

基于深度卷积神经网络的心音分类算法

孟丽楠, 谢红薇, 宁晨, 付阳

(太原理工大学 软件学院, 山西 晋中 030600)

摘要: 针对现有心音分类算法普适性差、依赖于对基本心音的精确分割、分类模型结构单一等问题, 提出采用大量未经过精确分割的心音二维特征图训练深度卷积神经网络(CNN)的方法; 首先采用滑动窗口方法和梅尔频率系数对心音信号进行预处理, 得到大量未经过精确分割的心音特征图; 然后利用深度CNN模型对心音特征图进行训练和测试; 根据卷积层间连接方式的不同, 设计了3种深度CNN模型: 基于单一连接的卷积神经网络、基于跳跃连接的卷积神经网络、基于密集连接的卷积神经网络; 实验结果表明, 基于密集连接的卷积神经网络比其他两种网络具备更大的潜力; 与其他心音分类算法相比, 该算法不依赖于对基本心音的精确分割, 且在分类准确率、敏感性和特异性方面均有提升。

关键词: 心音分类; 梅尔频率系数; 卷积神经网络; 密集连接

Heart Sound Classification Algorithm Based on Deep Convolutional Neural Network

MENG Linan, XIE Hongwei, NING Chen, FU Yang

(College of Software, Taiyuan University of Technology, Jinzhong 030600, China)

Abstract: Existing heart sound classification algorithms based on convolutional neural networks have the disadvantages of relying on precise segmentation of basic heart sounds, single classification model structure, and poor universality. So a method of training deep convolutional neural networks using a large number of two-dimensional heart sound feature maps that have not been accurately segmented is proposed. Firstly, the heart sound signal is preprocessed by the sliding window method and the Mel frequency coefficient to obtain a large number of heart sound feature maps that have not been accurately segmented. Then the deep CNN model is used to train and test the heart sound feature maps. According to the different connection modes between convolutional layers, three deep CNN models are designed: convolutional neural network based on single connection, convolutional neural network based on skip connection, and convolutional neural network based on dense connection. The experimental results show that the convolutional neural network based on dense connections has greater potential than based on single or skip connection. Compared with other heart sound classification algorithms, the algorithm we proposed does not rely on precise segmentation of basic heart sounds and has improved the accuracy, sensitivity and specificity of classification.

Keywords: heart sounds classification; Mel frequency spectral coefficients; convolutional neural network; densely connected

0 引言

根据国家心血管病中心发布的《中国心血管病报告 2018》, 心血管病患病人数达 2.9 亿, 且仍处于持续上升阶段^[1]。心音作为反应人体心脏活动及心血管功能的一种体征信号^[2], 可以通过心音听诊被专家评估, 进而得到心血管疾病的病理信息。另外, 心音信号所包含的一部分信息是心电图所不能体现的^[3]。因此, 心音信号分类是心血管疾病计算机辅助诊断研究和应用的关键技术。

传统的心音分类方法包含心音分割、特征提取和分类器设计 3 个步骤。在心音分割中, 一般会将会心音分割成基本心音^[4], 即第一心音、第二心音、收缩期和舒张期的声音。在特征提取过程中, 大多数研究者提取基于时间^[5]、

基于频率^[6]或基于时频^[7-8]的心音特征, 作为后续分类器的输入。在分类阶段, 选择支持向量机^[8-9]、K 近邻算法^[10]或决策树^[11]等分类器对提取的特征进行分类。传统心音分类方法过程简单、可解释性强, 但其也存在一些不足: 在心音分割阶段大多需要参考同时期心电图; 手工提取特征不够全面、耗时长; 分类器的分类效果严重依赖于分割的准确性和特征选取的有效性。

近年来, 卷积神经网络(CNN, convolutional neural network)以其强大特征提取能力在图像识别^[12]、声音识别^[13]等领域被广泛应用。于是, 近几年很多学者提出了基于 CNN 的心音分类算法。对于 CNN 模型而言, 多数采用浅层网络模型^[14-18]或者混合模型^[19-20]。对于 CNN 模型的

收稿日期: 2021-01-12; 修回日期: 2021-02-20。

基金项目: 国家自然科学基金(61872262); 山西省基础研究计划项目(201801D121143)。

作者简介: 孟丽楠(1994-), 女, 山西晋城人, 硕士, 主要从事人工智能、医学信息学方向的研究。

通讯作者: 谢红薇(1962-), 女, 江苏武进人, 硕士生导师, 教授, 主要从事人工智能、医学信息学方向的研究。

引用格式: 孟丽楠, 谢红薇, 宁晨, 等. 基于深度卷积神经网络的心音分类算法[J]. 计算机测量与控制, 2021, 29(8): 211-217, 222.

