

基于数据增强的可解释旋转机械故障诊断^{*}

张 昊 王海茹 马继东

(东北林业大学机电工程学院 哈尔滨 150036)

摘 要: 为解决机械故障诊断中稀疏标签、数据不足等问题,同时提升诊断性能与模型可解释性,本文提出了一种嵌入时频分析的时频卷积神经网络(TF-CNN)模型。该模型结合多种高效数据增强方法与嵌入时频分析的卷积神经网络,从振动信号中提取多尺度关键时频特征。时频卷积层利用时频变换的物理可解释性与卷积神经网络的自主特征提取能力,精准捕捉信号关键信息,并通过分类器实现故障诊断。在 CWRU 数据集上的实验结果表明,TF-CNN 模型的诊断准确率高达 99.8%,显著优于基线方法。此外,频率响应分析验证了模型对信号关键频段的关注能力,进一步增强了诊断过程的物理可解释性。本文提出的 TF-CNN 模型充分结合时频分析与深度学习优势,提供了一种高效、精准且可解释的解决方案,为工业故障诊断领域带来了新思路与实践指导。

关键词: 故障诊断;时频分析;数据增强;频率响应分析

中图分类号: TH17; TN911.7 **文献标识码:** A **国家标准学科分类代码:** 410.75

Interpretable rotating machinery fault diagnosis based on data enhancement

Zhang Hao Wang Hairu Ma Jidong

(College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150036, China)

Abstract: To address the challenges of sparse labels and insufficient data in mechanical fault diagnosis while enhancing diagnostic performance and model interpretability, this paper introduces a time-frequency convolutional neural network (TF-CNN) model embedded with time-frequency analysis. By integrating advanced data augmentation techniques with a convolutional architecture leveraging time-frequency transformation, the model effectively extracts multi-scale, key time-frequency features from vibration signals. The time-frequency convolutional layer combines the physical interpretability of time-frequency analysis with the autonomous feature extraction capabilities of convolutional neural networks, enabling precise identification of critical signal characteristics and fault diagnosis through classification mechanisms. Experimental validation on the CWRU dataset demonstrates that the TF-CNN model achieves a diagnostic accuracy of 99.8%, significantly outperforming baseline methods. Additionally, frequency response analysis confirms the model's ability to emphasize key signal frequency bands, further strengthening its physical interpretability. By seamlessly integrating the strengths of time-frequency analysis and deep learning, the TF-CNN model offers an innovative, efficient, and interpretable approach to industrial fault diagnosis. This work provides valuable insights and practical guidance for advancing fault diagnosis techniques, paving the way for robust applications in complex industrial scenarios.

Keywords: fault diagnosis; time-frequency analysis; data enhancement; frequency response analysis

0 引 言

旋转机械作为现代制造业的关键设备,由于长期处于恶劣环境中运行,例如腐蚀、高温、高负荷和高转速等条件,其核心部件如轴承和齿轮极易发生故障。这不仅显著增加了维修成本,甚至可能引发灾难性后果。因此,及时、准确

地预测旋转机械故障对降低经济损失、提高生产效率具有重要意义。随着传感器技术和工业物联网的发展,获取大量旋转机械运行数据变得更加便捷,通过对这些数据的有效处理,可以显著提升机械故障诊断的效率,从而推动数据驱动的分析方法在该领域的广泛应用。

数据驱动的故障诊断通常包括 3 个步骤:数据采集、特

征提取和故障分类,其中特征提取是故障诊断的核心。传统方法主要依赖信号处理技术,如稀疏编码^[1]、小波变换^[2]、模态分解^[3]等,通过对原始时域信号进行去噪,并转换至频域或时频域,以提取信号的内在特征。此外,一些研究尝试将传统数据驱动方法与机器学习相结合,以增强信号特征提取的能力。例如,连璞等^[4]提出的结合复合多尺度增长熵与优化核极限学习机的方法,考虑了信号幅值信息,提升了故障诊断的准确性和稳定性,特别是在处理非线性和复杂信号时显示了更好的效果。赵轲等^[5]提出基于模糊逻辑和领域适应技术的方案,通过生成模糊规则解决源域与目标域之间的数据分布差异,提升了跨域数据的利用效率。尽管传统数据驱动方法在一定程度上利用了领域知识与物理信息,使故障特征具备可解释性,但这些方法依然面临着许多局限性。首先,它们通常需要针对不同数据样本进行专门的预处理,限制了其在多样化数据环境中的泛化能力。其次,人工特征提取和选择既繁琐又耗时,尤其在面对海量异构数据时,这一过程尤为困难^[6]。因此,深度学习等数据驱动的方法已成为实现旋转机械智能运维的有效解决方案。

近年来,深度学习技术在异常检测、故障诊断和预测性维护等任务中取得了显著进展。深度神经网络通过对一维振动信号进行非线性映射,能够在信号的高维特征空间中提取故障信息,省去了传统方法中繁琐的人工特征提取过程,可以端到端地处理大量的信号数据。例如,Guo等^[7]提出的卷积注意力融合网络在小样本和增量学习任务中表现优异。Li等^[8]利用贝叶斯深度学习框架提高了模型精度和数据校准能力。张家安等^[9]融合卷积神经网络和长短期记忆网络对多传感器振动信号进行加权融合。冯肖亮等^[10]借助多头注意力机制学习数据的空间信息,建立交互联系。然而,尽管深度学习模型能从海量数据中学习复杂的模式,但其“黑箱”性质使得模型的决策过程难以解释,限制了其在高可靠性要求场景中的应用^[11]。此外,深度学习模型通常需要大量标注数据进行训练和微调,而实际应用中往往缺乏足够的标注数据,尤其是在某些行业中,数据标注的成本较高,这成为了深度学习应用的瓶颈^[12]。

为了解决上述两个问题,数据增强方法和可解释性神经网络得到了广泛关注。数据增强技术通过生成更多的训练样本来缓解数据不足的问题,减少对大量人工标注数据的依赖。同时,随着可解释性神经网络的发展,越来越多的技术被提出以增强深度学习模型的可解释性,例如,张龙等^[13]将连续小波变换与卷积层结合,设计了小波卷积层以提取可解释特征。Wang等^[14]结合深度小波自动编码器与极限学习机增强模型的可解释性。Li等^[15]采用连续小波卷积层替代标准卷积层以提取更具意义的核信息。尽管这些方法在一定程度上揭示了深度学习模型的诊断机制,但它们仍然依赖大量标注数据来微调模型,且多依赖于不同领域间的相似性预学习,难以同时解决数据稀疏性和可解

释性问题。

因此,本研究旨在解决旋转机械智能故障诊断中的两个关键问题:可解释性和标签信号不足。为此,本文提出了一种基于数据增强的时频卷积网络方法。该方法将增强后的时域和频域信号赋予不同的权重融合,输入到嵌入时频变换方法的传统卷积网络中以提取与故障相关的时频信息。在此基础上,我们将时频卷积视为预处理层,构建一个用于故障诊断任务的可解释性时频网络。这一方法不仅有效减少了对大量标注数据的依赖,还通过时频信息融合增强了模型的可解释性,为旋转机械的智能运维提供了一个切实可行的解决方案。本文的主要贡献为:

1) 本文设计了一套专门的时频域数据增强方法,减少了稀疏标签和数据不足的影响,有效提高旋转机械故障诊断模型的鲁棒性。

2) 将时频变换方法嵌入到卷积层中,结合传统卷积神经网络有效提取了振动信号的时频信息,提高模型的诊断性能。

3) 对训练好的预处理层进行频率响应分析,解释了模型如何处理不同频率的信息。并与现有基线模型进行对比实验及消融实验,结果说明模型在处理复杂信号时具有卓越的诊断性能。

1 理论基础

1.1 时频变换

时频变换是一种用于分析非平稳信号在时间域和频率域中分布的有效方法。通过内积运算,将信号从时间域映射到时频域,从而揭示信号在时间和频率上的动态变化特性^[16]。其中,短时傅里叶变换(short-time Fourier transform, STFT)是常用的时频分析方法之一,它通过引入窗口函数,能够在不同时间点上捕获信号在不同频率的能量分布,其频谱计算公式如下:

$$S(t_0, f) = \int_{-\infty}^{+\infty} \mathbf{x}(t) * \mathbf{w}(t - t_0) * e^{-2\pi jft} dt \quad (1)$$

其中,*代表卷积运算(后续公式中的运算同样遵循这一卷积计算方式), t_0 为时间信息,表示窗口中心。 f 表示频率信息。 $\mathbf{w}(t - t_0)$ 为窗口函数,能够捕获局部信息。 $e^{-2\pi jft}$ 为傅里叶变换的复指数基函数。

时频变换方法还包括小波变换(wavelet transform, WT)和Wigner-Ville分布(wigner-ville distribution, WVD)等,这些方法通过采用不同的窗函数进行卷积,适应各种信号特征的提取需求^[17]。小波变换通过将信号与一组缩放和平移的小波基函数进行卷积,能够有效分析信号的多尺度特征。而WVD则通过对信号的自相关函数进行傅里叶变换,提供高分辨率的时频分布,适用于非平稳信号的详细分析。其具体计算公式如下:

$$\mathbf{W}(a, b) = \int_{-\infty}^{+\infty} \mathbf{x}(t) * \varphi^* \left(\frac{t-b}{a} \right) dt \quad (2)$$

$$W_x(t_0, f) = \int_{-\infty}^{+\infty} x\left(t_0 + \frac{t}{2}\right)^* x\left(t_0 - \frac{t}{2}\right) * e^{-2\pi jft} dt \quad (3)$$

其中, $\varphi^*(t)$ 为小波基函数, a 和 b 分别为缩放因子和平移因子。时频变换通过不同的内积窗函数将非平稳信号从时间域投影到时频域,以揭示信号的时频分布特性。

如图 1 所示,输入的模拟振动信号表示加速度随时间的变化,其幅值反映了振动强度。时频分析能够提取信号随时间变化的频率成分,并生成各时刻的频谱幅值,提供更丰富的时变特征。将生成的时频图作为深度学习网络的输入,不仅能够同时捕获复杂非平稳信号的时间分辨率和频率分辨率,还可以提高特征提取的精度。因此,时频变换在机械故障诊断的特征提取过程中起着关键作用。

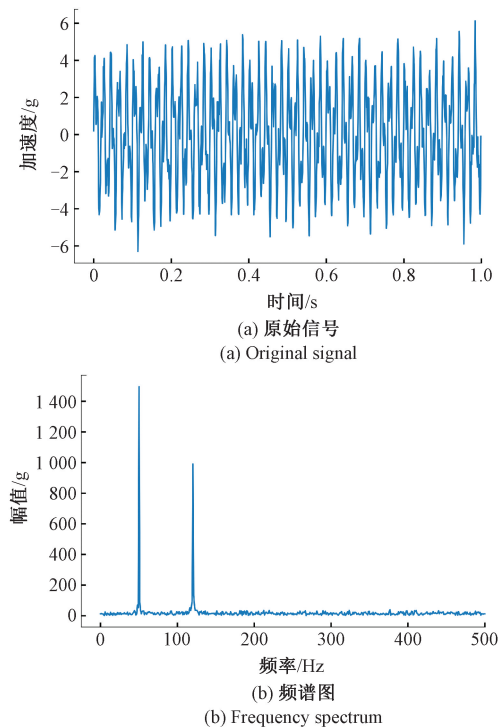


图 1 振动信号的频谱强度

Fig. 1 Spectral intensity of vibration signal

1.2 深度卷积网络

深度卷积网络(convolutional neural network, CNN)是一种具有卷积计算的深度神经网络,是深度学习中的代表性算法之一。由于其具有保留空间信息的能力,许多研究人员选择 CNN 作为故障诊断任务的强大特征提取器^[18]。CNN 模型通常由输入层、卷积层、池化层、全连接层和输出层组成,用于机械故障诊断时,输入数据多为一维信号,因此 CNN 模型也多为一维架构。在 CNN 模型中,输入层负责接收并预处理原始信号,为后续层提供标准化输入;卷积层通过卷积运算提取特征,是 CNN 的核心模块;池化层则通过下采样压缩数据,保留关键信息,同时降低计算复杂度;全连接层在特征提取的基础上进行整合,并

最终连接到输出层完成预测或分类任务。

卷积层的主要作用是使用卷积核作为滑动窗口,对输入信号进行局部加权求和,提取关键的局部特征值。对于多通道输入,特征映射的计算公式如下:

$$y[i] = \sum_{c=1}^C \sum_{j=1}^k w_c[j] * x_c[i+j-1] + b \quad (4)$$

其中, $y[i]$ 表示在位置 i 的卷积输出值, $w_c[j]$ 和 b 表示第 c 个通道的卷积核的权重和偏差, k 为卷积核的大小, $x_c[i+j-1]$ 为输入数据在滑动窗口内的值。

对所有通道的特征进行加权求和后,结合激活函数和降维操作,将原始信号逐步转化为高维特征。在分类阶段,这些高维特征经过扁平化处理,并输入全连接层以进一步提取类别信息。最终,通过 Softmax 函数计算每个类别的概率分布,并利用交叉熵损失函数评估分类性能,其公式为:

$$L(r, p) = - \sum_i r_i \cdot \log(p_i) \quad (5)$$

其中, i 表示类别数, r 和 p 分别表示真实值和预测概率。分类损失衡量真实标签与预测之间的差异。通过基于反向传播的训练策略和随机梯度下降等优化算法,模型的预测可以逐渐接近真实标签。

2 研究方法

本文提出了一种基于数据增强的时频卷积神经网络(TF-CNN)用于故障诊断任务,具体步骤如下:

数据预处理:从旋转机械中获取故障数据,将振动信号分割为长度为 1 024 的样本,并将其归一化至 0~1,确保数据分布的一致性。

数据增强:对时域数据采用多种增强方法,包括高斯噪声添加、水平翻转、垂直翻转、平移、掩码、衰减和缩放操作,以扩充数据集并提升模型的泛化能力。

时频特征提取:将增强后的信号输入到嵌入时频变换的卷积神经网络中,通过时频卷积层提取多尺度、多频段的关键时频特征。通过多层卷积和池化操作逐步提取高级特征,同时降低特征维度,提高模型的计算效率。

