

# An End-to-End 12-Leading Electrocardiogram Diagnosis System Based on Deformable Convolutional Neural Network With Good Antinoise Ability

Lang Qin<sup>ID</sup>, Yuntao Xie<sup>ID</sup>, Xinwen Liu<sup>ID</sup>, Xiangchi Yuan<sup>ID</sup>, and Huan Wang<sup>ID</sup>

**Abstract**—Electrocardiogram (ECG) is a tool to help judge heart activity. In recent years, the convolutional neural network (CNN) and various deep learning algorithms have been widely used in ECG diagnosis. CNN only considers the local feature. However, the ECG signal is susceptible to noise, and the waveform is complex, making it difficult for existing methods to get a good result. This article presents a novel neural network architecture for ECG diagnosis based on deformable CNN (Deform-CNN). The architecture makes good use of the feature-learning capability of deformable convolution to learn the time-domain and lead characteristics of multilead ECG signals. The proposed end-to-end method can achieve an overall diagnostic accuracy of 86.3% in the 12-lead ECG data of CPSC-2018, with good antinoise ability, which makes the method have a more competitive performance than other deep learning algorithms. The source code is publicly available at <https://github.com/HeartbeatAI/Deform-CNN>.

**Index Terms**—Convolutional neural network (CNN), deformable convolution, electrocardiogram (ECG), noise immunity.

## I. INTRODUCTION

HEART disease has become one of the most critical public health problems in the world. It has leaped to the top fatality rate. About one-third of the world's annual deaths are heart disease and its complications [1]. At the same time, most cardiac diseases can be initially diagnosed by electrocardiogram (ECG) records as demonstrated in Fig. 1, so timely and effective diagnosis of cardiac diseases based on ECG is important [2].

Automatic detection of ECG is an important topic in the field of cardiology. Therapeutic techniques based on the conventional ECG analysis are incredibly dependent on doctor's knowledge and experience, which is time-consuming and requires a high qualification level. This can lead to a dilemma

Manuscript received December 7, 2020; revised March 28, 2021; accepted April 2, 2021. Date of publication April 16, 2021; date of current version May 6, 2021. This work was supported in part by the Innovation Fund of Glasgow College, University of Electronic Science and Technology of China (UESTC) and in part by the Sichuan Science and Technology Program under Grant 2020108. The Associate Editor coordinating the review process for this article was Dr. Adam G. Polak. (*Lang Qin and Yuntao Xie contributed equally to this work.*) (*Corresponding author: Huan Wang.*)

Lang Qin, Yuntao Xie, Xinwen Liu, and Xiangchi Yuan are with the Glasgow College, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: qinlang51@126.com; yuntao.xie@hotmail.com; xinwenliu.carrie@qq.com; yuanxiangchi@126.com).

Huan Wang is with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wh.huanwang@gmail.com).

Digital Object Identifier 10.1109/TIM.2021.3073707

of under-resourced and overburdened senior doctors and inexperienced new doctors who find it difficult to gain experience. As ECG often contain various noise [3], manual diagnosis is often influenced by them and misdiagnosed. In recent years, many automatic analyses based on ECG have been proposed to solve the above problem.

Many studies on the diagnosis of cardiac diseases are based on traditional machine learning. Some studies have used empirical mode decomposition (EMD) to better solve the high-frequency signal distortion caused by the filter and the baseline drift in the stationary state, which cannot be solved by the wavelet transformation [4], [5]. In the feature extraction part, the ECG signal's relevant features can be extracted by many methods, such as wavelet transform, S-transform, and the Hermite interpolation basis function [6], [7]. There are also alternative schemes, including hybrid expert system, support vector machine, genetic algorithm, and other problems that depend on experience, model, parameters, and so on [8]. The most commonly used methods in data classification are neural networks [9], [10], the support vector machine [11], the K-nearest neighbor algorithm [12], the decision tree [13], and the random forest [14]. Although relevant studies have achieved good results, it needs more computing resources to apply to the requirements of instant ECG diagnosis [15]. However, due to the traditional methods' complexity, it is challenging to achieve an end-to-end diagnosis, and because of too many nonlinear components of the signal, the diagnosis performance is not well, so practical application has many limitations.

Deep learning technology simulates biological neurons to automatically extract and process data through supervised or unsupervised learning, leaving the complex manual data extraction process to the computer to learn, significantly reducing data processing complexity automatically. With the development of deep learning, it has become an emerging force in ECG diagnosis. Among them, the long short-term memory (LSTM), which can process the timing signal more effectively [16], [17], the attention mechanism, which can focus on a specific range of data [18], and the residual network, which can effectively utilize the redundant information, can all achieve a good result in ECG diagnosis [19].

The convolutional neural network (CNN) is a better performing neural network in deep learning, which can achieve data feature extraction through local receptive fields, weight

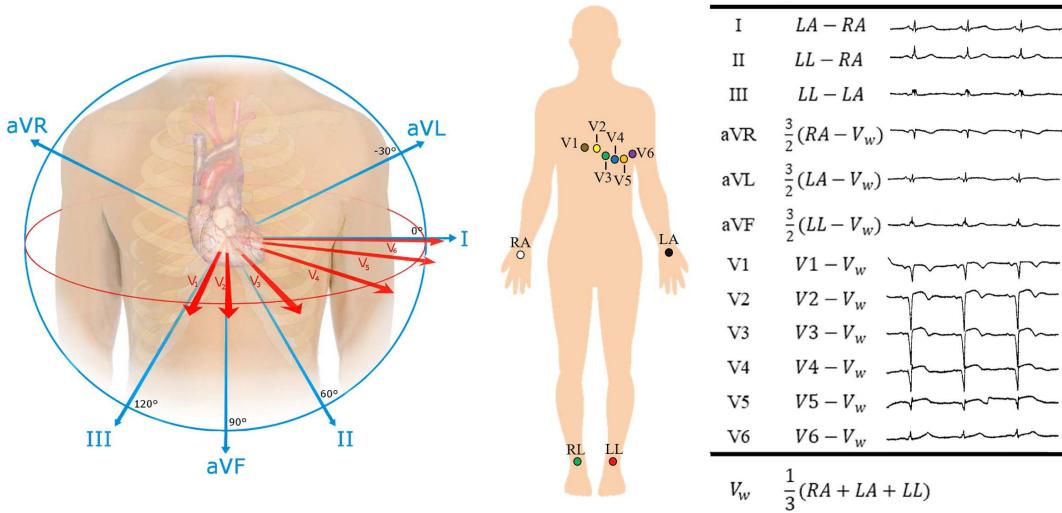


Fig. 1. Illustration of 12-lead ECG data.

sharing, downsampling, and other methods. CNN can simulate the mechanism of biological vision and is sensitive to small-scale features. Therefore, the existing deep learning methods based on CNN have a strong ability to extract small-scale features of images [20]. Compared with fully connected neural networks, it has fewer training parameters, better robustness, and diagnosis accuracy [21]. The research results of ECG diagnosis based on CNN can be divided into three categories. The first method is to use CNN as one stage of data processing using multistage processing, for example, the short-time Fourier transform (STFT), the frequency sliced wavelet transform, or the cross wavelet transform [22]. Huang *et al.* [23] transformed the time domain ECG data into the frequency domain using STFT and CNN. The second way is an improved end-to-end diagnostic method for CNN, such as 1-D CNN or multiscale CNN. Ince *et al.* [6] proposed an adaptive 1-D CNN based on beat detection and raw ECG morphology. The other is to combine it with different network structures, adding residual blocks, or attention module, using LSTM or bidirectional recurrent neural network (BRNN) to achieve good results. For example, Wang *et al.* [24] used the attention mechanism jointly with multiscale CNN and had a good result.

In multilead ECG, only local changes in a single lead are meaningful for 2-D CNN because only the intralead signal amplitudes are continuous, and the interlead signal amplitudes are discontinuous. Although multilead ECG signals are treated as 2-D matrices in this study, they differ from standard 2-D data, such as images. Diagnostic algorithms for cardiac diseases should focus on three attributes: integrity, diversity, and periodicity [15]. Integrity means that various leads reflect the cardiac status of the same patient. Thus, the intralead local changes in multilead ECG determine the diagnosis results. Diversity means that multilead ECG is synchronized ECG data based on various angles and distances of the patient's heart. Thus, it is reasonable to conclude that multilead ECG reflects various features of the heart. Periodicity refers to the widespread presence of periodicity with intralead and interlead. Because the ECG signal without signal processing is only a kind of time-domain signal, there

is no 2-D distance, and the data between leads are difficult to be processed by convolution, which is limited to the spatial relationship.

