

DOI: 10.13382/j.jemi.B2407629

基于 MADSC 和 SIDSwinT 的滚动轴承故障诊断*

赵小强^{1,2} 安贵财¹

(1. 兰州理工大学电气工程与信息工程学院 兰州 730050;
2. 兰州理工大学国家级电气与控制工程实验教学中心 兰州 730050)

摘要:针对卷积神经网络通过局部感受野对输入信号进行特征提取,在变负荷和变噪声条件下无法有效捕获全局上下文信息导致滚动轴承故障诊断精度较低的问题,提出了一种多尺度自适应深度可分离卷积(MADSC)和空间交互双流Swin Transformer(SIDSwinT)的滚动轴承故障诊断方法。首先,利用小波变换将一维振动信号转换成二维时频图以保留完整信息;接着,构建MADSC提取局部特征信息,捕捉不同尺度下滚动轴承振动信号的特征变化;然后,设计SIDSwinT提取全局特征信息,利用提出的空间交互模块(SIM)自适应地调整特征权重;同时,通过可变形注意力对采样信息进行加权消除工况波动造成的分布差异;最后,利用双向长短期记忆网络(BiLSTM)更好地理解上下文信息,提升诊断准确性和稳定性。使用两种不同数据集验证所提方法的故障诊断性能,实验结果表明,所提方法在信噪比为-4时准确率高于93.00%,在变负荷条件下准确率高于92.00%,验证了所提方法较对比方法具有更强的抗噪性能和泛化能力。

关键词: 滚动轴承;故障诊断;空间交互;可变形注意力;深度可分离卷积;SwinT

中图分类号: TH133.33; TN911.7 **文献标识码:** A **国家标准学科分类代码:** 460.2020

Rolling bearing fault diagnosis based on MADSC and SIDSwinT

Zhao Xiaoqiang^{1,2} An Guicai¹

(1. School of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. National Experimental Teaching Center of Electrical and Control Engineering,
Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: Aiming at the problem that the convolutional neural network extracts the features from the input signal through the local receptive field, and cannot effectively capture the global context information under variable load and noise environments, resulting in the low recognition accuracy of rolling bearing fault diagnosis, a rolling bearing fault diagnosis method based on multiscale adaptive depthwise separable convolution (MADSC) and spatial interaction double-stream Swin Transformer (SIDSwinT) is proposed. Firstly, one-dimensional vibration signals are converted into two-dimensional time-frequency maps using wavelet transform to retain the complete information. Next, MADSC is constructed to extract local feature information and capture the characteristic changes of rolling bearing vibration signals at different scales. After that, SIDSwinT is designed to extract the global feature information, and the proposed spatial interaction module (SIM) is utilized to adaptively adjust the feature weights, while the sampled information is weighted by the deformable attention to eliminate the distributional differences caused by fluctuations in working conditions. Finally, bidirectional long short-term memory (BiLSTM) is utilized to better understand the contextual information and to improve the diagnostic accuracy and stability. Two different datasets are used to verify the fault diagnosis performance of the proposed method, and the experimental results show that the accuracy of the proposed method is higher than 93.00% when the signal-to-noise ratio is -4, and the accuracy is higher than 92.00% under the condition of variable load, which verifies that the proposed method has a stronger anti-noise performance and generalization ability than the comparison methods.

Keywords: rolling bearings; fault diagnosis; spatial interaction; deformable attention; depth separable convolution; SwinT

收稿日期: 2024-06-27 Received Date: 2024-06-27

* 基金项目: 国家自然科学基金(62263021)、甘肃省教育厅产业支撑项目(2021CYZC-02)资助

0 引言

滚动轴承是一种精密部件,广泛用于各种旋转机械。由于滚动轴承大多处于非平稳、强噪声环境中,容易出现点蚀、断裂等各种故障,影响整机工作的稳定性和安全性。据统计,30%~40%的感应电机故障是由轴承故障引起的^[1],因此,滚动轴承故障诊断的研究就显得尤为重要。

滚动轴承故障诊断中信号处理方法主要有时域分析、频域分析和时频域分析,时域分析主要关注信号在时间域的波形特征,频域分析则是通过分析信号的频谱特征,如频率成分、频率分布等来诊断轴承的故障类型,而时频分析结合了时域和频域的优势,能够同时提供信号在时间和频率上的信息,适用于具有时变特性的信号,因此,时频分析被广泛应用于滚动轴承故障诊断。常用的时频分析方法包括短时傅里叶变换^[2]、小波变换^[3]、经验模态分解^[4]和奇异值分解^[5]等。一些传统的智能分类器算法,例如支持向量机^[6]、贝叶斯分类器^[7]和随机森林方法^[8]被用于识别故障类型和严重程度,但这些方法十分依赖信号处理和专家经验,很难自主挖掘海量数据中的有效信息,对于复杂的任务和大规模数据集,手动选择特征向量是非常困难且耗时的。

近年来,深度学习方法如深度置信网络^[9]、卷积神经网络^[10]、自编码器^[11]等被应用于故障诊断领域。其中卷积神经网络作为一种典型的深度学习方法,通过使用卷积层和池化层来自动提取图像或其他类型数据中的特征,解决了传统人工选择特征向量的问题。例如,李恒等^[12]提出了基于短时傅里叶变换和卷积神经网络的故障诊断方法,实现了端到端的故障模式识别。曲建玲等^[13]提出了自适应一维卷积神经网络(ACNN-FD)故障诊断算法,利用深度网络结构实现对原始振动信号特征的自适应层级化提取,取得较好的诊断准确率。同时,作为深度学习的后起之秀,Transformer^[14]凭借全局特征提取能力在许多任务中取得了非常好的效果,然而,Transformer 网络通常参数量非常大,尽管有学者在轻量级方面已经取得了一些突破,但仍有很大的进步空间。为了解决 Transformer 参数量大的问题,Li 等^[15]提出 SwinT 方法,该方法通过引入滑动窗口机制在处理大尺寸图像时能够实现更好的计算效率,同时通过局部注意力机制保持了 Transformer 模型对图像全局信息的关注。

虽然上述方法取得了不错的诊断效果,但是单一卷积神经网络和 Transformer 在变负荷和变噪声条件下无法有效捕获全局上下文信息而导致滚动轴承故障诊断精度不佳。因此,提出了一种多尺度自适应深度可分离卷积(multiscale adaptive depthwise separable convolution,

MADSC)和空间交互双流 Swin Transformer (spatial interaction double-stream swin transformer, SIDSwinT) 滚动轴承故障诊断方法,该方法首先通过小波变换将一维振动信号转换成二维时频图,充分利用信号的时频特性,并通过重采样技术提高时频分辨率以及减小频谱泄漏来增强方法的泛化能力。接着,构建 MADSC 模块灵活地提取不同尺度下的局部特征信息,有效捕捉滚动轴承振动信号在复杂工作条件下的细微变化,确保方法能够适应各种工况下的故障特征提取。在局部特征提取的基础上进一步设计 SIDSwinT 架构,它通过双流设计同时处理时频域的信息,利用空间交互模块(spatial interaction module, SIM)促进信息融合,从而实现对全局特征的高效捕捉。为克服工况波动带来的分布差异,本研究还引入了可变形注意力机制,通过对采样信息进行加权处理,模型能够自动调整关注关键特征,有效抑制噪声干扰。最后,利用双向长短时记忆网络(bidirectional long short-term memory, BiLSTM)对提取的特征进行整合,进一步提升故障诊断的精度和稳定性,确保在复杂工况下仍能维持高水平的诊断能力。实验结果表明,所提方法解决了强噪声和变负荷条件下滚动轴承故障诊断准确率较低的问题。