特征整合与诊断:在特征提取层后连接全连接层和分类器,对提取的时频特征进行进一步学习,并完成故障类别的分类诊断。

整体诊断框架如图 2 所示。

2.1 数据增强

采用数据增强方法可以从原始信号中学习潜在表示,能够提高模型从有限数据中学习的能力。数据增强方法在计算机领域发展成熟,但在机械信号分析中不太常见。Wang 等^[19]提出 4 种信号增强方法:水平翻转、垂直翻转、中心翻转以及高斯噪声。为了进一步扩展训练数据,本文引入了额外的信号增强方法:平移、掩码、衰减、缩放。

高斯噪声:与传统的直接在原始信号中添加随机高斯分布的噪声处理方法不同,本文同时控制了原始信号与噪

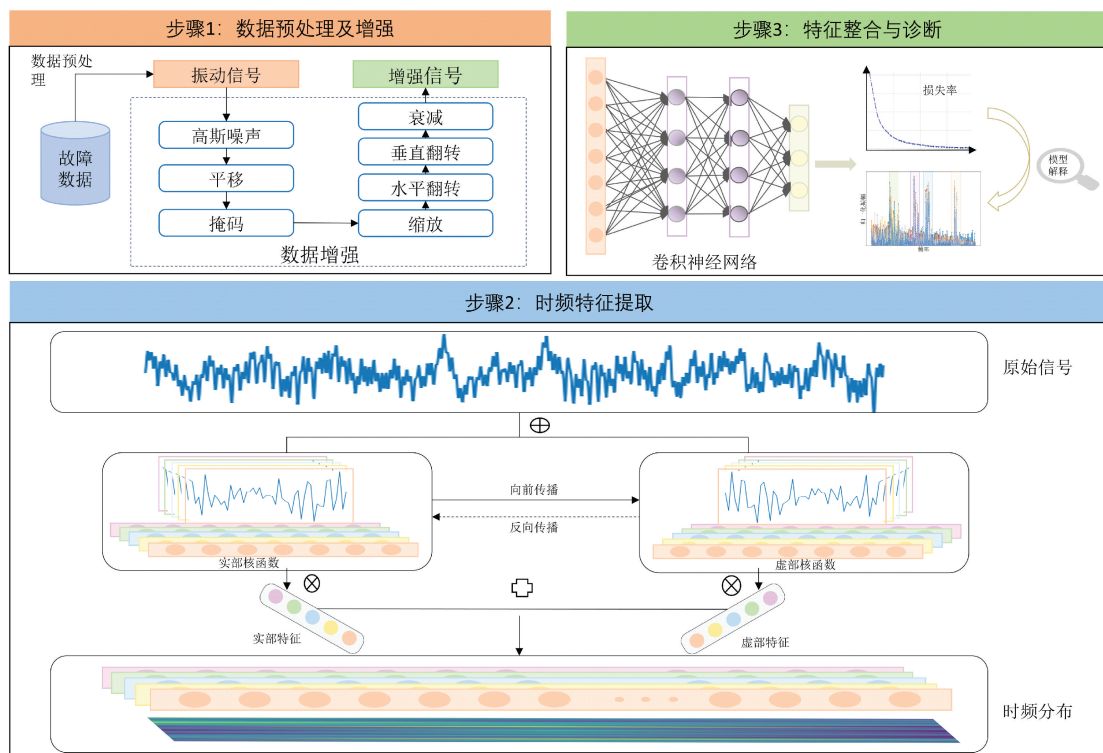


图 2 时频卷积神经网络框架

Fig. 2 Time-frequency convolutional neural network framework

声信号的信噪比,来模拟不同的噪声水平。能够很好的模拟真实工业场景中的噪声干扰。

平移:将振动信号沿时间轴进行平移操作,来模型时间上的延迟或提早现象。

掩码:随机设置 t_{mask} ,应用掩码将连续序列 $[t_0, t_0 + t_{\text{mask}}]$ 归零,以模拟部分时间段数据缺失的情况。

衰减:随机选择振动信号,对其振动幅度进行非线性的增大或缩小操作,以模拟信号强度逐渐增加或减弱的现象。

缩放:通过调整增益系数 ϑ 调整振动信号值,以模拟不同幅度条件下的数据。其中 $\vartheta \in [0.5, 1.5]$,增益系数 $\vartheta < 1$ 时会降低振动值, $\vartheta > 1$ 时会增加振动值。

水平翻转:将时间序列信号沿时间轴进行翻转,来模拟时间上完全对称的情况。

垂直翻转:将信号的正负号颠倒,来模拟相反极性条件下的数据。

图 3 说明了上述信号增强方法的效果。在训练过程中,7 种增强方法依次应用于原始时间信号数据。通过增加训练数据集的大小,能够很好的解决标记数据的数据或质量不足以训练深度学习模型的挑战。

2.2 嵌入时频变换的卷积神经网络

时频变换能够将信号分解为随时间变化的频率分量,具有良好的物理可解释性。然而,它依赖于预设的变换方法,对数据集的自适应能力较弱。相比之下,CNN 能够自

主、自适应地从输入信号中提取高维特征,减少对专业知识的依赖,同时提高故障检测的精度和鲁棒性。然而,CNN 的决策过程通常是“黑箱式”的,缺乏直观的可解释性。为了结合这两种方法的优势,本文将时频变换嵌入到卷积层中,用于提取时频特征。由于时频变换和卷积操作都基于内积计算,时频变换可以视为一种特殊的卷积操作,其中卷积核等效于时频变换的基函数。这种融合方法不仅保留了时频变换的物理可解释性,还充分发挥了 CNN 在特征提取上的自适应能力,为故障检测提供了一种更加有效且可解释的解决方案。

基于时频变换的卷积层采用两个独立的卷积核,分别对输入信号进行卷积处理:一个卷积核专门用于提取信号的实部特征,另一个卷积核用于提取信号的虚部特征。在处理机械振动信号 $x(t)$ 时,卷积层沿着信号的时域方向滑动卷积核,分别对信号的实部和虚部进行卷积操作。最终输出的实部特征和虚部特征是二维矩阵,它们分别表示信号在时域和频域中的特征组合。这种设计能够更有效地捕捉信号的时频信息,从而提升信号处理的准确性与效果。计算公式为:

$$\begin{cases} \mathbf{X}_{\text{real}}(t) = \sum_{t_0} \mathbf{x}(t_0) * \mathbf{h}_{\text{real}}(t - t_0) \\ \mathbf{X}_{\text{imag}}(t) = \sum_{t_0} \mathbf{x}(t_0) * \mathbf{h}_{\text{imag}}(t - t_0) \end{cases} \quad (6)$$