Considering the integrity, diversity, and periodicity of the 12-lead ECG signals, the introduction of deformable convolution can better adapt to ECG features by adaptively changing the receptive field [25]. Deformable convolution makes the convolution kernel disperse automatically by assigning position vectors for each convolution kernel weight so that the shape of the convolution kernel can also change automatically as the convolution kernel weight.

In this article, a novel ECG diagnosis system using deformable convolution is proposed to process ECG signals. Deformable CNN (Deform-CNN) is an end-to-end 12-lead ECG diagnostic algorithm without the need for complex data preprocessing and expert systems. Unlike traditional methods, it does not require detection of heartbeats or heart rhythms. The system proposed in this work is based on conventional convolution with deformable convolution and can be applied to long-term continuous cardiac detection and immediate cardiac disease diagnosis. Deform-CNN achieves a high rate of correct diagnosis. From the experiment, it has significant advantages over other deep learning methods in processing multilead ECG signals.

The contributions of this article are summarized as follows.

- 1) A novel end-to-end neural network model based on deformable convolution is proposed. It can take full advantage of the diversity between leads of ECG signals, integrity, and periodicity in single leads.
- 2) The deformable convolutional receptive field can change adaptively with the task, allowing the model to learn information in more complex geometric spatial structures. Compared to conventional CNN, it reduces the unobserved effects of the act of adjusting neural network parameters according to human cognition.
- 3) Explore the diagnostic performance of Deform-CNN in the face of noise. Three common noises were simulated to validate the noise immunity of Deform-CNN. Deform-CNN has good noise immunity as seen in the experimental results.

The article is organized as follows. In Section II, we will discuss the work related to Deform-CNN. The proposed Deform-CNN and the optimization process are described in detail in Section III. Section IV describes the experiment and gives the results of a comparison experiment between Deform-CNN and other structures. Section V gives the robustness of Deform-CNN to common ECG signal noises. Section VI summarizes the whole article.

## II. THEORETICAL BACKGROUND

The deformable convolution can select features according to the data characteristics of the actual signal. The kernels are not limited to the proximity of the signal in time. It is more like that the spatial transformer network (STN) disperses the original square convolution sliding window in an adaptive manner. The deformable convolution greatly preserving the original features in the signal and preventing the data extraction difficulties is caused by the separation of ECG data features [26].

As shown in Fig. 2, the conventional  $3 \times 3$  convolution's geometric structure is completely fixed, so its geometric spatial variability is limited. Each conventional  $3 \times 3$  convolution is used once, and the receptive field can be expanded from  $K \times K$  to  $(K + 2) \times (K + 2)$ . However, the spatial variability brought by multiple stacking is always limited. Many parameters brought about by conventional convolution are also a heavy burden on the computing equipment. As shown in Fig. 2(b), although dilated convolution can increase the receptive field faster, the receptive field's expansion rate increases linearly with the number of layers. The deformable convolution will break the fixed geometric structure. A deformable convolution layer can be realized by adding a deformable offset vector to the original position of the convolution kernel. As shown in Fig. 2(c), the convolution layer's data extraction ability will be enhanced, which significantly improves the generalization ability of the CNN and the capture of ECG.

The extraction of deformable convolution kernel is quite different from that of conventional convolution. We select a specific region and then add different weights to different sampling points in the region. Take a 2-D  $3 \times 3$  convolution kernel as an example. The expression of the convolution in an integral equation is shown in (1).

First, we sample the feature map  $x$  using a regular grid  $R$  and then summarize the sampled values weighted by  $w$ .  $p_0$  is the location of output feature map  $y$ , and  $p_n$  enumerates the location in  $R$ .  $R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_i). \quad (1)$$

However, for conventional convolutions, each  $p_n$  corresponds to the corresponding  $p_i$ , and there is no change in the position vector, so it can be expressed as

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n). \quad (2)$$

Deformable convolution has a position offset to be learned  $\{\Delta p_n | n = 1, 2, \dots, N-1, N\}$ , so there will be an offset vector,

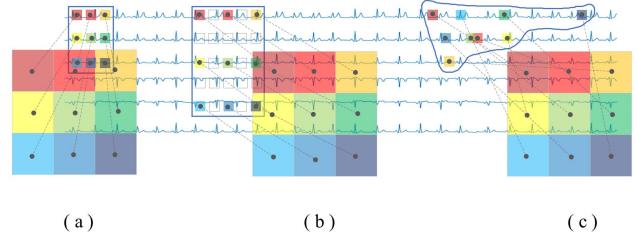


Fig. 2. Illustration of the sampling locations in  $3 \times 3$  conventional and deformable convolutions. (a) Regular sampling grid of conventional convolution. (b) Dilated convolution with dilation rate 2. (c) Deformed sampling locations with some offsets in deformable convolution.

where  $N = |R|$ , so the above formula changes as follows:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n). \quad (3)$$

However, due to decimals' unavoidable occurrence in the calculation, bilinear interpolation is used here for noninteger location information, so  $x(p_0 + p_n + \Delta p_n)$  here does not refer to the corresponding position. Instead, the position should be calculated by the following formula. Here,  $p = p_0 + p_n + \Delta p_n$  is the arbitrary position in the region,  $q$  is all spatial distribution vectors, and  $G(., .)$  represents a bilinear interpolation kernel, which can be decomposed into 1-D interpolation kernels in both directions of the  $x$ -axis and the  $y$ -axis because it is 2-D. Thus, it is calculated as in (4)–(6).  $n$  refers to the  $x$ -direction or the  $y$ -direction

$$x(p) = \sum_q G(q, p) \cdot x(q) \quad (4)$$

$$G(q, p) = g(q_x, p_x) \cdot g(q_y, p_y) \quad (5)$$

$$g(q_n, p_n) = \max(0, 1 - |q_n - p_n|). \quad (6)$$

The whole process is equivalent to setting the parameters to be learned in both sides of the (1). It learns the weights of each place and learns the location of them so that the network can adaptively diffuse the receptive field into the corresponding places.

As shown in Fig. 3(b) and (c), we divide the deformable convolution into three steps, starting with the part of the displacement part (boxed in blue). The complete displacement matrix offset comes from the  $x$ -direction displacement convolution  $pConv_x$  and the  $y$ -direction displacement convolution  $pConv_y$ . They are of the same spatial resolution and dilation as those of the current convolution layer. The next step is the corresponding convolution kernel weight matrix (boxed in yellowish) and similar to conventional convolution. Finally, the calculated displacement matrix is used to generate the actual displacement of the convolution kernel. The result is obtained by bilinear interpolation of the convolutional sum of the corresponding positions (boxed in red).

The ability to integrate more information due to the deformable convolution's larger receptive fields allows us to choose not to use large convolutions with many parameters [25]. Thus, the choice of deformable convolution improves accuracy by integrating information in a greater receptive field with only a slight increase in model complexity and computational effort, resulting in better performance in end-to-end

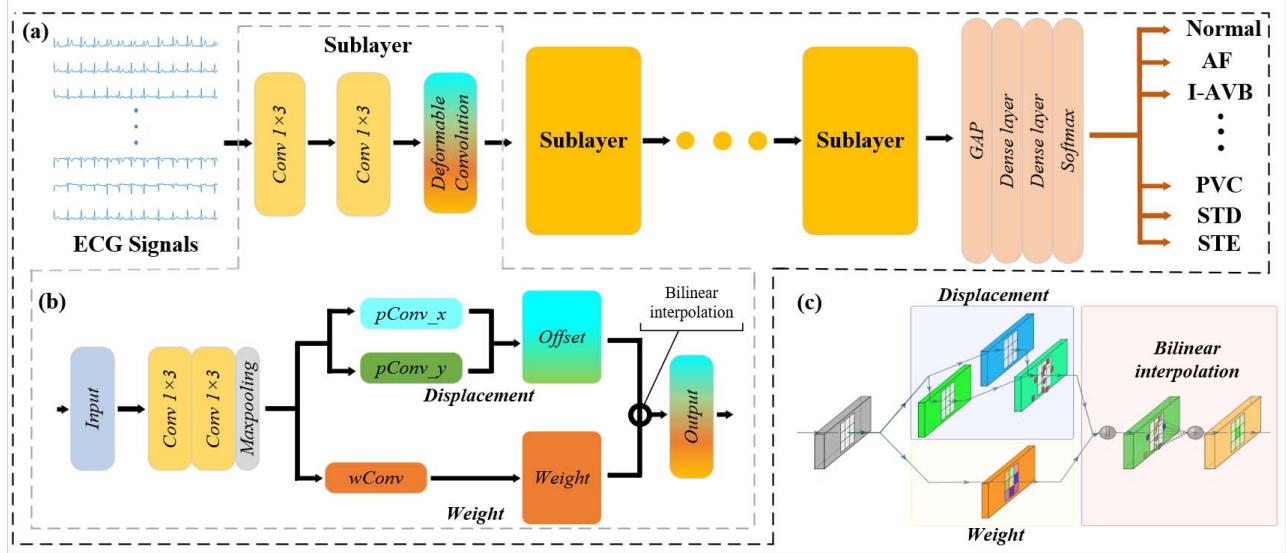


Fig. 3. Illustration of the structure of Deform-CNN. (a) Block diagram of the overall network flow. (b) Block diagram of sublayer in detail. (c) Influence of  $3 \times 3$  deformable convolution on the feature map.

