

一种通过增强图像编码和非对称卷积网络的心音分类算法

王晟懿¹ 杨宏波² 潘家华² 王威廉¹

1 云南大学信息学院 昆明 650504

2 云南省阜外心血管病医院 昆明 650102

(2814604624@qq.com)

摘要 文中提出了一种通过增强图像编码和非对称卷积网络的心音分类算法。与传统的基于统计特征和时频域特征提取心音的方法不同,该算法通过引入分数阶傅里叶变换(FrFT)分别对格拉姆角场(GAF)、马尔可夫场(MTF)、递归图(RP)3种图像编码方法进行增强,构成FrFT-GAF,FrFT-MTF,FrFT-RP图像编码模块。运用上述图像编码模块将一维心音信号转换为二维编码特征图,并利用计算机视觉技术在分类任务中的优势,采用非对称卷积网络(ACNet)对心音的二维编码特征图进行分析处理,从而实现于心音的有效分类。此外,还分别对上述图像编码模块的性能进行了评估和比较。实验结果表明,在心音二分类任务中,FrFT-RP模块的分类效果最好,在数据集1和数据集2(Physio Net/CinC 2016数据集)上的准确率分别为0.981和0.977,F1分别为0.989和0.974。FrFT-MTF和FrFT-GAF模块的效果次之。使用FrFT增强图像编码特征后较以往方法有明显提升,为心音信号分类提供了新的思路和方法,有望应用于先天性心脏病辅助诊断。

关键词:先天性心脏病;心音;图像编码;分数阶傅里叶变换;非对称卷积网络

中图分类号 TN912

Heart Sound Classification Algorithm Using Enhanced Image Coding and Asymmetric Convolutional Networks

WANG Shengyi¹, YANG Hongbo², PAN Jiahua² and WANG Weilian¹

1 School of Information Science and Engineering, Yunnan University, Kunming 650504, China

2 Kunming Medical University Affiliated Cardiovascular Hospital, Kunming 650102, China

Abstract This paper proposes a heart sound classification algorithm using enhanced image coding and asymmetric convolutional networks. Unlike traditional methods that extract heart sounds based on statistical features and time-frequency domain features, this algorithm enhances three image coding methods—Gramian angular field (GAF), Markov transition field (MTF), and recurrence plot (RP)—by introducing fractional Fourier transform (FrFT), which constitutes the image coding modules of FrFT-GAF, FrFT-MTF, and FrFT-RP, respectively. The one-dimensional heart sound signal is transformed into a two-dimensional encoded feature map using these image coding modules. An asymmetric convolutional network (ACNet) leverages computer vision advantages to analyze and process the two-dimensional encoded feature map for effective heart sound classification. In addition, the performance of the above image coding modules is evaluated and compared respectively. Experimental results demonstrate that the FrFT-RP module achieves the best classification performance in binary heart sound classification tasks, with an accuracy of 0.981 and 0.977, and F1 score of 0.989 and 0.974 on dataset 1 and dataset 2 (Physio Net/CinC 2016 dataset), respectively. The FrFT-MTF and FrFT-GAF modules show effective performance in that order. The performance of the method using FrFT to enhance image encoding features has significantly improved compared to previous methods, providing novel approaches and methods for heart sound signal classification, is expected to be applied in machine assisted diagnosis of congenital heart disease.

Keywords Congenital heart disease, Heart sound, Image coding, Fractional Fourier transform, Asymmetric convolutional network

1 引言

先天性心脏病 (Congenital Heart Disease, CHD) 简称先心病,是一种对青少年儿童身体健康构成严重威胁的心血管疾病,在全国多地均位居新生儿出生缺陷首位^[1]。CHD的初诊主要依靠专业医生进行心脏听诊,需要丰富的专业知识和临床经验,在早期大规模筛查中难以实现。因此,迫切需要

对心音图 (Phonocardiogram, PCG) 进行自动分析和分类,以克服人工听诊主观性的弊端,实现 CHD 的机器辅助诊断。

心音分类算法研究一般包括预处理、特征提取、分类识别3个步骤。其中,心音信号特征提取的丰富度是心音分类算法的关键,寻找高效且表征能力强的特征向量或特征图一直是研究人员关注的重点。常见的特征提取方法主要包括从时域、频域以及时频域中提取特征。时域特征主要是提取信号

基金项目:国家自然科学基金(81960067);云南省重大科技专项基金(2018ZF017)

This work was supported by the National Natural Science Foundation of China(81960067) and Major Science and Technology Projects of Yunnan Province in 2018(2018ZF017).

通信作者:王威廉(wlwang_47@126.com)

包络,包括希尔伯特包络^[2]、香农能量包络^[3]、同态包络^[4]等,以及定位 S1 和 S2 的位置,提取收缩期与舒张期持续时间的平均值和标准差^[5]、短时平均幅度差^[6]、信号能量^[7]等。频域特征指对信号进行频谱分析,通过数学变换将信号从时域转换到频域,如傅里叶变换^[8]、小波变换^[9]、经验模态分解^[10]等,从而描述信号在频率方面的特性。时频分析主要借鉴了语音处理方法,包括短时傅里叶变换(Short-Time Fourier Transform, STFT)^[11]、线性预测倒谱系数(Linear Prediction Cepstral Coefficients, LPCC)^[12]、梅尔倒谱系数(Mel-Frequency Cepstrum Coefficient, MFCC)^[13]、梅尔频谱系数(Mel-Frequency Spectral Coefficient, MFSC)^[14]、Wigner-Ville 分布^[15]、Bark 倒谱系数(Bark-Frequency Cepstrum Coefficient, BFCC)^[16]、动态 ERB 倒谱系数(Dynamic ERB-scale Frequency Cepstral Coefficients, DEFCC)^[17]等。

