DOI: 10.13382/j.jemi.B2407480

不平衡数据下的轻量化轴承故障诊断方法*

赵小强[1,2,3] 李 森[1]

(1. 兰州理工大学电气工程与信息工程学院 兰州 730050;2. 兰州理工大学甘肃省工业过程先进控制重点实验室 兰州 730050;3. 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:针对深层网络特征参数量大和故障类别样本数量不平衡导致轴承故障诊断效果差的问题,提出了一种不平衡数据下的轻量化轴承故障诊断方法。首先,将传感器所采集的一维振动信号重构为二维灰度图作为模型输入;其次,设计了非对称多尺度特征提取模块,利用不同尺度的卷积和空洞卷积对输入信号进行特征提取,并将一部分特征映射到原始空间用于去除噪声和还原原始数据结构;紧接着,被提取的丰富特征信息送入所设计的通道位置双加权模块利用反通道卷积和局部均值的方法对关键通道和关键位置特征进行双向加权;然后,设计了深度可分离卷积(DSC)密集残差结构,在保证网络轻量化的同时增加各层网络的特征融合,并通过快捷路径优化了反向传播性能;最后,利用焦点损失函数根据不同故障类别的重要性调整模型的学习过程,从而更好地适应不平衡的数据分布。利用美国凯斯西储大学轴承数据集和本实验数据集实验验证,结果表明,所提方法在不平衡数据集下故障诊断准确率最高,轻量化程度最好,并具有较好的抗噪性能。

关键词: 轴承;故障诊断;轻量化;特征加权;密集残差结构;焦点损失

中图分类号: TH133.3;TN911.7 **文献标识码**: A **国家标准学科分类代码**: 460.2020

Fault diagnosis method for lightweight bearings under unbalanced data

Zhao Xiaoqiang[1,2,3] Li Sen[1]

(1. School of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China; 2. Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou University of Technology, Lanzhou 730050, China; 3. National Experimental Teaching Center of Electrical and Control Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: To address the problem of poor bearing fault diagnosis due to the large amount of deep network feature parameters and the unbalanced number of fault category samples, this paper proposes a lightweight bearing fault diagnosis method under unbalanced data. Firstly, the one-dimensional vibration signals collected by the sensors are reconstructed into a two-dimensional grey scale map as model input. Secondly, an asymmetric multi-scale feature extraction module is designed to extract features from the input signal using convolution and null convolution at different scales, and a part of the features are mapped to the original space for removing noise and restoring the original data structure. Next, the extracted rich feature information is fed to the channel position bi-weighting module to bi-directionally weight the key channel and key position features using inverse channel convolution and local averaging. Then, a depthwise separable convolution (DSC) dense residual structure is designed to increase the feature fusion of each layer of the network while keeping the network lightweight and optimize the backpropagation performance through shortcut paths. Finally, the focal loss function is used to adjust the learning process of the model according to the importance of different fault categories, thus better adapting to the unbalanced data distribution.

Keywords: bearings; fault diagnosis; lightweighting; feature fusion; cascade residual structure; focal loss

收稿日期: 2024-04-29 Received Date: 2024-04-29

*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

滚动轴承作为旋转机械中应用最广泛的部件之一，其运行状态直接影响设备的整体性能、工作效率及使用寿命[1]。高速变载的恶劣工作环境容易使得轴承发生故障，为了避免其造成的巨大经济损失和不可预测的安全事故，进行轴承故障诊断工作是非常必要的[2-3]。

近年来，由深度学习所引领的轴承故障诊断技术突飞猛进，这些方法无需任何手工特征选择就可以自动提取关键故障特征。典型的深度学习方法包括去噪自编码器(denoising autoencoder，DAE)[4-5]、深度信念网络(deep belief net，DBN)[6-7]、稀疏自编码器(sparse autoencoder，SAE)[8-9]和卷积神经网络(convolutional neural networks，CNN)[10-11]等，其中 CNN 由于不需要引入无监督预训练过程来训练模型并且具有消除噪声的能力，因此，该方法在轴承故障诊断领域广泛应用。例如，Tian 等[12]利用混合粒子群优化 CNN 和长短时记忆神经网络的方法解决了非线性和多变量优化的问题，其中 CNN 作为 LSTM 模型的前端用于提取序列数据中的故障特征。Zhang 等[13]利用残差金字塔算法和 2D-CNN 实现了多个空间位置的声信号和振动信号的融合，并设计了一种具有删除机制的 AdaBoost 算法产生故障诊断结果。Jin 等[14]提出基于变分模态分解(variational mode decomposition，VMD)和改进卷积神经网络故障诊断方法，将经过 VMD 分解后的信号应用到 CNN 中实现轴承故障诊断。然而，上述方法以堆叠卷积的形式提取深层特征信息，这不仅使得模型特征参数量激增，而且造成模型梯度弥散和训练困难。

此外，上述方法构建数据集时设定为类别平衡，然而，实时采集的数据集中各类别是不平衡的。比如，正常样本的数量比故障样本的数量大得多；在高负载和高转速的条件下，滚动体易发生故障；在酸碱比例失衡的条件下，外圈极易发生故障。为了解决因不平衡数据导致少数类样本故障诊断精度低的问题，Wu 等[15]提出了一种全新的面向类不平衡故障诊断需求的模型，该模型以数据平衡采样器、平衡裕度损失和具有动态决策边界平衡的分类器增加少数类别的识别精度。郭俊锋等[16]提出一种基于 Wasserstein 距离条件梯度惩罚生成对抗网络的诊断方法，该方法以生成对抗网络扩充样本的形式弥补少数类样本。虽然上述方法取得了不错的结果，但是平衡采样和生成样本都会带来较高的计算成本，增加模型训练时间。

综上所述，以解决因数据不平衡问题导致故障诊断准确率低和深层网络特征参数量大的问题，提出了不平衡数据下的轻量化轴承故障诊断方法。首先，将采集的一维轴承振动信号重构为二维灰度图作为模型输入。其次，设计了非对称多尺度特征提取模块，利用不同尺度的普通卷积和空洞卷积对输入信号进行特征提取，并将提取的一部分特征映射到原始空间，用于去除噪声和还原原始数据结构。进而，设计了通道位置双加权模块，对关键通道和关键位置中的特征进行双向加权，剔除冗余信息。紧接着，设计了深度可分离卷积(depthwise separable convolution，DSC)密集残差结构，利用 DSC 大幅度降低参数量和计算复杂度，即在保证模型的轻量化的同时增加各层网络之间的特征融合。最后，由焦点损失函数调整模型的学习过程，从而更好地适应不平衡的数据分布。