ECG diagnostic tasks. Thus, it can be embedded in various structures as a reliable module in the overall neural network. Simultaneously, such a vast receptive field also compensates for the defects of CNN, showing strong adaptability.

Because the deformable convolution will extract corresponding features without artificially selecting some information, in this way, the loss of information can be reduced, and useful data can be obtained. Without additional monitoring signals, they can learn through the target task directly, reducing the complexity of the neural network and training costs. It can make the whole network truly meet the end-to-end requirements and achieve better results.

### III. PROPOSED DEFORM-CNN

#### A. Deform-CNN for ECG Diagnosis

Because conventional 2-D CNN does not fit the characteristics of ECG, meaningful intralead ECG local changes cannot be captured in conventional 2-D CNN. Considering the unique attributes of multilead ECG, Deform-CNN for the ECG diagnosis is proposed. The role of the deformable convolution kernel is summarized by Gao *et al.* [25] in two points: awareness of object scale and adaptation of effective receptive fields (EFRs). Thus, in ECG diagnostic tasks, the information between leads and the widely available periodic information can be well exploited by deformable convolution.

Fig. 3(a) illustrates the proposed Deform-CNN with the specific size of the feature map. Our network takes the 12-leading ECG as a  $12 \times 7500$  input (one lead per row) and outputs the diagnosis result in an end-to-end manner. It produces a set of feature maps with single-lead high resolution using CNN with a kernel size of  $1 \times 3$ . The feature maps after CNN have single-lead high resolutions but with fruitful detail information, while the feature maps after deformable convolution have low resolution but the information in the greater receptive field. Information in the greater receptive field can help to identify non-Euclidean data structures, such

as multilead ECG signals, and single-lead fine details enable the extraction of ECG details.

The structure of the Deform-CNN consists of four sublayers and a classification layer. The sublayers consist of two 1-D convolutions and a layer of deformable convolutions in series. Each sublayer takes the network input or the output of the previous sublayer as input to this layer. After obtaining the feature maps after two layers of 1-D CNN and 1-D max pooling, we implement deformable convolution by constructing displacement feature maps and weight feature maps, multiplying the shifted image of the displacement feature map with the weight feature map, and then calculating the shifted feature map after bilinear interpolation.

For such signal data, such as ECG, due to the spatial regularity of its features, the repeated stacking of deformable convolution brings too much spatial variability, which will lead to more difficulty in convergence during training. In terms of structure, we call the tandem structure of four layers of 1-D CNN, max-pooling, deformable convolution, and a sublayer. We construct the neural network by stacking sublayers four times to achieve integration of multilead ECG signal characteristics. After global pooling, the features are fed into the classification network. The detailed parameter table is shown in Table I.

#### B. Optimization Process of Deform-CNN

In Deform-CNN, the neural network structure with optimizable parameters consists of conventional convolutional layers and deformable convolutional layers. For conventional convolution, the optimization process for the weights is shown as follows:

$$W' \leftarrow W - \eta \nabla W \quad (7)$$

where  $W'$  is the updated matrix of weights,  $W$  is the matrix of weights before updating,  $\nabla W$  is the corresponding gradient of the error, and  $\eta$  is the learning rate.

TABLE I  
ARCHITECTURE OF THE DEFORM-CNN

Layer	Layer Details	Output
Input	—	12×7500
Sublayer_1	[Conv, 1×3, 32, stride=1]+ReLU	12×7500×32
	[Conv, 1×3, 32, stride=1]+ReLU	12×7500×32
	[Max Pooling, 1×4]	12×1875×32
	[Deform_Conv, 3×3, 32, stride=1]+ReLU	12×1875×32
Sublayer_2	[Conv, 1×3, 64, stride=1]+ReLU	12×1875×64
	[Conv, 1×3, 64, stride=1]+ReLU	12×1875×64
	[Max Pooling, 1×4]	12×469×64
	[Deform_Conv, 3×3, 64, stride=1]+ReLU	12×469×64
Sublayer_3	[Conv, 1×3, 128, stride=1]+ReLU	12×469×128
	[Conv, 1×3, 128, stride=1]+ReLU	12×469×128
	[Max Pooling, 1×4]	12×118×128
	[Deform_Conv, 3×3, 128, stride=1]+ReLU	12×118×128
Sublayer_4	[Conv, 1×3, 256, stride=1]+ReLU	12×118×256
	[Conv, 1×3, 256, stride=1]+ReLU	12×118×256
	[Max Pooling, 1×4]	12×30×256
	[Deform_Conv, 3×3, 256, stride=1]+ReLU	12×30×256
Classification	Global Average Pooling(GAP) [Dense layer]+Softmax types of ECG	256

As shown in Fig. 3(c), the deformable convolution is implemented by adding two displacement convolution layers to the conventional convolution to achieve displacement, with the displacement convolution kernel size matching the weight convolution kernel size and the output displacement and weights having the same resolution. The backward propagation includes two types of gradients:

- 1) the gradient of the weight kernels;
- 2) the gradient of the offset kernels.

The first type of gradients has the same form of the computation process as the conventional convolution. The flow of gradient of the offsets can be backpropagated through (4)–(6).

To be more specific, we can get the partial derivative of output  $O_i$  with respect to the  $x$  or  $y$  component of the offset, which can be computed as follows:

$$\frac{\partial O_i}{\partial \Delta p_x} = \sum_p x(p_i + p) \cdot \sum_q w_q \frac{\partial G(p + \Delta p, q)}{\partial \Delta p_x} \quad (8)$$

$$\frac{\partial G(q, p)}{\partial \Delta p_x} = g(q_y, p_y) \cdot \begin{cases} 0 & \text{if } |q_x - p_x| \geq 1 \\ 1 & \text{else if } q_x - p_x < 0 \\ -1 & \text{else if } q_x - p_x \geq 0 \end{cases} \quad (9)$$

$$\frac{\partial G(q, p)}{\partial \Delta p_y} = g(q_x, p_x) \cdot \begin{cases} 0 & \text{if } |q_y - p_y| \geq 1 \\ 1 & \text{else if } q_y - p_y < 0 \\ -1 & \text{else if } q_y - p_y \geq 0. \end{cases} \quad (10)$$

#### IV. EXPERIMENT VALIDATION

##### A. Data Sets' Description

The first data set used is from the 2018 China Physiological Signal Challenge's training set [27], named CPSC-2018. The data set contains nine categories: Normal, atrial fibrillation (AF), first-degree atrioventricular block (I-AVB), left bundle branch (LBBB), right bundle branch (RBBB), premature atrial contraction (PAC), premature ventricular contraction

TABLE II  
SERVER CONFIGURATION

Hardware information	CPU	Intel(R) Core(TM) i9-9900K
	RAM	32GB
	GPU	RTX 2080 8GB
	ROM	100 GB
Software information	NVIDIA driver version	418.67
	CUDA Version	10.1
	Python version	3.6.8
	Pytorch version	1.4.0+cu101

(PVC), ST-segment depression (STD), and ST-segment elevated (STE). Each recording is unequal in length, and the original data sampling rate is 500 Hz. We removed the unusual data, with 5850 recordings remaining. The remaining data of varying lengths are fully converted into signal data of 7500 lengths by reducing the sampling rate (down to 250 Hz) and repeating filling (making up 30 s of data with a total length of fewer than 30 s by copying the existing length.).

This second data set is unused data and validation set from the 2018 China Physiological Signal Challenge, named CPSC-Extra [28], which contains 73 different categories. Each record is unequal in length, and the original data sampling rate is 500 Hz. We selected data from nine categories, which is the same as CPSC-2018, and removed the unusual data, with 291 recordings remaining. In the experiments, this data set is used as an additional test set to verify the generalization of the model. Meanwhile, we used the same data preprocessing strategy as CPSC-2018.

##### B. Experiments Setup

Our network is implemented using Pytorch and Python 3.5, the network is trained and tested on a server with a 2080 GPU. The details of the server configuration were shown in Table II.

