

本文引用格式: 甘兆明,林家荣,杨其宇.基于知识蒸馏的房颤信号提取方法[J].自动化与信息工程,2023,44(3):29-33;51.

GAN Zhaoming, LIN Jiarong. Method for extracting atrial fibrillation signals based on knowledge distillation[J]. Automation & Information Engineering, 2023,44(3):29-33;51.

基于知识蒸馏的房颤信号提取方法

甘兆明 林家荣 杨其宇

(广东工业大学, 广东 广州 510006)

摘要: 针对基于双时域卷积网络的房颤信号提取网络模型存在的参数量大、运算资源要求高和实时性差等问题, 提出基于知识蒸馏的房颤信号提取方法。该方法的教师网络和学生网络分别采用 3 层、1 层的基于时域卷积网络 (TCN), 维度分别为 256 和 32。实验结果表明, 采用知识蒸馏的方法可以提高学生网络的性能, 且用蒸馏后的房颤信号提取的学生网络相比于教师网络, 其网络模型更小、运算资源要求更低、实时性更高, 为部署到资源有限的嵌入式设备提供了理论依据。

关键词: 知识蒸馏; 房颤信号提取; 时域卷积网络; 教师网络; 学生网络

中图分类号: R318;TN911.7

文献标志码: A

文章编号: 1674-2605(2023)03-0005-06

DOI: 10.3969/j.issn.1674-2605.2023.03.005

Method for Extracting Atrial Fibrillation Signals Based on Knowledge Distillation

GAN Zhaoming LIN Jiarong YANG Qiyu

(Guangdong University of Technology, Guangzhou 510006, China)

Abstract: A knowledge distillation based method for extracting atrial fibrillation signals is proposed to address the problems of large parameter quantities, high computational resource requirements, and poor real-time performance in the dual time domain convolutional network based atrial fibrillation signal extraction network model. The teacher network and student network of this method adopt a 3-layer and 1-layer time-domain convolutional network (TCN) with dimensions of 256 and 32, respectively. The experimental results show that using knowledge distillation can improve the performance of student networks, and compared to teacher networks, student networks extracted from distilled atrial fibrillation signals have smaller network models, lower computational resource requirements, and higher real-time performance, providing a theoretical basis for deploying to embedded devices with limited resources.

Keywords: knowledge distillation; extraction of atrial fibrillation signals; time domain convolutional network; teacher network; student network

0 引言

房颤是常见的心血管疾病^[1], 它会引发患者的并发症, 如脑卒中、心衰、心肌梗死或老年痴呆等。然而, 目前从患者体表采集的心电图 (electrocardiogram, ECG) 信号为心房和心室的混叠信号, 即房颤信号与心室信号在时域和频域上都存在混叠^[2]。因此, 对房颤信号的实时提取具有重要的现实意义。

房颤信号的提取方法分为多导联方法和单导联方法。其中, 多导联方法包括盲源分离算法^[3]、时空消除算法^[4]和神经网络算法^[5]等, 该方法需要连接多个电极, 存在较大的局限性; 单导联方法包括加权平均模板对消法^[6]、扩展卡尔曼滤波法^[7]等, 该方法存在对心室形态变异性敏感、依赖高精度 R 峰检测算法、需要人工调参等缺点。LU 等^[8]利用全卷积时域音频分离网络 (fully convolutional time domain audio separa-

tion network, Conv-Tasnet)，提出了双路时间卷积房颤信号提取网络（dual temporal convolution f-wave extraction network, DT-FENet），将 Conv-Tasnet 的单路编解码结构扩展为双路。DT-FENet 具有较高的房颤信号提取精度，但网络模型参数量大^[8]、运算资源要求高、实时性差，难以在资源受限的嵌入式设备中部署。

知识蒸馏作为一种深度神经网络的压缩方法，在行人检测、语义分割、图像超采样等应用场景都取得了良好的效果。为解决 DT-FENet 对运算资源要求高的问题，本文提出一种基于知识蒸馏的房颤信号提取方法，利用知识蒸馏对 DT-FENet 模型进行压缩。