输入而言,可选择心音信号的梅尔频谱系数^[14-15] (MFSC, Mel frequency spectral coefficients)、梅尔频率倒谱系数^[16] (MFCC, Mel-frequency cepstral coefficients)、功率谱密度图^[17-18] (PSD, power spectral density) 或一维波形信号^[21]等。Vykintas^[14]和谭朝文^[15]都将心音的 MFSC 以及一阶差分和二阶差分组织成心音的三通道特征图分别作为 6 层和 5 层 CNN 的输入,不同之处在于谭朝文^[15]需参考心电图信号。Rubin^[16]等人首先将心音分割成基本心音,进而提取基本心音的 MFCC,然后利用 4 层 CNN 进行分类。Nilanon^[17]和许春冬^[18]提取心音的 PSD,分别输入 4 层和 5 层的 CNN 进行训练与分类。Potes^[19]等人设计了 AdaBoost 和 CNN 结合的集成分类器,但其分类器的输入特征依赖于对基本心音的精确分割;李伟等人^[20]将 6 层 CNN 和 3 层循环神经网络相结合,分别提取心音的频域和时域特征进行分类。Bin^[21]等人仅仅利用了心音的一维时序特征,而其 CNN 深达 19 层,达到了目前最好的分类效果。

尽管近几年提出了一些基于 CNN 的心音分类方法,但是这些方法不同程度地存在以下三方面不足:1) 大多心音分类算法往往依赖于对基本心音的精确分割,过程繁琐,必然影响分类性能;2) 由于分类模型结构单一,未能充分挖掘和提取心音信号特征,导致分类性能仍有待提高。3) 心音分类算法往往在小样本中进行训练分类,其普适性无法得到有效验证。

针对以上不足,提出的大量未经过精确分割的心音二维特征图训练深度 CNN 模型的方法,消除了基本心音分割步骤产生的误差,同时可以显著提高分类模型的性能和普适性。该方法首先通过滑动窗口方法获取大量心音样本,再采用 MFSC 算法生成心音的二维特征图,最后利用深度 CNN 模型进行训练和测试。在深度 CNN 模型的设计中,根据卷积层间连接方式的不同,设计了 3 种深度 CNN 模型:基于单一连接的卷积神经网络 (SingleCNN, convolutional neural network based on single connection)、基于跳跃连接的卷积神经网络 (SkipCNN, convolutional neural network based on skip connection)、基于密集连接的卷积神经网络 (DenseCNN, convolutional neural network based on dense connection)。实验中,分别使用 Challenge 2016 的 29 384 个和 CHSC 数据集的 6 628 个心音样本作为 3 种深度 CNN 模型训练和测试数据,实验表明 DenseCNN 作为心音信号分类的深度 CNN 模型是最有潜力的。与目前已有的分类算法进行了分析比较,结果显示本文提出的心音分类算法不依赖于心音精确分割,且在分类准确率、敏感性和特异性方面均有优势。

1 数据预处理

1.1 滑动窗口方法

原始心音数据通常是一组不固定长度的时间序列,而深度 CNN 模型的输入往往需要固定大小的样本输入,因此,将原始心音信号分割为固定的心音片段是非常有必要

的。目前大多数心音分类算法选择将心音分割成基本心音,再进行特征提取和分类,这样做的弊端在于过程繁琐且特征提取和分类严重依赖于对基本心音的精准分割。本文采用滑动窗口的方法对原始数据进行截取,避免了基本心音分割带来的不确定性,同时也可以扩增样本数据量,为大量样本训练深度 CNN 模型奠定了基础。

滑动窗口方法是指对一段长度为 $n-1$ 的时间序列 $X = \{x[0], x[1], \dots, x[n-1]\}$ 借助窗口长度为 c 的滑动窗口被分割为多个长度为 c 的子时间序列 $Y = \{x[i], x[i+1], \dots, x[c+i-1]\}$ 。滑动窗口左端从起始点 $x[0]$ 开始,在时间方向上以步长为 s 的步幅向后迭代,直到到达整个时间序列 T 的末尾,序列末尾不足滑动窗口长度的舍弃。最终可以得到 $((n-1-c)/s+1)$ 个子序列。

图 1 展示了采用滑动窗口方法处理原始心音信号的过程。本文中一段原始心音信号即为时间序列 X ,采用滑动窗口方法处理后得到心音子序列 Y 作为后续步骤的样本数据。滑动窗口的长度 c 为数据集中最短的时间序列 X 的长度,步幅 s 小于 c 。

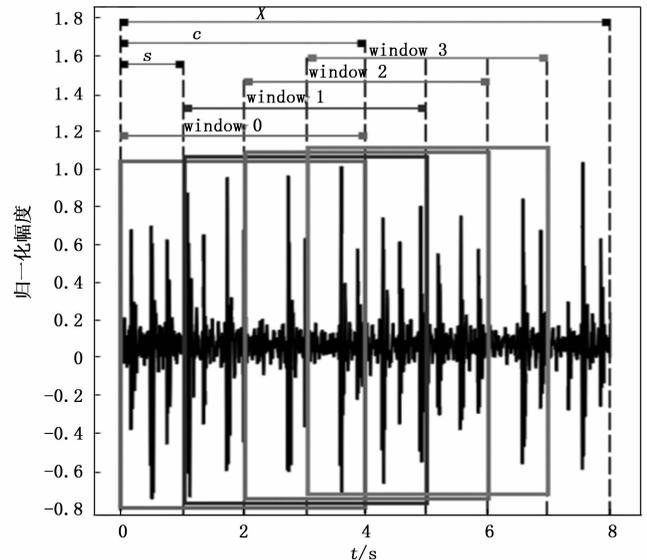


图 1 滑动窗口方法截取心音子序列

1.2 MFSC

MFSC 可以将心音的一维时序信号转换为二维特征图,便于深度 CNN 模型的特征提取和分类。MFSC 是去除了离散余弦变换步骤的 MFCC。MFCC 特征提取成功应用于语音识别^[22]、声纹识别^[23]以及各种语音模式识别^[24]任务中。然而,谭朝文等人^[25]指出:MFSC 相比较 MFCC 光谱能量能保持局部特性,MFSC 具有高度相关,在频谱上较为平滑,少量的 MFSC 特征即可代表绝大部分信号的信息,而且 Maknickas 等人^[14]也表明 MFSC 比 MFCC 在心音表征特征方面更具有优越性,故本文选用 MFSC 作为深度 CNN 模型的输入样本。