We use the stochastic gradient descent (SGD) optimizer with a learning rate of 0.0001. A learning rate decay operation is adopted with a momentum of 0.5. When the loss does not decrease for more than 5 epochs, the learning rate will decrease by 0.1%, and the minimum learning rate is 1e-7. In each experiment, the network is trained for 100 epochs, and the batch size is 8. The training set was put in a random order, including nine ECG types with different proportions. When training, the input data were randomly divided using the method of disordered training to avoid the influence of the training sequence on the results.

To better validate the performance of the proposed method, the data set used was randomly divided into the training set, the validation set, and the testing set, with the detailed data division shown in Fig. 4. For each experiment, the model with the best performance on the validation set was selected as the final testing model.

In order to verify the advantages of the proposed method, LSTM [16], Resnet [19], and visual geometry group (VGG) [20] are used for comparison. We also use two recently proposed well-performed ECG classification methods. Mostayed *et al.* [29] used a Bi-LSTM approach to

TABLE III  
OVERALL CLASSIFICATION RESULTS OF DIFFERENT STRUCTURES

Structure	3-C <sub>1</sub> D <sub>1</sub>	4-C <sub>1</sub> D <sub>1</sub>	5-C <sub>1</sub> D <sub>1</sub>	3-C <sub>2</sub> D <sub>1</sub>	4-C <sub>2</sub> D <sub>1</sub>	5-C <sub>2</sub> D <sub>1</sub>	3-C <sub>3</sub> D <sub>1</sub>	4-C <sub>3</sub> D <sub>1</sub>	5-C <sub>3</sub> D <sub>1</sub>
Accuracy	0.765	0.774	0.79	0.780	<b>0.863</b>	0.829	0.829	0.799	0.798

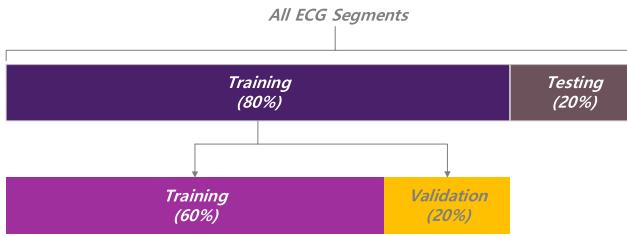


Fig. 4. Illustration of ECG data division for training, validation, and testing sets.

the diagnosis of ECG data. Liu *et al.* [30] proposed a modified Resnet framework, which includes a 17-layers CNN to detect deep features in ECG. Besides, conventional CNN consisting of 12 layers (the same number of layers as Deform-CNN) of  $3 \times 3$  conventional convolution is used for comparison as it has the same neural network depth as Deform-CNN. To fairly compare the performance of these methods, the training strategies and experimental environments of these methods are identical.

To measure the diagnostic performance of these methods from multiple perspectives, we introduce three performance indicators: accuracy,  $F_1$  score, and  $\kappa$  coefficient. Accuracy is a representation of the classification effect of the network structure.  $F_1$  score gives a more comprehensive picture of the actual diagnostic capability of the experimental results.  $\kappa$  coefficient is based on the confusion matrix and is commonly used for consistency testing, which is a measure of classification performance. These indicators can be defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (11)$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

$$F_1 \text{ score} = 2 \times \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (14)$$

$$\kappa = \frac{\text{Accuracy} - p_e}{1 - p_e}. \quad (15)$$

In the formulas, TP, FP, TN, and FN refer to the number of true positive samples, false positive samples, true negative samples, and false negative samples. Suppose that the number of real samples in each type to the total number of samples is  $p_{1x}, p_{2x}, p_{3x}, \dots$ , respectively, and the number of predicted samples in the corresponding type to the total number of samples is  $p_{x1}, p_{x2}, p_{x3}, \dots$ , respectively; then,  $p_e = \sum_{i=1}^9 p_{ix} p_{xi}$

$$\text{Noise-immunity} = \frac{\text{TT}}{\text{TT} + \text{TF}}. \quad (16)$$

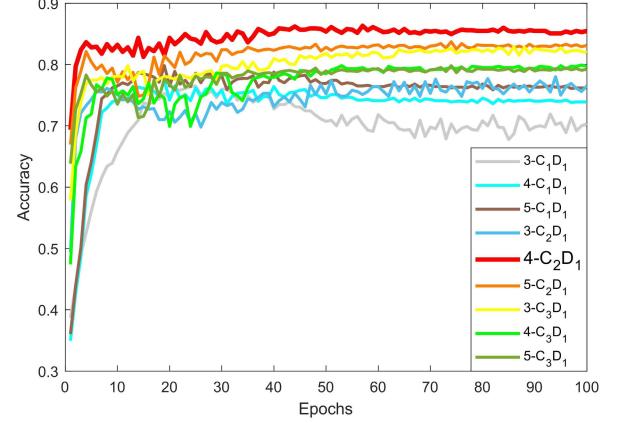


Fig. 5. Diagnostic accuracy for different Deform-CNN structures.

In addition, to measure the performance of the network in noisy condition, we introduce noise immunity. It is an indicator proposed by us to represent the tolerance to the noise of different network structures. TT refers to the number of samples that could be correctly classified before and after noise, and TF refers to those that could be classified before noise but could not be classified after noise.

### C. Exploration of the Deform-CNN Structure

Consider that different combinations of deformable and conventional convolution lead to different experimental results. The experiments were designed with different combination methods for the neural network structure. We have identified a more efficient basis for structural design  $i\text{-}C_aD_b$  through experiments conducted in preexperiment.  $C_aD_b$  represents the sublayer obtained by linearly concatenating  $a$  conventional convolution layer with the  $b$  deformable convolution layer;  $i$  represents the number of sublayers, which, together with the classification layer, forms the overall network.

As shown in Fig. 5 and Table III, the experiment was divided into nine groups, with different sublayer structures and different numbers of sublayers. The results are expressed by the diagnostic accuracy. From the diagnostic accuracy plot, all networks converge at around 50 training rounds, after which there is only a small oscillation in the diagnostic accuracy during training. Thus, we take the diagnostic accuracy after 100 epochs of training as a result. It can be seen from the diagnostic results that a network structure with a sublayer structure of  $C_2D_1$  has superior diagnostic performance in comparison. 4-C<sub>2</sub>D<sub>1</sub> had the highest diagnostic accuracy rate, reaching 86.3%. This is an improvement of 3.4% compared with 5-C<sub>2</sub>D<sub>1</sub>. Therefore, we use this structure as Deform-CNN in the following experiments.

TABLE IV  
COMPARISON WITH VALIDATED STRUCTURE IN DIFFERENT TESTING SETS

Data Source	Method	$F_1$ score for Different ECG Categories								Average $F_1$ score	Accuracy	$\kappa$ coefficient	
		Normal	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD				
CPSC-2018	LSTM [16]	0.730	0.792	0.763	0.848	0.909	0.268	0.763	0.800	0.105	0.664	0.773	0.731
	VGG-16[20]	0.750	0.861	0.874	0.857	0.918	0.333	0.859	<b>0.814</b>	0.462	0.748	0.813	0.779
	Resnet-18 [19]	0.730	0.882	0.877	0.786	0.905	0.487	0.733	0.784	0.444	0.737	0.798	0.763
	Resnet-50 [19]	0.712	0.876	0.845	0.760	0.904	0.395	0.716	0.772	0.500	0.720	0.779	0.740
	Mostayed <i>et al.</i> [29]	0.702	0.815	0.767	0.847	0.898	0.397	0.807	0.768	0.286	0.699	0.768	0.727
	Liu <i>et al.</i> [30]	0.768	0.887	0.880	0.867	0.912	0.600	0.798	0.784	0.353	0.761	0.797	0.789
	Conventional CNN	0.731	0.837	0.835	0.831	0.890	0.593	0.836	0.741	0.611	0.676	0.801	0.765
	Deform-CNN	<b>0.805</b>	<b>0.931</b>	<b>0.893</b>	<b>0.900</b>	<b>0.948</b>	<b>0.663</b>	<b>0.871</b>	0.800	<b>0.667</b>	<b>0.831</b>	<b>0.863</b>	<b>0.840</b>
CPSC-Extra	LSTM [16]	0.769	0.960	0.808	0.857	0.855	0.636	0.714	0.600	0.452	0.739	0.770	0.733
	VGG-16 [20]	0.795	0.889	0.714	<b>1.000</b>	0.942	0.276	0.906	0.786	0.545	0.761	0.810	0.779
	Resnet-18 [19]	0.905	0.914	0.863	<b>1.000</b>	0.967	0.571	0.787	0.754	0.571	0.815	0.832	0.807
	Resnet-50 [19]	0.764	0.895	0.809	0.750	0.906	0.524	0.846	<b>0.847</b>	0.667	0.779	0.814	0.783
	Mostayed <i>et al.</i> [29]	0.841	0.860	0.830	0.857	0.902	0.667	0.820	0.746	0.588	0.790	0.818	0.789
	Liu <i>et al.</i> [30]	0.800	0.866	0.792	0.824	0.915	0.571	<b>0.939</b>	0.714	<b>0.700</b>	0.791	0.822	0.791
	Conventional CNN	0.861	0.914	0.816	0.900	0.949	0.688	0.847	0.679	0.526	0.798	0.818	0.789
	Deform-CNN	<b>0.998</b>	<b>0.987</b>	<b>0.923</b>	0.947	<b>0.992</b>	<b>0.927</b>	0.703	0.841	0.647	<b>0.885</b>	<b>0.897</b>	<b>0.881</b>

The highest score for each class is in bold.