1 基础理论

1.1 小波变换

时频分析是现代信号处理领域的一项重要技术,滚动轴承的振动信号通常具有非平稳性,而小波变换以其独特的多分辨率特性非常适合非平稳信号分析^[16],小波函数的表达式如式(1)所示。

$$WT(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \times \varphi\left(\frac{t - \tau}{a}\right) dt \quad (1)$$

式中: a 是比例因子,控制函数的宽度, τ 是平移因子,控制函数在轴上的位置。

在小波变换中,小波基的选择非常重要,常见的小波基包括 Haar 小波、Daubechies 小波、Coiflets 小波、Morlet 小波等,每种小波基都有其独特的特性和优势。Daubechies 小波缺乏对瞬态特征的敏感性,适合于平滑信号的分析;Haar 小波由于其简单的阶跃函数形式,虽然能够快速检测信号中的突变点,但其时频分辨率较低,不适合捕捉复杂的轴承故障特征;Coiflets 小波设计时考虑了对称性和平滑性,适合于处理具有平滑过渡的信号;Morlet 小波具有良好的时频局部化特性。轴承在发生故障时会产生周期性的冲击脉冲信号,这些冲击信号具有短暂、高能量的特点,而 Morlet 小波通过其参数的调整,可以很好地匹配这些瞬态冲击信号的特性,从而实现对轴承故障特征的有效提取。因此,本研究选择的小波基

为 Morlet 小波。

1.2 深度可分离卷积神经网络

DSC 是一种轻量级的卷积神经网络结构,它通过将标准卷积操作分解为深度卷积和逐点卷积两个步骤来减少参数数量和计算量,其结构示意图如图 1 所示,步骤如下:

1) 深度卷积:深度卷积是在每个输入通道上分别应用卷积核,生成与输入通道数量相同的特征图。这个操作可以捕获输入数据的空间特征,但不进行通道之间的信息交互。

2) 逐点卷积:逐点卷积是在深度卷积的输出上应用 1×1 的卷积核,用于将不同通道之间的特征进行线性组合。这个操作可以实现通道之间的信息交互和特征融合。

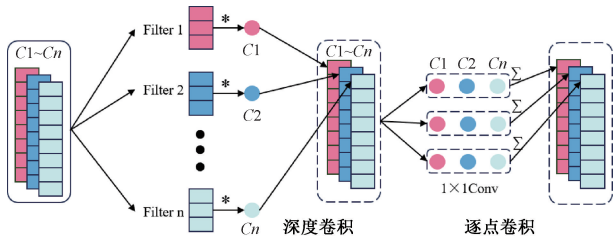


图 1 深度可分离卷积

Fig. 1 Deep separable convolution

1.3 Transformer

Transformer 是一种基于自注意力机制的神经网络架构,最初用于自然语言处理任务。随着 Transformer 模型在自然语言处理领域的成功应用,研究人员开始将 Transformer 应用于图像分类任务中,这些模型通常被称为 ViT,由于图像通常具有较高的分辨率和较大的尺寸,这导致了输入序列的长度非常大,从而增加了计算和内存的需求。为解决 ViT 在处理图像时需要大量的计算和内存资源的问题,提出了一种 SwinT 网络作为可行解决方案,结构如图 2 所示。

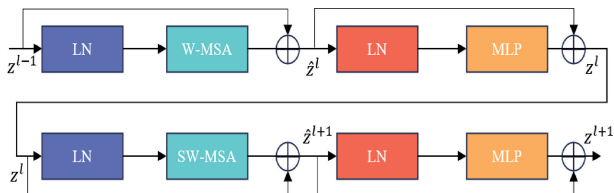


图 2 SwinT 结构

Fig. 2 SwinT structure

由图 2 可知, SwinT 通过引入分层的窗口注意力机制和跨窗口的位置编码,有效地解决了传统 Transformer 在处理大尺寸图像时的计算量大和内存需求过高问题。

2 滚动轴承故障诊断方法

在滚动轴承故障诊断领域,由于振动信号中的特征受到多重耦合作用,常呈现出复杂的多尺度特性。基于 CNN 的特征学习方法消除了传统故障诊断技术中人为特征提取的需要,允许端到端的信息处理。然而,使用单一模型处理信号可能不会产生理想的结果。因此,提出了一种基于 MADSC 和 SIDSwinT 滚动轴承故障诊断方法,结构如图 3 所示。

2.1 空间交互模块

注意力机制模拟了人类注意力的行为,它通过计算输入序列中每个位置的重要性权重,然后根据这些权重来加权计算输出结果。受 EMA^[17] 的启发,本文设计了一个空间交互模块 (spatial interaction module, SIM), 如图 4 所示,该模块采用 3 个并行路线来提取分组特征图的注意力权重。首先,在第一分支中采用平均池化操作对通道进行编码,在卷积之后加入 BN 层和 Relu 非线性激活函数,以此减少协变量偏移的影响和学习复杂的关系,并且在第 2 分支中仅堆叠单个 3×3 内核以捕获多尺度特征表示;此外,为了在第一分支中的两个并行路径之间实现不同的跨通道交互特征,通过乘法聚合每个组内的两个通道注意力图,这样, SIM 得到加权通道和空间信息;然后,利用最大池化操作对第一分支输出中的全局空间信息进行编码,为了高效计算,在最大池化处的输出采用非线性函数 Softmax,以拟合线性变换;最后,聚合输出特征图生成的空间注意力的权重值,它捕获像素级成对关系并突出显示所有像素的全局上下文。

2.2 空间交互双流 SwinT 网络结构

为了关注时域信息和频域信息,设计了 SIDSwinT 网络,其结构如图 5 所示。SIDSwinT 采用了分层的注意力机制,将输入图像分解为不同的块,在每个分层级别对这些块进行自注意力计算,使得模型能够更好地捕获全局和局部信息。同时,在 SIDSwinT 中引入了跨层连接,允许不同层之间直接进行信息交换和传递,这有助于提高模型的表示能力,在融合层通过 SIM 模块对通道和空间信息进行编码,自适应地分配特征权重以便获取更全面的故障特征。此外,虽然传统的注意力机制能够捕获全局上下文,但其计算复杂度随着输入序列长度的增加而迅速增长。相比之下,可变形注意力机制通过局部化采样和有限的偏移量学习,可以显著降低计算成本,同时保持良好的性能,因此,将 MSA 通过可变形注意力机制替换。最终, SIDSwinT 被计算为:

$$\hat{z}^l = SWMSA(LN(z^{l-1})) + z^{l-1} \quad (2)$$

$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l \quad (3)$$

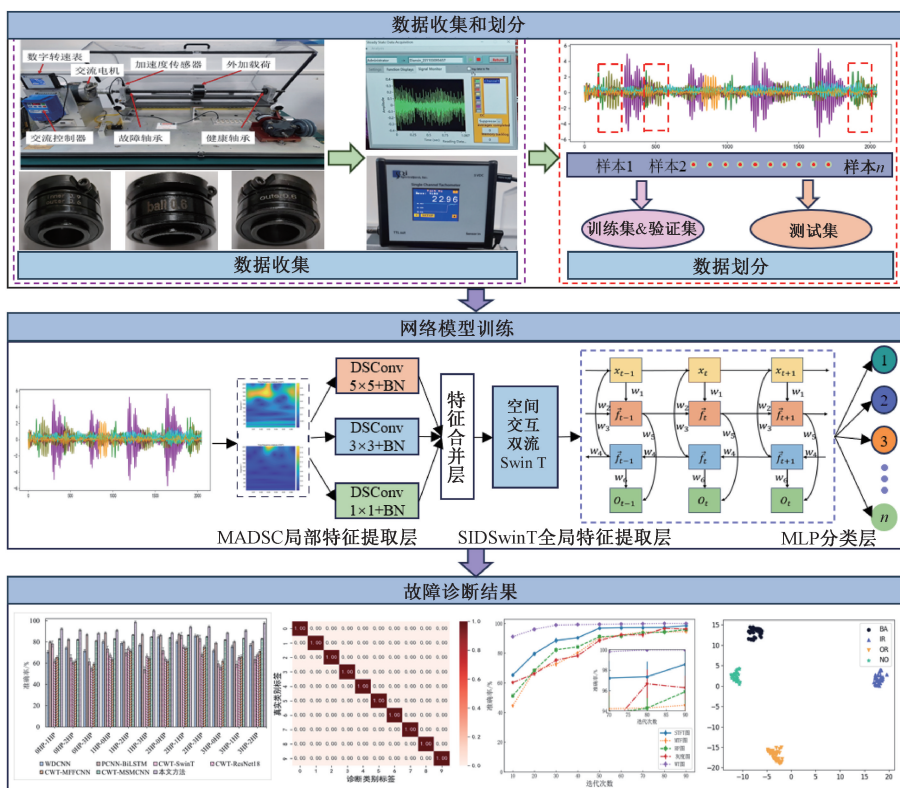


图 3 故障诊断方法

Fig. 3 Fault diagnosis method

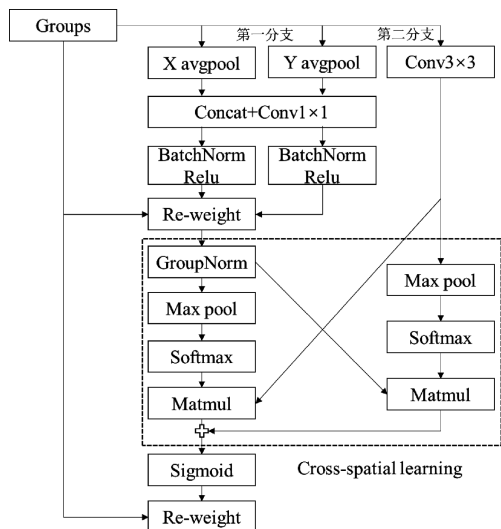


图 4 空间交互模块

Fig. 4 Spatial interaction module

2.3 故障诊断流程

所提出的滚动轴承故障诊断方法流程图如图 6 所示。首先,将生成的小波时频图数据集按 7 : 1.5 : 1.5 的比例划分成训练集、验证集和测试集;然后,进行参数初始化以提高模型的性能和训练效率,根据训练结果对参数进行优化,判断结果是否达到收敛,若结果收敛,则保存最佳模型;最后,在独立的测试集上评估模型对轴承故障类型的诊断准确性,以验证其泛化能力和实用性。

3 实验结果与分析

3.1 实验环境配置

本研究所有实验都在具有以下配置的环境运行: Torch2.0.0 和 Cuda117, 以及通过 Pytorch 框架运行的 Windows 10 操作系统,显卡为 GeForce RTX 4060 Ti,处理器为 13th Gen Intel(R) Core(TM) i5-13400F 2.50 GHz。优化器采用基于随机梯度下降算法的自适应 SGD 优化器,并使用反向传播来更新模型的参数,损失函数使用交叉熵损失函数,学习率设为 0.0001,批次大小设为 16。

$$\hat{z}^{l+1} = DA(LN(z^l)) + z^l \quad (4)$$

$$\hat{z}^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (5)$$

式中: $\hat{z}^l$ 和 z^l 分别表示块 L 的注意力模块和 MLP 模块的输出特征。

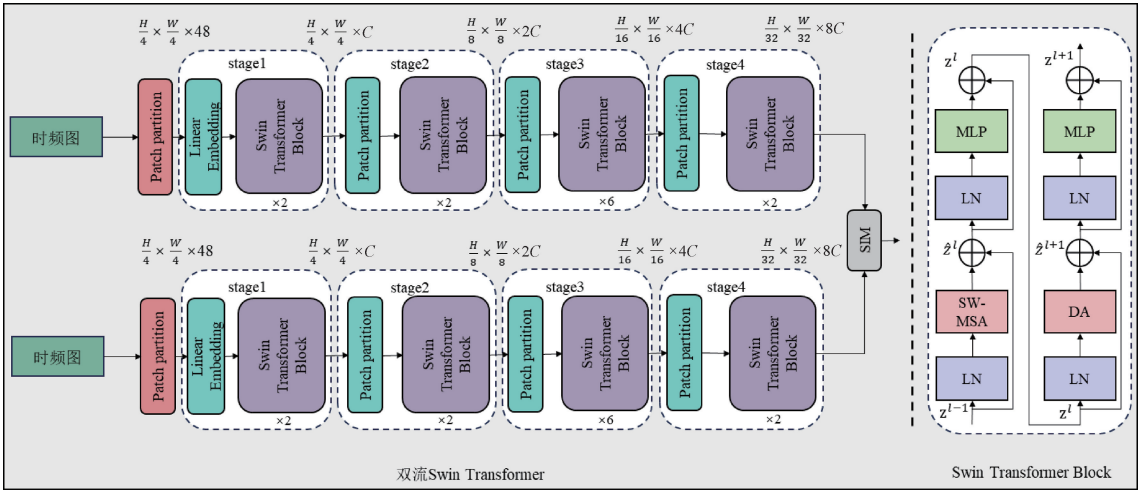


图 5 空间交互双流 SwinT
Fig. 5 Spatial interaction dual-stream SwinT

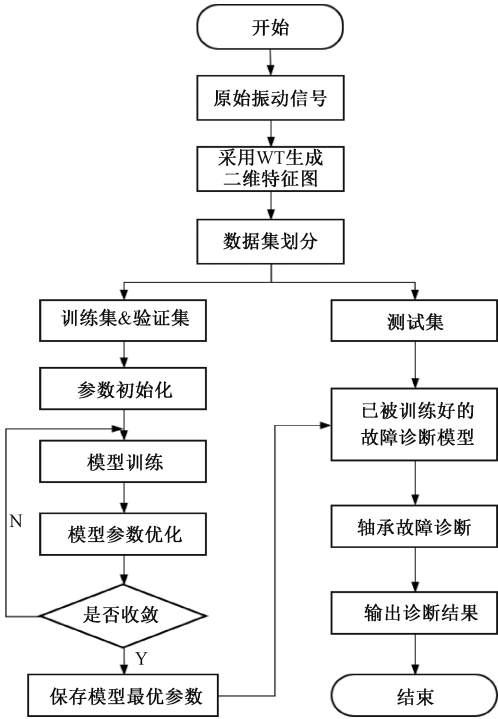


图 6 故障诊断流程
Fig. 6 Process of fault diagnosis

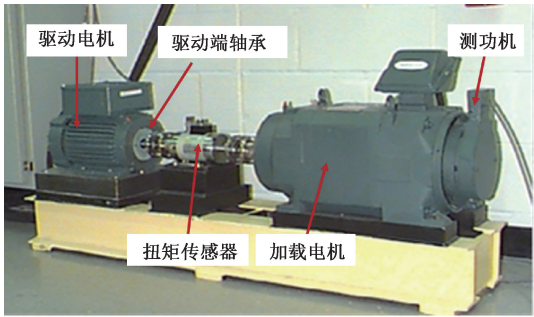


图 7 CWRU 实验台
Fig. 7 CWRU experimental bench

轴承为 SKF6203, 采样频率为 12 kHz。数据集由正常信号和故障信号组成, 故障信号采用电火花加工, 故障位置包括内圈、滚动体以及外圈^[19]。根据轴承不同故障类型, 将 CWRU 数据划分为 10 种类别, 如表 1 所示, 采样点数设置为 1 024, 分辨率为 224×224 像素。