1 相关理论基础

1.1 房颤心电信号模型

房颤患者的 ECG 信号 $\mathbf{x} = [x(1), x(2), \dots, x(T)] \in \mathbb{R}^{1 \times T}$ 由心室信号和房颤信号组成：

$$\mathbf{x} = \mathbf{x}_{VA} + \mathbf{x}_{AA} \quad (1)$$

式中： $\mathbf{x}_{VA} \in \mathbb{R}^{1 \times T}$ 为心室信号， $\mathbf{x}_{AA} \in \mathbb{R}^{1 \times T}$ 为房颤信号， T 为采样点数。

1.2 DT-FENet 方法

DT-FENet 方法是通过两组编解码器分别进行心室信号和房颤信号的编解码，以此进行心室信号和房颤信号的特征映射。DT-FENet 的网络结构如图 1 所示。

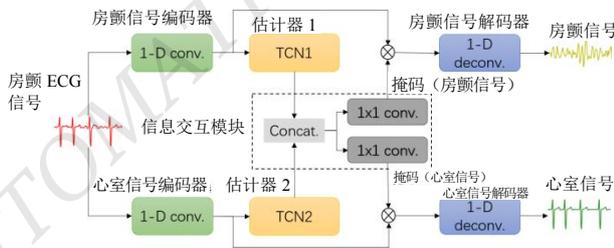


图 1 DT-FENet 网络结构图

将房颤患者的 ECG 信号 \mathbf{x} 输入到 DT-FENet 后，由房颤信号编码器和心室信号编码器分别输出房颤成分编码 F_{AA} 和心室成分编码 F_{VA} ； F_{AA} 和 F_{VA} 分

别送到估计器和解码器；信息交互模块利用房颤信号和心室信号的相关性，优化估计器输出的自注意力编码 M_{AA} 和 M_{VA} ；在 F_{AA} 上应用 M_{AA} 、在 F_{VA} 上应用 M_{VA} ，可分别得到带有自注意力特征的 \tilde{Z}_{AA} 和 \tilde{Z}_{VA} ；房颤信号解码器和心室信号解码器分别将高维的 \tilde{Z}_{AA} 和 \tilde{Z}_{VA} 解码，得到一维的房颤信号 \hat{Z}_{AA} 和心室信号 \hat{Z}_{VA} 。

1.3 知识蒸馏介绍

HINTON 等^[9]提出的知识蒸馏方法采用了“教师”和“学生”的概念，又被称为“教师-学生结构”。知识蒸馏方法中重量级的模型是结构复杂、规模大、拟合能力强的教师模型，且通常已被预先训练好；轻量化的学生模型模仿教师模型，吸收从教师模型中提炼出来的知识，具备更好的性能。ROMERO 等^[10]在 Hinton 等提出的知识蒸馏方法的基础上，提出了基于中间层的知识蒸馏。在中间层的知识蒸馏中，学生模型的学习目标是使自身的特征空间尽量靠近教师模型的特征空间，其损失函数为

$$\mathcal{L}(x, W) = H(y, \tilde{y}) + \mathcal{L}_{HT}(F_T, F_S) \quad (2)$$

$$\mathcal{L}_{HT}(F_T, F_S) = D(TF_t(F_T), TF_s(F_S)) \quad (3)$$

式中： H 为交叉熵、均方误差等的评价指标， y 和 \tilde{y} 分别为真实标签和模型输出， F_T 和 F_S 分别为教师模型和学生模型的中间层特征， TF_t 和 TF_s 为对应中间层的特征映射函数， D 为 L1 损失、L2 损失等距离函数。

2 基于知识蒸馏的房颤信号提取方法

2.1 框架设计

基于知识蒸馏的房颤信号提取方法以教师-学生结构为基本框架，以 DT-FENet 为教师网络，以缩小尺寸的 DT-FENet 为学生网络，对网络的中间层进行知识蒸馏，其框架如图 2 所示。

教师网络输出的编码特征、心房掩码、心室掩码用于指导学生网络，学生网络输出的 f 波和 QRST 波与数据集中的真实标签进行比较。学生网络在教师网络标签和真实标签的监督下进行训练。

