2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Dual Attention Convolutional Neural Network Based on Adaptive Parametric ReLU for Denoising ECG Signals with Strong Noise

Zixiao He, Xinwen Liu, Hao He, Huan Wang*

Abstract— **Electrocardiogram (ECG) signal is one of the most important methods for diagnosing cardiovascular diseases but is usually affected by noises. Denoising is therefore necessary before further analysis. Deep learning-related methods have been applied to image processing and other domains with great success but are rarely used for denoising ECG signals. This paper proposes an effective and simple model of encoder-decoder structure for denoising ECG signals (APR-CNN). Specifically, Adaptive Parametric ReLU (APReLU) and Dual Attention Module (DAM) are introduced in the model. Rectified Linear Unit (ReLU) is replaced with the APReLU for better negative information retainment. The DAM is an attention-based module consisting of a channel attention module and spatial attention module, through which the inter-spatial and inter-channel relationship of the input data are exploited. We tested our model on the MIT-BIH dataset, and the results show that the APR-CNN can handle ECG signals with a different signal-to-noise ratio (SNR). The comparative experiment proves our model is better than other deep learning and traditional methods.**

Clinical Relevance— **This paper proposed a method capable of denoising ECG signals with strong noise to alleviate difficulties for further medical analysis.**

I. Introduction

Cardiovascular disease has become the primary disease threatening human life. For a long time, research on cardiovascular disease is one of the main topics in the medical field. The ECG is a comprehensive manifestation of the electrical activity of the heart in the human epidermis and can be measured by sensors attached to the human's chest. Therefore, ECG has become a widely accepted method for analyzing cardiac conditions of human patients. However, ECG signals are usually contaminated by various kinds of noises. Common types of noises are power line interference, electrode contact noise, motion artifacts, muscle contractions, baseline wander. These noises differ in frequency and their power changes under different situations. Practical methods are required for dealing with these noises before further analysis.

A large scale of traditional denoising methods has been developed for a long time. These methods are mainly based on Empirical Mode Decomposition (EMD), Fourier transforms, discrete wavelet transform (DWT). However, experiments show that these traditional techniques can only handle slightly polluted signals or random noise [1]. Additionally, expertise is needed to tweak the algorithm's parameters for better performance, making application for traditional solutions difficult.

In recent years, we can observe the appearance of many deep learning-based algorithms [2, 3]. These methods can solve ineffectiveness with strong noise and inconvenient applications that traditional denoising methods are facing with neural networks. Poungponsri et al. [4] proposed an autoencoder used to denoise ECG signals. The structure of this network is simple, consisting of 6 convolutional layers and 6 deconvolutional layers. The network is tested on an SNR of 7dB, and the average SNR improvement is 9.7dB. Chiang et al. [5] designed a deep learning model that combines wavelet transform. By replacing ReLU with a set of wavelets as the activation functions, this model can increase the SNR by 3.4dB, with the SNR of input equals 17.7dB. To explore different network structures and find the one with better performance, Antczak et al. [6] proposed a deep recurrent neural network (DRNN). They tested different network structures and the effect of pre-training. After finding the best combination, they compared the model with a bandpass filter and discrete wavelet transform. The result proves this model is better than traditional methods. Arsene et al. [1] designed a Convolutional Neural Network (CNN) and a Long-Short-Term Memory (LSTM). They tested the two networks on different SNR values, and the result shows that CNN performs better than LSTM.

The quality of ECG signals collected in practical situations could be bad, making denoising a challenging task [7]. Therefore, methods that are also effective with intense noise are needed. In [8], a two-stage denoising method is proposed. An autoencoder model is used to eliminate noise in the first stage. Then a squeeze-and-excitation-based network is used to restore the signal. The experiment proves the validity of this method with strong noise. By designing deeper neural network models, the denoising performance is increased. Fan et al. [9] designed a combination of autoencoder and ResNet. By replacing the convolutional layers in the autoencoder with residual block, the model reaches 15 layers. The model can increase the SNR by 6.1dB at an 80% noise level.

In the networks mentioned above, ReLU [10] is commonly applied as the activation function. This kind of activation function will lead to information loss in the network layers. Also, these networks do not have special designs for strong noise, leading to their poor performance under small SNR.

This work was supported by Sichuan Science and Technology Program (2020108) and Glasgow College, University of Electronic Science and Technology of China. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board. (Corresponding author: Huan Wang).