表 1 CWRU 轴承数据划分

Table 1 Division of CWRU bearing data		
轴承状态	故障尺寸/mm	标签
正常	\	0
	0.177 8	1
	0.355 6	2
内圈故障	0.533 4	3
	0.177 8	4
	0.355 6	5
外圈故障	0.533 4	6
	0.177 8	7
	0.355 6	8
滚动体故障	0.533 4	9

3. 2 凯斯西储大学轴承数据集

1) 数据描述和划分

本次实验使用 CWRU 轴承数据集^[18]进行训练和测试, CWRU 实验台如图 7 所示, 由驱动电机、扭矩传感器、测功机和加载电机等部分组成。

该实验台采用加速度计收集对应转速负载条件分别为 3 HP、2 HP、1 HP 和 0 HP 下的轴承数据, 驱动端轴承型号为 SKF6205, 采样频率为 12 kHz 和 48 kHz。风扇端

2) 对比方法验证

为了验证所提方法的优越性, 将该方法应用于

CWRU 轴承数据集,并与 WDCNN^[20]、PCNN-BiLSTM^[21]、CWT-SwinT^[22]、CWT-ResNet18^[23]、CWT-MFFCNN^[24]以及 CWT-MSMCNN^[25]这几种先进的方法进行了比较。其中, WDCNN 和 PCNN-BiLSTM 中输入信号为原始时序信号。所有方法均进行 5 次独立实验取平均值,诊断准确率如图 8 所示。

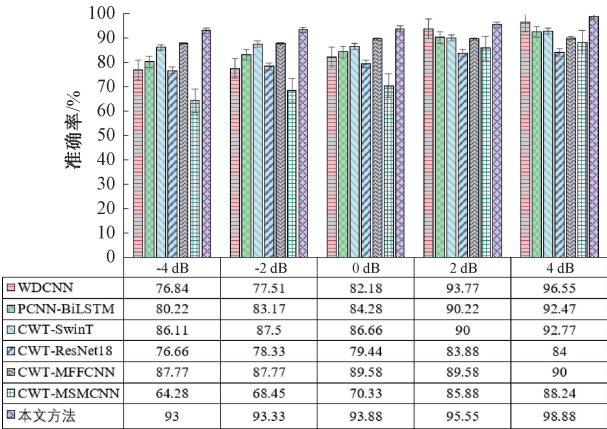


图 8 诊断准确率
Fig. 8 Diagnostic accuracy

由图 8 可以看出,在 SNR 值取-4 时,所提方法诊断准确率达到 95.33%,较 WDCNN 高 18.49 个百分点,较 PCNN-BiLSTM 高 15.11 个百分点,较 CWT-SwinT 高 12.47 个百分点,较 CWT-ResNet18 高 8.09 个百分点,较 CWT-MFFCNN 高 12.67 个百分点,较 CWT-MSMCNN 高 9.22 个百分点;当 SNR 值取 4 时,WDCNN 达到 96.55% 的诊断准确率,但是比所提方法低 1.55 个百分点。这表明所提方法较对比方法具有更好的稳定性和更强的抗噪性能。此外,在 SNR 为-4 dB 时,WDCNN 和 PCNN-BiLSTM 的诊断准确率要低于其他方法,这表明输入 CWT 信号比输入原始信号具有更强的抗噪性能。

为了验证所提方法对正常数据和故障数据的分离效果,使用混淆矩阵^[26]显示了在样本量为 100 且 SNR 分别为-4、-2、0、2 和 4 的详细诊断结果,如图 9 所示,在混淆矩阵中,主对角线上的每个元素表示该类别被正确分类的样本数占该类别所有样本数的比例。

由图 9 可知,SNR 为-4 时,所提方法的分类准确率达到 96.10%,SNR 为 4 时,分类准确率达到 100%,这表明所提方法在强噪声干扰下也有较高的分类准确率。

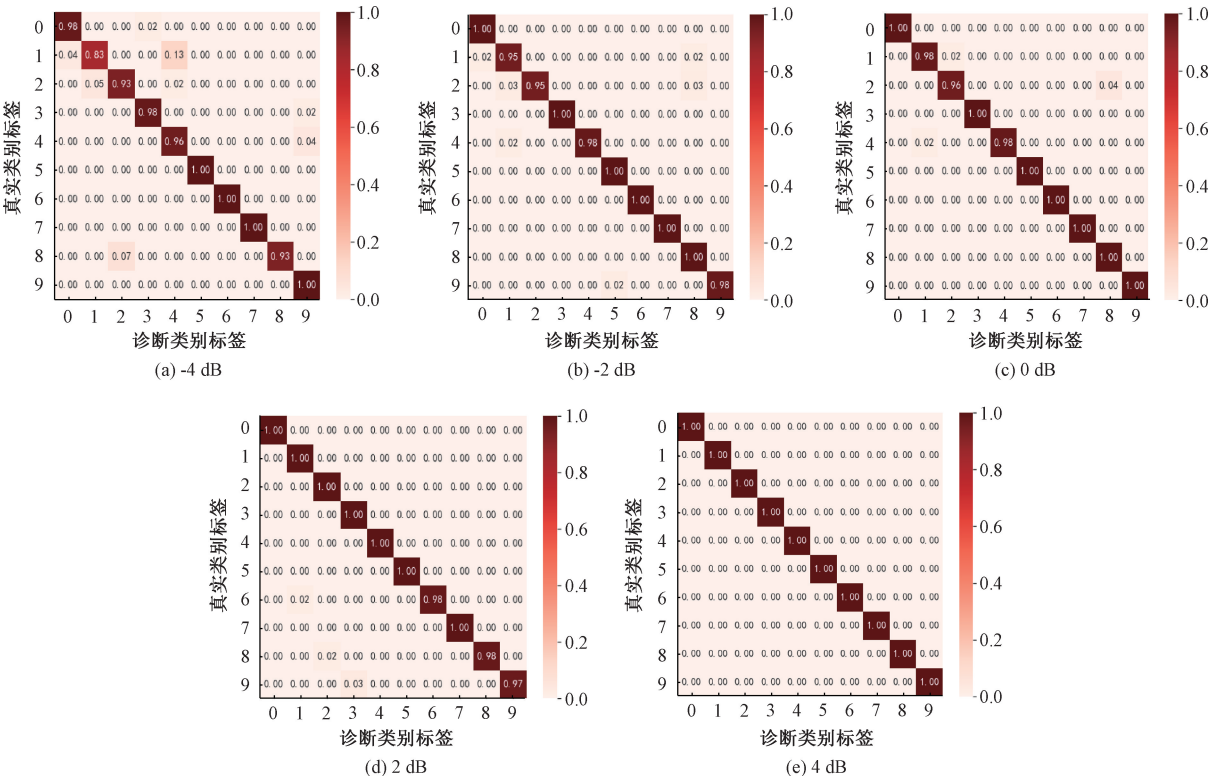


图 9 故障分类混淆矩阵
Fig. 9 Fault classification confusion matrix

3) 不同样本数量在噪声环境下对准确率的影响
滚动轴承在实际运行中,大部分时间都会处于健康

状态,这导致了可采集的故障数据有限,因此,数据缺乏情况下的故障诊断是目前面临的重要问题。为了观察模

型在少量样本下的性能,本研究使用样本量为 60、100、300 的样本来训练和验证模型的性能。同时,在工程实际环境中,振动信号往往无法避免噪声干扰,利用无噪声的训练集来评估模型的抗噪声能力,然后将噪声添加到测试集进行验证,实验结果如表 2 所示。噪声的信号强度用信噪比 (SNR)^[27] 来衡量,表示为:

$$SNR = 10 \lg \frac{P_{signal}}{P_{noise}} \tag{6}$$

式中: P_{signal} 表示信号功率, P_{noise} 表示噪声功率。

表 2 噪声环境下故障诊断准确率

Table 2 Fault diagnosis accuracy in noisy environment		SNR/dB					
类样本数	负载/HP	-4	-2	0	2	4	
300	0	94.22	94.35	95.11	95.77	96.44	
	1	95.33	95.55	96.44	97.11	98.00	
	2	91.33	97.77	98.00	99.33	99.55	
	3	92.66	95.55	98.22	97.33	98.88	
100	0	82.00	86.00	88.00	92.00	92.76	
	1	88.66	93.33	93.64	96.00	96.66	
	2	91.33	93.33	93.82	94.66	96.00	
	3	83.42	88.66	92.81	95.33	97.33	
60	0	76.66	82.22	92.96	94.00	96.33	
	1	71.58	80.00	85.55	88.88	91.24	
	2	76.44	78.88	91.11	93.24	95.55	
	3	73.85	77.86	83.89	83.83	88.88	

由表 2 可知,在样本数量为 60 条件下、负载 0 HP 且 SNR 为-4 时,所提方法准确率达到 76.66%,而当 SNR 为 4 时,准确率达到 96.33%;在样本量 300 条件下、负载

为 0 HP 且 SNR 为-4 时,准确率为 94.22%,当 SNR 为 4 时,准确率为增加至 96.44%。所提方法在样本量不变的条件下,随着 SNR 逐渐增加,诊断准确率也随之增加,当样本量从 60 增加到 300 时诊断准确率从 76.66%增加至 94.22%。因此,即使是少样本高噪声的条件下,依然能保持较高的诊断准确率。

4) 变负荷故障诊断结果

在实际工作环境中,轴承工作的条件会经常变化,因此,为了验证所提方法在不同负荷下的泛化能力,采用基于微调的迁移学习方法对不同负荷下的滚动轴承进行故障诊断,验证所提出的方法在不同负荷下的适应性。故障诊断结果如图 10 所示,其中,“0 HP~1 HP”表示模型在 0 HP 负荷下训练,1 HP 负荷下测试,依此类推。

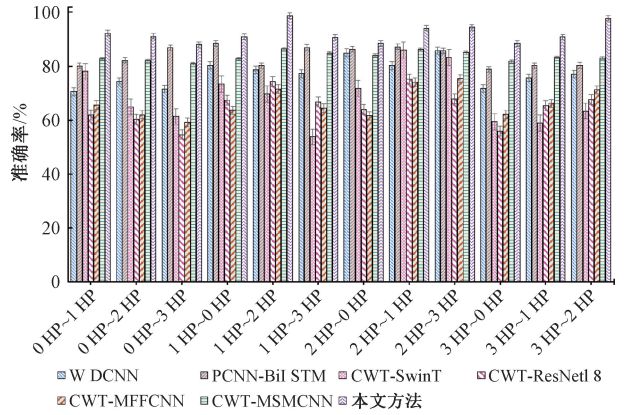


图 10 变负荷故障诊断结果

Fig. 10 Diagnostic results of variable load faults

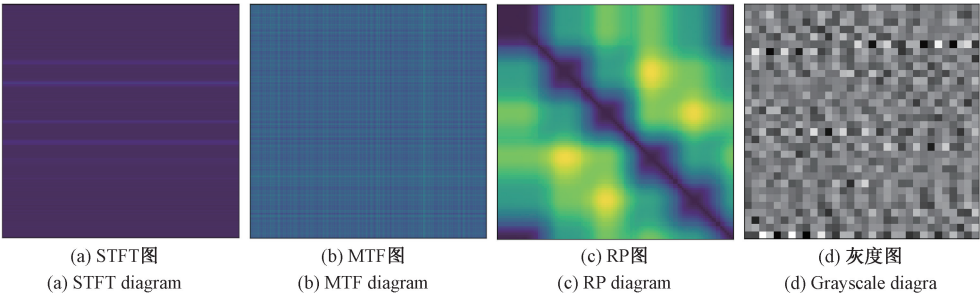


图 11 不同方式编码图

Fig. 11 Coding diagrams of different modalities

5) 与其他图像编码方式进行对比

为进一步验证 WT 图在故障诊断领域的优越性,本节选用以下图像输入方法与所提方法进行对比:(1) 利用短时傅里叶变换 (short-time fourier transform, STFT) 将振动信号转换为时频图作为输入;(2) 将原始振动信号转换为马尔科夫迁移场 (markov transition field, MTF) 图作为输入;(3) 以递归图 (recurrence plot, RP) 作为输入;(4) 将振动信号通过转换后的灰度图像作为输入。轴承振动信号采用 CWRU 数据集在 0 负载条件下的振动信

号,按照 7 : 1.5 : 1.5 的比例将数据划分为训练集、验证集和测试集,图像大小设为 224×224。将 4 种图像编码方式各取一个样本进行展示,如图 11 所示。

将上述 4 种图像和 WT 时频域图像分别作为本文网络模型的输入,在训练阶段,超参数选择跟上述实验保持一致,训练结果如图 12 所示。

从图 12 可以观察到,随着迭代次数的增长,模型很快收敛,在迭代 30 次之后,所提方法故障识别准确率曲线均趋于平稳,故障识别准确率可达 99.99%;STFT 图作

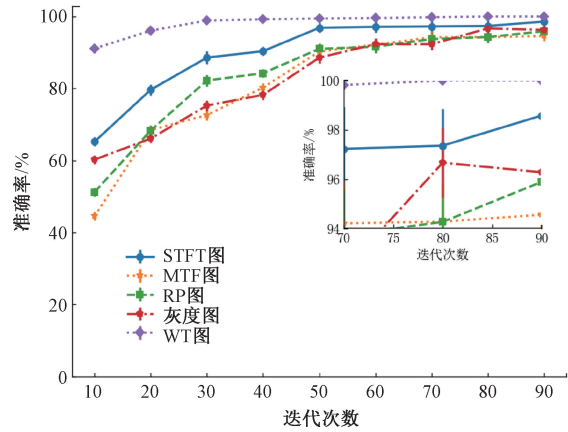


图 12 不同输入图像对诊断准确率的影响

Fig. 12 Effect of different input images on diagnostic accuracy

为网络输入时识别准确率分别为 98.57%，所提方法故障诊断准确率有 1.42% 的提高，STFT 准确率较低的原因是 STFT 在时间和频率上的分辨率是固定的，无法灵活地适应信号的，而 WT 具有多分辨率分析、时间和频率局部化、去噪能力以及非平稳信号分析等优势。RP 图、MTF 图以及灰度图作为网络输入时的识别准确率分别为 95.88%、94.55% 和 96.27%，均低于本文所提方法。相比较于 RP 图、MTF 图以及灰度图作为输入时故障诊断准确率，以 STFT 图作为输入，故障识别准确率有着明显的提高，这是因为 STFT 能够将信号在时间和频率上进行分析，生成的时频图能够融合信号的时域和频域信息，提供更全面的特征描述。

