-based and conventional CVD classifiers).

The architecture of the classifier is shown in Fig. 5. The classifier is composed of cascaded convolutional layers, a squeeze-and-excitation attention gate [32], and fully-connected (FC) layers with softmax output. The features modulated by the attention gate are flattened to a vector and then fed to the fully-connected layers to infer the probability of each disease.

The classifier is trained to minimize the cross-entropy loss between the predicted probability vector $\mathbf{p}$ of CVDs and the one-hot vector $\mathbf{l}$ of the ground-truth label.

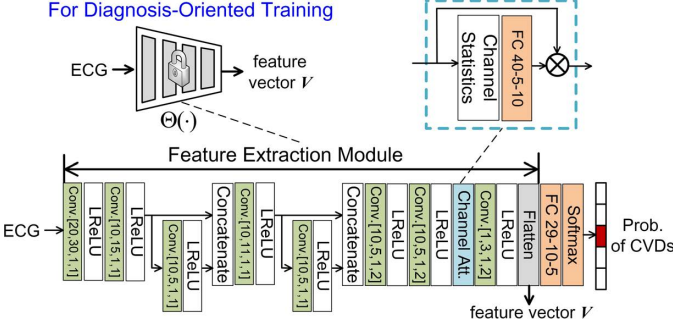


Fig. 5. Architecture of the CVD classifier for regularizing ECG inference.

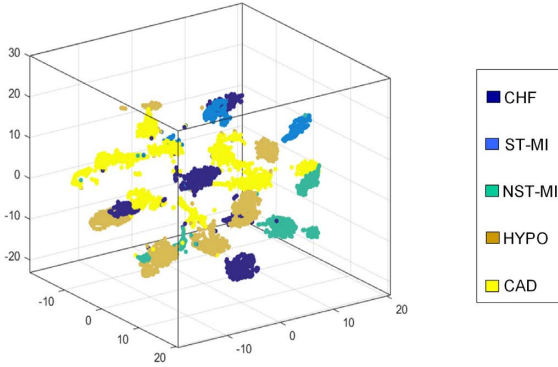


Fig. 6. Distribution of ECGs in the feature space.

Fig. 6 visualizes the distribution of the features extracted from the ECGs corresponding to five CVDs, including congestive heart failure (CHF), ST-elevated myocardial infarction (ST-MI), non-ST-elevated myocardial infarction (NST-MI), hypotension (HYPO), and coronary artery disease (CAD). The flattened feature vectors generated by the classifier were plotted in the 3D space using the t-distributed stochastic neighbor embedding (t-SNE) algorithm [33], and the classifier was trained using the ECG cycles in the Medical Information Mart for Intensive Care III (MIMIC-III) dataset [34]. We see that the features are discriminative, and those extracted from the ECGs corresponding to different CVDs are located in distinct clusters.

Considering the regularization effect of the classifier, it is desirable that the rationale behind its decision and the learned knowledge agree with clinical findings. We interpret the classifier using the axiomatic-attribution-based approach. We adopt the integrated gradient (IG) [35] method to calculate the contribution of each ECG sample point to the decision made by the classifier. Let us define by $\psi_j(\cdot) : \mathbb{R}^L \rightarrow [0, 1]$ the mapping from the input ECG $\mathbf{E} \in \mathbb{R}^L$ to the j -th dimension of the classifier's output (the probability of the j -th disease). The IG value of $\mathbf{E}[i]$ with respect to the disease is:

$$\text{IG}_i^j = \mathbf{E}[i] \times \int_0^1 \frac{\partial \psi_j(\alpha \mathbf{E})}{\partial \mathbf{E}[i]} d\alpha, \quad (i = 1, \dots, L). \quad (6)$$

The amplitude of IG_i^j quantifies the importance of $\mathbf{E}[i]$ to predicting the j -th disease. We use the Riemann sum to approximate the integral.

Fig. 7(a) plots an ECG cycle labeled as NST-MI and the amplitudes of IG values (represented by red bars). The contributions of ECG sample points to diagnosis are highly uneven, and the classifier pays more attention to several key segments, such as T-wave and P-wave. Note that the ECG cycle has an inverted T-wave, and this abnormal pattern is indicative of ischemia (i.e., decreased blood flow to the heart as a consequence of MI) [36]. We also inspected the intermediate features learned by the classifier. The feature maps produced by the attention gate are averaged over the channel direction and plotted in Fig. 7(b). We observe that the abnormal T-wave shows strong responses in feature maps. The results of model interpretation demonstrate the locality of the clinically significant patterns of ECG and the high sensitivity of the classifier to such patterns. A single per-point distance metric (e.g., the ℓ_1 norm between reconstructed and target ECG cycles), which gives equal weights to all sample points, cannot emphasize such patterns. To facilitate downstream diagnosis applications, we take advantage of the discriminating power of the classifier. The intermediate features are exploited to help the PPG2ECG network accurately represent the disease-related characteristics of ECG. After training the classifier, we freeze its parameters to define diagnosis-aware training objectives. Denote the feature extraction module of the pre-trained classifier (before the fully-connected layers) by $\Theta(\cdot)$ (see Fig. 5). Given a pair of reconstructed ECG cycle $\hat{\mathbf{E}}$ and the target $\mathbf{E}$, we compute their feature loss under $\Theta(\cdot)$:

$$\mathcal{L}_F = \|\Theta(\mathbf{E}) - \Theta(\hat{\mathbf{E}})\|_2^2. \quad (7)$$

Besides pulling $\hat{\mathbf{E}}$ close to $\mathbf{E}$ in the feature space, the training algorithm also pushes it far away from those with different CVD labels. As in contrastive learning [37], when CVD labels are available, we randomly sample N negative examples $\{\mathbf{E}_i^- | i = 1, \dots, N\}$ (i.e., the ECG cycles whose labels differ from that of $\mathbf{E}$) and compute their features using $\Theta(\cdot)$. Let us denote the features of $\hat{\mathbf{E}}$, $\mathbf{E}$, and the negative examples as $\hat{\mathbf{V}}$, $\mathbf{V}$, and $\{\mathbf{V}_i^- | i = 1, \dots, N\}$, respectively, and the features are all normalized to unit norm. To prevent class confusion, we use the following contrastive loss to regularize the distribution of ECG cycles in the feature space:

$$\mathcal{L}_C = \frac{-\exp(\hat{\mathbf{V}} \cdot \mathbf{V})}{\exp(\hat{\mathbf{V}} \cdot \mathbf{V}) + \sum_{i=1}^N \exp(\hat{\mathbf{V}} \cdot \mathbf{V}_i^-)}, \quad (8)$$

where $\cdot$ represents the dot product. In summary, the loss function for training the ECG reconstruction network is as follows:

$$\mathcal{L} = \|(\hat{\mathbf{E}} - \mathbf{E}) \otimes (\mathbf{1} + \mathbf{w})\|_1 + \lambda_F \mathcal{L}_F + \lambda_C \mathcal{L}_C, \quad (9)$$

where $\lambda_F = 2$ and $\lambda_C = 0.5$ are constant weights. As the training objective designed by Chiu *et al.* [10], we use a weighting vector $\mathbf{w} \in \mathbb{R}^L$ to enhance the QRS complex, and $\mathbf{1}$ is an all-ones vector. The weights in $\mathbf{w}$ are computed using a Gaussian function centered at the R-peak of ECG, and the variance of the Gaussian function is $\sigma^2 = 1$.

