IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 3, 2020, accepted January 15, 2020, date of publication January 20, 2020, date of current version January 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2967775

Robust QRS Detection Using High-Resolution Wavelet Packet Decomposition and Time-Attention Convolutional Neural Network

MENGHAN JIA[1], FEITENG LI[1], JIAQUAN WU[1], ZHIJIAN CHEN[1], AND YU PU[2]
[1]Institute of VLSI Design, Zhejiang University, Hangzhou 310027, China
[2]Alibaba DAMO Academy, Sunnyvale, CA 94085, USA

Corresponding author: Zhijian Chen (chenzj@vlsi.zju.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61801425.

ABSTRACT QRS detection is a crucial step in analyzing the electrocardiogram (ECG). For ECG collected by wearable devices, a robust QRS detection algorithm that yields high accuracy in spite of abnormal QRS morphologies and severe noise is needed. In this paper, we propose a QRS detection method based on high-resolution wavelet packet decomposition (HR-WPD) and convolutional neural network (CNN). Firstly, we design the HR-WPD that decomposes the ECG into multiple signals with different frequency bands to provide detailed QRS features. Secondly, all the decomposed signals are forwarded to a CNN for comprehensive morphology analysis and QRS prediction. To further improve the robustness, a time-attention module acting on the input signals is added to the CNN. Finally, a variable threshold is imposed to locate the QRS. The proposed method is validated by using two noisy databases (i.e., Telehealth Database (TELEDB) and MIT-BIH Noise Stress Test Database (NSTDB)) and one database with multiple ECG morphologies (i.e., MIT-BIH Arrhythmia Database (ARRDB)). The experiment results show that the proposed method achieves a comparable or even better performance compared with state-of-art methods on the TELEDB (SE 98.99%, P+ 95.57%, ER 5.61%, F1 97.25%), NSTDB (SE 99.25%, P+ 96.31%, ER 4.55%, F1 97.76%) and ARRDB (SE 99.89%, P+ 99.90%, ER 0.21%, F1 99.89%), suggesting that it is highly applicable to the QRS detection for ECG collected by wearable devices.

INDEX TERMS Electrocardiogram, convolutional neural network, wavelet packet decomposition, QRS detect.

I. INTRODUCTION

The electrocardiogram (ECG) is a graphical representation of the electric activity related to the functions of the heart. The conventional ECG comprises a sequence of P, Q, R, S, and T wave. Among them, the Q, R, and S waves compose the QRS-complex, which contains precious clinical information relates to cardiac health [1]. QRS detection directly affects heart rate variability measurement, heartbeat classification, diagnosis of heart diseases, etc [2].

Recently, as wearable ECG devices have been gaining popularity, ECG data can be recorded continuously for a long time to predict serious adverse events such as sudden cardiac death [3]. Automatic and computerized QRS detection is a promising research area because it could save experts' time. However, the morphological features of the QRS may change due to time, individuals, diseases or environments. Even worse, the ECG collected by wearable devices introduces severe noise [4], exacerbating the detection accuracy problem. Therefore, in the presence of multiple QRS morphologies and severe noise, a robust algorithm with high QRS detection performance is needed.

To solve these problems, various QRS detection algorithms have been proposed. Prior arts removes noise through filters, and highlight the QRS through means such as derivatives [5], [6], sixth power [7], wavelet transform [8]–[10], Hilbert transform [11], high-dimensional phase space [12], empirical mode decomposition [13], template matching [14], Savitzky-Golay filter [15], and multiple transforms [16]. All these algorithms generalize the QRS features manually. To ease feature generalization, only one frequency band of the ECG is selected. However, when there exists severe noise

The associate editor coordinating the review of this manuscript and approving it for publication was Junxiu Liu.

in this band, manually generalizing the features becomes quite challenging. Xiang *et al.* [17] used the difference to pre-process the ECG and convolutional neural network (CNN) to automatically generalize the QRS features. Though the difference operation suppressed some noise, it also restrained the network from obtaining useful low-frequency information hence being easily affected by high-frequency noise.

Aiming at more robust QRS detection, this paper proposes a method based on a high-resolution wavelet packet decomposition (HR-WPD) and a CNN. On the one hand, the HR-WPD decomposes the ECG into multiple signals with different frequency bands to expose the detailed QRS features in the frequency domain. On the other hand, the HR-WPD guarantees that the noise only affects parts of decomposed signals while the remaining signals can still facilitate the QRS detection, thus reducing the effects of the noise. All the decomposed signals are forwarded to a CNN to comprehensively analyze the morphologies and predict the QRS. Besides, we add a time-attention module in the CNN. By setting a specific weight to each moment, the proposed time-attention module helps the network allocate different attention to the different time and further improve the robustness. The robustness of the proposed algorithm is verified using two noisy databases, namely the Telehealth Database (TELEDB), MIT-BIH Noise Stress Test Database (NSTDB), and one database with multiple ECG morphologies, namely the MIT-BIH Arrhythmia Database (ARRDB).

The remainder of this paper is organized as follows. Section II briefly introduces the databases applied to this research. Section III presents the proposed QRS detection algorithm. The proposed method is evaluated with various ECG databases in Section IV. Section V discusses the experimental results. Section VI draws the conclusion.

II. RELEVANT DATABASES

Three public databases are used in this paper to evaluate the proposed QRS detection algorithm.

A. TELEHEALTH DATABASE

This ECG database is sampled at a rate of 500 Hz using dry metal Ag/AgCl plate electrodes (which the patient holds with each hand) and a reference electrode plate is positioned under the pad of the right hand. There are 300 ECG single lead-I signals recorded in a telehealth environment, 250 of which were selected randomly from 120 patients and the remaining 50 were manually selected from 168 patients to obtain a larger representation of poor quality data [18]. Three independent scorers annotated the data by identifying sections of artifact and QRSes. All scorers then annotated the signals as a group, to reconcile the individual annotations. Sections of the ECG signal which were less than 5s in duration were considered to be parts of the neighboring artifact sections and were subsequently masked. QRS annotations in the masked regions were discarded before the artifact mask and QRS locations being saved. Of the 300 telehealth ECG records in Redmond *et al.*, 50 records (including 29 of the 250 randomly selected records and 21 of the 50 manually selected records) were discarded as all annotated RR intervals within these records overlap with the annotated artifact mask and therefore, no heart rate can be calculated, which is required for measuring algorithm performance. The remaining 250 records will be referred to as the TELE database [16].

B. MIT-BIH ARRHYTHMIA DATABASE

The MIT-BIH Arrhythmia Database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979. 23 recordings were chosen at random from a set of 4000 24-hour ambulatory ECG recordings collected from a mixed population of inpatients (about 60%) and outpatients (about 40%) at Boston's Beth Israel Hospital; the remaining 25 recordings were selected from the same set to include less common but clinically significant arrhythmias that would not be well-represented in a small random sample. The recordings were digitized at 360 samples per second per channel with an 11-bit resolution over a 10 mV range. Two or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included in the database. The annotations have been modified several times since the database became public [19], [20].

