Towards Heart Sound Classification Without Segmentation Using Convolutional Neural Network

Wenjie Zhang[1], Jiqing Han[1]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

Abstract

Heart sound classification is an effective and convenient method for the preliminary diagnosis of heart diseases, it provides physiology and pathology information to determine whether further expert diagnosis is needed. However, one of the most difficult problems for heart sound classification is heart sound segmentation. In this study, we proposed a method for heart sound classification without segmentation using convolutional neural network (CNN). In the proposed method, the heart cycles with different start positions are intercepted from the heart sound signals during the training phase. Then the spectrograms of the intercepted heart cycles are scaled to a fixed size and input to the designed CNN architecture. Thus, the trained CNN is able to generate features of different start positions in the testing phase. Therefore, heart sound segmentation is not necessary for prediction in the proposed method. At last, the classification task is completed by support vector machine (SVM). Moreover, the proposed method is evaluated on two public datasets offered by the PASCAL classifying heart sounds challenge. The results show that the proposed method is competitive compared with the methods using heart sound segmentation information, especially that the performance improvement is not obvious when the segmentation information is used in the testing phase of the proposed method.

1.   Introduction

Heart diseases have long been one of the main causes of human death worldwide.It is of great value to extract the physiology and pathology information of the heart in the early stage which would provide useful information to determine whether further expert diagnosis is needed. Heart sound classification is one of the commonly used method for heart information extraction.It has been widely applied to automatic heart disease diagnosis. And it usually contains three steps: segmentation, feature extraction and classification.

The aim of segmentation is to detect the fundamental heart sounds (FHS) which are the important physical characteristics of the heart, including S1 and S2. Since feature extraction which greatly affects the classification performance usually is based on the heart sound segmentation results, heart sound segmentation plays an important role in the automatic heart sound classification [1]. However, due to the non stationary of the heart sound signal and the affection of the background noise, including murmurs and other artificial noise in the real world, heart sound segmentation is a difficult problem.

In fact, the primary task of heart sound classification can be accomplished without heart sound segmentation. The goal of the primary task is only to detect the presence of a disorder in the heart sound rather than to further identify it which requires to locate the precise position of the heart disease in the signal. In recent research, some methods towards automatic heart sound classification without segmentation have been proposed, such as the method of Yuenyong *et al.* [2] and our previous method [3]. However, the spectrogram information of the heart sound signals is not fully utilized in these methods [2,3]. In order to extract more discriminative features from the spectrogram information, CNN is used in this study since it has been applied with great success to recognizing visual images [4].

In this study, we proposed a novel method for heart sound classification without segmentation using CNN. In the proposed method, the heart cycles with different start positions are intercepted from the heart sound signals during the training phase. Then the spectrograms of the intercepted heart cycles are scaled to a fixed size and input to the designed CNN architecture. Thus, the trained CNN is able to generate features of different start positions in the testing phase. Therefore, heart sound segmentation is not necessary for prediction in the proposed method. We also note that Pylypenko proposed the similar idea [5]. However, during the training phase, the start positions are not balanced and the extracted spectrograms are not from equal heart cycles in the method of Pylypenko [5]. As a result, the designed CNN of Pylypenko is easier to overfitting. Finally, the performance is evaluated by SVM and the results show that the proposed method is competitive.

2. Methods

2.1. Data collection

The datasets used in this paper, including Dataset-A and Dataset-B [6], are collected from the PASCAL classifying heart sounds challenge competition. Dataset-A is collected by volunteers using iStethoscope and is grouped into four categories: *Normal*, *Murmur*, *Extra Heart Sound* and *Artifact* [6]. Dataset-B is collected at the Real Hospital Portugues using a Littmann Model 3100 electronic stethoscope with a 4000 Hz sampling frequency and is grouped into three categories: *Normal*, *Murmur* and *Extrasystole* [6]. The number of samples used in training and testing dataset is shown in Table 1.

