Region Aggregation Network: Improving Convolutional Neural Network for ECG Characteristic Detection*

Ming Chen[1], GuiJin Wang[1], PengWei Xie[1], ZhenHua Sang[2], TingTing Lv[2], Ping Zhang[2], HuaZhong Yang[1]

***Abstract*— Detection of ECG characteristic points serves as the first step in automated ECG analysis techniques. We propose a novel end-to-end deep learning scheme called Region Aggregation Network (RAN) for ECG characteristic points detection. A 1D Convolutional Neural Network (CNN) is adopted to automatically process ECG signals. A novel strategy of Region Aggregation is proposed to replace the conventional fully connected layer as regressor. Our work provides robust and accurate detection performance on public ECG database. The evaluation results of our method on QT database show comparable detection accuracy compared with state-of-the-art works.**

I. INTRODUCTION

Automatic annotation algorithm of electrocardiogram (ECG) has received wide attention because of its vital role in clinical diagnosis of various cardiac diseases [1]. Most of the useful clinical information of ECGs can be obtained from various intervals and magnitudes of the ECG characteristic points (the limits and peaks of QRS complex, P wave and T wave).

The precise detection of ECG characteristic points is a challenging task due to the large changes in ECG waveforms. And there is currently no universal rule for determining the range of individual component waveforms.

Traditional delineation methods include low-pass differentiation (LPD) [2], hidden Markov models [3], spline representation [4] and wavelet-based methods [5][6]. Among them, wavelet transform is the most popular method. Inspired by methods in [5], Martnez et al. developed a single-lead ECG delineation system based on wavelet transform in [6]. The evaluation results on public ECG databases show acceptable detection accuracy. P wave detection and delineation is also achieved with phase free stationary wavelet transform in [7]. However, the parameters of these methods often require manual and empirical adjustment, which may significantly effect the performance. Besides, they are often data-sensitive and varying from different sources. Therefore, these methods lack sufficient generalization ability.

In recent years, many machine learning methods have been developed in ECG characteristic points detection. Saini et al. in [8] proposed a K-Nearest Neighbor classification approach for ECG recognition. Improved in [9] his K-NN with gradient feature method was evaluated on multiple ECG databases. However, the K-NN method suffers from the curse of dimensionality when the feature dimension is high, and the trained K-NN classifier model is memory consuming since it needs to store all of the training data. Based on partially collapsed Gibbs sampler, Bayesian method was proposed in [10][11]. By exploiting the strong local dependency of ECG signals, the method showed relatively high detection rate on QT database. Gao et al [12] proposed randomly selected signal pair difference (RSSPD) feature extracted from time domain signals and used a random forest classifier and some post processing to generate final results. The evaluation results of their method on QT database show acceptable performance. However, conventional machine learning methods demand complex feature engineering. Besides, the feature extraction and classification or regression process are performed separately.

In this paper, we propose a novel end-to-end deep learning based ECG characteristic point detection algorithm, named as Region Aggregation Network (RAN). To our knowledge, this is the first time that a deep learning method is presented to detect ECG characteristic points. The RAN consists of a 1D Convolutional Neural Network (CNN) and a novel region aggregation module. The region aggregation module is designed to replace the simple fully connected layers which usually play the role of regressor. The whole network can be trained elegantly by using stochastic gradient descent, which means that the feature extraction and regression process are preformed concurrently in our scheme. The evaluation results of our algorithm achieve comparable performance with state-of-the-art works on QT database [13].

II. FRAMEWORK AND PREPROCESSING

Fig. 1 shows the framework of our scheme. The ECG signal is denoised to remove high frequency noise and baseline wandering, as is described in section A. Then, the denoised signal is segmented into individual beats which will be fed into the CNN. After segmentation, we normalize the segments according to the maximum and minimum amplitude value in individual segment by

$$x_i' = \frac{x_i}{(S_{max} - S_{min})}, i = 1, 2, ..., L \quad (1)$$

Where x_i is the amplitude of i-th sample point in individual segment, L is the length of the segment, S_{max} and S_{min} is the maximum and minimum signal amplitude value. The purpose of normalization is to make the model more robust to variations of ECG amplitude. Then we resample each segment to 325 since CNN can only take fixed size tensor as input.

