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

Automatic 12-Leading Electrocardiogram Classification Network with Deformable Convolution

Yuntao Xie, Lang Qin, Hongcheng Tan, Xinyang Li, Bisen Liu and Huang Wang*

Abstract— **Electrocardiogram (ECG) is an electrical signal that helps monitor the physiology of the heart. A complete ECG record includes 12 leads, each reflecting features from a different angle of the heart. In recent years, various deep learning algorithms, especially convolutional neural networks (CNN), have been applied to detect ECG features. However, the conventional CNN can only extract the local features and cannot extract the data correlation across the leads of ECG. Based on deformable convolution networks (DCN), this article proposes a new neural network structure (DCNet) to detect ECG features. The network architecture consists of four DCN blocks and a classification layer. For the ECG classification task, in a DCN block, the combination of normal convolution and deformable convolution with better effect was testified by the experiments. Based on the feature learning capability of DCN, the architecture can better extract the characteristics between leads. Using the public 12-leading ECG data in CPSC-2018, the diagnostic accuracy of this architecture is the highest, reaching 86.3%, which is superior to other common network architectures with good results in ECG signal classification.**

Clinical relevance—**In this paper, we proposed an effective automatic ECG classification model that can reduce medical staff workload.**

I. INTRODUCTION

Disease-related to the heart is one of the significant causes of death, with approximately 17.9 million people dying from cardiovascular disease in 2016, making up one-third of all humankind deaths [1]. Therefore, heart-related diseases require our attention, and it is significant to detect the disease through diagnosis and take timely and appropriate treatment. Non-invasive electrocardiogram (ECG) records can obtain the physiological condition of the heart. A complete ECG record consists of 12 leads (I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6), responding to characteristics of the heart at different angles [2], which play a significant part in the detection and prevention of heart problem. However, ECG signals are characterized by high noise and complexity [3], which makes it challenging even for cardiologists to identify abnormal ECG signals in people with heart problems [4]. The diagnostic analysis of ECG wave forms through human observation is extremely dependent on doctors' knowledge base and experience, which is time-consuming and challenging to ensure real-time. An effective automatic ECG diagnosis method can ensure real-time diagnosis and reduce the workload of medical staff.

Many automated ECG diagnostic methods based on deep learning have been proposed [5][6][7] in recent years because of the fast improvement of deep learning techniques, including Convolutional Neural Network (CNN) [8], Long Short-Term Memory (LSTM) [9], Encoder-Decoder [10], have been proposed and have made some progress. Chandra et al. [11] proposed an encoder-decoder model based on Kullback-Leibler Divergence to detect anomalies on ECG signals. Mostayed et al. [12] proposed a 12-Leading ECG signal classifier that including two Bi-LSTM layers. Liu et al. [13] proposed a network that combines ECG signal features obtained from a Resnet-based 17-layer one-dimensional CNN with expert features for ECG classification.

Convolutional neural network (CNN) [8], as a model widely used in the field of deep learning, effectively extracts features within the proximity interval in space through the fixed local receptive field, and the sharing of parameters between different neurons and downsampling decreases the number of parameters and decreases the difficulty of training the model [14]. However, conventional CNN has shortcomings in automatic ECG diagnosis. The normal convolution is limited by its fixed convolution kernel structure, which can only extract the local features but cannot extract the data correlation across the leads of ECG.

Deformable convolution breaks up the originally fixed convolution kernel by introducing displacement weights so that it can be not limited to a square sampling area [15], and the sampling points in the convolution kernel can be shifted based on the displacement weights, which can be used to adjust the receptive field automatically and extract the desired features by training the weights [16]. Based on CNN, the introduction of deformable convolution can better handle the correlation between leads. The shape of the deformable convolution's kernel can be adjusted to select the desired features according to the actual situation, so as to extract the characteristics between different leads better.

We design a neural network based on deformable convolution according to the characteristics of ECG data. Our method does not need complex preprocessing of ECG data and has high classification accuracy.