其中, t_0 为卷积窗口内的采样点, \mathbf{h}_{real} 与 \mathbf{h}_{imag} 分别为

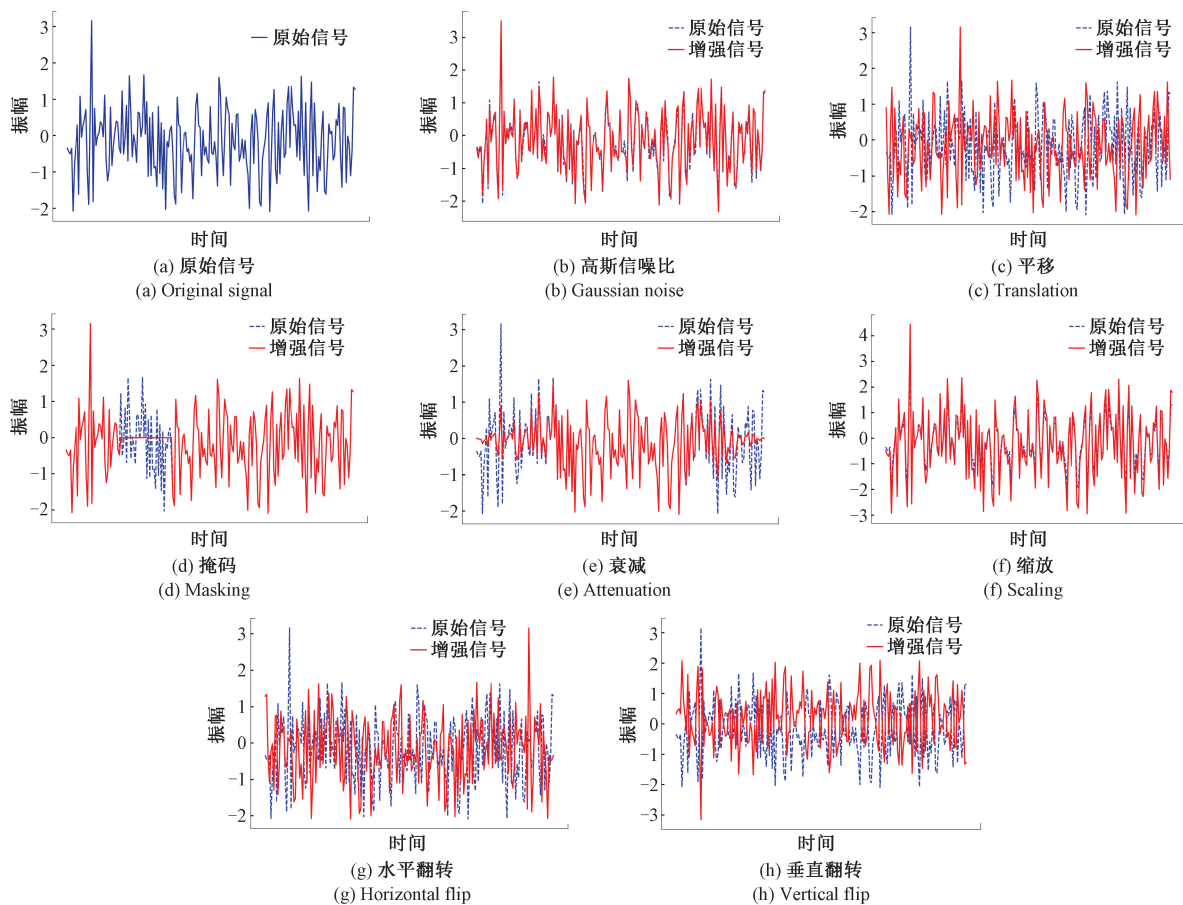


图 3 7 种增强方法的可视化

Fig. 3 Visualization of seven enhancement methods

短时傅里叶变换的实部核函数和虚拟核函数。

最终通过模值计算反映信号在特定时刻和频率下的总能量,输出的时频分布可以表示为:

$$\mathbf{X}_{\text{mod}}(t) = \sqrt{(\mathbf{x}(t) * \mathbf{h}_{\text{real}}(t))^2 + (\mathbf{x}(t) * \mathbf{h}_{\text{imag}}(t))^2} \quad (7)$$

综上,本文所提出的嵌入时频变换的卷积层结构如图 4 所示。

本文通过对训练后的卷积层进行频率响应分析来解释卷积层在时频域上的行为。频率响应表示卷积核对不同频率信号的反应能力。具体步骤为:

1) 使用傅里叶变换将卷积核的时域特性转换为频域特性,得到每个卷积核的频率响应。

2) 计算所有卷积核的频率响应的平均值,以此表示卷积层整体对不同频率成分的响应能力。

最终本文将可解释的卷积层作为预处理层,与主干 CNN 相结合,形成新的网络。该网络可以从原始振动信号中提取与故障相关的时频信息,高效诊断机械设备的故障状态。

3 实验验证及结果分析

3.1 数据集

本文使用广泛用于机械故障诊断领域的 CWRU 数据集。数据集在 4 种电机负载下,以 12 kHz 和 48 kHz 采集正常轴承和具有单点缺陷的损坏轴承的振动信号。包含有 3 种不同的轴承故障诊断类型:滚动体、内圈和外圈故障。对于每种故障类型,分别考虑不同的故障直径,即 0.007、0.014 和 0.021 英寸。在表 1 中,将健康轴承和 3 种故障模式(包括内圈故障、滚动体故障和外圈故障)根据不同的故障大小分为 10 类(一个健康状态和 9 种故障状态)。

3.2 实验设置

所提出的框架完全使用 Pytorch 实现,并在具有 Windows11、16 核 CPU(32 GB RAM)和 NVIDIA RTX 4060 GPU(12 GB VRAM)的系统上进行评估。本文将振动信号截断为长度 1 024 的样本,每种类别有 470 个样本。训练集、验证集与测试集的比例为 6:2:2,对每个类别的样

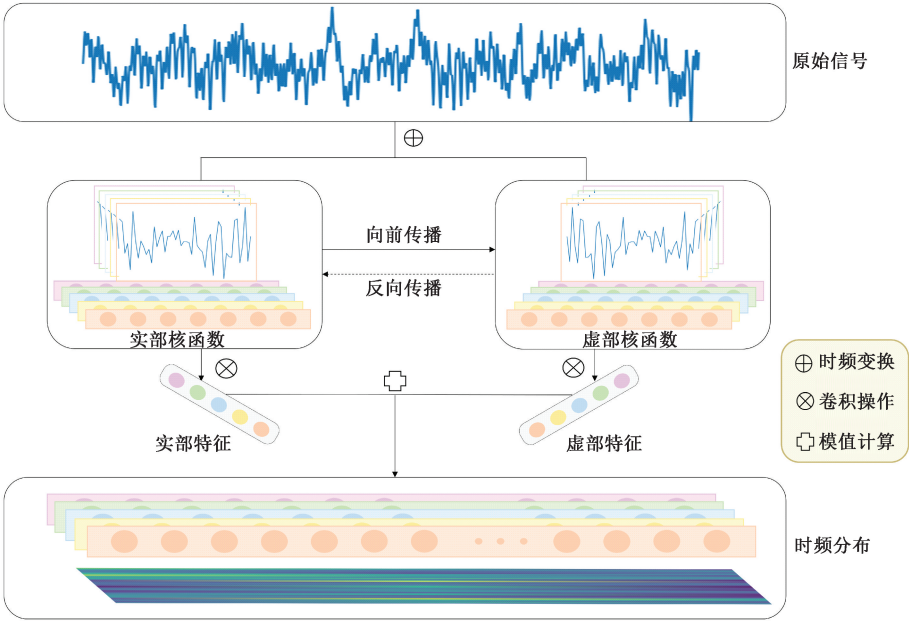


图 4 嵌入时频变换的卷积层结构

Fig. 4 Convolutional layer structure embedded in time-frequency transform

表 1 CWRU 数据集描述

Table 1 Description of CWRU dataset

标签	故障类型	描述
N	健康状况	1 791 rpm 和 0 HP 时的正常轴承
I1	内圈 1	1 797 rpm 和 0 HP 时内圈故障为 0.007 英寸
I2	内圈 2	1 797 rpm 和 0 HP 时内圈故障为 0.014 英寸
I3	内圈 3	1 797 rpm 和 0 HP 时内圈故障为 0.021 英寸
B1	滚动体 1	1 797 rpm 和 0 HP 时滚动体故障为 0.007 英寸
B2	滚动体 2	1 797 rpm 和 0 HP 时滚动体故障为 0.014 英寸
B3	滚动提 3	1 797 rpm 和 0 HP 时滚动体故障为 0.021 英寸
O1	外圈 1	1 797 rpm 和 0 HP 时外圈故障为 0.007 英寸
O2	外圈 2	1 797 rpm 和 0 HP 时外圈故障为 0.014 英寸
O3	外圈 3	1 797 rpm 和 0 HP 时外圈故障为 0.021 英寸

