MADNN: A Multi-scale Attention Deep Neural Network for Arrythmia Classification

Ran Duan, Xiaodong He, Zhuoran Ouyang

Edan Diagnostics ltd., Shenzhen, China

Abstract

The morphological features for arrhythmia diagnosis are normally identified in different sizes. Attention-based deep neural networks have been proven to boost meaningful features on different scales, while suppressing the weak features. To boost capability of extracting the ECG features on different scales, we proposed MADNN: a multi-scale attention deep neural network combining kernel-wise and branch-wise attention modules. In the PhysioNet challenge 2020 for arrhythmia classification, MADNN achieved an overall score of 0.446 on the hidden testing-set.

1. Introduction

For diagnosis of arrhythmia, doctors often pay more attention to certain segments of an electrocardiogram (ECG) signal and pay less attention to the others [1]. In the area of deep learning, attention-based deep neural networks (DNNs) have been introduced to imitate the similar process [2].

The attention-based DNNs is proved to be capable of learning from global information and identifying areas to be focused on. Features related to the focused area are assigned more weights during the learning process, and useless information is suppressed [3].

Attention-based DNNs have been introduced with good performances, such as point-wise spatial attention network, squeeze & excitation module (SENet) [4] and selective kernel network (SKNet) [5]. These networks are designed to apply attention-based module in features on single-scale (respectively on channel-wise and branch-wise scales).

Recently, certain DNNs, such as convolutional block attention module (CBAM) [6] and split-attention networks (ResNeSt) [7], explored the potential of combining attention modules on different scales. These networks performed better then single-scale attention modules in the task of image classification.

Combining attention modules on different scales bring the increasing need for computational ability and make the DNN architecture more complicated. The residual modules in Aggregated Residual Transformations for Deep Neural Networks (ResNeXt) [8] reduce the computational ability and the difficulty of hyper-parameter tuning. The main idea of ResNeXt is feeding the inputs to branches containing the residual modules in the same kernel size.

Therefore, we proposed a 1-dimentional CNN for ECG signal classification, combining kernel-wise and branch-wise attention modules. The architecture of splitting branches in our proposed model was modified based on the architecture of ResNeXt.

2. Methods

2.1. Data retrieval

43134 samples were obtained from PhysioNet Challenge 2020 [9] [10]. Each sample contained one ECG signal and one diagnosis. All ECG signals were recorded in the sample rates of 500 Hz or 1000 Hz. The length of each signal varied from 10 seconds to 30 minutes (Figure 1).

Each ECG recording has one or more labels including normality and 111 types of abnormalities (Fig. 2). In our particular task, 27 types of labels were planed for classification and scored according to PhysioNet challenge 2020.

2.2. Data preprocessing

The DNN contains full-connected layers which requires a fixed size of inputs. To fixed the size of inputs for our model, we resampled the signals to a fixed rate of 500 Hz, and padded the signals to a length of 30000.

First, a fast Fourier transform was applied to resample all signals to 500hz. All signals were then padded to 30s. For the signal shorter than 60s, it was copied end to end to the length of 30s. On the contrary, the data longer than 30s

was truncated to the length of 30s from the end of the signal.

At the end of data preprocessing, all samples were randomly shuffled and split into a training-set and a testing-set (respectively 90% and 10% of the total).

2.3. Data augmentation

To increase the randomness of the data and reduce overfitting. We randomly cropped each padded signal (Figure 1). 60 sampling points were reserved at the beginning of the signal as the starting interval for random clipping, and the last 60 sampling points of the signal as the ending interval for random clipping.

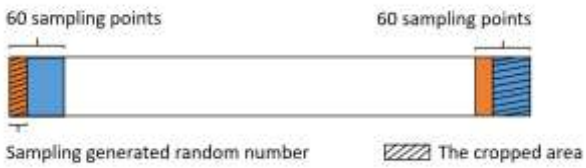

Figure 1. The process of random cropping.

