



# (12) 发明专利申请

(10) 申请公布号 CN 116304861 A

(43) 申请公布日 2023. 06. 23

(21) 申请号 202310108116.6

G06F 18/25 (2023.01)

(22) 申请日 2023.02.09

G06N 3/0464 (2023.01)

(71) 申请人 河北工业大学

G06N 3/0442 (2023.01)

地址 300400 天津市北辰区双口镇西平道  
5340号

G06N 3/048 (2023.01)

G06N 3/084 (2023.01)

G01M 13/045 (2019.01)

(72) 发明人 刘晶 张明辉 季海鹏 赵佳  
夏淑媛

(74) 专利代理机构 天津麦芽知识产权代理有限  
公司 12269

专利代理师 祝国庆

(51) Int. Cl.

G06F 18/241 (2023.01)

G06F 18/2431 (2023.01)

G06F 18/214 (2023.01)

G06F 18/21 (2023.01)

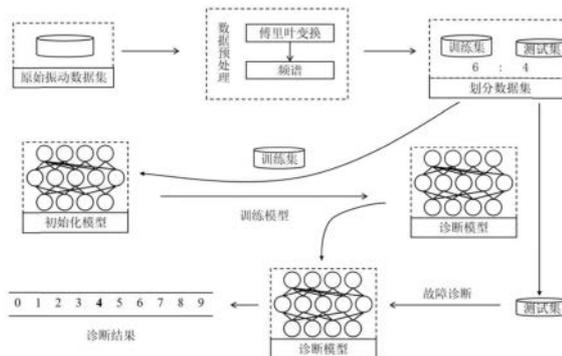
权利要求书3页 说明书10页 附图6页

## (54) 发明名称

一种基于自注意力的时频特征融合故障诊断方法

## (57) 摘要

本发明提供了一种基于自注意力的时频特征融合故障诊断方法,属于轴承设备故障诊断的技术领域,包括下述步骤:1.对各种状态下运行的轴承采集其振动信号;2.对轴承振动信号进行数据预处理,获得振动信号的频谱,并划分训练集和测试集;3.将原始振动信号和相应频谱同时输入到模型中进行训练;4.当有新的振动数据数据时,采用相同的方法进行预处理并输入到模型中,实现故障类型的分类诊断。本发明提出一种新的时频域特征融合方法,能够学习时、频域两方面的特征,相较于依赖单一时域或频域特征的故障诊断模型具有更高的诊断精度,具有较好的理论基础和应用价值。



1. 一种基于自注意力的时频特征融合故障诊断方法,其特征在于,包括下述步骤:

步骤1:采用电火花技术对待测轴承设置内圈、外圈、滚动体三种故障,每处故障直径又分为0.007、0.014、0.021inch三种,分别在4种不同负载下进行测试,由放置在驱动端和风扇端的加速度传感器分别以12kHz、48kHz的采样频率记录振动数据,最终采集到正常样本8个、外圈故障样本53个、内圈故障23个以及滚动体故障样本11个;

步骤2:采用48kHz采样频率下的3种不同负载的数据作为原始数据集,故障类别共10类;

步骤3:对原始振动信号进行数据预处理,采用傅里叶变换分析原始振动信号频率成分,获得频谱作为模型的频域输入;

步骤4:按6:4的比例划分训练集与测试集,每个工况下样本数量为4750个,即每类样本475个,单个样本长度800个数据点,对应频谱为400个数据点;

步骤5:将原始信号和相应频谱分别作为时域特征和频域特征输入到模型中,经多层基于自注意力的时频特征融合模块,进行全局特征提取的同时进行特征融合获得深度融合的时频融合特征;

步骤6:将时频融合特征输入到改进的多尺度残差网络进一步提取局部特征,并输出分类结果;

步骤7:使用Adam优化器对整个模型中的相关参数进行优化,得到使损失函数达到最小的最优参数;

步骤8:将未知故障数据以相同的数据预处理方法进行处理,并输入到模型中,得到最终分类结果。

2. 根据权利要求1所述的基于自注意力的时频特征融合故障诊断方法,其特征在于,步骤5中基于自注意力的时频特征融合模块:

基于自注意力的时频特征融合模块接受1维序列输入,轴承振动的信号的时、频域特征 $x, y$ 都属于一维信号,即 $x \in \mathbf{R}^{H_1 \times 1}, y \in \mathbf{R}^{H_2 \times 1}$ ,其中 $H_1$ 为时域信号采样点数, $H_2$ 为频域信号数据点数,原始振动信号和频谱分别作为时、频域特征输入,由于频谱的数据点数通常为原始振动信号的一半,分割为 $N$ 个Patch后数据点数量是不同的,为使时、频特征的数据维度一致,首先进行Patch Embedding,即经过一个全连接层将每个Patch统一映射到 $D$ 维度,然后使用标准的可学习1D位置嵌入,以保留每个Patch的位置信息,

$$z_0 = (z_0^x, z_0^y)$$

$$z_0^x = [x^1 \mathbf{E}; x^2 \mathbf{E}; \dots; x^N \mathbf{E}] + \mathbf{E}_{pos}, \quad \mathbf{E} \in \mathbf{R}^{(P_1 \square) \times D}$$

$$z_0^y = [y^1 \mathbf{E}; y^2 \mathbf{E}; \dots; y^N \mathbf{E}] + \mathbf{E}_{pos}, \quad \mathbf{E} \in \mathbf{R}^{(P_2 \square) \times D}$$

式中 $z_0$ 为模型输入,包括时域输入 $z_0^x$ 和频域输入 $z_0^y$ ,每个Patch乘以 $\mathbf{E}$ 代表Patch Embedding操作, $\mathbf{E}_{pos} \in \mathbf{R}^{N \times D}$ 为位置嵌入向量, $D$ 是TFF内部处理数据的维度,设置 $D=512$ ;

TFF由时频混合注意力MTFA和MLP模块组成,在每个模块之前应用层标准化,并且在每个模块之后使用残差连接,MLP包含两层线性层,采用GELU激活函数,相比于卷积网络中常用的RELU对小于0的值全部输出为零,GELU对于小于零的值会输出小的负值,有利于稳定网络梯度流,不容易造成梯度爆炸和梯度消失,TFF的层数 $L$ 可以由人为设定,层数越多意味着

模型中参数数量越多,训练就越困难,而且层数过多的模型对小数据集的效果反而更差, $L=3$ 时能够获得较好的诊断效果的同时模型的时间、空间性能都比较好;

$$z'_1 = \text{MTFA}(\text{LN}(z_{1-1})) + z_{1-1}$$

$$z_1 = \text{MLP}(\text{LN}(z_{1-1})) + z_{1-1}$$

式中 $z'_1$ 为第1层中间状态, $z_1$ 为第1层输出;

MTFA表示了单个Patch的处理流程,根据时域、频域输入分别生成 $(q_t^i, k_t^i, v_t^i)$ 和 $(q_f^i, k_f^i, v_f^i)$ ,其中 $q_t^i$ 为 $Q_t$ 中的一列,即 $Q_t = \mathbf{W}^q z^x = [q_t^1, q_t^2, \dots, q_t^N]$ ,将多个Patch分别做同样的操作得到 $(Q, K, V)$ ,然后做时频混合注意力,实现特征融合, $\mathbf{W}^q$ 为可学习的参数矩阵,随着模型训练不断优化

$$\text{MTFA}(z^x, z^y) = \left( \text{softmax}\left(\frac{Q_t K_f^T}{\sqrt{d_k}}\right) V_f, \text{softmax}\left(\frac{Q_f K_t^T}{\sqrt{d_k}}\right) V_t \right)$$