Zixiao He, Xinwen Liu, Hao He, and Huan Wang are with the Glasgow College, University of Electronic Science and Technology of China, Chengdu, 611731, China. Huan Wang is also with the Department of Industrial Engineering, Tsinghua University, China (email: wh.huanwang @gmail.com).

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

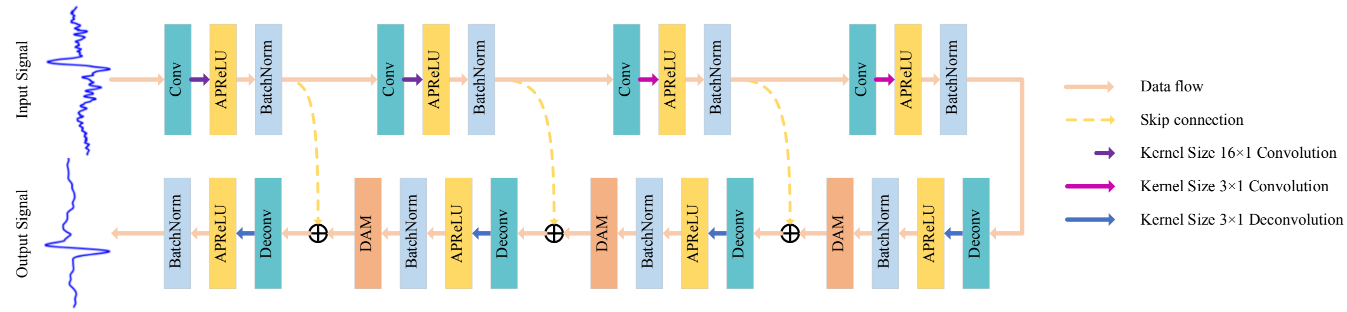

Fig. 1. The structure of the proposed network.

To solve these problems, this paper designed a general deep learning model to denoise the ECG signals (APR-CNN). The backbone of the model is of encoder-decoder structure with a skip connection embedded. APReLU is used to replace ReLU as the activation function for better retainment of negative information. To optimize performance under strong noise, we introduced the attention mechanism by designing the Dual Attention Module (DAM). The network is trained and tested on the MIT-BIH dataset [11]. Real ECG signals and noise with different SNRs are used in the experiment. The results show that our network can improve the SNR by 9.56dB under -4dB noise.

II. METHODOLOGY

A. *The Architecture of the proposed model*

Our network is a deep convolutional network of encoder-decoder structure. APReLU is applied as the activation function. DAM is inserted behind each deconvolutional layer in the decoder. We introduced skip connections between the layers in the encoder and decoder. The overall structure of our model is presented in Fig. 1.

The encoder-decoder structure will refine the features from input data in the encoder part and reconstruct and output the denoised signal in the decoder part. This model has four compression stages and four reconstruction stages, and multiple of each is two. In the encoder part, we set 16×1 as the kernel size for the first two layers. The experiment results in [12] show that selecting a large kernel size for the first few convolutional layers would significantly remove baseline drift noise. But to control the total size of our network and avoid over-fitting, the kernel size of the last two layers is set to three. In the decoder part, we obtained the best kernel size of the deconvolutional layers is 16 by experiments. To further improve reconstruction performance, we added the DAM at the back of each layer. This module can extract the inter-channel and intra-channel attention and help reconstruct the signal. In the entire model, we used APReLU as an activation function, enhancing nonlinear transformation flexibility. In addition, skip connection is added and connects layers in the encoder to corresponding layers in the decoder directly. This extra path allows low-level features to be combined with high-level features, which helps alleviate information loss in the decoder.

B. *Improve activation function with APReLU*

Typically, the activation function in each layer in a convolutional neural network is rectifier linear units (ReLU) [10]. Using such activation functions would damage the negative information contained in the input. Also, the transformations applied to each input will be identical, which limits the feature learning ability.

Adaptive parametric ReLU (APReLU) [13] is developed for these problems. A typical APReLU is shown in Fig. 2. This module starts by separating the positive and negative parts of the input. Then the global average pooled value of the two parts is calculated and generates two 1-D vectors. Then the vectors are concatenated and propagated through a fully connected network, as shown in the graph. This yields a vector α containing the slope for the negative part of the input. Each slope corresponds to a channel and is different from the other. The output of this module is computed as:

$$y = \max(x,0) + \alpha \times \min(x,0) \tag{1}$$

where x and y represent the input and output. Channel-wise multiplication is performed on the α and the negative part. A specific set of weights is produced for each input feature, so the nonlinear transformation is different. This module brought highly flexible transformation to the activation function and strengthened the denoising capability of our model.

