

# 基于两级神经网络的心音分割

冯正伟, 全海燕

(昆明理工大学信息工程与自动化学院, 昆明 650031)

**摘要:** 心音信号是分析诊断心脏疾病的重要信号, 而心音分割是对其进行分析处理之前必不可少的一步。本文通过将心音分割任务分离为定位与识别两个子任务, 提出一种两级卷积神经网络, 由定位网络和判别网络两级构成, 分别完成心音信号的识别与定位。首先将原始信号通过滑动窗口进行分帧, 然后通过短时傅里叶变换得到其频谱, 再通过梅尔滤波器得到其梅尔频谱系数(Mel frequency spectral coefficient, MFSC)特征, 输入第1个定位网络对其是否为心音段进行判断, 如果是的话, 再输入判别神经网络, 识别第一心音与第二心音, 从而实现心音的分割。最后利用多帧结果投票, 减小误判。同时, 在卷积神经网络中引入空间注意力机制, 实验结果表明, 这种加入了注意力机制的两级神经网络模型在心音分割任务上比使用单个卷积神经网络分类模型的准确率更高, 也使得模型更加简单, 轻量化。

**关键词:** 心音分割; 短时傅里叶变换; 梅尔倒谱; 卷积神经网络; 空间注意力机制

**中图分类号:** R318.04      **文献标志码:** A

## Heart Sound Segmentation Based on Two-Stage Neural Network

FENG Zhengwei, QUAN Haiyan

(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650031, China)

**Abstract:** Heart sound signal is an important signal for analyzing and diagnosing heart problems, and heart sound segmentation is an essential part before analyzing and processing it. By separating the heart sound segmentation task into two sub tasks of localization and recognition, this paper proposes a two-stage convolutional neural network, which is composed of localization network and discrimination network to complete the recognition and localization of heart sound signals respectively. First, the original signal is divided into frames through a sliding window, then the spectrum is obtained by short time Fourier transform, and then the Mel frequency spectral coefficient (MFSC) characteristics are obtained by Mel filter. The first localization model is input to judge whether it is a heart sound segment. If so, the discrimination neural network is input to identify the first heart sound and the second heart sound, so as to achieve heart sound segmentation. At last, multi frame voting results are used to reduce the misjudgment. At the same time, the spatial attention mechanism is introduced into the convolutional neural network. Experimental results show that this two-stage neural network model with attention mechanism has higher

accuracy in heart sound segmentation tasks than a single convolutional neural network classification model, and also makes the model more simple and lightweight.

**Key words:** heart sound segmentation; short time Fourier transform; Mel frequency spectral coefficient (MFSC); convolutional neural network; spatial attention mechanism

## 引 言

心血管疾病是我国最为常见的一类疾病,也是中老年人死亡率最高的疾病。心音信号由于包含大量的心脏信息,能充分反映出心脏的生理状态,如心脏瓣膜的工作状态、血液流动情况等,因此对于诊断分析心血管疾病有着很大的帮助<sup>[1]</sup>。心音信号每隔相同时间便会反复出现,该时间间隔叫做一个心动周期。每个心动周期包含第一心音(S1)、收缩期、第二心音(S2)、舒张期4个心音成分<sup>[2]</sup>,心脏瓣膜周期性的打开与关闭、心房与心室壁的收缩与舒张、血管的扩张以及狭窄引发的压力都是产生心音的原因,心音的精确分割是对心音进行后续分析的基础,因此对心音信号进行分割具有重要的研究意义。

传统的方法常用心音的包络来进行分割<sup>[3]</sup>,如郭兴明等<sup>[4]</sup>用希尔伯特变换提取心音的包络,但是这种方法存在太多毛刺,不利于后期分割。人们利用心音时间上的联系,使用马尔可夫模型来进行分割,如奎皓然等<sup>[5]</sup>采用隐马尔可夫模型来对心音的4个状态进行分割。针对使用希尔伯特变换求取的包络存在不够光滑的问题,许春冬等<sup>[6]</sup>使用香农包络来解决包络的毛刺问题,再使用隐半马尔可夫模型(Hidden semi-Markov model, HSMM)来识别心音状态。也有研究人员采用机器学习算法进行学习,如Moukadem等<sup>[7]</sup>提出了基于心音S矩阵奇异值分解(Singular value decomposition, SVD)提取心音特征,并使用k最近邻(k-nearest neighbor, KNN)算法对其进行分类。但KNN算法需要将待测样本与数据库中全部样本都计算一次距离,因此当数据库较大时,计算力要求太高,对于实时处理不利。赵湛等<sup>[8]</sup>利用心音信号的周期特点,利用其自相关函数来对心音进行分割,但是信号中的噪声对算法分割效果的影响很大。