式中 $d_k$ 为 $Q, K$ 的维度, $z^x, z^y$ 分别为为时、频域输入,因为softmax的输入很大时梯度趋于零,为了使梯度不至于太小,所以除以 $\sqrt{d_k}$ ,使 $\frac{QK}{\sqrt{d_k}}$ 的分布尽可能接近均值为0,方差为1的正态分布;

MLP的结构包括两层线性层并采用GELU激活函数,中间隐层的维度为 $\text{mlp\_dim}=1024$ ,在时频混合注意力后加入这一结构的作用是为了进行特征提取,增强模型的表达能力,由于MTFA本质上是矩阵乘法运算,即都属于线性变换,线性变换的学习能力是远不如非线性变换的,所以通过加入激活函数为模型提高模型的学习能力。

3. 根据权利要求1所述的基于自注意力的时频特征融合故障诊断方法,其特征在于,步骤6中改进的多尺度残差网络:

为避免全注意力网络参数数量过大,造成模型不易训练且容易过拟合的现象,而卷积网络结构能够利用很少的参数获得较好诊断效果,所以选择基于卷积结构构建局部特征提取模块,能够提高模型泛化能力同时使整个模型参数数量不至于太大,传统的卷积神经网络依赖单一尺寸卷积核,由于卷积核只能关注感受野范围内的特征,其特征提取能力对于一些复杂任务往往效果不佳,尤其是轴承振动信号本身具有非线性、非平稳性,进一步造成隐性特征难提取,为了提取更为丰富的特征,提出改进的多尺度残差网络MSRN,由多个多尺度残差块MSRB组成,MSRB内部包含一个残差连接,1×3卷积、1×5卷积和1×7卷积,多个尺寸的卷积核能够关注不同尺度的局部特征,同时残差连接能够使输出中包含部分输入的信息,缓解因网络层数加深带来的梯度消失问题,加快模型收敛速度,MSRN堆叠3层多尺度残差块,然后连接两层全连接层进行降维,最后进行softmax处理得到概率分布,概率最大的类别即为模型预测的分类;

卷积层的作用是提取特征,具有参数数量少、计算速度快、拟合能力强的特点,许多领域的优秀模型都是基于卷积构建的,CNN通过卷积核在数据上的滑动计算获得抽象程度更高的特征,过程可表示为:

$$F = f(w * x + b)$$

式中 $F$ 表示卷积层输出, $f$ 表示卷积操作, $w$ 和 $b$ 分别表示卷积核权重和偏置向量;

激活层的作用是对输入特征做非线性映射,将输入特征映射到高维非线性空间,能够增强网络的表达能力,MSRB中采用的ReLU激活函数,经过大量模型检验,是卷积神经网络中

最常用的一种,ReLU的稀疏性能够有效缓解模型过拟合;

$$\text{ReLU}(x) = \max(0, x)$$

MSRB中大量使用批标准化层,目的是保证各层输入特征的分布相对稳定,BN层能够将数据的均值和方差控制在相对稳定的区间,这就使得学习过程更加平稳,学习到的模型对于数据的适应性更高。

## 一种基于自注意力的时频特征融合故障诊断方法

### 技术领域

[0001] 本发明涉及轴承设备故障诊断技术领域,尤其是涉及一种基于自注意力的时频特征融合故障诊断方法。

### 背景技术

[0002] 随着人工智能和大数据技术的发展以及工业4.0的提出,推动制造业进入智能制造阶段。轴承作为旋转机械设备的重要部件,其健康状态对于设备的安全稳定运行具有重要意义。据统计,旋转机械设备故障中超过30%的故障是轴承故障,然而早期轴承故障特征并不明显,一旦出现故障,轻则降低生产效率,重则造成严重的安全事故,所以高效、可靠的轴承故障诊断成为了研究热点。

[0003] 随着设备逐渐向智能化发展,获得设备生产过程中的运行数据变得越来越容易,也推动了基于数据驱动的故障诊断的发展,其中包括人工神经网络、卷积神经网络、长短时记忆网络等,都取得了较好的效果。文章[袁建虎等.基于小波时频图和CNN的滚动轴承智能故障诊断方法[J].机械设计与研究,2017,33(02):93-97.]采用连续小波变换将原始振动信号转换为二维小波时频图,然后输入到卷积神经网络进行故障分类,取得了较好效果;文章[彭成等.融合时空特征的滚动轴承多位置多类型故障诊断方法[J/OL].计算机集成制造系统:1-20.]分别采用长短时记忆网络和一维全卷积网络提取信号时序特征、空间特征,然后在全连接层进行融合,依据融合特征进行分类大大提高了诊断准确率;文章[田科位等.基于改进深度残差网络的轴承故障诊断方法[J].振动与冲击,2021,40(20):247-254.]将基于注意力机制的挤压与激励网络结构引入到残差神经网络残差块之中建立特征提取通道之间的联系,提出改进的残差网络,经实验验证模型在变工况、强噪声环境下均有良好的表现。文章[雷春丽等.小样本下基于MTF与SSCAM-MSCNN的滚动轴承变工况故障诊断方法[J/OL].计算机集成制造系统:1-19.]针对故障样本不足以及故障类别不平衡的问题,提出一种小样本的故障诊断方法,采用马尔科夫转移场将一维振动信号转化为二维特征图,其次提出条纹自校正注意力机制,能够有效捕捉全局信息,在实际数据及加噪数据中均有较好的诊断效果;综上所述,采用CNN、LSTM可以有效提取故障特征,且取得了较好的效果,但是由于CNN依赖卷积核进行特征提取,而卷积核只能关注感受野范围内的特征,从而缺乏全局特征提取能力,对于具有强非线性、非平稳性的轴承振动信号难免丢失部分隐性故障特征,影响故障诊断效果;而LSTM的优势在于时序特征提取能力,但是其记忆机制只能保持短期记忆,对于轴承振动信号这种长序列无法关注全局特征。Transformer是在2017年提出的用于自然语言处理(Natural Language Processing,NLP)相关任务的网络模型,完全依赖自注意力处理输入数据。NLP领域中将句子中的每个单词编码为独热向量并进行位置嵌入后输入到模型中,自注意力处理每个向量时关注了整个输入序列,给语言模型提供了考虑上下文的能力。自注意力对于长序列数据的特征提取具有很好的效果,所以为轴承故障诊断提供了一条新的有效途径。

## 发明内容

[0004] 本发明的目的在于提供一种基于自注意力的时频特征融合故障诊断方法,以原始振动信号和相应频谱作为输入,首先经过多层时频混合注意力进行全局特征提取的同时产生深度融合的时频融合特征;然后采用改进的多尺度残差网络提取不同尺度的、更为全面的局部特征。相较于传统基于CNN和LSTM的故障诊断方法故障诊断准确率有显著提升,且泛化能力大大增强,具有优秀的跨工况诊断能力。

[0005] 本发明提供一种基于自注意力的时频特征融合故障诊断方法,包括下述步骤:

[0006] 步骤1:采用电火花技术对待测轴承设置内圈、外圈、滚动体三种故障,每处故障直径又分为0.007、0.014、0.021inch三种,分别在4种不同负载下进行测试,由放置在驱动端和风扇端的加速度传感器分别以12kHz、48kHz的采样频率记录振动数据,最终采集到正常样本8个、外圈故障样本53个、内圈故障23个以及滚动体故障样本11个;

[0007] 步骤2:采用48kHz采样频率下的3种不同负载的数据作为原始数据集,故障类别共10类;