At the beginning of preparing each input batch for training, a random integer was generated in a uniform distributed interval from 0 to 60. We cropped the area before the random number of sampling points in the starting interval, and cropped the ending interval after the random number of sampling points.

The classes of signals, with a smaller number of samples, contributes less in weights (weights of neurons in our DNN). Therefore, focal loss [11] was adopted as our loss function in the process of training. It helped in balancing the contribution of classes in different numbers of samples.

Also, an extraction weight for each class of data was generated according to Formula-1 for balancing the importance of different classes (W_i denoted the extraction weight of class I; μ denoted the number of samples in total; μ_i denoted the mean number of samples labeled as class i). Before feeding each sample to our model, a random float ranged from 0 to 1 was generated. If the random float is less than the extraction weight of the corresponding class of the sample, the sample was skipped in training.

$$W_i = \frac{\mu}{2\mu_i} \qquad \text{(Formula 1)}$$

2.4. DNN architecture

Kernel-wise attention modules (Figure 2-a) from SE modules and branch-wise attention modules (Figure 2-b) from SKNet were combined and modified as the multi-scale attention modules (Figure 2-c). The essence of our proposed multi-scale attention modules is to attached a Kernel-wise attention module after each branch-wise attention module.

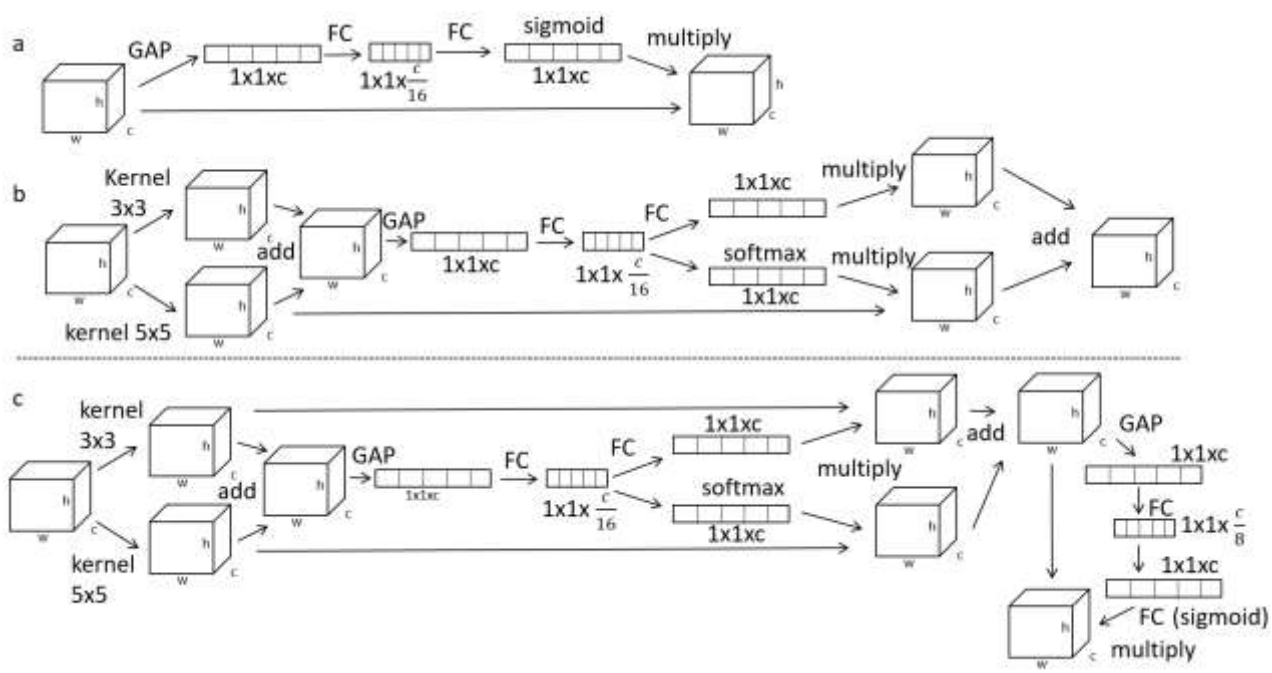

Figure 2. Comparison of SE Module (a), SK module (b) and our proposed multi-scale attention module (c). FC denoted full-connected layer.