近年来,深度学习的方法被用到各种研究当中,研究人员也开始用深度学习来对心音信号进行分割。侯雷静<sup>[9]</sup>提取心音的梅尔倒谱系数(Mel-frequency cepstral coefficients, MFCCs),再使用深度神经网络(Deep neural network, DNN)对S1和S2进行识别,准确率达到88.63%。Chen等<sup>[10]</sup>先提取出心音信号的MFCC特征,然后使用K-means聚类算法将每一个心音的MFCC特征汇聚分类,计算每个类的中心向量,将两个中心向量组合为一个超向量,然后将此超向量用于训练神经网络,进而对心音中的S1和S2进行分类,该模型准确率可达到91%。随着卷积神经网络(Convolutional neural network, CNN)的兴起,研究人员也开始用其对心音的时频谱进行处理分析<sup>[11-12]</sup>,如申玉静等<sup>[13]</sup>将心音信号中的S1与S2提取出来进行识别,用其时频谱作为特征,然后采用卷积神经网络进行分类,该方法计算简单,但是需要将第一心音与第二心音提出来进行识别。文献[14-15]将长短时间记忆网络与卷积神经网络结合起来处理,更好地利用了心音在时间上的关联性,但是也使模型变得臃肿。文献[16]提出通过结合经验模型分解(Empirical mode decomposition, EMD)与小波分解的方法实现心音的降噪,但缺点是运行用时太长,不利于实时分析。文献[17]提出对心音进行降采样操作,然后再使用离散小波阈值对降采样之后的信号进行降噪,该处理可以有效去除心音信号中的高频噪声。

在以上研究的基础上,本文提出了一种两级卷积神经网络。首先将心音信号通过小波阈值去噪,再用滑动窗口将原始信号分割为等长的片段,之后再提取片段的梅尔频谱系数(Mel frequency spectral coefficient, MFSC)特征作为卷积神经网络的输入。通过将心音分割分解为定位与识别两个任务,训练

两个分支网络分别对应心音定位和第一心音与第二心音识别两个任务,用定位网络判断当前帧含有心音的概率,没有超过阈值的话说明当前为收缩期或者舒张期,超过阈值的话再用第二个判别卷积神经网络对第一心音和第二心音进行识别,最后使用多帧结果投票减少判别模型对第一心音和第二心音的误判。实验结果表明,该模型在保证高分割准确率的同时,有效减小了模型参数规模。

## 1 实验方法

### 1.1 算法流程

本文算法流程如图1所示,主要步骤包括预处理、分帧、提取MFSC特征、心音定位、第一心音与第二心音识别和多帧结果融合。

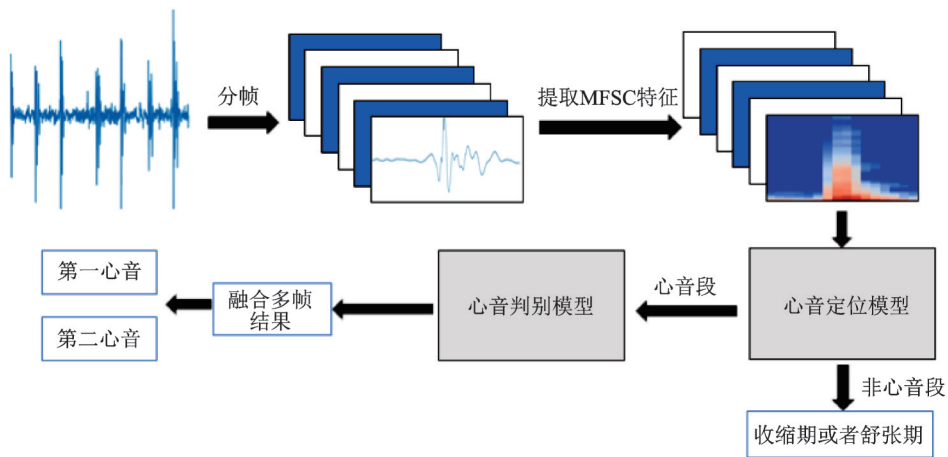


图1 算法流程图

Fig.1 Algorithm flow chart

### 1.2 心音信号预处理

在对心音信号进行识别分割之前,针对原始信号长短不一,并且还存在噪声干扰等问题,需要对其进行预处理。

#### 1.2.1 归一化

由于心音信号的幅值之间差异过大,对于神经网络来说波动过大的输入会降低其特征提取能力,降低收敛速度,所以需要对其进行归一化,使得输入数据更加规范,加快模型收敛。对于心音分割而言,主要信息集中在第一心音与第二心音,而两者幅值差异不大,因此本实验采用的是绝对值归一化,即

$$x'(n) = \frac{x(n)}{\max|x(n)|} \quad (1)$$

式中: $x(n)$ 为原始信号; $x'(n)$ 为归一化之后的信号。

#### 1.2.2 滤波

心音信号中存在着许多噪声,如呼吸音、血流声和外部杂音等,为了避免噪声对后续处理的影响,需要对其进行滤波,经过之前的研究,本文采用离散小波降噪<sup>[18]</sup>。选用db 8小波基函数,采用硬阈值法,通过将信号进行五层小波分解,再将其中小于设定阈值的分量去掉,再进行小波重构,可以在保留有用信号的同时去除掉心音信号中的大部分噪声,最终去噪声效果如图2所示。