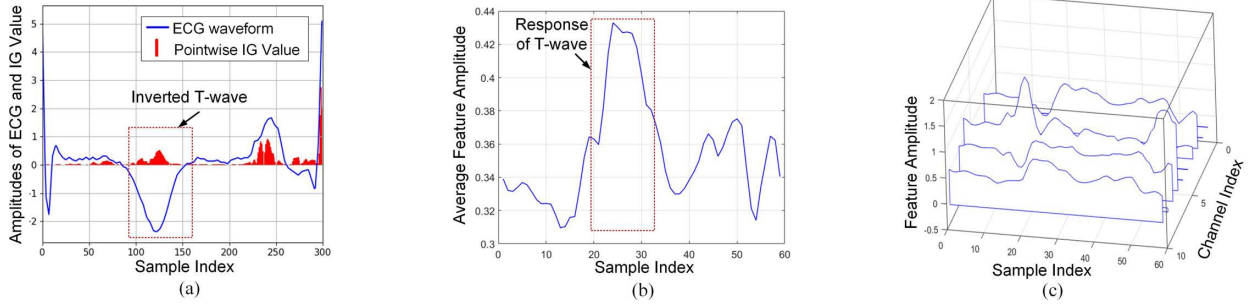


Fig. 7. Feature visualization and interpretation of the CVD classifier. (a) An ECG cycle (with inverted T-wave) labeled as NST-MI and the IG values showing the contribution of each sample point to the classification result. (b) Average feature response of ECG sub-waves. The curve in (b) is plotted by averaging the feature maps in (c) over the channel direction.

E. Model Compression for Mobile Applications

To better accommodate the stringent memory requirement of mobile devices, we compress the PPG2ECG network using parameter re-use and knowledge distillation.

The cascaded ENC_s and DEC_s take up more than 95% of the total parameters. The modules have similar architectures but different parameters. If we require the input and output of an arbitrary module to have the same size, the feed-forward pass through M cascaded modules can be simplified by the R -depth ($R \leq M$) recursion of one module [38]:

$$\mathbf{Y}_{\text{out}} = \underbrace{T \circ \dots \circ T}_{R}(\mathbf{X}_{\text{in}}), \quad (10)$$

where $T(\cdot)$ represents the module (ENC or DEC). Take ENC for example, (10) is equivalent to repeatedly applying $T(\cdot)$ on the input PPG for R times. In this way, low-level and high-level features are extracted using the same set of kernels. Since the patterns of PPG and ECG are relatively simple, re-using kernels does not noticeably degrade the expressive power of the PPG2ECG network. We have observed in experiments that with the aid of distillation, recursion can reduce over 60% parameters while maintaining the quality of ECG reconstruction. Given a pre-trained PPG2ECG network (i.e., teacher network), we construct a lightweight student network by replacing the cascaded ENC_s and DEC_s with recursive ones, and the student network is then trained to extract knowledge from the teacher network.

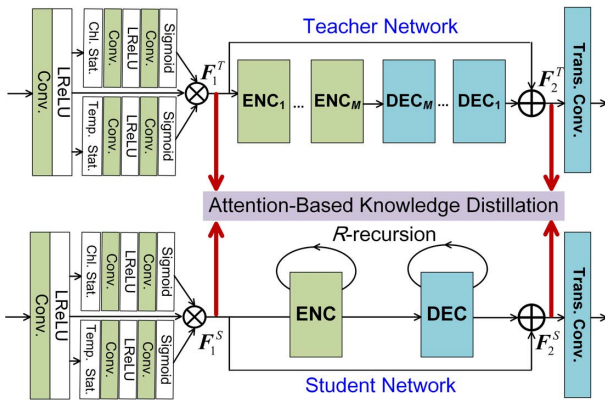


Fig. 8. Knowledge distillation for training a thin student PPG2ECG network.

The student network needs to mimic the internal behaviors of the teacher network to exhibit similar performance. To

this end, we select several key positions from the pre-trained teacher network and peek into its internal behaviors from feature maps. For the feature maps generated at each selected position, we apply max-pooling over the channel direction to obtain the maximum feature response at each time instant. The information reflects the salient parts of the feature map and tells where the teacher network pays attention [39]. This work selects two positions for knowledge distillation: the inputs to the first ENC and the ECG reconstruction layer (see the red arrows in Fig. 8). The distillation algorithm forces the student network to reproduce the information distilled from the teacher network while synthesizing ECG. The discrepancy between the two networks is measured by the cosine similarity between the max-pooling results of their intermediate features, leading to the following distillation loss:

$$\mathcal{L}_D = 1 - \frac{1}{2} \sum_{i=1}^2 \frac{\mathbf{M}_i^T \cdot \mathbf{M}_i^S}{\|\mathbf{M}_i^T\|_2 \|\mathbf{M}_i^S\|_2}, \quad (11)$$

where $\mathbf{M}_i^T = \text{MaxPool}(\mathbf{F}_i^T)$ and $\mathbf{M}_i^S = \text{MaxPool}(\mathbf{F}_i^S)$, $\mathbf{F}_i^T$ and $\mathbf{F}_i^S$ are the features sampled from the teacher and student networks at the i -th position. When training the student network, we combine $\mathcal{L}_D$ with the fidelity and feature losses in (9), where the ground-truth ECG in (9) is replaced by the output of the teacher network.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Datasets, Parameter Settings, and Competing Algorithms

Algorithms were trained and tested on three publicly available datasets. The primary reason for choosing online datasets is to facilitate open benchmarking. The training and testing PPG/ECG waveforms were all sensed from real-world subjects across a broad range of ages, ethnicities, and health conditions, including patients with common CVDs and healthy individuals without any known CVDs. The recording devices include professional monitors in hospitals and consumer-level wearable sensors during physical exercise. The artifacts and sensing noise in raw data ensure comprehensive coverage of real-world scenarios.

MIMIC-III: MIMIC-III [34] was chosen as the benchmark dataset for its richness of waveforms and CVD types, detailed diagnostic results from patients, public availability, and the real-world nature. MIMIC-III has a full coverage of pathological patterns related to major CVDs, and noisy data were intentionally preserved to reflect real-world healthcare settings. The

waveforms in MIMIC-III allow for evaluating the diagnosis of reconstructed ECG. Per our best knowledge, other publicly available datasets do not have comparable sizes and richness of PPG-ECG patterns and CVD types. Following the practice in previous studies [3], [4], [8], [9], we used the waveforms in Folder 35. The waveforms were screened using the signal quality assessment function in the PhysioNet Cardiovascular Signal Toolbox [40], and those labeled as “unacceptable (Q)” were discarded.

BIDMC: The BIDMC PPG and Respiration Dataset was acquired from 53 patients (aged from 19 to 90+) admitted by the Beth Israel Deaconess Medical Centre (BIDMC) [41]. It provides the 8-min recordings of the PPG, ECG, and physiological parameters (e.g., heart rate and respiratory rate) of each patient. PPG and ECG signals were sampled at a frequency of 125 Hz. The BIDMC and MIMIC-III datasets have no overlap in subjects.

Motion PPG-ECG: The PPG and ECG waveforms in this dataset were collected during physical exercise [42]. The subjects were asked to run on a treadmill at varying speeds (6km/h, 8km/h, 12km/h, and 15km/h), and there are two rest sessions of 30s at the beginning and end of each recording. The PPG waveforms were measured by a wrist-type sensor with a green LED, and ECG waveforms were sensed by a wet sensor attached to the chest. Moreover, the dataset also provides the acceleration signals simultaneously measured by an accelerometer placed on the wrist. The dataset was created for PPG-based heart rate estimation. The data in the testing part were not used in our experiment since ECG waveforms are not available.