1　DSC

DSC[17]是一种轻量级的卷积操作，它独特的分步计算方式不仅减少了特征参数量和计算复杂度，而且还降低了模型过拟合的风险。

DSC 的第 1 步计算称为深度卷积，它是将输入特征沿着通道的方向进行卷积操作，对于每一个通道，使用一个卷积核进行卷积，得到单独的卷积结果。假设输入数据为 $D_I \times D_I \times M$，输出通道数目为 N，则该操作得到的特征参数量为 $D_K \times D_K \times M \times D_I \times D_I$。第 2 步计算称为逐点卷积，它是利用 1×1 的卷积核对每个通道的卷积结果进行融合处理，该操作并不是在空间维度上平移卷积核，而是对每个位置上每个通道的特征信息进行加权求和，该操作中特征参数量为 $M \times N \times D_I \times D_I$，由此得到 DSC 的总特征参数量为 $D_K \times D_K \times M \times D_I \times D_I + M \times N \times D_I \times D_I$，而常规卷积操作的特征参数量 $M \times N \times D_I \times D_K \times D_K$。将 DSC 和常规卷积的特征参数量进行对比，如式(1)所示[18]，得出 DSC 所减少的特征参数量比例与卷积核大小的平方和输出通道数成正比。

$$\frac{D_K \times D_K \times M \times D_I \times D_I + M \times N \times D_I \times D_I}{M \times N \times D_I \times D_K \times D_K} = \frac{1}{N} + \frac{1}{D_K \times D_K} \tag{1}$$

式中：$D_K \times D_K$ 为卷积核的大小；D_I 为输出特征图尺寸；M 为输入通道数目；N 为输出通道数目。

2　轴承故障诊断方法

针对数据不平衡问题导致故障诊断准确率低和深层网络特征参数量大的问题，提出了一种不平衡数据下的轻量化轴承故障诊断方法，结构如图 1 所示。该方法主要由非对称多尺度特征提取模块、通道位置双加权模块、DSC 密集残差结构和焦点损失函数组成。

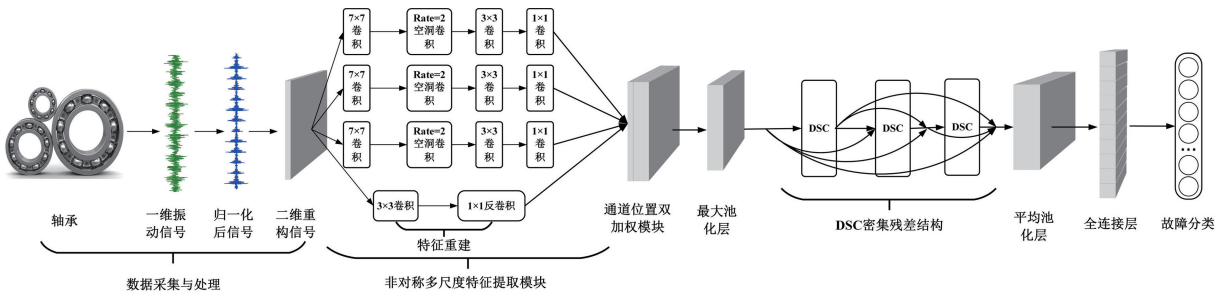

图1　所提方法结构

Fig. 1　Structure of the proposed methodology

2.1　非对称多尺度特征提取模块

滚动轴承常年在强噪声环境中工作,这导致轴承振动信号与干扰信号混叠在一起,在这种情况下,传统的单尺度网络所提取的特征十分有限,因此,为了最大程度地获取故障特征信息,设计了非对称多尺度特征提取模块,结构如图2所示。首先,第1层利用卷积核为7×7的卷积大范围捕捉混杂在干扰信号中的故障特征;第2层利用空洞率为2的空洞卷积继续提取故障特征,在保证参数量不变的同时获得较大的感受野;第3层利用卷积核为3×3的卷积在小范围内提取局部细节特征信息;最后利用1×1的卷积层在保留细节信息的同时减少模型的参数,降低模型过拟合的风险。

此外,该模块通过侧支路将提取到的一部分特征重新映射到原始输入空间中进行特征重建,迫使模型学习到更为鲁棒和稳定的表示。为了提升模型的性能,该模块对每一个卷积层的输出都进行了批量标准化和Relu激活函数处理。

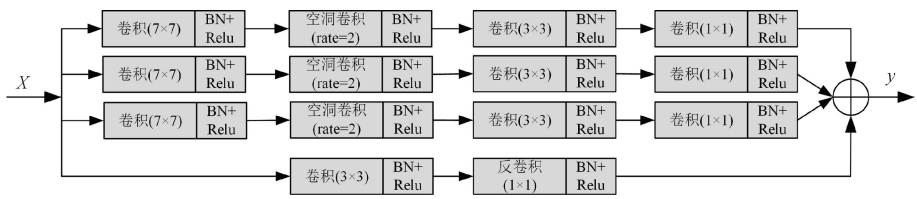

图2　非对称多尺度特征提取模块结构

Fig. 2　Structure of asymmetric multi-scale feature extraction module

2.2　通道位置双加权模块

经过非对称多尺度特征提取模块可以提取到丰富的故障特征,但不同故障所表示的特征分布各有差异,传统的通道注意力机制只能够实现不同通道的特征加权,而将通道中不同位置特征之间的差异所忽略。因此,为了实现通道和位置两个层面的特征双加权,设计了通道位置双加权模块,结构如图3所示。

