

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

Ran Duan, Xiaodong He, Zhuoran Ouyang

Edan Diagnostics ltd., Shenzhen, China

Abstract

The morphological ECG features for arrhythmia diagnosis are usually identified and combined on different scales. For example, morphological ECG features can be identified on the scale of length or amplitude of QRS waves. Professionals can then make a diagnosis based on the combination of these identified features. Attention-based deep neural networks have been proved to boost meaningful features on different scales and suppress weak features. To boost and combine ECG features on different scales for arrhythmia classification, we proposed MADNN: a multi-scale attention deep neural network for arrhythmia classification. Our proposed network was designed combining kernel-wise and branch-wise attention modules based on a backbone of 1-dimensional convolutional neural networks. MADNN with properly tuned hyper-parameters was tested for arrhythmia classification in the PhysioNet/Computing in Cardiology Challenge 2020. In this challenge, MADNN officially achieved a validation score of 0.446, and a full test set score of 0.236. Our team named Minibus ranked the 18th out of 41 teams.

1. Introduction

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

Attention-based DNNs are capable of learning from global information and identifying features to be focused on. Essential features are assigned more weights in training, and useless information is suppressed [3].

Many attention-based DNNs have been validated with good performances, such as point-wise spatial attention network [4] and squeeze & excitation module (SENet) [5]. They were designed to apply the attention-based modules to features on a single-scale (respectively in spatial-wise and kernel-wise).

State-of-the-art combined-attention DNNs, such as selective kernel network (SKNet) [6], convolutional block

attention module (CBAM) [7] and split-attention networks (ResNeSt) [8], explored the potential of combining attention modules on different scales. These networks performed better than the reported individual single-scale attention networks.

Combined-attention modules bring extra model complexity and increase the need for computational ability in model inference. The residual modules in Aggregated Residual Transformations for Deep Neural Networks (ResNeXt) [9] was designed to reduce the computational ability and simplify hyper-parameter tuning. ResNeXt was proved to be a robust backbone network in SKNet.

Although applying a ResNeXt backbone simplify the network complexity, the kernel-wise attention modules in SENet were simplified in the ResNeXt-based SKNet. Unlike SENet, each branch of kernels in a SKNet attention module was assigned to the same attention weight.

Another multi-scale attention network, ResNeSt, was recently proved with better performance than SKNet in image classification. In ResNeSt, the outputs of several attentive modules were concatenated in kernel-wise. Attention weights for each kernel were permuted across different attentive modules, and this added more diversity to the overall attention weights.

Compared with SKNet, it is difficult to design the architecture of ResNeSt and tune the hyper-parameters, especially for a task that is different from image classification (such as ECG signals classification). This aroused our attention to design a novel DNN that each type of attention module can be optimized separately with diverse attention weights to the outputs of different convolutional kernels.

Thus, we proposed MADNN: a network sequentially combines kernel-wise attention modules in SENet and kernel-selective modules in SKNet. Also, we modified and tested the proposed model on the task of ECG signal classification.

2. Methods

2.1. Data retrieval

43134 samples were obtained from the PhysioNet

Challenge 2020 [10] [11]. Each sample contained one ECG signal and one or more corresponding diagnoses. All ECGs were recorded at the sample rates of 500 Hz or 1000 Hz. The length of each ECG signal varied from 10 seconds to 30 minutes.

Each ECG record contained one or more diagnoses, including normality and 111 types of abnormalities. 27 types of diagnoses were labeled in our study because they were scored in the PhysioNet Challenge 2020.

2.2. Data preprocessing

The fully-connected layers in our DNN required a fixed-size input. The input signals were resampled to a fixed sample rate at 500 Hz and padded to a fixed length of 30000 sample points.

First, fast Fourier transform was applied to resample all signals to 500hz. All signals were then padded to 60 seconds. The signals which was shorter than 60 seconds were copied end to end to the length of 60 seconds. On the contrary, the signals longer than 60 seconds were truncated to the length of 60 seconds from the end.