Table 1. The number of samples in the training and testing dataset

Dataset	Category	Training	Testing
Dataset-A	*Normal*	31	14
	Murmur	34	14
	Extra Heart Sound	19	8
	Artifact	40	16
Dataset-B	*Normal*	200	136
	Murmur	66	39
	Extrasystole	46	20

2.2. Preprocessing

Before further processing, the heart sound signal is first pre-processed. Since the collected heart sound signal is often contaminated with high frequency noise, the heart sound signal is resampled to 2000 Hz. Additionally, the resampled signal is filtered with a band-pass (20-950 Hz), 6th-order Butterworth filter to further eliminate the noise. Next the heart sound signal is normalized to a fixed scale of [-1 1].

2.3. Heart cycle detection and interception

In order to have a fair comparison between different heart sound samples in the training phase, the heart cycles are detected based on the average heart cycle duration estimation and S1 alignment methods [7]. Then the detected heart cycles with different start positions are intercepted from the heart sound signals. Suppose the detected heart cycles are denoted by $\{h_s|s = 1, 2, \cdots, S\}$, the corresponding heart cycle durations are $\{d_s|s = 1, 2, \cdots, S\}$ and each heart cycle with P start positions is intercepted from the heart sound signal. The intercepted heart cycles with different start positions from each h_s have the same duration d_s. And the intercepted $P \cdot S$ heart cycles are denoted by $\{h_s^p|s = 1, 2, \cdots, S; p = 1, 2, \cdots, P\}$.

2.4. Spectrogram extraction and scaling

In order to extract more discriminative features, the heart sounds are represented in the transformed time-frequency domain by spectrograms. The short time Fourier transform is used to extract the spectrograms of heart sound signals in the time-frequency domain in this study since it is easy to implement and scale. Moreover, the spectrograms of the intercepted heart cycles usually are not the same size due to the variation of the heart sound signals. To have a fair comparison, the spectrograms of the heart cycles h_s^p are scaled to a fixed size using the bilinear interpolation method [8]. The scaling size is 30×20 which is the same as the one in our previous research [8].

2.5. Convolutional neural networks

The scaled spectrograms are put into the CNN to recognize different heart sound samples with different start positions. A standard CNN architecture is used in this research, consisting of four convolutional layers, each two connected convolutional layers followed by a max-pooling layer, followed by two fully connected layers before final classification. To prevent the CNN from overfitting, dropout technique [9] is added to the designed CNN architecture. The CNN architecture for heart sound classification is depicted in Figure 1. It can be seen that the heart cycles with different start positions are intercepted from the heart sound signal. Then the spectrograms are extracted and scaled to a fixed size. Next, the scaled spectrograms are put into the CNN. Thus the designed CNN architecture is able to extract features from different heart sound samples with different start positions. Not that the number O of the output nodes is determined by the number of start positions P and heart sound categories C. It is obtained by,

$$O = P \cdot C \tag{1}$$

Moreover, the bottleneck features from CNN contain inherently discriminative information [10], it is used as the classifier input in this study.

In the designed CNN architecture, the first and second convolutional layers learn 16 3×3 kernels individually, using same padding. This is followed by applying a 2×2 max-pooling filter. The third and fourth convolutional layers learn 32 3×3 kernels individually, once again using same padding. Then following a 2×2 max-pooling filter. In the above two max-pooling filter, the strides both are (2,2). Furthermore, a flattening operation is applied to the max-pooling filter which unrolls the 32 7×5 feature maps into a single dimension vector of size 1120. Then the flattened features are fully connected by 128 hidden units. At last, the hidden units are connected with the output nodes.