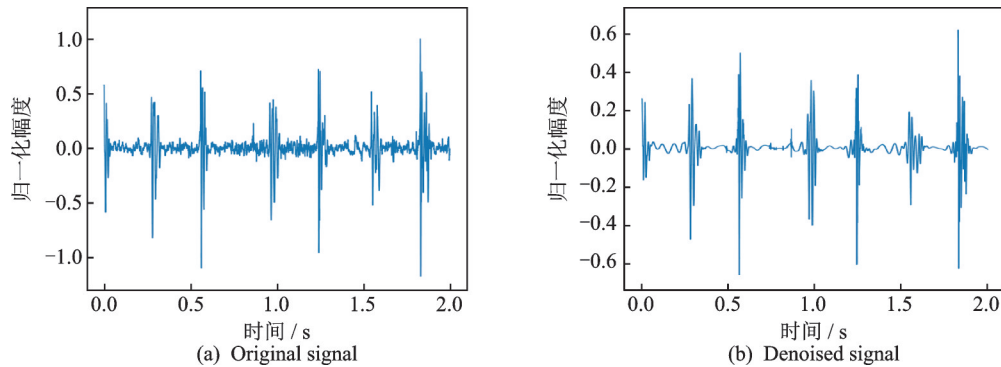


图2 小波降噪

Fig.2 Wavelet denoising

### 1.3 分帧与 MFSC 特征提取

MFCC 特征是心音研究中经常使用的特征,而 MFSC 特征是去除了离散余弦变换的 MFCC,保留了更多的局部特征,因此本实验采用其作为输入特征。为了提取心音的 MFSC 特征,本实验通过窗长为 125 ms,帧移为 25 ms 的滑动窗口对原始信号进行分帧,这个窗口长度可以完整地包含第一心音或者第二心音,以便于后期处理。对于上面得到的 125 ms 的块,其分帧后可近似看为平稳信号,对其使用短时傅里叶变换(Short time Fourier transform, STFT),其包括分帧、加窗、快速傅里叶变换等步骤,具体可表示为

$$X(i, k) = \sum_{m=-\infty}^{+\infty} x(m)w(i-m)e^{-jkm} \quad (2)$$

式中: $x(m)$ 为输入信号; $w(m)$ 为窗函数; $X(i, k)$ 为得到的时频谱。

利用 STFT 得到的信号时频谱在频域上分布不饱满,将其通过梅尔滤波器组滤波之后,可得到更饱满且符合人耳听觉的 MFSC 特征。在这之前,需要得到信号的功率谱,其计算方法为

$$E(i, k) = |X(i, k)|^2 \quad (3)$$

得到功率谱之后,计算其梅尔能量,计算过程为

$$S(i, m) = \sum_{k=1}^n E(i, k)H_m \quad (4)$$

式中: $H_m$ 为第  $m$  个梅尔滤波器的频率响应; $m$ 为梅尔滤波器的数量,一般为 26~40。

将得到的梅尔能量取对数组合便可得到 MFSC 特征。在实验中,原始心音信号的采样率为 8 k,实验中 STFT 采用的是 256 点快速傅里叶变换,帧移为 64 点,窗函数为汉明窗,梅尔滤波器数量为 40,最后得到的谱图如图 3 所示。

### 1.4 两级神经网络

#### 1.4.1 空间注意力

本实验的两个网络模型使用了 CBMA 注意力机制<sup>[19-20]</sup>中的空间注意力机制。所谓注意力机制就是使模型更加关注局部信息,对于空间注意力机制

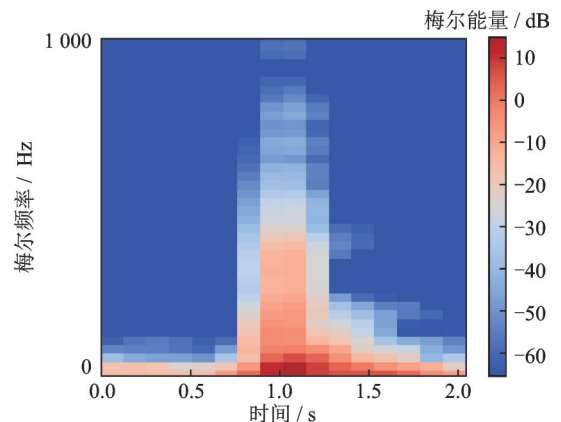


图3 心音 MFSC 特征

Fig.3 MFSC characteristics of heart sounds

来说,其目的就是为了模型聚焦于局部的像素点,而心音信号在频谱上的信息分布也是局部的,因此本实验采用空间注意力机制,使模型更加关注心音信号频谱中 useful 信息的区域。空间注意力是利用同一像素的所有通道的信息来计算注意力权重。本实验采用输入的均值和最大值两个特征来计算空间注意力权重,最后会生成一个与输入特征相同形状的掩码矩阵。矩阵中每一个元素代表了输入特征与其对应位置的像素的注意力