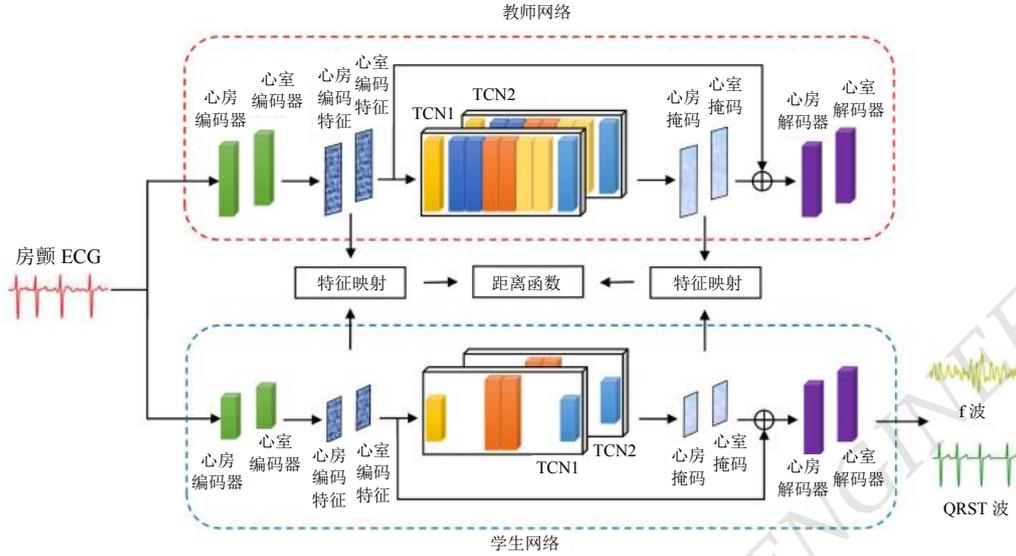


图2 DT-FENet 的知识蒸馏框架

2.2 知识蒸馏方法

本文采用的知识蒸馏方法包括 FitNet 方法和 AT 方法。FitNet 方法的教师模型和学生模型之间使用一维卷积进行特征映射，即将维数较低的学生模型中间层特征，通过单层的一维卷积网络映射到与教师模型中间层一致的维数，其损失函数为

$$\mathcal{L}_{\text{FitNet}}(F_T, F_S) = \left(\frac{\sum_{j=1}^n \sum_{i=1}^c \|F_T^i - r(F_S^i)\|_2^2}{C} \right) / n \quad (4)$$

式中：非线性函数 r 为一维卷积网络， C 为教师模型中间层的通道数， n 为样本长度。

AT 方法的教师模型和学生模型都加入了注意力机制，作为教师模型和学生模型的特征转换。距离函数选择均方误差 (mean square error, MSE)，其损失函数为

$$\mathcal{L}_{\text{AT}}(Q_T, Q_S) = \left(\sum_{j=1}^n \left\| \frac{Q_T^j}{\|Q_T^j\|_2} - \frac{Q_S^j}{\|Q_S^j\|_2} \right\|_2^2 \right) / n \quad (5)$$

$$Q_T(F_T) = \sum_{i=1}^{C_T} \|F_T^i\|^2 \quad (6)$$

$$Q_S(F_S) = \sum_{i=1}^{C_S} \|F_S^i\|^2 \quad (7)$$

式中： C_T 为教师模型的中间层通道数， C_S 为学生模型的中间层通道数， Q_T 和 Q_S 分别为教师模型和学生模型的注意力图， n 为样本长度。

AT 方法的注意力机制将中间层各通道的值求和并进行 L_2 范数归一化，使该层神经元整体的激活程度成为权重，令样本的不同区域对应不同程度的注意力，并将该注意力信息通过知识蒸馏传递给学生网络。

3 实验和结果分析

3.1 数据集

本实验的训练集和验证集均源自 Castilla-La Manch 数据库 (CLMDB)^[11]，测试集来源于 PhysioNet 的 MIT-BIH Atrial Fibrillation 数据库 (AFDB)^[12] 和 PhysioNet/Computing in Cardiology Challenge 2017 数据库 (Challenge 2017)^[13]。CLMDB 中的样本被随机划分为训练集 (70%，168 条) 和验证集 (30%，72 条)，并被切割成时长为 30 s 的片段。为了抑制测试集样本的基线漂移，采用截止频率为 0.01 Hz 的一阶巴特沃兹滤波器对 AFDB 和 Challenge 2017 中的样本进行高通

