Noise Reduction in ECG Signals Using Fully Convolutional Denoising Autoencoders

HSIN-TIEN CHIANG[1], YI-YEN HSIEH[1], SZU-WEI FU[2], KUO-HSUAN HUNG[2], YU TSAO[2], AND SHAO-YI CHIEN[1], (Senior Member, IEEE)

[1]Graduate Institute of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan
[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei 11529, Taiwan

Corresponding author: Yu Tsao (yu.tsao@citi.sinica.edu.tw)

ABSTRACT The electrocardiogram (ECG) is an efficient and noninvasive indicator for arrhythmia detection and prevention. In real-world scenarios, ECG signals are prone to be contaminated with various noises, which may lead to wrong interpretation. Therefore, significant attention has been paid on denoising of ECG for accurate diagnosis and analysis. A denoising autoencoder (DAE) can be applied to reconstruct the clean data from its noisy version. In this paper, a DAE using the fully convolutional network (FCN) is proposed for ECG signal denoising. Meanwhile, the proposed FCN-based DAE can perform compression with regard to the DAE architecture. The proposed approach is applied to ECG signals from the MIT-BIH Arrhythmia database and the added noise signals are obtained from the MIT-BIH Noise Stress Test database. The denoising performance is evaluated using the root-mean-square error (RMSE), percentage-root-mean-square difference (PRD), and improvement in signal-to-noise ratio (SNR_{imp}). The results of the experiments conducted on noisy ECG signals of different levels of input SNR show that the FCN acquires better performance as compared to the deep fully connected neural network- and convolutional neural network-based denoising models. Moreover, the proposed FCN-based DAE reduces the size of the input ECG signals, where the compressed data is 32 times smaller than the original. The results of the study demonstrate the superiority of FCN in denoising, with lower RMSE and PRD, as well as higher SNR_{imp}. According to the results, we believe that the proposed FCN-based DAE has a good application prospect in clinical practice.

INDEX TERMS Electrocardiography, signal denoising, artificial neural networks, denoising autoencoders, fully convolutional network.

I. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of death in the world according to the World Health Organization (WHO) [1]. The American Heart Association (AHA) recently reported that CVDs accounted for approximately one out of every three deaths in the United States (US) in 2017 [2]. Among all types of CVDs, arrhythmia is most related to the risk of sudden death [3]. An arrhythmia is an irregular rate or rhythm of the heartbeat. It happens when the natural rhythm coordinates incorrectly with the electrical impulses in patients' hearts. The ECG is an efficient, noninvasive, and low cost indicator that is widely used for effective analysis and diagnosis of arrhythmia. However, ECG signals

are prone to be contaminated by different kinds of noise, such as baseline wander (BW), muscle artifact (MA), and electrode motion (EM) [4]. Baseline wander is a low frequency artifact in the ECG that arises from breathing, electrically charged electrodes, or subject movement [5]. Muscle artifacts are generated because of skeletal muscle activity [6], and electrode motion is caused by changes in electrode-skin impedance and changes in skin potential [7]. All these noises may cause deformations on ECG waveforms and mask tiny features that are important for diagnosis. Accordingly, the removal of noises from ECG signals becomes necessary.

In order to prevent noisy inference, several approaches have been reported to denoise ECG signals based on adaptive filtering [8]–[10], wavelet methods [11]–[13], and empirical mode decomposition (EMD) [14]–[16]. In [9], various adaptive filters based on the error nonlinear signed regres-

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

sor LMS (ENSRLMS) algorithm, the error nonlinear sign LMS (ENSLMS) algorithm, and the error nonlinear sign-sign LMS (ENSSLMS) algorithm, have been proposed. The simulation results show that the performance of sign-based algorithms is better than the LMS counterpart. However, adaptive filters often require noise reference signals as an input that are difficult to obtain with the ECG signal acquisition system. Wavelet denoising techniques are popular and widely implemented for noise cancellation by decomposing a signal in the time-frequency domain. The techniques deal with wavelet coefficients utilizing different thresholding methods mainly including hard and soft thresholding. Although being confirmed to yield satisfactory denoising performance in general, a potential limitation of the hard thresholding is that it may lead to oscillations of the reconstructed ECG signal, known as the Pseudo-Gibbs phenomenon [13]. The soft thresholding, on the other hand, can generate ECG signals that are relatively smoother than those generated by hard thresholding, as well as better continuity. However, soft thresholding distorts the amplitudes of the reconstructed waveform, and particularly the amplitudes of the R waves in QRS complexes that are important for diagnosis [12]. In EMD-based denoising methods, the noisy signals are decomposed into some intrinsic mode functions (IMFs), and then the IMFs containing the most noise are removed. Finally, the signal is reconstructed with the remaining IMFs. As the high frequency noise is embedded in the first few IMFs, the EMD method may not perfectly distinguish between high frequency noise and the QRS complexes. In [15], the portions of the first few IMFs corresponding to the QRS complexes are preserved using proper windowing. In [16], the authors proposed a hybrid approach based on EMD and wavelet methods to obtain a further improvement on denoising. Meanwhile, it is reported that the Hilbert transform used in EMD could not separate similar frequency signals perfectly. Hence, P-waves and T-waves may be filtered out from the signals, leading to misdiagnosis [4].