力,该权重会反向传播自动学习,也就是由网络自动学习聚焦于哪些地方, CBMA 空间注意力的网络模型如图 4 所示。

图 4 中掩码矩阵的具体计算方式如式(5)所示。

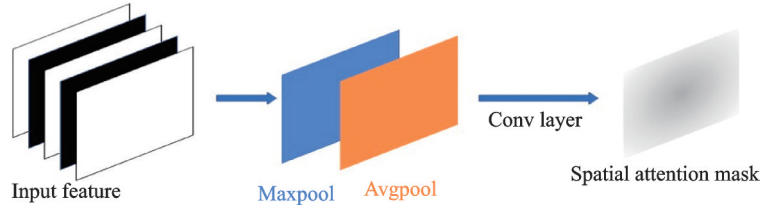


图4 CBMA 空间注意力网络模型

Fig.4 CBMA spatial attention network model

$$\begin{cases}
 \mathbf{X}_I = \begin{pmatrix} x_{11} & \cdots & x_{1w} \\ \vdots & \ddots & \vdots \\ x_{h1} & \cdots & x_{hw} \end{pmatrix}_{\times n} \\
 m_{ij} = \frac{1}{n} \sum_{c=1}^n x_{ij}(c) \\
 M_{ij} = \max(x_{ij}(1), \dots, x_{ij}(n)) \\
 w_{ij} = f(w_{\text{mean}} \cdot m_{ij} + w_{\text{max}} \cdot M_{ij} + b) \\
 \mathbf{W} = \begin{pmatrix} w_{11} & \cdots & w_{1w} \\ \vdots & \ddots & \vdots \\ w_{h1} & \cdots & w_{hw} \end{pmatrix} \\
 \mathbf{X}_O = \mathbf{W} \odot \mathbf{X}_I
 \end{cases} \quad (5)$$

式中: $f$ 为 Relu 激活函数; $n$ 为样本的输入特征通道数; $x_{ij}(c)$ 为第  $c$  个通道  $i$  行  $j$  列的元素; $m_{ij}$ 为每一元素所有通道的均值; $M_{ij}$ 为每一元素所有通道的最大值; $\mathbf{W}$ 为空间注意力权重矩阵; $\odot$ 表示矩阵对应元素相乘。

1.4.2 定位网络

定位网络输入心音的 MFSC 特征,输出为输入含有心音的概率,其模型架构如图 5 所示。

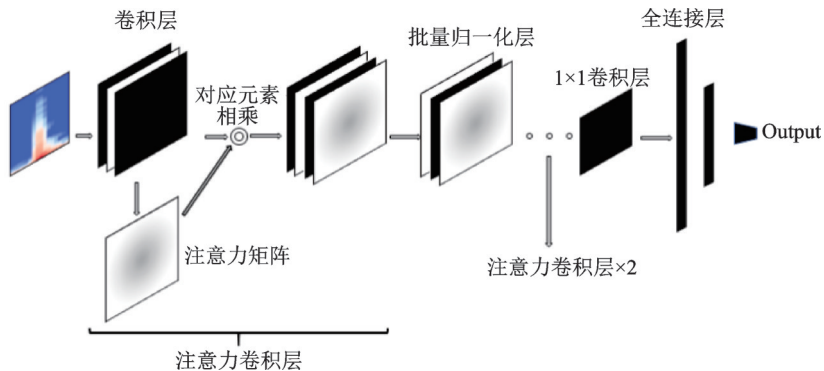


图5 定位模型

Fig.5 Position model

定位模型主要由注意力卷积层构成,其中卷积层包括16个 $3 \times 3$ 的卷积核,步长为1。卷积层提取特征之后,通过注意力机制聚焦有用特征,再用池化层进行特征压缩,之后重复一次注意力卷积层和池化层,最后输出层使用两层全连接层来分类,输出使用的是Sigmoid激活函数,其输出为输入含有心音的概率。