表 2 超参数设置

Table 2 Hyperparameter settings

超参数	设置
optimizer	Adam
lr_scheduler	stepLR
learning rate	1×10^{-4}
gamma	0.9
kernel_size	11
mid_channel	32
n_epochs	200
clamp_flag	True

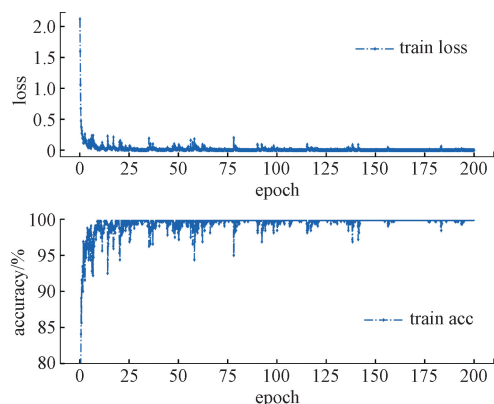
本数据进行随机分割。本文沿用传统 CNN 超参数设置,选用 Adam 作优化器,学习率衰减因子 (gamma) 设置为 0.9。具体来说,TF-CNN 的超参数总结如表 2 所示。对于其他对比方法,本文使用了其公开的最优参数设置,以确保公平比较。此外,为避免诊断结果中的偶然性,在不改变参数的情况下重复实验 15 次后,对每个模型的准确性进行平均。

3.3 实验结果

使用所选的超参数,TF-CNN 模型在训练集、验证集及测试集上的表现如图 5 所示。在模型训练的早期阶段,验证集和测试集的损失率波动较大,但在约 50 次迭代后,准确率和损失率逐渐趋于稳定,且波动幅度显著降低。这表明,本文提出的模型具有良好的收敛性,能够快速适应

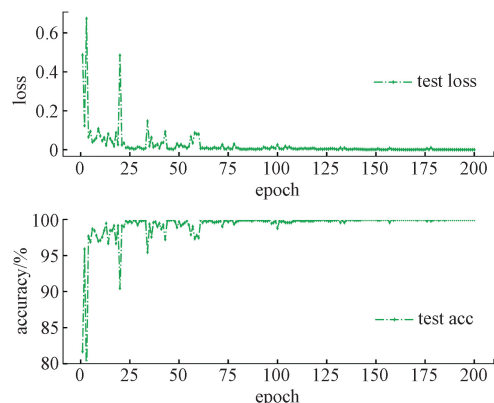
训练过程并达到稳定状态。从整体趋势来看,验证集和测试集的准确率呈现出缓慢而稳定的增长,这反映了模型在这两个数据集上的表现逐渐优化,并表明优化器的参数选择是合理的,优化过程能够有效引导模型逐步提升性能。在训练后期,准确率和损失率分别维持在较高和较低的恒定值,进一步证明了 TF-CNN 模型在训练集上的良好表现和在测试集上的稳定性,表明该模型具有较强的泛化能力。

图 6 展示了模型训练阶段、验证阶段和测试阶段的混淆矩阵。在训练阶段,模型实现了几乎完美的分类结果。在验证阶段,虽然故障编号 B3 出现少量识别错误,但错误占比极低,且其他类别的识别准确率仍超过 95%,表明模型在处理复杂故障信号时具备高效的特征提取和分类能力。在进行 15 次重复实验后,模型在测试集上的平均表现非常优异,最终准确率达到 99.8%。这一结果验证了本



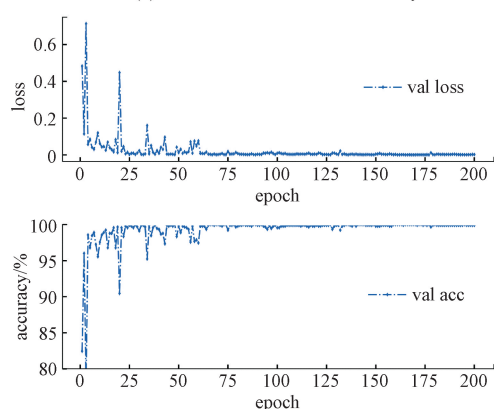
(a) 训练损失率和训练准确率

(a) The training loss rate and training accuracy



(b) 测试损失率和测试准确率

(b) The test loss rate and test accuracy



(c) 验证损失率和测试准确率

(c) The validation loss rate and test accuracy

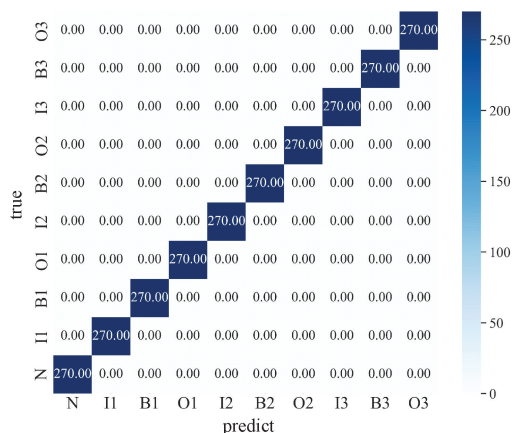
图 5 TF-CNN 模型的表现

Fig. 5 Performance of the TF-CNN model

文提出的 TF-CNN 模型在故障诊断任务中的卓越性能,以及其在复杂工业信号处理中出色的泛化性和鲁棒性。

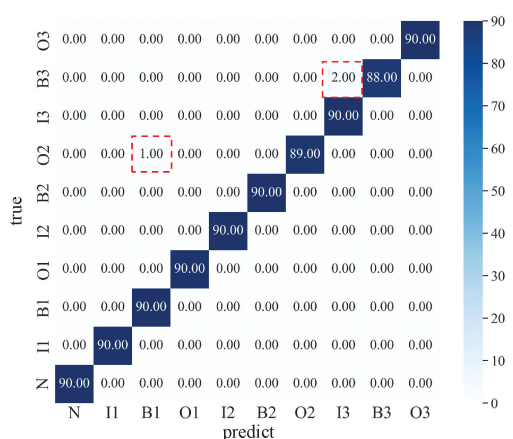
3.4 有效性分析

为了评估所提出的 TF-CNN 模型的故障检测性能,本文基于以下标准将其与几种基于深度学习的故障诊断方法进行了对比实验:1)在以往的研究过程中,许多学者将振动信号转换为二维图片,输入到 CNN 中进行故障诊断。