The above analyses show that there are still rooms to further improve the existing ECG denoising methods. More recently, denoising algorithms based on denoising autoencoder (DAE) have been recently shown to have better performance than conventional denoising algorithms, such as those in [17], [18] and [4]. While these studies put more efforts on denoising signals, we also take into account of ECG compression which helps to reduce the cost and increase the efficiency of signal processing. Our goal is to show the superiority of clinical practice with improved denoising performance and better compression performance. We propose a novel denoising algorithm for ECG signals utilizing FCN. To the best of our knowledge, our work is the first to bring the advantage of FCN to process ECG signals. Meanwhile, the proposed FCN model can compress ECG signals due to the DAE architecture.

The remainder of this paper is organized as follows. We introduce the basic concepts of the DAE and the FCN model in Section II. Then we describe the details of the proposed FCN-based DAE method in Section III. The experimental results obtained from the MIT-BIH Arrhythmia database are presented in Section IV. The discussion of experimental results is provided in Section V. Finally, we conclude this work in Section VI.

II. METHODOLOGY

In [19], it has been shown that DAE is powerful in learning low-dimensional representations and can be used to recover noise-corrupted input. Here, we propose a FCN-based DAE to remove noises from noise-corrupted ECG signals.

A. DENOISING AUTOENCODER

An autoencoder (AE) is a machine learning model that aims to reproduce input data as close as possible. An AE generally comprises two parts: encoder and decoder. The encoder maps an input x to a hidden representation z via a nonlinear transformation. Then, the decoder maps the hidden representation z back to reconstructed data $\hat{x}$ via another nonlinear transformation:

$$z = f(Wx + b) \tag{1}$$
$$\hat{x} = g(\hat{W}z + \hat{b}) \tag{2}$$

where W and b represent the weight and bias matrices of the encoder, respectively, while $\hat{W}$ and $\hat{b}$ represent the weight and bias matrices of the decoder, respectively. Meanwhile, f and g denote non-linear activation functions. There are many alternatives for f and g, such as a sigmoid function, hyperbolic tangent, and rectified linear function. The parameters are determined by optimizing the following objective function as:

$$L(\theta) = \sum_i ||x - \hat{x}||_2^2 \tag{3}$$

where $\theta = \left\{ W, b, \hat{W}, \hat{b} \right\}$ is the parameter set.

DAE, proposed by Vincent *et al.* [19], is a stochastic extension to classic AE. DAE tries to reconstruct a clean input from its corrupted version. As shown in Fig. 1, the initial input x is corrupted to $\tilde{x}$ by a stochastic mapping $\tilde{x} \sim q(\tilde{x}|x)$. Subsequently, DAE uses the corrupted $\tilde{x}$ as input data, and then maps to the corresponding hidden representation z and ultimately to its reconstruction $\hat{x}$.

Due to their powerful nonlinear mapping capabilities, the AE and DAE models have been popularly used for data compression [20], [21] and noise reduction on speech signals [22], [23] and medical images [24].

B. FULLY CONVOLUTIONAL NETWORK

FCN is a special type of CNN. CNN normally consists of convolutional layers, activation functions, max-pooling layers, and a fully connected layer. Convolutional layers consist of a set of filters that can extract feature maps to describe the characteristics of input data. Different feature maps in a layer use different parameters of filters with a feature map sharing the same parameters. Compared to the fully connected layer
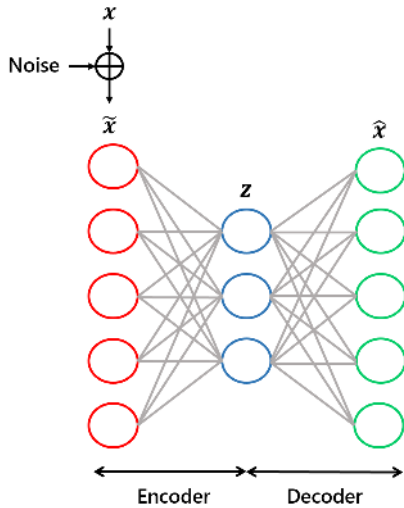
FIGURE 1. Schematic diagram of DAE.

FIGURE 2. Architecture of the proposed FCN-based DAE.

where every neuron is connected to all outputs of the previous layer, convolutional layers highly reduce the number of parameters by sharing parameters with filters. Pooling is used to achieve translation and rotation invariance, and a typical pooling operation is max-pooling. Max-pooling layers perform a down-sampling process by taking the maximum value of a certain scope from the mapping space, leading to a reduced dimension. The fully connected layer aims to perform regression or classification. The difference between FCN and CNN is that the fully connected layer in CNN is removed in the FCN [25]. By discarding the fully connected layers, the number of parameters is reduced, which enables simpler hardware implementation. In addition, FCN enables each output sample to preserve the locally-spatial information of neighboring input regions, whereas fully connected layers do not properly maintain this kind of information from the previous layers [26], [27]. In order to obtain exact signal alignment of input and output, our FCN model contains no pooling layer, as pooling layers may cause the network to lose information about the detailed structure and textures [28].

III. PROPOSED FCN-BASED DAE
This study proposes an approach based on DAE for signal denoising, as well as for compressing the size of ECG waveforms. The proposed FCN-based DAE consists of an encoder and a decoder with 13 layers, and is shown in Fig. 2. In the encoder, the size of ECG signals is reduced, and the signals are encoded into low dimensional features. The decoder tries to reconstruct an output depending on the low dimensional features. We employed exponential linear units (ELU) [29] as activation functions for hidden layers, and there is no activation function for the output layer in the FCN model. In addition, each hidden layer is equipped with batch normalization [30].