The datasets consist of 71,098 pairs of ECG and PPG cycles, amounting to 47.3 hours of recordings. The signals were normalized to a fixed length of $L = 300$. For a fair performance comparison, we followed the same testing protocols as the prior work of Zhu *et al.* [3], [4] and Tian *et al.* [8], [9]. The PPG-ECG waveforms of each subject exist in both training and testing datasets. The first 80% and last 20% waveforms of each subject were allocated for training and testing, respectively. The parameter settings of the PPG2ECG network are presented in Fig. 3. The network was implemented in Pytorch and trained for 60 epochs with a batch size of ten using the Adam optimizer (with default parameter settings). The initial learning rate was set to 10^{-3} and then decayed by 0.6 every 18 epochs. The proposed algorithm was compared with five pieces of prior work on PPG-based ECG inference, covering both hand-crafted and deep-learning-based approaches, that are the DCT and linear regression algorithm (DCT) [4], the cross-domain joint dictionary learning algorithm (XDJDL) [8], the follow-up work that applies label-consistency regularization to XDJDL using CVD labels (LC-XDJDL) [9], the W-shaped neural network (W-Net) [12], and the transformed attentional neural network (TANN) [10]. W-Net was designed for the PPG signals with 1024 samples, so the testing data were first up-sampled to 1024 using bi-linear interpolation, and the reconstructed ECG cycles were then down-sampled to 300 samples for performance evaluation. TANN was tested using the codes posted online. We also adapted and trained a U-Net [43] as an additional baseline, and our implementation

of the 1D U-Net consists of four convolutional layers and four transposed convolutional layers with 60 kernels each. The kernel lengths of the convolutional layers are 30, 15, 10, and 5, respectively, and stride and dilation were all set to one. Two mirroring convolutional and transposed convolutional layers have the same parameter settings and are linked with a residual connection.

B. Comparison on the Quality of ECG Inference, Model Size, Complexity, and Explainability

We use the Pearson correlation coefficient and the relative Root Mean Squared Error (rRMSE) to evaluate the fidelity of inferred ECG $\hat{E}$:

$$\rho = \frac{(\mathbf{E} - \mu[\mathbf{E}])^T (\hat{\mathbf{E}} - \mu[\hat{\mathbf{E}}])}{\|\mathbf{E} - \mu[\mathbf{E}]\|_2 \|\hat{\mathbf{E}} - \mu[\hat{\mathbf{E}}]\|_2}, \quad (12)$$

$$\text{rRMSE} = \frac{\|\mathbf{E} - \hat{\mathbf{E}}\|_2}{\|\mathbf{E}\|_2}, \quad (13)$$

where $\mathbf{E}$ is the ground truth, and $\mu[\cdot]$ represents the element-wise mean value of a vector. Table I lists the statistics of the quality scores measured on each dataset². Table II gives an overall comparison on average quality scores, parameter amounts, computational complexities, and explainabilities of the testing algorithms. We take the number of floating-point operations (FLOPs) as a platform-independent measure of complexity, and the results in Table II were obtained by counting the operations for reconstructing an ECG cycle with 300 sample points. Two trends can be observed from the comparison. First, compared with the generic orthogonal bases of DCT, data-dependent bases (convolutional kernels and sparse coding atoms) better suit the underlying structures of PPG and ECG. Second, deep-learning-based approaches outperform DCT and XDJDL, which synthesize ECG using global bases. The correlation coefficients and rRMSEs in Table I indicate that the ECG cycles generated by the proposed algorithm have the highest fidelity on all datasets. Our ECG reconstruction model has the least amount of parameters. It also has the highest efficiency among sparse-coding- and deep-learning-based algorithms, and its FLOPs value is nearly half that of U-Net.

The multi-scale and attention mechanisms improve the sensitivity of the proposed network to the subtle difference among PPG waveforms. As can be seen from Fig. 9, although the waveforms of PPG are quite similar, the network can represent the distinct morphological difference among ECG waveforms. The PPG2ECG network can faithfully infer the fine detail and abnormal morphology of ECG, such as the inverted QRS complex in Fig. 9(c). The accurate reconstruction of ECG from PPG confirms the intrinsic association between the electrical and mechanical activities of the heart and the existence of CVD-related information in PPG. Besides local morphologies,

²LC-XDJDL applies additional label-consistency regularization to XDJDL. It exhibits equal performance as XDJDL on the BIDMC and Motion PPG-ECG datasets where CVD labels are unavailable, so its performance scores on these datasets are not separately listed in Table I.



Fig. 9. Comparison of the ECG sequences inferred by different approaches. The waveforms in (a), (b), and (c) were sensed from the subjects with coronary artery disease, myocardial infarction, and congestive heart failure, respectively.

TABLE I

QUANTITATIVE QUALITIES OF RECONSTRUCTED ECGs MEASURED ON EACH TESTING DATASET (TOP RESULTS ARE DISPLAYED IN BOLD; $\uparrow$ MEANS HIGH VALUE DESIRED, AND VICE VERSA.)

Datasets	Algorithms	$\rho \uparrow$			rRMSE $\downarrow$		
		μ	med	σ	μ	med	σ
MIMIC-III	DCT [4]	0.71	0.83	0.31	0.67	0.60	0.26
	XDJDL [8]	0.88	0.96	0.23	0.39	0.29	0.31
	LC-XDJDL [9]	0.92	0.97	0.17	0.33	0.26	0.25
	U-Net	0.93	0.96	0.13	0.32	0.27	0.19
	W-Net [12]	0.92	0.96	0.13	0.33	0.29	0.18
	TANN [10]	0.94	0.97	0.13	0.28	0.23	0.19
	Proposed	0.94	0.98	0.13	0.27	0.22	0.20
BIDMC	DCT [4]	0.70	0.83	0.35	0.66	0.60	0.23
	XDJDL [8]	0.82	0.94	0.27	0.48	0.35	0.36
	U-Net	0.86	0.93	0.17	0.43	0.37	0.25
	W-Net [12]	0.89	0.95	0.14	0.42	0.37	0.22
	TANN [10]	0.90	0.97	0.17	0.35	0.26	0.27
Motion PPG-ECG	Proposed	0.90	0.97	0.16	0.35	0.26	0.26
	DCT [4]	0.55	0.79	0.56	0.78	0.69	0.26
	XDJDL [8]	0.51	0.75	0.53	0.79	0.71	0.30
	U-Net	0.80	0.83	0.14	0.58	0.56	0.16
	W-Net [12]	0.80	0.82	0.12	0.59	0.58	0.13
	TANN [10]	0.80	0.83	0.14	0.58	0.56	0.17
	Proposed	0.81	0.83	0.13	0.57	0.56	0.16

μ : mean value; med: median value; std: standard deviation.

the timing information of ECG also reflects the health conditions of the heart. For example, a prolonged PR interval indicates slow conduction between the atria and ventricles. We also assessed the accuracy of the PPG2ECG network in representing the timing information. For the waveforms in the MIMIC-III dataset sampled at a frequency of 125Hz, the Mean Absolute Errors (MAE) between the durations of the PR, QRS, and QT intervals measured from reconstructed and ground-truth waveforms are 4ms, 5ms, and 9ms, respectively. The MAE is smaller than a quarter of the horizontal length of a small square (40ms) on ECG graph paper. As can be seen from Fig. 9, the timing information of the reconstructed and

TABLE II

COMPARISON ON AVERAGE QUALITY SCORES, PARAMETER AMOUNT, COMPUTATIONAL COMPLEXITY, AND EXPLAINABILITY

Algorithms	#Para.(M) $\downarrow$	FLOPs(M) $\downarrow$	$\bar{\rho} \uparrow$	rRMSE $\downarrow$	Explain.
DCT [4]	0.27	0.36	0.65	0.70	High
XDJDL [8]	5.67	60.21	0.74	0.55	High
LC-XDJDL [9]	5.90	64.72	0.75	0.53	High
U-Net	0.22	55.99	0.86	0.44	Low
W-Net [12]	2.37	35.97	0.87	0.45	Low
TANN [10]	10.98	518.79	0.88	0.40	Low
Proposed	0.12	28.32	0.88	0.40	High

ground-truth waveforms show good agreement. However, it is clear from Table I that the ECG inference in the ambulatory setting is more challenging. Due to the interference of motion on PPG measurement, all algorithms show degradation on the Motion PPG-ECG dataset. In the following subsection, we will describe a method that exploits auxiliary information for more robust ECG inference.