[0008] 步骤3:对原始振动信号进行数据预处理,采用傅里叶变换分析原始振动信号频率成分,获得频谱作为模型的频域输入;

[0009] 步骤4:按6:4的比例划分训练集与测试集,每个工况下样本数量为4750个,即每类样本475个,单个样本长度800个数据点,对应频谱为400个数据点;

[0010] 步骤5:将原始信号和相应频谱分别作为时域特征和频域特征输入到模型中,经多层基于自注意力的时频特征融合模块,进行全局特征提取的同时进行特征融合获得深度融合的时频融合特征;

[0011] 步骤6:将时频融合特征输入到改进的多尺度残差网络进一步提取局部特征,并输出分类结果;

[0012] 步骤7:使用Adam优化器对整个模型中的相关参数进行优化,得到使损失函数达到最小的最优参数;

[0013] 步骤8:将未知故障数据以相同的数据预处理方法进行处理,并输入到模型中,得到最终的分类结果。

[0014] 作为优选的,所述的基于自注意力的时频特征融合模块为:

[0015] 基于自注意力的时频特征融合模块接受1维序列输入,轴承振动的信号的时、频域特征 $x, y$ 都属于一维信号,即 $x \in \mathbf{R}^{H_1 \times 1}, y \in \mathbf{R}^{H_2 \times 1}$ ,其中 $H_1$ 为时域信号采样点数, $H_2$ 为频域信号数据点数,原始振动信号和频谱分别作为时、频域特征输入,由于频谱的数据点数通常为原始振动信号的一半,分割为 $N$ 个Patch后数据点数量是不同的,为使时、频特征的数据维度一致,首先进行Patch Embedding,即经过一个全连接层将每个Patch统一映射到 $D$ 维度,然后使用标准的可学习1D位置嵌入,以保留每个Patch的位置信息,

$$[0016] \quad z_0 = (z_0^x, z_0^y)$$

$$[0017] \quad z_0^x = [x^1 \mathbf{E}; x^2 \mathbf{E}; \dots; x^N \mathbf{E}] + \mathbf{E}_{pos}, \quad \mathbf{E} \in \mathbf{R}^{(P_1 \Pi) \times D}$$

$$[0018] \quad z_0^y = [y^1 \mathbf{E}; y^l \mathbf{E}; \dots; y^N \mathbf{E}] + \mathbf{E}_{pos}, \quad \mathbf{E} \in \mathbf{R}^{(P_2 \Pi) \times D}$$

[0019] 式中 $z_0$ 为模型输入,包括时域输入 $z_0^x$ 和频域输入 $z_0^y$ ,每个Patch乘以 $E$ 代表Patch

Embedding操作,  $E_{\text{pos}} \in \mathbb{R}^{N \times D}$  为位置嵌入向量,  $D$  是TFF内部处理数据的维度, 设置  $D=512$ ;

[0020] TFF由时频混合注意力MTFA和MLP模块组成, 在每个模块之前应用层标准化, 并且在每个模块之后使用残差连接, MLP包含两层线性层, 采用GELU激活函数, 相比于卷积网络中常用的RELU对小于0的值全部输出为零, GELU对于小于零的值会输出小的负值, 有利于稳定网络梯度流, 不容易造成梯度爆炸和梯度消失, TFF的层数  $L$  可以由人为设定, 层数越多意味着模型中参数数量越多, 训练就越困难, 而且层数过多的模型对小数据集的效果反而更差,  $L=3$  时能够获得较好的诊断效果的同时模型的时间、空间性能都比较好;

$$[0021] \quad z_1' = \text{MTFA}(\text{LN}(z_{1-1})) + z_{1-1}$$

$$[0022] \quad z_1 = \text{MLP}(\text{LN}(z_{1-1})) + z_{1-1}$$

[0023] 式中  $z_1'$  为第1层中间状态,  $z_1$  为第1层输出;

[0024] MTFA表示了单个Patch的处理流程, 根据时域、频域输入分别生成  $(q_t^i, k_t^i, v_t^i)$  和  $(q_f^i, k_f^i, v_f^i)$ , 其中  $q_t^i$  为  $Q_t$  中的一列, 即  $Q_t = \mathbf{W}^q z^x = [q_t^1, q_t^2, \dots, q_t^N]$ , 将多个Patch分别做同样的操作得到  $(Q, K, V)$ , 然后做时频混合注意力, 实现特征融合,  $\mathbf{W}^q$  为可学习的参数矩阵, 随着模型训练不断优化

$$[0025] \quad \text{MTFA}(z^x, z^y) = \left( \text{softmax}\left(\frac{Q_t K_f^T}{\sqrt{d_k}}\right) V_f, \text{softmax}\left(\frac{Q_f K_t^T}{\sqrt{d_k}}\right) V_t \right)$$

[0026] 式中  $d_k$  为  $Q, K$  的维度,  $z^x, z^y$  分别为为时、频域输入, 因为softmax的输入很大时梯度趋于零, 为了使梯度不至于太小, 所以除以  $\sqrt{d_k}$ , 使  $\frac{QK}{\sqrt{d_k}}$  的分布尽可能接近均值为0, 方差为1的正态分布;

[0027] MLP的结构包括两层线性层并采用GELU激活函数, 中间隐层的维度为  $\text{mlp\_dim}=1024$ , 在时频混合注意力后加入这一结构的作用是为了进行特征提取, 增强模型的表达能力, 由于MTFA本质上是矩阵乘法运算, 即都属于线性变换, 线性变换的学习能力是远不如非线性变换的, 所以通过加入激活函数为模型提高模型的学习能力。

[0028] 作为优选的, 所述的改进的多尺度残差网络为:

[0029] 为避免全注意力网络参数量过大, 造成模型不易训练且容易过拟合的现象, 而卷积网络结构能够利用很少的参数获得较好诊断效果, 所以选择基于卷积结构构建局部特征提取模块, 能够提高模型泛化能力同时使整个模型参数数量不至于太大, 传统的卷积神经网络依赖单一尺寸卷积核, 由于卷积核只能关注感受野范围内的特征, 其特征提取能力对于一些复杂任务往往效果不佳, 尤其是轴承振动信号本身具有非线性、非平稳性, 进一步造成隐性特征难提取, 为了提取更为丰富的特征, 提出改进的多尺度残差网络 (Multiscale Residual Net, MSRN), 由多个多尺度残差块 (Multiscale Residual Block, MSRB) 组成, MSRB内部包含一个残差连接,  $1 \times 3$  卷积、 $1 \times 5$  卷积和  $1 \times 7$  卷积, 多个尺寸的卷积核能够关注不同尺度的局部特征, 同时残差连接能够使输出中包含部分输入的信息, 缓解因网络层数加深带来的梯度消失问题, 加快模型收敛速度, MSRN堆叠3层多尺度残差块, 然后连接两层全连接层进行降维, 最后进行softmax处理得到概率分布, 概率最大的类别即为模型预测的分类;

[0030] 卷积层的作用是提取特征, 具有参数数量少、计算速度快、拟合能力强的特点, 许

多领域的优秀模型都是基于卷积构建的,CNN通过卷积核在数据上的滑动计算获得抽象程度更高的特征,过程可表示为:

[0031]  $F=f(w*x+b)$

[0032] 式中F表示卷积层输出,f表示卷积操作, $w$ 和b分别表示卷积核权重和偏置向量;