#### D. Comparison With Existing Methods

To verify our proposed method's performance, we compared our proposed Deform-CNN with recently proposed well-performing ECG classification methods and basic neural networks.

In Table IV, we compare the  $F_1$  score for different ECG categories of eight different methods. As shown in Table IV, the proposed Deform-CNN has an  $F_1$  score of 0.831, which is comparatively better than the other methods. For each individual ECG category, our model has the highest  $F_1$  score in eight of the total nine categories. To be specific, the conventional CNN and LSTM [16] methods are surpassed by our approach by 0.155 and 0.167 in  $F_1$  score. Furthermore, the  $F_1$  scores of all categories performed well than others except for STD, which was slightly lower than the VGG-16.

When it comes to the  $F_1$  scores of CPSC-Extra for each model, the Deform-CNN performs well than other methods. To be specific, the proposed method reaches the highest  $F_1$  score of 0.885, which is 0.070 higher than the second-ranked method, Resnet18 [19], which has an  $F_1$  score of 0.815. Compared with conventional CNN, our Deform-CNN has a 0.087 improvement. Besides, the  $F_1$  scores of Normal, AF, I-AVB, RBBB, and PAC are higher than others.

We also list the accuracy and  $\kappa$  coefficients for each network in Table IV, and our proposed network has the highest  $\kappa$  coefficient and accuracy in both CPSC-2018 and CPSC-Extra compared to other methods. In detail, the accuracy of Deform-CNN in CPSC-2018 is 0.863. Compared with the second-ranked VGG-16 [20], conventional CNN, and Mostayed *et al.* [29] in the last place, our Deform-CNN has 0.05, 0.062, and 0.095 improvements, respectively. In CPSC-Extra, our method reaches the highest accuracy of 0.897, the increment is between 0.065 and 0.127 compare with other methods, and the average increase amounted to 0.085. The  $\kappa$  coefficients of Deform-CNN are 0.840 and 0.881 in CPSC-2018 and CPSC-Extra, respectively, which is also higher than others. The proposed method without complex preprocessing can obtain high  $F_1$  scores,  $\kappa$  coefficients, and

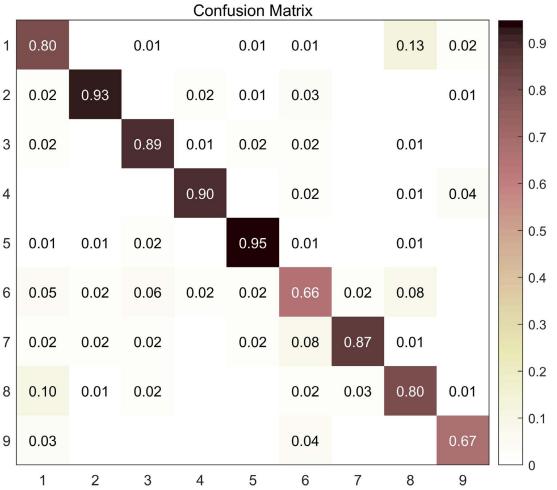


Fig. 6. Normalized confusion matrix of Deform-CNN.

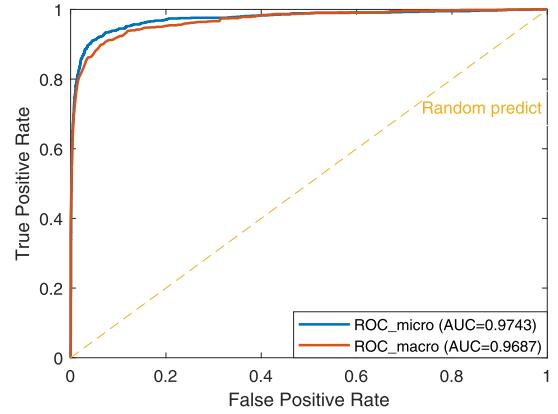


Fig. 7. ROC curve of Deform-CNN.

accuracy on two different data sets. Our method has good performance and generalization ability.

Also, we present the confusion matrix, receiver operating characteristic (ROC) curve, and the area under curve (AUC) for Deformable-CNN in Figs. 6 and 7, respectively. It can

TABLE V  
COMPUTATIONAL COMPLEXITY OF THE VALIDATED STRUCTURE

Method	LSTM [16]	VGG-16 [20]	Resnet-18 [19]	Resnet-50 [19]	Mostayed et al. [29]	Liu et al. [30]	Conventional CNN	<b>Deform-CNN</b>
Training Time(s/epoch)	368	347	48	91	341	12	13	50
Testing Time(s/epoch)	20	26	5	7	27	<1	<1	3
Parameters(M)	0.203	134.296	11.175	23.530	0.327	7.190	2.025	1.375
Flops(G)	1.579	95.481	4.849	10.854	2.537	1.551	0.803	0.825
Accuracy	0.773	0.813	0.798	0.779	0.768	0.797	0.801	0.863

be seen from the confusion matrix that Deform-CNN has a good classification for almost nine types of ECG. The ROC curve refers to the line of points obtained from the network output at different classification criteria with FP and TP as the axis coordinates. It can be a good reflection of the classification performance of the network structure itself. Since both micro-AUC and macro-AUC from the ROC curve are greater than 95%, it can be concluded that Deform-CNN has an excellent performance in classifying different cardiac diseases.

#### E. Analysis of the Complexity

In ECG diagnosis tasks, the immediacy of diagnosis and the number of parameters is also important criteria for judging the goodness of the network structure. Therefore, we give the training/testing time and the number of parameters for Deform-CNN and the validated structure.

Results are recorded in Table V. We can see from Table V that the complexity of Deform-CNN in time scale (training time and testing time) is higher than that of Conventional CNN and Liu *et al.* [30]; it has a clear advantage over the rest of the other validated methods. The number of parameters is lower than other validated methods due to the large number of conventional  $1 \times 3$  convolutions in Deform-CNN; the same applies to Flops. Remarkably, deformable convolutional layers require more parameters and computation time than conventional convolutional layers. However, from the experimental results, the increase in parameters and computation time is not significant, and that of the specific network can be even reduced by combining the deformable convolution with simpler modules.

## V. ANTINOISE PERFORMANCE

#### A. Noisy Data Construction

Noise is one of the major problems when recording the signal. The existing methods are more or less disadvantageous and difficult to meet end-to-end requirements [31], [32]. Instead of separating the filter from the neural network, it is better to find a neural network structure with better noise immunity and to fuse a simple filter to achieve better results.

Simulation experiments are needed to test and evaluate the performance of the denoising approach. We need to produce the noise before the simulation. However, it is not a time-consuming choice if we include other relative algorithms in order to assess the author's method objectively. Thus, we choose the method used in the breakthrough work of

Donoho [33] and Donoho and Johnstone [34] in the field of wavelet denoising technique.

The approach used by Donoho to make the noisy signal assumes all noise as additive noise. The noises were generated through a bandpass Chebyshev type II filter in the corresponding noise frequency range according to the corresponding SNR. The “clean” signal is normalized by the standard deviation, multiplied by square-root SNR (SQRT-SNR). Then, the noisy signal with the prespecified SQRT-SNR is the normalized signal plus processed Gaussian noise. The method is widely adopted by many researchers and became a standard function of the wavelet toolbox in MATLAB.

We set  $f'$ ,  $\Delta$ , and  $f$  to denote the “clean” signal, the additive Gaussian noise with variance  $\sigma^2 = 1$ , and the noisy signal, respectively.  $N$  is set to be the length of  $f'$ ,  $STD(f')$  is the standard variation of “clean” signal, and SNR is the prespecified SQRT-SNR.