上述特征提取方法在心音分类研究中取得了一定成果,由于心音信号是典型的非线性非平稳信号,传统的对心音信号时频域的映射和分析存在一定的局限性,无法充分捕捉心音信号的分量和变化趋势。已有研究提出将时间序列数据转换为图像形式,然后利用计算机视觉技术进行有效的特征提取和分类,在生物医学信号处理、金融时间序列分析、语音识别等多个领域中都有广泛的应用。1987年, Eckmann 等^[18]提出了递归图(Recurrence Plot, RP),其是一种适用于非线性动力系统的可视化工具,能够揭示数据中的隐藏动态模式、非线性结构和非平稳性,被广泛应用于物理学、气象学、生物医学等领域。2015年, Wang 等^[19]提出格拉姆角场(Gramian Angular Field, GAF)和马尔可夫场(Markov Transition Field, MTF),旨在将时间序列数据编码为不同类型的图像,以促进计算机视觉方法在时间序列分类任务中的应用,现已在轴承故障诊断、金融、脑心电图分析等多个领域取得了显著的成果。

GAF, MTF, RP 虽然在特征提取的侧重点和具体实现上有所不同,但三者都是将时间序列转换成矩阵,以图像的形式表示时间序列中的模式和特征,从而进一步进行图像处理和分类。GAF 和 MTF 最初被用于按时间索引的单变量序列数据中, Zhou 等^[20]将 GAF 和 MTF 扩展到以频域为索引的单变量序列数据中,其首先对心音信号进行傅里叶变换得到频谱,然后生成 GAF 和 MTF 图像,该方法在小儿心音多分类中取得了一定的效果,但其对心音信号的表征仍然不够全面,丢失了时域部分的信息。为此,本文创新性地引入了分数阶傅里叶变换(Fractional Fourier Transform, FrFT)分别对 GAF, MTF, RP 图像编码进行增强。与傅里叶变换相比, FrFT 描述了信号的局部时频变化,使心音信号的能量集中在特定的时频区域,在处理非平稳信号时对噪声较为敏感,抗干

扰性能强,从而提升心音信号特征提取的效果和分类性能。

在计算机视觉任务中,尤其是在图像分类识别任务中,卷积神经网络(Convolutional Neural Network, CNN)早已发展成为一种核心且高效的方法,其通过多层卷积层自动学习并提取图像特征以实现分类。例如, Riccio 等^[21]利用分区迭代函数系统提取心音信号特征,然后使用 CNN 进行分类,其准确率为 85%。Nguyen 等^[22]提取对数-梅尔频谱图特征,使用长短时记忆网络(Long Short-Term Memory, LSTM)和 CNN 对心音进行分类。Ranipa 等^[23]融合 MFCC 和广义频域特征,提出一种多模态注意力网络 MACNN,在 PhysioNet/CinC 2016 数据集上的准确率达到 0.955。但上述算法的普适性仍有待验证,文献[21]在对信号进行转换时对输入信号强加了固定长度,如果信号长于或短于预定长度则必须进行截断或复制,且其分类的准确率还有待提高。文献[22]未使用 PhysioNet/CinC 2016 公开数据集进行验证,其泛化性能还需要进一步评估。文献[23]提出的 MACNN 虽然参数量相对较少,但需要训练多个 CNN 特征提取器进行特征融合,增加了模型的计算复杂度。

基于以上情况,本文提出了一种通过增强图像编码和非对称卷积网络的心音分类算法。通过引入 FrFT 分别对 GAF, MTF, RP 这 3 种图像编码方法进行增强,构建 FrFT-GAF, FrFT-MTF, FrFT-RP 图像编码模块。运用图像编码模块将按心动周期分割后的一维心音信号分别转换为 FrFT-GAF, FrFT-MTF, FrFT-RP 二维编码特征图,然后将非对称卷积块(Asymmetric Convolutional Block, ACBlock)引入到 CNN 中,构建非对称卷积网络(Asymmetric Convolutional Network, ACNet)^[24],利用计算机视觉技术在分类任务中的优势,对心音的二维编码特征图进行分析处理,实现了对心音的有效分类。本文对上述各种图像编码模块的性能进行评估和比较,进一步表明所提方法的可行性和有效性,为心音信号分类提供了一种新的思路和方法,可用于 CHD 的机器辅助诊断中。

2 本文方法

2.1 整体框架

本文所提出的心音分类算法的流程如图 1 所示,其中图像编码模块可替换,包含以下 3 个模块选项:1) FrFT-GAF 模块;2) FrFT-MTF 模块;3) FrFT-RP 模块。具体的实现步骤为:1)使用双阈值自适应分割方法对心音信号按心动周期进行分割;2)对分割后的信号,应用 FrFT-GAF, FrFT-MTF, FrFT-RP 图像编码模块将经 FrFT 变换处理后的整个心动周期信号转换成二维编码特征图;3)使用 ACNet 模型进行训练及分类,并对各种图像编码模块的分类结果进行评估和比较。

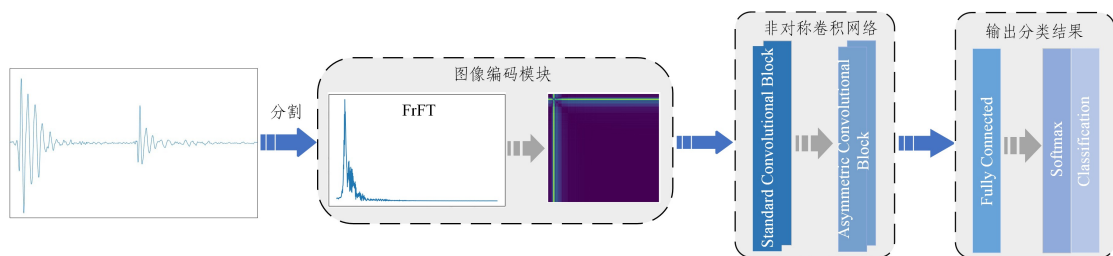


图 1 心音分类算法的框图

Fig. 1 Block diagram of heart sound classification algorithm

2.2 预处理

首先将心音样本下采样到 2000 Hz,然后归一化到 -1 和 1 之间,以确保心音样本幅度的一致性,加速收敛。接着使用双阈值自适应分割方法对心音样本按心动周期进行分割^[25],该方法根据 S1 和 S2 的能量相较于收缩期和舒张期的能量显著更高,通过归一化短时能量和频谱扩散来计算包络,从而确定 S1, S2 开始和结束的位置,实现自适应的心音信号分割。