MFSC 包括以下步骤:预加重、分帧加窗、快速傅里叶变换、功率谱计算、梅尔滤波、对数能量计算。

1.2.1 预加重

语音信号处理中, 预加重是将语音信号通过一个高通滤波器, 以达到补偿高频分量, 放大高频共振峰的目的。预加重信号的计算方式如式 (1):

$$x'(n) = x(n) - \partial x(n-1) \quad (1)$$

其中: $x(n)$ 为原始信号, $x'(n)$ 为预加重后信号, n 为离散时间点, ∂ 为预加重因子, 一般取 $0.9 \sim 1.0$ 。

1.2.2 分帧、加窗

心音信号是一种短时非平稳信号, 研究起来困难, 分帧可以使信号近似为平稳信号。加窗可以防止信号相邻两帧之间变化过大, 使得相邻两帧有重叠部分, 克服频谱泄露现象^[15]。分帧后, 心音信号的帧数 Z 的计算公式为式 (2):

$$Z = \frac{c \times f_s - F_l}{F_m} + 1 \quad (2)$$

其中: c 为截取的心音子序列时间长度, F_l 为帧长, F_m 为帧移, f_s 为采样频率。 F_l 通常取 $10 \sim 30$ ms 信号所包含的点数, F_m 通常为 0.5 倍的 F_l 。

1.2.3 快速傅里叶变换与功率谱计算

利用快速傅里叶变换将信号从时域转换到频域是信号处理的一般方法。对心音的时域信号 $x'(n)$ 进行快速傅里叶变换, 可获得心音信号的频谱特征 $X(k)$, 计算公式为式 (3):

$$X(k) = \sum_{n=1}^N x'(n) e^{-j2\pi kn/N}, \quad 1 \leq n, k \leq N \quad (3)$$

其中: N 为每帧心音信号所包含的点数。在此基础上计算每帧信号的功率谱 $P(k)$, 计算公式为式 (4):

$$P(k) = \frac{1}{N} |X(k)|^2 \quad (4)$$

1.2.4 梅尔滤波与对数能量计算

梅尔滤波与对数能量计算过程可以使得声音信号更符合人类的听觉系统, 有助于心音分类。每帧心音信号的功率谱经过梅尔滤波器组, 并计算每帧信号在每个通道上的对数能量 $E(m)$, 计算方法如式 (5)。梅尔滤波器的个数 M 通常取 $20 \sim 26$ 。

$$E(m) = \ln \left[\sum_{k=0}^{N-1} P(k) H_m(k) \right], \quad 0 \leq m \leq M \quad (5)$$

其中: $H_m(k)$ 表示第 m 个滤波器第 k 条谱线对应的强度。

每帧信号通过一组梅尔滤波器即可得到该帧信号的一组梅尔系数, 每组梅尔系数以每帧信号所指向的时间为序进行排列, 可得到梅尔系数矩阵, 即心音信号经过 MFSC 算法后, 从一维变换到二维, 每帧信号包含频率信息, 而每帧之间以时间为序, 因而 MFSC 特征图包含频域和时域信息。

如图 2, 心音子序列 Y 经过 MFSC 算法后转换为 $M \times Z$ 特征图, 特征图的横轴方向表示心音信号时间域方向的帧数, 纵轴方向表示第一个到第 M 个梅尔滤波器对应的频率通道。

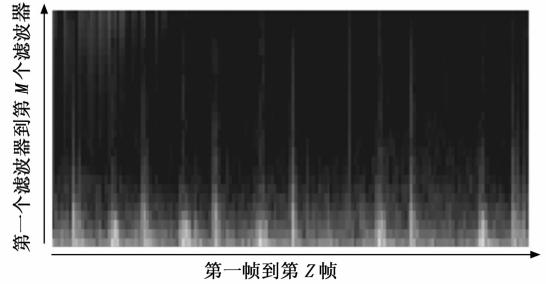


图 2 心音子序列的 MFSC 特征图

2 深度 CNN 模型

本文提出的用于心音分类的深度 CNN 模型的总体架构如图 3 所示。主要由普通卷积层 (Conv_1, Conv_2, Conv_3)、复合卷积层 (Com_1, Com_2, Com_3)、池化层 (Pool_1, Pool_2)、全局池化层 (GlobPool_1, GlobPool_2, GlobPool_3) 和全连接层 (FC) 组成。

第一个普通卷积层 Conv_1 用 64 个 3×3 卷积核和修正线性单元 (Rectified Linear Unit, ReLU) 提取了心音信号特征图 X_0 的 64 个特征组成 X_{1_c} , 为后续网络层提取心音特征奠定了基础。池化层 Pool_1 和 Pool_2 之前的卷积层 Conv_2 和 Conv_3 包含 1×1 卷积核和 ReLU 函数, 增加了网络的非线性, 使得本文深度 CNN 模型可以表达更加复杂的心音信号特征。复合卷积层 Com_1 和 Com_2 的输出特征图 X_{1_m} 和 X_{2_m} 在网络中有两种流向, 一种是流入下一个普通卷积层, 另一种是与当前复合卷积层的输入特征图 X_{1_c} 在通道方向进行连接, 作为全局池化层 GlobPool_1 和 GlobPool_2 的输入特征图 X_{CM_1} 和 X_{CM_2} 。Com_3 的输出特征图仅有第二种流向。池化层的作用在于对前一层的输出特征图进行简化, 达到缩减特征图尺度的目的。全连接层 FC 将经 3 个全局池化层压缩后的心音特征进行融合, 并用 softmax 分类。