Thus, the formula for the noisy signal is constructed as follows:

$$STD(f') = \left\{ \frac{1}{N-1} \sum_{i=1}^N [f'(i) - \bar{f'}]^2 \right\}^{\frac{1}{2}} \quad (17)$$

$$f = \frac{SNR}{STD(f')} \times f' + \Delta. \quad (18)$$

#### B. Robustness of Deform-CNN Against Powerline Frequency Interference

The powerline frequency noise is caused by the interference of the capacitance and electrode lead loop in the human body by the magnetic field and so on [35], resulting in a 50–60-Hz sinusoidal signal and its various harmonics. Because there are a large number of small changes in the ECG signal, the presence of powerline frequency noise will mask these small changes, making the SNR of the ECG signal decrease [36], thus affecting the judgment of the heart disease, according to different circumstances and its amplitude. This is roughly equivalent to 5%–40% of the *R*-wave amplitude. Fig. 8 illustrates the effect of powerline frequency interference on the ECG signal.

This part of the experiment is divided into two parts. The first is a comparison of the recognition accuracy of the Deform-CNN with other seven outstanding deep learning algorithms for normal data and noisy data. The second is the test between Deform-CNN and some of the previously proposed structures to investigate the effect of the deformable convolutional module on noise immunity.

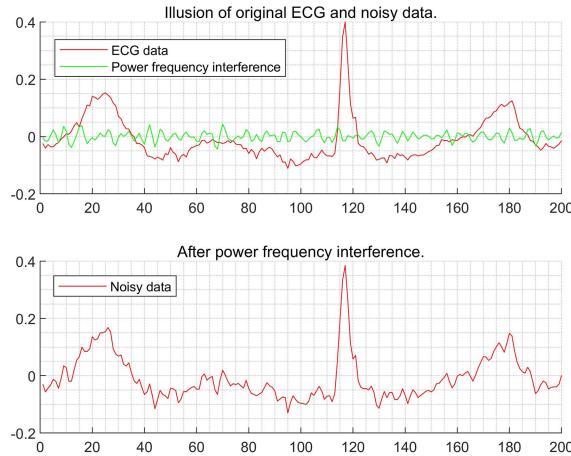


Fig. 8. Illustration of the effect of powerline frequency interference on ECG signals.

TABLE VI  
COMPARISON OF DIAGNOSTIC ACCURACY AFTER POWERLINE FREQUENCY INTERFERENCE

Method	Original Data	Noisy Data
LSTM [16]	0.773	0.772
VGG-16 [20]	0.813	0.788
Resnet-18 [19]	0.798	0.762
Resnet-50 [19]	0.779	0.749
Mostayed <i>et al.</i> [29]	0.768	0.766
Liu <i>et al.</i> [30]	0.797	0.790
Conventional CNN	0.801	0.782
Deform-CNN	<b>0.863</b>	<b>0.853</b>

The following experimental data are presented in Table VI for comparison with Deform-CNN using seven types of other deep learning structures. Table VI gives the accuracy of ECG diagnosis for each neural network.

Conventional CNN is used here to represent a Deform-CNN in which the deformable convolution is completely replaced by conventional convolution, the overall structure of which is equivalent to a 12 layers  $3 \times 3$  CNN. While LSTM [16] is a structure with two recurrent layers, each containing 100 recurrent cells in which the recurrent cells are bidirectional LSTM cells.

As can be seen from Table VI, both conventional CNN, Resnet, and VGG have a relatively significant performance degradation, with an average decrease in diagnosis accuracy of 2%–3%, which means that VGG and Resnet cannot improve the poor noise immunity of baseline drift. However, the diagnosis performance of Deform-CNN was only reduced by 1%, indicating that the deformable convolution can help CNN to effectively reduce the impact of powerline frequency interference. LSTM was hardly affected by the noise and the diagnosis accuracy rate dropped by only 0.1%. The structures of Mostayed *et al.* [29] and Liu *et al.* [30] also have strong noise immunity with 0.7% and 0.2% decreases in the correct diagnostic rate, respectively. However, because of their low diagnosis accuracy of the original data, the lower performance

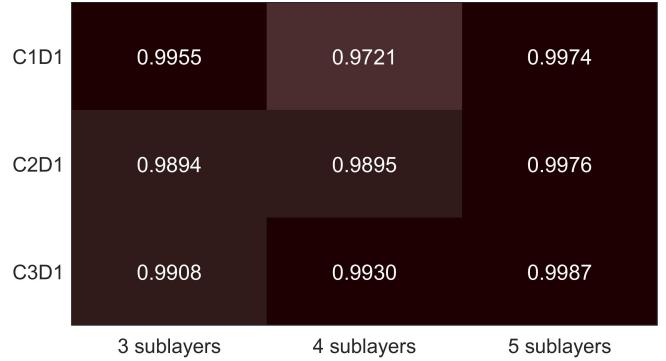


Fig. 9. Noise immunity of different structures to powerline frequency interference.

degradation does not result in a good diagnostic accuracy of noisy data.

To investigate the relationship between the structure and its noise immunity, we used structures in Section IV-C for this experiment. Noise immunity was used to characterize the performance of different structures facing noise. For ease of viewing the results, a color shade has been used to represent how much of this is accounted for; the higher the percentage, the darker the color. The result was shown in Fig. 9.

As can be seen from Fig. 9, the overall noise immunity of all structures is high, with a noise immunity of almost 0.98 or higher, indicating that the inclusion of deformable convolution can effectively suppress the powerline frequency interference. In general, noise immunity increases with the number of sublayers and increases slightly with the number of conventional convolution layers in sublayers. The shallow layers of the neural network are susceptible to this because of the higher frequency of such noise [37]. As the number of layers increases and the receptive field of the neural network expands, the effect of such high-frequency noise is reduced. This process can be accelerated by adding deformable convolution due to its large receptive field. Noise immunity can be improved by the changes of sublayer structures, while it can be better improved by an increase in the number of sublayers.

### C. Robustness of Deform-CNN Against Baseline Drift

In the original signal acquisition, the original ECG signal drift amplitude reaches 0.1–0.2 times of the maximum amplitude due to the low-frequency noise caused by the interference of signal recording and electronic equipment or respiratory interference and motion artifacts [38], which has the greatest impact on the ECG signal. The ECG signal disturbed by such noise becomes distorted and jitters up and down; thus, it is difficult to determine the position of signal feature points. Its frequency range is generally set at 0.15–0.3 Hz, and the amplitude is set at about 15 of the maximum amplitude. Fig. 10 illustrates the effect of baseline drift on the ECG signal.

This experiment can also be divided into two parts. Table VII compared the performance of Deform-CNN with seven other deep learning algorithms, giving their recognition accuracy for both normal and noisy data.

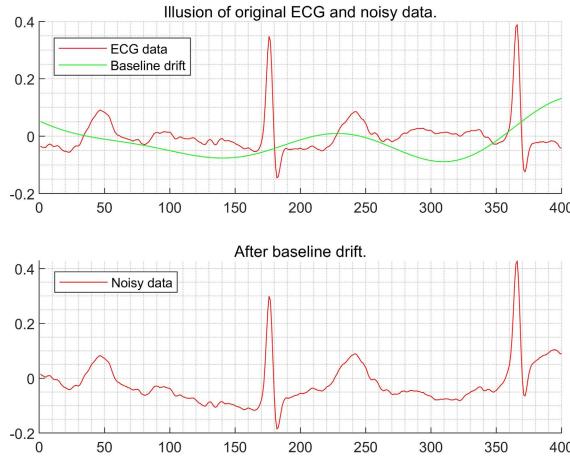


Fig. 10. Illustration of the effect of baseline drift on ECG signals.

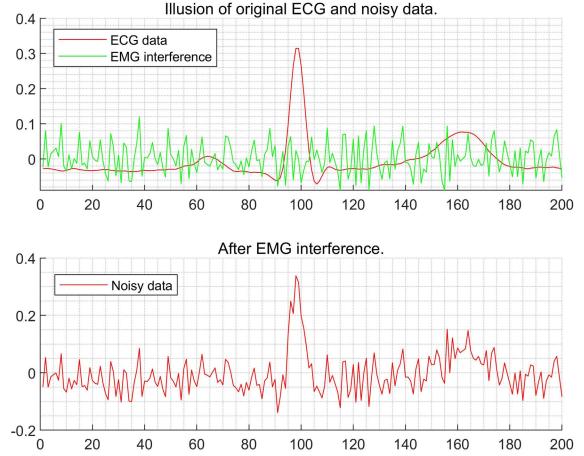


Fig. 12. Illustration of the effect of EMG interference on ECG signals.