The encoder contains a series of layers, where each individual layer is composed of a convolutional layer, a batch normalization layer, and an activation layer. In the encoder
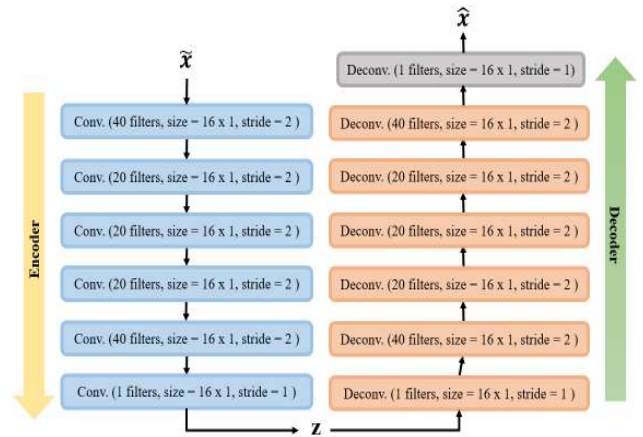
of the model, the original signals with size of 1024×1 are taken as input, and a convolutional process with 40 filters of size 16×1 and stride of 2 is applied on the first layer. The next three convolutional layers all have 20 filters of size 16×1 with a stride of 2. Then, the next layer consists of 40 filters of size 16×1 with a stride of 2. The last layer has 1 filter of size 16×1 with a stride of 1. The down-sampling process is achieved using a stride of 2. Through the encoding process, a 32×1 dimensional feature map is obtained. This feature map also represents the compressed data and is $32\times$ smaller than the original size. The decoder part is inversely symmetric to the encoder. Here, the deconvolutional layers proceed to up-sample the feature maps, and to recover structural details. Contrary to convolutional layers which connect multiple input activations to a single activation, deconvolutional layers project a single input activation into multiple outputs [31]. As for the output layer, a deconvolutional layer with 1 filter of size 16×1 and stride of 1 produces the output signal.

IV. EXPERIMENTS
A. PERFORMANCE EVALUATION CRITERIA
In this study, root mean square error (RMSE), percentage root mean square difference (PRD), and improvement in SNR (SNR_{imp}) are used as quantitative performance estimators.

The RMSE is used for determining the variance between the output predicted by the model and the actual output. A smaller value of RMSE corresponds to a smaller difference and better performance, and is defined as:

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{n=1}^{N}(x_i - \hat{x}_i)^2} \quad (4)$$

The PRD indicates the recovery quality of the compressed signal by measuring the error between the original signal and the resultant signal after reconstruction. A lower PRD represents a better quality of reconstructed signal. PRD is

obtained by the following expression:

$$PRD = \sqrt{\frac{\sum_{n=1}^{N}(x_i - \hat{x}_i)^2}{\sum_{n=1}^{N} x_i^2}} \times 100 \qquad (5)$$

The SNR_{imp} indicates the difference between the SNR after noise reduction and the original input signal SNR. The greater the SNR_{imp} is, the better denoising performance is achieved. SNR_{imp} is described as the following expression:

$$SNR_{imp} = SNR_{out} - SNR_{in} \qquad (6)$$

where SNR_{in} and SNR_{out} are formulated as follows:

$$SNR_{in} = 10 \times \log_{10}\left(\frac{\sum_{n=1}^{N} x_i^2}{\sum_{n=1}^{N}(\tilde{x}_i - x_i)^2}\right) \qquad (7)$$

$$SNR_{out} = 10 \times \log_{10}\left(\frac{\sum_{n=1}^{N} x_i^2}{\sum_{n=1}^{N}(\hat{x}_i - x_i)^2}\right) \qquad (8)$$

In the above equations, x_i is the value of sampling point i in the original ECG signal, $\tilde{x}_i$ is the value of sampling point i in the noisy ECG signal, $\hat{x}_i$ is the value of sampling point i in the denoised waveform and N is the length of the ECG signal.

B. EXPERIMENTAL DATA
We employed the MIT-BIH Arrhythmia database [32] to prepare the clean ECG signals. The database consists of 48 ECG records with lengths of 30 minutes. All data utilized here are sampled at 360Hz and quantized with 11-bit resolution. As each ECG signal comprises information of two electrodes in the database, we focus on the data of lead II, owing to its extensive adoption for ambulatory or wireless body sensor network- (WBSN-) based ECG applications [33].

The proposed approach was applied to all patients. For each record, we excerpted 200 fragments, each with a length of 1024 samples, and because long duration signals were used, there was no need for QRS detection [34]. Real noises including BW, MA, and EM were collected from the MIT-BIH Noise Stress Test Database (NSTDB). The noise samples were also divided into three parts, one for the training set, one for the validation set, and one for the testing set. Therefore, the noise components in the three sets were similar, but not completely the same. We considered this setup similar to the real-world scenarios, where the noise signals may be informed when building the denoising systems.

Because each noise source (of BW, MA, and EM) was a sequence of samples, we cut a segment of samples from the entire noise sources to obtain the noise samples, which were then used to generate noisy ECG signals. To increase the randomness, the starting point to segment was randomly indicated each time. Next, we prepared a combined noise by adding the noise samples of BW, MA, and EM collectively with equal weights. Finally, the noisy signals were generated by artificially adding the combined noise signals to the clean ECG signals.