[0033] 激活层的作用是对输入特征做非线性映射,将输入特征映射到高维非线性空间,能够增强网络的表达能力,MSRB中采用的ReLU激活函数,经过大量模型检验,是卷积神经网络中最常用的一种,ReLU的稀疏性能够有效缓解模型过拟合;

[0034]  $\text{ReLU}(x)=\max(0,x)$

[0035] MSRB中大量使用批标准化层(Batch Normalization,BN),目的是保证各层输入特征的分布相对稳定,BN层能够将数据的均值和方差控制在相对稳定的区间,这就使得学习过程更加平稳,学习到的模型对于数据的适应性更高。

[0036] 本发明的有益效果为:

[0037] 针对传统轴承故障诊断模型缺乏关注全局特征能力,对于非线性、非平稳的轴承振动信号存在特征提取不充分的问题,提出基于自注意力的时频特征融合模块,将原始信号和相应频谱同时输入到模型中,经过多次时频混合注意力能够在特征融合同时进行全局特征提取,获得深度融合的时频融合特征,进一步提高了轴承故障诊断的准确性。为进一步提高模型泛化能力同时避免全注意力网络参数量过大,提出改进的多尺度残差网络,基于时频融合特征采用不同大小的卷积核进一步提取更全面的局部特征,构建先全局后局部的模型结构。相较于传统基于CNN和LSTM的故障诊断方法故障诊断准确率有显著提升,且泛化能力大大增强,具有优秀的跨工况诊断能力。

## 附图说明

[0038] 为了更清楚地说明本发明具体实施方式或现有技术中的技术方案,下面将对具体实施方式或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施方式,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0039] 图1是本发明基于自注意力的时频特征融合故障诊断方法学习过程图;

[0040] 图2是本发明的诊断流程图;

[0041] 图3是本发明基于自注意力的时频特征融合模块结构图;

[0042] 图4是本发明时频混合注意力示意图;

[0043] 图5是本发明改进的多尺度残差块结构图;

[0044] 图6是本发明网络结构图;

[0045] 图7-8是本发明同工况训练集和测试集准确率变化图;

[0046] 图9是本发明同工况故障诊断准确率对比图;

[0047] 图10是本发明跨工况故障诊断准确率对比图。

## 具体实施方式

[0048] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合附图对本发明的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是

全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0049] 如图1-2所示，本发明实施例提供一种基于自注意力的时频特征融合故障诊断方法包括如下步骤：

[0050] 步骤1：采用电火花技术对待测轴承设置内圈 (Inner Raceway, IR)、外圈 (Out Raceway, OR)、滚动体 (Ball, BO) 三种故障，每处故障直径又分为0.007、0.014、0.021inch (1inch=2.54cm) 三种。分别在4种不同负载 (0、1、2、3马力) 下进行测试。由放置在驱动端和风扇端的加速度传感器分别以12kHz、48kHz的采样频率记录振动数据，最终采集到正常样本8个、外圈故障样本53个、内圈故障23个以及滚动体故障样本11个；

[0051] 步骤2：采用48kHz采样频率下的3种不同负载 (1、2、3马力) 的数据作为原始数据集，故障类别共10类；

[0052] 步骤3：对原始振动信号进行数据预处理，采用傅里叶变换分析原始振动信号频率成分，获得频谱作为模型的频域输入；

[0053] 步骤4：按6:4的比例划分训练集与测试集。每个工况下样本数量为4750个，即每类样本475个，单个样本长度800个数据点，对应频谱为400个数据点；

[0054] 步骤5：将原始信号和相应频谱分别作为时域特征和频域特征输入到模型中，经多层基于自注意力的时频特征融合模块，进行全局特征提取的同时进行特征融合获得深度融合的时频融合特征；

[0055] 步骤6：将时频融合特征输入到改进的多尺度残差网络进一步提取局部特征，并输出分类结果；

[0056] 步骤7：使用Adam优化器对整个模型中的相关参数进行优化，得到使损失函数达到最小的最优参数；

[0057] 步骤8：将未知故障数据以相同的数据预处理方法进行处理，并输入到模型中，得到最终的分类结果。

[0058] 基于自注意力的时频特征融合模块为：

[0059] 基于自注意力的时频特征融合模块如图3所示，接受1维序列输入，轴承振动的信号的时、频域特征 $x, y$ 都属于一维信号，即 $x \in \mathbf{R}^{H_1 \times 1}, y \in \mathbf{R}^{H_2 \times 1}$ ，其中 $H_1$ 为时域信号采样点数， $H_2$ 为频域信号数据点数。原始振动信号和频谱分别作为时、频域特征输入，由于频谱的数据点数通常为原始振动信号的一半，分割为 $N$ 个Patch后数据点数量是不同的，为使时、频特征的数据维度一致，首先进行Patch Embedding，即经过一个全连接层将每个Patch统一映射到 $D$ 维度。然后使用标准的可学习1D位置嵌入 (Position Embedding)，以保留每个Patch的位置信息。

[0060]  $z_0 = (z_0^x, z_0^y)$

[0061]  $z_0^x = [x^1 \mathbf{E}; x^2 \mathbf{E}; \dots; x^N \mathbf{E}] + \mathbf{E}_{pos}$ ,  $\mathbf{E} \in \mathbf{R}^{(P_1 \Pi) \times D}$

[0062]  $z_0^y = [y^1 \mathbf{E}; y^2 \mathbf{E}; \dots; y^N \mathbf{E}] + \mathbf{E}_{pos}$ ,  $\mathbf{E} \in \mathbf{R}^{(P_2 \Pi) \times D}$

[0063] 式中 $z_0$ 为模型输入，包括时域输入 $z_0^x$ 和频域输入 $z_0^y$ ，每个Patch乘以 $\mathbf{E}$ 代表Patch Embedding操作， $\mathbf{E}_{pos} \in \mathbf{R}^{N \times D}$ 为位置嵌入向量， $D$ 是TFF内部处理数据的维度，设置 $D=512$ 。

[0064] TFF由时频混合注意力 (Mixed time-frequency attention, MTFA) 和MLP模块组

成。在每个模块之前应用层标准化 (Layer Norm, LN), 并且在每个模块之后使用残差连接。MLP包含两层线性层, 采用GELU激活函数, 相比于卷积网络中常用的RELU对小于0的值全部输出为零, GELU对于小于零的值会输出小的负值, 有利于稳定网络梯度流, 不容易造成梯度爆炸和梯度消失。TFF的层数L可以由人为设定, 层数越多意味着模型中参数数量越多, 训练就越困难, 而且层数过多的模型对小数据集的效果反而更差。L=3时能够获得较好的诊断效果的同时模型的时间、空间性能都比较好。

$$[0065] \quad z'_1 = \text{MTFA}(\text{LN}(z_{1-1})) + z_{1-1}$$

$$[0066] \quad z_1 = \text{MLP}(\text{LN}(z_{1-1})) + z_{1-1}$$

[0067] 式中 $z'_1$ 为第1层中间状态,  $z_1$ 为第1层输出。

[0068] MTFA如图4所示, 展示了单个Patch的处理流程, 根据时域、频域输入分别生成 $(q_t^i, k_t^i, v_t^i)$ 和 $(q_f^i, k_f^i, v_f^i)$ , 其中 $q_t^i$ 为 $Q_t$ 中的一列, 即 $Q_t = \mathbf{W}^q z^x = [q_t^1, q_t^2, \dots, q_t^N]$ , 将多个Patch分别做同样的操作得到 $(Q, K, V)$ , 然后做时频混合注意力, 实现特征融合。 $\mathbf{W}^q$ 为可学习的参数矩阵, 随着模型训练不断优化。

$$[0069] \quad \text{MTFA}(z^x, z^y) = \left( \text{softmax}\left(\frac{Q_t K_f^T}{\sqrt{d_k}}\right) V_f, \text{softmax}\left(\frac{Q_f K_t^T}{\sqrt{d_k}}\right) V_t \right)$$