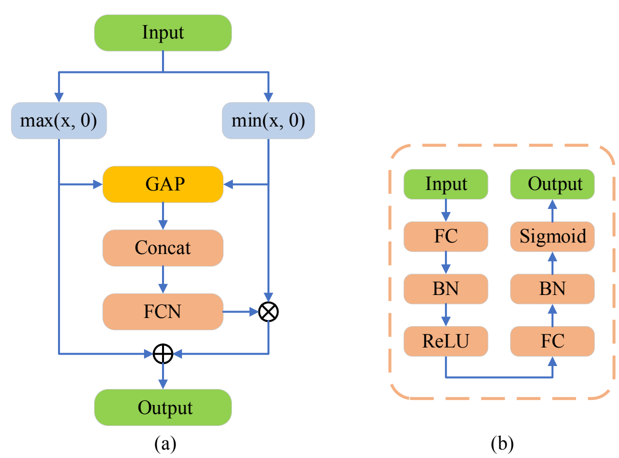

Fig. 2. (a) The structure of APReLU. (b) The inner structure of the FCN.

C. *Dual Attention Module*

Attention mechanism plays an essential role in human perception. Several attempts are made in recent studies trying to introduce this mechanism into deep learning models [14, 15]. Inspired by the Squeeze-and-Excitation module proposed in [16], we designed the Dual Attention Module (DAM) for better ECG signal feature extraction capability. As shown in Fig. 3, the DAM consists of two serially connected submodules, the channel attention module and the spatial attention module. The two submodules will exploit the inter-channel and inter-spatial relationship of features, respectively. The GMP and GAP values are calculated and fed into a fully connected network in the channel attention module. The two outputs are then added together, passed through the sigmoid function, and yield the channel attention. The CMP and CAP value is calculated in the spatial attention module, generating two 2D feature maps. The maps are then concatenated by channel and propagate through a convolutional layer followed by a sigmoid function, which yields spatial attention. The output of the DAM is computed by:

$$\begin{aligned} \mu &= \alpha_c \otimes x \\ y &= \alpha_s \otimes \mu \end{aligned} \tag{2}$$

where x and y represent DAM input and output, μ is an intermediate variable, α_c and α_s represent the channel attention and spatial attention. This module is a black-box model with one input and output and can be inserted into any part of the network. Also, the DAM significantly improved the performance without increasing network parameters because of the simple inner structure.

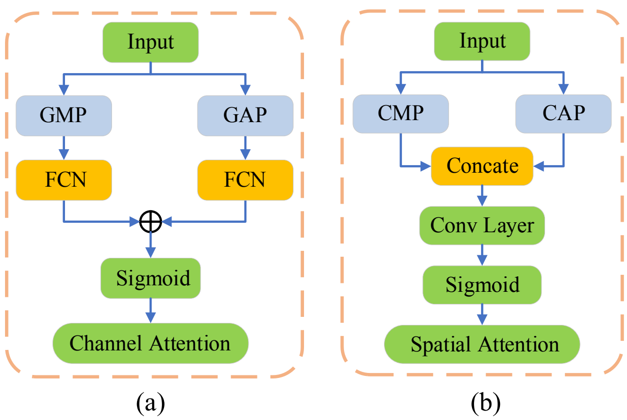

Fig. 3. (a) Architecture of the channel attention module. GMP and GAP mean global mean pooling and global average pooling. FCN is a two-layers fully connected network. (b) Architecture of the spatial attention module. CMP and CAP represent channel mean pooling and channel average pooling.

III. Experiment & Results

A. Dataset and Data Preparation

Real ECG signals are used in the experiment, which came from MIT-BIH [11] Arrhythmia Database. This dataset contains 48 half-hours of two-channel ambulatory ECG recordings from 47 subjects, with a 360Hz sampling frequency. ECG signals from this dataset are the network target output. We manually added noise into it. Noise is real ECG noise from MIT-BIH Stress Test Database. Three kinds of noise are recorded in this dataset, containing baseline wander, muscle artifact, and electro motion artifact.