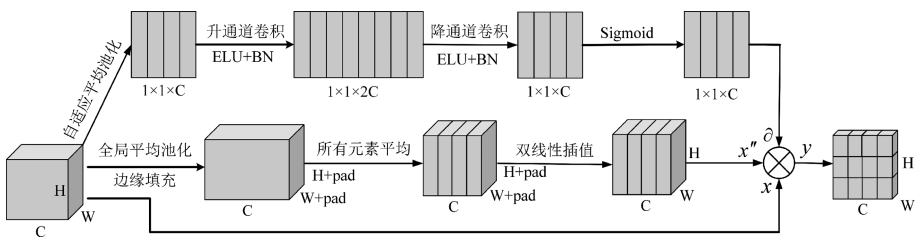

图3　通道位置双加权模块结构

Fig. 3　Channel position dual weighting module structure

针对通道层面的特征增强问题,引入了非线性变换和参数约束。首先,该模块将输入张量 $\boldsymbol{x}$ 利用自适应平均池化操作将空间维度压缩至1×1,然后利用变通道卷积先降低通道数,后增加通道数,这不仅帮助网络更好地学习特征的抽象表示,而且降低了模型的计算复杂度。为了提升模型的性能,该结构在每个卷积层的后面使用了BN层和ELU激活函数,其中BN层用于加速训练过程,ELU激活函数用于解决传统RELU函数神经元死亡

的问题,这得益于 ELU 函数在负数部分的输出不会被截断为 0,这能够更好地保持神经元的活跃性,从而改善网络训练中的梯度消失问题。在经过双层变通道卷积操作之后得到通道注意力权重 ∂,接着对该权重进行 Sigmoid 函数处理,使得最终的 ∂ 处于(0,1)范围内。上述步骤的推导过程如式(2)和式(3)所示[19]。

$$z_c = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{x}(i,j) \tag{2}$$

$$s_c = f\{\sigma[W_2\sigma(W_1 z_c)]\} \tag{3}$$

式中:c 为压缩的通道序号;z_c 为第 c 个通道压缩后的值;$\boldsymbol{x}(i,j)$ 为 $i \times j$ 维的二维矩阵;s_c 为激活后的信号;$f(\cdot)$ 为 Sigmoid 函数;σ 为 ELU 激活函数;W_1、W_2 为通道卷积的权重参数。

针对位置层面的特征增强问题,引入了局部均值。首先,该模块将输入张量 $\boldsymbol{x}$ 在空间维度上进行反射模式的边缘填充,并通过全局平均池化计算每个点周围邻域的平均值得到张量 $\boldsymbol{x}'$。然后,将 $\boldsymbol{x}'$ 除以所有元素的平均值得到张量 $\boldsymbol{x}''$,并使用双线性插值使得 $\boldsymbol{x}''$ 的大小与 $\boldsymbol{x}$ 相同,此时的 $\boldsymbol{x}''$ 代表每个点周围邻域的局部均值权重。上述步骤的推导过程如下所示:

$$\boldsymbol{x}' = \frac{1}{HW}\sum_{i=1}^{H+padW+pad}\sum_{j=1}\boldsymbol{x}(i,j) \tag{4}$$

$$\boldsymbol{x}'' = \delta\left(\frac{\boldsymbol{x}'}{mean(\boldsymbol{x}')}\right) \tag{5}$$

式中:pad 为边缘填充值;$mean$ 代表所有元素的平均值;$\delta(\cdot)$ 代表双线性插值算法。

最后,将通道和位置权重相结合,将计算得到的局部均值权重 $\boldsymbol{x}''$、∂ 和 $\boldsymbol{x}$ 相乘得到加权后的特征图 y,如式(6)所示,该步骤实现了关键通道和关键位置中重要特征的双重加权。

$$y = \boldsymbol{x} \cdot \boldsymbol{x}'' \cdot \partial \tag{6}$$

2.3　DSC 密集残差结构

为了有效地提取网络深层特征,传统的残差结构是通过增加网络的深度增强提取能力,然而,这不仅会导致特征参数量激增,而且使得模型的学习能力变差。因此,设计了 DSC 密集残差结构。

首先,该结构采用多级残差连接的方式改变了特征信息的流通路线,即前一层网络的输入被增加至后面各层网络的输入之前,这不仅增加了各级网络之间的特征融合,而且保留了原始故障的特征。当模型在正向传播时,该结构不仅可以将输入数据映射为不同的特征表示,还可以从不同的隐藏层中提取更高级别的特征表示。当模型在反向传播时,特征信息通过残差结构的快捷路径流通,这避免了梯度消失和梯度爆炸的问题。除此之外,该结构利用 DSC 独特的分步计算特性大幅度降低参数量。因此,该结构不但保证了模型的轻量化,而且使得模型的学习能力更强,抑制了退化现象。该结构如图 4 所示。

设该结构的输入为 x,经过卷积 1 提取的特征为 $f_1(x)$,而卷积 2 的输入此时为 $f_1(x)+x$,以这种方式依次进行特征表示,最终得到输出 y,步骤推导过程如下所示[20]。

$$f_1(x) = F_{RELU}(f_{BN}(DSConv(x))) \tag{7}$$

$$f_2(x) = F_{RELU}(f_{BN}(DSConv(f_1(x)))) + x \tag{8}$$

$$y = F_{RELU}(f_{BN}(DSConv(f_2(x)))) + x + f_1(x) \tag{9}$$

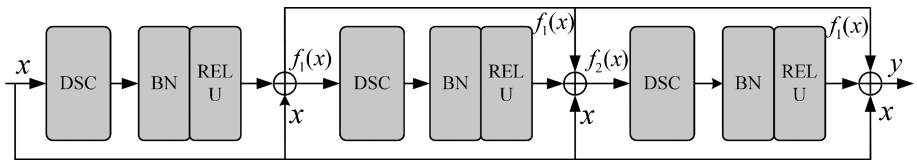

图 4　DSC 密集残差结构图

Fig. 4　DSC dense residual structure diagram

2.4　焦点损失函数

为了解决不平衡数据集造成样本不均匀和类别不平衡的问题,利用焦点损失函数根据不同类别的重要性调整模型的学习过程,从而更好地适应不平衡的数据分布,公式如式(10)所示[21]。

传统的交叉熵损失函数只通过对每一个样本的交叉熵损失进行求和或者取平均来得到整个训练集上的平均交叉熵损失。而焦点损失函数则是通过两个可调整的参数,将每个样本权重进行衰减,衰减的程度取决于分类器的置信度,即正确分类的样本置信度越大,则损失函数的权重越小,对模型的影响就越小;而错误分类的样本置信度越低,则损失函数的权重越大,对模型的影响就越大。焦点损失函数通过该方式使得模型更加关注那些难以分类的样本,从而提高模型的分类能力和泛化能力。

$$FL(P_j) = -\sum_{i=1}^{n}\partial_j(1-p_j^i)^{\gamma}\log_2(p_j^i) \tag{10}$$

式中:p_j^i 是第 i 个样本被预测为类别 j 的概率;n 是训练集中样本的数量;∂_j 是样本的系数;$\gamma \geqslant 0$ 是一个可控的超参数,用于平衡容易分类和困难分类样本之间的损失权重。