滤波。

3.2 损失函数

LU 等在提出 DT-FENet 方法时, 选择信噪比 (signal noise ratio, SNR) 为训练网络时的损失函数, 其定义为

$$\mathcal{L} = \sum_{i=1}^{n_{tr}} \text{SNR}_{AA}^i - \alpha \sum_{i=1}^{n_{tr}} \text{SNR}_{VA}^i \quad (8)$$

$$\text{SNR}_{AA}^i = 10 \log_{10} \frac{\|x_{AA}^i\|_2^2}{\|\tilde{x}_{AA}^i - x_{AA}^i\|_2^2} \quad (9)$$

$$\text{SNR}_{VA}^i = 10 \log_{10} \frac{\|x_{VA}^i\|_2^2}{\|\tilde{x}_{VA}^i - x_{VA}^i\|_2^2} \quad (10)$$

式中: n_{tr} 为训练集的批尺寸, SNR_{AA}^i 和 SNR_{VA}^i 分别为从每条训练样本提取的房颤信号和心室信号的信噪比, x_{AA}^i 和 \tilde{x}_{AA}^i 分别为训练样本的真实房颤信号和模型提取的房颤信号, x_{VA}^i 和 \tilde{x}_{VA}^i 分别为训练样本的真实心室信号和模型提取的心室信号, $\alpha = 0.05$ 是心房通道和心室通道之间的折衷系数。

值得注意的是, 房颤信号提取模型的训练目标是最大化信噪比值, 而神经网络的训练过程只能使目标值单调递减, 所以在信噪比前加负号, 将模型的训练方向设置为 SNR 值的最大化。这一设计是损失函数的值恒小于 0 且其绝对值逐渐变大。然而, 知识蒸馏方法的距离函数采用了 MSE, 在进行知识蒸馏时的优化目标是最小化 MSE, 即损失函数的值恒大于 0 且其绝对值逐渐变小。因此, 为了正常训练损失和知识蒸馏损失有一致的趋势, 有必要对正常训练损失进行调整。考虑到归一化均方误差 (normalized mean squared error, NMSE) 和 SNR 的计算公式在形式上有倒数的关系, 本文将正常训练损失由 SNR 改为 NMSE, 其定义为

$$\mathcal{L}_{train} = \sum_{i=1}^{n_{tr}} \text{NMSE}_{AA}^i + \alpha \sum_{i=1}^{n_{tr}} \text{NMSE}_{VA}^i \quad (11)$$

$$\text{NMSE}_{AA}^i = \frac{\|x_{AA}^i - \tilde{x}_{AA}^i\|_2^2}{\|x_{AA}^i\|_2^2} \quad (12)$$

$$\text{NMSE}_{VA}^i = \frac{\|x_{VA}^i - \tilde{x}_{VA}^i\|_2^2}{\|x_{VA}^i\|_2^2} \quad (13)$$

综上所述, 知识蒸馏的损失函数由正常训练损失 \mathcal{L}_{train} 和知识蒸馏损失 \mathcal{L}_{KD} 两部分组成。采用 FitNet 方法时, \mathcal{L}_{KD} 对应公式(15); 采用 AT 方法时, \mathcal{L}_{KD} 对应公式(16):

$$\mathcal{L}_{distill} = \mathcal{L}_{train} + \mathcal{L}_{KD} \quad (14)$$

$$\mathcal{L}_{KD} = \gamma \cdot (\mathcal{L}_{FitNet}^{enc-AA} + \alpha \mathcal{L}_{FitNet}^{enc-VA}) + \beta \cdot (\mathcal{L}_{FitNet}^{mask-AA} + \alpha \mathcal{L}_{FitNet}^{mask-VA}) \quad (15)$$

$$\mathcal{L}_{KD} = \gamma \cdot (\mathcal{L}_{AT}^{enc-AA} + \alpha \mathcal{L}_{AT}^{enc-VA}) + \beta \cdot (\mathcal{L}_{AT}^{mask-AA} + \alpha \mathcal{L}_{AT}^{mask-VA}) \quad (16)$$

式中: α 、 β 、 γ 均为折衷系数。为了减少超参数的数量, 2 个掩码通道的折衷系数与 α 相同, 即 $\alpha = \beta = \gamma = 0.05$ 。

对于 FitNet 方法, $\mathcal{L}_{FitNet}^{enc-AA}$ 和 $\mathcal{L}_{FitNet}^{enc-VA}$ 分别是编码器输出的心房编码蒸馏特征和心室编码蒸馏特征, $\mathcal{L}_{FitNet}^{mask-AA}$ 和 $\mathcal{L}_{FitNet}^{mask-VA}$ 分别是 TCN 网络输出的心房编码蒸馏特征和心室编码蒸馏特征。AT 方法类似, 这里不再重复叙述。