C. MIT-BIH NOISE STRESS TEST DATABASE

This database includes 12 half-hour ECG recordings and 3 half-hour recordings of noise typical in ambulatory ECG recordings. The three noise records were assembled from the recordings by selecting intervals that contained predominantly baseline wander (BW) (in record 'bw'), muscle artifact (MA) (in record 'ma'), and electrode motion artifact (EM) (in record 'em'). Electrode motion artifact is generally considered the most troublesome, since it can mimic the appearance of ectopic beats and cannot be removed easily by simple filters, as can noise of other types. The ECG recordings were created using two clean recordings (118 and 119) from the aforementioned MIT-BIH Arrhythmia Database [19], to which calibrated amounts of noise from record 'em' were added using ''nst'' tool provided by the PhysioNet [21]. The noise was added at the beginning after the first 5 minutes of each record, during two-minute segments alternating with two-minute clean segments. The signal-to-noise ratios (SNRs) during the noisy segments of these records are 24, 18, 12, 6, 0, −6dB [20], [21].

III. THE PROPOSED METHOD

The proposed method consists of three major steps: (a) preprocess the ECG with the HR-WPD method, (b) predict the QRS by the time-attention CNN and (c) locate the QRS through a variable threshold. The block diagram of the proposed method is shown in Fig. 1.

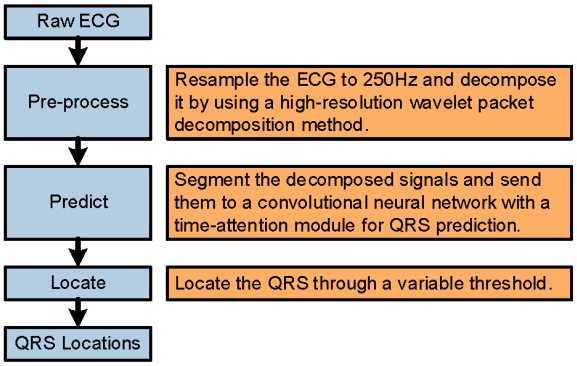


FIGURE 1. The block diagram of the proposed QRS detection method.

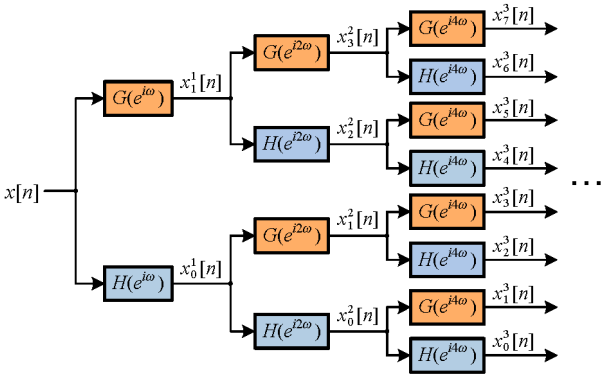


FIGURE 2. The implementation of the proposed HR-WPD.

A. PRE-PROCESS WITH THE HR-WPD

Different ECG sensors use different sampling rates, which brings challenges to QRS detection. To solve the issue, we resample the original ECG to 250Hz to unify the QRS detection parameters. To provide detailed QRS features in the frequency domain without losing time resolution, inspired by [9], we propose a high-resolution wavelet packet decomposition (HR-WPD) method, as shown in Fig. 2. In the HR-WPD, the input is first filtered by a low-pass filter $H(e^{i\omega})$ and a high-pass filter $G(e^{i\omega})$ from the wavelet to get two decomposed signals x_0^1 and x_1^1. Then the impulse responses of the filters are zero interpolated and all the signals are further decomposed by the new filters. This step is repeated to recursively decompose the signal. Compared with the conventional wavelet packet decomposition [22], in which a $2\times$ downsampling step is required for each decomposition, the proposed HR-WPD removes the downsampling step and achieves higher time resolution, which helps predict the accurate QRS occurrence time.

We decompose the signal by using the filters from the quadratic spline wavelet [23]. Because of the unique filtering features, these filters are considered suitable for QRS analysis in [9]. The corresponding low-pass filter $H(e^{i\omega})$ and high-pass filter $G(e^{i\omega})$ are defined as follows,

$$\begin{aligned} H(e^{i\omega}) &= e^{i\omega/2}\left(\cos\frac{\omega}{2}\right)^3 \\ G(e^{i\omega}) &= 4ie^{i\omega/2}\left(\sin\frac{\omega}{2}\right) \end{aligned} \tag{1}$$

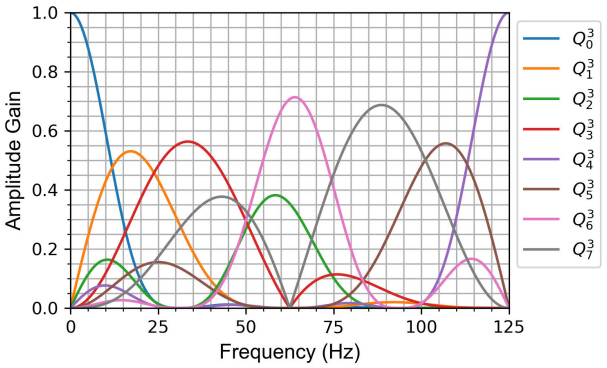


FIGURE 3. The amplitude-frequency responses of the equivalent filters $Q_p^3(e^{i\omega})$ for the HR-WPD at 250Hz sampling rate.

By decomposing the ECG 3 times, we get 8 signals (i.e., x_0^3 to x_7^3). Let $Q_p^3(e^{i\omega})$ be the transform functions of the equivalent filters in the HR-WPD shown in Fig. 2. For example, $Q_0^3(e^{i\omega}) = H(e^{i\omega})H(e^{i2\omega})H(e^{i4\omega})$, and $Q_1^3(e^{i\omega}) = H(e^{i\omega})H(e^{i2\omega})G(e^{i4\omega})$. The amplitude-frequency responses of $Q_p^3(e^{i\omega})$ at 250Hz sampling rate are shown in Fig. 3.

Each decomposed signal mainly contains the information of a specific frequency band, and all the signals provide more detailed QRS features in the frequency domain. Besides, when severe noise distributes in a certain frequency region, the HR-WPD guarantees that the noise only exists in a part of decomposed signals. As a result, the signals which are less polluted provide more accurate QRS features, thus reducing the effect of noise.

To cancel the impact of the overall signal scaling, we standardize the decomposed signals, i.e. set the average of x_p^3 to "0", variance to "1" and then get x'^3_p. Fig. 4 shows three examples of the ECG and their decomposed signals. For ECG with some occasional premature ventricular contractions (PVCs) (as shown in Fig. 4 (a)), despite the fact that x'^3_6 and x'^3_7 are noisy and the QRS in x'^3_0 is not obvious, the QRS features in other signals are significant, especially in x'^3_3 and x'^3_5. For ECG with high-frequency noise, as shown in Fig. 4 (b), the proposed HR-WPD method obtains some decomposed signals which are less polluted by noise, such as x'^3_0, x'^3_1, x'^3_4, thus helping to analyze the morphological features and predict the QRS. For ECG with the EM noise, x'^3_3 and x'^3_7 are less affected by the motion artifacts, thus helping to detect the QRS, as shown in Fig. 4 (c).