The dataset of ECG was split into 80% training and 10% validation, and the remaining 10% were used to evaluate the
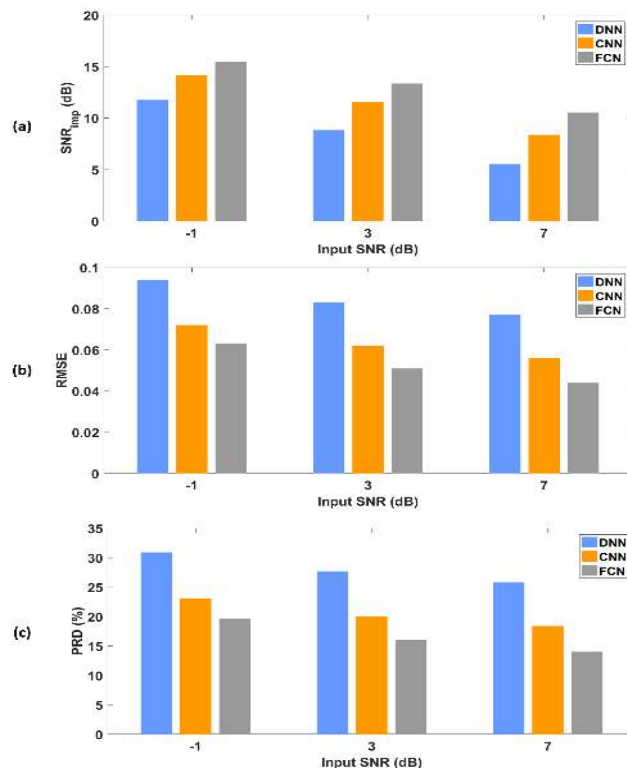

FIGURE 3. Comparison of the denoising performance of all evaluated methods at different input SNR levels (a) SNR_{imp} of DNN, CNN and FCN with varying input SNR levels (b) RMSE of DNN, CNN and FCN with varying input SNR levels (c) PRD of DNN, CNN and FCN with varying input SNR levels.

denoising models. In total, the dataset contained 7680 fragments for training, 960 for validation, and 960 for testing. All training data were corrupted with input SNR of -2.5, 0, 2.5, 5, and 7.5 dB. Both training and validation sets shared the same SNR levels, and the testing set consisted of 3 different levels of input SNR of -1, 3 and 7 dB. All signals were normalized as a preliminary operation, so that the amplitudes of the sampling points laid between 0 and 1 [34].

C. EXPERIMENTAL RESULTS
The denoising performance is compared to deep fully connected neural network (termed DNN for simplicity in the following discussion) based and CNN based DAE. For a fair comparison, the input and output ECG signals with size of 1024×1 are taken in the three neural networks. The DNN also has 13 layers, and the numbers of nodes are 512, 256, 128, 64, 32, 32, 32, 64, 128, 256, 512, 1024, and 1024. The same as FCN, the DNN compresses the input signal $32\times$ to the original size in the middlemost layer. We carefully designed this setup so that the features are down-sampling two times in each layer during the encoding processing and then up-sampling two times in each layer during the decoding process. In this way, the quantity of information processed by the DNN and the FCN models are comparable. We applied the dropout technique with a rate of 0.5 for all layers, which gave the best performance in the validation set. CNN has the
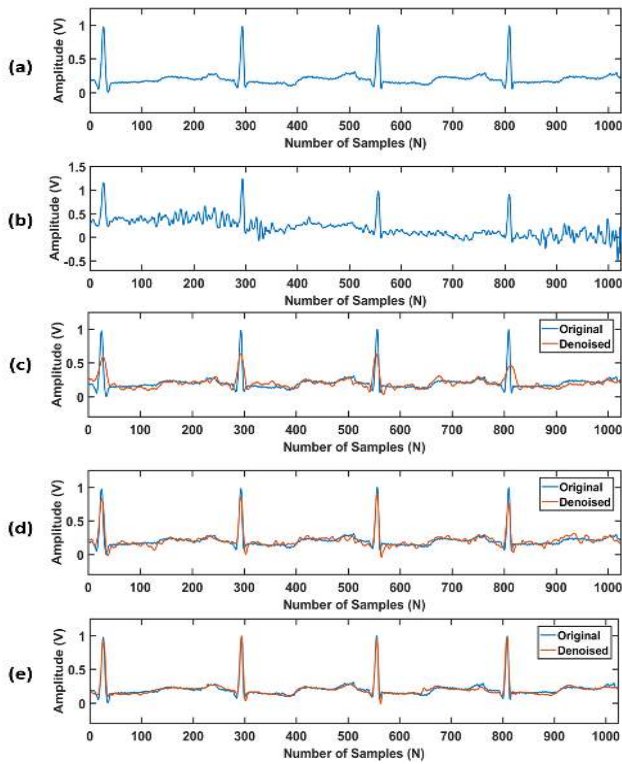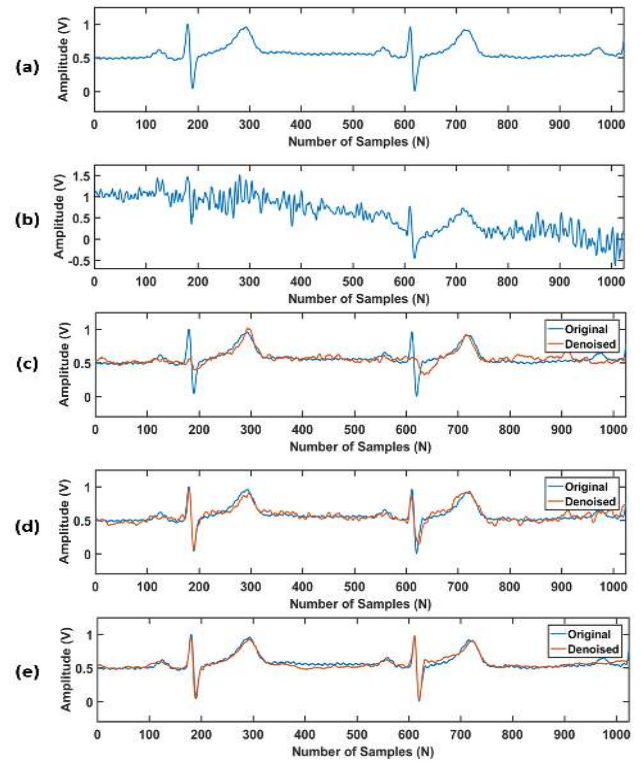
FIGURE 4. Comparison of denoised results of all evaluated methods on record 100 (a) original ECG signal (b) noisy ECG signal with an input SNR of 3dB (c) DNN (d) CNN (e) FCN.