The categorical cross entropy loss function is utilized to train the designed CNN architecture. Then the bottleneck
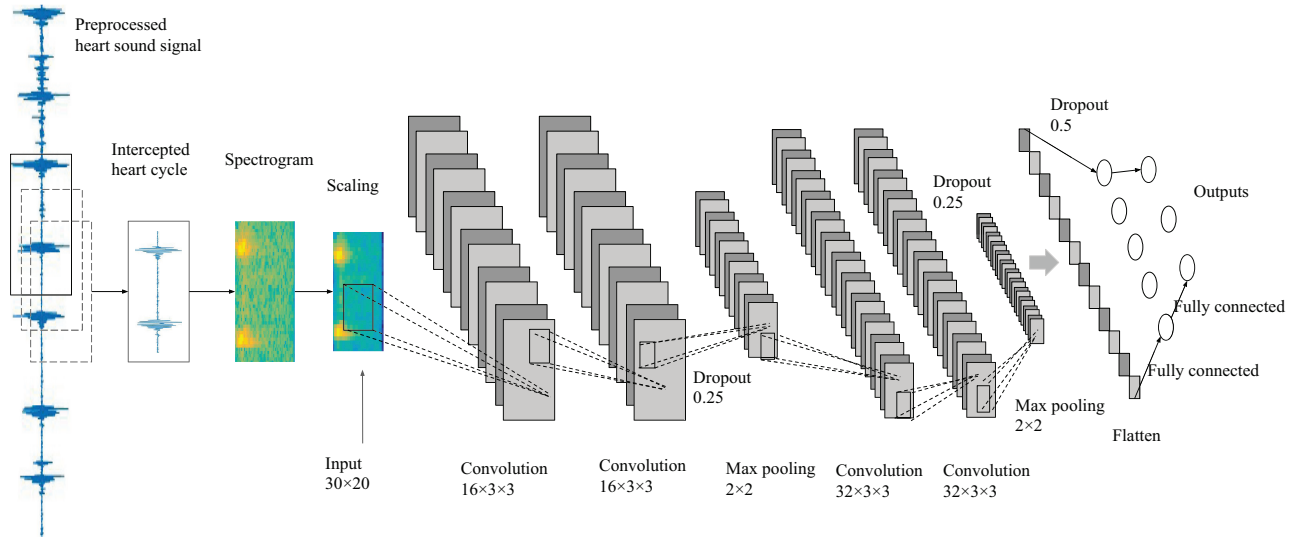
Figure 1. Convolutional neural network architecture for heart sound classification.

features in the training dataset are extracted and used as the classifier input. In order to predict the physiology and pathology information of the heart sound signal, it is necessary to extract features in the testing dataset. Since the heart cycles with different start positions have been fused to the CNN during the training phase, the features of the heart cycles at any start positions are able to be extracted in the testing dataset. Therefore, it is not necessary to perform heart sound segmentation. To have a fair comparison, the average heart cycle duration is estimated using the average magnitude difference function [7]. Then equal estimated average heart cycles are intercepted from the heart sound signal and put into the trained CNN in the testing dataset. The generated bottleneck features are used as the extracted features of the testing sample.

3.   Results and discussion

Using the Dataset-A and Dataset-B, the proposed method is compared with the method of Yuenyong et al. [2] which is abbreviated as ASHA, our preivous method SVM DM [3] and the method of Pylypenko [5] which is abbreviated as PCNN. For convience, the proposed method is represented by WSCNN. Moreover, to further evaluate the proposed method, the features extracted with segmentation in the testing phase are utilized and the method is represented by SCNN.

In the SVM classifier, the slack variable of the radial basis kernel function is set as 0.01 and the regularization parameter is 150. The evaluation criteria used in this paper are the precision of each heart sound category in the test sets and the evaluation of the problem heart sound categories. More detailed evaluation criteria definition is de-

scribed in [8].

The results of Dataset-A are shown in Table 2. It can be seen that the highest precision of *Murmur* and *Artifact* categories are obtained by our method WSCNN. Although the highest precision of *Normal*, *Extra Heart Sound* categories are achieved by SVM MD method. WSCNN obtains the best normalized precision performance. Moreover, WSCNN method also achieves the highest score for Youden index of *Artifact*. Overall, the proposed method is competitive. The results of Dataset-B are shown in Table 3. It can be seen that the highest precision of *Normal* category is obtained by our method WSCNN. The highest score for Youden index of *Artifact* is also obtained by WSCNN method. Therefore, the proposed method has competitive performance for detecting the problem heart sound categories and *Normal* heart sound category.