This work was supported by Glasgow College, UESTC. Yuntao Xie and Lang Qin are co-first authors of this article. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board. Corresponding author: Huan Wang (wh.huanwang@gmail.com)

Yuntao Xie, Lang Qin, Hongcheng Tan, Xinyang Li are with the Glasgow College, UESTC, Chengdu, Sichuan 611731, China yuntao.xie@hotmail.com; qinlang51@126.com; tanhc804@163.com; Xiang_Li0502@126.com

Bisen Liu is with the Department of Computational Mathematics and Center for Applied Mathematics, UESTC, Chengdu, Sichuan 611731, China bisen_liu@163.com

Huan Wang is with the Department of Industrial Engineering, Tsinghua University, China wh.huanwang@gmail.com

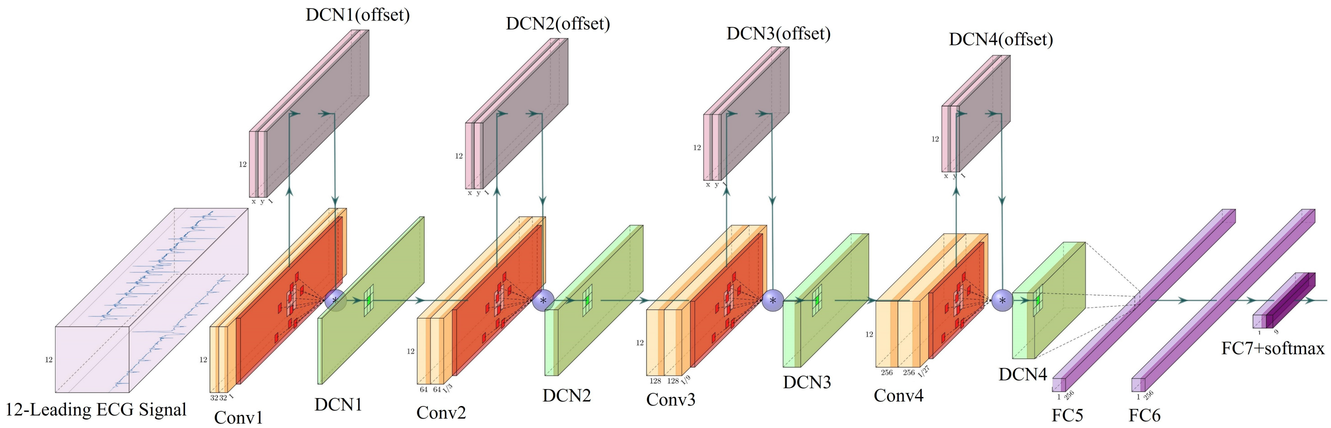

Fig. 1. Diagram of the whole structure of DCNet.

II. METHOLOGY

A. Deformable convolution

The convolution kernel of deformable convolution is not limited to be a fixed square. However, it can change the convolution shape adaptively to obtain a larger receptive field and automatically integrate helpful information within it to better adapt to the characteristics of ECG. The procedure for calculating the deformable convolution is described in detail as follows.

First, from the perspective of the normal convolution, a sliding summation of the data with the convolution kernels at fixed positions gives a feature map incorporating neighborhood information.

$$O_i^n = \sum_{\Delta i \in K} O_{i+\Delta i}^{n-1} \cdot W_{\Delta i}^n \tag{1}$$

where, O_i^n represents the value corresponding to position i of the *nth* layer feature map. $W_{\Delta i}^n$ represents the convolution kernel weight corresponding to the nth layer, for a regular 3×3 convolution $K = \{(-1,-1), (-1,0), \cdots, (0,1), (1,1)\}$.