*This work is partially supported by the National Key Research Project of People's Republic of China (SQ2017YFB140080-04 and 2016YFC0900901) and the Self-determined Project of Tsinghua University(20161080077)

[1]Department of Electronic Engineering, Tsinghua University, Beijing, China

[2]Beijing Tsinghua Changgung Hospital, Tsinghua, Beijing, China

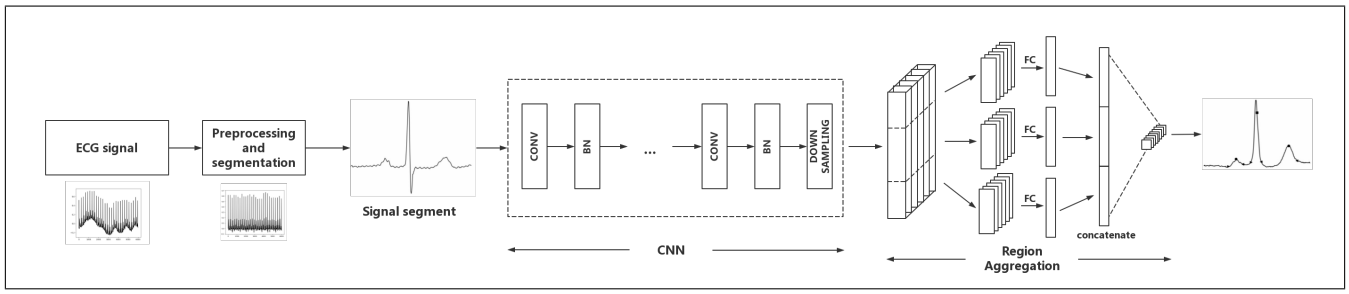

Fig. 1. The proposed Region Aggregation Network.

We use a 1D CNN and a novel region aggregation module to extract features from single-beat ECG segments and generate the characteristic points concurrently. The network can automatically learn the essential features in ECG waveforms and avoid complicated manual feature extraction. We directly use denoised ECG signal as input so that we may utilize more information than using manually extracted features. In the region aggregation module, the feature maps generated by CNN are uniformly cropped into several regions and fed into fully connected layers respectively as branches. Features from the last FC layers of all regions are concatenated and used to infer the positions with an extra regression layer. The details of region aggregation and model architecture are elucidated in section III.

Our scheme is divided into two stages: training and testing. In the training stage, RAN takes ECG signal segment as input and generates 8 positions (the begining, peak and end of P wave and QRS complex, the peak and end of T wave). Then we define a loss function to measure the distance between generated positions and labels. We adjust all parameters in RAN by minimizing the loss function with stochastic gradient descent strategy. In the testing stage, we just feed a segment into RAN and directly obtain 8 positions. In this stage, only one time forward propagation is processed, which means that the testing phase is computationally efficient. The feature extraction and regression process can be performed simultaneously without separation both in training and testing stage.

A. ECG signal preprocessing

Inspired by [14], we apply a preprocessing method combining wavelet-based denosing and median filtering. ECG signals are decomposed into 11 scales using Dual-Tree Complex Wavelet Transform (DTCWT). Information in scale 2 to 9 is retained to reconstruct the signals, and the others are regarded as noise and set to zero. In order to remove the baseline wander effectively, as in [15], a 200ms width median filter removing P wave and QRS complex, then a 600ms width median filter removing T waves are used to fit the baselines. Then we subtract the fitted baseline from the denoised signal.

B. Segmentation and data augmentation

We apply DPI algorithm [16] to obtain all locations of QRS peaks in each record and calculate an average RR interval RR_{ave}. We use a series of window functions $g_i(x)$ to segment the ECG waveform. The $g_i(x)$ is defined by

$$g_i(x) = \begin{cases} 1, & |x - x_i| \le 0.7 \times RR_{ave} \\ 0, & otherwise \end{cases} \tag{2}$$

Where x_i is the label of i-th QRS peak and x is position of arbitrary sample point.

