



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331, CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目: 深度卷积神经网络在心音分类方法中的应用  
作者: 陈伟, 孙强, 齐月月, 徐晨  
网络首发日期: 2020-10-22  
引用格式: 陈伟, 孙强, 齐月月, 徐晨. 深度卷积神经网络在心音分类方法中的应用. 计算机工程与应用.  
<https://kns.cnki.net/kcms/detail/11.2127.TP.20201021.1741.016.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 深度卷积神经网络在心音分类方法中的应用

陈伟<sup>1,2</sup>, 孙强<sup>2</sup>, 齐月月<sup>2</sup>, 徐晨<sup>2</sup>

1. 南通大学医学院(护理学院), 江苏南通 226001
2. 南通大学信息科学技术学院, 江苏南通 226001

**摘要:**通过分析心音信号,对心脏早期的病理状态进行确诊具有重要的意义。本文提出了一种基于深度卷积神经网络的心音分类方法。首先,将心音信号转化成具有时频特性的梅尔频谱图(Mel-frequency spectral coefficients, MFSC)特征图,将其作为深度卷积神经网络模型的输入;然后,利用深度卷积神经网络对MFSC特征图进行训练,引入中心损失函数建立最优的深度神经网络模型;测试阶段,先将心音信号转换成多张二维MFSC特征图,然后利用训练好的深度学习模型对其进行分类,最后利用多数表决原则判断心音信号的类别。针对人工标注的训练样本有限,导致模型训练正确率不高的问题,本文以心音的二维MFSC特征图为对象分别从时间域和频率域进行随机屏蔽处理进而扩充训练样本。实验结果表明,本文方法在PASCAL心音数据集上进行测试对正常、杂音、早搏三种心音进行分类其性能明显的优于现有最好的方法。

**关键词:**心音分类;深度卷积神经网络;数据扩充

文献标志码:A 中图分类号:TP391 doi: 10.3778/j.issn.1002-8331.2005-0037

陈伟,孙强,齐月月,等.深度卷积神经网络在心音分类方法中的应用.计算机工程与应用

CHEN Wei, SUN Qiang, QI Yueyue, et al. Deep Convolutional Neural Networks for Heart Sound Classification. Computer Engineering and Applications

## Deep Convolutional Neural Networks for Heart Sound Classification

CHEN Wei<sup>1,2</sup>, SUN Qiang<sup>2</sup>, QI Yueyue<sup>2</sup>, XU Chen<sup>2</sup>

1. Medical School of Nantong University, Nantong University, Nantong, Jiangsu 226001, China
2. School of Information Science and Technology, Nantong University, Nantong, Jiangsu 226001, China

**Abstract:** It is of great significance to diagnose the early pathological state of the heart by analyzing the heart sound signals. This paper presents a heart sound classification method based on Deep Convolution Neural Network (DCNN). Firstly, the heart sound signal is transformed into Mel feature maps with time-frequency characteristics, which are used as the input of the DCNN model. Then, the DCNN model is used to train the Mel feature maps, and the center loss function is introduced to establish the optimal deep learning model. In the testing stage, the heart sound signal is first converted into several two-dimensional Mel feature maps. Then, the feature maps are classified by the pre-trained deep learning model. Finally, the classification of heart sound signal is judged by the principle of majority voting. Due to the limited number of labeled samples, the accuracy of model is not high. In this paper, two-dimensional Mel feature maps of heart sound are randomly shielded in time domain and frequency domain in order to augment the training datasets. The experimental results show that the performance of this method is better than the state-of-the-art methods in the PASCAL heart sound datasets, which aims to classify Normal, Murmur and Extrasystole heart sound in the test samples.

**Key words:** heart sound classification; Deep Convolution Neural Network; data augment

**基金项目:** 国家自然科学基金面上项目(No.61971467);南通市科技计划项目(No.JC2018133, No.JC2018128, No.JC2019123);江苏省产学研合作项目(No.BY2019120)。

**作者简介:** 陈伟(1987-),男,博士在读,研究领域为人工智能、生物医学信号处理;孙强(1980-),男,副教授,研究领域为人工智能、智能通信;齐月月(1994-),男,硕士研究生,研究领域为深度学习;徐晨(1960-),男,通讯作者,教授,研究领域为人工智能、智能信息处理, E-mail: xuchen@ntu.edu.cn。

## 1 引言

心血管疾病是全球的头号杀手，每年死于心血管疾病的人数多于任何其他死因。据报道，2016年约有1790万人死于心血管疾病，占全球死亡总数的31%<sup>[1]</sup>。早发现、早治疗是降低心血管疾病危害的重要途径。因此提高心血管疾病的早期诊断是一项尤其重要的任务。

心脏听诊可以直接反映心脏血流动力、心室收缩状况。有些心脏疾病在临床表现之前，心音往往会出现异常，通过听诊可以确诊某些异常。但是，人的耳朵在低频范围内对声音强度变化的不足及其对低频声音的敏感，使得心脏听诊任务变得困难。其次，心脏听诊对医生的要求特别高，需要拥有丰富的临床经验和技巧。因此，通过计算机自动提取心音信号(phonocardiogram, PCG)的特征进行定量分析，来辅助诊断心脏疾病具有重要的意义。