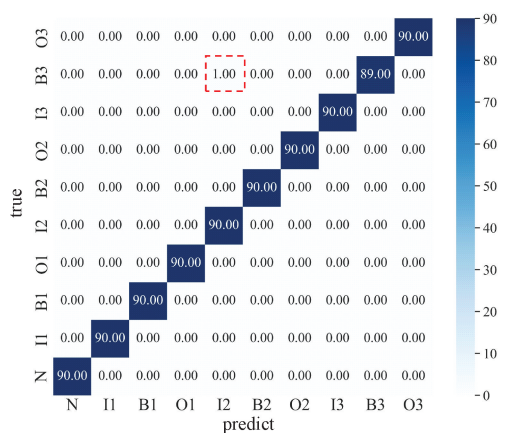
(a) 训练阶段的混淆矩阵

(a) The confusion matrix of the training phase



(b) 测试阶段的混淆矩阵

(b) The confusion matrix of the test phase



(c) 验证阶段的混淆矩阵

(c) The confusion matrix of the verification phase

图 6 训练阶段、验证阶段和测试阶段的混淆矩阵

Fig. 6 Confusion matrix of the training phase, verification phase and the test phase

因此,选择经典的二维卷积神经网络模型作为 TF-CNN 模型的对比方法。2)考虑到振动信号的时间和频率特性,选取了融合多种网络架构、能够提取多尺度信息的复合模型作为 TF-CNN 模型的对比方法。3)引入具有物理意义的

先验知识是增强模型可解释性的经典方法。因此,选择基于物理先验的可解释性增强模型作为 TF-CNN 模型的对比方法。此外,本文所选算法已在顶级会议上发表,并在故障诊断领域被广泛引用。这些算法已在多个领域和数据集中进行了广泛的测试和验证,能够为本研究提供具有可靠性和稳定性的比较基准。具体而言,基线方法包括:

1)多分支卷积神经网络(MBCNN)^[20]:利用快速谱峰度将一维信号转换为二维特征图,作为卷积神经网络的输入,以捕获故障特征。

2)CNN-LSTM^[21]:结合卷积神经网络和长短期记忆网络(LSTM),以挖掘去噪处理后的传感器数据中的隐藏特征。

3)SincNet^[22]:采用参数化的 sinc 函数作为卷积神经网络的预处理层,提升频域特征学习能力。

此外,为验证 TF-CNN 模型各关键组成部分的有效性,本文还进行了消融实验,具体如下:

1)wo/enh:评估高效数据增强方法在缓解稀疏标签问题上的作用。实验中仅采用基础数据增强方法,即添加高斯噪声对数据进行预处理,并将增强后的数据输入嵌入时

频卷积的卷积神经网络。

2)BCNN:验证卷积层中嵌入时频分析是否能够更准确地捕捉信号特征。实验中使用传统卷积神经网络的卷积核作为预处理层以提取数据特征。

上述模型均包含 4 个不同通道数的预处理层:16、32、64 和 128。在 CWRU 数据集上的对比实验与消融实验结果如图 7 所示。各模型的性能通过混淆矩阵表示,具体结果如图 8 所示。