#### 1.4.3 判别网络

判别网络输入心音的MFSC特征,判别其为第一心音还是第二心音,其模型架构如图6所示。

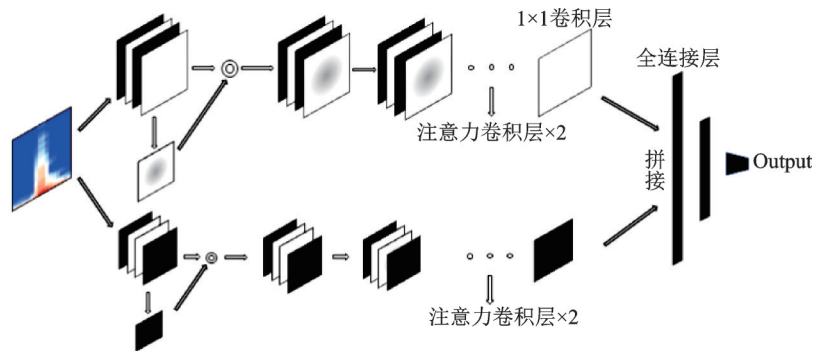


图6 判别模型

Fig.6 Discriminator model

在判别网络中,为了提取更多细节特征,卷积层分别使用了16个 $3 \times 3$ 的卷积核与16个 $2 \times 2$ 的卷积核两种大小的卷积核提取特征,最后再将特征拼接。在卷积层之后,为了保留细节特征,没有使用池化层,换为批量归一化层。在全连接层之前,使用 $1 \times 1$ 卷积,将通道数减少为1,可以有效减小模型参数数量。最后的激活函数使用Softmax激活函数,其输出为第一心音与第二心音的概率。

#### 1.4.4 心音定位

使用定位网络判断前帧是否为心音段之后,若当前帧为心音段,需要计算心音位置用于分割。

在分帧之后,由于每帧存在交叠部分,相邻的几帧都会被预测为心音段,因此本实验取相邻心音帧的中间帧的中心位置 $P$ 作为最后输出的心音位置,假设当前为第 $n$ 帧,则帧中心位置 $P$ 计算方式为

$$P = n \times s + \frac{l_{win}}{2} \quad (6)$$

式中: $l_{win}$ 为滑动窗口长度; $s$ 为滑动窗移。

#### 1.4.5 多帧结果融合

在使用心音判别模型判定时,相邻几帧包含的是同一个心音,输出结果本该一致。但是由于第一心音与第二心音特征相近,模型很可能会出现误判,因此本实验采用投票机制,选择相邻多帧中预测结果中占多数的判断结果为最后的输出,这可以有效减少模型对于第一心音与第二心音的误判。其具体的计算流程如图7所示。

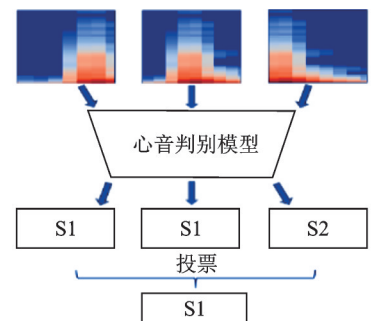


图7 投票机制

Fig.7 Voting mechanism

## 2 实验与结果分析

### 2.1 数据集及其预处理

本实验采用Classifying Heart Sound Challenge中的第一个心音分割挑战赛数据集(<http://www.peterjbentley.com/heartchallenge>),其提供了临床真实情况下采集到的320个心音数据,并提供部分心音位置的标注,共有1472个心音标注,以csv格式文件保存。原始数据

的采样率为 8 kHz,为减小数据量,降采样为 4 kHz,然后采用滑动窗口方法对初始数据进行分帧。窗口大小为 125 ms,帧移为 25 ms,经过分帧后,一共得到了 6 047 个样本,其中包含有心音数据(S1与S2)的共 3 462 个样本。将上述样本按照 8:2 的比例分割为训练集与测试集。对于心音判别模型,本实验以心音位置为中心,取 125 ms 的帧长,得到 1 417 个样本,为了增加样本的普适性,在其心音中心左右 25 ms 范围内取随机值作为新的样本中心,在保证完整包含心音的同时,随机移动心音的位置,可以有效提升模型的泛化性能,再得到 2 737 个样本,共 4 154 个样本作为心音判别网络的数据集。

## 2.2 实验设置

实验中,两级网络中的两个模型使用不同的数据集训练,对于定位模型,batch\_size 设置为 512,学习率为 0.000 02,优化算法为 Adam 优化算法,迭代次数 25 轮;对于心音判别模型,batch\_size 设置为 256,学习率为 0.000 2,优化算法同样为 Adam 优化算法,迭代次数为 50 轮,最后将两个预训练模型组合为两级网络模型,再使用第 3 个数据集进行测试。以上实验内容均是基于 Window10 操作系统,Python 平台,Pytorch 深度学习框架上进行,计算机主要硬件配置如下:内存 DDR4-16 GB,固态硬盘 1 TB,CPU 为 Intel-i7-10875H,GPU 为 NVIDA-RTX-2060。

## 2.3 结果分析

本实验采取准确率(Accuracy)、精确率(Precision)和召回率(Recall)指标对实验结果进行评估,其计算方式分别为

$$\begin{cases} \text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \end{cases} \quad (7)$$