Due to the lack of available training data, data augmentation is necessary. We implement data augmentation by changing the window functions as

$$g_i(x) = \begin{cases} 1, & 0 \le x - x_i \le R_{right} \times RR_{ave} \\ 1, & 0 \le x_i - x \le R_{left} \times RR_{ave} \\ 0, & otherwise \end{cases} \tag{3}$$

We set the range of R_{right} and R_{left} between 0.5 and 0.8 and apply uniformly random sampling. We repeat the segmentation process for 30 times and ensure that R_{right} and R_{left} are different each time. After segmentation, all segments are resampled to 325. This augmentation strategy is equivalent to implementing waveform translation, stretching and compression. In addition, we also artificially add Gaussian noise to the segments to improve the anti-interference ability of our model.

III. Region Aggregation Network

As is shown in Fig. 1, we employ a 1D CNN and region aggregation module to automatically extract features and generate characteristic points. In this section, the details of our network architecture and process of region aggregation are described.

A. Network architecture

Our convolutional network consists of 22 convolutional layers and 3 down sampling layers. Each convolutional layer is followed by a batch normalization (BN) layer and a Rectified Linear Unit (ReLU) activation function. We use convolution layers with a kernel size of 2 and a stride of 2 instead of pooling layers to implement downsampling. The architecture of our CNN is shown in Table. I.The input size is 325×1 and the output size is 21×128. It is not necessary to select 325 as input size, other values may also be used as long as the used value is fixed in both training and testing stage. The input is segmented single-lead ECG signal and the output is a fixed-size feature map. In region aggregation module, we use fully connected layers of size

TABLE I
NETWORK ARCHITECTURE

Layer	Size
Input	325×1
16×5 conv, stride 1	325×16
16×5 conv, stride 1	325×16
16×5 conv, stride 1	325×16
16×5 conv, stride 1	325×16
16×5 conv, stride 1	325×16
Down Sampling 1	163×16
32×5 conv, stride 1	163×32
32×5 conv, stride 1	163×32
32×5 conv, stride 1	163×32
32×5 conv, stride 1	163×32
32×5 conv, stride 1	163×32
Down Sampling	82×32
64×5 conv, stride 1	82×64
64×5 conv, stride 1	82×64
64×5 conv, stride 1	82×64
64×5 conv, stride 1	82×64
64×5 conv, stride 1	82×64
Down Sampling	41×64
128×5 conv, stride 1	41×128
128×5 conv, stride 1	41×128
128×5 conv, stride 1	41×128
128×5 conv, stride 1	41×128
128×5 conv, stride 1	41×128
128×5 conv, stride 1	41×128
128×5 conv, stride 1	41×128
Down Sampling	21×128

256 and dropout with a keep rate of 0.5 in each branch. The output features of all branches are concatenated into a large vector and fed into an extra fully connected layer. The output of the final fully connected layer is a vector of length 8, representing the positions of 8 ECG characteristic points within a cardiac cycle. According to the length of segment before and after resampling, we can calculate the positions in original waveform by

$$p_i' = \frac{p_i \times 325}{l_{raw}}, i = 1,2,...,8 \tag{4}$$

Where p_i' is original position and p_i is predicted position. l_{raw} is the length of segment recorded in segmentation process before resampling.

B. Region aggregation

The region aggregation module is a tree-shaped fully connected structure, containing multiple branches with the same size. Different from the simple fully connected layers, we evently divide the feature maps into N regions at the end of CNN. For each region, we feed it into the FC layers respectively as branches. Then features from the last FC layers of all regions are concatenated and used to infer the characteristic points with an extra regression layer, as is shown in Fig. 1. The region aggregation module utilizes each region to separately predict the positions of all characteristic points and combine the results. Different regions in the feature maps from CNN represent different regions of the input ECG segment. This process is basically equal to a kind of multi-view voting strategy [17], which takes multiple different regions in the same ECG segment as inputs, then averages all results generated from these regions. Compared simple fully connected layer, region aggregation module can more effectively extract the essential information according to mutiple regions in the same signal, whcih means it can achieve more robust performance. In order to restrict the position in the segment, we use the hyperbolic tangent function to limit the range of all elements of the output variable from 0 to 324 by

$$x' = (tanh(x) + 1) \times 162 \tag{5}$$

The whole network can be trained end-to-end by minimizing the regression loss. We set N = 3 to balance the trade-off between performance and efficiency For each region we use FC layers with the same size 256. We do not adopt multi-scale regions because it will lead to imbalanced parameter number in FC layers. Region aggregation can achieve similar performance as multi-view voting, which means RAN is more robust than simple CNN. End-to-end training allows the model to automatically adjust the contributions of different regions in the same signal to the final detection results. We use concatenation instead of averaging in the last layer so that the regressor can effectively utilize information from different FC layers, making the model more powerful.