Note that, the performance of the method WSCNN shows no obvious degradation compared with method SCNN. Therefore, it is not necessary for segmentation during the testing phase in the proposed method.

4.   Conclusion

In this study, a novel method for heart sound classification without segmentation using CNN is proposed. With the advantage of the designed CNN architecture, the features of the heart cycles with different start positions are fused in the network. Thus the segmentation is not necessary to perform in the testing phase. Finally, the proposed method is evaluated on the public datasets from the PASCAL classifying heart sounds challenge. The experiments show that our method is competitive.

Table 2. Results on Dataset-A

Evaluation criteria	ASHA	SVM DM	PCNN	SCNN	WSCNN
Precision of *Normal*	0.34	**0.62**	0.50	0.55	0.61
Precision of *Murmur*	0.82	**0.91**	0.55	**0.91**	**0.91**
Precision of *Extra Heart Sound*	0.00	**1.00**	0.75	0.50	0.50
Precision of *Artifact*	0.80	0.64	0.67	**0.94**	**0.94**
Artifact sensitivity	0.75	**1.00**	**1.00**	**1.00**	**1.00**
Artifact specificity	0.50	0.58	0.44	0.64	**0.67**
Youden index of *Artifact*	0.25	0.58	0.44	0.64	**0.67**
F-score	0.56	**0.74**	0.59	0.66	0.67
Total precision	1.96	**3.17**	2.47	2.90	2.96
Normalized precision	0.56	0.76	0.60	0.76	**0.77**

Table 3. Results on Dataset-B

Evaluation criteria	ASHA	SVM DM	PCNN	SCNN	WSCNN
Precision of *Normal*	0.74	0.77	0.72	0.78	**0.81**
Precision of *Murmur*	0.52	**0.76**	0.41	0.63	0.67
Precision of *Extrasystole*	0.00	**0.50**	0.12	0.21	0.14
Heart problem sensitivity	0.25	0.34	0.35	0.46	**0.51**
Heart problem specificity	0.91	**0.95**	0.67	0.82	0.80
Youden index of heart problem	0.16	0.29	0.03	0.27	**0.31**
Discriminant power	0.30	**0.54**	0.03	0.32	0.34
Total precision	1.26	**2.03**	1.25	1.62	1.62
Normalized precision	0.62	**0.74**	0.60	0.69	0.71

5. Acknowledgements

References

[1] Springer DB, Tarassenko L, Clifford GD. Logistic regression-hsmm-based heart sound segmentation. IEEE Transactions on Biomedical Engineering 2016;63(4):822–832.

[2] Yuenyong S, Nishihara A, Kongprawechnon W, Tungpimolrut K. A framework for automatic heart sound analysis without segmentation. Biomedical Engineering Online 2011;10(1):1–23.

[3] Deng SW, Han JQ. Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. Future Generation Computer Systems 2016;60:13–21.

[4] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems. 2012; 1097–1105.

[5] Pylypenko K. Heart sounds classification using convolutional neural networks. In International Conference on Electronics and Nanotechnology. 2016; 232.

[6] Bentley P, Nordehn G, Coimbra M, Mannor S. The pascal classifying heart sounds challenge. See http://www.peterjbentley.com/heartchallenge/index.html, 2011.

[7] Zhang W, Han J, Deng S. Heart sound classification based on scaled spectrogram and partial least squares regression. Biomedical Signal Processing and Control 2017;32:20–28.

[8] Zhang W, Han J, Deng S. Heart sound classification based on scaled spectrogram and tensor decomposition. Expert Systems with Applications 2017;84:220–231.

[9] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 2014;15(1):1929–1958.

[10] Su R, Liu X, Wang L. Convolutional neural network bottleneck features for bi-directional generalized variable parameter hmms. In International Conference on Information and Automation. IEEE, 2016; 1126–1131.

Address for correspondence:

Jiqing Han
School of Computer Science and Technology / Harbin Institute of Technology
92 Xidazhi St /Nangang, Harbin / China
jqhan@hit.edu.cn