FIGURE 5. Comparison of denoised results of all evaluated methods on record 117 (a) original ECG signal (b) noisy ECG signal with an input SNR of 3dB (c) DNN (d) CNN (e) FCN.

same structure as FCN, except that the last two convolutional layers in FCN are replaced with two fully connected layers. Additionally, dropout was only used in fully connected layers with a rate of 0.5. All models employed batch normalization, followed by ELU activation functions after every hidden layer except for the output layer. We trained all models using Adam [35] to minimize the mean square error between the clean and denoised waveforms.

Fig. 3(a) illustrates the average SNR_{imp} scores over the testing data at specific input SNRs. As it can be seen, all models provide results with a similar trend, i.e., the SNR_{imp} increases when the input SNR is low, while the SNR_{imp} gives a lower value when the input SNR increases. For example, with input SNR of -1 dB, DNN, CNN, and FCN, respectively, have a SNR_{imp} of 11.75, 14.14, and 15.49 dB on average. On the other hand, as the input SNR increases up to 7 dB, the values of SNR_{imp} of DNN, CNN, and FCN become lower, with values of 5.56, 8.37, and 10.54 dB, respectively. In addition, it is observed that FCN has the SNR_{imp} higher than DNN and CNN for all input SNRs. The better the SNR_{imp} the higher its resemblance to a clean signal. Accordingly, FCN performs better denoising of the ECG signal as compared to the two other approaches, as the higher SNR_{imp} denotes more resemblance to clean signals.

Figs. 3(b) and (c), respectively, show the average RMSE and PRD scores over the testing data at specific input SNRs. The lower the RMSE and PRD, the closer the denoised signal is to the original signal, indicating less distortion.

We can observe that FCN outperforms CNN, and CNN outperforms DNN. More specifically, FCN has the best denoising performance in all noise levels. For DNN and CNN, both RMSEs and PRDs are higher than that of FCN: DNN achieved RMSE = 0.094 with PRD = 30.89%, and CNN achieved RMSE = 0.072 with PRD = 23.09% with input SNR = -1 dB. On the other hand, FCN yields RMSE = 0.063 with PRD = 19.68%. For a higher input SNR, the proposed model also has smaller RMSE and PRD values as compared to DNN and CNN. For example, with input SNR = 7 dB, DNN and CNN yield RMSE = 0.077 and 0.056, respectively, which are 0.033 and 0.012 higher than FCN with RMSE = 0.044. Meanwhile, DNN and CNN obtain PRD = 25.79% and 18.4%, respectively, which are 11.74% and 4.35% higher than FCN with PRD = 14.05%. Overall, the resultant SNR_{imp} value for the proposed DAE with FCN is significantly higher than that with DNN and CNN. In addition, the RMSE and PRD values are lower for the proposed work as desired. These results demonstrate that FCN can achieve promising performance in reconstructing a denoised output signal from an original ECG in all noise levels as compared to DNN and CNN.
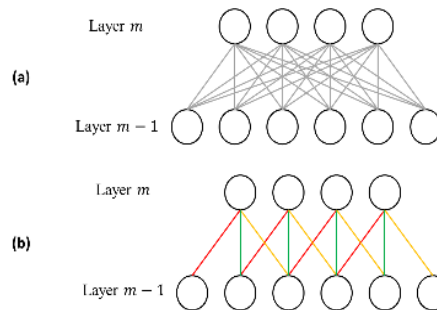
As most ECG studies often refer to records 100, 117, and 119 [34] for visual assessment, we included the waveforms and results of these records. Figs. 4 to 6 plot the original and denoised signals after reconstruction. For each figure, (a) denotes the original ECG signal, and (b) shows the ECG signal corrupted by the combined noise for an input SNR
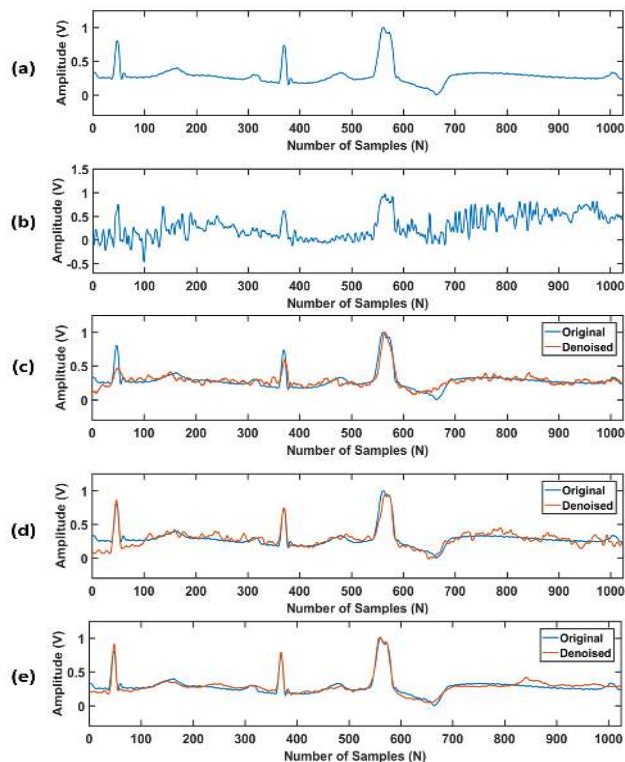
FIGURE 6. Comparison of denoised results of all evaluated methods on record 119 (a) original ECG signal (b) noisy ECG signal with an input SNR of 3dB (c) DNN (d) CNN (e) FCN.