2.3 特征提取

本小节介绍了特征提取部分所使用的图像编码模块的数学模型原理。

2.3.1 格拉姆角场(GAF)

GAF 是一种结合坐标变换和格拉姆矩阵(Gram Matrix)实现将时间序列变换为图像的编码方法,在保留信号完整信息的同时,还有效解决了时间序列依赖性的问题,非常适合心音信号等非平稳信号的处理。GAF 有两种编码方式,格拉姆角和场(Gramian Angular Summation Field, GASF)和格拉姆角差场(Gramian Angular Difference Field, GADF)。GASF 基于余弦函数,通过计算时间序列数据点之间角和的余弦值来生成矩阵,用于捕捉时间序列的累积特性;GADF 基于正弦函数,通过计算其角差的正弦值生成矩阵,强调数据点之间的变化趋势。具体实现步骤如下:

步骤 1 对信号进行缩放,将笛卡尔坐标系下给定的时间序列 $X = \{x_1, x_2, \dots, x_n\}$ 通过式(1)缩放至 $[-1, 1]$ 。

$$\tilde{x}_i^1 = \frac{(x_i - \min(X)) + (x_i - \min(X))}{\max(X) - \min(X)} \quad (1)$$

其中, x_i 表示第 i 个采样点的值; \tilde{x}_i^1 为第 i 个归一化后采样点的值; $\max(X)$ 和 $\min(X)$ 为时间序列 X 中所有采样点的最大、最小值。

步骤 2 转化为极坐标,通过式(2)将缩放后的数值 \tilde{x}_i^1 编码为角余弦 ϕ_i , 将时间节点 t_i 编码为半径 r_i , 使心音信号中的每一个值在极坐标中都有唯一的表示。

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i^1), & -1 \leq \tilde{x}_i^1 \leq 1 \\ r_i = \frac{t_i}{N}, & t_i \in N \end{cases} \quad (2)$$

其中, \tilde{x}_i^1 为归一化后的时间序列; t_i 为时间步; N 为极坐标下的正则化常数系数。

步骤 3 根据在编码过程中所使用的三角函数,得到 G_{GASF} 和 G_{GADF} , 其定义分别如式(3)和式(4)所示:

$$G_{\text{GASF}} = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} \quad (3)$$

$$G_{\text{GADF}} = \begin{bmatrix} \sin(\phi_1 - \phi_1) & \cdots & \sin(\phi_1 - \phi_n) \\ \sin(\phi_2 - \phi_1) & \cdots & \sin(\phi_2 - \phi_n) \\ \vdots & & \vdots \\ \sin(\phi_n - \phi_1) & \cdots & \sin(\phi_n - \phi_n) \end{bmatrix} \quad (4)$$

其中, $\cos(\phi_i + \phi_j)$ 为角度 ϕ_i 和角度 ϕ_j 之间的余弦和值; $\sin(\phi_i - \phi_j)$ 为角度 ϕ_i 和角度 ϕ_j 之间的正弦差值。

GAF 矩阵中主对角元素包含所有时间戳对应的值,非主对角元素(第 i 行、第 j 列, $i \neq j$)表示时间序列中第 i 个时间点和第 j 个时间点之间的数据相关性。在生成的图像中,颜色的深浅变化用于直观地表示两时刻之间相似度的高低,颜

色越深表示相似度越高,以此识别出心音信号中的相似模式和潜在变化趋势,从而更全面地对心音信号进行分析。

2.3.2 马尔可夫场(MTF)

MTF 是一种基于马尔可夫链特性的时间序列数据编码方法,将其一维序列数据视为一阶马尔可夫链,通过离散化和计算转移概率,将时间序列转化为图像。具体实现步骤如下:

步骤 1 离散化数据,给定时序信号 $X = \{x_1, x_2, \dots, x_n\}$, 根据不同时刻的信号幅值将 X 分为 Q 分位数区域,然后将时序信号的每个采样点 x_i 根据其幅值映射到相应的分位数区域 q_j ($j \in \{1, 2, \dots, Q\}$), 每一个采样点 x_i 对应唯一的 q_j 。

步骤 2 构建转移矩阵,沿时间轴方向,将离散化后的时序信号视为一阶马尔可夫链,对其计算各分位点之间的转移概率,由此构建 $Q \times Q$ 的马尔可夫转移矩阵 W , 其定义如式(5)所示:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1Q} \\ w_{21} & w_{22} & \cdots & w_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ w_{Q1} & w_{Q2} & \cdots & w_{QQ} \end{bmatrix} \quad (5)$$

其中, w_{ij} 为分位数区域 q_i 中数据点转移到分位数区域 q_j 的概率。

步骤 3 构建 MTF 矩阵,为了弥补 W 忽略了时间序列 X 的分布与时间步长 t 之间的时间依赖关系,MTF 将状态转移的概率按时间顺序排列来拓展转移矩阵 W , 生成马尔可夫矩阵 M , 其定义如式(6)所示:

$$M = \begin{bmatrix} w_{ij|x_1 \in q_i, x_1 \in q_j} & w_{ij|x_1 \in q_i, x_2 \in q_j} & \cdots & w_{ij|x_1 \in q_i, x_n \in q_j} \\ w_{ij|x_2 \in q_i, x_1 \in q_j} & w_{ij|x_2 \in q_i, x_2 \in q_j} & \cdots & w_{ij|x_2 \in q_i, x_n \in q_j} \\ \vdots & \vdots & \ddots & \vdots \\ w_{ij|x_n \in q_i, x_1 \in q_j} & w_{ij|x_n \in q_i, x_2 \in q_j} & \cdots & w_{ij|x_n \in q_i, x_n \in q_j} \end{bmatrix} \quad (6)$$