近些年，国内外很多学者利用传统的机器学习方法和信号处理的手段对心音分类的问题进行了大量的研究<sup>[2-13]</sup>。

M. Abdollahpur 等提取了 90 种心音信号的时域、频域和梅尔频率倒谱系数(MFCC)特征利用三个前馈神经网络和一个投票系统对正常和异常心音进行分类正确率达到了 0.826<sup>[2]</sup>。唐洪等在文献[3]研究的基础上提出了一种利用多域特征和支持向量机(support vector machine, SVM)的方法识别正常和异常心音信号<sup>[4]</sup>。该方法提取了大量的心音特征，其中状态频谱、能量谱和熵谱结合 SVM 分类器证实为最有效方法，相比文献[3]正确率提高了 6%。Noguera<sup>[5]</sup>等借助心电图(electrocardiogram, ECG)信号利用逻辑回归隐半马尔可夫模型(logistic regression-hidden semi-markov models, HSMM)<sup>[6]</sup>对心音信号分段，在此基础上提取了相关时域特征和频域特征然后分别用 SVM、卷积神经网络(convolutional neural networks, CNN)及随机森林分类器对心音信号分类。在所有的特征及分类器组合中 MFCC 特征及部分的时域特征结合 SVM 分类器的模型性能最优，其敏感性、特异性及最终得分分别是 0.8737、0.7907、0.8322。Chakir 等<sup>[7]</sup>在分段的心音信号中提取一系列的时频特征然后结合判别式分类器对正常、杂音、早搏三种心音分类最终总精确度为 1.58。上述研究中必须从分段的心音信号中提取大量的特征。而心音的准确分割很大程度上取决于心音信号的质量，这无疑

影响到了心音分类模型的性能。Zabihi 等<sup>[8]</sup>在未分段的心音信号中提取 18 种时频域特征，设计了一个含有 20 个前馈神经网络的分类器。该方法在 2016 年举行的全球生理测量挑战赛中获得第二名成绩<sup>[9, 10]</sup>。W.Zhang 等<sup>[11]</sup>提出了一种未分段的心音分类方法。首先，利用离散小波分解(discrete wavelet transform, DWT)从心电信号的子带包络中提取自心音信号的相关特征；然后，将自相关特征进行融合得到扩散图的统一特征表示；最后，用 SVM 分类器对其分类。该方法在 PASCAL 心音分类挑战赛上取得了出色的表现<sup>[12]</sup>。W. Zhang 等<sup>[13]</sup>提出了一种基于尺度谱图和偏最小二乘回归(partial least square regression, PLSR)的心音信号分类方法。首先，采用香农能量包络的短时平均幅度差估计心动周期；然后，利用双线性插值法对频谱图进行插值得到尺寸固定的频谱图；接着，利用 PLSR 方法对含有大量冗余信息的频谱图进行降维；最后，设计 SVM 分类器对降维后的特征分类。所提方法对正常、杂音、早搏三种心音分类总精确度为 1.75，是目前心音多类研究中精确度最高的方法。

上述基于传统机器学习的心音分类方法性能很大程度上依赖特征的提取，而且不同的特征对分类器的选择也不尽相同。深度学习作为人工智能的核心技术善于从海量数据中挖掘抽象的特征且这些特征具有良好的泛化性能。近年来，已经出现了一些文献将深度学习应用于心音的分类研究中典型的方法如卷积神经网络(CNN)<sup>[14-15]</sup>及循环神经网络(recurrent neural network, RNN)<sup>[16]</sup>。

Rubin 等<sup>[14]</sup>提出了一种利用 CNN 对正常和异常心音进行自动分类的深度学习方法。作者使用文献[6]提出的分割算法对 3 秒心音信号进行分割，然后将这些固定长度的心音信号转换为二维热图用于训练模型。但是，该团队在全球生理测量挑战赛中的成绩得分并不高只有 0.7278。Potes 等<sup>[15]</sup>集成 Adaboost 和 CNN 分类器对正常和异常的心音进行分类。在该研究中，提取了 124 种时频特征来训练 Adaboost 分类器，心音信号被分解为四个频段用于训练 CNN 分类器，这种集成的方法在该比赛中表现最佳，其特异性、敏感性和总成绩分别为 0.7781、0.9424 和 0.8602。Latif 等<sup>[16]</sup>提出了一种基于 RNN 的心搏异常检测方法，该方法基于 HSMM 方法<sup>[6]</sup>将心音分割具有 5 个心动周期序列并提取 13 种 MFCC 特征作为 RNN 模型的训练数据，其优化的模型对正常、异常的识别总正确率达到 97.63%。

本文提出了一种基于深度卷积神经网络的心音分类方法。其主要贡献有以下两点：第一，目前现有的心音分类方法绝大部分依赖对心音的正确分段，而我们所提出的方法直接对原始的未分段的心音信号进行识别。在减少计算机的计算成本情况下，提高了心音分类的识别精度。第二，本文提出了一种间接的手段对心音二维 MFSC 特征图进行扩充解决了深度学习对训练样本的需求，一定程度上提高了模型性能。

本文结构安排如下，引言介绍了研究现状和本文研究内容，第 1 节分析心音产生机理，第 2 节介绍本文提出的方法，第 3 节分析实验结果，第 4 节为结论及未来的工作。

## 2 心音产生机理

心音揭示了各种心脏生理、病理信息。它是由心肌的收缩和舒张、瓣膜启闭、血流冲击心室壁和大动脉等因素引起的机械振动而产生的声音。完整的心动周期包括两个部分：心脏收缩期（Systole）和舒张期（Diastole）。二尖瓣和三尖瓣关闭标志着心脏收缩期的开始，此期间心脏射血；收缩期又分为两个部分，第一部分分为等容收缩期和快速心室射血期。等容收缩期间以二尖瓣和三尖瓣关闭后心室内压力升高开始。快速心室射血期间，心室压力超过主动脉和肺动脉压力使得主动脉和肺动脉开放，导致血液快速从心室射出。心室收缩期的第二部分，心室压力下降，心室射血变少直到心室舒张期开始。主动脉瓣和肺动脉瓣关闭标志着心室舒张期开始，此期间心脏充盈。舒张期的前三分之一分为等容舒张期和快速充盈期。等容舒张期间血液没有进入心室，心室容积不发生变化。快速充盈期间心房压力大于心室压力，血液快速进入心室。舒张期中间的三分之一时期，几乎没有血液进入心室，此时心房和心室都处在舒张的状态。心室舒张的最后三分之一称为晚期充盈期，此期间血液从心房挤压到心室。心音包含四个成分包括第一心音(S1)、第二心音(S2)、第三心音(S3)及第四心音(S4)<sup>[17]</sup>。图