FIGURE 7. Fully connected layer versus convolutional layer. (a) In a fully connected layer, each neuron in layer *m* is connected to all neurons of layer *m* − 1 (b) In a convolutional layer, each neuron in layer *m* only connects to a small region called a receptive filed with the previous layer. The receptive field has width of 3 in this example. Furthermore, the weights are shared for these connections in a convolutional layer.

of 3dB. The results obtained by using DNN, CNN and FCN are shown in (c), (d), and (e), respectively. We can see that the noise influences more on P and T waves and the morphology of the ECG signal is not affected too much in Fig. 4(b), whereas in Figs. 5(b) and 6(b), the noises corrupt the morphology of the ECG signals severely, such as in the disappearance of the P and T waves and QRS complexes' distortions. In each figures, we can clearly observe that DNN appears to show a severe loss of amplitudes of R peaks. The waveforms generated by DNN have smaller R peaks that fail to reconstruct accurate ECG signals. This phenomenon can also be observed in CNN, but is not as serious as in DNN. In contrast, FCN can generally maintain the shape of QRS complexes better than the compared methods. With high-fidelity QRS complexes, an FCN preserves more clinically relevant information. Meanwhile, the denoised signals reconstructed by FCN are smoother than the original ones. This indicates that FCN extracts the original signal from the noisy input signal by learning to capture the main morphological features of ECG signals, in contrast to two other approaches which tend to follow the noisy signal rather than the morphological information of the signal. Overall, both quantitative and visual comparisons demonstrate that FCN gains an advantage over compared methods in terms of noise reduction and clinical detail preservation.

Please note that in previous studies [4], [17], it has been shown that the DNN-based DAE can perform better than several conventional ECG denoising approaches. In this

study, we focus our attention on comparing the proposed FCN-based DAE with the other two deep learning based models (DNN and CNN). Please also note that throughout the experiments in this study, we designed the three deep learning models (DNN, CNN, and FCN) to yield the same compression ratio (=32) when performing signal denoising for a fair comparison.

V. DISCUSSION

The experimental results have found that FCN outperforms DNN and CNN, especially with DNN having the worst performance. More specifically, the QRS complexes are not well reconstructed by DNN, leading to loss of clinical information. In this section, we discuss this problem in detail.

We use time-series data as input for all models. However, under the characteristics of raw waveform, a sampling point alone in the time domain does not carry much information. Instead, it must combine with its neighbors to gain further information. Fu *et al.* pointed out that for speech signal processing, this interdependency may make DNN improper to model waveforms because fully connected layers average out the relation between neighboring samples [26]. In other words, when generating waveforms, fully connected layers result in the loss of spatial information. This also gives explanation of why CNN has relatively minor distortions on QRS complexes when compared to DNN, because the fully connected layers in CNN are less than that of DNN.

In DNN and CNN, the output layer and last hidden layer are fully connected, where each neuron has complete connections to all the neurons in the previous layer. As shown in Fig. 7(a), in output layer *m*, each neuron has full connection with all the units in the last hidden layer *m* − 1. Because the weight vectors in fully connected layers have high correlation with each other, spatial information in DNN is lost. In contrast, the convolutional layers in FCN have the property of local connectivity. Each neuron only depends on a small region of the previous layer, called the receptive field. The input features share the same weights within the receptive field, leading to translation invariance. In Fig. 7(b), neurons in the output layer *m* have receptive fields of width 3, and thus
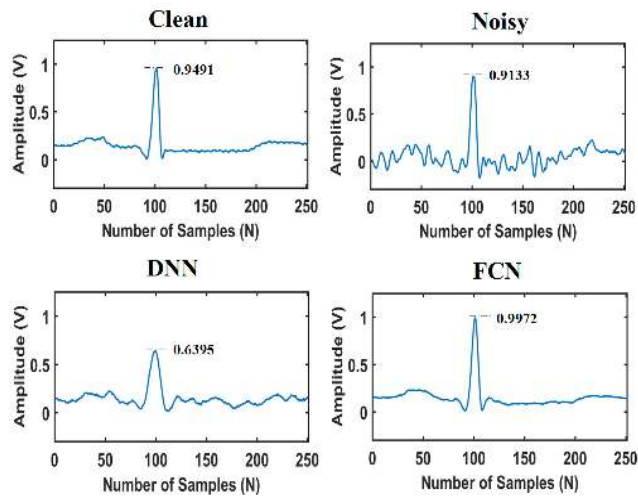
FIGURE 8. Example of waveform generated by DNN and FCN.

are only connected to 3 adjacent neurons in the last hidden layer $m - 1$. This architecture confines local patterns, and enables FCN to have the ability to extract and preserve local information effectively.