其中, w_{ij} 为分位数区域 q_i 中数据点转移到分位数区域 q_j 的概率。

MTF 图像由不同颜色的矩阵格点组成,每个格点的颜色表示信号在不同频率状态之间的转移概率,颜色越深表示转移概率越高,颜色越浅表示转移概率越低。通过可视化的方法直观展示出心音信号的频率状态转移情况,捕捉了心音信号的动态特征,能有效地区分出正常信号和 CHD 信号。

2.3.3 递归图(RP)

RP 主要基于时间序列数据的自相似性原理,通过比较时间序列中不同时间点上的状态相似性,将时间序列在状态空间中的循环或重复模式编码成二维图像,以直观地分析时间序列数据的非平稳性、混沌性以及潜在的周期性特征。具体实现步骤如下:

步骤 1 相空间重构,给定一维时序数据 $X = \{x_1, x_2, \dots, x_n\}$, 通过计算嵌入维度 m 和延迟时间 τ , 将其转化为相空间向量,如式(7)所示:

$$X_i = (x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}) \quad (7)$$

步骤 2 计算重构后两向量之间的距离,对于每一对向量 X_i 和 X_j , 通过式(8)计算它们之间的欧氏距离。

$$d_{ij} = \|X_i - X_j\| \quad (8)$$

步骤 3 确定阈值,选择一个阈值 ϵ , 用来判断状态是否相似。如果两个状态之间的距离小于或等于 ϵ , 则认为是相似的。

步骤4 生成递归图,将所选阈值 ϵ 与计算得到的 d_{ij} 作差,然后将结果输入到 Heaviside 函数中求得递归矩阵上每个点的值,定义如式(9)所示:

$$R_{ij} = \Theta(\epsilon - \| \vec{x}_i - \vec{x}_j \|) \quad (9)$$

其中, Θ 是 Heaviside 函数,定义如式(10)所示:

$$\Theta(t) = \begin{cases} 0, & t < 1 \\ 1, & \text{其他} \end{cases} \quad (10)$$

递归矩阵 \mathbf{R} 的每个元素 R_{ij} 表示时间点 i 和的特征是否相似,当 $d_{ij} \leq \epsilon$ 时, $R_{ij} = 1$,在递归图中对应的坐标位置上绘制一个点,表明此时的两个序列特征高度相似;否则 $R_{ij} = 0$,在递归图中不绘制点,表明两个序列特征明显不同。这些点在二维平面上组成图像,用于揭示一维序列中的结构和内在模式。

2.3.4 图像编码模块

FrFT 是对傅里叶变换的扩展形式,如果将傅里叶变换视为信号从时间轴逆时针旋转 $\pi/2$ 至频率轴上,则 FrFT 可将信号旋转任意角度到 u 轴上^[26],如图 2 所示,其定义如式(11)所示:

$$f_a(u) = F^a(u) = \int_{-\infty}^{+\infty} K_a(t,u) f(t) dt \quad (11)$$

其中, F^a 是 FrFT 算子; $K_a(t,u)$ 为 FrFT 的核函数,其定义如式(12)所示:

$$K_a(t,u) = \begin{cases} \sqrt{|1-j \cot(a)|} \exp\{j\pi[(t^2+u^2) \cot(a) - 2ut \csc(a)]\}, & a \neq n\pi \\ \delta(t-u), & a = 2n\pi \\ \delta(t+u), & a = (2n+1)\pi \end{cases} \quad (12)$$

其中, $\delta(u)$ 为冲激函数; n 为任意整数; a 表示时间轴以 O 为原点逆时针旋转的角度, $a = 0$ 时, FrFT 表现为时域; $a = \pi/2$ 时, FrFT 即为传统的傅里叶变换,信号从时域变换到频域。

FrFT 通过引入旋转角度的概念,将传统的傅里叶变换扩

展到了时域和频域之间的任意分数域中,其旋转的角度与 FrFT 的阶次相关,寻找最优阶次可以获取更多的特征。Fan 等^[27]对心音信号进行 FrFT 变换,根据 FrFT 的幅值谱呈现能量集中这一特性,得到最优阶次为 0.994,因此本文选取阶次 0.994。

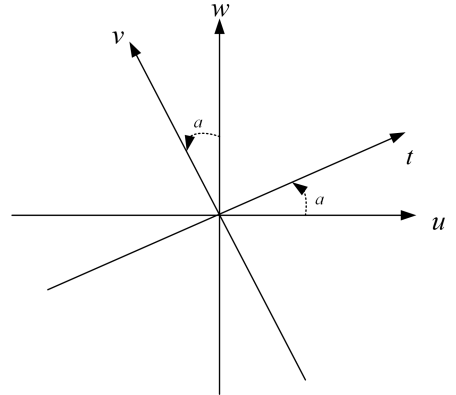


图 2 时频轴旋转 a 角度的分数阶 Fourier 域平面

Fig. 2 Fractional Fourier domain plane with time-frequency axis rotated by a

本文将 FrFT 应用于 GAF, MTF, RP 中,用于增强图像编码特征,构建了 FrFT-GAF, FrFT-MTF, FrFT-RP 图像编码模块,利用 FrFT 的时频混合特性增强原始图像编码生成的图像特征,相应地生成 FrFT-GAF, FrFT-MTF, FrFT-RP 二维编码特征图。由于 GAF 有两种编码方式,FrFT-GAF 进一步分为 FrFT-GASF 和 FrFT-GADF 特征图。图 3 即为将整个心动周期的心音信号经过图像编码模块处理后得到的编码特征图, n 点的一维心音信号转换为 $n \times n$ 点阵的二维编码特征图,本文使用双线性插值将得到的编码特征图的大小调整为 128×128 点阵。