The two datasets are divided into segments, each with 250 data points, approximately one ECG cycle. Signal and noise are randomly selected from the two datasets. The signal strength of the noise is changed to satisfy the specified SNR. Then the signal and noise are added together as the input of the model. And original ECG signal is used as the target output. The generated data, all of which is normalized, contains 10000 segments and is partitioned according to the rate of 8:1:1 to form the training, testing, and validating set.

B. Experimental Setup

The networks in the experiment were developed based on the Pytorch framework. The model is trained for 80 epochs, with 256 samples per batch. The training algorithm used was stochastic gradient descent with adaptive momentum (Adam). The learning rate was set to 1e-3 initially and decreased to 1e-4 after 40 epochs of training. During the procedure of testing, batch size was set to 200, and a total of 2000 samples were used for testing.

We selected Mean Square Error (MSE) and Signal Noise Ratio (SNR) as the evaluation index. They are computed by:

$$MSE = \sum_{i=1}^{N} (s_i - x_i)^2 \Big/ N \tag{3}$$

$$SNR_{dB} = 10 \log_{10} \frac{\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} (s_i - x_i)^2} \tag{4}$$

where s represents the output signal from the network and x represents the target signal, which is the original noise-free ECG signal. MSE smaller is better, SNR greater is better.

C. Effectiveness of APReLU and DAM

To verify that the two applied modules are effective in improving the denoising performance, we selected three models with similar structures for comparison. The backbone network of the four models is a deep convolutional neural network (DCNN) of encoder-decoder structure with skip connection. One model with APReLU introduced, one with DAM introduced, and one with no modules introduced. To compare the performance of the four models under different SNR, we trained and tested the four models on datasets with SNR equals -4dB, -2dB, 0dB, and 4dB. The SNR of data is kept the same in training and testing. The result is shown in TABLE I.

TABLE I. Testing result of the four models.

Model	SNR_{dB}				MSE			
	-4	**-2**	**0**	**4**	**-4**	**-2**	**0**	**4**
DCNN	1.844	2.532	3.441	5.822	0.127	0.107	0.093	0.055
DCNN+APReLU	4.755	5.501	7.132	9.548	0.127	0.112	0.096	0.056
DCNN+DAM	4.380	5.227	6.745	9.211	0.065	0.060	0.054	0.031
APR-CNN	**5.564**	**5.966**	**7.449**	**9.841**	**0.061**	**0.057**	**0.035**	**0.027**

From TABLE I, a considerable advancement can be seen from the two networks with APReLU, or DAM introduced. The average improvement of the two networks is about 3.15dB. The APReLU can better retain negative information, while DAM can improve feature learning ability by extracting attention. The test result of DCNN with APReLU is better than with DAM, proving the importance of negative information. By combining the two modules, our network outperforms other networks. The average improvement is about 3.79dB compared with DCNN. Especially under -4dB, the improvement is 3.72dB, showing the effectiveness of the designed structures. However, the improvement with weak noise is not so distinct. This is because the network has restored most of the information in the signal, and further improvement could lead to overfitting.

B. Comparative experiment

For comparison purposes, a deep convolutional denoising autoencoder described in [5] and traditional denoising method - discrete wavelet transform (DWT) thresholding are tested along with our model. The convolutional autoencoder has 12 layers, six layers of encoder, and six layers of decoder —each layer with kernel size equal to 16 and ReLU as activation. As for DWT, "db8" is selected as the wavelet base, and soft thresholding is applied.

TABLE II. Comparative experiment results.

Model	SNR of the signal				MSE			
	-4	**-2**	**0**	**4**	**-4**	**-2**	**0**	**4**
Autoencoder	2.322	2.845	3.341	4.965	0.128	0.115	0.100	0.073
DWT	-3.850	-1.505	0.426	4.227	0.464	0.302	0.182	0.074
APR-CNN	**5.564**	**5.966**	**7.449**	**9.841**	**0.061**	**0.057**	**0.035**	**0.027**

From TABLE II, one can observe that our network is better under all values of SNR. The autoencoder only has a little denoising effect and is less effective with weak noise. However, DWT thresholding is not working with the data used in this experiment. The two evaluation indexes are nearly unchanged after denoising. Fig. 4 presents a visual example of the denoising result. Noise is almost reduced after passing through the proposed model, except for slight imperfection at extreme points. Signals are smoothed after DWT thresholding, but SNR and mean squared error are not reduced. This is mainly because of the existence of drifting noise. On the contrary, the proposed network learned the characteristics of ECG signals and can handle this kind of noise with ultra-low frequency.