B. CNN PREDICTION

Taking advantage of the fact that CNN is good at integrating information from multiple signals, we design a CNN to comprehensively analyze the morphologies from all the decomposed signals and predict the QRS. Fig. 5(a) shows the architecture of the proposed CNN. The input signals consist of 8 channels which are segmented from the signals x'^3_p by a window. We expect that the network input contains at least one QRS, which is beneficial for morphological analysis. On further considering the size of the network and the ease of

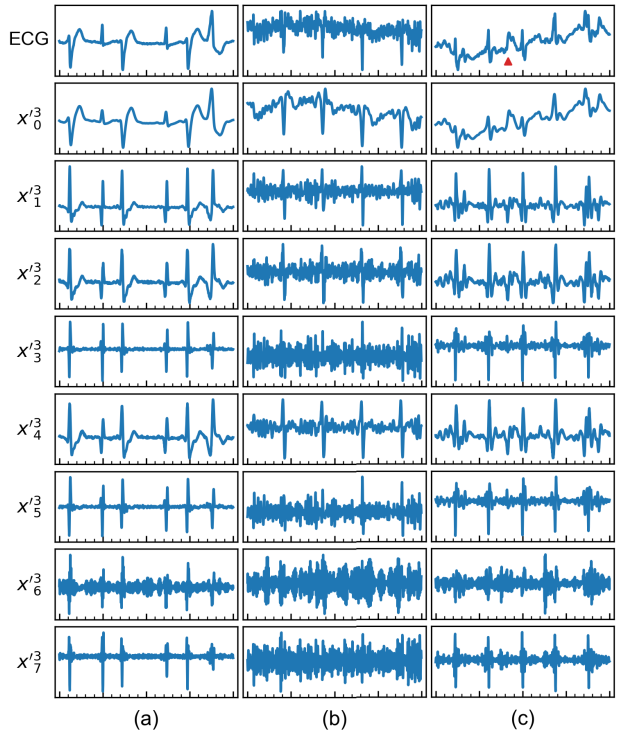


FIGURE 4. Three examples of the ECG and the decomposed signals by using the proposed HR-WPD. The ECG in (a) is from the recording 200 in the ARRDB, including some occasional PVCs. The ECG in (b) is from the recording 051 in the TELEDB, including high-frequency noise. The ECG in (c) is from the recording 118 including 6dB EM noise in the NSTDB. The red triangle in (c) indicates a sudden fluctuation of the EM noise which looks like an ectopic heartbeat.

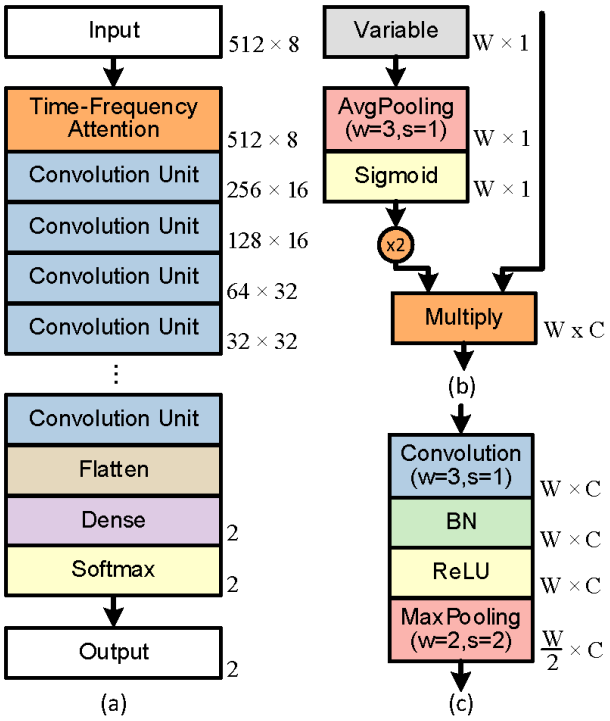


FIGURE 5. (a) The architecture of the proposed network. (b) The architecture of the time-attention module. (c) The structure of the Convolutional unit. Where "BN" stands for the batch-normalization layer, "ReLU" stands for the rectified linear activation unit, "W" corresponds to the width of the feature map, "C" corresponds to the channel number, "w" corresponds to the width of the convolution kernels or pooling, and "s" corresponds to the stride of the convolution or pooling.

pooling data, we select 512 points (2.048s) as the width of the network input (also the width of the segmentation window). The network only classifies the midpoint (set as the 256th point) of the input each time and the segmentation window moves one point per time to reduce the effect of individual prediction errors on the QRS detection. If the distance between the midpoint and the nearest R-peak occurrence time is $\leq$15 points (60ms), this point will be regarded in the QRS region and classified as "1", otherwise it will be regarded out of the QRS region and classified as "0", as shown in Fig. 6. This classification method forces the network to predict each QRS region multiple times and keeps the distance between two consecutive QRS regions large enough to avoid the region overlap.

Since the network only classifies the midpoint, the importances vary with different input time. However, due to the calculation characteristic of the convolution, it transforms the signals from different time in the same way. Therefore, we add a time-attention module at the begin of the network to automatically emphasize the important time. The structure of the proposed time-attention module is shown in Fig. 5(b).

The time-attention $\mathbf{A} = [a_1, a_2, \ldots a_W]$ provides a coefficient for each moment and rescales the feature map $\mathbf{U}$ by the coefficient, $\hat{\mathbf{U}} = \mathbf{A} \cdot \mathbf{U}$, where $\mathbf{U}, \hat{\mathbf{U}} \in \mathbb{R}^{W \times C}$. W corresponds to the width of the feature map, and C corresponds to the channel number. Since we only apply the time-attention module to the network input, the W and C are set as 512 and 8, respectively. We add a group of trainable variables $\mathbf{V} = [v_1, v_2, \ldots v_W]$ in the time-attention module. Each variable $v_j, j = 1 \ldots W$ corresponds to a certain moment. Given the continuity of the ECG, we first use an average pooling operation with a width of 3 and stride of 1 to avoid excessive attention differences between adjacent points. Then we employ a single gating mechanism with a sigmoid activation and multiply the data by 2 to obtain the final attention, as shown below:

$$a_i = 2\sigma\left(\frac{1}{3}\sum_{i-1}^{i+1} v_i\right) \tag{2}$$

All the variables in the time-attention module are initialized to "0", in which case $\hat{\mathbf{U}} = \mathbf{U}$ and the network pays the same attention to all the moments before training. As the training goes on, the network redistributes the attention automatically.