A few modifications were applied in designing the multi-scale attention modules. To extract features from signals in 1-dimention, 2-dimentional CNNs in SE modules and SKNet were converted to 1-dimentional CNNs correspondingly. Unlike the traditional SK modules, the input feature maps of our modules were fed to different branches of CNNs in the same kernel size. This idea was inspired by ResNeXt, and proven to be able to improve the accuracy and reduce the effort in hyper-parameters tuning.

The overall architecture of MADNN is a network connected a stem module in ResNeXt, four of our modified multi-scales attention modules, a global averaged pooling layer and a full-connected output layer (Figure 3).

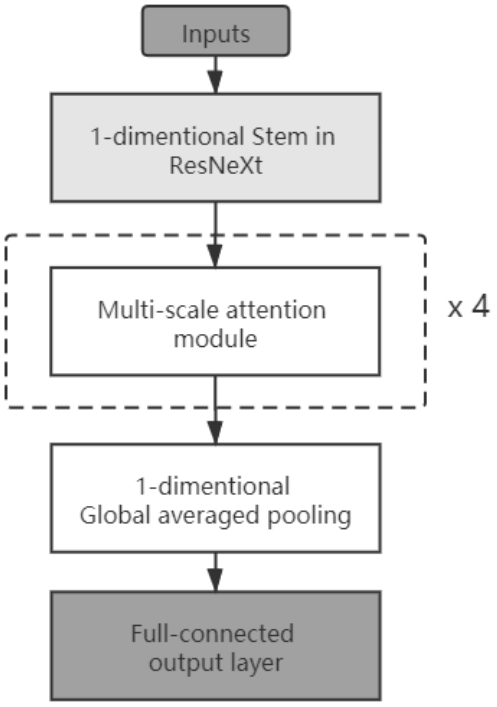

Figure 3. The overall architecture of MADNN.

3. Results

MADNN achieved an overall score of 0.446 in the hidden testing-set in PhysioNet Challenge 2020 for arrhythmia classification. The overall score was calculated according to a new scoring metric that awards partial credit to misdiagnoses as cardiologists [9].

References

[1] Vecht R, Gatzoulis MA, Peters N. ECG diagnosis in clinical practice. Springer Science & Business Media; 2009 Apr 21.

[2] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. InAdvances in neural information processing systems 2017 (pp. 5998-6008).

[3] Vig J. A multiscale visualization of attention in the transformer model. arXiv preprint arXiv:1906.05714. 2019 Jun 12.

[4] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 7132-7141).

[5] Li X, Wang W, Hu X, Yang J. Selective kernel networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2019 (pp. 510-519).

[6] Woo S, Park J, Lee JY, So Kweon I. Cbam: Convolutional block attention module. InProceedings of the European conference on computer vision (ECCV) 2018 (pp. 3-19).

[7] Zhang H, Wu C, Zhang Z, Zhu Y, Zhang Z, Lin H, Sun Y, He T, Mueller J, Manmatha R, Li M. Resnest: Split-attention networks. arXiv preprint arXiv:2004.08955. 2020 Apr 19.

[8] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1492-1500).

[9] Perez Alday E A, Gu A, Shah A, Liu C, Sharma A, Seyedi S, Bahrami Rad A, Reyna M, Clifford G. Classification of 12-lead ECGs: the PhysioNet - Computing in Cardiology Challenge 2020 (version 1.0.1). PhysioNet. 2020.

[10] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov PC, Mark R, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[11] Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. InProceedings of the IEEE international conference on computer vision 2017 (pp. 2980-2988).

Address for correspondence:

Ran Duan.
No.15 Jinhui Rd.,
Shenzhen, 518122,
P.R.China
413677671@qq.com