IV. EXPERIMENT

In this section we describe the experiments to evaluate our method. We show both qualitative and quantitative results on QT database, the detection accuracy is quantified and compared with bayesian methods [10][11], wavelet transform methods proposed in [6][7] and machine learning methods in [12]. For comparison, we implement a baseline which employs the same CNN architecture and simple FC layers as regressor.

The QT database include 105 records chosen from existing ECG database, including the MIT-BIH Arrhythmia Database, ST-T Database and several other ECG databases. At least 30 beats in each record, 3622 beats in all, were manually annotated in the database. The annotation includes the beginning, peak and end of the P wave, the beginning and end of the QRS complex, and the peak and end of the T wave. Since the RAN has to accept a fixed-size input and generate a fixed-size output, we only use records containing annotations of eight points as testing or training set, accounting for 97 out of 105. We divide the 97 records into two parts: training set and test set, in which the training set contains 74 records and the test set contains 23 records.

We perform 10 independent experiments with the test and training sets randomly divided and take the average of 10 test results as the evaluation result. The performance is measured by the average M and standard deviation SD of error values, which indicates the detection accuracy of the proposed method. For each expert label, the error is defined as the distance between detection output and the current expert label.

The comparison between our method and other studies are provided in Table. II. As can be seen from the results, our

TABLE II

DETECTION ACCURACY ON THE QT DATABASE

Method	Metric	P-on	P-peak	P-off	QRS-on	QRS-off	T-peak	T-off
RAN(ours)	M±SD (ms)	**0.4±14.4**	**-0.4±10.1**	**-2.0±12.7**	**-0.7±10.9**	**-4.8±13.1**	**-3.0±10.5**	**-0.3±18.5**
CNN(Baseline)	M±SD (ms)	6.6±17.8	3.9±14.2	2.4±18.6	-0.3±14.0	-6.6±15.2	-4.5±17.2	-6.1±26.4
RSSPD([12])	M±SD (ms)	0.4±22.0	N/A	2.1±12.9	0.2±10.2	0.5±14.4	N/A	1.4±17.2
Beat-to-beat BGS([11])	M±SD (ms)	3.4±14.2	1.1±5.3	2.7±9.8	N/A	N/A	0.8±4.1	-3.9±14.0
Multi-beat PCGS([10])	M±SD (ms)	1.7±10.8	2.7±8.1	2.5±11.2	N/A	N/A	0.7±9.6	2.7±13.5
Phase free SWT([7])	M±SD (ms)	-0.3±12.2	N/A	5.8±9.1	N/A	N/A	N/A	N/A
WT([6])	M±SD(ms)	2.0±14.8	3.6±13.2	3.5±18.0	4.6±7.7	0.8±8.7	0.2±13.9	-1.6±18.1

[1]N/A indicates the statistics is not available.

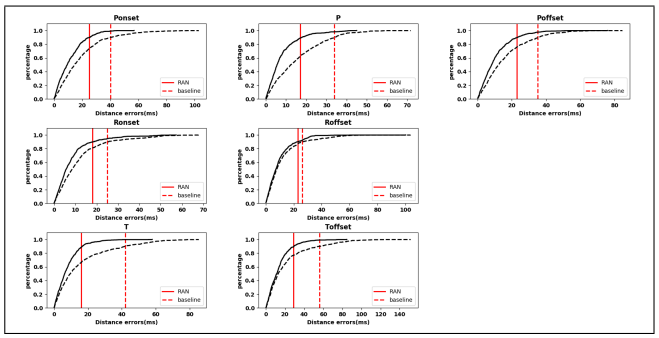

Fig. 2. Accumulate percentage of detection error values and 90% confidence interval (red lines).

method has achieved comparable performance with state-of-the-art works. It proves the deep learning method is effective in detection of ECG characteristic points.