We use multiple convolution units arranged in serial to preliminary extract morphology features of the signal. The structure of the proposed convolution unit is shown in Fig. 5(c). The batch-normalization (BN) and rectified linear activation units (ReLU) help the network get higher learning rates [24], [25], and the maxpooling operations are used to

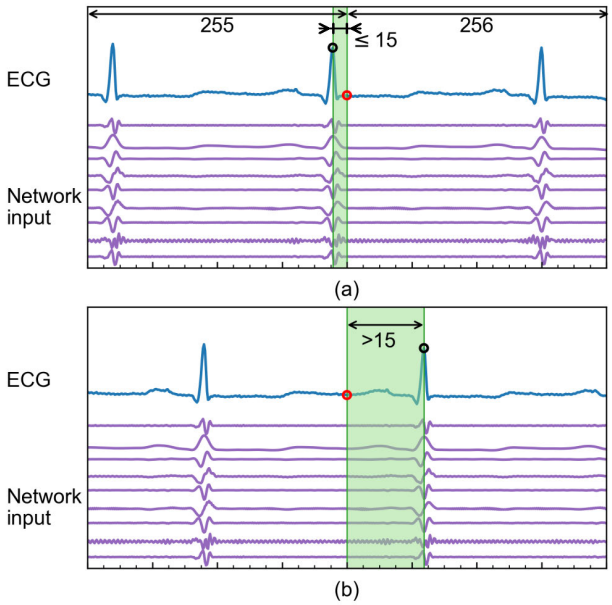


FIGURE 6. Two cases of the network classification. The black circle is the manually identified R-peak occurrence time. The red circle is the midpoint of the input, which is the current classification point. The time region colored in green indicates the distance between the midpoint and the nearest R-peak occurrence time. The midpoint shown in (a) is in the QRS region and labeled "1" and the midpoint shown in (b) is out of the QRS region and labeled "0".

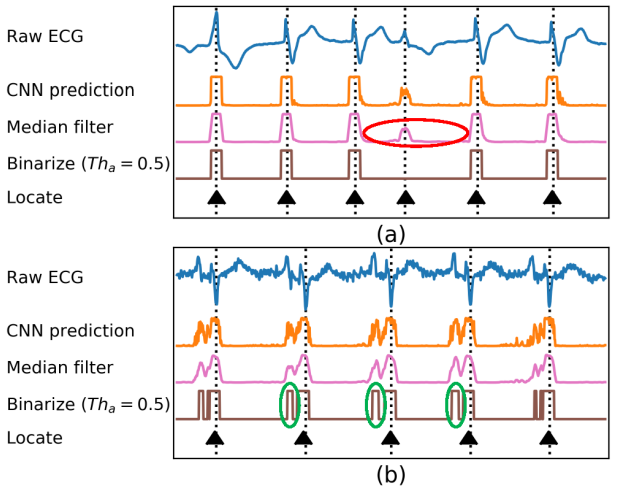


FIGURE 7. Two examples of using the proposed variable threshold method for QRS location from (a) recording 217 in the ARRDB (b) recording 108 in the ARRDB. The dotted lines are the manually identified R-peak occurrence time.

remove redundant information, reduce the network size and to broaden the horizon of the next convolutional layer [26]. 8 convolution units in the first two layers are adopted to extract the limited local morphological features such as the slope feature. Each time the data pass through two convolutional layers, the number of the convolution units is doubled to obtain more global and combined information of the QRS. After the convolutional layers, we add a dense layer (also called the fully-connected layer) with 2 nodes to analyze the convolutional outputs altogether and predict whether the midpoint of the inputs is in the QRS region. The output of the dense layer is fed into a softmax activation layer to produce the final prediction probabilities. During training, we maintain the training class balance by randomly removing redundant segments to avoid the CNN tending to predict as the majority class [27].

C. VARIABLE THRESHOLD FOR QRS LOCATION

After the CNN prediction, we get the probabilities of each point belonging to the QRS regions, as shown as the orange line in Fig. 7. As there may be burrs in the sequence, we first pass the predictions to a median filter. Then, we binarize the predictions with an amplitude threshold $Th_a = 0.5$ and select all the positive time regions with the width larger than 30ms (half the defined QRS region width) to distinguish the QRS regions from false predictions. Since the normal heart rate is greater than 50 beats per minute [28] (that is, the average interval between two adjacent heartbeats is less than 1.2s), when the distance between two adjacent QRS regions is larger than 1.2s, we have likely missed some QRS. So we slightly lower Th_a to 0.3 in this duration and search for the QRS again, as shown in the duration circled in red in Fig. 7(a). This variable threshold method reduces the probability of missed detection, thus improving the detection performance. When the distance between two predicted QRS regions is less than 0.2s which rarely happens in the real world [29], we remove the one with a smaller width, as shown in the region circled in green in Fig. 7(b). Finally, the midpoints of all the QRS regions are considered to be the QRS locations.

IV. EXPERIMENTAL RESULTS

A. EVALUATION APPROACH

The proposed HR-WPD and QRS location algorithm are implemented in Python, and the proposed time-attention CNN is trained by Keras [30], which is a high-level Python library. Keras allows for easy and fast prototyping of neural networks through user-friendliness, modularity, and extensibility.

We use the TELEDB, NSTDB, and ARRDB to evaluate the robustness of the proposed method. Among them, the TELEDB includes the ECG with true noise and the NSTDB includes the ECG with EM noise which is generally considered the most troublesome [21] and the ARRDB includes the ECG with arrhythmia, abnormal QRS morphologies, and high amplitude P, T waves.

We adopt the 5-fold randomized cross-validation [31] to accurately evaluate the performance of this method on all the recordings in the TELEDB and ARRDB. For the NSTDB, since all the evaluation recordings are mixed with the same EM noise, adopting the K-fold cross-validation will cause data leakage, so we extract the unused parts of the EM noise provided in the NSTDB and add it to recording 200-234 in the ARRDB as the training set. All evaluation recordings provided in the NSTDB are set as the testing set.

TABLE 1. The F1 of different networks with different convolutional layers on the TELEDB, NSTDB and ARRDB and the p-values obtained by the independent samples t-test of F1.

N_l	Trainable Parameters	TELEDB		NSTDB		ARRDB	
		F1(%)	p-value	F1(%)	p-value	F1(%)	p-value
2	5,826	96.35	—	96.36	—	99.80	—
3	7,426	96.56	0.001	96.99	0.001	99.84	0.000
4	8,514	96.81	0.000	97.15	0.118	99.85	0.275
5	14,786	97.02	0.000	97.28	0.282	99.86	0.587
6	26,178	97.25	0.000	97.76	0.000	99.89	0.004
7	51,010	97.27	0.884	97.79	0.540	99.89	0.709
8	99,906	97.22	0.154	97.64	0.022	99.86	0.000

Following ANSI/AAML EC38 [32] and EC57 [33], if the predicted QRS is within 150ms of an annotated QRS, this prediction is categorized as a true positive (TP, indicating that a QRS is detected correctly). The QRS missed by the algorithm is classified as a false negative (FN) and the false detection by the algorithm is classified as a false positive (FP). Sensitivity (SE), positive predictivity (P+), error rate (ER) and F1-score (F1) are used as four evaluation metrics. They are defined as equation 3. SE is the ratio of the QRS correctly detected using the proposed method to the true beats (TB) by the given annotation; P+ is the ratio of the TP to all the detected QRS; ER is the ratio of false detections (FP and FN) over the TB by the given annotation [15]; F1 provides a comprehensive result based on SE and P+. Since ER and F1 both consider the information of TP, FP, and FN detections in a signal, they can be regarded as comprehensive indicators of the QRS detection performance.

$$SE = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

$$P+ = \frac{TP}{TP + FP} \times 100\% \tag{4}$$

$$ER = \frac{FP + FN}{TB} \times 100\% \tag{5}$$

$$F1 = \frac{2 \cdot SE \cdot P+}{SE + P+} \times 100\% \tag{6}$$

B. PERFORMANCE COMPARISON OF THE NETWORKS WITH DIFFERENT CONVOLUTIONAL LAYERS

To obtain the most suitable network structure for QRS detection, we adjust the number of the convolutional layers (defined as N_l) in the CNN as 2-8 and use F1 values as the comprehensive indicators to evaluate the performance of different networks with different convolutional layers on the TELEDB, NSTDB, ARRDB. In addition, we apply the independent samples t-test to evaluate the significance of the F1 change for each additional convolutional layer. The results are shown in Table 1.