[0070] 式中 $d_k$ 为Q, K的维度,  $z^x, z^y$ 分别为为时、频域输入。因为softmax的输入很大时梯度趋于零, 为了使梯度不至于太小, 所以除以 $\sqrt{d_k}$ , 使 $\frac{QK}{\sqrt{d_k}}$ 的分布尽可能接近均值为0, 方差为1的正态分布。

[0071] MLP的结构包括两层线性层并采用GELU激活函数, 中间隐层的维度为mlp\_dim=1024。在时频混合注意力后加入这一结构的作用是为了进行特征提取, 增强模型的表达能力。由于MTFA本质上是矩阵乘法运算, 即都属于线性变换, 线性变换的学习能力是远不如非线性变换的, 所以通过加入激活函数为模型提高模型的学习能力。

[0072] 所述的改进的多尺度残差网络为:

[0073] 为避免全注意力网络参数量过大, 造成模型不易训练且容易过拟合的现象, 而卷积网络结构能够利用很少的参数获得较好诊断效果, 所以选择基于卷积结构构建局部特征提取模块, 能够提高模型泛化能力同时使整个模型参数数量不至于太大。传统的卷积神经网络依赖单一尺寸卷积核, 由于卷积核只能关注感受野范围内的特征, 其特征提取能力对于一些复杂任务往往效果不佳, 尤其是轴承振动信号本身具有非线性、非平稳性, 进一步造成隐性特征难提取。为了提取更为丰富的特征, 提出改进的多尺度残差网络 (Multiscale Residual Net, MSRN), 由多个多尺度残差块 (Multiscale Residual Block, MSRB) 组成。MSRB如图5所示, 内部包含一个残差连接,  $1 \times 3$ 卷积、 $1 \times 5$ 卷积和 $1 \times 7$ 卷积。多个尺寸的卷积核能够关注不同尺度的局部特征, 同时残差连接能够使输出中包含部分输入的信息, 缓解因网络层数加深带来的梯度消失问题, 加快模型收敛速度。MSRN堆叠3层多尺度残差块, 然后连接两层全连接层进行降维, 最后进行softmax处理得到概率分布, 概率最大的类别即为模型预测的分类。

[0074] 卷积层的作用是提取特征, 具有参数数量少、计算速度快、拟合能力强的特点, 许多领域的优秀模型都是基于卷积构建的。CNN通过卷积核在数据上的滑动计算获得抽象程度更高的特征, 过程可表示为:

[0075]  $F=f(w*x+b)$

[0076] 式中F表示卷积层输出,f表示卷积操作,w和b分别表示卷积核权重和偏置向量。

[0077] 激活层的作用是对输入特征做非线性映射,将输入特征映射到高维非线性空间,能够增强网络的表达能力。MSRB中采用的ReLU激活函数,经过大量模型检验,是卷积神经网络中最常用的一种,ReLU的稀疏性能够有效缓解模型过拟合。

[0078]  $ReLU(x)=\max(0,x)$

[0079] MSRB中大量使用批标准化层(Batch Normalization,BN),目的是保证各层输入特征的分布相对稳定。BN层能够将数据的均值和方差控制在相对稳定的区间,这就使得学习过程更加平稳,学习到的模型对于数据的适应性更高。

[0080] 如图6所示,基于自注意力的时频特征融合故障诊断方法的试验验证如下:

[0081] 1、数据描述

[0082] 驱动端待测轴承型号为SKF6205深沟球轴承、风扇端为SKF6203深沟球轴承,采用电火花技术对待测轴承设置内圈(Inner Raceway,IR)、外圈(Out Raceway,OR)、滚动体(Ball,B0)三种故障,每处故障直径又分为0.007、0.014、0.021inch(1inch=2.54cm)三种。分别在4种不同负载(0、1、2、3马力)下进行测试。由放置在驱动端和风扇端的加速度传感器分别以12kHz、48kHz的采样频率记录振动数据,最终采集到正常样本8个、外圈故障样本53个、内圈故障23个以及滚动体故障样本11个;由于数据存在不平衡现象,所以选择48kHz采样率下采集的1、2、3马力负载下的驱动轴承振动信号,作为原始数据集,然后将数据切割为800个数据点的样本,每种工况下共得到4750条样本,故障类别为十类,即每类故障有475条样本。训练集与测试集按6:4的比例划分。

[0083] 表1轴承故障数据集描述

故障编号	故障标签	故障位置	故障直径
0	DE_BO_07	Ball	0.007
1	DE_BO_14	Ball	0.014
2	DE_BO_21	Ball	0.021
3	DE_IR_07	Inner Raceway	0.007
[0084] 4	DE_IR_14	Inner Raceway	0.014
5	DE_IR_21	Inner Raceway	0.021
6	DE_OR_07	Out Raceway	0.007
7	DE_OR_14	Out Raceway	0.014
8	DE_OR_21	Out Raceway	0.021
9	Normal	---	---

[0085] 2、模型结构

[0086] TFF的层数L可以由人为设定,层数越多意味着模型中参数数量越多,训练就越困

难,而且层数过多的模型对小数据集的效果反而更差。选择 $L=3$ ,获得很好的诊断效果的同时模型的时间、空间性能都比较好。为避免全注意力网络参数量过大,造成模型不易训练且容易过拟合的现象,而卷积网络结构能够利用很少的参数获得较好诊断效果,所以选择基于卷积结构构建局部特征提取模块,能够提高模型泛化能力同时使整个模型参数数量不至于太大。传统的卷积神经网络依赖单一尺寸卷积核,由于卷积核只能关注感受野范围内的特征,其特征提取能力对于一些复杂任务往往效果不佳,尤其是轴承振动信号本身具有非线性、非平稳性,进一步造成隐性特征难提取。选择3层多尺度残差块来进行进一步局部特征提取。

[0087] 表2模型参数描述

参数名称	参数描述
时、频域输入 (x, y)	x.shape=(800, 1), y.shape=(400, 1)
Dropout	MLP(rate=0.1), emb(rate=0.1)
[0088] 内部数据维度	D=512、mlp_dim=1024
损失函数和优化器	CrossEntropyLoss, Adam
学习率	0.00005

[0089] 3、结果分析

[0090] 为验证所提出模型的有效性,采用以下4种模型进行对比实验,将故障诊断准确率(accyracy)作为重要指标。

[0091] TTF-MRSN:即所提出的模型,模型主体结构包括三层Transformer Encoder和三层MSRB组成的多尺度残差网络,模型参数设置如表3.1所示。

[0092] TFF-Res:将模型(1)的MSRN部分替换为3层单尺度残差网络,卷积核大小为5。

[0093] TTF:全注意力网络结构,即将所提出模型的MSRN部分删除掉,其余参数与模型(1)相同。

[0094] MSRN-6:采用6层堆叠的MSRB组成多尺度残差网络,由于无法进行特征融合,所以采用频谱作为模型输入。

[0095] 首先,将模型在1马力(Horse Power,HP)负载下的数据集上进行训练,训练集与测试集之比为6:4,模型训练20轮的准确率曲线如图7-8所示。避免训练结果的偶然性,重复5次实验后取平均值。MSRN网络在测试集上准确率95%左右,MSRN网络无论在诊断准确度和模型收敛速度方面都明显更差,原因主要有两点:(1)依赖卷积结构的模型特征提取能力不足,不能充分挖掘隐性特征;(2)比一般的情况设置更少的训练集,更多的测试集,特征提取能力较差的模型表现就会相对更差。其余三种模型都是基于提出的TFF构建的,最终测试集准确率相差不大,都能达到98%以上,单就同工况诊断准确度三种模型均表现较好,但是由于TFF和TFF-Res模型相较于模型参数数量更少,所以在模型收敛速度上略强。