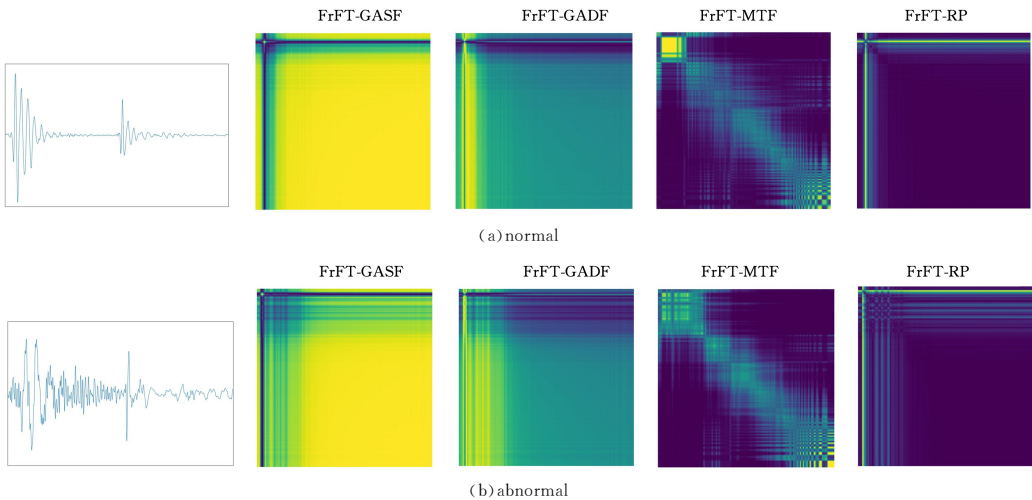


图 3 心音信号的时域图和图像编码特征图

Fig. 3 Time domain map and image coding feature map of heart sound signals

图 3 中,第 1 列为心音信号时域图,第 2 列为 FrFT-GASF 编码特征图,第 3 列为 FrFT-GADF 编码特征图,第 4 列为 FrFT-MTF 编码特征图,第 5 列为 FrFT-RP 编码特征图,所有编码特征图均为 128×128 点阵。

2.4 分类模型

传统的 CNN 在参数量、计算效率和特征提取能力方面

存在一定的局限性,尤其是使用的二维卷积核在处理特定方向上的特征时,无法充分考虑到其特征的差异性。为了提高 CNN 网络的性能,Ding 等^[24]在 2019 年提出了一种使用一维非对称卷积块替代标准的二维卷积核的非对称卷积网络 AC-Net,其 ACBlock 将 $d \times d$ 的卷积核替换为 $d \times d, d \times 1$ 和 $1 \times d$ 的 3 个卷积核,并将其输出进行算术求和,结构如图 4 所

示。通过结合对称和非对称卷积核,ACNet 增强了图像在水平和垂直方向上的特征,尤其适用于本文所使用的图像编码模块特征,提高了特征表达的丰富性,且 ACBlock 可以简单、便捷地替换成熟网络体系中的二维卷积核,而不需要额外的引入超参数。现已在情绪识别、生物医学图像处理、图像处理等多个领域取得显著成效^[28-30]。

本文使用 ACNet 对心音信号进行分类,网络结构如图 5 所示。该模型包含两个普通卷积层和两个非对称卷积层,普通卷积层的卷积核设置为 3×3 ,非对称卷积层包含 $3 \times 3, 3 \times 1$ 和 1×3 这 3 个卷积核,并选用线性整流函数(Rectified Linear unit,ReLU)作为激活函数;之后使用池化核为 2×2 的最大池化层进行特征降维;经过卷积、池化处理后将数据进行展

平(Flatten),接着接入 2 个全连接层(Dense),并使用 Dropout 和 L2 权重正则化来防止模型过拟合;最后使用 softmax 作为最后一层的激活函数,获得分类结果的概率分布。