All samples were then randomly shuffled and split into a training set, a validation set, and a testing set (respectively 80%, 10%, and 10% of the total).

2.3. Data augmentation

To increase the randomness and reduce overfitting, we randomly cropped each padded signal (Figure 1) after

signal padding. In detail, 60 sample points were reserved at the beginning of a signal as the starting interval, and the last 60 sample points of the signal were reserved as the ending interval for random cropping.

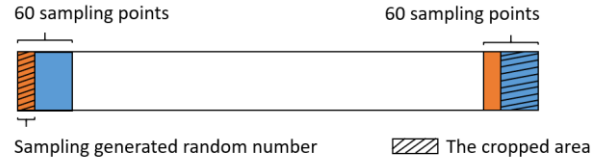


Figure 1. The process of random cropping.

When preparing each batch of inputs, a uniformly distributed random variable was generated ranged from 0 to 60. We cropped the random variable of sample points from the beginning of the starting interval. Similarly, 60 minus the random variable of sample points were cropped from the end of the ending interval.

To avoid the imbalance contribution of different labeled classes (imbalance update of weights in DNN neurons), focal loss [12] was adopted as our loss function in the process of model training.

Also, a balance factor for each label group was calculated according to Formula 1 (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 a sample to our model, a random float ranged from 0 to 1 was generated. If the generated float was smaller than the corresponding balance

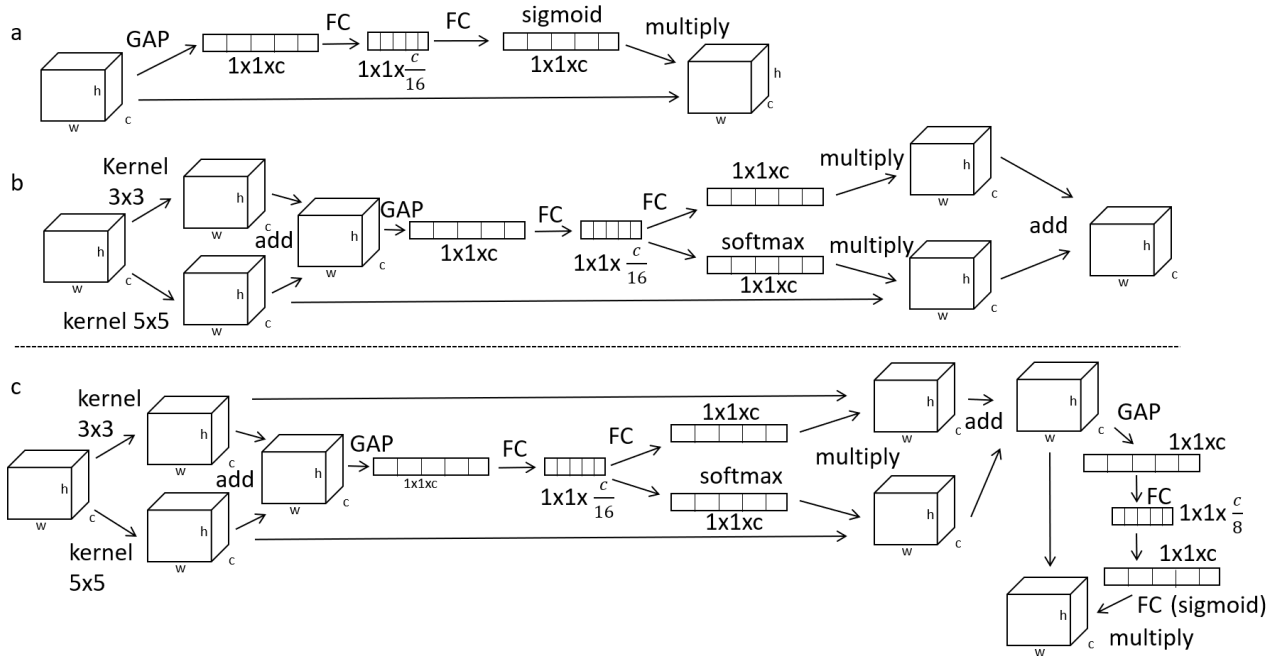


Figure 2. Comparison of SE Module (a), SK module (b), and our proposed multi-scale attention module (c). FC denoted a full-connected layer.

factor, the sample would be skipped for training.