1 描述了各个心音之间的关系示意图。

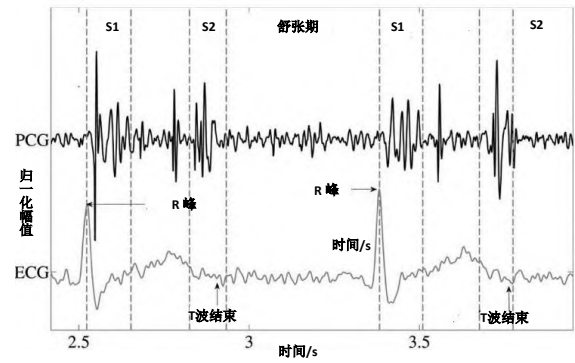


图 1 同步的心电及心音信号之间的关系示意图<sup>[18]</sup>

Fig.1 PCG signal with simultaneously recorded ECG and the four states of the PCG recording: S1, Systole, S2 and Diastole<sup>[18] Fig.4]</sup>

## 3 基于深度卷积神经网络的心音分类方法

### 3.1 系统结构

本文提出的心音分类的系统结构如图 2 所示，主要包括训练及测试部分。

模型训练阶段，包括以下几个过程。第一步，对 PCG 信号预处理。预处理过程包括，1.心音信号下采样，采样频率为 2000Hz；2.利用 6 阶巴特沃斯滤波器带宽限制在 25 至 900Hz 范围对心音滤波处理。3.经滤波的信号进行幅值归一化处理，使得提取的特征具有更强的鲁棒性。第二步，将预处理后的 PCG 信号，生成具有三个通道的二维 MFSC 特征图，具体的过程如 2.3 节介绍。第三步，均衡及扩充训练样本。由于现有心音数据集的样本数量有限及不同类别心音数量不均衡的问题，我们采用了 2.4 节的方法对样本的数量进行了有效的扩充。第四步，微调超参数训练深度卷积神经网络。

心音测试阶段，包括以下几个过程。首先，对心音信号采取与训练阶段同样的预处理过程。接着，将其转换成多张二维 MFSC 特征图。然后，用训练好的模型对每张二维 MFSC 特征图进行测预，得到不同标签的心音 MFSC 特征图。最后，利用多数表决原则决定心音的类别。

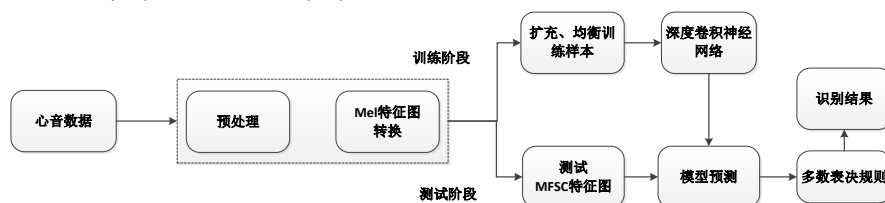


图 2 心音分类系统结构

Fig.2 Structure of Heart Sound Classification System



## 3.2 深度卷积神经网络结构

### 3.2.1 Inception-ResNet-64 网络

深度卷积神经网络是深度学习的重要模型之一，最初应用于在计算机视觉领域<sup>[14]</sup>。它是一种特殊的深度前馈神经网络，因其具有权值共享和稀疏连接等特点可大大降低网络模型的参数规模。理论上，网络的深度和宽度的增加均可提高网络的表征能力。

结合 ResNet 与 GoogLeNet 特性，Christian Szegedy 等提出了 Inception-ResNet 网络并在 ImageNet 数据集上取得了出色表现<sup>[19]</sup>。ResNet 网络不仅加快了网络收敛速度同时也提高了 Inception 网络的准确率。在此模型启发下，我们设计了 Inception-ResNet-64 网络并将其应用到心音分类任务中。

原始的模型中 Stem 模块的输入特征的尺寸为 299×299，而本文的特征尺寸只有 64×64，直接在原始的模型上训练会出现维度不匹配的问题，修改后的 Inception-ResNet-64 网络结构如图 3 所示。其中 Modified-Stem 模块是由不同尺寸的卷积操作堆叠构成的深度卷积层，其中 1×1 的卷积不仅起到了降维的作用同时提升了网络的非线性特性。

Inception-ResNet 模块不改变输出特征的参数尺寸，它不仅加深了整个网络的深度，同时加快了网络训练的速度；Reduction-A 及 Reduction-B 模块分别将输入特征参数由 31×31、15×15 降维到 15×15、7×7；卷积层深度卷积神经网络的最后分别由 Dropout 及 AveragePooling 模块组成。最终的目标函数由 Softmax Loss 和中心损失（Center Loss）构成的数据损失和模型参数 L2 正则化损失（Regularization Loss）共同组成。其中，Modified-Stem 结构如图 4 所示，图中分别给出了各个网络输出特征图的尺寸及卷积或者池化操作的参数，其中“S”、“V”代表卷积或者池化的填充方式分别为“SAME”、“VALID”。

### 3.2.2 损失函数

对修改后的深度卷积神经网络模型优化过程中，我们引入了中心损失函数，它能增加不同类别之间的类间距离，同时缩小同类别的类内距离，总损失函数如公式 1 所示：

$$L = L_s + \lambda L_c + \frac{\alpha}{2m} \sum_i \omega_i^2 \quad (1)$$

其中， $L_s$  和  $L_c$  代表 Softmax 损失和中心损失，分别由式（2）（3）给出。 $\lambda$  为平衡中心损失在总损失中的系数，式中第三项为加入到总损失的 L2 正则