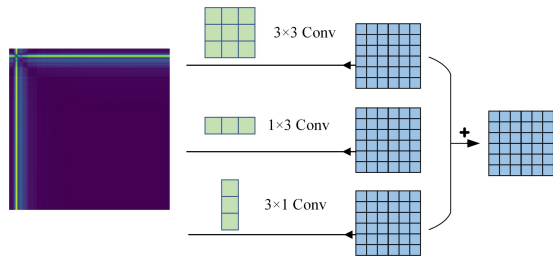


图 4 非对称卷积模块示意图

Fig. 4 Schematic diagram of ACBlock

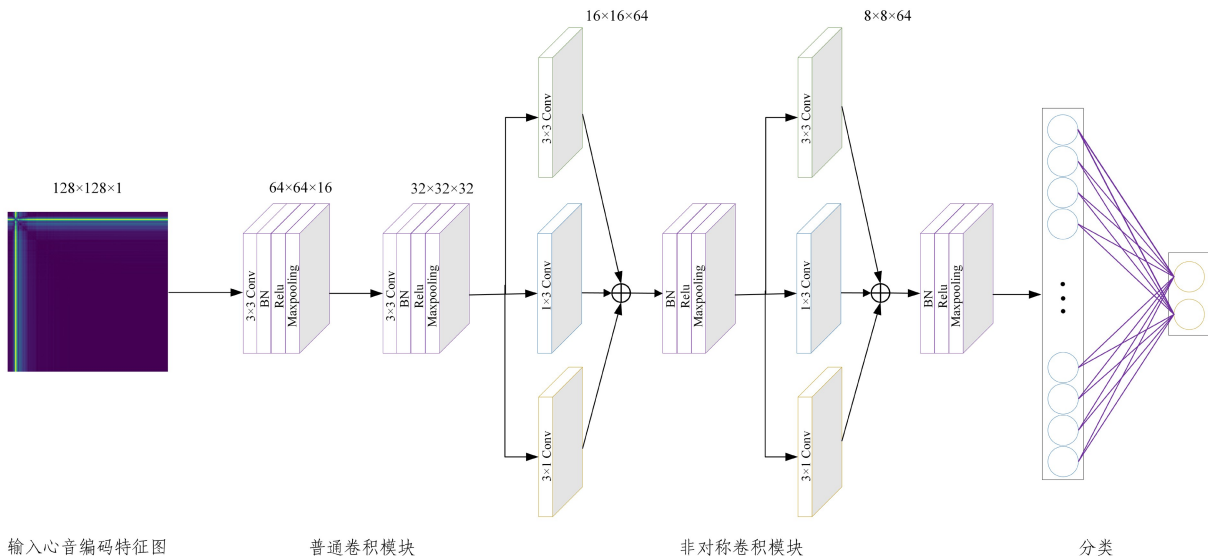


图 5 ACNet 模型的结构图(水平放置)

Fig. 5 Structure diagram of ACNet model(horizontally placed)

3 实验结果与分析

3.1 数据来源

本文使用两个数据集。数据集 1 为云南大学与云南省阜外心血管病医院合作构建,数据源于从临床和先心病筛查现场采集所得,患者、志愿者年龄段为 6 个月至 18 岁,所有数据均带标签。心音样本的采集和使用得到了云南大学和云南省阜外心血管病医院医学伦理委员会的批准。所采集的样本包括心脏区域的 5 个临床听诊点位,即二尖瓣区(M)、肺动脉瓣区(P)、主动脉瓣区(A)、主动脉瓣第二听诊区(E)和三尖瓣区(T),其采样频率为 5000 Hz,时长 20 s。本实验共使用 1600 例心音样本,其中正常和异常(CHD)样本比例为 1:1,异常样本为几类常见的 CHD,包括房间隔缺损、室间隔缺损、动脉导管未闭,实验将数据集分为训练集、验证集和测试集,比例为 8:1:1。

数据集 2 为 PhysioNet/CinC 2016 心音数据库,共包含 3240 例心音样本,采样频率为 2000 Hz,时长为 5~120 s 不等,其中正常样本 2575 例,异常样本 665 例,实验按照 8:1:1 的比例划分为训练集、验证集和测试集。

3.2 实验环境

本实验所用硬件配置为:中央处理器(Intel Xeon W-2102 @ 2.90 GHz),独立显卡(NVIDIA GeForce RTX 2080)。所

使用的深度学习框架为 TensorFlow 2.4,编程语言为 Python 3.7,编程软件为 Pycharm 2022,所有实验均在同一台设备进行。

3.3 评价指标

为评估算法的性能,本文采用准确率(Accuracy, Acc)、灵敏度(Sensitivity, S_e)、特异度(Specificity, S_p)和 F1 分数进行评估,其中 F1 分数是精确率(Precision, P)和召回率(Recall, R)的调和平均数,其计算式如式(13)~式(18)所示:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (13)$$

$$S_e = \frac{TP}{TP + FN} \times 100 \quad (14)$$

$$S_p = \frac{TN}{TN + FP} \times 100\% \quad (15)$$

$$R = S_e \quad (16)$$

$$P = \frac{TP}{TP + FP} \times 100\% \quad (17)$$

$$F_1 = \frac{2PR}{P+R} \times 100\% \quad (18)$$

其中, TP 代表真阳性,指正确识别异常样本的数目; TN 代表真阴性,指正确识别正常样本的数目; FP 代表假阳性,指正正常样本分类错误的数目; FN 代表假阴性,指异常样本分类错误的数目。

3.4 实验结果与分析

为验证本文所提出的心音分类算法的准确性,使用 FrFT-GAF, FrFT-MTF, FrFT-RP 图像编码模块对一维心音信号进行编码,然后分别输入到 ACNet 分类模型中进行训练,并在测试集上对上述图像编码模块的性能进行评估和比较。

首先,探究 FrFT-GAF, FrFT-MTF, FrFT-RP 图像编码模块与 GAF, MTF, RP 原始图像编码在数据集 1 中使用同一分类模型中的分类结果,结果如表 1 所列。由表 1 可知,原始图像编码方法中使用 RP, MTF, GAF 效果依次递减,整体的分类准确率不高。使用 FrFT 增强图像编码后,心音分类的准确率明显提高,其中 FrFT-RP 模块的分类效果最好,在数据集 1 上的准确率可达 0.981, F1 为 0.989; FrFT-MTF, FrFT-GAF 模块的效果依次,在 FrFT-GAF 两种编码方式中 FrFT-GASF 模块的效果优于 FrFT-GADF 模块。

表 2 不同分类算法的对比

Table 2 Comparison of different classification methods

算法	特征	分类方法	评价指标			
			Accuracy	Sensitivity	Specificity	F1
Riccio 等 ^[21]	PIFS	CNN	0.850	—	—	0.860
Ranipa 等 ^[23]	MFCs+ Mel + CST	MACNN	0.955	0.973	0.884	0.972
Wang 等 ^[31]	second-order spectral	PCTMF-Net	0.930	—	—	0.927
Arnab 等 ^[32]	Spectrogram	YAMNet	0.922	—	—	—
Zhang 等 ^[33]	原始信号	MDFNet	0.944	0.898	0.957	0.870
本文算法	FrFT-GASF	ACNet	0.967	0.987	0.951	0.963
	FrFT-GADF	ACNet	0.963	0.991	0.940	0.959
	FrFT-MTF	ACNet	0.972	0.971	0.972	0.968
	FrFT-RP	ACNet	0.977	0.983	0.972	0.974