From the comparison in Table II, we note that the network exhibits comparable or superior performance than TANN on all the datasets, while its parameter amount and computational load are orders of magnitude lower. It only has about 1% the parameter amount and 5% the FLOPs as TANN. The DCT-based algorithm is the most efficient since DCT and linear regression can be implemented through a few matrix-vector multiplications. The computational load required by the proposed algorithm does not impose a substantial burden on mainstream devices. As a reference for comparison, the mobile processor Exynos 7 Dual released in 2016 for wearable devices can perform up to 15G FLOPs per second. We will show later in Section IV-F that the model size and computational load can be further reduced via model compression.

We take explainability as another dimension of comparison.

Linear models, including DCT [4], XDJDL [8], and LC-XDJDL [9], are inherently interpretable. For example, LC-XDJDL uses linear transforms to infer the sparse codes of ECG and CVD labels from the sparse codes of PPG. Hence, the pathological patterns of PPG can be discovered by identifying the atoms that frequently co-occur with a specific CVD in sparse representation. Unlike other deep learning algorithms that mainly focus on the data-fitting aspect of PPG-based ECG inference, the proposed work also addresses the interpretation aspect. More specifically, we take PPG-to-ECG mapping and model interpretation as avenues for understanding the influence of CVDs on PPG, as will be discussed later in Section IV-E.

C. Motion-Information-Aided ECG Reconstruction

Motion artifacts degrade the accuracy of ECG reconstruction. From the physiological perspective, PPG is shaped by the electrical response of the heart. While under intensive physical exercise, motion also interferes with the optical sensing of PPG signal and becomes another causing factor. For validation, we use causal inference to examine the factors that affect PPG, the insights from which help to improve the robustness of ECG reconstruction.

We use directed information [44] to analyze the causal influence among PPG, ECG, and motion. Directed information (DI) is an information-theoretic metric for examining the causal influence of one time series on another. Given two L -length sequences $\mathbf{X}$ and $\mathbf{Y}$, the directed information from $\mathbf{X}$ to $\mathbf{Y}$ is denoted by $I(\mathbf{X} \rightarrow \mathbf{Y})$, and

$$I(\mathbf{X} \rightarrow \mathbf{Y}) = H(\mathbf{Y}) - \sum_{i=1}^L H(\mathbf{Y}[i] | \mathbf{Y}[1:i-1], \mathbf{X}[1:i]), \quad (14)$$

where $H(\cdot)$ represents entropy, $H(\cdot|\cdot)$ represents conditional entropy, and the colon operator has the same meaning as in Matlab.

The DI in the reverse direction is denoted by $I(\mathbf{Y} \rightarrow \mathbf{X})$, and the metric is not symmetric. The relationship between the DI values in two directions asserts the causal influence between two series [44]:

- 1) $I(\mathbf{X} \rightarrow \mathbf{Y}) \gg I(\mathbf{Y}[1:L-1] \rightarrow \mathbf{X})$ indicates that $\mathbf{X}$ causes $\mathbf{Y}$,
- 2) $I(\mathbf{X} \rightarrow \mathbf{Y}) \ll I(\mathbf{Y}[1:L-1] \rightarrow \mathbf{X})$ indicates that $\mathbf{Y}$ causes $\mathbf{X}$,
- 3) $I(\mathbf{X} \rightarrow \mathbf{Y}) \approx I(\mathbf{Y}[1:L-1] \rightarrow \mathbf{X}) \rightarrow 0$ implies the independence of $\mathbf{X}$ and $\mathbf{Y}$, and accordingly, there is no causal influence in any direction.

In Fig. 10(a), we plot the curves of the DI values between PPG and ECG waveforms along two directions, and the waveforms were extracted from the Motion PPG-ECG dataset. An obvious causal influence of ECG on PPG can be observed, indicating that the electrical activities of the heart (represented by ECG) cause blood circulation (represented by PPG), which is consistent with known facts. To examine the impact of motion on PPG measurement, we also estimated the DI values between PPG and motion information. The motion information was measured at the subject's wrist using an accelerometer

during running. Fig. 10(b) shows the curves of the DI values between PPG and the x-axis acceleration signal. The causal influence of motion on PPG is quite remarkable, suggesting that motion indeed affects the optical sensing of PPG. In this setting, the physiological behavior of the heart is not the sole decisive factor of PPG. Hence, for better robustness of ECG reconstruction, it is necessary to use reference motion information to counteract the artifacts of PPG.

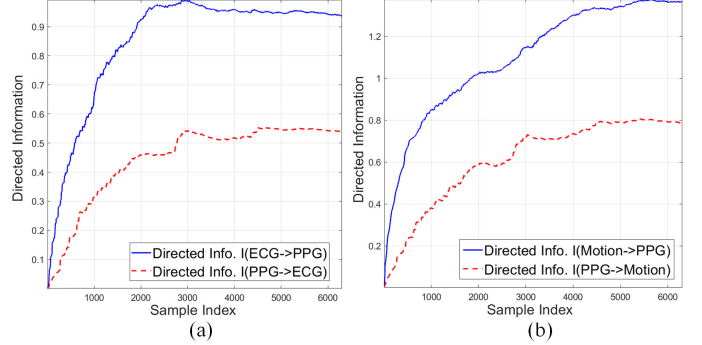


Fig. 10. Causal effects of ECG and motion on PPG. (a) Curves of DI between ECG and PPG. (b) Curves of DI between acceleration signal and PPG.

Motion information is readily available since most devices use a built-in accelerometer for fitness tracking. Acceleration signals have been exploited to correct the motion artifacts in PPG. Some PPG-based heart rate estimation algorithms de-noise PPG signals by taking acceleration signal as the reference of adaptive filter [45]. It is usually assumed that motion artifacts are additive in the raw signal space, but the influence of motion on PPG is far more complicated. In this work, we do not directly de-noise PPG. Instead, the three-dimensional acceleration signal is concatenated with PPG along the channel direction as auxiliary inputs. This gives the PG2ECG network more flexibility to alleviate motion artifacts, not only in the raw input space but also in the feature spaces. Table III compares the qualities of the ECGs synthesized with and without the auxiliary motion information. The PG2ECG network aided by acceleration signals achieves more accurate ECG reconstruction, and the average rRMSE drops to 0.51 from 0.57. This finding suggests that using the side information provided by the accelerometer can enhance the robustness of PPG-based cardiac monitoring when a subject performs intensive exercise.