2.5　诊断流程

图5是所提方法的故障诊断流程图,其详细诊断步骤如下所示:首先将一维振动信号归一化处理后划分为训练集和测试集,然后将训练集数据转换为灰度图输入到模型中利用反向传播算法对参数进行更新优化,若训练到所设置批次即保存模型。最后,将测试集数据送入保存模型中输出诊断结果。

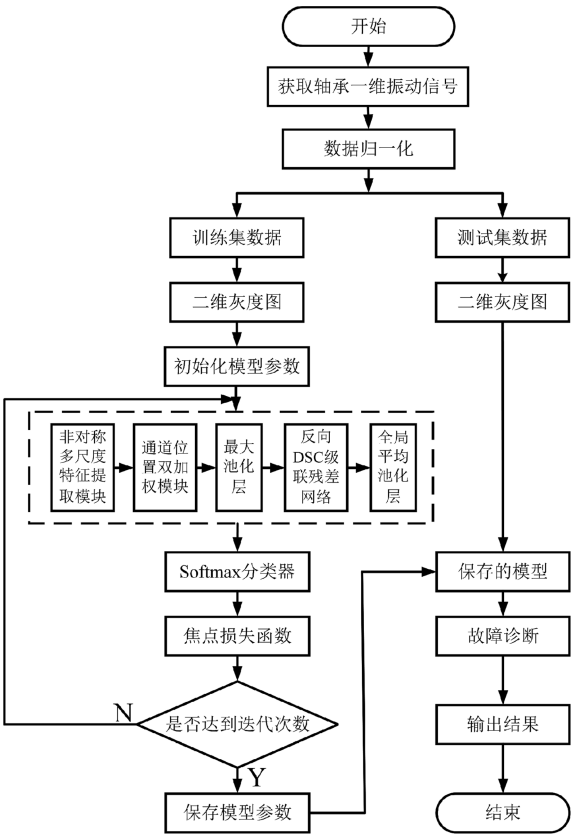

图5　所提方法流程图

Fig. 5　Flowchart of the proposed method

3　实验验证与分析

实验是在Windows10系统,酷睿i5-8300H处理器的计算机上进行,深度学习框架为Tensorflow。设置焦点损失中的 ∂ 值为0.25,β 值为2,学习率为0.001,Dropout层的弃权值为0.5,衰减率为0.9,模型利用Adam优化算法优化模型训练,迭代批数设置为200,批次为64。

3.1　案例1

1)数据集介绍

案例1的数据来源于美国凯斯西储大学(CWRU)的轴承数据中心[22]。轴承型号为SKF6202,轴承故障由电火花加工而成。该数据集记录了电机在4种转速下的运行数据。该数据集记录了4种不同位置的故障,分别为轴承滚动体、内圈、外圈6点钟方向,损伤直径分别为0.177 8、0.355 6和0.533 4 mm。依据轴承的损伤状态,将该数据集划分为10种状态,正常状态和故障状态的损伤程度对应标签如表1所示。本节实验随机选择每类故障360个样本进行实验,并按照3∶1的比例将样本划分为训练集和测试集,即训练集包含2 700个样本,测试集包含900个样本。

表1　损伤直径位置与状态对应表

Table 1　Corresponding table of damage diameter position and status

标签	损伤直径/mm	损伤位置
0	/	/
1	0.177 8	滚动体
2	0.177 8	内圈
3	0.177 8	外圈6点钟
4	0.355 6	滚动体
5	0.355 6	内圈
6	0.355 6	外圈6点钟
7	0.533 4	滚动体
8	0.533 4	内圈
9	0.533 4	外圈6点钟

2)数据预处理

由传感器所获取的数据为一维振动信号,一方面为了减轻模型的计算复杂度,另一方面为了能够利用二维卷积的空间特征提取能力有效地进行下采样操作,将长度为784的一维时间序列数据进行等长截取,获得每段具有28的数据点的28段数据,即重构为[28,28]的二维灰度图,具体操作示意如图6所示。

进一步为了增加模型的训练速度,对输入数据进行标准化处理,这有效消除了变量量纲和变异范围的影响,该操作的公式如式(11)所示[23]:

$$X = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{11}$$

式中:X 为输入数据;$X_{\min}$ 为输入数据中的最小值;$X_{\max}$ 为输入数据中的最大值。

3)比较方法

为验证所提方法在不平衡数据集等实验中是否有较好的故障诊断效果,将所提方法与4种较为主流的方法进行对比,分别为:文献[24]中的ResNet18-SVM,该方法是利用ResNet18中的残差学习完成特征提取,并使用支持向量机完成分类;文献[25]中的PS-DAN,该方法是利用一种参数共享对抗域网络完成故障诊断;文献[26]中的IRC方法,该方法是利用基于Inception网络设计的数据池化层和带跳跃连接线的残差块完成故障诊断;文献中[27]的MWCNN-ViT方法,该方法首先是利用多尺度卷积网络提取局部特征信息,然后利用Vision transformer的多头注意力机制提取全局特征信息,并依靠编码器内