式中:TP 表示正样本被预测为正;FN 表示正样本被预测为负;TN 表示负样本预测为负;FP 表示负样本被预测为正。

实验共对比了以下 6 种算法:全连接网络(FCN)、单模型三分类卷积神经网络(无空间注意力)(CNN 1)、单模型三分类卷积神经网络(含空间注意力)(CNN 2)、单模型三分类残差连接卷积神经网络(含空间注意力)(Resnet)、两级卷积神经网络(无空间注意力)(Two-stage-CNN 1)和两级卷积神经网络(含有空间注意力)(Two-stage-CNN 2)。具体结构对比如表 1 所示,其中 Linear 为全连接层,Conv 为卷积层,Sa-cov 为注意力卷积层,Res-cov 为残差连接卷积层。

表 1 算法对比

Table1 Algorithm comparison

算法结构	FCN	CNN1	CNN 2	Resnet	Two-stage CNN 1		Two-stage CNN 2	
					Net 1	Net 2	Net 1	Net 2
Layer 1	Linear	Conv	Sa-cov	Res-cov	Conv	Conv	Sa-cov	Sa-cov
Layer 2	Linear	Conv	Sa-cov	Res-cov	Conv	Conv	Sa-cov	Sa-cov
Layer 3	Linear	Conv	Sa-cov	Res-cov	Conv	Conv	Sa-cov	Sa-cov
Layer 4	Linear	Conv	Sa-cov	Res-cov	Linear	Linear	Linear	Linear
Layer 5	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Layer 6	Linear	Linear	Linear	Linear				

将表1中的6种方法依次进行实验。两级卷积神经网络模型按照2.2节所述进行训练,其他4种方法均采用第一个数据集训练,测试均采用第一个测试数据集,评估标准采用每一类分类结果的平均值。6种方法的实验效果对比如表2所示。

表2 算法性能对比

Table 2 Algorithm performance comparison

评估指标	FCN	CNN 1	CNN 2	Resnet	Two-stage-CNN 1	Two-stage-CNN 2
Accuracy	71.41	79.21	81.02	80.74	90.78	95.56
Precision	69.12	76.62	78.51	77.65	89.45	94.32
Recall	74.13	72.36	76.12	76.12	92.41	96.56

由表2得出,全连接网络(FCN)在所有模型中效果最差,主要原因是分帧得到的心音位置在每一帧中不一样,特征区域不固定,导致全连接网络难以学习到特征,而卷积神经网络在提取特征时具有平移不变性,刚好适合这类任务。单个卷积神经网络(CNN)进行三分类效果尽管较全连接网络有提升,但是很有限。原因可能在于除去第一心音与第二心音之外,第三类成分复杂,如含有杂音,或者部分心音等,情况复杂,这就导致三分类模型难以学习到其特征。残差连接卷积神经网络效果与卷积神经网络没有太多差异,该网络模型本是为了解决网络太深时的梯度消失问题,而本实验使用的模型层数不深,因此其相对于卷积神经网络提升不大。本文提出的两级神经网络将任务进行分离,定位与分离由两个模型分别完成,定位模型只需定位心音段,判别模型只需学习S1与S2的二分类任务,有效降低了任务复杂度,因此各项性能指标都提升了10%左右。同时,空间注意力机制的应用进一步使准确率提升了4.22%。

为了测试网络参数对性能的影响,本文对卷积神经网络的不同参数之间也进行了对比,结果如表3所示。

表3 定位模型参数对比

Table 3 Comparison of position model parameters

定位模型参数	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Sa-conv层数量	1	2	2	2	3	3
Dense层数量	1	1	2	2	2	2
卷积核大小	3×3	3×3	3×3	2×2	2×2	3×3
池化层数量	2	2	2	2	2	2
Accuracy/%	92.42	93.28	96.84	95.56	91.63	89.67
Precision/%	93.89	92.56	97.68	95.78	89.65	88.23
Recall/%	92.98	91.23	96.21	96.48	92.14	90.12

从表3可以看出,Model 1与Model 2对比,最后的全连接层的数量使用两层效果更好,其拥有更好的拟合能力;Model 4与Model 5对比得出,模型的空间注意力卷积层不是越深越好,实验结果在两层左右得到最好的效果,原因是MFSC的输入特征尺寸不大,仅有 $40 \times 16$ ,且使用了池化层压缩特征,太深的网络会导致特征图不断变小,识别效果降低。

心音判别模型的参数分析如表5所示,通过Model 1和Model 2对比得出,加入池化层会使模型性能降低。其原因是对于判别模型而言,使用池化层会使得部分信息丢失,而第一心音与第二心音很多特征类似,细节信息减少会导致模型识别能力下降,因此在判别模型中移除了池化层,改用批量归一化