TABLE III
COMPARISON BETWEEN THE PERFORMANCE OF ECG RECONSTRUCTION UNDER EXERCISE WITH AND WITHOUT THE ASSISTANCE OF ACCELERATION SIGNAL

Modes	$\rho \uparrow$			rRMSE $\downarrow$		
	μ	med	σ	μ	med	σ
With Acc. Signal	0.84	0.88	0.15	0.51	0.47	0.22
Without Acc. Signal	0.81	0.83	0.13	0.57	0.56	0.16

Given the effectiveness of the motion-information-aided scheme, we seek to understand how neural network utilizes motion information in the ambulatory setting using the IG-based model interpretation [35]. As in interpreting the CVD classifier, we calculated the IG value of the i -th dimension of the input with respect to the j -th dimension of the output ECG

($\hat{\mathbf{E}} \in \mathbb{R}^L$) and denote the result as IG_i^j . The contribution of the i -th input point to reconstructing the whole ECG cycle is measured by summing up the amplitudes of $\{\text{IG}_i^j | j = 1, \dots, L\}$, as $S_i = \sum_{j=1}^L |\text{IG}_i^j|$.

By summing up S_i over each input modality (PPG or acceleration signal), we found that the ratio of the contribution given by the acceleration signal in generating the output ECG is 27%. This is in line with our intuitive understanding. Despite the interference of motion, the activities of the heart are still the principal causing factor of PPG. Hence, a vast amount of the information for inferring ECG comes from PPG, while the acceleration signal only helps to reduce the motion artifacts of PPG. Let us further examine under what conditions the acceleration signal plays a more positive role in assisting ECG synthesis. In Fig. 11, we highlight the top 20% sample points in the acceleration signal with the largest contribution. It is clear that the PPG2ECG network pays more attention to the auxiliary information when there are strong directional changes in acceleration. In such moments, the gap between the skin and the surface of the pulse oximeter changes rapidly, and this is the primary source of motion artifacts [42].

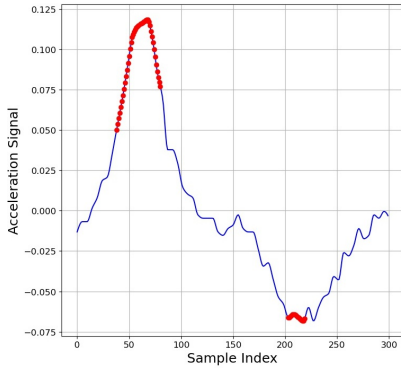


Fig. 11. Visualization of the model interpretation results about the acceleration signal. The sample points highlighted in red are the top 20% ones that contribute most to the ECG reconstruction under exercise.

D. Effects of Diagnosis-Oriented Training

We also evaluated the utility of reconstructed ECGs in diagnosing CVDs. The experiments were conducted on the MIMIC-III dataset where the testing waveforms are annotated with disease labels. The testing data cover five common CVDs, as listed in Section III-D. Reconstructed ECGs were fed to the CVD classifier described in Section III-D, and the classifier was trained using the original ECGs in the training subset of MIMIC-III. The diagnostic precision was evaluated at the cycle level. Diagnosis can also be made at the sequence level using majority voting and will exhibit higher precision. To set a more stringent criterion and eliminate the error-correcting effect of majority voting, we did not adopt the sequence-level precision for performance assessment. The disease label of each ECG cycle was predicted by picking the one with the highest probability. The diagnostic results were compared with the annotations provided by clinical staff, showing that 93% of reconstructed ECGs are correctly classified. Table IV shows the disease-specific F_1 scores. For all kinds of diseases, the F_1

scores are over 0.95, indicating that the PPG2ECG network is able to infer the pathological ECG patterns from PPG.

TABLE IV
DISEASE-SPECIFIC DIAGNOSTIC ACCURACY MEASURED BY F_1 SCORE

CHF	ST-MI	NST-MI	HYPO	CAD
0.96	0.96	0.95	0.96	0.95

Ablation experiments were designed to demonstrate the effects of the diagnosis-oriented training (DOT) strategy. We re-trained the PPG2ECG network by removing $\mathcal{L}_F$ and $\mathcal{L}_C$ from (9), only leaving the weighted ℓ_1 norm. In Table V, we compare the performance of the networks trained with different approaches. After disabling DOT, the precision of CVD diagnosis on reconstructed ECGs drops to 0.90 from 0.93. It is worth mentioning that the training scheme also improves the fidelity of reconstructed signals, as verified by the comparison on ρ and rRMSE in Table V. The feature loss and contrastive loss supervise the PPG2ECG network using the knowledge about the clustering structures of ECGs related to CVDs, so the network needs to represent the pathological changes of ECG with higher fidelity.

TABLE V
EFFECTS OF DOT MEASURED ON THE PROPOSED PPG2ECG NETWORK

Training Alg.	$\bar{\rho} \uparrow$	rRMSE $\downarrow$	$P_{NN} \uparrow$	$P_{SVM} \uparrow$
With DOT	0.94	0.27	0.93	0.88
Without DOT	0.93	0.29	0.90	0.83

We are curious if the benefits brought by DOT can extend to a different classifier that has not been used for regularizing ECG reconstruction, so experiments were also conducted in a model-agnostic manner by taking a multi-class support vector machine (SVM) as the classifier. The precisions attained by SVM on the ECGs reconstructed by the networks with and without DOT are 0.88 and 0.83, respectively (shown in the last column of Table V). We conjecture that the diagnostic cues captured by different classifiers have some overlaps, so the benefits brought by the regularization effect of one classifier are transferable to another unseen one.

The DOT algorithm is independent of the architecture of the ECG reconstruction model and can serve as a generic performance-boosting approach. In another experiment, we applied it to U-Net, and the network reinforced by our training scheme was compared with the one tested in Section IV-B (see Table VI). As in the above ablation experiment, performance gains in terms of the diagnostic precision of CVDs and the quality of output signals can be observed.

TABLE VI
EFFECTS OF DOT MEASURED ON U-NET

Training Alg.	$\bar{\rho} \uparrow$	rRMSE $\downarrow$	$P_{NN} \uparrow$	$P_{SVM} \uparrow$
With DOT	0.93	0.30	0.91	0.87
Without DOT	0.93	0.32	0.86	0.80

A discriminative CVD classifier plays a strong effect in regularizing the PPG2ECG model. To examine the discriminating capability of the classifier, we further evaluated it on a more challenging dataset, PTB-XL [46], which should be the largest

publicly available ECG dataset to date. It contains 21,837 12-lead ECG waveforms measured from 18,885 patients over seven years, and each recording is of ten seconds. The dataset is a rich source of real-world cases, where sensing noise, drift, and other artifacts were all preserved in the raw waveforms. PTB-XL contains both normal and pathological ECG recordings, and the CVDs covered include conduction disturbance, myocardial infarction, hypertrophy, and ST/T change. Some diseases co-occur in a patient, making CVD prediction a multi-label classification problem, so the CVD prediction task on PTB-XL is more challenging than that on MIMIC-III. Accordingly, the number of output nodes in the classifier was set to five, each representing the probability of a specific label. We fed the 12-lead ECG waveforms to the classifier, so the input channel number was set to 12. To adapt the expressive power of the classifier to the diversity and scale of the dataset, we increase the kernel numbers of all convolutional layers by three times.