从表 1 可以看出,采用原始的图像编码方法处理心音信号时,整体分类准确率较低。分析其可能的原因在于, GAF 在将信号映射到极坐标并计算所有成对向量的格拉姆矩阵时,由于其转换过程中的复杂性而丢失了部分关键信息,而对于 GASF 性能优于 GADF 可能是因为心音信号具有大量的重复性成分(即 S1 和 S2),信号中不同点之间的值相同, GADF 通过计算角度差的余弦值,使其相关性计算为 0,从而忽略了一些正相关的信息,增加了模型学习的难度。MTF 主要基于状态间的转移概率矩阵,虽然能揭示信号的一些动态特性,但无法充分捕捉到心音信号的特征,导致图像特征信息不足。RP 通过识别心音序列在相空间中不同时刻的状态相似性来生成图像,虽然能够揭示信号的一些周期性或重复性模式,但单纯依赖相空间信息可能无法充分反映心音信号的复杂特征。通过 FrFT-GAF, FrFT-MTF, FrFT-RP 图像编码模块后,不仅保留了原始图像编码方法的优势,还结合了时域和频域信息,使得生成的编码特征图可以更全面地反映心音信号的本质特性,从而显著提升了整体分类性能。

从表 2 可以看出,传统的特征提取方法对心音样本自身特征的表达存在一定局限性,无法捕捉到心音的复杂、非线性特征。本文所提出的通过增强图像编码的方法分类性能更优,更好地揭示了心音复杂的动态特性和时间依赖性,为提升心音分类任务的准确性和可靠性开辟了新的途径。同时,数据集 1 的数据来自临床和先心病筛查现场,涵盖 5 个临床听诊点位,且在较为安静的环境下采集,心音样本质量更高,使其分类准确率高于数据集 2。

表 1 数据集 1 的分类结果

Table 1 Classification results of dataset 1

分类方法	评价指标			
	Accuracy	Sensitivity	Specificity	F1
GASF+ACNet	0.920	0.847	0.994	0.918
GADF+ACNet	0.907	0.821	0.994	0.899
MTF+ACNet	0.931	0.871	0.991	0.927
RP+ACNet	0.943	0.946	0.941	0.936
FrFT-GASF+ACNet	0.969	0.941	0.998	0.968
FrFT-GADF+ACNet	0.966	0.965	0.968	0.967
FrFT-MTF+ACNet	0.973	0.957	0.989	0.973
FrFT-RP+ACNet	0.981	0.979	0.998	0.989

其次,为验证本文算法的有效性 with 可靠性,使用公开数据集 2 与近两年的心音信号特征提取方法进行对比,其结果如表 2 所列。由表 2 可知,本文所提出的心音分类算法的准确率和 F1 值最高,整体分类效果优于其他特征,其中 FrFT-RP 模块准确率可达 0.977, F1 为 0.974。

此外,为了探究 ACBlock 在网络模型中的影响,以 FrFT-RP 模块为例进行消融实验,分别使用传统的 CNN 和带有 ACBlock 的 ACNet 进行实验,其结果如表 3 所列。由表 3 可知,使用 ACNet 进行分类的效果更好,其在数据集 1 和数据集 2 上的准确率比传统的 CNN 分别提升了 3.1%, 2.2%。ACNet 中的 3×1 和 1×3 卷积核,使网络增强了水平和垂直方向上的特征信息,提高了模型的特征提取能力,且计算复杂度低,使用灵活,在心音信号分类任务中展现出了其性能优势。

表 3 消融实验

Table 3 Ablation experiment

数据集	是否使用 ACBlock	评价指标			
		Accuracy	Sensitivity	Specificity	F1
1	否	0.950	0.905	0.996	0.948
	是	0.981	0.977	0.987	0.982
2	否	0.955	0.965	0.947	0.949
	是	0.977	0.983	0.972	0.974

结束语 本文提出了一种通过增强图像编码和非对称卷积网络的心音分类算法。首先将心音信号按心动周期进行分割,然后引入 FrFT 对 GAF, MTF, RP 原始图像编码方法进行增强,构建了 FrFT-GAF, FrFT-MTF, FrFT-RP 图像编码模块,进而将分割后的一维心音信号编码为二维编码特征图,并使用 ACNet 模型对心音进行分类研究。此外,还对增强后的各种图像编码模块的性能进行了评估和比较。实验结果显示,在心音二分类任务中, FrFT-RP 模块的分类效果最好, FrFT-MTF, FrFT-GAF 模块的效果依次,使用 FrFT 增强图

像编码特征后较以往方法有明显提升,为心音信号分类提供了新的思路和方法,有望应用于先心病机器辅助诊断。本文算法已成功实现正常和异常心音的二分类,下一步的研究重点是对常见先心病的多分类进行分析研究。