The accumulate percentage of error values of the 10 test rounds is shown in Fig. 2. The red lines in the figure indicates the 90% confidence intervals. The solid curves of RAN indicate that most of the error values of proposed method lies within 20ms. Compared with the baseline, our RAN method achieves a confidence interval performance improvement of over 15 ms in most points' detection. The experiments show that our RAN has achieved a more stable detection effect than simple CNN.

V. CONCLUSIONS

We have proposed a novel end-to-end deep learning scheme called Region Aggregation Network to detect ECG characteristic points. It is the first time deep learning method is adopted in this field. The new architecture consists of a 1D convolutional network and a novel region aggregation module. In training stage, we employ data augmentation to make up for the lack of data in QT database. The proposed scheme can detect ECG characteristic points with high detection accuracy. It proves that deep learning method is effective in analyzing ECG signals.

REFERENCES

[1] A. Gacek and W. Pedrycz, *ECG signal processing, classification and interpretation: a comprehensive framework of computational intelligence*. Springer Science & Business Media, 2011.

[2] P. Laguna, R. Jané, and P. Caminal, "Automatic detection of wave boundaries in multilead ecg signals: Validation with the cse database," *Computers and biomedical research*, vol. 27, no. 1, pp. 45–60, 1994.

[3] R. V. Andreao, B. Dorizzi, and J. Boudy, "Ecg signal analysis through hidden markov models," *IEEE Transactions on Biomedical engineering*, vol. 53, no. 8, pp. 1541–1549, 2006.

[4] F. Guilak and J. McNames, "A bayesian-optimized spline representation of the electrocardiogram," *Physiological measurement*, vol. 34, no. 11, p. 1467, 2013.

[5] C. Li, C. Zheng, and C. Tai, "Detection of ecg characteristic points using wavelet transforms," *IEEE Transactions on biomedical Engineering*, vol. 42, no. 1, pp. 21–28, 1995.

[6] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ecg delineator: evaluation on standard databases," *IEEE Transactions on biomedical engineering*, vol. 51, no. 4, pp. 570–581, 2004.

[7] G. Lenis, N. Pilia, T. Oesterlein, A. Luik, C. Schmitt, and O. Dössel, "P wave detection and delineation in the ecg based on the phase free stationary wavelet transform and using intracardiac atrial electrograms as reference," *Biomedical Engineering/Biomedizinische Technik*, vol. 61, no. 1, pp. 37–56, 2016.

[8] I. Saini, D. Singh, and A. Khosla, "Delineation of ecg wave components using k-nearest neighbor (knn) algorithm: Ecg wave delineation using knn," in *Information Technology: New Generations (ITNG), 2013 Tenth International Conference on*. IEEE, 2013, pp. 712–717.

[9] ——, "K-nearest neighbour-based algorithm for p-and t-waves detection and delineation," *Journal of medical engineering & technology*, vol. 38, no. 3, pp. 115–124, 2014.

[10] C. Lin, C. Mailhes, and J.-Y. Tourneret, "P-and t-wave delineation in ecg signals using a bayesian approach and a partially collapsed gibbs sampler," *IEEE transactions on biomedical engineering*, vol. 57, no. 12, pp. 2840–2849, 2010.

[11] C. Lin, G. Kail, A. Giremus, C. Mailhes, J.-Y. Tourneret, and F. Hlawatsch, "Sequential beat-to-beat p and t wave delineation and waveform estimation in ecg signals: Block gibbs sampler and marginalized particle filter," *Signal Processing*, vol. 104, pp. 174–187, 2014.

[12] P. Gao, J. Zhao, G. Wang, and H. Guo, "Real time ecg characteristic point detection with randomly selected signal pair difference (rsspd) feature and random forest classifier," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016, pp. 732–735.

[13] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg," in *Computers in Cardiology 1997*. IEEE, 1997, pp. 673–676.

[14] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin, and H. Yang, "Patient-specific ecg classification based on recurrent neural networks and clustering technique," in *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on*. IEEE, 2017, pp. 63–67.

[15] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.

[16] A. Ramakrishnan, A. Prathosh, and T. Ananthapadmanabha, "Threshold-independent qrs detection using the dynamic plosion index," *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 554–558, 2014.

[17] H. Guo, G. Wang, X. Chen, C. Zhang, F. Qiao, and H. Yang, "Region ensemble network: Improving convolutional network for hand pose estimation," *arXiv preprint arXiv:1702.02447*, 2017.