3.3 评价指标

因为测试集 AFDB 和 Challenge 2017 都来源于临床的 ECG 监测, 仅包含心房和心室的混合信号, 无法提供房颤信号作为真实值, 所以在测试过程中无法直接使用模型提取的波形与真实值的拟合误差来表示模型的提取精度。本文采用频谱集中度 (spectral concentration, SC) [14] 替代 SNR 和 NMSE 来作为评价指标, 房颤信号的频谱较窄, 绝大部分频谱能量集中

在 3~12 Hz 的区间内, 若提取的房颤信号频谱集中度较高, 说明房颤信号的失真小, 即算法的提取精度较高。SC 的计算公式为

$$SC = \frac{\sum_{f_i=3}^{12} P_{AA}(f_i)}{\sum_{f_i=0}^{F_s/2} P_{AA}(f_i)} \quad (17)$$

式中: F_s 为样本的采样率, P_{AA} 为房颤信号的功率谱。

3.4 实验设置

教师模型和学生模型的参数设置如表 1 所示。

表 1 教师模型和学生模型的参数设置

参数	教师模型	学生模型
编码器卷积核数量/个	256	32
估计器残差块堆叠层数/层	3	1
解码器卷积核数量/个	256	32

模型参数更新采用 Adam 优化器, 初始化学习率为 0.001, 训练的总数为 80 个 epoch, 批量大小为 2。引入梯度裁剪防止模型发生梯度爆炸, 模型梯度 L_2 范数的阈值为 5。FitNet 方法的 $\gamma = 0.01$, $\beta = 100$ 。AT 方法的 $\gamma = 1 \times 10^4$, $\beta = 1 \times 10^4$ 。

3.5 实验结果和分析

为验证知识蒸馏对房颤信号提取的有效性, 对比教师网络、FitNet 学生网络和 AT 学生网络的房颤信号提取效果, 其在验证集和测试集的提取精度分别如表 2、表 3 所示。

表 2 教师网络和学生网络在验证集上的提取精度

方法	SNR/dB	NMSE(10^2)	SC(10^2)
教师网络	7.84	24.01	82.09
FitNet	8.27	22.00	82.53
AT	8.00	23.36	82.29

由表 2 可以看出, SC 与 SNR、NMSE 具有较好的一致性, 从一定程度上证明了 SC 是一种有效的房颤信号提取精度评估指标, 适用于数据集未提供实值时的房颤信号提取精度的评估。

表 3 教师网络和学生网络在测试集上的提取精度

方法	SC(10^2)	
	AFDB	Challenge2017
教师网络	65.22	75.48
FitNet	69.61	74.33
AT	66.30	75.02

由表 3 可以看出, 使用 FitNet 方法知识蒸馏所得的学生网络, 在 AFDB 上的房颤信号提取精度优于 AT 方法网络和教师网络的提取精度。在 Challenge 2017 上, 使用 FitNet 方法和 AT 方法的知识蒸馏所得学生网络的 SC 与正常训练所得学生网络的 SC 相当。

教师网络、FitNet 学生网络和 AT 学生网络模型的资源占用情况如表 4 所示。

表 4 网络模型的资源占用情况

网络模型	总参数	FLOPs
教师网络	~2.82 MB	~92.74 MB
FitNet	~840 kB	~31.23 MB
AT	~838 kB	~31.12 MB

由表 4 可知, 学生模型的总参数量和总浮点运算数均是教师模型的三分之一。

结合表 3 可知, 学生网络用三分之一的资源开销在 Challenge 2017 测试集上取得了和教师网络相当的效果, 证明了基于知识蒸馏的房颤信号提取方法的有效性。

4 结论

针对现有的房颤信号提取网络 DT-FENet 存在规模大、消耗资源多的问题, 本文通过知识蒸馏的方法对 DT-FENet 网络模型进行压缩, 学生网络用三分之一的资源开销在 Challenge 2017 测试集上取得了和教师网络相当的效果, 解决了 DT-FENet 难以部署在嵌入式设备的问题。本文只利用知识蒸馏对 DT-FENet 进行压缩, 还可以利用剪枝和量化等方法进行下一步研究。

参考文献

- [1] 国家心血管病中心. 中国心血管健康与疾病报告 2020[J]. 心血管病杂志, 2021, 40(10): 5. (下转第 51 页)