参 考 文 献

- [1] MA L Y, WANG Z W, FAN J, et al. Highlights of the China Cardiovascular Health and Disease Report 2022 [J]. Chinese Family Medicine, 2023, 26(32): 3975-3994.
- [2] SUN S P, WANG H B, JIANG Z W, et al. Segmentation-based heart sound feature extraction combined with classifier models for a VSD diagnosis system[J]. Expert Systems with Applications, 2014, 41(4): 1769-1780.
- [3] CHOWDHURY M, POUDEL K, HU Y. Detecting abnormal PCG signals and extracting cardiac information employing deep learning and the shannon energy envelope[C]//2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE, 2020: 1-4.
- [4] KAMSON A P, SHARMA L N, DANDAPAT S. Enhancement of the heart sound envelope using the logistic function amplitude moderation method[J]. Computer Methods and Programs in Biomedicine, 2020, 187: 105239.
- [5] VARGHEES V N, RAMACHANDRAN K I. Effective Heart Sound Segmentation and Murmur Classification Using Empirical Wavelet Transform and Instantaneous Phase for Electronic Stethoscope[J]. IEEE Sensors Journal, 2017, 17(12): 3861-3872.
- [6] ZHANG W J, HAN J Q, DENG S W. Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation[J]. Biomedical Signal Processing and Control, 2019, 53: 101560.
- [7] IBRAHIM N, JAMAL N, SHA'ABANI A H, et al. A Comparative Study of Heart Sound Signal Classification Based on Temporal, Spectral and Geometric Features[C]//2020 IEEE-EMB-SCongress on Biomedical Engineering and Sciences(IECBES). 2021: 24-29.
- [8] ESLAMIZADEH G, BARATI R. Heart murmur detection based on wavelet transformation and a synergy between artificial neural network and modified neighbor annealing methods[J]. Artificial Intelligence in Medicine, 2017, 78: 23-40.
- [9] BHATIKAR S R, DEGROFF C, MAHAJAN R L. A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics[J]. Artificial Intelligence in Medicine, 2005, 33(3): 251-260.
- [10] ZENG W, LIN Z X, YUAN C Z, et al. Detection of heart valve disorders from PCG signals using TQWT, FA-MVEMD, Shannon energy envelope and deterministic learning[J]. Artificial Intelligence Review, 2021(7): 1-38.
- [11] CHEN J X, GUO Z H, XU X, et al. A Robust Deep Learning Framework Based on Spectrograms for Heart Sound Classification[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2024, 21(4): 936-947.
- [12] JAMIL S, ROY A M. An efficient and robust Phonocardiography(PCG)-based Valvular Heart Diseases(VHD) detection framework using Vision Transformer(ViT)[J]. Computers in Biology and Medicine, 2023, 158: 106734.
- [13] 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.
- [14] KUI H R, PAN J H, ZONG R, et al. Heart sound classification based on log Mel-frequency spectral coefficients features and convolutional neural networks[J]. Biomedical Signal Processing and Control, 2021, 69: 102893.
- [15] OBAIDAT M S. Phonocardiogram signal analysis: Techniques and performance comparison[J]. Journal of Medical Engineering & Technology, Taylor & Francis, 1993, 17(6): 221-227.
- [16] LUBIS C, GONDAWIJAYA F. Heart Sound Diagnose System with BFCC, MFCC, and Backpropagation Neural Network[J]. IOP Conference Series: Materials Science and Engineering, 2019, 508(1): 012119.
- [17] WANG Y L, SUN J, YANG H B, et al. Classification Model of Heart Sounds in Pulmonary Hypertension Based on Time-Frequency Fusion Features[J]. Computer Science, 2024, 51(S1): 387-393.
- [18] ECKMANN J P, KAMPHORST S O, RUELE D. Recurrence Plots of Dynamical Systems[J]. Europhys Lett, 1987, 4(9): 973-977.
- [19] WANG Z G, OATES T. Imaging Time-Series to Improve Classification and Imputation [J]. arXiv. org e-Print archive, 2015, arXiv:1506.00327.
- [20] ZHOU G, CHIEN C, CHEN J, et al. Identifying pediatric heart murmurs and distinguishing innocent from pathologic using deep learning [J]. Artificial Intelligence in Medicine, 2024, 153: 102867.
- [21] RICCIO D, BRANCATI N, SANNINO G, et al. CNN-based classification of phonocardiograms using fractal techniques[J]. Biomedical Signal Processing and Control, 2023, 86: 105186.
- [22] NGUYEN M T, LIN W W, HUANG J H. Heart Sound Classification Using Deep Learning Techniques Based on Log-mel Spectrogram [J]. Circuits, Systems, and Signal Processing, 2023, 42(1): 344-360.
- [23] RANIPA K, ZHU W P, SWAMY M N S. A novel feature-level fusion scheme with multimodal attention CNN for heart sound classification[J]. Computer Methods and Programs in Biomedicine, 2024, 248: 108122.
- [24] DING X H, GUO Y C, DING G G, et al. ACNet: Strengthening the Kernel Skeletons for Powerful CNN via Asymmetric Convolution Blocks[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 1911-1920.
- [25] GE B B, YANG H B, MAP Y, et al. Detection of pulmonary arterial hypertension associated with congenital heart disease based on time-frequency domain and deep learning features[J]. Biomedical Signal Processing and Control, 2023, 81: 104451.
- [26] 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]. Elsevier, 2020, 57: 101788.
- [27] FAN Q L, YANG H B, GUO T, et al. FrFT-Bark domain fea-

ture extraction and CNN residual shrinkage network heart sound classification algorithm[J]. Journal of Yunnan University: Natural Science Edition, 2023, 45(3): 564-574.

- [28] WANG P F, GONG X G, GUO Q, et al. Children's Expression Recognition Based on Multi-Scale Asymmetric Convolutional Neural Network[J]. International Journal of Advanced Computer Science and Applications(IJACSA), 2024, 15(7): 437.
- [29] HOU C Q, LI J S, WANG W, et al. Second-order asymmetric convolution network for breast cancer histopathology image classification [J]. Journal of Biophotonics, 2022, 15(5): e202100370.
- [30] WU J, WANG Y X, ZHANG X G. Lightweight Asymmetric Convolutional Distillation Network for Single Image Super-Resolution[J]. IEEE Signal Processing Letters, 2023, 30: 733-737.
- [31] WANG R S, DUAN Y F, LI Y K, et al. PCTMF-Net: heart sound classification with parallel CNNs-transformer and second-order spectral analysis[J]. The Visual Computer, 2023, 39(8): 3811-3822.
- [32] MAITY A, PATHAK A, SAHA G. Transfer learning based

heart valve disease classification from Phonocardiogram signal [J]. Biomedical signal Processing and Control, 2023, 85: 104805.

- [33] ZHANG H B, ZHANG P, WANG Z W, et al. Multi-Feature Decision Fusion Network for Heart Sound Abnormality Detection and Classification[J]. IEEE Journals & Magazine, 2023: 1386-1397.



WANG Shengyi, born in 1999, postgraduate. Her main research interests include signal processing and machine learning.



WANG Weilian, born in 1947, bachelor, professor. His main research interests include signal processing and pattern recognition, biological signal processing, digital-analog hybrid IC and ASIC design.