We can see that with N_l increasing, the F1 improves at first. When N_l is 6 and 7, the network obtains the highest F1 on all the evaluated databases. When N_l continues to increase, the F1 drops. This may be because the output data width of the 8th convolutional layer has been reduced to 2, resulting in the loss of time information. As N_l increases from 6 to 7, the p-values of the three databases are all larger than 0.5, meaning that the QRS detection performance of the network does not improve significantly. Taking both the detection performance and computation amount into consideration, we choose the network with 6 convolutional layers as the most suitable network for QRS detection. The following experimental results are all based on the network with 6 convolutional layers.

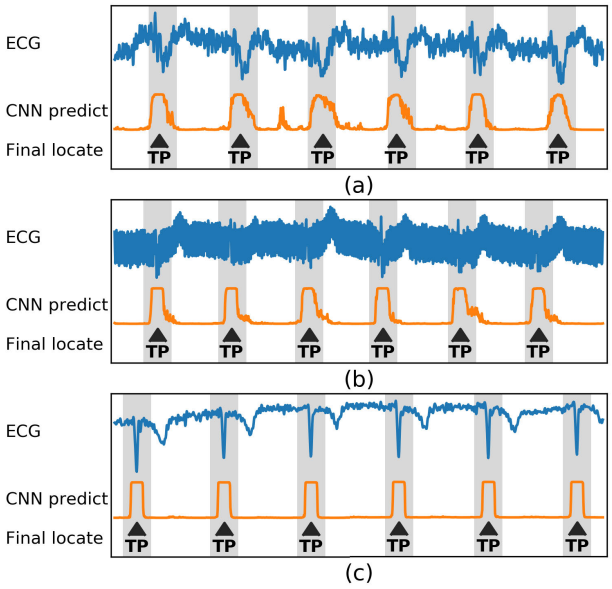


FIGURE 8. Examples of the proposed QRS detection on the TELEDB from (a) recording 051, which contains high-intensity mixed noise, (b) recording 169, which contains high-intensity powerline interference noise and (c) recording 007, which contains both noise and high-amplitude T-waves. The QRS detected during the time region colored in grey is regarded as a TP.

C. PERFORMANCE ON THE TELEHEALTH DATABASE

This method achieves 98.99% SE, 95.57% P+, 5.61% ER and 97.25% F1 on the TELEDB. Three representative examples of the proposed QRS detection method on the TELEDB are shown in Fig. 8. When the ECG contains high-intensity noise or contains both noise and high-amplitude T-waves, the proposed method still predicts the QRS correctly and achieves high robustness.

Table. 2 compares the performance of the proposed QRS detection method with several other existing methods. Since the TELEDB was available in 2016, only a few papers used

TABLE 2. The comparison of the QRS detection performance on the TELEDB. Where the SE, P+, ER, and F1 are calculated by the summarized TP, FP, FN, and TB from all the ECG recordings.

Method	TB	SE(%)	P+(%)	ER(%)	F1(%)
Pan *et al.* [5] [a]	6708	96.2	88.53	16.26	92.21
Arzeno *et al.* [6] [a]	6708	93.9	90.6	15.85	92.22
Manikandan *et al.* [34] [a]	6708	88.18	78.22	36.37	82.90
Khamis *et al.* [16]	6708	98.05	**95.75**	6.31	96.89
Proposed	6708	**98.99**	95.57	**5.61**	**97.25**

[a] Reproduced by Kashif *et al* [4].

TABLE 3. The detection performance of the proposed QRS detection method on the NSTDB.

SNR	Rec	TB	FN	FP	SE(%)	P+(%)	ER(%)	F1(%)
24	118	2278	0	3	100	99.87	0.13	99.93
	119	1987	0	1	100	99.95	0.05	99.97
18	118	2278	0	3	100	99.78	0.22	99.89
	119	1987	0	19	100	99.6	0.4	99.80
12	118	2278	0	14	100	99.69	0.31	99.85
	119	1987	0	71	100	98.32	1.71	99.15
6	118	2278	0	31	99.96	99.09	0.97	99.52
	119	1987	2	108	100	96.83	3.27	98.39
0	118	2278	10	91	99.56	96.51	4.04	98.01
	119	1987	8	168	99.7	93.66	7.05	96.59
-6	118	2278	99	296	94.6	87.46	18.96	90.89
	119	1987	48	296	97.43	86.43	17.87	91.60
Sum	—	25590	167	1099	99.25	96.31	4.55	97.76

this database to evaluate the robustness of their algorithms. To compare the proposed method with more methods, we add the methods from [5], [6], [34], [35] reproduced by Kashif *et al* in [4] for comparison. In Table. 2, though [16] achieves higher P+, the proposed method achieves the highest SE, F1 and the lowest ER, showing the best comprehensive detection performance.

D. PERFORMANCE ON THE MIT-BIH NOISE STRESS TEST DATABASE

Table. 3 shows the results of the proposed method on the NSTDB. In general, our method achieves 99.25% SE, 96.31% P+, 4.55% ER and 97.76% F1. Fig. 9 shows a representative example of the QRS detection using the proposed method on the NSTDB. When the ECG is contaminated by 0dB EM noise, this method can still distinguish the QRS-complex from the motion artifacts.

Table. 4 compares the performance of the proposed QRS detection method with several other existing methods. The detection performance of the proposed method on the NSTDB is much better than other methods.

E. PERFORMANCE ON THE MIT-BIH ARRHYTHMIA DATABASE

The proposed method achieves 99.89% SE, 99.90% P+, 0.21% ER and 99.89% F1 on the ARRDB in general. Fig. 10 shows some critical cases for QRS detection. When the evaluated ECG contains PVCs, high-amplitude P, T waves or varying R-peak amplitudes, the proposed method keeps high robustness and detects the QRS correctly.

Table. 5 compares the performance of the proposed method with several other existing methods. Xiang *et al* [17] excluded

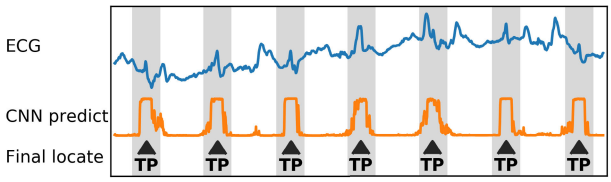


FIGURE 9. An Example of the proposed QRS detection method on the NSTDB from recording 118 containing 0dB EM noise. The QRS detected during the time region colored in grey is regarded as a TP.