化

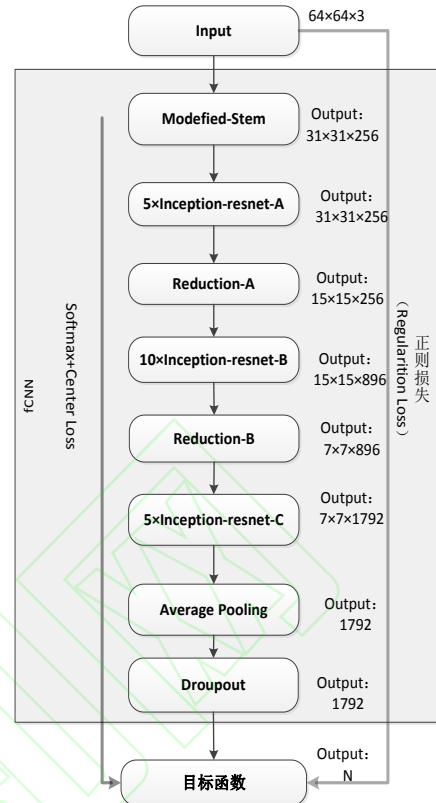


图 3 Inception-ResNet-64 模型结构

Fig.3 Inception-ResNet-64 model structure

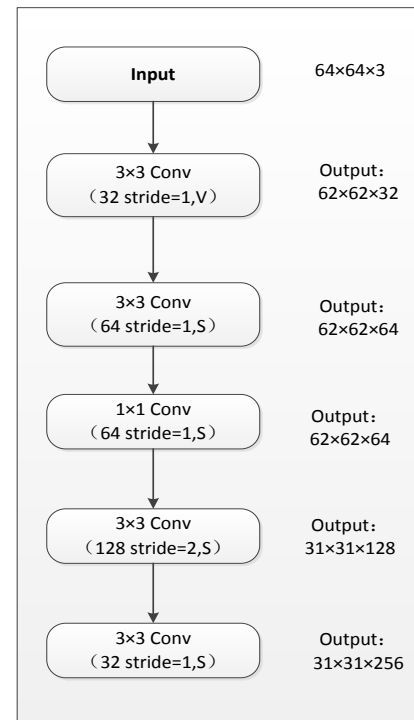


图 4 Modified-Stem 结构

Fig.4 Modified-Stem model structure

$m$  表示每批训练的样本数目。将 L2 正则化项加入到项，其由网络中的权重 $\omega$ 决定， $\alpha$ 为权重衰减系数，总损失中，可以使神经网络获得更小的权重值，从而在一定的程度上避免过拟合问题。

$$L_s = -\sum_{i=1}^m \log \frac{e^{\alpha_i}}{\sum_{j=1}^N e^{\alpha_j}} \quad (2)$$

$$L_c = \frac{1}{2} \sum_{i=0}^m \|x_i - c_{y_i}\|_2^2 \quad (3)$$

公式 (2、3) 中， $e^{\alpha_j}$ 表示来自上一层的第  $j$  的维度的值， $N$  表示分类的类别数即上一层输出的维度大小， $x_i \in R^d$ 表示第  $i$  个训练样本全连接层输出的特征向量， $y_i$ 表示第  $i$  个训练样本的实际的类标签， $c_{y_i} \in R^d$ 表示为类标签 $y_i$ 维护的该类的中心。

### 3.3 心音 MFSC 特征图的生成

声谱图作为一种可视化的声学特征广泛应用于语音信号处理领域<sup>[20, 21]</sup>。心音信号和语音信号相似也是一种典型的非平稳信号，将更符合人耳听觉特征的梅尔滤波器组应用于心音的声谱图得到的二维特征我们称之 MFSC 特征图，并将其应用于心音分类研究中<sup>[22]</sup>。心音 MFSC 特征图生成的过程如图 5 所示。首先，将预处理后的心音信号进行预加重处理即用一阶高通滤波器提升高频部分使得信号频谱更加平坦。然后，进行分帧、加窗、短时傅里叶变换等处理，其中窗口长度、帧移分别为 40ms、20ms。接着，将每一帧功率谱估计的结果经一组有 64 个 Mel 滤波器组进行处理，分别和每个滤波器进行加权求和从而获得 64 个特征能量数值。对这 64 个特征值取对数，得到 MFSC 滤波器对数能量称之为静态 MFSC 特征。对以上静态特

征进行一阶差分和二阶差分运算作为动态特征，然后将这三种特征映射到图像 RGB 通道组合成 RGB 三通道图像。由于 PCG 信号持续时间不等，而深度卷积神经网络的输入特征尺寸是固定的，我们将 MFSC 特征图在时域方向上以 64 帧即 1.28s 为一个单位对 MFSC 特征图进行连续不重叠剪切。因此，可以将一张尺寸不固定的 MFSC 特征图剪切成若干个 64×64 大小的二维图像，且 PCG 信号时间越长，产生的 MFSC 特征图也就越多。

图 6、7 分别给出了正常、杂音、早搏三种心音的时域波形图及对应的 MFSC 特征图。图 7 横坐标表示心音时间域方向的帧数，纵坐标表示心音 MFSC 频率的通道。

### 3.4 基于 MFSC 特征图的数据扩充方法

增加训练样本的数量是提高深度学习模型性能的一个有效的方法。本文间接地以心音的二维 MFSC 特征图为对象，分别从频域、时域进行处理：频域信息的部分丢失及随机的屏蔽时域的信息。假设  $g(x, y)$  为待扩充的 Mel 特征图，通过设计不同的掩模函数  $h(x, y)$  与  $g(x, y)$  相乘，可以获得扩充后的 MFSC 特征图  $f(x, y)$ ，如公式 (4) 所示其中  $x, y$  分别表示 MFSC 特征图的时域和频域坐标。

$$f(x, y) = g(x, y)h(x, y) \quad (4)$$

(1) 频率域信息的部分丢失。

$$h(x, y) = \begin{cases} 0 & (f \leq y \leq f_0 + f) \\ 1 & (\text{其他}) \end{cases} \quad (5)$$