层。Model 2与Model 3对比得出,使用32个卷积核比16个卷积核在精确率和准确率上都有提升,更多的卷积核能使模型提取更多细节特征。对比Model 4与Model 5得出,相对于单尺度卷积核,多尺度的卷积核能提取更多特征。

表4 判别模型参数对比

Table 4 Comparison of discriminator model parameters

判别模型参数	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Sa-conv层数量	1	1	1	2	3	3
Dense层数量	2	2	2	2	2	2
卷积核数量	16	16	32	32	32	32
卷积核大小	3×3	3×3	3×3	3×3,2×2	3×3,2×2	3×3,2×2
池化层数量	1	0	0	0	0	0
Accuracy/%	78.43	82.26	88.12	93.12	<b>95.12</b>	79.12
Precision/%	76.12	81.25	85.12	96.52	<b>97.32</b>	80.12
Recall/%	78.98	80.43	86.13	<b>92.13</b>	91.80	82.13

在心音判别模型之后,本文对比使用了多帧结果投票与单帧预测结果,其结果如表5所示。相对于只使用单帧的结果,在使用多帧结果投票之后,准确率提升了2个百分点,召回率提升了3个百分点。由于投票机制结合了相邻多帧的信息,因此有效减少了心音判别模型对第一心音与第二心音的误判。

表5 投票机制

Table 5 Volting mechanism

评估指标	单帧结果	多帧结果投票
Accuracy	95.96	97.98
Recall	94.32	97.24
Precision	95.56	95.94

## 2.4 心音分割效果

训练好的两级网络模型的实际分割效果如图8所示,其中Systole为收缩期,Diastole为舒张期。从图中可以看出两级网络模型不仅能较好地分割出心音心动周期的4个时期,而且能较为准确地识别出第一心音与第二心音。但是在分割边缘的精度上还有待提高,主要原因是本方法的分割精度是由分帧的帧移来决定的,而帧移的精度有限,这也是后续需要解决的问题。

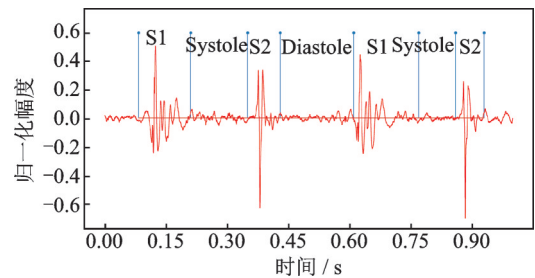


图8 实验结果图

Fig.8 Experimental results

## 3 结束语

目前针对心音分割的深度学习方法多采用深层网络模型,网络参数量庞大,不利于在实际中部署。本研究通过将心音分割任务分解为定位与判别两个子任务,以此分别构建两个更加简单的模型,并采用投票机制,融合相邻几帧的信息,有效减少了模型误判。实验结果表明,这种方法相对普通卷积神经网络模型在平均准确率上提升了10%,在添加了注意力机制之后进一步提升了4.4%。此外,两级网络模型参数规模更小,推理速度快,更有利于在实际临床条件下进行部署使用。在未来的工作中,希望能解决分帧导致的定位精度不足的问题,同时能对心音分割之后的数据进行进一步的处理,通过深度学习的方法进行降噪、识别,以此协助医护人员更好地诊断疾病。