TABLE VII  
COMPARISON OF DIAGNOSTIC ACCURACY AFTER BASELINE DRIFT

Method	Original Data	Noisy Data
LSTM [16]	0.773	0.671
VGG-16 [20]	0.813	0.673
Resnet-18 [19]	0.798	0.688
Resnet-50 [19]	0.779	0.612
Mostayed et al. [29]	0.768	0.671
Liu et al. [30]	0.797	0.751
Conventional CNN	0.801	0.628
Deform-CNN	<b>0.863</b>	<b>0.813</b>

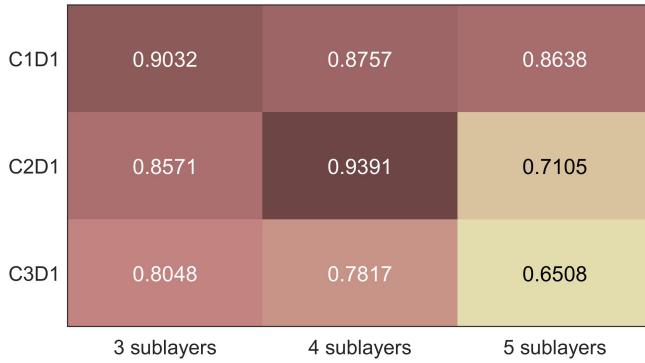


Fig. 11. Noise immunity of different structures to baseline drift.

As can be seen from Table VII, the baseline drift had a large impact on the experimental results. Average performance degradation is up to 10% to 20%. In comparison, the Deform-CNN's noise immunity decreased by only 5%, indicating that the Deform-CNN is effective against such noise. The remaining structures in the experiment have a large performance degradation except for the structure proposed by Liu *et al.* [30] with only a small decrease by 4.6%.

As shown in Fig. 11, noise immunity of structures is lower dealing with the data after baseline drifts except for 4-C<sub>2</sub>D<sub>1</sub>. Thus, 4-C<sub>2</sub>D<sub>1</sub>, with the noise immunity of 0.9391, has some superiority in this aspect. The overall trend roughly shows that the noise immunity decreases as the number of conven-

tional convolution layers in the sublayers increases except for 4-C<sub>2</sub>D<sub>1</sub>. Since the baseline drift is a type of low-frequency noise with a large amplitude, it is difficult to filter out this type of low-frequency information in the small receptive fields and can even lead to more severe distortion of the original data. Although the deformable convolution has a large receptive field, the data distortion caused by the sublayer interferes with the data acquisition of the deformable convolution because the deformable convolution is located after the conventional convolution in the sublayer and has few sampling points. This results in a decrease in noise immunity as the number of sublayers and the number of conventional convolution layers in the sublayers increase. Also, due to the deformable convolution with few sample points, the structure of the neural network designed in some way will not significantly be affected by the data distortion caused by conventional convolution, but rather the baseline drift is better filtered out by the large receptive fields of deformable convolution.

#### D. Robustness of Deform-CNN Against Electromyography Interference

Electromyography (EMG) interference is due to the diversity of signals in the human body [39]. Some biomass is a signal in one case and noise in another case. The bioelectrical signals caused by the contraction of numerous muscle fibers are intermingled, and the noise is formed by the action of ECG electrodes. It is mainly due to the trembling of muscle fibers, resulting in changes in body surface potential. The potential difference measured by electrode patches on the body surface is affected, resulting in a short duration, which is the noise of ECG producing tiny ripples. Due to its diverse sources, this kind of noise is similar to white noise with no fixed incidence, a wide range of frequency distribution, relatively more between 30 and 300 Hz, and the data generally remain around 25 mV. Fig. 12 illustrates the effect of electromyography interference on the ECG signal.

This experiment consists of three parts. Table VIII gives the accuracy of different structures for normal and noisy data. Fig. 13 verifies whether a Gaussian filter can effectively

TABLE VIII  
COMPARISON OF DIAGNOSTIC ACCURACY AFTER EMG INTERFERENCE

Method	Original Data	Noisy Data
LSTM [16]	0.773	0.766
VGG-16 [20]	0.813	0.678
Resnet-18 [19]	0.798	0.611
Resnet-50 [19]	0.779	0.608
Mostayed et al. [29]	0.768	0.759
Liu et al. [30]	0.797	0.757
Conventional CNN	0.801	0.644
Deform-CNN	<b>0.863</b>	<b>0.660(0.828)</b>

The number in brackets represents the recognition accuracy of Deform-CNN after the noise addition data has been filtered by a Gaussian filter.

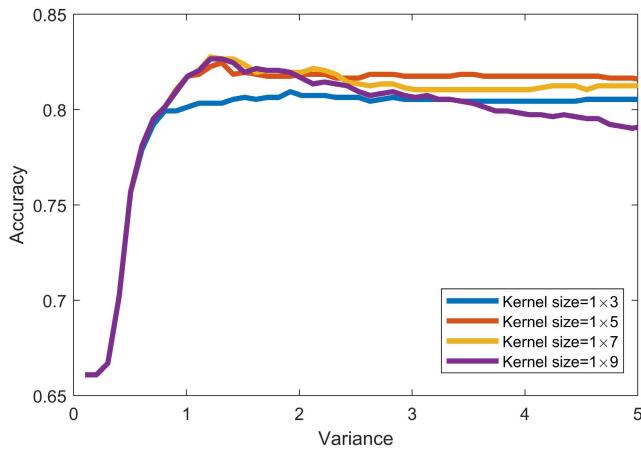


Fig. 13. Classification accuracy of Deform-CNN for noisy data after different Gaussian filters.

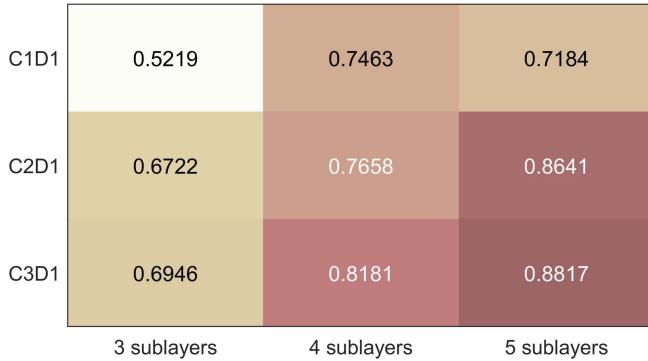


Fig. 14. Noise immunity of different structures to EMG interference.

improve the noise immunity of Deform-CNN to EMG noise. Fig. 14 indicates the noise immunity of proposed structures in Section IV-C.

Table VIII compared the performance of Deform-CNN with seven other deep learning algorithms, giving their recognition accuracy for both normal and noisy data.

It can be seen from Table VIII that the EMG noise has a strong attack performance against all network structures except for LSTM [16]. Considering the excellent performance of LSTM [16] in EMG noise and powerline frequency interference, we will consider introducing the idea of LSTM [16] in future work to improve the noise performance of networks.

Deform-CNN performance drop of almost 20% is the most pronounced here. It is, therefore, necessary to investigate this phenomenon.

In consideration of the weak defenses of Deform-CNN against broadband high-frequency noise like EMG, we try to use a low-pass filter to reduce such effects. Therefore, an attempt was made to classify the noise-added data again with Deform-CNN after passing it through a Gaussian filter. We experimented with four different Gaussian filter sizes and set up 50 sets of experiments from 0.1 to 5 with 0.1 as an interval. The classification results are shown in Fig. 13.

From Fig. 13, a  $1 \times 7$  Gaussian filter with a variance of 0.400 can increase the accuracy to 82.8%. This greatly improves its classification performance. Therefore, the use of a filter largely reduces the impact of EMG on recognition accuracy.

Fig. 14 gives a comparison of the noise immunity of different structures against EMG.

When dealing with data adding EMG interference, all the noise immunities are low, with few above 0.8. It can be seen from Fig. 14 that the overall noise immunity increases significantly with the number of sublayers and the number of convolution layers in the sublayers. As the number of sublayers increases, the change in noise immunity is more pronounced. EMG interference is a class of the high-frequency noise, such as powerline frequency interference, but it has a much wider frequency range, which makes it more difficult to filter out. The increase in deformable convolution and the increase in the number of convolution layers can improve the diagnosis accuracy of the data after noise addition, but it is difficult to achieve high noise immunity.

## VI. CONCLUSION

In this article, Deform-CNN was designed by introducing deformable convolution in CNN, where we use CPSC-2018 as the training set and obtain good classification results, which means that Deform-CNN is better suited to ECG signal characteristics and can perform ECG classification tasks effectively. In ECG classification tasks, it is possible to combine data features more efficiently than in other network structures, achieving higher recognition accuracy with a smaller number of parameters.

In ECG classification tasks, Deform-CNN can synthesize feature data more efficiently than commonly used LSTM, the attention module, and many other methods. It is based on the periodicity of the ECG itself and the correlation of data between leads, showing greater adaptability in ECG diagnosis. Deform-CNN can also be made more robust by adding a Gaussian filter. Compared with other structures, Deform-CNN can effectively identify data containing all three ECG noise types, so it has a much broader application potential.