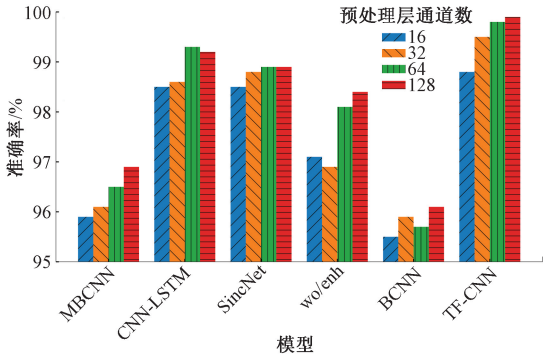


图 7 有效性分析测试准确率结果
Fig. 7 Validity analysis test accuracy results

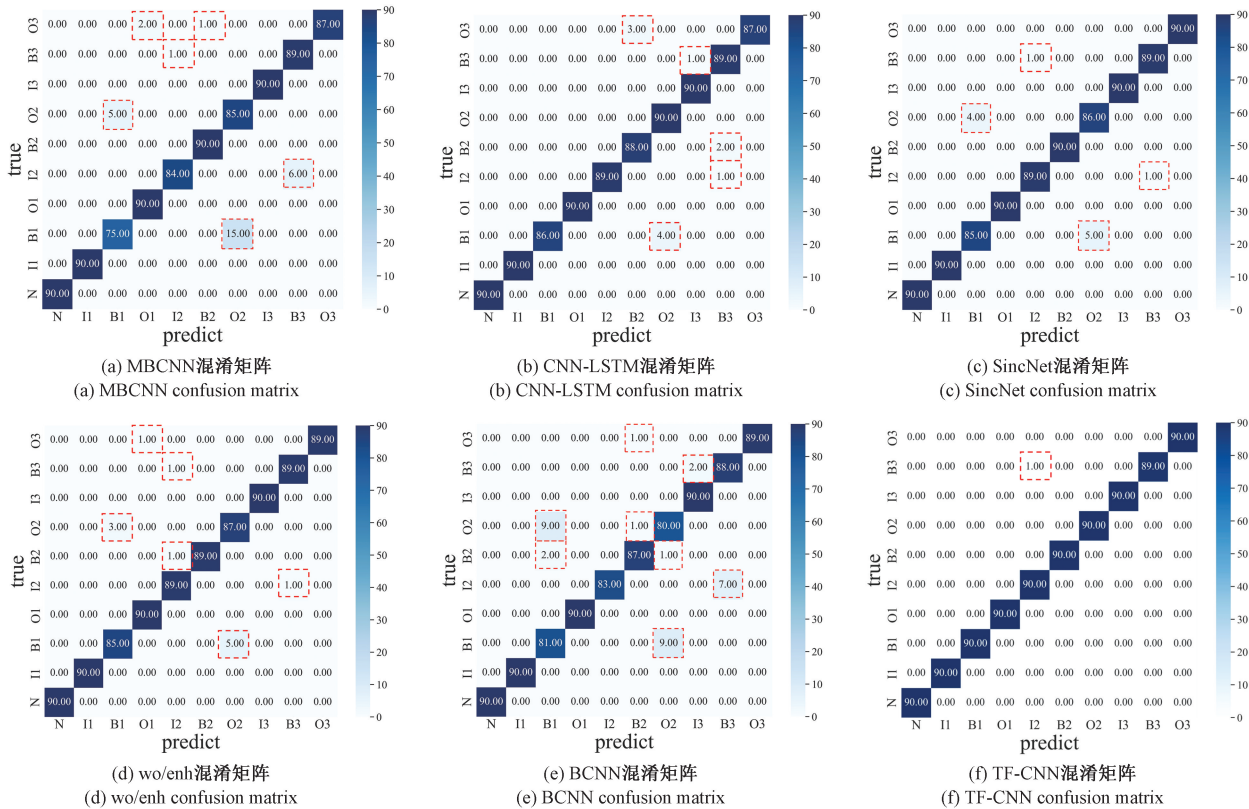


图 8 有效性分析混淆矩阵

Fig. 8 Validity analysis confusion matrix

总体来看,在对比实验中,MBCNN 的诊断结果最差,而 CNN-LSTM 和 SincNet 的准确率次之。这主要是由于

信号数据未经过有效清洗,导致在转换为二维图像时丢失了部分细节特征。此外,在信号干扰的情况下,卷积神经

网络的局部特征提取能力可能不足以捕捉全局性或长时间尺度的模式。尽管 CNN-LSTM 通过结合卷积网络和递归网络优化了结构设计,但模型的复杂度显著增加,且没有关注数据的频率信息,在数据量有限或特征表达不足的情况下,模型难以学习到数据的本质规律。SincNet 则主要关注频域特征提取,但振动信号中许多关键特征可能存在于时域。如果频域信息与时域特征之间无法有效对齐,模型的诊断能力会受到限制。相比之下,本文提出的模型通过使用多种数据增强方法提升了数据的多样性,有效缓解了有限数据和稀疏标签带来的挑战。此外,嵌入时频分析的机制使模型能够更好地学习振动信号的时间域和频率域特征,从而实现优异的诊断性能。

在消融实验中,仅采用传统数据增强方法对数据进行预处理,未结合振动信号的物理特性。这种方式无法有效生成与故障诊断相关的有意义特征,导致时频卷积层难以准确提取清晰的频谱特征。传统卷积层表现出最低的准确率,因为其卷积核仅能捕捉空间特征,对频率域信息缺乏直接感知能力,无法全面捕捉信号中的复杂特性。这进一步验证了本文提出的 TF-CNN 模型在故障诊断任务中的有效性。

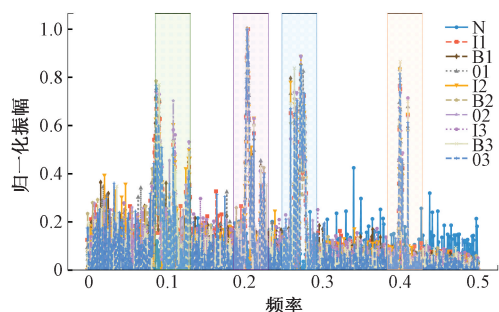
此外,可以发现通道数越多,模型诊断准确性越高,这种现象在 TF-CNN 模型中更为明显。当通道数为 16 时,TF-CNN 模型与其他模型的诊断准确性相似,但随着通道数的增加,TF-CNN 模型的诊断精度显著高于其他方法。说明具有更多通道的时频卷积层能够提取更多的时频信息,从而提高诊断精度。

TF-CNN 的时频卷积层能够将原始振动信号转化为与故障相关的时频特征,使得模型在处理复杂信号时具有卓越的诊断性能。因此,结合扩展的数据增强方法和嵌入时频分析的卷积模型,不仅能够缓解数据稀疏问题,还能准确捕捉与故障诊断相关的关键特征,从而显著提升模型性能。

3.5 可解释性分析

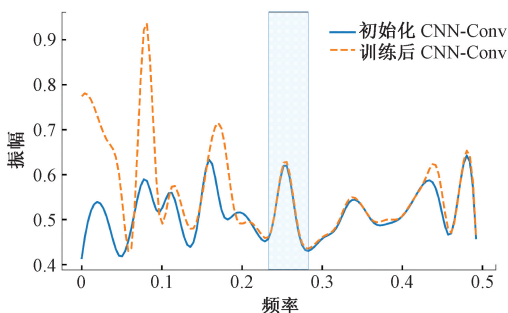
本小节通过 2.2 节中提供的频率响应分析,阐释了模型如何处理不同频率的信息,并从频域角度验证模型的故障诊断能力。图 9 展示了 CWRU 数据集的频谱信息,以及传统卷积核和时频卷积核的频率响应特性。分析表明,数据集的频率幅值主要集中在 4 个关键信息频段,这表明轴承故障通常与这些频段内的频率特征密切相关。因此,一个理想的检测模型应该能够准确关注并提取这些频段的特征,以实现较高的诊断准确性。

从图 9(b)和图 9(c)中可以看出,使用传统卷积核的网络仅在第 3 频段出现显著振幅响应,这说明其在提取关键频段信息方面存在不足,难以全面覆盖数据集中主要的频率信息。这种局限性可能导致模型在故障诊断任务中表现不佳,无法充分挖掘数据中的有用特征。相比之下,经过训练后的时频卷积核能够很好地捕捉到 1、2、3、4 四个关



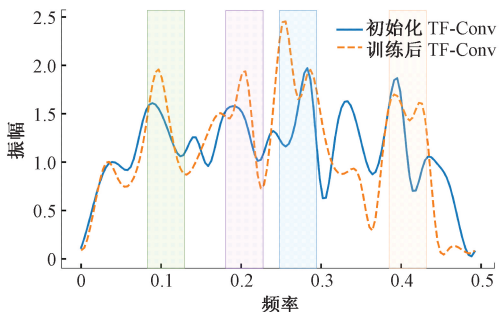
(a) CWRU数据集的频谱信息

(a) Spectrum information of the CWRU dataset



(b) 传统卷积核的频率响应结果

(b) Frequency response of traditional convolution keme



(c) 时频卷积核的频率响应结果

(c) Frequency response of the time frequency convolution kernel

图 9 频率响应分析结果

Fig. 9 Frequency response analysis results

键频段的频率特征,且在信息带之外未出现显著的幅值峰值。这表明时频卷积核能够更精准地过滤掉与故障无关的噪声信号,聚焦于与轴承故障相关的关键频率信息,展现出较强的频谱分辨能力。因此,TF-CNN 不仅在频率特征提取方面实现了显著优化,还具备最佳的物理可解释性。

通过对训练后预处理层频率响应的深入分析,本文进一步验证了 TF-CNN 在模型可解释性上的卓越性能。经过训练的时频卷积核,其频谱分布与 CWRU 数据集的频谱特征高度匹配,验证了本文的假设:嵌入时频分析的模型能够有效关注振动信号中的不同频率特征,并在训练过程中更显著地聚焦于与故障相关的频率。这种能力使得 TF-CNN 能够在复杂的工业信号诊断任务中实现优异的性能,既提升了模型的诊断准确性,又为模型输出提供了

更强的解释性支持。

4 结 论

本研究提出了一种嵌入时频分析的时频卷积神经网络(TF-CNN)模型,在机械故障诊断任务中展现出显著的性能优势与物理可解释性。通过将时频变换嵌入卷积神经网络结构,结合数据增强策略,从多尺度、多频段提取关键特征,并利用频率响应分析验证模型对信号重要频段的关注能力。这种方法不仅克服了传统方法对人工特征设计的依赖,还在特征提取的自动化与高效性方面实现了突破。相比现有文献中仅依赖时频分析或深度学习的单一方法,本文的 TF-CNN 模型有效结合了二者的优势,显著提升了诊断性能,同时增强了结果的物理可解释性,为复杂工业场景中的故障诊断提供了参考。然而,本文也存在局限,如模型的适用性尚未在更复杂的数据集上验证,实时性和跨场景泛化能力有待进一步提升。未来,研究将着眼于优化模型计算效率,扩展多场景应用,并引入多维度可解释性分析,以进一步提升模型在工业故障诊断中的实用性和可信度,为复杂系统的健康监测与预测提供可靠支持。

参考文献

[1] MA H J, LI SH M, LU J T, et al. Impulsive wavelet based probability sparse coding model for bearing fault diagnosis[J]. Measurement, 2022, 194: 110969.