6) 消融实验

为了验证所提方法的计算效率和资源消耗，记录了训练样本量为 100 时，模型迭代 100 次所需的平均训练时间与模型参数量，结果如表 3 所示。

表 3 消融实验结果

方法	准确率/%	训练时间/s	参数量/ 10^6
WDCNN	85.84	47	0.57
PCNN-BiLSTM	87.21	84	1.14
CWT-SwinT	94.33	671	8.18
CWT-ResNet18	88.37	203	2.27
CWT-MFFCNN	92.18	330	3.52
CWT-MSMCNN	93.61	278	3.57
所提方法	99.27	221	2.41

由表 3 可知，WDCNN 作为一种典型的网络结构，其训练时间和参数量均为最低，但准确率仅为 85.84%；PCNN-BiLSTM 模型由于网络深度的增加，训练时间和参数量都有所上升；CWT-ResNet18 网络由于从时频维度上提取特征，导致输入数据维度增加，因此其训练时间和参数量相较于 WDCNN 和 PCNN-BiLSTM 都有所增加；

CWT-MFFCNN 和 CWT-MSMCNN 网络的训练时间和参数量显著增加，主要是因为常规卷积在生成每个新的特征图时都需要一组新的权重参数，从而导致了参数量的增长。所提方法通过结合 MADSC 和 SIDSwinT，在保证高准确率的同时有效避免了通道冗余计算，优化了计算资源的使用。

3.3 MFS 轴承数据集

1) 数据描述和划分

为进一步验证所提方法在不同轴承数据上的泛化性，采用 MFS 轴承数据集进行实验，MFS 实验台主要由加速度传感器、交流电机、交流控制器和信号采集器等部分组成，如图 13 所示。

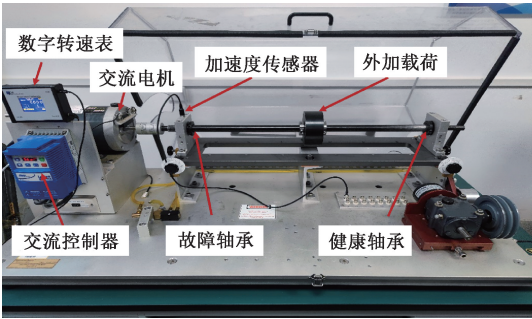


图 13 MFS 实验台

Fig. 13 MFS experimental bench

轴承类型为深沟球轴承，型号为 ER-16K。该轴承故障均是由激光刻蚀技术加工所得。在正常状态下模拟了 4 个位置故障，对转速分别为 1 130 r/min、1 251 r/min、1 348 r/min 和 1 449 r/min 的信号进行收集，采样频率为 15.6 kHz，样本以训练集、测试集和验证集按 7 : 1.5 : 1.5 的比例进行划分，如表 4 所示。

表 4 MFS 轴承数据描述

轴承状态	故障直径/mm	采样频率/Hz	标签
滚球故障	0.6	15 600	BA
内圈故障	2.4	15 600	IR
外圈故障	0.6	15 600	OR
复合故障	in0.6&ou0.9	15 600	NO

2) 方法对比

为了全面比较所提方法抗噪性能，在转速 1 130 r/min 时将 SNR 值为 -4、-2、0、2 和 4 的高斯白噪声添加到测试集中，取 5 次重复实验的平均值作为实验结果，每种实验方法的诊断结果如图 14 所示。

由图 14 可知，随着 SNR 值降低，对比方法的分类精度显著下降，特别是 ResNet18 方法受噪声影响最大。在 SNR 为 -4 时，所提方法准确率达到 93.00%，这表明所提方法具有良好的抗干扰能力。

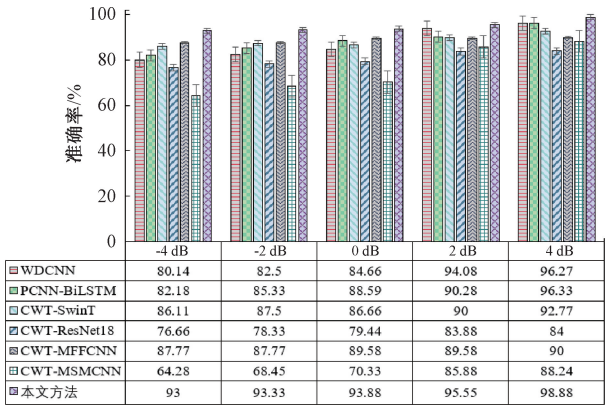


图 14 诊断准确率
Fig. 14 Diagnostic accuracy

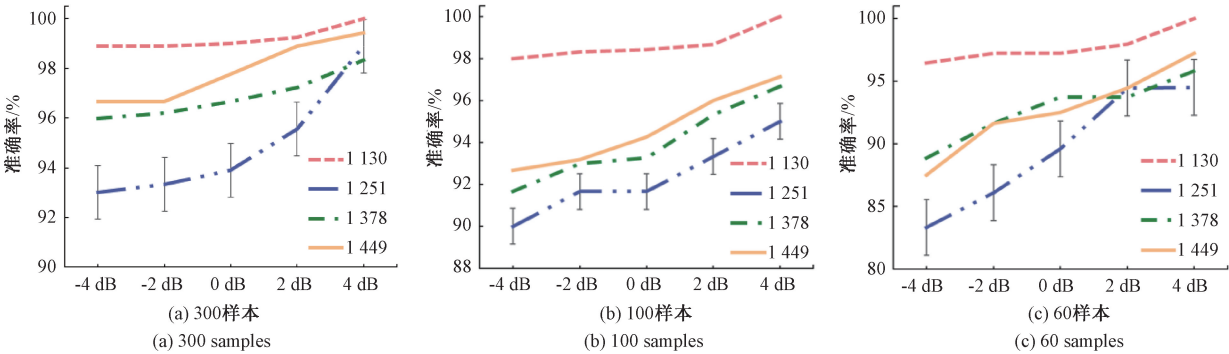


图 15 不同转速下故障诊断准确率
Fig. 15 Accuracy of fault diagnosis at different speeds

由图 15 可知,在样本数为 300 时,高转速和低转速对准确率的影响并不明显,SNR 为 -4 时准确率均在 93% 以上;但是样本数为 60 的条件下,在转速为 1 130 r/min 时,准确率达到 96.44%,在转速为 1 251 r/min、1 378 r/min 和 1 449 r/min 时的准确率分别为 83.33%、88.88% 和 87.50%,在高转速下准确率降低的原因是某些故障可能只在低转速下才会出现,而在高转速下不明显或完全消失,因此,在高转速下进行故障诊断时,可能会错过一些故障的信息。此外,在样本量为 60 且 SNR 为 4 时诊断准确率均高于 94.00%。以上实验数据都验证了本文方法在小样本条件下也具有较强的抗噪性能。

4) 可视化

为了观察这些方法的特征提取能力,在测试集上使用 t-SNE 可视化最后分类层的输出,图 16 显示了转速 1 449 r/min 下 SNR 值为 0 时的结果。

从图 16 可以看出 WDCNN、CWT-SwinT 和 CWT-ResNet18 方法在特征空间中存在明显的重叠,导致小样本故障诊断的准确性较差。相比之下,PCNN-BiLSTM、CWT-MFFCNN 和 CWT-MSMCNN 方法提高了学习特征的