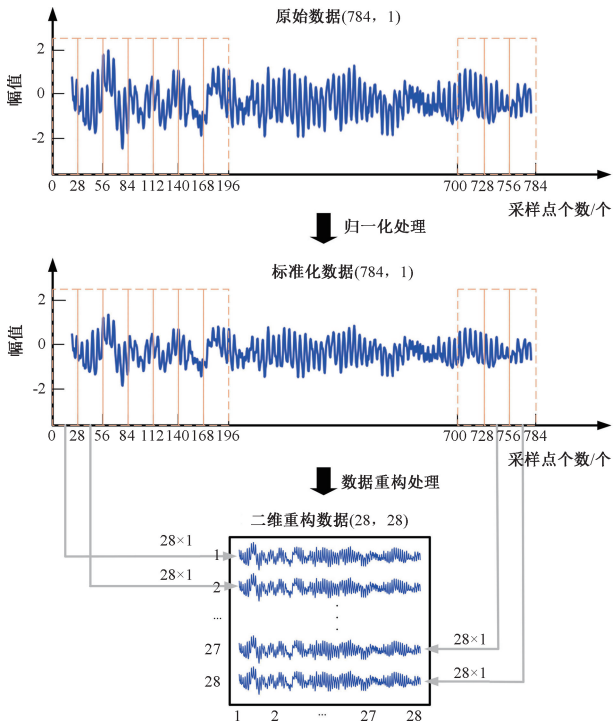

图 6　数据预处理方式图

Fig. 6　Map of data preprocessing methods

部的分类头完成故障分类。

4)轻量化实验验证与分析

模型训练时间和特征参数量是评价轻量化的重要指标,为了验证所提方法在轻量化方面的优越性,设计轻量化对比实验,对比结果如表 2 所示,可以看出,在保证故障诊断效果的同时,所提方法的训练时间和特征参数量完全优于对比方法,这归根于所提方法在保证充分特征提取的同时,尽可能减少特征参数量。首先,在模型的输入形式方面,在众多的二维特征形式中选择了最轻量化的灰度图;其次,在设计的非对称多尺度特征提取模块中使用了空洞卷积,在参数量保持不变的情况下,获得更大的感受野,避免使用多层卷积增加参数量;最后,在设计的 DSC 密集残差结构中使用 DSC 大幅度降低参数量。而 ResNet18-SVM 使用小波时频图像作为模型输入,过分专注于故障诊断准确率而忽略了模型的计算复杂度,MWCNN-ViT 中由于多层的编码器和解码器导致参数量巨大,IRC 深层的网络结构导致模型特征参数量大,PS-DAN 由于共享特征提取器和分类器之间需要在源域和目标域之间共享知识,因此特征参数量较大。

5)在不平衡数据集上的性能测试

现实工作场景中所采集数据集中的正常样本与故障样本之间的比例是严重失衡的,在这种情况下,大多数方法的性能都会受到不同程度的限制。为了验证所提方法在不平衡数据下的优越性,本节设置新的故障数据集,即原始数据集中的故障样本数量不变,而正常样本数量按照不平衡率扩大。为了降低样本选择的偶然性,本节以 5 次实验的平均值作为最终结果,如图 7 所示。可以看出,当不平衡率为 50∶1 时,所提方法的故障诊断准确率比 4 种对比方法分别高出 9.77、17.06、4.49、21.18 个百分点,这完全得益于所提方法中的焦点损失函数利用两个可调参数平衡分类样本之间的损失权重;随着不平衡率的逐渐减小,各方法的故障诊断准确率在逐渐增大,当不平衡率为 5∶1 时,所提方法的故障诊断准确率最高,仅存在 0.24%错误分类,IRC 与 PS-DAN 的故障诊断效果一般,故障诊断率分别为 96.74%和 96.35%,而 CWT-ViT 和 ResNet18-SVM 的故障诊断效果较差,这是因为 SVM 分类器和 ViT 中的分类头极易受到样本不平衡率的制约。

表 2　不同方法轻量化对比

Table 2　Comparison results of different methods of lightweighting

方法	训练时间/s	特征参数量/个
IRC	178.95	2 839 537
ResNet18-SVM	1 684.73	10 279 753
PSDAN	267.98	5 478 739
MWCNN-ViT	1 983.04	12 289 928
所提方法	144.29	1 599 317

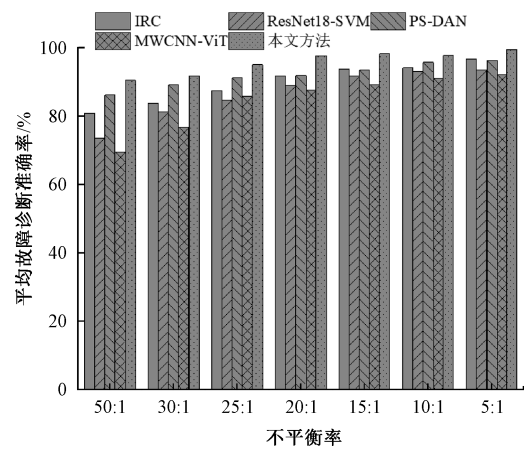

图 7　不平衡数据集下对比结果

Fig. 7　Comparison results under unbalanced dataset

6)抗噪性能验证与分析

在实际的工作环境中,由于各零部件之间的摩擦和周围环境所产生的强噪声导致诊断方法的分类效果受到影响,因此,只有抗噪性能极强的诊断方法才能实现高效的故障诊断。为此,本小节对所提方法进行了抗噪性能实验,即在测试样本中添加不同信噪比(SNR)的高斯白噪声获得不同强度的噪声样本。SNR 是评价信号中所含噪声的重要指标,其表达式如式(12)所示[28]。

$$SNR = 10 \cdot \lg \frac{P_{singal}}{P_{noise}} \tag{12}$$

式中：P_{singal} 为信号的有效功率；P_{noise} 为噪声功率。

本实验选取转速为 1 772 r/min 下的样本作为训练样本和测试样本，其中，训练集样本不变，而向测试集样本中分别引入 6、9、12 dB 的高斯白噪声获得不同程度的噪声样本，接着对模型进行训练和测试获得混淆矩阵结果如图 8 所示。由图可知，当信噪比为 6 dB 时，准确率达到 94.50% 及以上，当信噪比为 9 dB 时，准确率为 99.00%，存在第 2 个标签和第 9 个标签的故障判断错误。当信噪比为 12 dB 时，准确率为 99.95%，仅存在第 9 个标签 0.05% 的错误分类。因此，可以得出结论，所提方法在不同的噪声环境下仍然具有较强的故障诊断能力。