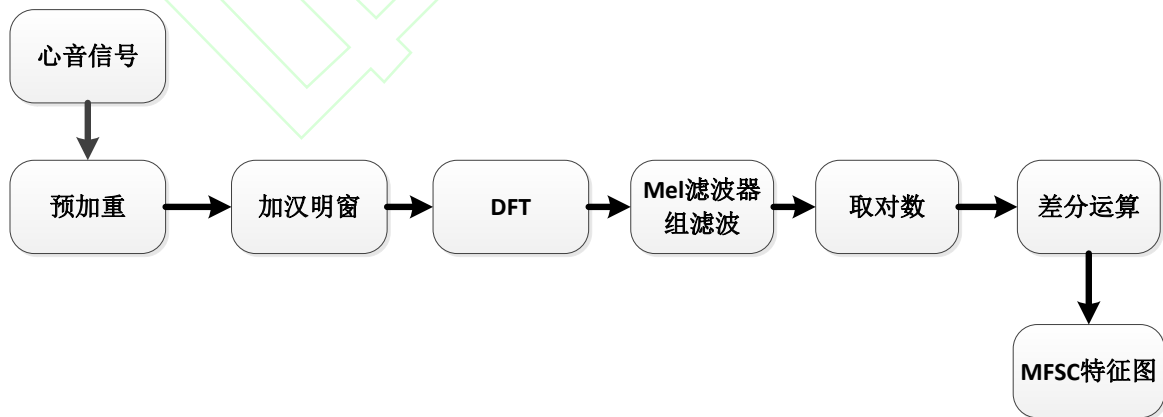


图 5 心音 MFSC 特征图产生过程

Fig.5 The process of MFSC feature maps

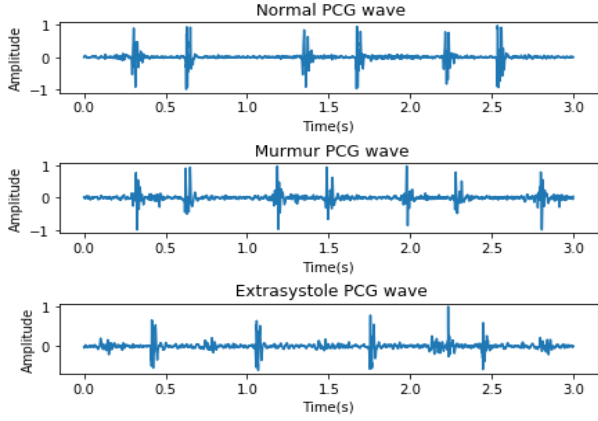


图 6 3种心音波形

Fig.6 Three heart sounds wav

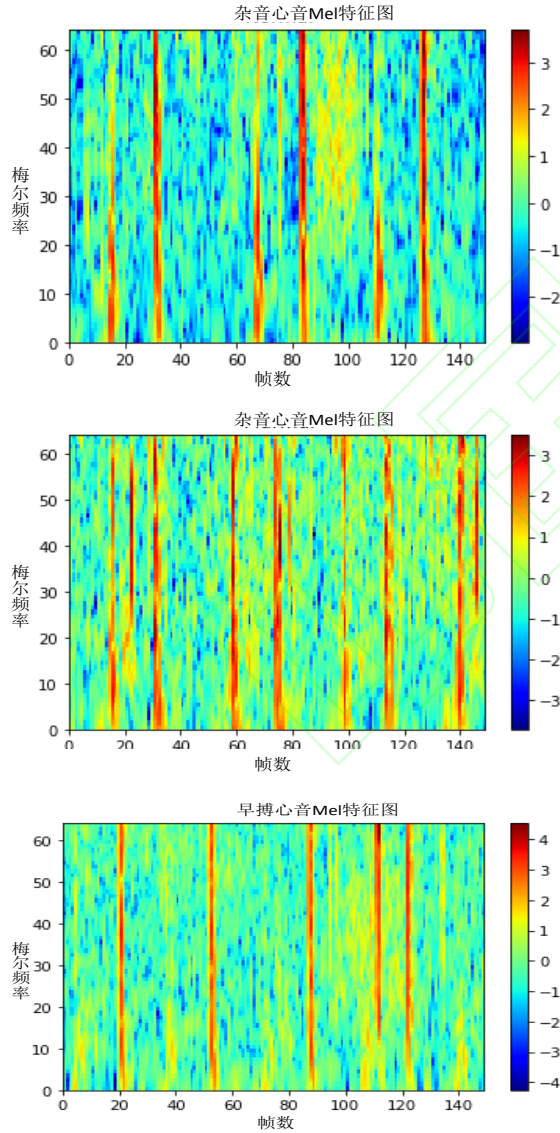


图 7 3例心音 MFSC 特征图

Fig.7 Three MFSC feature maps

其中,  $f$  是小于  $F-f_0$  的随机整数,  $F$  是 MFSC 特征图的最大通道数值,  $f_0$  是小于  $F_0$  的随机整数。

(2) 随机屏蔽时域信息。

$$h(x,y) = \begin{cases} 0(t \leq x \leq t_0 + t) \\ 1(\text{其他}) \end{cases} \quad (6)$$

其中,  $t$  是小于  $T-t_0$  的随机整数,  $T$  为 MFSC 特征图时域上最大的帧数,  $t_0$  是小于  $T_0$  的随机整数。实验中通过表 1 中的策略对 MFSC 特征图进行扩充, 其中  $N_f$ 、 $N_t$  表示掩膜函数的个数。

图 8 (a) 为一例 6s 时长的的心音的 MFSC 特征图, 该 MFSC 特征图经归一化后以零均值单位方差分布。图 8 (b)、图 8 (c) 分别是经过以上两种方法扩充后的 MFSC 特征图, 图 8 (d) 同时采用了这两种掩膜函数后的 MFSC 特征图。