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

2.4. DNN architecture

A customized multi-scale attention module was proposed in our study to combine features on different scales. A kernel-wise attention module (Figure 2-a) from SE modules and a branch-wise attention module (Figure 2-b) from SKNet were combined and modified into a multi-scale attention module (Figure 2-c). In each multi-scale attention module, a kernel-wise attention module was attached to the end of a branch-wise attention module.

For the purpose of extracting features from 1-dimensional signals, the original 2-dimensional CNNs in SENet and SKNet were modified to 1-dimensional CNNs correspondingly. Inspired by ResNeXt, each branch of convolutional layers in the proposed multi-scale attention module shared the same kernel size. The kernel size was changed to 3 from 7 in the stem convolutional layers in comparison to ResNeXt. Applying convolutional layers with large kernel size will suppress the features of high-frequency in ECG signals.

Overall, MADNN consisted of a stem module in ResNeXt, four modified multi-scale attention modules attached one by another, a global averaged pooling layer, and a fully-connected output layer (Figure 3).

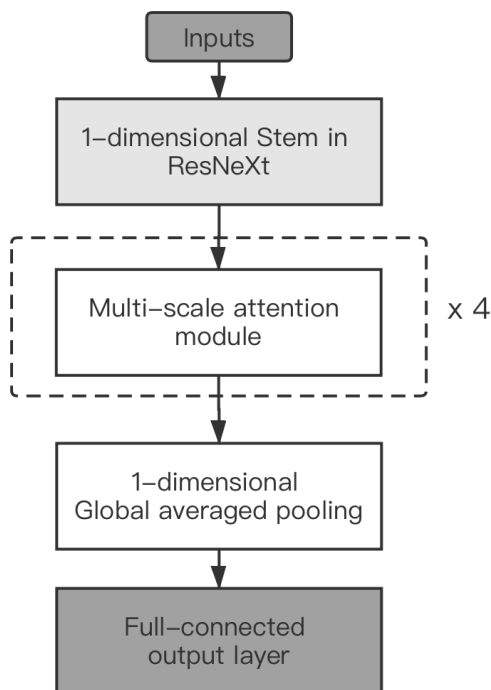


Figure 3. The overall architecture of MADNN.

2.5. Training Methodology

Model weights were updated based on the binary cross-entropy loss for each label. The model was trained with an Adam [13] optimizer of an initial learning rate of 0.0001, beta1 of 0.9, beta2 of 0.999, and epsilon of 1e-08. The learning rate was automatically multiplied with a reduction factor of 0.5 when the validation loss has stopped improving for 5 epochs.

A dropout layer with a 0.1 dropout rate was attached to each multi-scale attention module to avoid overfitting. When training each batch of data, 10% of the weights in its previous convolutional layer were randomly selected not to be updated. The dropout rate in our study was set to a smaller value compared with the dropout rate of 0.2 in ResNeSt. This is mainly because a moderate strategy was preferred for medical image classification.

The training was stopped after the validation loss stopped optimizing for 20 epochs. An optimal model was carefully selected from models in the 20 epochs based on the validation result.

2.6. Ensemble learning

We also implemented an Xgboost [14] classifier to aggregating the predictions from an ensemble of the optimized MADNN, an Xception model [15], and a customized VGG-like model [16] (Figure 4). All individual models were trained on the same challenge data.