Fig. 8 shows an example of modeling waveforms by DNN and FCN with record 100. We can observe that DNN generates a smaller QRS complex. This failing may be attributed to the limitation of fully connected layers. Because the amplitude of QRS complexes is much higher than the remainder of ECG signals, DNN may be influenced by global properties and sacrifice this component during optimization process. This may be unsuitable to utilize feature representations from fully connected layers, because of the spatial information loss [36]. Consequently, DNN may not reconstruct ECG waveforms as precisely as FCN.

VI. CONCLUSION

This paper proposes a novel FCN-based DAE for denoising noisy ECG signals, contaminated by baseline wander, muscle artifact, and electrode motion. To the best of our knowledge, this is the first study on 1-D ECG signals using FCN-based DAE for the process of noise reduction. Performances of our algorithm shows higher SNR$_{imp}$, and lower RMSE and lower PRD compared to DNN- and CNN-based DAEs with the same compression ratio. With input SNR of -1 dB, the FCN achieves notable 31.83% [=(15.49-11.75)/11.75] and 9.55% [=(15.49-14.14)/14.14] relative SNR$_{imp}$ improvements and 32.98% [=(0.094-0.063)/0.094] and 12.50% [=(0.072-0.063)/0.072] relative RMSE reductions and 36.29% [=(30.89-19.68)/30.89] and 14.77% [=(23.09-19.68)/23.09] relative PRD reductions, when compared with the DNN and CNN, respectively. In the meanwhile, visual comparisons also indicate that the FCN-based DAE has better noise suppression and detail preservation than the other two models. Additionally, the proposed method obtains high compression performance, where each ECG signals with 1024 samples can be successfully reconstructed by representing only 32 dimensions. With high

noise reduction and low signal distortion, the practicality and superiority of the proposed method is suitable for clinical diagnosis.

REFERENCES

[1] A. Alwan, *Global Status Report on Noncommunicable Diseases 2010*. Geneva, Switzerland: World Health Org., 2011.

[2] E. J. Benjamin *et al.*, "Heart disease and stroke statistics-2017 update: A report from the American heart association," *Circulation*, vol. 135, no. 10, pp. e146–e603, Mar. 2017.

[3] R. M. John *et al.*, "Ventricular arrhythmias and sudden cardiac death," *Lancet*, vol. 380, no. 9852, pp. 1520–1529, Oct. 2012.

[4] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu, "ECG signal enhancement based on improved denoising auto-encoder," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 194–202, Jun. 2016.

[5] Y. Luo *et al.*, "A hierarchical method for removal of baseline drift from biomedical signals: Application in ECG analysis," *Sci. World J.*, vol. 2013, Apr. 2013, Art. no. 896056.

[6] V. de Pinto, "Filters for the reduction of baseline wander and muscle artifact in the ECG," *J. Electrocardiol.*, vol. 25, Supplement, p. 4048, 1992.

[7] M. J. Rooijakkers *et al.*, "Influence of electrode placement on signal quality for ambulatory pregnancy monitoring," *Comput. Math. Methods Med.*, vol. 2014, Feb. 2014, Art. no. 960980.

[8] C. Chandrakar and M. K. Kowar, "Denoising ECG signals using adaptive filter algorithm," *Int. J. Soft Comput. Eng.*, vol. 2, no. 1, pp. 120–123, 2012.

[9] M. Z. U. Rahman, R. A. Shaik, and D. V. R. K. Reddy, "Efficient and simplified adaptive noise cancelers for ECG sensor based remote health monitoring," *IEEE Sensors J.*, vol. 12, no. 3, pp. 566–573, Mar. 2012.

[10] M. H. Moradi, M. A. Rad, and R. B. Khezerloo, "ECG signal enhancement using adaptive Kalman filter and signal averaging," *Int. J. Cardiol.*, vol. 173, no. 3, pp. 553–555, May 2014.

[11] G. U. Reddy, M. Muralidhar, and S. Varadarajan, "ECG de-noising using improved thresholding based on wavelet transforms," *Int. J. Comput. Sci. Netw. Secur.*, vol. 9, no. 9, pp. 221–225, Sep. 2009.

[12] S. Liu, Y. Li, X. Hu, L. Liu, and D. Hao, "A novel thresholding method in removing noises of electrocardiogram based on wavelet transform," *J. Inf. Comput. Sci.*, vol. 10, no. 15, pp. 5031–5041, Oct. 2013.

[13] O. El B'Charri, L. Rachid, K. Elmansouri, A. Abenaou, and W. Jenkal, "ECG signal performance de-noising assessment based on threshold tuning of dual-tree wavelet transform," *Biomed. Eng. Online*, vol. 16, no. 1, p. 26, Feb. 2017.

[14] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[15] B. Weng, M. Blanco-Velasco, and K. E. Barner, "ECG denoising based on the empirical mode decomposition," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2006, pp. 1–4.

[16] M. A. Kabir and C. Shahnaz, "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains," *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 481–489, Sep. 2012.

[17] P. Xiong, H. Wang, M. Liu, and X. Liu, "Denoising autoencoder for eletrocardiogram signal enhancement," *J. Med. Imaging Health Inform.*, vol. 5, no. 8, pp. 1804–1810, Dec. 2015.

[18] P. Xiong, H. Wang, M. Liu, F. Lin, Z. Hou, and X. Liu, "A stacked contractive denoising auto-encoder for ECG signal denoising," *Physiol. Meas.*, vol. 37, no. 12, p. 2214, Nov. 2016.

[19] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, Jul. 2008, pp. 1096–1103.

[20] L. Deng, M. Seltzer, D. Yu, A. Acero, A. Mohamed, and G. Hinton, "Binary coding of speech spectrograms using a deep auto-encoder," in *Proc. Interspeech*, 2010, pp. 1692–1695.