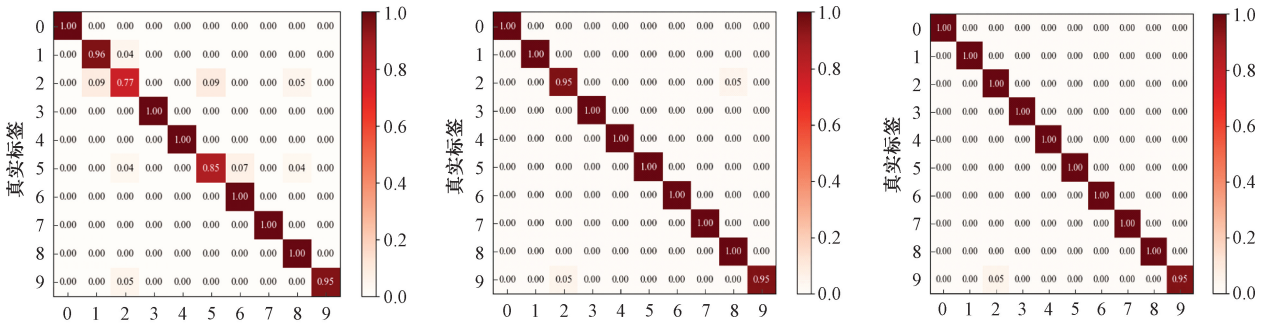

图 8　不同噪声下的测试集结果图

Fig. 8　Plot of test set results under different noises

7）消融实验结果与分析

为了验证所提方法主要创新点的有效性，设置非对称多尺度特征提取模块、DSC 密集残差结构、通道位置双加权模块为变量进行消融实验。当验证非对称多尺度特征提取模块时，使用单尺度卷积网络代替该模块；当验证 DSC 密集残差结构时，使用传统残差结构代替该模块；当验证通道位置双加权模块时，使用传统通道注意力机制代替该模块。

结果如表 3 所示，当所提模型中缺少非对称多尺度特征提取模块时，平均故障诊断准确率仅为 91.47%，这说明非对称多尺度特征提取模块比传统的单尺度卷积网络提取特征信息更加充分；当所提模型中缺少 DSC 密集残差结构时，平均故障诊断准确率仅为 90.51%，这说明了该结构改变特征传输路线和增加各层之间的特征融合是有优势的；当所提模型中缺少通道位置双加权模块时，平均故障诊断率为 97.73%，这说明该模块对提取的特征信息进行了有利筛选，使得模型更加关注关键通道和关键位置的特征信息。当 3 个模块都存在时，平均故障诊断准确率为 98.35%，这说明只有当这 3 个模块相互配合才能完成恶劣条件下的故障诊断工作。

表 3　消融实验故障诊断准确率对比表

Table 3　Comparison table of troubleshooting accuracy of ablation experiments

非对称多尺度特征提取模块	DSC 密集残差网络	通道位置双加权模块	准确率/%			
			3 dB	6 dB	9 dB	12 dB
	√	√	83.70	92.40	94.47	95.32
√		√	77.70	90.80	96.40	97.14
√	√		95.13	96.90	99.27	99.62
√	√	√	96.01	97.42	99.98	100.0

3.2　案例 2

1）数据集介绍

为了进一步验证所提方法的有效性，案例 2 利用本实验室的 MFS 轴承数据集进行实验验证，采集装置如图 9 所示，该平台主要由驱动端、加速度传感器、交流电机、交流控制器和信号采集器组成。轴承类型为深沟球轴承，型号为 ER-16K，故障缺陷是采用激光刻蚀技术加工得到。数据是通过数据线连接信号采集器和加速度传感器，并通过 USB 接口将信号传输到计算机采集到的。该数据集记录了 4 种不同的故障状态，分别为滚动体、内圈、外圈、内外圈混合故障，故障部位如图 10 所示。本节根据损伤位置和损伤直径将该数据集划分为 4 种

标签,如表 4 所示。本节实验随机选择每类故障 900 个样本进行实验,并按照 3∶1 的比例将样本划分为训练集和测试集,即训练集包含 2 700 个样本,测试集包含 900 个样本。

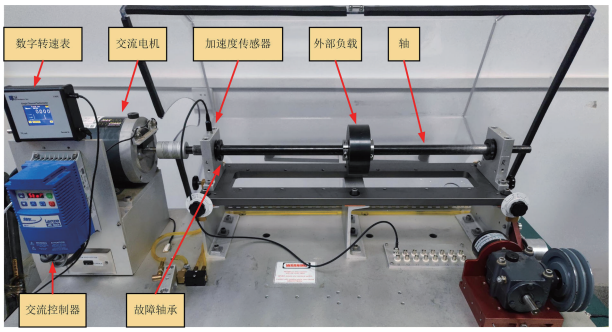

图 9　MFS 轴承实验台

Fig. 9　MFS Bearing Test Stand

(a) 滚动体故障　(b) 内圈故障

(a) Ball fault　(b) Inner fault

(c) 外圈故障　(d) 混合故障

(c) Outer fault　(d) Combine fault

图 10　故障部位图

Fig. 10　Fault location diagram

表 4　MFS 数据集标签故障对应关系

Table 4　MFS dataset label fault correspondence

标签	损伤直径/mm	故障类型
0	0.6	滚动体故障
1	2.4	内圈故障
2	0.6	外圈故障
3	(0.6)和外圈(0.9)	内外圈混合故障

2)T-SNE 可视化实验

为了验证所提方法在 MFS 数据集下的故障分类效果,本小节对模型中 4 个不同位置进行了 T-SNE 可视化,结果如图 11 所示。可以看出,输入层数据杂乱无章,相互堆叠。经过非对称多尺度特征提取模块和通道位置双加权模块之后,相同类的故障相互聚集,不同类的故障相互分离,但未完全聚合或者分离。经过全连接层和分类层之后,不同类故障已经完全分离。这说明所提方法通过各模块之间的相互配合使得各类故障最终被准确识别。

3)在不平衡数据集上的性能测试

为了进一步验证所提方法在不平衡数据集上的优越性,本节实验在 MFS 数据集下重新设置数据集,即原数据集中的故障样本数量不变,而正常样本数量按照不平衡率扩大,并以 5 次实验的平均值作为最终的结果,如表 5 所示。当不平衡率为 50∶1 时,所提方法的平均故障诊断率比对比方法分别高出 7.83、12.22、6.38、9.07 个百分点。随着不平衡率的逐渐减小,各方法的故障诊断准确率在逐渐增大,当不平衡率为 5∶1 时,所提方法已无错误分类,而其余方法均有不同程度的分类错误。因此,可以得出结论,所提方法在不同的轴承设备下依然具有较好的分类效果。