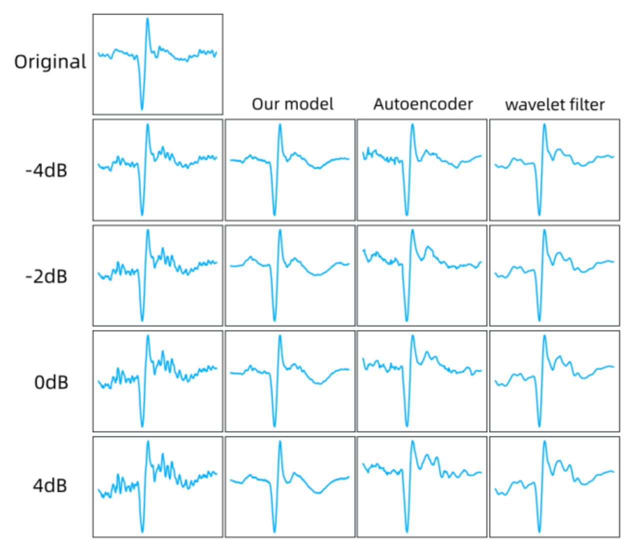

Fig. 4. Visualization of the denoised signal.

IV. Conclusion

In this paper, a deep convolutional network for denoising ECG signals is proposed. We applied APReLU as the activation function, through which different sets of transformation are assigned to input and thus improved flexibility and performance of the network. DAM is designed in this paper to improve the denoising capability further. The DAM draws on attention mechanisms, applying feature refinement with channel attention and spatial attention modules. This module considerably improved the denoising capability while keeping the overhead small. The network architecture comparison experiment shows the effectiveness of the two introduced modules. And the proposed network yields 8dB improvement under -4dB noise, which is way more superior to other networks and traditional methods.

References

[1] C. T. Arsene, R. Hankins, and H. Yin, "Deep learning models for denoising ECG signals." pp. 1-5.

[2] L. Qin, Y. Xie, X. Liu, X. Yuan, and H. Wang, "An End-to-End 12-Leading Electrocardiogram Diagnosis System Based on Deformable Convolutional Neural Network With Good Antinoise Ability," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-13, 2021.

[3] Y. Gao, H. Wang, and Z. Liu, "An end-to-end atrial fibrillation detection by a novel residual-based temporal attention convolutional neural network with exponential nonlinearity loss," *Knowledge-Based Systems,* vol. 212, pp. 106589, 2021.

[4] S. Poungponsri, and X.-H. Yu, "Electrocardiogram (ECG) signal modeling and noise reduction using wavelet neural networks." pp. 394-398.

[5] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y. Chien, "Noise reduction in ECG signals using fully convolutional denoising autoencoders," *IEEE Access,* vol. 7, pp. 60806-60813, 2019.

[6] K. Antczak, "Deep recurrent neural networks for ECG signal denoising," *arXiv preprint arXiv:1807.11551*, 2018.

[7] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," *Knowledge-Based Systems*, pp. 107187, 2021.

[8] L. Qiu, W. Cai, J. Yu, J. Zhong, Y. Wang, W. Li, Y. Chen, and L. Wang, "A two-stage ECG signal denoising method based on deep convolutional network," *bioRxiv*, 2020.

[9] G. Fan, J. Li, and H. Hao, "Vibration signal denoising for structural health monitoring by residual convolutional neural networks," *Measurement,* vol. 157, pp. 107651, 2020.

[10] V. Nair, and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines."

[11] A. L. Goldberger, L. A. 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, 2000.

[12] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine,* vol. 25, no. 1, pp. 65-69, 2019.

[13] M. Zhao, S. Zhong, X. Fu, B. Tang, S. Dong, and M. Pecht, "Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis," *IEEE Transactions on Industrial Electronics,* vol. 68, no. 3, pp. 2587-2597, 2020.

[14] R. A. Rensink, "The dynamic representation of scenes," *Visual cognition,* vol. 7, no. 1-3, pp. 17-42, 2000.

[15] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 20, no. 11, pp. 1254-1259, 1998.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks." pp. 7132-7141.