The deformable convolution adds an adaptively changeable position vector *j* to the original position. Deformation is achieved by optimizing the parameters with the method of back-propagation in a similar way to the normal convolution.

$$O_i^n = f_B(\sum_{\Delta i \in K} O_{i+\Delta i}^{n-1} \cdot W_{\Delta i}^n) \tag{2}$$

B. Network architecture design

We arrange 12 leads of ECG data in parallel rows as the network's input and use one-dimensional convolution to extract the local feature of the ECG signal within single leads. Since the correlation between leads is independent of the order of the leads, the normal convolution cannot handle the correlation between leads well and can only extract the local features within a single lead. Therefore, we use adaptive respective field adjustment, a prominent feature of deformable convolution, to automatically select useful local features. Furthermore, the sampling points of its convolution kernel can adjust the position of the samples with trainable displacement weights, enabling intra-lead and inter-lead sampling to extract correlation between leads and periodic characteristics of the ECG signal.

Our proposed DCNet consists of four DCN blocks and a classification layer, and each DCN block contains two one-dimensional convolutional layers and one deformable convolutional layer, shown in Fig. 1. We use ReLU as an activation function to add non-linearity to the model and append it after each convolutional layer or deformable convolutional layer. The first DCN block takes the 12-leading ECG data as the input of this module, and the remaining DCN blocks take the output of the previous block as the current blocks' input. In each DCN block, a feature map is obtained through two layers of 1×3 convolution and 1×4 maximum pooling, which is then inputted to the deformable convolution layer. The DCN block is used four times to extract the ECG signal feature, and Global Average Pooling (GAP) is performed in the classification layer and input to a Dense layer with softmax function. Finally, the classification possibilities of ECG types are the output. The specific network structure and parameters are shown in Table I, "Convolution" denotes the normal convolution layer, "DConv" denotes the deformable convolution layer.

C. Implementation Details

We use the Pytorch 1.4.0 framework to implementing our code and use a Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.5 and choose 0.0001 as the initial learning rate. We trained all the models for 100 epochs with a batch size is eight and chose the highest accuracy model on the validation set within 100 epochs. The server we used for training has an Intel i9-9900K CPU, 32GB memory, and an Nvidia RTX 2080 GPU with 8GB RAM and runs an Ubuntu 18.04 system with GPU driver version 418.67.

III. EXPERIMENTAL VALIDATION

A. Dataset and Preprocessing

We used the publicly available dataset from China Physiological Signal Challenge 2018 (CPSC-2018) [13] to train, validate and test the effectiveness of our model. The dataset was obtained from 11 different hospitals, some of which were used for competition scoring and therefore not available to the public. In the publicly available data, the sampling rate

TABLE I

THE DETAILS OF THE DCNET

Input	Layer Details		Intput	
	-		12-leading ECG data	

	Layer Details	Output		Layer Details	Output
DCN block 1	[32×1×3 Convolution, stride=1]+Relu	12×7500×32	DCN block 2	[64×1×3 Convolution, stride=1]+Relu	12×1875×64
	[32×1×3 Convolution, stride=1]+Relu	12×7500×32		[64×1×3 Convolution, stride=1]+Relu	12×1875×64
	[1×4 1D Max-pooling]	12×1875×32		[1×4 1D Max-pooling]	12×469×64
	[32×3×3 DConvolution, stride=1]+Relu	12×1875×32		[64×3×3 DConv, stride=1]+Relu	12×469×64
	Layer Details	Output		Layer Details	Output
DCN block 3	[128×1×3 Convolution, stride=1]+Relu	12×469×128	DCN block 4	[256×1×3 Convolution, stride=1]+Relu	12×118×256
	[128×1×3 Convolution, stride=1]+Relu	12×469×128		[256×1×3 Convolution, stride=1]+Relu	12×118×256
	[1×4 1D Max-pooling]	12×118×128		[1×4 1D Max-pooling]	12×30×256
	[128×3×3 DConv, stride=1]+Relu	12×118×128		[256×3×3 DConv, stride=1]+Relu	12×30×256

Classification	Layer Details	Output
	Global Average Pooling Dense+Softmax	Classification possibilities of different ECG types

of the signal is 500 Hz, each record containing 12 leads, and the duration of the records is inconsistent, with the shortest record being 6 seconds and the longest record being 144 seconds. There are 9 different categories in total, which are Normal, AF, I-AVB, LBBB, RBBB, PAC, PVC, STD, and STE. We ignored a small portion of the data that were too exceptional and downsampled the remaining data to 250 Hz to reduce the data volume and increase training efficiency. To adapt to the characteristics of DCN, we also adjusted the sampling time of data to be 30 seconds equally by repetitive padding and finally obtained 5850 ECG data with the length of 7500. The ECG data were randomly divided into 3510 data as the training set to adjust parameters of the model, 1170 data as the validation set to make an initial assessment of the model's capabilities which can also monitor the sample to avoid overfitting, and 1170 data as the test set which does not participate in the training model parameters to assess the generalization capability of the final model. The order of the input ECG data is random when entering the network.