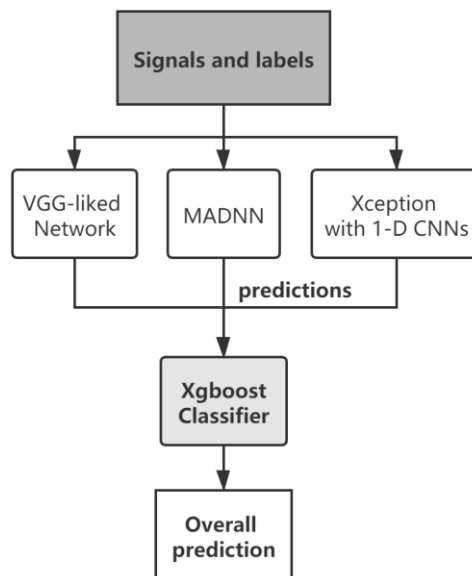


Figure 4. The workflow of ensemble learning.

In specific, the Xgboost classifier was designed under a random search [17] of several hyper-parameters: the

number of estimators ranged from 80 to 200 in 4 trails; the learning rate ranged from 0.001 to 2 in 20 trails; the subsample ranged from 0.6 to 0.9 in 20 trails; the feature sample rate by column ranged from 0.5 to 0.98 in 10 trails;

3. Results and discussion

MADNN without ensemble learning was tested on the hidden datasets in the PhysioNet/Computing in Cardiology Challenge 2020. Scores were calculated according to a new scoring metric that awards partial credit to misdiagnoses as cardiologists [10]. The MADNN officially achieved a validation score of 0.446. It was also tested on a full hidden test set and officially achieved a score of 0.236. Our official score ranked 18th out of the scores of 41 teams in this challenge (our team named Minibus in the challenge).

The ensemble classifier was tested on the local testing set, which is split by cross-validation. The training score was 0.6 higher than the score of independent MADNN. The ensemble classifier was not submitted and tested on the hidden testing set because its inference speed was relatively low.

In this study, we addressed the potential of a simple solution to combining attention modules on different scales. The model performed good in the challenge, but more tests could be implemented in future works to compare the performance of MADNN with other similar DNNs.

Also, label smoothing and stacking ensemble are potential optimization methods. Regarding the limited computational ability of ECG machines, an extension for MADNN could be simplifying the current model.

References

- [1] Vecht R, Gatzoulis MA, Peters N, "ECG diagnosis in clinical practice," *Springer Science & Business Media*, 2009, pp 67-111.
- [2] Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, Rueckert D, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, Apr, 2019, pp. 197-207.
- [3] Vig JA, "multiscale visualization of attention in the transformer model," *arXiv preprint*, arXiv:1906.05714, Jun, 2019.
- [4] Zhao H, Zhang Y, Liu S, Shi J, Change Loy C, Lin D, Jia J, "Psanet: Point-wise spatial attention network for scene parsing," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 267-283.
- [5] Hu J, Shen L, Sun G, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132-7141.
- [6] Li X, Wang W, Hu X, Yang J, "Selective kernel networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 510-519.
- [7] Woo S, Park J, Lee JY, So Kweon I. "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3-19.
- [8] 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, Apr, 2020.
- [9] Xie S, Girshick R, Dollár P, Tu Z, He K, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1492-1500.
- [10] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, Liu F, Rad BA, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD, Reyna MA, "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," *Physiol. Meas.*, 2020 (Under Review).
- [11] 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.
- [12] Lin TY, Goyal P, Girshick R, He K, Dollár P, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980-2988.
- [13] Kingma DP, Ba J, "Adam: A method for stochastic optimization," *arXiv preprint*, arXiv:1412.6980, Dec, 2014.
- [14] Chen T, Guestrin C, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug, 2016, pp. 785-794.
- [15] Chollet F, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251-1258.
- [16] Simonyan K, Zisserman A, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, arXiv:1409.1556, Sep, 2014.
- [17] Bergstra J, Bengio Y, "Random search for hyper-parameter optimization," *The Journal of Machine Learning Research*, Feb, 2012, pp. 281-305.

Address for correspondence:

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