表 5　不平衡数据集下的对比结果

Table 5　Comparison results under unbalanced dataset

不平衡率	故障诊断准确率/%				
	IRC	ResNet18-SVN	PS-DAN	MWCNN-ViT	本文方法
50∶1	80.56	76.17	82.01	79.32	88.39
30∶1	83.87	79.44	85.52	83.19	91.90
25∶1	86.61	82.95	87.01	85.73	93.61
20∶1	90.96	86.28	90.26	88.64	98.75
15∶1	92.60	90.06	93.66	90.39	98.81
10∶1	94.05	91.17	94.69	94.02	99.86
5∶1	95.34	93.49	95.35	95.25	100.0

4　结　论

针对深层网络特征参数量大和故障类别样本数量不平衡导致轴承故障诊断效果差的问题,提出了一种不平衡数据下的轻量化轴承故障诊断方法。首先,设计了非对称多尺度特征提取模块,利用不同尺度的卷积对输入信号进行特征提取和特征映射,去除噪声和还原原始数据结构。其次,设计了通道位置双加权模块,利用变通道卷积和局部均值从通道和位置两个层面进行特征加权,剔除了冗余信息。进而,设计了 DSC 密集残差结构,利用 DSC 和级联残差网络在保证网络轻量化的同时增加各层网络之间的特征融合,并通过残差的快捷路径减少了信息丢失,提高了反向传播的能力。最后,使用焦点损失函数衡量分类器的预测准确性,增强了分类效果。同时,后续将结合轴承工况变换频繁的问题深入域自适应研究。

参考文献

[1]　杨建华, 韩帅, 张帅, 等. 强噪声背景下滚动轴承微

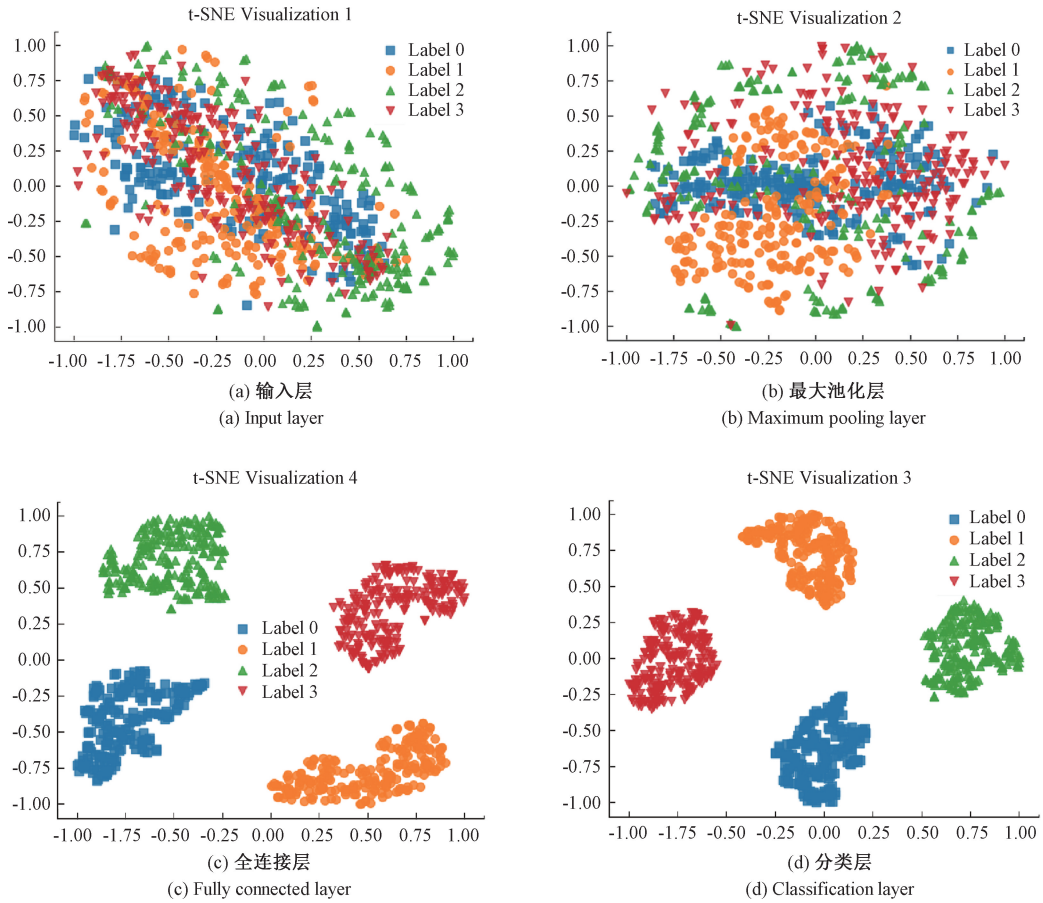

图 11　不同方法 T-SNE 图

Fig. 11　T-SNE plots of different methods

弱故障特征信号的经验模态分解 [J]. 振动工程学报, 2020, 33(3): 582-589.
YANG J H, HAN SH, ZHANG SH, et al. Empirical modal decomposition of weak fault signature signals of rolling bearings in strong noise backgrounds[J]. Journal of Vibration Engineering, 2020, 33(3): 582-589.

[2] 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,2022,36(11):170-177.
GAO L L, CUI B ZH, WANG H N. Fault diagnosis of rolling bearings under multiple operating conditions using MSDNet with the Adabelief optimizer [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(11):170-177.

[3] 王爽,丁传仓,曹懿,等. 加权多尺度卷积稀疏表示及其在滚动轴承复合故障诊断中的应用[J]. 仪器仪表学报,2024,45(5):197-207.
WANG SH, DING CH C, CAO R, et al. Weighted multiscale convolutional sparse representation and its application in rolling bearing compound fault diagnosis[J]. Chinese Journal of Scientific Instrument, 2024,45(5): 197-207.

[4] ZHANG J L, CHEN ZH G, DU X L, et al. Application of stack marginalised sparse denoising auto-encoder in fault diagnosis of rolling bearing [J]. The Journal of Engineering,2018,2018(16):1772-1777.

[5] CHE CH CH, WANG H W, NI X M, et al. Intelligent fault diagnosis method of rolling bearing based on stacked denoising autoencoder and convolutional neural network[J]. Industrial Lubrication and Tribology, 2020, 72(7): 947-953.