其中, 复合卷积层是本文重点探索和研究的模块。在 CNN 发展过程中, 卷积层之间经典的连接方式有单一连接^[26]、跳跃连接^[27]以及密集连接^[28]。在此基础上, 设计了 3 种复合卷积层单元结构分别构建 3 种深度 CNN 分类模型, 期望找到最优的用于心音分类的深度 CNN 模型。

2.1 基于单一连接的复合卷积层

单一连接, 即 CNN 的第 l 层的输出特征图作为第 $l+1$ 层的输入特征图^[26], 可以用式 (6) 表示:

$$X_{l+1} = H_{l+1}(X_l) \quad (6)$$

其中: X_l 表示卷积神经网络第 l 层的输出, X_{l+1} 表示第 $l+1$ 层的输出, H_{l+1} 表示第 $l+1$ 层的非线性变换函数。以 5 个卷积层为例, 其结构如图 4 所示。此结构中, 特征信息流从低层逐层流入高层。

2.2 基于跳跃连接的复合卷积层

跳跃连接最早出现在 ResNet^[27] 中, 表示当前卷积层的输入特征图来源于前两层输出特征图的叠加和。卷积层之间的关系可用式 (7) 表示:

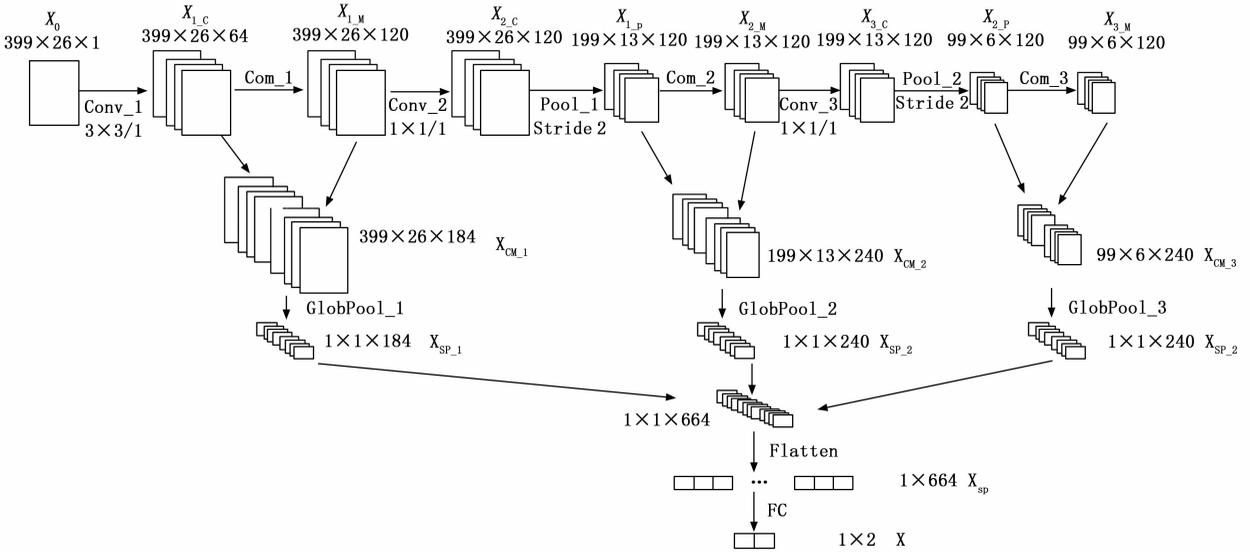


图 3 本文深度 CNN 模型的总体架构图

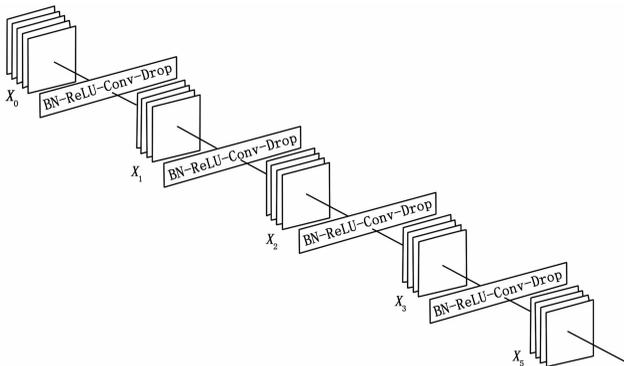


图 4 基于单一连接的复合卷积层

$$X_i = H_i(X_{i-1}) + X_{i-1} \quad (7)$$

图 5 展示了基于跳跃连接的复合卷积层。此结构中，梯度信息可以直接通过恒等映射 X_{i-1} 从上层传输到后层网络^[27]，优点是网络层次很深时，减少了连续求导运算导致的梯度消失。