[2] 刘飞,荆晓远,韩光信,等.基于残差网络和时频域特征融合的滚动轴承故障诊断方法[J]. 机床与液压,2024, 52(15):226-232.

LIU F, JING X Y, HAN G X, et al. Bearing fault diagnosis method based on residual network and time-frequency feature fusion [J]. Machine Tool & Hydraulics, 2024, 52(15): 226-232.

[3] 施亦非,黄宇峰,王锋,等.基于参数优化特征模态分解的强背景噪声下滚动轴承故障诊断[J]. 振动与冲击, 2024,43(21):107-115.

SHI Y F, HUANG Y F, WANG F, et al. Bearing fault diagnosis under strong background noise based on parameter-optimized feature modal decomposition[J]. Vibration & Shock, 2024, 43(21): 107-115.

[4] 连璞,吴磊,伍永豪.基于 CMIE 与参数优化 KELM 的旋转机械故障诊断策略[J]. 机电工程,2024, 41(1): 62-71.

LIAN P, WU L, WU Y H. Fault diagnosis strategy for rotating machinery based on CMIE and parameter-optimized KELM[J]. Electromechanical Engineering, 2024, 41(1): 62-71.

[5] 赵轲,叶敏,王瑞欣,等.基于模糊域自适应的源自由域旋转机械故障诊断方法[J]. 机械工程学报,2024,

60(18):43-52.

ZHAO K, YE M, WANG R X, et al. Source-free domain adaptation-based fault diagnosis method for rotating machinery [J]. Journal of Mechanical Engineering, 2024, 60(18): 43-52.

[6] 赵小强,安贵财.基于 MADSC 和 SIDSwinT 的滚动轴承故障诊断[J]. 电子测量与仪器学报,2024,38(11): 58-69.

ZHAO X Q, AN G C. Bearing fault diagnosis based on MADSC and SIDSwinT[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(11): 58-69.

[7] GUO ZH, DU W L, LIU ZH P, et al. Few-shot sample multi-class incremental fault diagnosis for gearbox based on convolutional-attention fusion network [J]. Expert Systems with Applications, 2024, 264:125918.

[8] LI H, JIAO J Y, LIU Z Y, et al. Trustworthy Bayesian deep learning framework for uncertainty quantification and confidence calibration: Application in machinery fault diagnosis [J]. Reliability Engineering & System Safety, 2025, 225: 110657.

[9] 张家安,邓强,马增强,等.基于 IMHSA-MSCNN-BiLSTM 的风机轴承故障诊断[J]. 电子测量技术, 2024,47(7):170-176.

ZHANG J AN, DENG Q, MA Z Q, et al. Wind turbine bearing fault diagnosis based on IMHSA-MSCNN-BiLSTM [J]. Electronic Measurement Technology, 2024, 47(7): 170-176.

[10] 冯肖亮,赵广.基于交互式多头注意力的机械故障诊断方法[J]. 计算机应用与软件,2024,41(11): 108-116,152.

FENG X L, ZHAO G. Machinery fault diagnosis method based on interactive multi-head attention[J]. Computer Applications & Software, 2024, 41(11): 108-116,152.

[11] 严如强,商佐港,王志颖,等.可解释人工智能在工业智能诊断中的挑战和机遇:先验赋能[J]. 机械工程学报,2024,60(12):1-20.

YAN R Q, SHANG Z G, WANG ZH Y, et al. Challenges and opportunities of explainable artificial intelligence in industrial intelligent diagnosis: Prior knowledge empowerment[J]. Journal of Mechanical Engineering, 2024, 60(12): 1-20.

[12] 路飞宇,佟庆彬,姜学东,等.一种基于条件度量迁移学习的机械故障诊断可解释方法[J]. 仪器仪表学报, 2024,45(7):250-262.

LU F Y, TONG Q B, JIANG X D, et al. An

- interpretable method for mechanical fault diagnosis based on conditional metric transfer learning [J]. Chinese Journal of Scientific Instrument, 2024, 45(7): 250-262.
- [13] 张龙,肖逸文,周神赐,等. 一种双阻尼小波赋能的可解释卷积神经网络在轴承故障诊断中的应用[J/OL]. 西安交通大学学报, 1-13 [2025-02-16]. <https://kns-cnki-net.webvpn.nefu.edu.cn/kcms/detail/61.1069.t.20240918.1529.002.html>.
ZHANG L, XIAO Y W, ZHOU SH C, et al. Application of an interpretable convolutional neural network empowered by dual-damping wavelet in bearing fault diagnosis [J/OL]. Journal of Xi'an Jiaotong University, 1-13 [2025-02-16]. <https://kns-cnki-net.webvpn.nefu.edu.cn/kcms/detail/61.1069.t.20240918.1529.002.html>.
- [14] WANG D, CHEN Y K, SHEN CH Q, et al. Fully interpretable neural network for locating resonance frequency bands for machine condition monitoring[J]. Mechanical Systems and Signal Processing, 2022, 168: 108673.
- [15] LI T F, ZHAO ZH B, SUN CH, et al. WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 52(4): 2302-2312.
- [16] ZHANG J Y, BAO Y F, AOKI T, et al. Refined linear chirplet transform for time-frequency analysis of non-stationary signals [J]. Mechanical Systems and Signal Processing, 2025, 223: 111785.
- [17] MARTINEZ-HERRERA A L, LEDESMA-CARRILLO L M, FERRUCHO-ALVAREZ E R, et al. Tunable hyperbolic Cohen-class kernel for cross-term diminishing in time-frequency distributions[J]. Mechanical Systems and Signal Processing, 2023, 186: 109850.
- [18] DASH B M, BOUAMAMA B O, BOUKERDJA M, et al. Bond graph-CNN based hybrid fault diagnosis with minimum labeled data [J]. Engineering Applications of Artificial Intelligence, 2024, 131: 107734.
- [19] WANG W CH, LI CH, LI AI M, et al. One-stage self-supervised momentum contrastive learning network for open-set cross-domain fault diagnosis[J]. Knowledge-Based Systems, 2023, 275: 110692.
- [20] ZHANG J Q, XU B M, WANG ZH Y, et al. An FSK-MBCNN based method for compound fault diagnosis in wind turbine gearboxes [J]. Measurement, 2021, 172: 108933.
- [21] ZHI ZH, LIU L SH, LIU D T, et al. Fault detection of the harmonic reducer based on CNN-LSTM with a novel denoising algorithm[J]. IEEE Sensors Journal, 2021, 22(3): 2572-2581.
- [22] RAVANELLI M, BENGIO Y. Interpretable convolutional filters with sincnet[J]. ArXiv preprint arXiv:1811.09725, 2018.

作者简介

张昊,硕士研究生,主要研究方向为故障诊断。

E-mail:2466161259@qq.com

王海茹,硕士研究生,主要研究方向为故障诊断。

E-mail:1962858458@qq.com

马继东(通信作者),副教授,博士,主要研究方向为系统仿真和优化。

E-mail:david@nefu.edu.cn