Since Deform-CNN's input data do not require excessive preprocess, it also has high accuracy in ECG diagnosis, low overall network complexity, and high stability, which makes end-to-end ECG diagnosis possible.

Deform-CNN has a good antinoise ability, but some noise types can still lead to significant damage to the network's performance. The addition of filters can effectively improve

this situation, but this is not sufficient for end-to-end requirements. In future work, we will explore in depth the relationship between antinoise ability and the network structure and explore deep learning-based ECG denoising algorithms and ECG diagnostic framework with better antinoise performance.

## REFERENCES

- [1] WHO | *Cardiovascular Diseases (CVDs) Factsheet*, World Health Organization, Geneva, Switzerland, 2013.
- [2] S. Kaplan Berkaya, A. K. Uysal, E. Sora Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," *Biomed. Signal Process. Control*, vol. 43, pp. 216–235, May 2018.
- [3] G. M. Friesen, T. C. Jannett, M. A. Jadallah, S. L. Yates, S. R. Quint, and H. T. Nagle, "A comparison of the noise sensitivity of nine QRS detection algorithms," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 1, pp. 85–98, Jan. 1990.
- [4] S. Aziz, M. U. Khan, Z. Ahmad Choudhry, A. Aymin, and A. Usman, "ECG-based biometric authentication using empirical mode decomposition and support vector machines," in *Proc. IEEE 10th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2019, pp. 0906–0912.
- [5] M. Abdelazez, S. Rajan, and A. D. C. Chan, "Detection of atrial fibrillation in compressively sensed electrocardiogram measurements," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
- [6] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.
- [7] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012.
- [8] M. Moavenian and H. Khorrami, "A qualitative comparison of artificial neural networks and support vector machines in ECG arrhythmias classification," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3088–3093, Apr. 2010.
- [9] M. Hammad, A. M. Ilyasu, A. Subasi, E. S. L. Ho, and A. A. A. El-Latif, "A multitier deep learning model for arrhythmia detection," *IEEE Trans. Instrum. Meas.*, vol. 70, 2020, Art. no. 2502809.
- [10] N. A. Bhaskar, "Performance analysis of support vector machine and neural networks in detection of myocardial infarction," *Procedia Comput. Sci.*, vol. 46, pp. 20–30, 2015.
- [11] R. Thilagavathy, R. Srivatsan, S. Sreekarun, D. Sudeshna, P. L. Priya, and B. Venkataramani, "Real-time ECG signal feature extraction and classification using support vector machine," in *Proc. Int. Conf. Comput. Appl. (IC3A)*, Feb. 2020, pp. 44–48.
- [12] I. Saini, D. Singh, and A. Khosla, "Delineation of ECG wave components using K-nearest neighbor (KNN) algorithm: ECG wave delineation using KNN," in *Proc. 10th Int. Conf. Inf. Technol.: New Generations*, Apr. 2013, pp. 712–717.
- [13] L. Zhang, H. Peng, and C. Yu, "An approach for ECG classification based on wavelet feature extraction and decision tree," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2010, pp. 1–4.
- [14] E. Manibardo *et al.*, "ECG-based random forest classifier for cardiac arrest rhythms," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 1504–1508.
- [15] W. Liu *et al.*, "Real-time multilead convolutional neural network for myocardial infarction detection," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1434–1444, Sep. 2018.
- [16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [17] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM-based auto-encoder model for ECG arrhythmias classification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1232–1240, Apr. 2020.
- [18] R. Wang, Q. Yao, X. Fan, and Y. Li, "Multi-class arrhythmia detection based on neural network with multi-stage features fusion," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 4082–4087.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, Apr. 2015. [Online]. Available: [https://www.robots.ox.ac.uk/~vgg/research/very\\_deep/](https://www.robots.ox.ac.uk/~vgg/research/very_deep/)
- [21] B. Pourbabae, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.
- [22] S. Banerjee and M. Mitra, "Application of cross wavelet transform for ECG pattern analysis and classification," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 326–333, Feb. 2014.
- [23] J. Huang, B. Chen, B. Yao, and W. He, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019.
- [24] R. Wang, J. Fan, and Y. Li, "Deep multi-scale fusion neural network for multi-class arrhythmia detection," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 9, pp. 2461–2472, Sep. 2020.
- [25] H. Gao, X. Zhu, S. Lin, and J. Dai, "Deformable kernels: Adapting effective receptive fields for object deformation," in *Proc. ICLR*, Feb. 2020. [Online]. Available: <http://people.eecs.berkeley.edu/~hangg/deformable-kernels/>
- [26] J. Dai *et al.*, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.
- [27] F. Liu *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *J. Med. Imag. Health Informat.*, vol. 8, no. 7, pp. 1368–1373, Sep. 2018.
- [28] E. A. Perez Alday *et al.*, "Classification of 12-lead ECGs: The physionet/computing in cardiology challenge 2020," *Physiological Meas.*, vol. 41, no. 12, Jan. 2021, Art. no. 124003.
- [29] A. Mostayed, J. Luo, X. Shu, and W. Wee, "Classification of 12-lead ECG signals with Bi-directional LSTM network," 2018, *arXiv:1811.02090*. [Online]. Available: <https://arxiv.org/abs/1811.02090>
- [30] Z. Liu, X. Meng, J. Cui, Z. Huang, and J. Wu, "Automatic identification of abnormalities in 12-lead ECGs using expert features and convolutional neural networks," in *Proc. Int. Conf. Sensor Netw. Signal Process. (SNSP)*, Los Alamitos, CA, USA, Oct. 2018, pp. 163–167.
- [31] C. D. Mcmanus, K.-D. Neubert, and E. Cramer, "Characterization and elimination of AC noise in electrocardiograms: A comparison of digital filtering methods," *Comput. Biomed. Res.*, vol. 26, no. 1, pp. 48–67, Feb. 1993.
- [32] J. A. Van Alste and T. S. Schilder, "Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 12, pp. 1052–1060, Dec. 1985.
- [33] D. L. Donoho, "Ideal spacial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [34] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.
- [35] S. Mishra, D. Das, R. Kumar, and P. Sumathi, "A power-line interference canceler based on sliding DFT phase locking scheme for ECG signals," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 132–142, Jan. 2015.
- [36] G. J. J. Warmerdam, R. Vullings, L. Schmitt, J. O. E. H. Van Laar, and J. W. M. Bergmans, "A fixed-lag Kalman smoother to filter power line interference in electrocardiogram recordings," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1852–1861, Aug. 2017.
- [37] R. Kumar, A. Banerjee, B. C. Vemuri, and H. Pfister, "Trainable convolution filters and their application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1423–1436, Jul. 2012.
- [38] V. S. Chouhan and S. S. Mehta, "Total removal of baseline drift from ECG signal," in *Proc. Int. Conf. Computing: Theory Appl. (ICCTA)*, Mar. 2007, pp. 512–515.
- [39] E. Stålberg, J. Chu, V. Bril, S. Nandedkar, S. Stålberg, and M. Ericsson, "Automatic analysis of the EMG interference pattern," *Electroencephalogr. Clin. Neurophysiol.*, vol. 56, pp. 672–681, Dec. 1984.



**Lang Qin** was born in Sichuan, China, in 2000. He is currently pursuing the B.S. degree in communication engineering with the University of Electronic Science and Technology of China, Chengdu, China.

His research interests include natural language processing, image recognition, deep learning, and machine learning.



**Yuntao Xie** was born in Guangdong, China, in 2000. He is currently pursuing the B.S. degree in communication engineering with the University of Electronic Science and Technology of China, Chengdu, China.

His research interests include cloud computing, the Internet of Things, and machine learning.



**Xiangchi Yuan** was born in Anhui, China. He is currently pursuing the B.S. degree with the Glasgow College, University of Electronic Science and Technology of China, Chengdu, China.

His research interests include antinoise for image segmentation and machine learning.



**Xinwen Liu** was born in Tianjin, China, in 2000. She is currently pursuing the B.S. degree in communication engineering with the University of Electronic Science and Technology of China, Chengdu, China.

Her research interests include deep learning and health management.



**Huan Wang** was born in Hunan, China, in 1994. He received the B.S. and M.S. degrees in mechanical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2016 and 2021, respectively. He is currently pursuing the Ph.D. degree in management science and engineering with Tsinghua University, Beijing, China.

He has authored or coauthored more than 20 publication papers (SCI/EI), including eight SCI-Indexed journal papers and one ESI highly cited paper. His research interests include prognostics and health management (PHM) for rotating machinery, biological signal processing [electrocardiogram (ECG), EEG], medical image analysis, deep learning, and machine learning.