B. Evaluation Metrics

We used accuracy as an evaluation metric to judge the effectiveness of our network, and the accuracy was calculated as shown in Eq (3).

$$\text{Accurary} = \frac{TP}{TP + \sum_{i=1}^{8} FN_i} \tag{3}$$

where TP represents the amount of true-positive samples, and FN represents the quantity of false-negative samples.

C. Comparison of different combinations

TABLE II

COMPARISON OF DIFFERENT COMBINATIONS

Accuracy		Number of DCN block		
		3	4	5
Number of Convolution Layer in each DCN block	1	0.765	0.774	0.790
	2	0.780	**0.863**	0.829
	3	0.829	0.799	0.798

We have tried several different combinations of normal and deformable convolutional networks, and experimentally obtained one that works well for the ECG classification task, which we call the DCN block. DCN block consists of two parts:

- 1) Several normal convolution layers for extracting local features within leads.
- 2) One deformable convolution layer for information integration and cross-lead information extraction.

We experimented with different numbers of DCN blocks stacked together and the effect of the number of normal convolution layers contained in each DCN block on the final network classification results.

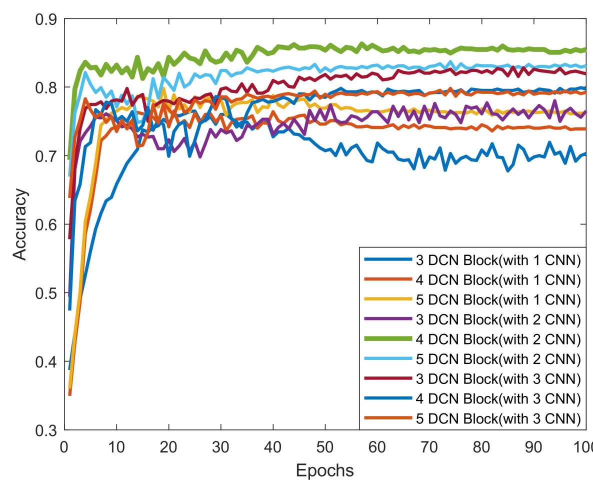

Fig. 2. Diagnostic accuracy for different combination.

From Fig. 2, we can see that the proposed combination method converges fast, which converges and stabilizes at about 50 epochs. We chose the highest accuracy among 100 epochs as the final accuracy result of this combination method and recorded it in Table II. From Table II, it can be observed that the combination method with a stack of four DCN blocks and each DCN block containing two normal convolution layers has the best performance of 86.3% accuracy, so we use this combination method in our DCNet.

TABLE III

COMPARISON WITH DCNET AND OTHER COMMON NETWORK STRUCTURES

Data Source		CPSC-2018					
Method		LSTM[9]	VGG_{16}[17]	$Resnet_{18}$[18]	$Resnet_{50}$[18]	Conventional CNN	DCNet
Accuracy for Different ECG Categories (amount of data)	Normal (856)	0.725	0.765	0.698	0.779	0.658	**0.781**
	AF (910)	0.905	0.899	0.854	0.918	0.880	**0.945**
	I-AVB (650)	0.774	0.849	**0.877**	0.802	0.764	0.874
	LBBB (169)	**0.903**	0.871	0.710	0.613	0.871	0.886
	RBBB (1444)	0.925	0.941	0.900	0.908	0.929	**0.942**
	PAC (429)	0.183	0.244	0.451	0.402	0.488	**0.716**
	PVC (477)	0.718	0.824	0.824	0.682	0.871	**0.937**
	STD (754)	0.838	**0.853**	**0.853**	0.735	0.801	0.774
	STE (161)	0.071	0.643	0.429	0.429	**0.786**	0.667
Weighted average accuracy		0.773	0.813	0.798	0.779	0.801	**0.863**

* The numbers in gray in brackets are the total number of ECG recordings for each category.
* The accuracy in the table are calculated based on the recordings in test set.