[21] A. Gogna, A. Majumdar, and R. Ward, "Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2196–2205, Sep. 2017.

[22] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," in *Proc. Interspeech*, Aug. 2013, pp. 436–440.

[23] Y. H. Lai, F. Chen, S.-S. Wang, X. Lu, Y. Tsao, and C.-H. Lee, "A deep denoising autoencoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1568–1578, 2017.

[24] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 241–246.

[25] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[26] S.-W. Fu, Y. Tsao, X. Lu, and H. Kawai. (Mar. 2017). "Raw waveform-based speech enhancement by fully convolutional networks." [Online]. Available: https://arxiv.org/abs/1703.02205

[27] S.-W. Fu, T.-W. Wang, Y. Tsao, X. Lu, and H. Kawai, "End-to-end wave-form utterance enhancement for direct evaluation metrics optimization by fully convolutional neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 9, pp. 1570–1584, Sep. 2018.

[28] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[29] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. (Nov. 2015). "Fast and accurate deep network learning by exponential linear units (ELUs)." [Online]. Available: https://arxiv.org/abs/1511.07289

[30] S. Ioffe and C. Szegedy. (Feb. 2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167

[31] X.-J. Mao, C. Shen, and Y.-B. Yang. (Jun. 2016). "Image restoration using convolutional auto-encoders with symmetric skip connections." [Online]. Available: https://arxiv.org/abs/1606.08921

[32] G. B. Moody, R. G. Mark and A. L. Goldberger, "Physionet: A Web-based resource for the study of physiologic signals," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 70–75, 2001.

[33] K. Luo, J. Li, Z. Wang, and A. Cuschieri, "Patient-specific deep archi-tectural model for ECG classification," *J. Healthcare Eng.*, vol. 2017, May 2017, Art. no. 4108720.

[34] O. Yildirim, T. R. San, and U. R. Acharya, "An efficient compression of ECG signals using deep convolutional autoencoders," *Cognit. Syst. Res.*, vol. 52, pp. 198–211, Dec. 2018.

[35] D. P. Kingma and J. Ba. (Dec. 2014). "Adam: A method for stochastic optimization," [Online]. Available: https://arxiv.org/abs/1412.6980

[36] L. Jin, X. Shu, K. Li, Z. Li, G.-J. Qi, and J. Tang. (May 2018). "Deep ordinal hashing with spatial attention." [Online]. Available: https://arxiv.org/abs/1805.02459

HSIN-TIEN CHIANG received the B.S. degree from National Chiao Tung University, in 2016, and the M.S. degree from National Taiwan University, Taipei, Taiwan, in 2018. She is cur-rently a Research Assistant with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. Her research interests include biomedical signal processing and deep learning.

YI-YEN HSIEH received the B.S. and M.S. degrees in electrical engineering and electron-ics engineering from National Taiwan University, Taipei, Taiwan, in 2015 and 2017, respectively, where she is currently pursuing the Ph.D. degree with the Graduate Institute of Electronics Engi-neering. She is also a Research Assistant with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. Her research interests include biomedical signal processing, hardware accelerator design, machine learning, and deep learning.

SZU-WEI FU received the B.S. and M.S. degrees from the Department of Engineering Science and Ocean Engineering, Graduate Institute of Commu-nication Engineering, National Taiwan University, Taipei, Taiwan, in 2012 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the Department of Computer Science and Information Engineering. He is also a Research Assistant with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. His research interests include speech processing, speech enhancement, machine learning, and deep learning.

KUO-HSUAN HUNG received the B.S. degree from National Chiao Tung University, in 2015, and the M.S. degree from National Central Uni-versity, Taiwan, in 2017. He is currently pur-suing the Ph.D. degree with the Department of Biomedical Engineering, National Taiwan Uni-versity. He is also a Research Assistant with the Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan. His research interests include biomedical signal pro-cessing, noise reduction, source separation, and deep learning.

YU TSAO received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Insti-tute of Technology, Atlanta, GA, USA, in 2008. From 2009 to 2011, he was a Researcher with the National Institute of Information and Communi-cations Technology, Tokyo, Japan, where he was involved in research and product development of automatic speech recognition for multilingual speech-to-speech translation. He is currently an Associate Research Fellow with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. His research interests include speech and speaker recognition, acoustic and language modeling, audio-coding, and biosignal processing. He was a recipient of the Academia Sinica Career Development Award, in 2017, and the National Innovation Award, in 2018.

SHAO-YI CHIEN received the B.S. and Ph.D. degrees from the Department of Electrical Engi-neering, National Taiwan University (NTU), Taipei, Taiwan, in 1999 and 2003, respectively.

From 2003 to 2004, he was a Research Staff with the Quanta Research Institute, Taoyuan, Taiwan. In 2004, he joined the Department of Electrical Engineering, Graduate Institute of Elec-tronics Engineering, National Taiwan University, as an Assistant Professor. Since 2012, he has been a Professor. He served as the Vice Chairperson of the Department of Electrical Engineering, National Taiwan University, from 2013 to 2016. His research interests include computer graphics, real-time image/video processing, video coding, computer vision, and the associated VLSI and processor architectures. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, from 2009 to 2016, and *Circuits, Systems and Signal Processing* (CSSP, Springer), from 2009 to 2015. He also served as the Guest Editor for the Journal of Signal Processing Systems (Springer), in 2008, and an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: REGULAR PAPERS, from 2012 to 2013. He also serves on the Technical Program Committees of several conferences, including ISCAS, ICME, AICAS, SiPS, A-SSCC, and VLSI-DAT.

• • •