The training and testing sets were split using the method recommended by the creators of PTB-XL. Specifically, we used the waveforms in the tenth folder for testing and the remaining ones for training. Experiments were carried out on down-sampled versions of the ECG waveforms (with a sampling rate of 100Hz). We use the subject-centered metric to measure the performance of multi-label classification. Prediction results are obtained by comparing the estimated likelihood of each state with a decision threshold. For a given threshold, we calculate the precision and recall rates from the prediction results, and an F_1 score is computed for each subject. As the threshold varies from 0 to 1, the average F_1 score of all subjects observed at the optimal threshold is 0.71.

E. Exploring CVD-Related Signs on PPG Using Model Interpretation

The CVD-related ECG abnormalities have been extensively studied, while the representations of CVDs on PPG are less understood. The pathological changes of the heart affect its pumping power and blood circulation, so PPG may reflect such abnormality. As shown earlier, the ECG signals reconstructed from PPG show encouraging performance in predicting CVDs, suggesting that PPG carries some diagnostic cues. Interpreting the mechanisms underlying data-driven models can offer medical practitioners complementary support and enrich the clinical knowledge base [47]. Recall that we have constructed two models that establish the connections between PPG and ECG and between ECG and CVDs. The joint interpretation of these models has the potential to bridge the knowledge gap between PPG and CVDs.

Cascading the PPG2ECG network $G_{P \rightarrow E}(\cdot)$ and the ECG-based CVD classification network $\psi(\cdot)$ can produce a holistic classifier that directly predicts CVDs from PPG: $\Psi(\cdot) = \psi \circ G_{P \rightarrow E}(\cdot)$. As in Section III-D, we attribute the diagnostic result to each PPG point using the IG-based model interpretation [35]. The IG values tell which parts of PPG are most influential to the decision.

In Fig. 12, we show two PPG cycles measured from the subjects diagnosed with coronary artery disease (CAD)

and congestive heart failure (CHF), respectively, and the red dots mark the top 20% sample points contributing most to diagnosing the diseases. We find that for both diseases, the regions around the peak are more informative. It agrees with the finding that the increased rounding or triangulation and asymmetry are more likely to appear in the PPG waveforms of CVD patients [48]. For CAD, the front of the ascending slope of PPG (corresponding to the moments when blood flows out of the heart) also receives high focus. CAD is caused by the plaque deposited on the inner walls of the arteries. According to the Moens-Korteweg equation [49], the velocity of pulse wave partially depends on the elasticity of the arterial wall and the radius of the artery. The narrowing and increased stiffness of arteries affect the dynamics of blood flow. We conjecture that neural network learns to detect such changes from the increasing rate of blood volume. In the case of CHF, the heart becomes too weak to relax properly, resulting in reduced blood flow to the heart and the buildup of blood in other organs. From Fig. 12(b), the sample points on the tail of PPG also contribute much to predicting CHF.

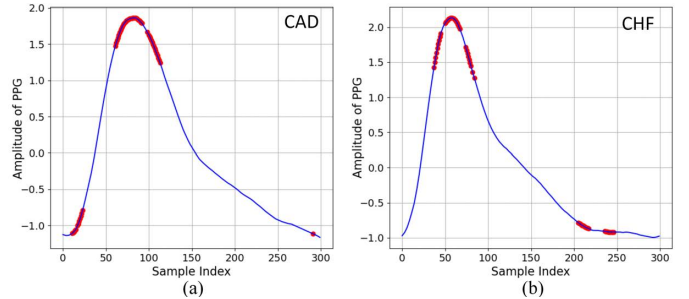


Fig. 12. Visualization of the association between CVDs and PPG. (a) and (b) are the PPG cycles measured from two subjects diagnosed with CAD and CHF, respectively. For each disease, the sample points highlighted in red are the top 20% most important ones for diagnosing the disease.

F. Performance of Model Compression

In the model compression experiment, we replaced the four cascaded ENC and DEC in the pre-trained PPG2ECG network with 2-depth recursive ENC and DEC to obtain a thinner student network. The kernel lengths of the three convolutional (or transposed convolutional) layers in the recursive modules are 13, 15, and 17. The student network approximates the distilled intermediate features and outputs of the pre-trained teacher network to minimize their performance gap. The training process only relies on a pre-trained network and input PPG signals without using the corresponding ECG signals. In this way, when adapting a pre-trained PPG2ECG network to lower-end devices with less memory and computational resources via distillation, there is no need to collect paired PPG and ECG signals, which alleviates the burden of data collection and reduces the information leakage in training data. In Table VII, we compare the model size, FLOPs, and average quality scores of the compressed network with other algorithms. Table VIII lists the quantitative performance of the full and compressed networks measured on each dataset. The compression cuts 67% of parameters and 34% of FLOPs. The compressed network only has 40K parameters and is about 5 ~ 273 times smaller than the comparative ones. From

Table VIII, the variations of average ρ and rRMSE measured on all datasets after model compression are 0.01 and 0.02, respectively. Moreover, the loss of diagnostic precision is also maintained at a low level. The average F_1 scores of the neural-network-based CVD classifier measured on the waveforms reconstructed by the full and compressed networks are 0.96 and 0.94, respectively.

TABLE VII
COMPARISON OF THE COMPRESSED PPG2ECG NETWORK WITH OTHER ALGORITHMS

Algorithms	#Para.(M)↓	FLOPs(M)↓	$\bar{\rho}$ ↑	rRMSE ↓
DCT [4]	0.27	0.36	0.65	0.70
XDJDL [8]	5.67	60.21	0.74	0.55
LC-XDJDL [9]	5.90	64.72	0.75	0.53
U-Net	0.22	55.99	0.86	0.44
W-Net [12]	2.37	35.97	0.87	0.45
TANN [10]	10.98	518.79	0.88	0.40
Full	0.12	28.32	0.88	0.40
Compressed	0.04	18.63	0.87	0.42

TABLE VIII
COMPARISON BETWEEN THE FULL AND COMPRESSED NETWORKS

Datasets	Algorithms	ρ ↑			rRMSE ↓		
		μ	med	σ	μ	med	σ
MIMIC-III	Full	0.94	0.98	0.13	0.27	0.22	0.20
	Compressed	0.93	0.97	0.14	0.31	0.25	0.20
BIDMC	Full	0.90	0.97	0.16	0.35	0.26	0.26
	Compressed	0.89	0.95	0.16	0.39	0.31	0.25
Motion PPG-ECG	Full	0.81	0.83	0.13	0.57	0.56	0.16
	Compressed	0.80	0.83	0.13	0.57	0.57	0.16

G. Ablation Study on Neural Network Architectures

We first examine the effects of the attention gates and the channel shuffle layers in ENC and DEC. Three ablated networks were constructed by removing each and both kinds of modules from the PPG2ECG network. Table IX reports the average performance of each ablated network measured on the testing datasets. Compared with the original network, all the ablated networks show inferior performance. The average rRMSE between reconstructed and original ECG waveforms rises to 0.44 after discarding both modules. As mentioned above in Section III-C, the modules were designed to emphasize the informative parts of feature maps and maximize the fusion of multi-scale features, respectively. These mechanisms are beneficial to synthesizing fine-granular ECG structures. To control the parameter budget, the modules have quite few or no parameters. All the attention gates contain 432 parameters (approximately 0.3% of the parameter amount of the PPG2ECG network), and the channel shuffle layers have no parameter.

TABLE IX
COMPARISON BETWEEN THE ABLATED AND FULL NETWORKS

Attention Gates	Channel Shuffle	$\bar{\rho}$ ↑	rRMSE ↓
×	×	0.86	0.44
✓	×	0.86	0.43
×	✓	0.87	0.42
✓	✓	0.88	0.40

The kernel sizes of the convolutional layers in each ENC or DEC are set to different values to diversify the receptive fields in feature learning and ECG synthesis. We also conducted ablation experiments to verify the effectiveness of this design, where the kernel sizes were all set to nine (i.e., the average kernel size in the first ENC). Degradation in the qualities of reconstructed ECG cycles has been observed, where the average ρ declines to 0.87 from 0.88, and the average rRMSE rises to 0.42 from 0.40.