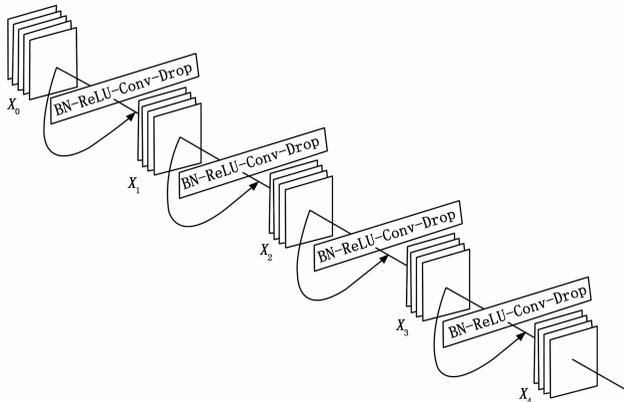


图 5 基于跳跃连接的复合卷积层

2.3 基于密集连接的复合卷积层

密集连接在 DenseNet^[28] 中首次提出，其利用特征重用的方式加强卷积层间有效信息的流动。此结构中，任意卷积层均和其后的所有层建立直接连接，即网络中第 l 层的输入特征图来自其前面所有层的输出特征图。其卷积层间的关系可用式 (8) 表示：

$$X_l = H_l([X_0, X_1, X_2, \dots, X_{l-1}]) \quad (8)$$

其中： $[X_0, X_1, \dots, X_{l-1}]$ 指第 0、 \dots 、 $l-1$ 层的输出特征图在通道方向直接连接。图 6 展示了密集连接结构。此结构可以使输出层获得之前所有网络层的信息，增大了特征图信息的利用率。

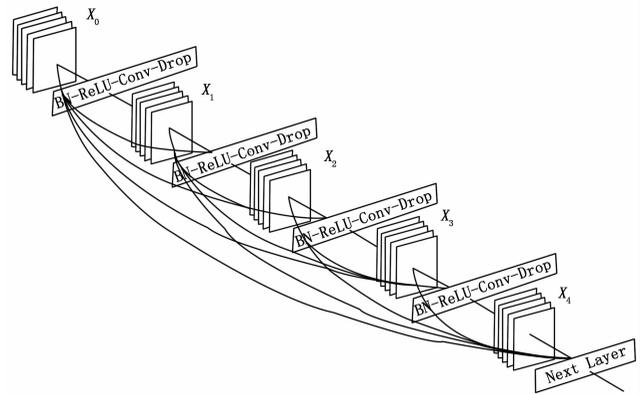


图 6 基于密集连接的复合卷积层

3 种复合卷积层的每层网络均包含批归一化 (BN, batch normalization)、ReLU、卷积 (Convolution, Conv) 和 Dropout 操作。用上述 3 种复合卷积层构建本文心音分类的深度 CNN 模型的总体架构中的复合卷积层 Com_1、Com_2 和 Com_3，便得到了本文的 3 个候选的深度 CNN 模型：SingleCNN、SkipCNN 和 DenseCNN。

3 实验结果与分析

3.1 数据准备及预处理

3.1.1 Challenge 2016 数据集

实验中使用的第一个心音数据集来自 PhysioNet/CinC Challenge 2016 数据库^[29]。该数据库由来自不同研究小组的 9 个子数据库组成, 均来自健康的和患有各种心脏疾病的受试者, 总共包含 3 240 段心音录音。这些录音是通过各种型号的电子听诊器所采集的, 长度从 4~120 s 不等, 采样频率为 2 kHz。

在 Challenge2016 的数据集上, 采用滑动窗口方法处理数据时, 窗长 c 为 4 s, 步幅 s 为 1 s, 最终得到 61 624 个心音子序列, 包括 14 692 个负类子序列和 46 932 个正类子序列。其中, 正类子序列表示正常心音信号, 负类子序列表示异常心音信号。为得到正负类子序列平衡的数据样本集, 本文随机舍弃了多于负类子序列的正类子序列, 最终得到 14 692 个正类子序列和 14 692 个负类子序列。

将心音子序列转换为 MFSC 特征图时, 选择 F_l 为 40 点, F_m 为 20 点, 梅尔滤波器个数为 26, 故可得到尺寸为 $399 \times 26 \times 1$ 的 MFSC 特征图。

3.1.2 CHSC 数据集

实验中使用的第二个数据集来自心音分类挑战赛 (CHSC, classifying heart sounds challenge)。该数据集有两大类: 第一类是通过 iStethoscope Pro iPhone 应用程序从大众人群获得, 采样频率为 44.1 kHz; 第二类使用数字听诊器从医院听诊试验中收集, 采样频率为 4 kHz。整个数据集包含 585 条 1~30 s 不等的心音录音。使用窗长 c 为 1 s、步幅 s 为 0.5 s 的滑动窗口对原始心音信号进行截取, 然后平衡正负类子序列数量, 得到 6 628 个心音子序列。

将心音子序列转换为 MFSC 特征图时, 第一类数据样本选择 F_l 为 882 点, F_m 为 441 点; 第二类数据样本选择 F_l 为 80 点, F_m 为 40 点, 梅尔滤波器个数为均为 26, 故可得到尺寸为 $99 \times 26 \times 1$ 的 MFSC 特征图。

对两个数据集的 MFSC 特征图均按照 4:1 划分训练集和测试集, 并作为本文深度 CNN 模型的输入数据。

3.2 实验环境设置和网络参数配置

1) 软件环境。本文实验在 python2.7 环境下完成。MFSC 特征提取用到 python_speech_features 依赖包, 卷积神经网络主要用到 tensorflow-gpu 1.14.0。