[6] SHIRKE S D, RAJA BHUSHNAM C. Optimization driven deep belief network using chronological monarch butterfly optimization for iris recognition at-a-distance [J]. International Journal of Knowledge-based and Intelligent Engineering Systems, 2022, 26(1): 17-35.

[7] GAO SH ZH, XU L T, ZHANG Y M, ET AL. Rolling bearing fault diagnosis based on SSA optimized self-adaptive DBN [J]. ISA Transactions, 2022, 128: 485-502.

[8] XU F, TSE P W. Automatic roller bearings fault diagnosis using DSAE in deep learning and CFS algorithm[J]. Soft Computing, 2019, 23: 5117-5128.

[9] ZHANG W H, LIU ZH F, LIAO ZH Q. Digital twin inspired intelligent bearing fault diagnosis method based on adaptive correlation filtering and improved SAE classification model [J]. Mathematical Problems in Engineering, 2022, 2022(1): 8767974.

[10] JIA L, CHOW T W S, YUAN Y. GTFE-Net: A gramian time frequency enhancement CNN for bearing fault diagnosis [J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105794.

[11] CUI K, LIU M, MENG Y N. A new fault diagnosis of rolling bearing on FFT image coding and L-CNN[J]. Measurement Science and Technology, 2024, 35(7):076108.

[12] TIAN H, FAN H C, FENG M W, et al. Fault diagnosis of rolling bearing based on hpso algorithm optimized cnn-lstm neural network[J]. Sensors, 2023, 23(14): 6508.

[13] ZHANG K, GAO T H, SHI H T. Bearing fault diagnosis method based on multi-source heterogeneous information fusion[J]. Measurement Science and Technology, 2022, 33(7): 075901.

[14] JIN ZH ZH, CHEN D, HE D Q, et al. Bearing Fault Diagnosis Based on VMD and Improved CNN [J]. Journal of Failure Analysis and Prevention, 2023, 23(1): 165-175.

[15] WU J Y, ZHAO ZH B, SUN CH, et al. Learning from class-imbalanced data with a model-agnostic framework for machine intelligent diagnosis [J]. Reliability Engineering & System Safety, 2021, 216: 107934.

[16] 郭俊锋,王淼生,孙磊,等. 基于生成对抗网络的滚动轴承不平衡数据集故障诊断新方法[J]. 计算机集成制造系统,2022,28(9):2825-2835.
GUO J F, WANG M SH, SUN L, et al. New method for fault diagnosis for rolling bearing unbalance dataset based on generative adversarial network [J]. Computer Integrated Manufacturing Systems, 2022, 28 (9): 2825-2835.

[17] LI X Y, YUAN P, WANG X K, et al. An unsupervised transfer learning bearing fault diagnosis method based on depthwise separable convolution [J]. Measurement Science and Technology, 2023, 34(9): 095401.

[18] 金钰森,丁飞,陈竺,等. 基于深度可分离卷积神经网络的轴承故障诊断模型[J]. 无线电通信技术,2024,50(1):193-202.
JIN Y S, DING F, CHEN ZH, et al. Bearing fault diagnosis based on deep convolutional neural network[J]. Radio Communications Technology, 2024, 50 (1): 193-202.

[19] 张洪亮,余其源,秦超群,等. 基于信息融合及双连接注意力残差网络的轴承故障诊断[J]. 振动与冲击,2023,42(20):114-123.
ZHANG H L, YU Q Y, QIN CH Q, et al. Bearing fault diagnosis based on double-connected attention residual network and information fusion[J]. Journal of Vibration and Shock, 2023,42(20):114-123.

[20] ZHAO X Q, GUO H K. Rolling bearing fault diagnosis model based on DSCB-NFAM[J]. Measurement Science and Technology, 2023, 35(1): 015029.

[21] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.

[22] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the case Western Reserve University data: A benchmark study[J]. Mechanical Systems and Signal Processing, 2015, 64: 100-131.

[23] 赵小强,张亚洲. 利用改进卷积神经网络的滚动轴承变工况故障诊断方法[J]. 西安交通大学学报,2021,55(12):108-118.
ZHAO X Q, ZHANG Y ZH. Improved CNN-based fault diagnosis method for rolling bearings under variable working conditions [J]. Journal of Xi′ an Jiaotong University, 2021,55(12):108-118.

[24] YUAN L H, LIAN D SH, KANG X, et al. Rolling bearing fault diagnosis based on convolutional neural network and support vector machine[J]. IEEE Access, 2020,8: 137395-137406.

[25] QIN Y, YAO Q W, WANG Y, et al. Parameter sharing adversarial domain adaptation networks for fault transfer diagnosis of planetary gearboxes[J]. Mechanical Systems and Signal Processing, 2021, 160: 107936.

[26] 赵小强, 梁浩鹏. 使用改进残差神经网络的滚动轴承变工况故障诊断方法 [J]. 西安交通大学学报,

2020, 54(9): 23-31.
ZHAO X Q, LIANG H P. Fault diagnosis of rolling bearings under variable operating conditions using improved residual neural network[J]. Journal of Xi′an Jiaotong University, 2020, 54(9): 23-31.

[27] LI S, ZHAO X Q. A lightweight multi-feature fusion vision transformer bearing fault diagnosis method with strong local sensing ability in complex environments[J]. Measurement Science and Technology, 2024, 35(6): 065104.

[28] 邢自扬,赵荣珍,吴耀春,等. 多尺度卷积神经网络小样本轴承故障辨识方法[J]. 振动. 测试与诊断,2023, 43(5):915-922,1037-1038.
XING Z Y, ZHAO R HZ, WU Y CH, et al. Small sample bearing fault identification method using novel multi-scale convolutional neural network[J]. Journal of Vibration, Measurement & Diagnosis, 2023,43(5):915-922,1037-1038.

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 PhD degree from Zhejiang University in 2006, and is now a professor at Lanzhou University of Technology, with main research interests in fault diagnosis, image processing, and data mining etc.

李森,2022 年于兰州理工大学获得学士学位,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: xqzhao@lut.edu.cn

Li Sen received his B. S. degree from Lanzhou University of Technology, Lanzhou, China, in 2022, and is currently a master's student at Lanzhou University of Technology, Lanzhou, China, with a major research interest in fault diagnosis.