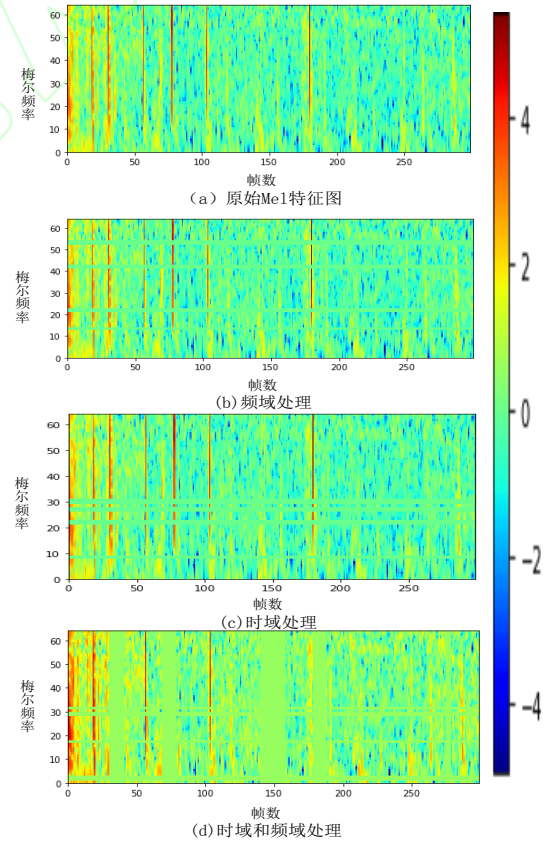


图 8 MFSC 特征图扩充结果

Fig.8 The results of augment on MFSC feature map

## 4 实验与分析

### 4.1 实验数据

为验证本文方法的效果，在 PASCAL 心音数据集（Dataset-B）上测试，与文献[7、12、13]方法比较。此数据集均来自临床，包括 320 例正常心音、95 例杂音及 46 例早搏三种心音类型。每例心音的时长不等最长有 12s，最短只有 1.2s，采样频率为 8000Hz，均为 wav 格式。表 2 给出了训练样本和测试样本的统计信息，其中测试样本的标签是未公开的。

用 2.3 节介绍的方法对训练集心音信号转换成 MFSC 特征图，转换之后的 MFSC 特征图称为原始数据集。为了验证本文所提数据扩充方法在深度卷积神经网络上的效果，分别用表 1 中的策略对原始的数据集进行扩充，扩充后的数据集分别为数据集 A、B、C 具体信息如表 3 所示。

表 1 Mel 特征图扩充策略

Table 1 Police of augment on Mel feature maps

策略	$N_f$	$F_0$	$N_t$	$T_0$
$F_1$	1	4	0	0
$F_3$	3	4	0	0
$T_1$	0	0	1	4
$T_3$	0	0	3	4

表 2 训练集和测试集样本统计

Table 2 Statistics of training and test datasets

类别	正常	杂音	早搏
训练集	200	66	46
测试集	136	39	20

表 3 Mel 特征图统计

Table 3 Statistics of Mel feature maps

数据集	策略	正常	杂音	早搏
原始	无扩充	1431	545	193
A	$T_1+T_3$	1431	1431	1431
B	$F_1+F_3$	1431	1431	1431
C	$T_1+T_3+F_1+F_3$	2844	2862	2862

### 4.2 实验指标

（1）精确度（Precision）衡量了分类器的预测的精度，用公式（7）表示。

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

其中， $TP$  表示正确划分为正例的个数， $FP$  表示错误地将负例划分为正例的个数。

（2）Youden's Index ( $r$ ) 衡量了诊断病理性心音

如杂音和早搏两种疾病的能力，用公式（8）表示。

$$r = Se - (1 - Sp) \quad (8)$$

其中， $Sp = \frac{TN}{FP+TN} \times 100\%$ ， $Se = \frac{TP}{TP+FN} \times 100\%$ 。

$TN$  表示正确划分为正例的个数， $FN$  表示错误地将正例划分为负例的个数。

（3）分类器的预测正负样本的能力用 Discriminant Power ( $DP$ ) 指标衡量。 $DP$  的数值越大分类器能够区分正负样本的能力越强。特别的，当特异性或灵敏性为 1 时，表明分类器区分正负样本的能力相当好。

$$DP = \frac{\sqrt{3}}{\pi * (\lg X) + \lg Y} \quad (9)$$

其中， $X = Se/1 - Se$ ， $Y = Sp/1 - Sp$

### 4.3 实验环境

实验系统环境为 Ubuntu16.04，内存为 64GB 的 NVIDIA GeForce GTX 1080Ti GPU，深度学习框架为 TensorFlow 1.7.0，编程语言为 python 2.3。

分别在原始数据集、A 数据集、B 数据集及 C 数据集下实验，验证集与训练集的划分比例为 1 比 4。本文的深度卷积神经网络模型配置如下：Adam 优化器优化目标函数；学习率采用阶梯下降的方式初始学习率为 0.3，每 20 次迭代学习率除以 10 直到迭代 100 次为止；L2 正则化的权重衰减设置为  $5 \times 10^{-4}$ ；Dropout 层的保持概率为 0.8；其中公式 (1) 中的标量  $\lambda$  被设置为 1， $\alpha$  设为  $5 \times 10^{-4}$ ， $m$  设为 64。