D. Comparative experiment

We compared our network with other common network structures that work better in ECG signal classification, namely Resnet [18], VGG [17], LSTM [9], and networks with the same number of layers as our network but each layer is composed of a 3×3 normal convolution. All network structures were compared on the same CPSC-2018 dataset using the same dataset preprocessing and segmentation methods.

In Table III, we compared the performance of DCNet with five counterparts using the accuracy for nine ECG categories. From the table, the proposed model has the highest weighted average accuracy (0.863) among six networks, surpassing the second-ranked method (VGG_{16}) by 0.05. Compared with Conventional CNN and $Resnet_{18}$, 0.062 and 0.065 accuracy improvements are obtained by the proposed method, respectively. In addition, even more gains (0.084 and 0.09) are obtained compared with the other two methods.

Furthermore, the accuracy of DCNet exceeds those of the other five methods in Normal, AF, RBBB, PAC, and PVC. Specifically, for the category of PVC, the proposed method gains 0.219 and 0.255 increases compared with LSTM and $Resnet_{50}$, respectively. In addition, for the category of PAC, LSTM is surpassed by the proposed method by a large gap (0.533). While the existing methods already perform well in the category of RBBB, the proposed method also obtains an increase compared with these methods.

IV. CONCLUSION

We propose a network combining deformable convolution and normal convolution based on ECG features. This structure obtains better classification results on the CPSC-2018 dataset and has an advantage in classification accuracy compared with other networks. This shows that our proposed network can effectively select and integrate ECG features and does not require much preprocessing of the input ECG signal, and can achieve the goal of end-to-end ECG automatic diagnosis. In clinical applications, it can provide a reference for doctors to determine the type of ECG and reduce the workload of medical staff. In the following work, we will set about training and validating the effectiveness of our model on more ECG datasets.

REFERENCES

[1] WHO. (2017) Who fact sheets, 2017. from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] F. N. Wilson, C. E. Kossmann, G. E. Burch *et al.*, "Recommendations for standardization of electrocardiographic and vectorcardiographic leads," *Circulation*, vol. 10, pp. 564–573, 10 1954.

[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 Transactions on Biomedical Engineering*, vol. 37, no. 1, pp. 85–98, 1990.

[4] S. M. Salerno, P. C. Alguire, and H. S. Waxman, "Competency in Interpretation of 12-Lead Electrocardiograms: A Summary and Appraisal of Published Evidence," pp. 751–760, 2003.

[5] L. Qin, Y. Xie, X. Liu, X. Yuan, and H. Wang, "An end-to-end 12-leading electrocardiogram diagnosis system based on deformable convolutional neural network with good antinoise ability," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[6] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," *Knowledge-Based Systems*, vol. 227, p. 107187, 2021.

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

[8] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[10] K. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, pp. 1724–1734.

[11] A. Chandra and R. Kala, "Regularised encoder-decoder architecture for anomaly detection in ECG time signals," 2019, pp. 1–6.

[12] A. Mostayed, J. Luo, X. Shu, and W. Wee, "Classification of 12-Lead ECG signals with bi-directional LSTM network," 2018.

[13] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, and E. N. Y. Kwee, "An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, 2018.

[14] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature extraction using convolution neural networks (CNN) and deep learning," 2018, pp. 2319–2323.

[15] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable convnets V2: More deformable, better results," vol. 2019-June, 2019, pp. 9300–9308.

[16] H. Gao, X. Zhu, S. Lin, and J. Dai, "Deformable kernels: Adapting effective receptive fields for object deformation," 2019.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," vol. 2016-December, 2016, pp. 770–778.