为进一步验证以 WT 图像编码方式的优越性,使用跟 3.2 节中 5) 的 4 种图像编码方式与 WT 进行对照。

表 5 不同输入方式的平均故障识别准确率

Table 5 Average fault recognition accuracy for different input methods					
输入图像方式	STFT	MTF	RP	灰度图	WT
平均诊断准确率/%	98.00	94.86	96.52	97.18	99.58

实验结果如表 5 所示,可以看到以 WT 的图像编码方式仍然可以取得最高的故障识别准确率,进一步验证了 WT 的优越性。

3) 不同样本数量在噪声环境下对准确率的影响

为了验证所提方法在 MFS 数据集上的泛化性能,在模型迭代 100 次之后的实验结果如图 15 所示。

可分离性,成功分离了大多数轴承状态。然而,在分析复合故障时,仍然存在部分重叠,与上述方法相比,所提方法在同一簇内的样本之间的距离相对较近,而簇之间的距离较远。因此,即使在强噪声环境下,所提方法也能实现高精度的滚动轴承故障分类且优于其他几种对比方法。

4 结 论

为了解决卷积神经网络在变负荷和变噪声环境下无法有效捕获全局上下文信息导致滚动轴承故障诊断准确率低的问题,提出了一种新的滚动轴承故障诊断方法。

该方法首先利用小波变换将一维振动信号转换成二维时频图以保留原始信号的时频特性。然后,构建 MADSC 提取局部特征并进一步设计 SIDSwinT 结构有效地捕获细粒度特征和建模全局依赖关系。此外,设计 SIM 对通道间信息进行编码以自适应调整不同通道的重要性;同时,构建可变形注意力对采样信息进行加权克服了工况波动引起的分布差异,进一步提高了模型的性能;

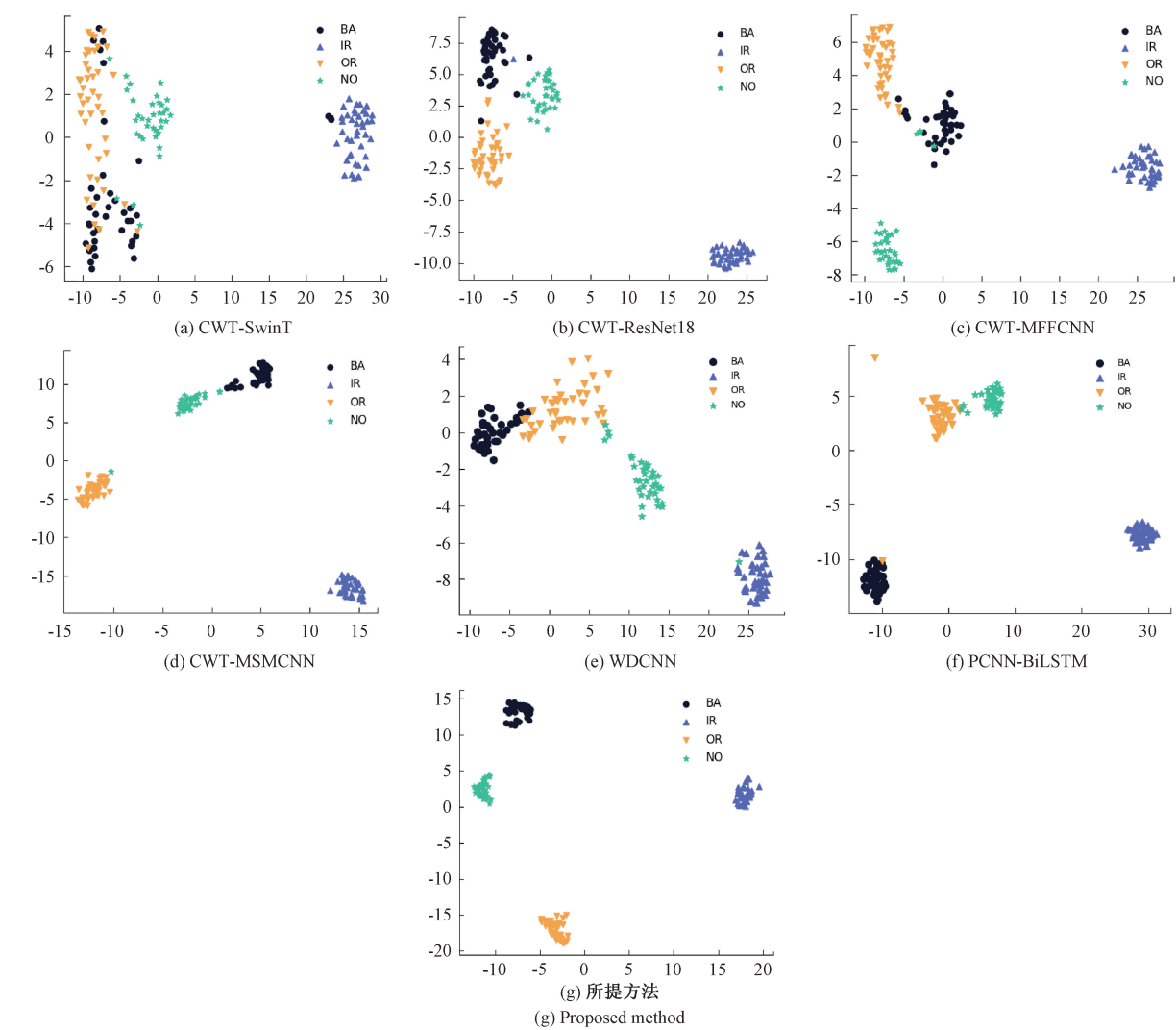


图 16 聚类可视化结果

Fig. 16 Clustering visualization results

最后,通过两个轴承数据集进行实验验证,在变负荷和变噪声情况下所提方法均具有较高的诊断准确率和较强的泛化能力。

虽然该方法在故障诊断中取得了良好的效果,但实验结果依赖于数据服从同一分布的假设。因此,在以后的工作中会考虑域自适应故障诊断方法研究。

参考文献

[1] CHEN F, ZHOU X, XU B B, et al. Instantaneous square current signal analysis for motors using vision transformer for the fault diagnosis of rolling bearings[J]. Applied Sciences, 2023, 13(16): 9349.

[2] HE M, HE D. Deep learning based approach for bearing fault diagnosis [J]. IEEE Transactions on Industry Applications, 2017, 53(3): 3057-3065.

[3] 康守强, 杨加伟, 王玉静, 等. 基于联邦多表示域适

应的不同工况下滚动轴承故障诊断方法[J]. 仪器仪表学报, 2023, 44(6): 165-176.

KANG SH Q, YANG J W, WANG Y J, et al. Fault diagnosis method of rolling bearings under different working conditions based on federated multi-representation domain adaptation[J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 165-176.

[4] 江莉, 向世召. 基于 CEEMDAN-VSSLMS 的滚动轴承故障诊断 [J]. 计算机集成制造系统, 2024, 30(3): 1138.

JIANG L, XIANG SH ZH. Fault diagnosis of rolling bearings based on CEEMDAN-VSSLMS [J]. Computer Integrated Manufacturing Systems, 2024, 30(3): 1138.