2) 硬件环境。所有训练和测试都是在服务器中完成, 处理器 Intel (R) Xeon (R) Silver 4210 CPU@2.2 Hz, 内存 128 G, 显卡 GeForce RTX 2080 Ti 10 GiB。

3) 网络参数配置: 本文 CNN 模型的初始学习率为 0.1; 总迭代次数为 100 次, 随着迭代次数增加, 学习率会随之动态变化 (总迭代次数的三分之一和四分之三时降为原来的 0.1 倍); 使用带有动量的随机梯度下降算法训练和优化模型, 动量为 0.9, 权重衰减为 10^{-4} , Dropout 率为 0.5。

3.3 实验结果与分析

本文性能评估中, 主要使用准确率 (Acc, accuracy)、敏感性 (Sen, sensitivity)、特异性 (Spe, specificity) 和损失值 (loss) 4 种评价指标对不同深度 CNN 模型的性能进行评价, 计算方式如式 (9)~(12):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$Sen = \frac{TP}{TP + FN} \quad (10)$$

$$Spe = \frac{TN}{FP + TN} \quad (11)$$

其中: TP 、 TN 、 FP 和 FN 为二分类问题中所有可能的预测结果, 分别表示正样本被预测为正类、负样本被预测为负类、负样本被预测为正类和正样本被预测为负类。

损失值的计算用了交叉熵损失函数, 其定义为公式 (12), 其中 $Loss$ 是损失值, y_n 为第 n 个样本的真实值, \hat{y}_n 为第 n 个样本的预测值, N 为样本总数。

$$Loss = \frac{1}{N} \sum_{n=1}^N y_n \lg \hat{y}_n + (1 - y_n) \lg (1 - \hat{y}_n) \quad (12)$$

3.3.1 本文 3 种深度 CNN 模型的比较

表 1 是 3 种已优化 CNN 模型的网络架构参数配置, 表中 $nblock$ 表示复合卷积层个数、 $nFilter$ 表示每个复合卷积层内每层卷积层过滤器个数、 $nLayer$ 表示每个复合卷积层内卷积层个数、 $totalLayer$ 表示网络总层数。

表 1 本文模型架构参数配置

模型	$nblock$	$nFilter$	$nLayer$	$totalLayer$
SingleCNN	5	24	5	31
SkipCNN	5	24	5	31
DenseCNN	3	24	5	19

图 7 展示了 3 种 CNN 模型在训练过程中 Acc 和 $Loss$ 在 Challenge2016 训练集和测试集上的变化趋势。

3 种模型在整个训练过程中均迭代 100 次, 在第 33 次迭代之前, 3 种模型在训练集上 Acc 迅速上升、 $Loss$ 迅速下降, 其原因在于初始学习率较大, 可以让模型快速趋于最优状态, 但是从测试集来看, Acc 和 $Loss$ 震荡较严重, 此时的模型并不是真的达到最优, 而是陷入了局部最优值。迭代 33 次以后, 学习率自动调整为原来的 0.1 倍, 模型在训练集上进一步精细优化, 在测试集上表现出微小震荡; 到第 75 次迭代之后, 模型在训练集和测试集均达到稳定收敛状态。不同的是, DenseCNN 从第一次迭代开始就表现出了自身挖掘和提取心音特征的优势, 其分类效果在整个训练过程中的 Acc 和 $Loss$ 一直明显优于其余两种 CNN 模型, SkipCNN 的 Acc 和 $Loss$ 仅次于 DenseCNN。在 DenseCNN 和 SkipCNN 在测试集上趋于稳定时, SingleCNN 还存在微小震荡, 相比之下分类效果最差。