H. Leave-One-Out Mode Versus Personalized ECG Inference

The results in Table I corroborate the generic correlations between PPG and ECG. It is also worth noting that the electrical and mechanical properties of the heart vary from individual to individual. Following the practice in another work [9], we compared two training setups: the group mode and the leave-one-out mode, where the historical waveforms of a subject are included in and excluded from the training set, respectively. For a randomly selected subject in the BIDMC dataset, we measured the performance of the PPG2ECG networks trained using the two modes. For the group mode, the average ρ measured on the subject is 0.92, and the value falls to 0.82 in the leave-one-out mode. This result is consistent with the observations that due to physiological variances, the rules learned from average population may not work best for all subjects [9] [24]. Patient-specific modeling (PSM) [50] is a promising strategy for tackling this challenge. In light of individual variability and the difficulties of obtaining one-fit-all models, PSM advocates using the data of a target patient to develop individualized computational models for more accurate clinical outcomes. To meet the emerging trends of precision medicine and cardio-physiological digital twin, we also tested ECG inference in the PSM manner. Personalized ECG inference can be achieved by fine-tuning a pre-trained (in the leave-one-out mode) model using the historical data of an individual, which is more practical than training an individualized model from scratch. After two epochs of fine-tuning, the average ρ rises to 0.97. Personalized fine-tuning benefits the subjects whose cardiovascular systems show some attributes rarely seen in the average population.

V. CONCLUSIONS

This paper has presented a computational approach for continuous ECG monitoring using optical sensing data. We have shown the feasibility of using a lightweight neural network to synthesize ECG waveforms from the blood volume variation signal measured by a PPG sensor. To facilitate CVD diagnosis, we have developed a task-aware training algorithm to ensure the precise representation of the clinically significant features of ECG. Compared with prior arts, the proposed algorithm demonstrates comparable or superior performance with fewer parameters. Our study also addresses the data analytic problems associated with PPG-based ECG synthesis and CVD diagnosis. We have leveraged model interpretation and causal analysis techniques to reveal the abnormal patterns of ECG and PPG related to CVDs, as well as the physiological and physical factors affecting PPG measurement. The insights

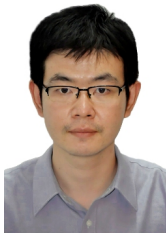
gained from such exploration can support more transparent and reliable cardiac monitoring. We anticipate that the personalized PPG2ECG model may act as a building block of the cardio-physiological digital twin that enables personalized precision healthcare.

The current work has some limitations that need to be addressed in further studies. First, the current algorithm uses paired ECG and PPG waveforms in training, while synchronized or paired data are generally limited in practice, and a large volume of physiological data collected during routine monitoring are unpaired. It would be beneficial to investigate how to harness unpaired data to model the activities of the heart and overcome the scarcity of paired training data. Second, this work has not fully exploited the health information of a patient in ECG inference (e.g., age, weight, and medical history generally available in the health record). Such factors are known to be associated with certain CVDs and the pumping power of the heart. The ECG inference model conditioning on auxiliary health information may achieve more accurate and personalized cardiac monitoring.

REFERENCES

- [1] M. Heron, "Heart disease facts," *Nat. Vital Statist. Rep.*, vol. 68, no. 6, pp. 1–77, 2019.
- [2] A. Torbal, E. Boersma, J. Kors, G. Herpen, J. Deckers, D. Kuip, B. Stricker, A. Hofman, and J. Witteman, "Incidence of recognized and unrecognized myocardial infarction in men and women aged 55 and older: The Rotterdam study," *Euro. Heart J.*, vol. 27, pp. 729–736, 2006.
- [3] Q. Zhu, X. Tian, C. Wong, and M. Wu, "ECG reconstruction via PPG: A pilot study," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2019, pp. 1–4.
- [4] —, "Learning your heart actions from pulse: ECG waveform reconstruction from PPG," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16 734–16 748, 2021.
- [5] M. V. Perez, K. W. Mahaffey, H. Hedlin, J. S. Rumsfeld, A. Garcia, T. Ferris, V. Balasubramanian, A. M. Russo, A. Rajmane, L. Cheung *et al.*, "Large-scale assessment of a smartwatch to identify atrial fibrillation," *New Engl. J. Med.*, vol. 381, no. 20, pp. 1909–1917, 2019.
- [6] K. Bruynseels, F. Santoni de Sio, and J. van den Hoven, "Digital twins in health care: Ethical implications of an emerging engineering paradigm," *Front. Genet.*, vol. 9, p. 31, 2018.
- [7] N. Selvaraj, A. K. Jaryal, J. Santhosh, K. K. Deepak, and S. Anand, "Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography," *J. Med. Eng. Technol.*, vol. 32, pp. 479–484, 2008.
- [8] X. Tian, Q. Zhu, Y. Li, and M. Wu, "Cross-domain joint dictionary learning for ECG reconstruction from PPG," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 936–940.
- [9] —, "Cross-domain joint dictionary learning for ECG inference from PPG," *IEEE Internet Things J.*, vol. 10, no. 9, pp. 8140–8154, 2023.
- [10] H. Chiu, H. Shuai, and P. C. P. Chao, "Reconstructing QRS complex from PPG by transformed attentional neural networks," *IEEE Sensors J.*, vol. 20, no. 20, pp. 12 374–12 383, 2020.
- [11] K. Vo, E. K. Nacini, A. Naderi, D. Jilani, A. M. Rahmani, N. Dutt, and H. Cao, "P2E-WGAN: ECG waveform synthesis from PPG with conditional wasserstein generative adversarial networks," in *Proc. Annu. ACM Symp. Appl. Comput.*, 2021, pp. 1030–1036.
- [12] Q. Tang, Z. Chen, R. Ward, C. Menon, and M. Elgendi, "PPG2ECGps: An end-to-end subject-specific deep neural network model for electrocardiogram reconstruction from photoplethysmography signals without pulse arrival time adjustments," *Bioeng.*, vol. 10, no. 6, 2023.
- [13] Q. Tang, Z. Chen, Y. Guo, Y. Liang, R. Ward, C. Menon, and M. Elgendi, "Robust reconstruction of electrocardiogram using photoplethysmography: A subject-based model," *Frontiers in Physiology*, vol. 13, 2022.
- [14] E. Lan, "Performer: A novel PPG-to-ECG reconstruction transformer for a digital biomarker of cardiovascular disease detection," in *IEEE Winter Conf. Appl. Comput. Vis.*, 2023, pp. 1990–1998.
- [15] R. Banerjee, A. Sinha, A. Dutta Choudhury, and A. Visvanathan, "PhotoECG: Photoplethysmography to estimate ECG parameters," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2014, pp. 4404–4408.
- [16] P. Wang, C. Han, F. Yu, Y. Zheng, X. Huang, and L. Cui, "Rec-AUNet: Attentive unet for reconstruction of ECG from BCG," in *IEEE Int. Conf. Bioinfo. Biomed.*, 2022, pp. 1241–1248.
- [17] Y. Cao, F. Li, H. Chen, X. Liu, L. Zhang, and Y. Wang, "Guard your heart silently: Continuous electrocardiogram waveform monitoring with wrist-worn motion sensor," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 3, 2022.
- [18] D. Toda, R. Anzai, K. Ichige, R. Saito, and D. Ueki, "ECG signal reconstruction using FMCW radar and convolutional neural network," in *Proc. Int. Symp. Commun. Inf. Technol.*, 2021, pp. 176–181.
- [19] A. Hannun, P. Rajpurkar, M. Haghighpanahi, G. Tison, C. Bourn, M. Turakhia, and A. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, pp. 65–69, 2019.
- [20] T. Yang, L. Yu, Q. Jin, L. Wu, and B. He, "Localization of origins of premature ventricular contraction by means of convolutional neural network from 12-lead ECG," *IEEE Trans. Bio-Med. Eng.*, vol. 65, no. 7, pp. 1662–1671, 2018.
- [21] A. Sbröllini, M. C. De Jongh, C. C. T. Haar, R. W. Treskes, S. Man, L. Burattini, and C. A. Swenne, "Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: A deep-learning approach," *Bio. Eng. Online*, vol. 18, no. 1, p. 15, 2019.
- [22] Y. Liang, Z. Chen, R. K. Ward, and M. Elgendi, "Photoplethysmography and deep learning: Enhancing hypertension risk stratification," *Biosensors*, vol. 8, no. 4, p. 101, 2018.
- [23] C. T. Arsene, R. Hankins, and H. Yin, "Deep learning models for denoising ECG signals," in *Proc. Eur. Signal Process. Conf.*, 2019, pp. 1–5.
- [24] T. Golany and K. Radinsky, "PGANs: Personalized generative adversarial networks for ECG synthesis to improve patient-specific deep ECG classification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 557–564.
- [25] J. C. Ruiz-Rodríguez, A. Ruiz-Sanmartín, V. Ribas, J. Caballero, A. García-Roche, J. Riera, X. Nuviols, M. Nadal, O. Sola-Morales, J. Serra *et al.*, "Innovative continuous non-invasive cuffless blood pressure monitoring based on photoplethysmography technology," *Intensive Care Med.*, vol. 39, no. 9, pp. 1618–1625, 2013.
- [26] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, "Deep PPG: Large-scale heart rate estimation with convolutional neural networks," *Sensors*, vol. 19, no. 14, p. 3079, 2019.
- [27] C. Yu, Z. Gao, W. Zhang, G. Yang, S. Zhao, H. Zhang, Y. Zhang, and S. Li, "Multitask learning for estimating multitype cardiac indices in MRI and CT based on adversarial reverse mapping," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 493–506, 2020.
- [28] S. Wu, S. Wei, C. Chang, A. L. Swindlehurst, and J. Chiu, "A scalable open-set ECG identification system based on compressed CNNs," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, 2021.
- [29] N. Trayanova and J. Rice, "Cardiac electromechanical models: From cell to organ," *Front. Physiol.*, vol. 2, 2011.
- [30] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
- [33] G. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 15, 2002, pp. 833–840.
- [34] A. Johnson, T. Pollard, L. Shen, L. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Celi, and R. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, p. 160035, 2016.
- [35] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3319–3328.
- [36] "Clinical ECG Interpretation (eBook)," Available: <https://ecgwaves.com/>, Accessed: Dec. 10, 2021.
- [37] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 18 661–18 673.
- [38] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.