[5] LI H, LIU T, WU X, et al. A bearing fault diagnosis method based on enhanced singular value decomposition[J]. IEEE Transactions on Industrial Informatics, 2020, 17(5):

- 3220-3230.
- [6] WANG B, QIU W, HU X, et al. A rolling bearing fault diagnosis technique based on recurrence quantification analysis and Bayesian optimization SVM[J]. *Applied Soft Computing*, 2024, 156: 111506.
- [7] 石怀涛, 尚亚俊, 白晓天. 基于贝叶斯优化的SWDAE-LSTM滚动轴承早期故障预测方法研究[J]. *振动与冲击*, 2021, 40(18): 286-297.
- SHI H T, SHANG Y J, BAI X T. Research on early failure prediction method of swdae-lstm rolling bearings based on bayesian optimization[J]. *Journal of Vibration and Shock*, 2021, 40(18): 286-297.
- [8] 王进花, 周德义, 曹洁, 等. 基于多特征融合与RF的球磨机滚动轴承故障诊断[J]. *北京航空航天大学学报*, 2022, 49(12): 3253-3264.
- WANG J H, ZHOU D Y, CAO J, et al. Fault diagnosis of rolling bearings in ball mill based on multi-feature fusion and RF[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2022, 49(12): 3253-3264.
- [9] 李巍华, 单外平, 曾雪琼. 基于深度信念网络的轴承故障分类识别[J]. *振动工程学报*, 2016, 29(2): 340-347.
- LI W H, SHAN W P, ZENG X Q. Bearing fault classification and identification based on deep belief network[J]. *Journal of Vibration Engineering*, 2016, 29(2): 340-347.
- [10] 杨洁, 万安平, 王景霖, 等. 基于多传感器融合卷积神经网络的航空发动机轴承故障诊断[J]. *中国电机工程学报*, 2022, 42(13): 4933-4942.
- YANG J, WAN AN P, WANG J L, et al. Fault diagnosis of aero engine bearing based on multi-sensor fusion convolutional neural network[J]. *Proceedings of the CSEE*, 2022, 42(13): 4933-4942.
- [11] LIU H, ZHOU J, ZHENG Y, et al. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders [J]. *ISA Transactions*, 2018, 77: 167-178.
- [12] 李恒, 张氢, 秦仙蓉, 等. 基于短时傅里叶变换和卷积神经网络的轴承故障诊断方法[J]. *振动与冲击*, 2018, 37(19): 124-131.
- LI H, ZHANG Q, QIN X R, et al. Bearing fault diagnosis method based on short-time Fourier transform and convolutional neural network [J]. *Vibration and Shock*, 2018, 37(19): 124-131.
- [13] 曲建岭, 余路, 袁涛, 等. 基于一维卷积神经网络的滚动轴承自适应故障诊断算法[J]. *仪器仪表学报*, 2018, 39(7): 134-143.
- QU J L, YU L, YUAN T, et al. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network [J]. *Chinese Journal of Scientific Instrument*, 2018, 39(7): 134-143.
- [14] HAN K, XIAO AN, WU EN H, et al. Transformer in transformer [J]. *Advances in Neural Information Processing Systems*, 2021, 34: 15908-15919.
- [15] LI J, BAO Y, LIU W X, et al. Twins transformer: Cross-attention based two-branch transformer network for rotating bearing fault diagnosis[J]. *Measurement*, 2023, 223: 113687.
- [16] MO ZH L, ZHANG H, SHEN Y, et al. Conditional empirical wavelet transform with modified ratio of cyclic content for bearing fault diagnosis[J]. *ISA Transactions*, 2023, 133: 597-611.
- [17] CHEN Y, WANG G, CHEN R. Efficient multi-scale cosine attention transformer for image super-resolution[J]. *IEEE Signal Processing Letters*, 2023.
- [18] 肖茂华, 周爽, 黄天逸, 等. 参数优化VMD-MPE和PSO-CS-Elman神经网络在滚动轴承故障诊断中的应用研究[J]. *振动工程学报*, 2023, 36(3): 861-874.
- XIAO M H, ZHOU SH, HUANG T Y, et al. Application of parameter optimization vmd-mpe and pso-cs-elman neural network in fault diagnosis of rolling bearings[J]. *Journal of Vibration Engineering*, 2023, 36(3): 861-874.
- [19] JIN Y H, HOU L, CHEN Y SH. A time series transformer based method for the rotating machinery fault diagnosis[J]. *Neurocomputing*, 2022, 494: 379-395.
- [20] CHEN ZH Y, GRYLLIAS K, LI W H. Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network[J]. *IEEE Transactions on Industrial Informatics*, 2019, 16(1): 339-349.
- [21] SONG B Y, LIUI Y Y, FANG J ZH, et al. An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples[J]. *Neurocomputing*, 2024, 574: 127284.
- [22] PENG F J, MU L H, FANG CH K. Fault diagnosis of shipboard medium-voltage alternating current power system with fault recording data-driven Se-Resnet 18-1 model [J]. *IEEJ Transactions on Electrical and Electronic Engineering*, 2024, 19(3): 403-413.
- [23] GAO H F, MA J, ZHANG ZH H, et al. Bearing fault diagnosis method based on attention mechanism and multi-channel feature fusion[J]. *IEEE Access*, 2024, DOI: 10.1109/ACCESS.2024.3381618.
- [24] 段晓燕, 焦孟莹, 雷春丽, 等. 基于MTF-MSMCNN的小样本滚动轴承故障诊断方法[J]. *航空动力学报*, 2024, 39(1): 240-252.

DUAN X Y, JIAO M X, LEI CH L, et al. Fault diagnosis method for small sample rolling bearings based on MTF-MSMCNN [J]. Journal of Aerospace Power, 2024, 39(1): 240-252.

[25] 徐鹏, 皋军, 邵星. 基于 AMCNN-BiGRU 的滚动轴承故障诊断方法研究[J]. 振动与冲击, 2023, 42(18): 71-80.

XU P, GAO J, SHAO X. Research on fault diagnosis method of rolling bearing based on AMCNN-BiGRU[J]. Vibration and Shock, 2023, 42(18): 71-80.

[26] CHALESHTORI A E, AGHAIE A. A novel bearing fault diagnosis approach using the Gaussian mixture model and the weighted principal component analysis [J]. Reliability Engineering & System Safety, 2024, 242: 109720.

[27] 高乐乐, 崔宝珍, 王浩楠. 应用 Adabelief 优化器的 MSDNet 在多工况下 滚动轴承的故障诊断[J]. 电子测量与仪器学报, 2023, 36(11): 170-177.

GAO L L, CUI B ZH, WANG H N. MSDNet using adabelief optimizer for fault diagnosis of rolling bearings under multiple working conditions [J]. Journal of

Electronic Measurement and Instrumentation, 2023, 36(11): 170-177.

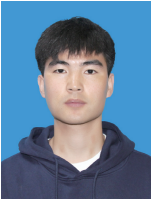
作者简介



赵小强 (通信作者), 2006 年于浙江大学获得博士学位, 现为兰州理工大学教授, 主要研究方向为故障诊断、图像处理、数据挖掘等。

E-mail: xqzhao@lut.edu.cn

Zhao Xiaoqiang (Corresponding author), received his Ph. D. degree from Zhejiang University in 2006. He is now a professor at Lanzhou University of Technology. His main research interests include fault diagnosis, image processing, and data mining etc.



安贵财, 2022 年于河南工业大学获得学士学位, 现为兰州理工大学硕士研究生, 主要研究方向为故障诊断。

E-mail: 2623934045@qq.com

An Guicai received his B. Sc. degree from Henan University of Technology in 2022. He is now a professor at Lanzhou University of Technology. His main research interest includes fault diagnosis.