### 4.4 实验结果与分析

#### 4.4.1 其他模型对比

实验中将 J48、MLP、CS UCL<sup>[12]</sup> 和 SS-PLSR<sup>[13]</sup>、PASM<sup>[7]</sup> 几种心音分类方法与本文方法进行比较分析，各评价指标如 3.2 节所述。其中 J48、MLP、CS UCL 是当年比赛中性能最好的方法。J48 方法和 MLP 方法一样只提取了心音的时域特征，不同的是 J48 采用决策树分类器，MLP 利用多层感知机作为分类器。而 CSUCL 利用决策树分类器，小波分解和频谱图作为特征矢量。SS-PLSR、PASM 分别于 2017 年、2018 年发表是目前为止性能最好的方法。

从表 4 中我们得出，无论是精确度总精确度指标、诊断病理心音的能力  $r$  指标还是分类器的预测正负样本的能力  $DP$  指标均优于上述几种方法。同时，该模型的正常精确度、杂音精确度及早搏精确



度指标分别为 0.8047、1、1 说明了模型能够将测试集中 136 个正样本（正常心音）全部识别出来，对正负样本的区分能力表现相当的好。同时，从表中我们发现 SS-PLSR 方法的性能尤其是  $r$ 、DP 指标好于 PASM、J48、MLP 及 CSUCL 方法，一方面说明了该方法利用双线性插值对频谱图处理得到固定大小的频谱图特征更有利于 SVM 分类器的分

类。也进一步说明了传统分类器分类性能对心音特征提取的有效性非常敏感。而本文基于深度学习的方法能从心音信号的 MFSC 特征图中自动提取更有效的特征并准确预测所属的类别。

因此，本文方法与其他基于人工提取特征的方法相比明显地提高了心音分类识别的能力。

表 4 模型性能对比

Table 4 Model performance comparison

方法	数据集	正常 精确度	杂音 精确度	早搏 精确度	Heart Problem Sensitivity	Heart problem Specificity	$r$	DP	总 精确度
SS-PLSR	Dataset-B	0.76	0.65	0.33	0.34	0.90	0.24	0.36	1.75
PASM	Dataset-B	0.82	0.59	0.18	0.49	0.66	0.15	0.15	1.58
J48	Dataset-B	0.72	0.32	0.33	0.22	0.82	0.04	0.05	1.37
MLP	Dataset-B	0.70	0.30	0.67	0.19	0.84	0.02	0.04	1.67
CS UCL	Dataset-B	0.77	0.37	0.17	0.51	0.59	0.01	0.09	1.31
Inception-ResNet64	Dataset-B	0.7778	0.9524	0.3333	0.3559	0.9780	0.3339	0.7659	2.0635
Inception-ResNet64	A	0.8121	0.9310	1	0.4745	0.9853	0.4599	0.9324	2.7432
Inception-ResNet64	B	0.7972	0.9231	1	0.4237	0.9853	0.4090	0.933	2.720
Inception-ResNet64	C	0.8047	1	1	0.4408	1	0.4408	其他 (好)	2.804

4.4.2 数据扩充效果

为了验证本文提出的心音扩充方法的效果，分别比较了在原始训练集、训练集 A、训练集 B 及训练集 C 下训练的模型性能。图 9 给出了四种模型的验证集上的正确率曲线。四条曲线表明了随着迭代次数的增加模型在验证集上的正确率越来越高最终趋于稳定。在训练集 A、B、C 下的模型最终的验证集正确率分别为 0.9750、0.9871、0.9953 明显的高于原始训练集训练的模型的正确率 0.7650。训练集 C 下的模型最优是由于同时使用了时间域及频域扩充的样本，其训练集数量比训练集 A 或 B 高出了 2 倍的数据量，有效的增加了训练样本进而提高了深度卷积神经网络的学习的性能。图 10 给出了四种模型在测试样本上的总精确度指标。在训练集 A、B、C 下的模型在测试样本上的总精确度指标分别为 2.7432、2.72、2.804 均高于原始训练集训练的模型的总精确度指标。其中，训练集 A 的模型及训练集 B 的模型分别是在时间域及频率域上扩充的样本，在验证集及测试集中的表现相当说明了无论是频域屏蔽还是时间域上的随机丢弃均能有效的增加训练样本进而提高模型的性能。

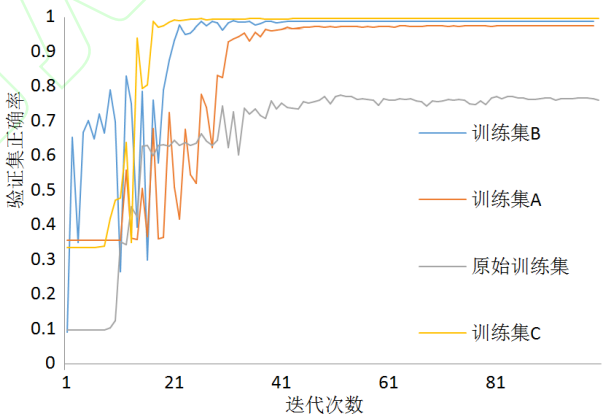


图 9 四种模型在验证集上正确率曲线对比  
Fig.9 Comparison of validation accuracy curves of Four Models

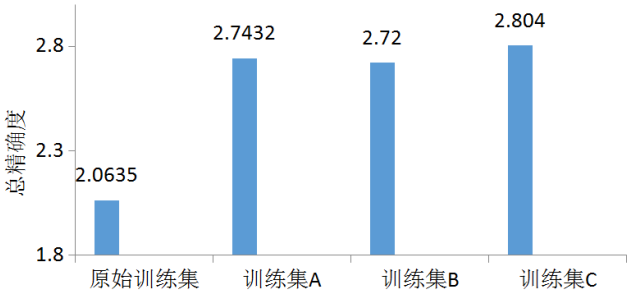


图 10 四种模型在测试集上总精确度对比  
Fig.10 Total Precision comparison of four models on test datasets