[0096] 为进一步对比同工况故障诊断准确度,将各模型在3种工况下后10轮测试集上的诊断准确率的平均值作为模型准确率,如表3所示。三种基于TFF的模型准确率均能达到

98%的水平,且模型在各工况下诊断准确率均优于对比模型,MSRN平均准确率只有94%左右,如图9所示。说明TFF模块有效提高了模型的诊断准确率。

[0097] 表3同工况故障诊断准确率对比

		1HP	2HP	3HP
	TFF-MSRN	98.38	98.21	98.51
[0098]	TFF-Res	98.04	97.99	97.94
	TFF	98.25	97.73	98.38
	MSRN	94.06	93.97	93.48

[0099] 为了适应实际生产条件下的故障诊断,必须让模型具备更强的泛化能力,同工况诊断准确率并不能衡量一个模型泛化能力的高低。通过跨工况故障诊断准确率来测试模型泛化能力,如图10所示,其中1-2表示在1HP数据上训练的模型,采用2HP数据进行故障诊断测试。由表4可知,模型TFF-MSRN在各跨工况条件下均取得最优效果,1-2条件下准确率与TFF-Res相近能达到95%的水平,但是1-3条件下有89%的准确率,比TFF-Res高3%以上,说明提出的MSRN优于单一尺度的残差网络。TFF模型中没有关注局部特征的结构,所以跨工况诊断准确率较TFF-MSRN、TFF-Res更差。平均准确率方面,由1HP数据训练的模型跨工况准确率>2HP>3HP,造成这一现象的原因可能由于轴承负载越大,振动信号中混合的噪声就越多,其非线性、非平稳性就越明显,所以模型能学习到的特征就越少,在进行构建诊断模型时尽量选择低负载的数据进行训练,有助于模型泛化能力的提高。综上所述,TFF-MSRN模型同工况、跨工况条件下具有更高的诊断准确率以及更好的泛化能力。

[0100] 表4跨工况故障诊断准确率对比

		1-2	1-3	2-1	2-3	3-1	3-2
	TFF-MSRN	95.30	89.94	91.49	83.94	80.91	89.10
[0101]	TFF-Res	95.06	86.35	91.11	78.92	80.12	86.68
	TFF	91.29	79.56	83.24	78.37	81.06	86.83
	MSRN	88.09	81.55	86.48	83.97	72.48	82.45

[0102] 4、结论

[0103] 针对传统轴承故障诊断模型特征提取不充分的问题,提出基于自注意力的时频特征融合模块,获得时频融合特征,结合改进的多尺度残差网络进一步提高模型泛化能力,提出TFF-MSRN故障诊断模型,并运用CWRU轴承数据集和工厂减速机轴承故障数据集进行验证,实验结果表明,所提出方法具有较高的诊断准确度和较强泛化能力。同时得出以下结论:

[0104] (1) TFF模块基于自注意力,同时关注时、频域全局特征,能够有效提取振动信号的深层特征,大大提高了模型诊断准确度。

[0105] (2) MSRN采用多个不同大小的卷积核进一步提取更为全面的局部特征,有效提高

了模型的泛化能力,具体表现为模型跨工况诊断能力大大提升。

[0106] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,简称ROM)、随机存取存储器(Random Access Memory,简称RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0107] 最后应说明的是:以上所述实施例,仅为本发明的具体实施方式,用以说明本发明的技术方案,而非对其限制,本发明的保护范围并不局限于此,尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,其依然可以对前述实施例所记载的技术方案进行修改或可轻易想到变化,或者对其中部分技术特征进行等同替换;而这些修改、变化或者替换,并不使相应技术方案的本质脱离本发明实施例技术方案的范围。都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求的保护范围为准。