TABLE 4. The comparison of the QRS detection performance on the NSTDB. Where the SE, P+, ER, and F1 are calculated by the summarized TP, FP, FN, and TB from all the ECG recordings.

Method	TB	SE(%)	P+(%)	ER(%)	F1(%)
Dohare *et al.* [7]	25590	88.09	88.93	N/R	88.51
Merah *et al.* [10]	25590	95.30	93.98	10.81	94.64
Khamis *et al.* [16]	25590	93.14	86.23	21.73	89.55
Lee *et al.* [15]	25590	94.08	93.45	12.57	93.76
Proposed	25590	**99.25**	**96.31**	**4.55**	**97.76**

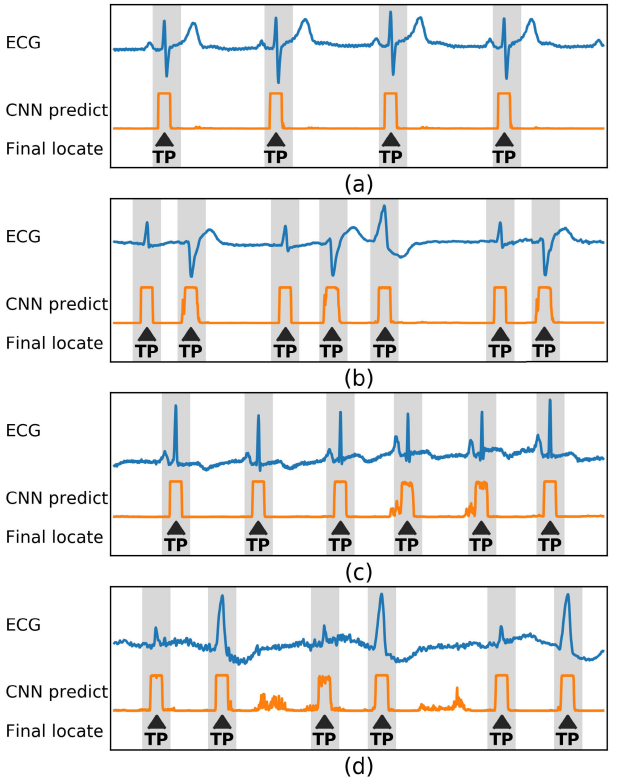


FIGURE 10. Examples of the proposed QRS detection method on the ARRDB from (a) recording 117, which contains high amplitude of the T-waves, (b) recording 200, which contains some PVCs, (c) recording 222, which contains the high amplitude P-waves and (d) recording 228, which contains greatly varying R-peak amplitudes. The QRS detected during the time colored in grey is regarded as TP.

recording 102 and 104 when evaluating their algorithm with the ARRDB. To compare the proposed method with more methods, we use both all the recordings and the recordings excluded 102 and 104 in the ARRDB for evaluation, as listed in Table. 5. Since the ARRDB contains less noise, most methods have high detection performance. The ER and F1 of

TABLE 5. The comparison of the QRS detection performance on the ARRDB. Where the SE, P+, ER, and F1 are calculated by the summarized TP, FP, FN, and TB from the evaluated recordings in the ARRDB.

Method	If Evaluated on Other Noisy Databases	TB	SE(%)	P+(%)	ER(%)	F1(%)
Pan *et al*. [5]	TELEDB by [4]	109809 [a]	99.56	99.76	0.68	99.66
Arzeno *et al*. [6]	TELEDB by [4]	109456	99.92	99.24	1.48	99.58
Manikandan *et al*. [34]	TELEDB by [4]	109496	**99.93**	99.86	**0.21**	**99.89**
Dohare *et al*. [7]	NSTDB	109966	99.21	99.34	1.45	99.27
Merah *et al*. [10]	NSTDB	109494	99.84	99.88	0.28	99.86
Khamis *et al*. [16]	TELEDB & NSTDB	109494	99.76	99.80	0.44	99.78
Xiang *et al*. [17]	No	105078 [b]	99.77	**99.91**	0.32	99.84
Proposed	TELEDB & NSTDB	105078 [b]	99.88	99.90	0.22	99.89
		109494	99.89	99.90	**0.21**	**99.89**

[a] Values computed according to the record-by-record table in the referred work since there is a discrepancy between total values and the sum of the individual ones.
[c] These evaluations excluded recording 102 and 104.

the proposed method are the same as the lowest ER and highest F1 in Table. 5, proving that this method yields a comparable result to state-of-art methods for ECG with less noise and more abnormal morphologies. Compared with [17], which also used a CNN for the QRS detection, although [17] has got higher P+, its SE is lower, resulting in a lower comprehensive QRS detection performance than the proposed method. Although [34] also achieved a very high QRS detection performance on the ARRDB, [4] reproduced this method and found that its noise tolerance is poor, as shown in Table. 2.

V. DISCUSSION

A. SPECIFIC EFFECT OF THE TIME-ATTENTION MODULE

To enhance the influence of the input in important time and reduce the interference in unimportant time, a time-attention model is added in this paper. After training, we get the time-attention **A** as described in Equ. 2 from the network. Fig. 11 draws the time-attention when using the TELEDB, NSTDB, and ARRDB for evaluation. It shows that the network pays more attention to the time regions which are most relevant to the classification, i.e., within 15 points (60ms) from the midpoint. When the distance from the midpoint is more than 60ms, the time-attention drops rapidly, which helps the CNN classify the boundary of the QRS region. Besides, the network also pays attention to the time regions of 0.2s to 0.36s from the midpoint. It may mean that the P, T waves help the network to predict the QRS. This is because, when a heartbeat occurs at the midpoint, the adjacent P-wave and T-wave occurs likely in these two time regions. To further explore the effectiveness of the time-attention module, we remove the time-attention module from the network with 6 convolutional layers and use the same pre-processing and post-processing algorithms for QRS detection. The QRS detection results, the F1 difference compared with the network with a time-attention module and the corresponding p-values of the independent samples t-test on F1 are shown in Table. 6. Compared with Table 2,4,5, we can see that the time-attention module results in F1 improvements from 0.01% to 0.14%.

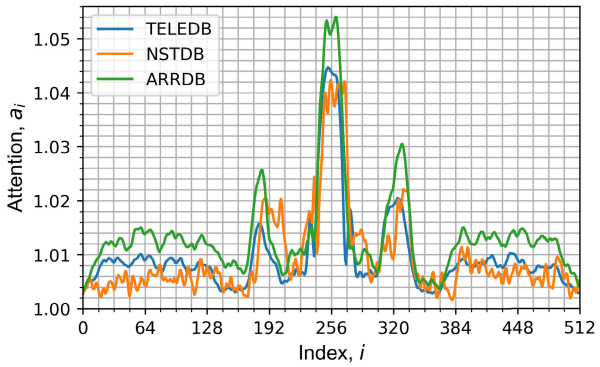


FIGURE 11. The after-trained time-attention A when using the TELEDB, NSTDB, and ARRDB for evaluation. Where the time-attention curve of the TELEDB or ARRDB is obtained by calculating the average of all the time-attention curves from the 5-fold cross-validation, and the time-attention curve of the NSTDB is trained by recording 200-234 in the ARRDB mixed with the EM noise in the NSTDB, as described in Section IV.A.