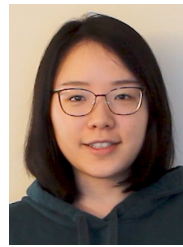
- [39] N. Komodakis and S. Zagoruyko, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," in *Proc. Int. Conf. Learn. Rep.*, 2017.
- [40] A. Vest, Q. Li, C. Liu, S. Nemati, G. Da Poian, A. Shah, and G. Clifford, "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiol. Meas.*, vol. 39, no. 10, 2018.
- [41] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1914–1923, 2017.
- [42] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Trans. on Biomed. Eng.*, vol. 62, no. 2, pp. 522–531, 2015.
- [43] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Medical Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.
- [44] J. Jiao, H. H. Permuter, L. Zhao, Y. Kim, and T. Weissman, "Universal estimation of directed information," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6220–6242, 2013.
- [45] S. H. Kim, D. W. Ryo, and C. Bae, "Adaptive noise cancellation using accelerometers for the PPG signal from forehead," in *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2007, pp. 2564–2567.
- [46] P. Wagner, N. Strothoff, R.-D. Boussejot, D. Kreisler, F. I. Lunze, W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset," *Sci. Data*, vol. 7, no. 1, p. 154, 2020.
- [47] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, 2021.
- [48] J. Allen, K. Overbeck, G. Stansby, and A. Murray, "Photoplethysmography assessments in cardiovascular disease," *Meas. Control*, vol. 39, no. 3, pp. 80–83, 2006.
- [49] B. Brown, R. Smallwood, D. Barber, P. Lawford, and D. Hose, *Medical Physics and Biomedical Engineering*. Taylor & Francis, 1998.
- [50] M. L. Neal and R. Kerckhoffs, "Current progress in patient-specific modeling," *Brief. Bioinform.*, vol. 11, no. 1, pp. 111–126, 2010.



Yuenan Li (Senior Member, IEEE) received his B.E. and M.E. degrees in measurement and automation, and the Ph.D. degree in information and communication engineering, in 2004, 2006, and 2010, respectively, all from Harbin Institute of Technology, Harbin, China.

He is currently an Associate Professor at the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. From 2019 to 2020, he was a visiting researcher at the University of Maryland, College Park, USA. His research inter-

ests include multimedia signal processing, information security and forensics, computer vision, and physiological signal processing.



Xin Tian (Member, IEEE) received the B.E. degree in optoelectronic information science and engineering from the Huazhong University of Science and Technology of China, Wuhan, China, in 2017, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA, in 2022.

She has been a Research Scientist with Meta Inc., Menlo Park, CA, USA, since 2022. Her research interests are signal processing, data science, and machine learning in smart health.

Dr. Tian was selected as an Outstanding Graduate Assistant and a Future Faculty Fellow by the University of Maryland in 2020 and 2021, respectively.



Qiang Zhu (Member, IEEE) received the B.E. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2010, the M.S. degree in control science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2020.

He has been a Research Scientist with Meta Inc., Menlo Park, CA, USA, since 2020. His research interests are signal processing, machine learning,

and information retrieval.

Dr. Zhu received the Distinguished Teaching Assistance Award in 2016 from the University of Maryland.



Min Wu (Fellow, IEEE) received the B.E. degree in automation (Highest Hons.) and the B.A. degree in economics from Tsinghua University, Beijing, China, in 1996, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2001.

She is a Christine Y. Kim Eminent Professor in Information Technology and an Associate Dean of Engineering, and a Distinguished Scholar-Teacher at the University of Maryland, College Park, MD, USA. Her research interests include information security and forensics, multimedia signal processing, and applications of data science and machine learning in health and IoT.

Prof. Wu received a U.S. NSF CAREER Award, a U.S. ONR Young Investigator Award, a TR100 Young Innovator Award from the MIT Technology Review, an IEEE Harriett B. Rigas Education Award, and an IEEE Signal Processing Society Meritorious Service Award. She has served as the Chair of the IEEE Technical Committee on Information Forensics and Security, the Vice President—Finance of the IEEE Signal Processing Society, and the Editor-in-Chief of the *IEEE Signal Processing Magazine*. Currently, she is serving as the 2024-2025 President of the IEEE Signal Processing Society. She is a Fellow of AAAS and the U.S. National Academy of Inventors.