## 5 结论

本文提出了一种利用深度学习结合心音 MFSC 特征图的方法对心音信号分类识别。基于 MFSC 滤波器组的声学特征更适合人耳的听觉特性,因此将心音转换为 MFSC 特征图作为深度卷积神经网络的输入是将深度卷积神经网络应用于心音分类的切入点。基于深度学习的心音识别系统可以自动的从 MFSC 特征图中学习到能够表征不同类别心音的深层特征并通过最小化事先定义的损失函数来实现深度学习模型的训练,从而实现整个心音识别系统的优化。针对现有人工标注的数据集不多导致模型训练精度不高的问题,提出了对心音的 MFSC 特征图进行时域和频域的屏蔽处理的方法。所提出的方法大大的扩充了训练样本的数量,进一步提高了模型的性能。

深度学习有着比传统机器学习方法无可比拟的优势,将其应用于心脏疾病的智能诊断有助于提高自动化检测效率。未来我们的工作会与临床专家合作,采集、标注更加丰富的心音数据以及进一步的评估、优化我们的模型。

## 参考文献:

- [1]WHO. Cardiovascular diseases (CVDs) [EB/OL]. May 1,2020.[https://www.who.int/zh/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/zh/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
- [2]Abdollahpur M,Ghaffari A,Ghiasi S,et al.Detection of pathological heart sounds[J].Physiological Measurement, 2017,38.
- [3]Tang H, Chen H, Li T, et al.Classification of normal/abnormal heart sound recordings based on multi-domain features and back propagation neural network[C]//2016 Computing in Cardiology Conference (CinC). IEEE, 2016: 593-596.
- [4]Hong T,Ziyin D, Yuanlin J,et al.PCG Classification Using Multidomain Features and SVM Classifier[J].BioMed Research International,2018,1-14.
- [5]Nogueira D M,Ferreira C A,Gomes E F,et al.Classifying Heart Sounds Using Images of Motifs, MFCC and Temporal Features[J]. Journal of medical systems,2019,43(6): 168.
- [6]Springer D B,Tarassenko L,Clifford G D. Logistic regression-HSMM-based heart sound segmentation[J].IEEE Transactions on Biomedical Engineering,2015, 63(4): 822-832.
- [7]Chakir F,Jilbab A,Nacir C,et al.Phonocardiogram signals processing approach for PASCAL Classifying Heart Sounds Challenge[J]. Signal, Image and Video Processing, 2018, 12(6): 1149-1155.
- [8]Zabihi M, Rad A B, Kiranyaz S, et al. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation[C]//2016 Computing in Cardiology Conference (CinC). IEEE, 2016: 613-616.
- [9]C.Liu,D.Springer,Q. Li,B. Moody, R.A. Juan,F.J.Chorro,et al.An open access database for the evaluation of heart sound algorithms[J].Physiol. Meas.,2016,37(12):2181-2213.
- [10]Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, et al.Physionet 2016 Classification of normal/abnormal heart sound recordings-The PhysioNet/Computing in Cardiology Challenge 2016[EB/OL].<https://physionet.org/challenge/2016/>.
- [11]Zhang W, Han J. Towards heart sound classification without segmentation using convolutional neural network[C]// 2017 Computing in Cardiology (CinC). IEEE, 2017: 1-4.
- [12] G. N. P. Bentley, M. Coimbra, S. Mannor, The pascal classifying heart sounds challenge[DB/OL]. Available online: <http://www.peterjbentley.com/heartchallenge/index.html> (accessed on May 26, 2020).
- [13]Zhang W,Han J,Deng S.Heart sound classification based on scaled spectrogram and partial least squares regression[J]. Biomedical Signal Processing and Control, 2017, 32:20-28.
- [14]Rubin J, Abreu R, Ganguli A,et al. Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients[C]//2016 Computing in Cardiology Conference (CinC). IEEE, 2016: 813-816.
- [15]C.Potes,S.Parvaneh,A.Rahman,and B.Conroy.Ensemble of feature based and deep learning-based classifiers for detection of abnormal heart sounds[C]//in Proc.Comput.Cardiol. Conf. (CinC),Sep.2016:621-624.
- [16]Latif S,Usman M, Rana R, et al. Phonocardiographic sensing using deep learning for abnormal heartbeat detection[J]. IEEE Sensors Journal, 2018, 18(22): 9393-9400.
- [17]胡玉良,王海滨, 陈健,等. 心音时域分析的新方法研究[J]. 生物医学工程学杂志, 2010,27(2):425-428.
- Hu Y L, Wang H B, Chen J et al. A New Method for Heart Sound Analysis in Time Domain[J], Journal of Biomedical Engineering, 2010,27(2):425-428.
- [18]Lehner R J,Rangayyan R M. A three-channel micro computer system for segmentation and characterization of the

- phonocardiogram[J]. IEEE Transactions on Biomedical Engineering, 1987 (6): 485-489.
- [19] KRIZHEVAKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems, Nevada, USA, 2012:1097-1105.
- [20] 李虹, 徐小力, 吴国新, 等. 基于 MFCC 的语音情感特征提取研究[J]. 电子测量与仪器学报, 2017, 31(3):448-453.
- Li H, Xu X L, Wu G X et al. Research on Speech Emotion Feature Extraction Based on MFCC[J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(3):448-453.
- [21] 张涛涛, 陈丽萍, 蒋兵, 等. 采用深度神经网络的说话人特征提取方法[J]. 小型微型计算机系统, 2017, 38(1):142-146.
- Zhang T T, Chen L P, Jiang B et al. Novel Method for Speaker Feature Extraction Using Deep Neural Network[J]. Journal of Chinese Computer Systems, 2017, 38(1):142-146.
- [22] 谭朝文, 王威廉, 宗容, 潘家华, 朱莉莉. 卷积神经网络应用于先心病心音信号分类研究[J]. 计算机工程与应用, 2019, 55(12): 174-180.
- Tan Z W, Wang W L, Zong R et al. Research on Classification of Congenital Heart Disease Heart Sound Signal Using Convolutional Neural Network[J]. CEA, 2019, 55(12): 174-180.