表 2 展示了 3 种 CNN 模型在 Challenge2016 数据集上

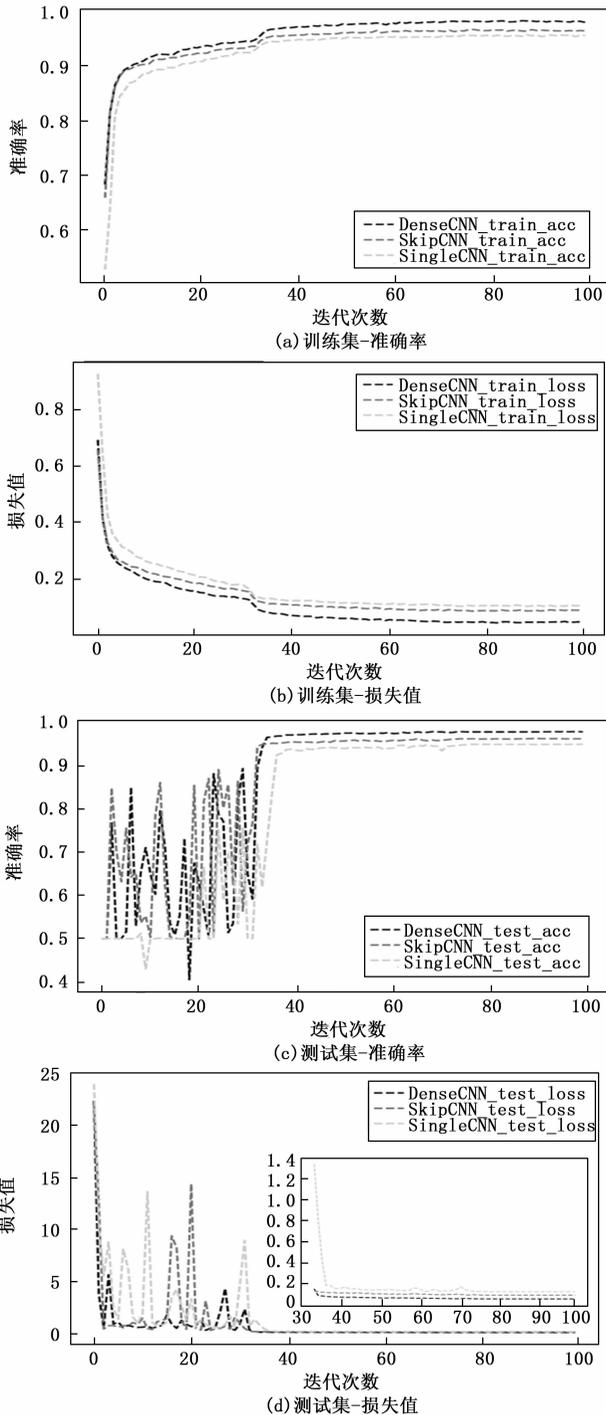


图 7 SingleCNN、SkipCNN 和 DenseCNN 训练和测试过程

的 *Acc*、*Sen*、*Spe* 以及 *Loss*。其中，DenseCNN 的分类效果在 3 个模型中更为突出，拥有最高的 *Acc*、*Sen* 和 *Spe*。DenseCNN 比 SingleCNN 和 SkipCNN 在 *Acc* 方面分别高 3.30% 和 2.49%，在 *Sen* 方面分别高 1.44% 和 2.39%，在 *Spe* 方面分别高 4.18% 和 1.77%；在 *Loss* 方面，DenseCNN 的 *Loss* 仅为 SingleCNN 和 SkipCNN 的近 0.5 倍。SkipCNN 的 *Acc* 和 *Spe* 均高于 SingleCNN，*Loss* 略低于

SingleCNN 的 *Loss*，但是 *Sen* 不如 SingleCNN。

表 3 展示了 3 种 CNN 模型在 CHSC 数据集上的分类结果。其中，DenseCNN 模型的 *Acc*、*Sen* 和 *Spe* 均高于 SingleCNN 和 SkipCNN 模型，*Loss* 也是 3 种模型中最小的。

表 2 各模型分类效果对比(Challenge2016 数据集)

模型	<i>Acc</i> /%	<i>Sen</i> /%	<i>Spe</i> /%	<i>Loss</i>
SingleCNN	94.92	97.01	93.00	0.124 1
SkipCNN	95.73	96.06	95.41	0.119 0
DenseCNN	98.22	98.45	97.18	0.055 7

表 3 各模型分类效果对比(CHSC 数据集)

模型	<i>Acc</i> /%	<i>Sen</i> /%	<i>Spe</i> /%	<i>Loss</i>
SingleCNN	74.52	77.41	73.40	0.523 1
SkipCNN	75.53	75.86	75.51	0.419 0
DenseCNN	78.22	78.33	77.24	0.312 7

结合 3 种 CNN 模型的架构参数来分析，SingleCNN 和 SkipCNN 达到最好的分类效果时网络深度一样，但是 SkipCNN 的表现略胜于 SingleCNN，这是因为 SkipCNN 中的跳跃连接能够在深层网络中减少梯度消失，使得更多有效信息可以被网络学习到。DenseCNN 在达到最好的分类效果时，层数相较于 SkipCNN 和 SingleCNN 达到各自结构下最好分类效果时的层数浅，但是其表现出了更好的分类效果。这是由于 DenseCNN 具有密集连接，采用此结构可以实现心音特征的重用，使得网络以更少的层数就能够充分挖掘和提取更多的心音特征。故相比采用单一连接或跳跃连接结构的 CNN，采用密集连接结构的 CNN 模型更有利于心音分类。

3.3.2 本文方法与已有方法的比较

本小节实验分别在两个数据集上与其他心音分类算法进行比较，以验证本文算法的优越性。

表 4 比较了 3 种已有心音分类算法与本文的最优算法，这 3 种心音分类算法所用数据集均为 Challenge 2016。

表 4 本文算法与其他算法分类效果对比
(Challenge2016 数据集)

算法	<i>Acc</i> /%	<i>Sen</i> /%	<i>Spe</i> /%
MFSC-CNN[14]	90.10	89.33	91.76
AdaBoost-CNN[19]	90.88	95.54	88.81
1-D CNN[21]	94.10	88.05	95.22
DenseCNN	98.22	98.45	97.18

从表中数据来看，本文心音分类算法在 *Acc*、灵敏度和 *Spe* 方面都有优势。本文算法的 *Acc* 为 98.22%，比 MFSC-CNN^[14]、AdaBoost-CNN^[19]、1-D CNN^[21] 分别提高了 8.12%、7.34%、4.12%；本文算法的 *Sen* 为 98.45%，比 MFSC-CNN^[14]、AdaBoost-CNN^[19]、1-D CNN^[21] 分别

提高了 9.12%、2.91%、10.4%；本文分类算法的 *Spe* 为 97.18%，比 MFSC-CNN^[14]、AdaBoost-CNN^[19]、1-D CNN^[21] 分别提高了 5.42%、8.37%、1.96%。

表 5 展示了 3 种心音分类算法与本文算法在 CHSC 数据集上的分类效果。本文的 DenseCNN 模型在 *Acc*、*Sen* 和 *Spe* 方面相比其他分类算法均有所提升。