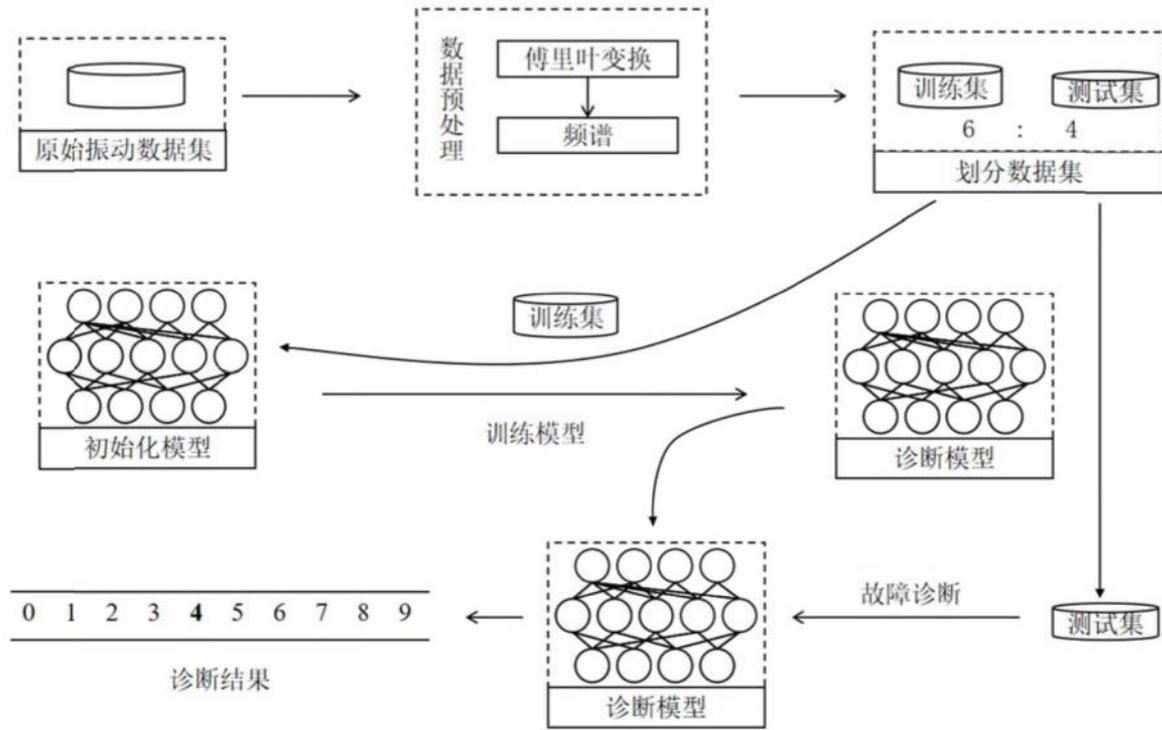


图1

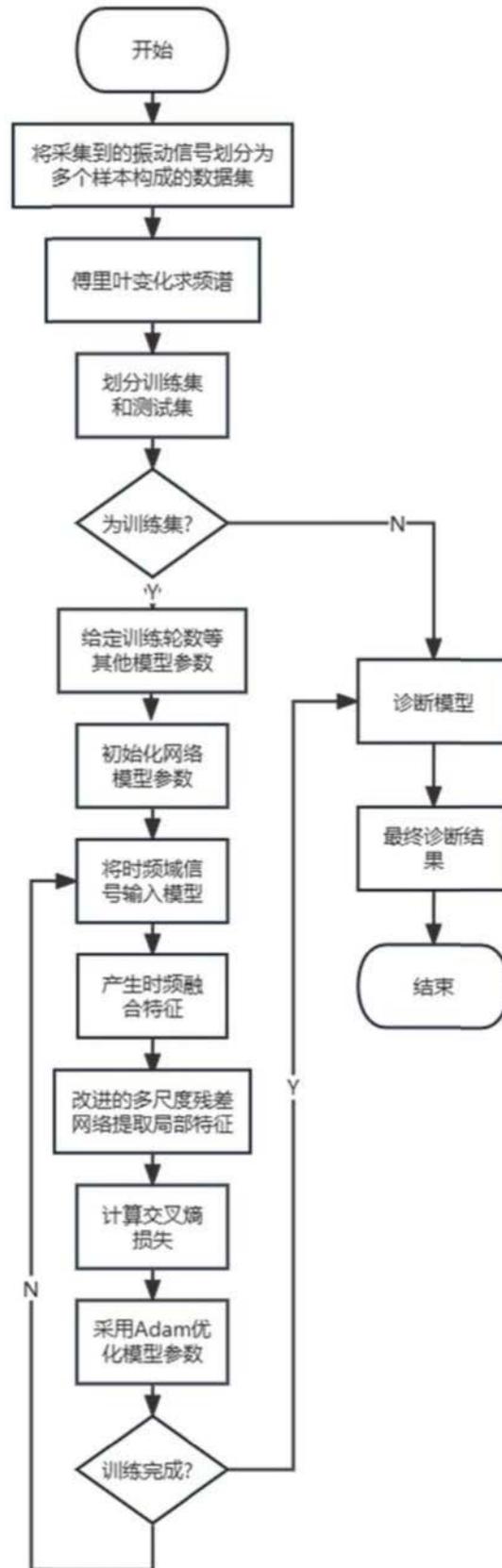


图2

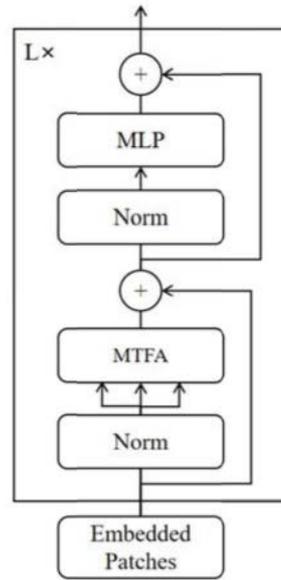


图3

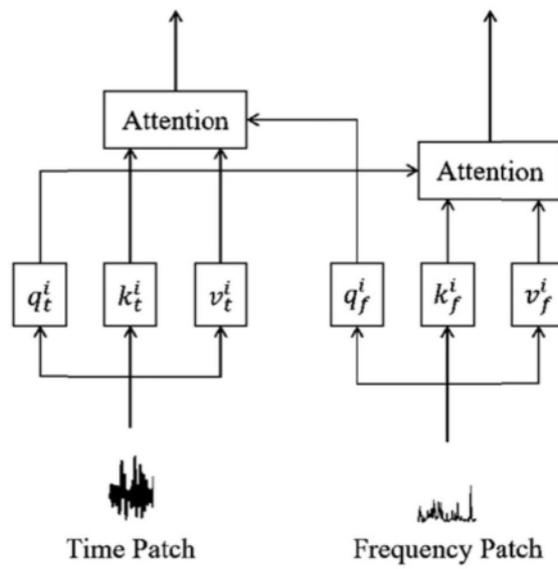


图4

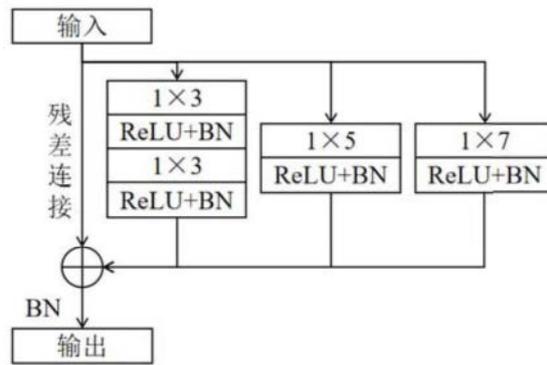


图5