## 参考文献:

- [1] 巩燕,胡杰,高彬,等.心血管疾病即时检测技术的研究进展[J].中国科学:技术科学,2016,46(11):1116-1134.  
GONG Yan, HU Jie, GAO Bin, et al. Advances in monitoring of cardiovascular diseases at the point of care[J]. *Scientia Sinica Technologica*, 2016, 46(11): 1116-1134.
- [2] 侯雷静,郭婷婷,孙燕,等.面向心音分割的个性化高斯混合建模方法[J].声学学报,2019,44(1):20-27.  
HOU Leijing, GUO Tingting, SUN Yan, et al. Personalized Gaussian mixture modeling method for heart sound segmentation [J]. *Acta Acustica Sinica*, 2019, 44 (1): 20-27.
- [3] CHEN H, YUAN X, LEE J, et al. Automatic multi-level in-exhale segmentation and enhanced generalized S-transform for wheezing detection[J]. *Computer Methods and Programs in Biomedicine*, 2019, 178(12): 163-173.
- [4] 郭兴明,蒋鸿,郑伊能.基于改进的维奥拉积分方法提取心音信号包络[J].仪器仪表学报,2016,37(10):2352-2358.  
GUO Xingming, JIANG Hong, ZHENG Yineng. Heart sound signal envelope extraction based on modified Viola integral method[J]. *Chinese Journal of Scientific Instrument*, 2016, 37(10): 2352-2358.
- [5] 奎皓然,潘家华,容容,等.基于持续时间隐马尔可夫模型的心音分割算法[J].生物医学工程学杂志,2020,37(5):765-774.  
KUI Haoran, PAN Jiahua, ZONG Rong, et al. Heart sound segmentation algorithm based on duration hidden Markov model [J]. *Journal of Biomedical Engineering*, 2020, 37 (5): 765-774.
- [6] 许春冬,林海.结合SVM和香农能量的HSMM心音分割算法[J].数据采集与处理,2021,36(5):950-959.  
XU Chundong, LIN Hai. HSMM heart sound segmentation algorithm based on SVM and Shannon energy[J]. *Journal of Data Acquisition and Processing*, 2021, 36 (5): 950-959.
- [7] MOUKADEM A, DIETERLEN A, HUEBER N, et al. A robust heart sounds segmentation module based on S-transform[J]. *Biomed Signal Proces*, 2013, 8: 273-281.
- [8] 赵湛,张旭茹,方震,等.基于心动周期估计的心音分割及异常心音筛查算法[J].电子与信息学报,2017,39(11):2677-2683.  
ZHAO Zhan, ZHANG Xuru, FANG Zhen, et al. Heart sound segmentation and abnormal heart sound screening algorithm based on cardiac cycle estimation[J]. *Journal of Electronics and Information*, 2017, 39 (11): 2677-2683.
- [9] 侯雷静.心音信号的周期分析与状态分割[D].北京:中国科学院大学,2018.  
HOU Leijing. Periodic analysis and state segmentation of heart sound signal[D]. Beijing: University of Chinese Academy of Sciences, 2018.
- [10] CHEN T E, YANG S I, HO L T, et al. S1 and S2 heart sound recognition using deep neural networks[J]. *IEEE Transactions on Biomedical Engineering*, 2017, 64(2): 372-380.
- [11] SHAUGHNESSY D O. Recognition and processing of speech signals using neural networks[J]. *Circuits, Systems, and Signal Processing*, 2019, 38: 3454-3481.
- [12] RENNA F, OLIVEIRA J, COIMBRAM T. Deep convolutional neural networks for heart sound segmentation[J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(6): 2435-2445.
- [13] 申玉静,王寻,唐闽,等.使用二分支卷积神经网络识别第一心音与第二心音[J].生物医学工程学杂志,2021,38(1):138-144,153.  
SHEN Yujing, WANG Xun, TANG Min, et al. Recognition of first heart sound and second heart sound using two branch convolutional neural network[J]. *Journal of Biomedical Engineering*, 2021, 38(1): 138-144, 153.
- [14] 刘伟伟,桑胜波,张宏鹏.基于CNN+LSTM的改进心音分类模型研究[J].电子设计工程,2022,30(2):38-42.  
LIU Weiwei, SANG Shengbo, ZHANG Hongpeng. Research on improved heart sound classification model based on CNN+LSTM[J]. *Electronic Design Engineering*, 2022, 30 (2): 38-42.
- [15] 李学日.基于双通道卷积长短时记忆神经网络的肺音识别方法研究[D].广州:广东技术师范大学,2021.  
LI Xueri. Research on lung sound recognition method based on two channel convolution long and short-term memory neural network[D]. Guangzhou: Guangdong Polytechnic Normal University, 2021.
- [16] 李伟,江晓林,陈海波,等.基于EEMD\_Hankel\_SVD的矿山微震信号降噪方法[J].煤炭学报,2018,43(7):1910-1917.  
LI Wei, JIANG Xiaolin, CHEN Haibo, et al. Denoising method of mine microseismic signal based on EEMD\_Hankel\_SVD[J]. *Journal of China Coal Society*, 2018, 43 (7): 1910-1917.

- [17] JAIN P K, TIWARI A K. An adaptive thresholding method for the wavelet based denoising of phonocardiogram signal[J]. Biomedical Signal Processing and Control, 2017, 38(1): 388-399.
- [18] 陈娟. 基于小波阈值法的心音信号降噪处理[J]. 内蒙古科技与经济, 2021(20): 104-106.  
CHEN Juan. Heart sound signal denoising based on wavelet threshold method[J]. Inner Mongolia Science Technology & Economy, 2021 (20): 104-106.
- [19] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of European Conference on Computer Vision (ECCV). Munich, Germany: [s.n.], 2018: 3-19.
- [20] 张俊飞, 张贵英. 基于频域角度改进CBAM注意力机制[J]. 自动化与仪器仪表, 2021(9): 87-90, 95.  
ZHANG Junfei, ZHANG Guiying. Improving CBAM attention mechanism based on frequency domain[J]. Automation and Instrumentation, 2021 (9): 87-90, 95.

作者简介:



冯正伟(1999-),男,硕士研究生,研究方向:心音信号处理, E-mail: 2773063343@qq.com。



全海燕(1970-),通信作者,男,博士,副教授,硕士生导师,研究方向:智能信号处理, E-mail: quanhaiyan@163.com。

(编辑:王静)