表 5 本文算法与其他算法分类效果对比(CHSC 数据集)

算法	<i>Acc</i> /%	<i>Sen</i> /%	<i>Spe</i> /%
MFSC-CNN ^[14]	76.32	75.66	69.10
AdaBoost-CNN ^[19]	75.22	74.44	67.81
1-D CNN ^[21]	77.51	76.10	75.00
DenseCNN	78.22	78.33	77.24

在两个数据集上均已验证本文算法的有效性，分析其原因，文献 [14] 所用算法同本文整体思路一致，都用了心音的 MFSC 特征图和 CNN 模型，但在 CNN 方面，文献 [14] 的 CNN 由两个卷积层和 3 个全连接层堆叠而成，本文 CNN 模型层次更深、结构更复杂，提取有鉴别性特征的能力更强。文献 [19] 不仅需对基本心音信号的精准分割，还要提取基本心音的 9 个频带和 MFCC 特征，其分类效果严重依赖精确分割；而本文算法采用滑动窗口方法得到心音的固定片段，消除了基本心音分割不准确带来的隐患，同时丰富了样本多样性，使得模型普适性和鲁棒性更强。文献 [21] 和本文提出的心音分类是从深度模型角度解决心音分类问题，但其模型输入使用心音的一维时序信号，而本文使用的是兼具时域和频域信息的二维 MFSC 特征图，可以充分发挥深度 CNN 模型提取特征的能力。

4 结束语

针对现有的基于卷积神经网络的心音分类算法普适性差、依赖于对基本心音的精确分割、分类模型结构单一的问题，本文提出采用大量未经过精确分割的心音二维特征图训练深度卷积神经网络的方法用于心音分类。在数据预处理方面，通过滑动窗口方法获取大量的心音样本，又采用 MFSC 算法将心音的一维时序信号变换为兼具时域和频域信息的二维特征图，作为后续深度 CNN 模型的输入。深度 CNN 模型方面，首先设计了分类模型总体架构图，然后利用 3 种不同的复合卷积层构建模型细节，得到 3 种深度 CNN 分类模型。实验表明，基于密集连接的卷积网络模型由于其特征重用的功能在心音分类任务中展示了明显的优势；基于跳跃连接的卷积神经网络分类效果仅仅次于基于密集连接的卷积神经网络，基于单一连接的卷积神经网络分类效果在三者中表现最差。相比其他已有心音分类算法，本文心音分类算法不依赖于基本心音分割，训练样本量更大，模型结构表征能力更强，使得心音分类算法普适性和鲁棒性更强，分类性能更佳。在后续的研究中，可以探索

在 MFSC 中增加一阶差分和二阶差分作为 MDenseCNN 模型的输入特征，进一步提高心音分类效果。

参考文献:

- [1] 胡盛寿, 高润霖, 刘力生, 等. 《中国心血管病报告 2018》概要 [J]. 中国循环杂志, 2019, 34 (3): 209-220.
- [2] 陈 剑, 郭兴明, 肖守中. 心音信号识别的意义及其方法的研究 [J]. 国外医学 (生物医学工程分册), 2004 (2): 87-89.
- [3] 于云之, 聂邦畿. 心音的临床意义及研究现状 [J]. 现代医学仪器与应用, 1997 (3): 9-12.
- [4] SPRINGER D B, LIONEL T, CLIFFORD G D. Logistic regression-HSMM-based heart sound segmentation [J]. IEEE Transactions on Bio-medical Engineering, 2016, 63 (4): 822-832.
- [5] FATEMEH S, SHYAMALA D, AZREEN A, et al. Multi-level basis selection of wavelet packet decomposition tree for heart sound classification [J]. Computers in Biology and medicine, 2013, 43 (10): 1407-1414.
- [6] SUMETH Y, AKINORI N, WAREE K, et al. A framework for automatic heart sound analysis without segmentation. [J]. Biomedical engineering online, 2011, 10 (1): 13-36.
- [7] ARI S, HEMBRAM K, SAHA G. Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier [J]. Expert Systems With Applications, 2010, 37 (12): 8019-8026.
- [8] ABDUH Z, NEHARY E A, WAHED M A, et al. Classification of heart sounds using fractional fourier transform based Mel-frequency spectral coefficients and traditional classifiers [J]. Biomedical Signal Processing and Control, 2020, 57 (101788): 1-11.
- [9] WHITAKER B M, ANDERSON D V. Heart sound classification via sparse coding [C] //2016 Computing in Cardiology Conference (CinC), IEEE, 2016: 805-808.
- [10] QUICENO-MANRIQUE A F, GODINO-LLORENTE J I, BLANCO-VELASCO M, et al. Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiographic signals [J]. Annals of biomedical engineering, 2010, 38 (1): 118-137.
- [11] BANERJEE R, BISWAS S, BANERJEE S, et al. Time-frequency analysis of phonocardiogram for classifying heart disease [C] //2016 Computing in Cardiology Conference (CinC), IEEE, 2016: 573-576.
- [12] 曹 辉, 翁理国, 张德正. 基于密集连接空洞卷积神经网络的青藏地区云雪图像分类 [J]. 计算机测量与控制, 2019, 27 (9): 169-173.
- [13] 刘英伶, 李伟红, 龚卫国. 可变形特征图残差网络用于城市声音识别 [J/OL]. 计算机辅助设计与图形学学报, 2020, 32 (11): 1853-1862.

(下转第 222 页)