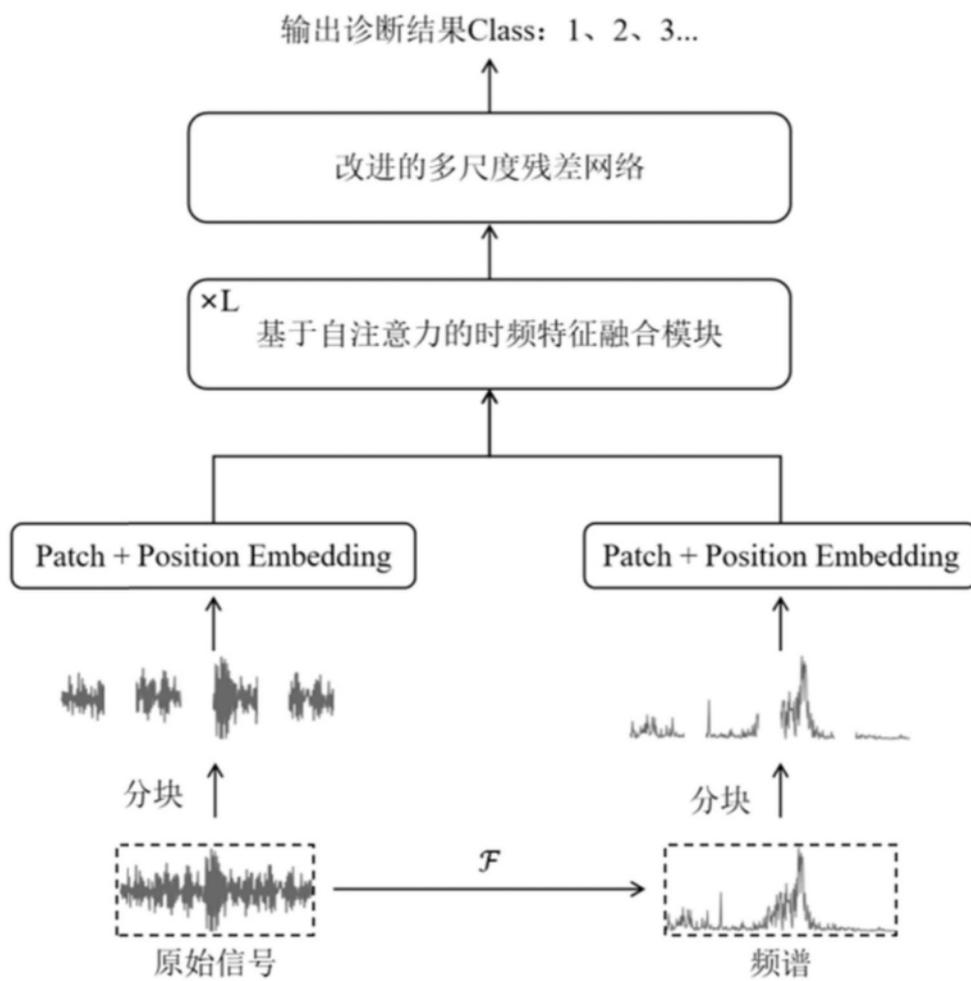


图6

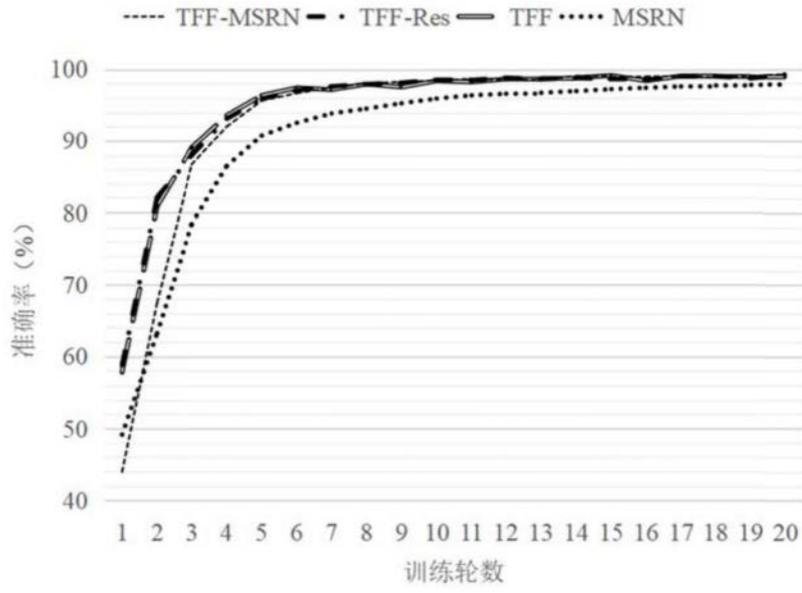


图7

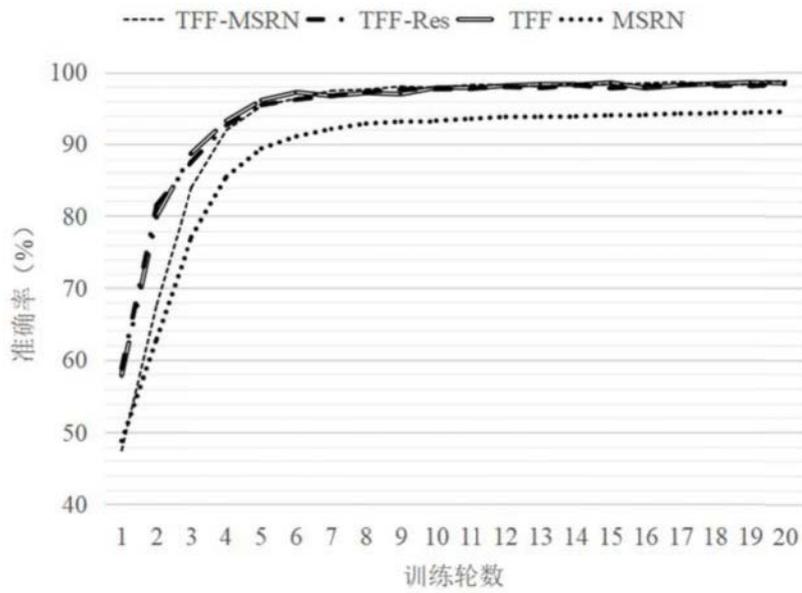


图8

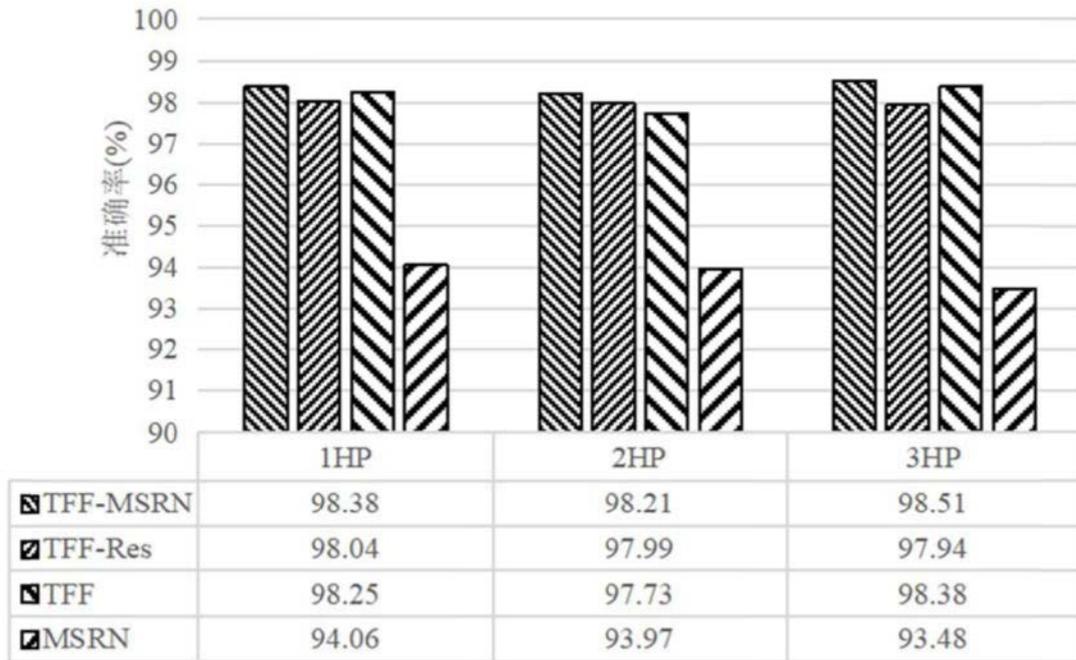


图9

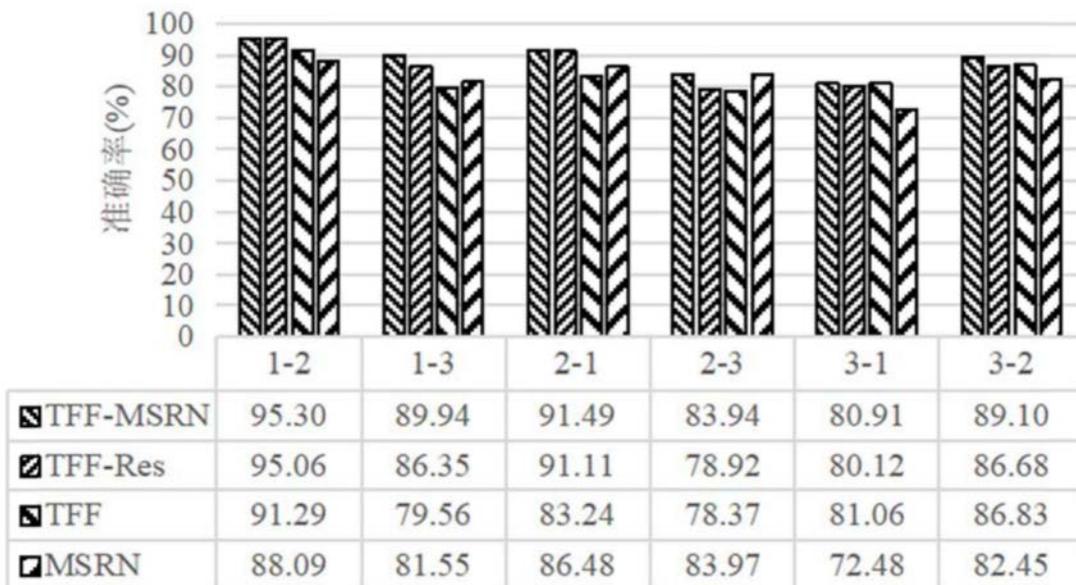


图10