B. THEORETICAL ANALYSIS FOR THE RESULTS IN DIFFERENT DATABASES

For the TELEDB, in which there are many recordings with true noise [16], the compared methods in Table 2 applied bandpass filters for pre-processing to filter out the high-frequency and low-frequency noise. However, using the bandpass filter loses both the low-frequency and high-frequency information of the QRS, thus dropping the detection performance of these methods will drop when the noise occurs in the passband of the filter. Although the proposed method retains noise, the HR-WPD can acquire the signals less affected by the noise when it only distributed in certain frequency bands, and the CNN can comprehensively analyze these signals and detect the QRS. Therefore, this method reaches the lowest ER and the highest SE and F1.

For the NSTDB, because the EM noise widely distributed in the NSTDB distributes in about 1-10Hz [36], the SNR of the ECG with the EM noise in the high-frequency band is higher than that in the low-frequency band. Losing the high-frequency information, the QRS detection algorithms with the conventional bandpass filters (such as [15], [16]) or the wavelet transforms similar to bandpass filters (such as [10])

TABLE 6. The detection performance of the proposed QRS detection method without the time-attention module on the TELEDB, NSTDB, and ARRDB, the F1 differences compared with the network with the time-attention module and the corresponding p-values of the independent samples t-test on F1.

Database	SE(%)	P+(%)	ER(%)	F1(%)	F1 diff(%)	F1 p-value
TELEDB	98.96	95.40	5.81	97.15	-0.10	0.030
NSTDB	99.45	95.85	4.86	97.62	-0.14	0.262
ARRDB	99.85	99.92	0.23	99.88	-0.01	0.164

all have poor robustness for ECG with the EM noise. This method, however, needs to comprehensively analyze the high-frequency part and low-frequency part of the ECG, demonstrating higher robustness against the EM noise.

For the ARRDB, compared with the difference operation applied before the CNN in [17], the proposed HR-WPD is more advantageous for the CNN to obtain full frequency band QRS features. Besides, the proposed time-attention model helps the CNN focus on more important time regions. So this method results in higher F1 and lower ER than [17]. However, from Table 5, we can see that the QRS detection performance of this method does not improve a lot in the case of a significant increase in the computation amount. The first reason is that the simple band-pass filter is enough to filter out the high amplitude P, T waves and retain the QRS. The second reason is that different types of the QRS are unevenly distributed in the ARRDB, (for example, all the 16 atrial escape beats distribute in the recording 223 [19],) which reduces the representative of the training samples in K-fold cross-validation, thus reducing the QRS detection performance of the proposed method.

C. LIMITATIONS OF THE PROPOSED METHOD

Though the proposed method shows comparable or better performance on multiple public databases, its limitations should be recognized. First, as a machine learning algorithm, the proposed method is highly dependent on the training data. When there is a big difference between the training set and the testing set, this method may not be robust enough. Second, when the ECG does not contain noise, this method does not improve the detection performance significantly compared with state-of-art methods. Third, although we strive to control the size of the network, the proposed method still needs a larger amount of computation compared with traditional methods. Therefore, this method is more suitable for detecting the QRS on the cloud/edge server from the ECG data collected by wearable devices at present. Nonetheless, with the advancement of the neural network accelerator, this method could eventually work on wearable devices in the future.

VI. CONCLUSION

In this paper, we have proposed a robust QRS detection method including an HR-WPD for pre-processing and a CNN with a time-attention module for prediction. The HR-WPD decomposes the ECG to multiple signals and induces the generalization of QRS features in different frequency bands to reduce the effects of noise. Through setting a specific weight to each moment, the time-attention module helps the neural network allocate different attention to different time, hence further improving the robustness. The proposed method is evaluated on the TELEDB, NSTDB, and ARRDB. The experiment results indicate that our method achieves comparable and even better performance compared with state-of-art methods on the TELEDB (SE 98.99%, P+ 95.57%, ER 5.61%, F1 97.25%), NSTDB (SE 99.25%, P+ 96.31%, ER 4.55%, F1 97.76%) and ARRDB (SE 99.89%, P+ 99.90%, ER 0.21%, F1 99.89%). For ECG with abnormal QRS morphologies, high amplitude P, T waves and severe noise, the proposed method yields consistently accurate QRS detection results, suggesting that it is highly applicable to the QRS detection from the ECG signals collected by wearable end-points.

REFERENCES

[1] G. F. Fletcher, P. A. Ades, P. Kligfield, R. Arena, G. J. Balady, V. A. Bittner, L. A. Coke, J. L. Fleg, D. E. Forman, T. C. Gerber, M. Gulati, K. Madan, J. Rhodes, P. D. Thompson, and M. A. Williams, "Exercise standards for testing and training: A scientific statement from the American heart association," *Circulation*, vol. 128, no. 8, pp. 873–934, Aug. 2013.

[2] C. Nayak, S. K. Saha, R. Kar, and D. Mandal, "Automated QRS complex detection using MFO-based DFOD," *IET Signal Process.*, vol. 12, no. 9, pp. 1172–1184, Dec. 2018.

[3] C. W. Israel, "Mechanisms of sudden cardiac death," *Indian Heart J.*, vol. 66, pp. S10–S17, Jan. 2014, doi: 10.1016/j.ihj.2014.01.005.

[4] M. Kashif, S. M. Jonas, and T. M. Deserno, "Deterioration of R-wave detection in pathology and noise: A comprehensive analysis using simultaneous truth and performance level estimation," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2163–2175, Sep. 2017.

[5] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vols. BME-32, no. 3, pp. 230–236, Mar. 1985.

[6] N. M. Arzeno, Z.-D. Deng, and C.-S. Poon, "Analysis of first-derivative based QRS detection algorithms," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 478–484, Feb. 2008.

[7] A. K. Dohare, V. Kumar, and R. Kumar, "An efficient new method for the detection of QRS in electrocardiogram," *Comput. Electr. Eng.*, vol. 40, no. 5, pp. 1717–1730, Jul. 2014.

[8] C. Li, C. Zheng, and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 1, pp. 21–28, 1995.

[9] J. Martinez, R. Almeida, S. Olmos, A. Rocha, and P. Laguna, "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.

[10] M. Merah, T. Abdelmalik, and B. Larbi, "R-peaks detection based on stationary wavelet transform," *Comput. Methods Programs Biomed.*, vol. 121, no. 3, pp. 149–160, Oct. 2015.

[11] D. Benitez, P. Gaydecki, A. Zaidi, and A. Fitzpatrick, "A new QRS detection algorithm based on the Hilbert transform," in *Proc. Comput. Cardiol.*, Cambridge, MA, USA, vol. 27, 2000, pp. 379–382.

[12] Q. Zhang, D. Zhou, and X. Zeng, "A novel framework for motion-tolerant instantaneous heart rate estimation by phase-domain multiview dynamic time warping," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 11, pp. 2562–2574, Nov. 2017.

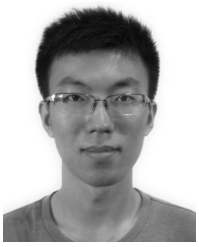
[13] H. Li, X. Wang, L. Chen, and E. Li, "Denoising and R-peak detection of electrocardiogram signal based on EMD and improved approximate envelope," *Circuits Syst Signal Process*, vol. 33, no. 4, pp. 1261–1276, Apr. 2014.

[14] Y. Nakai, S. Izumi, M. Nakano, K. Yamashita, T. Fujii, H. Kawaguchi, and M. Yoshimoto, "Noise tolerant QRS detection using template matching with short-term autocorrelation," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, USA, Aug. 2014, pp. 34–37.

[15] M. Lee, D. Park, S.-Y. Dong, and I. Youn, "A novel R peak detection method for mobile environments," *IEEE Access*, vol. 6, pp. 51227–51237, 2018.

[16] H. Khamis, R. Weiss, Y. Xie, C.-W. Chang, N. H. Lovell, and S. J. Redmond, "QRS detection algorithm for telehealth electrocardiogram recordings," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1377–1388, Jul. 2016.

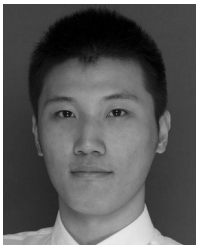
[17] Y. Xiang, Z. Lin, and J. Meng, "Automatic QRS complex detection using two-level convolutional neural network," *Biomed. Eng. OnLine*, vol. 17, no. 1, p. 13, Dec. 2018.
[18] S. J. Redmond, Y. Xie, D. Chang, J. Basilakis, and N. H. Lovell, "Electrocardiogram signal quality measures for unsupervised telehealth environments," *Physiol. Meas.*, vol. 33, no. 9, pp. 1517–1533, Sep. 2012.
[19] G. Moody and R. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.
[20] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
[21] G. B. Moody, W. K. Muldrow, and R. G. Mark, "A noise stress test for arrhythmia detectors," *Comput. Cardiol.*, vol. 11, no. 3, pp. 381–384, 1984.
[22] K. Teotrakool, M. Devaney, and L. Eren, "Adjustable-speed drive bearing-fault detection via wavelet packet decomposition," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 8, pp. 2747–2754, Aug. 2009.
[23] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 710–732, Jul. 1992.
[24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 37, 2015, pp. 448–456.
[25] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1. Red Hook, New York, USA: Curran Associates, 2012, pp. 1097–1105.
[27] P. Hensman and D. Masko, "The impact of imbalanced training data for convolutional neural networks," Degree Project Comput. Sci., KTH Roy. Inst. Technol., Stockholm, Sweden, 2015.
[28] D. H. Spodick, "Survey of selected cardiologists for an operational definition of normal sinus heart rate," *Amer. J. Cardiol.*, vol. 72, no. 5, pp. 487–488, Aug. 1993.
[29] J. Karvonen and T. Vuorimaa, "Heart rate and exercise intensity during sports activities," *Sports Med.*, vol. 5, no. 5, pp. 303–312, May 1988.
[30] (Dec. 2019). *Keras-Team/Keras*. [Online]. Available: https://github.com/keras-team/keras
[31] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, no. 0, pp. 40–79, 2010.
[32] *Ambulatory Electrocardiographs*, Standard ANSI/AAMI EC38, Association for the Advancement of Medical Instrumentation, 1998.
[33] *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, Standard ANSI/AAMI EC57, Association for the Advancement of Medical Instrumentation, 1998.
[34] M. Manikandan and K. Soman, "A novel method for detecting R-peaks in electrocardiogram (ECG) signal," *Biomed. Signal Process. Control*, vol. 7, no. 2, pp. 118–128, Mar. 2012.
[35] X. Liu, J. Yang, X. Zhu, S. Zhou, H. Wang, and H. Zhang, "A novel R-peak detection method combining energy and wavelet transform in electrocardiogram signal," *Biomed. Eng. Appl. Basis Commun.*, vol. 26, no. 01, Feb. 2014, Art. no. 1450007.
[36] A. Sulthana, M. Z. U. Rahman, and S. S. Mirza, "An efficient Kalman noise canceller for cardiac signal analysis in modern telecardiology systems," *IEEE Access*, vol. 6, pp. 34616–34630, 2018.



MENGHAN JIA received the B.S. degree in information engineering from Zhejiang University, Hangzhou, China, in 2015, where he is currently pursuing the Ph.D. degree from the Institute of VLSI Design. His current research interests include ECG signal analysis and ECG processor.

FEITENG LI received the B.S. degree in Internet of Things engineering from the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China, in 2014. He is currently pursuing the Ph.D. degree with the College of Electrical Engineering, Zhejiang University, Hangzhou, China.

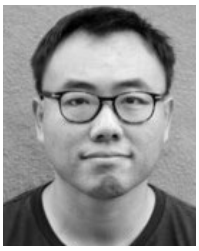
His research interests include physiological signal processing with machine learning and ultra-lower-power neural network accelerator.



JIAQUAN WU received the Bachelor of Engineering degree in electronic and information engineering from Zhejiang University, in 2015, where he is currently pursuing the Ph.D. degree with the Institute of VLSI Design. His research interests include biomedical signal processing and neural network accelerating.

ZHIJIAN CHEN received the B.S. and Ph.D. degrees from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2006 and 2011, respectively.

From 2011 to 2013, he was a Postdoctoral Researcher with the College of Electrical Engineering, Zhejiang University, where he has been a Lecturer with the College of Information Science and Electronic Engineering, since 2013. His research interest is ultra-low-power physiological signal processor design.



YU PU received the B.S. degree from Zhejiang University, Hangzhou, China, in 2004, and the Ph.D. degree in electrical engineering from the Eindhoven University of Technology, The Netherlands, in association with the NXP Research in 2009. From 2009 to 2011, he was a Research Assistant Professor with Sakurai Lab, The University of Tokyo, Japan. From 2011 to 2012, he was a Research Scientist with the Accelerator Team, IBM Research Zurich, Switzerland. From 2012 to 2013, he was a Principal Scientist with NXP Research, where he led R&D in ultra-low-power MCUs. From 2014 to 2019, he was with Qualcomm Research, San Diego, CA, USA and led R&D in always-on Android wearable SoC from concept to mass production. Since 2019, he has been with Alibaba DAMO computing research lab, Sunnyvale, CA, USA. He is currently an Adjunct Full Professor with Shanghai Jiaotong University and Zhejiang University, China. He has authored or coauthored over 30 scientific publications and held more than 15 U.S. patents. Dr. Pu is the Technical Program Committee (TPC) Chair and a member of various IEEE/ACM conferences. He is also an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (TCAS-I and TCAS-II).