

DOI: 10.19650/j.cnki.cjsi.J2209762

基于孪生网络的小样本滚动轴承故障诊断研究*

徐卓飞¹, 李旭东¹, 张婵婵¹, 侯和平¹, 张武²

(1. 西安理工大学印刷包装与数字媒体学院 西安 710048; 2. 西安科技大学机械工程学院 西安 710054)

摘要:针对小样本和强噪声条件下的滚动轴承故障诊断问题,提出了一种孪生网络模型:首先,对于滚动轴承故障信号进行连续小波变换以获得时频图像,引入卷积神经网络模型以实现故障图像模式识别;进而,对故障样本进行交叉配对以重新组合,实现了少量故障样本的大幅扩容;同时,针对扩容后样本对数据构建了包含两个子模型的孪生网络模型;最后,为了实现强噪声、小样本条件下滚动轴承故障诊断,设计了孪生网络末端专用分类器,在加噪声数据库和机械故障实验中对方法进行测试,分别达到了96.25%和97.08%正确率。所提出模型能够依靠少量样本完成训练并实现轴承故障准确诊断,所需每类样本的数量可减少至20个,与经典卷积神经网络模型相比具有明显优势。

关键词:孪生网络;卷积神经网络;连续小波变换;故障诊断

中图分类号: TH17 **文献标识码:** A **国家标准学科分类代码:** 460.2020 460.5510 520.6080

Fault diagnosis of rolling bearings with limited samples based on siamese network

Xu Zhuofei¹, Li Xudong¹, Zhang Chanchan¹, Hou Heping¹, Zhang Wu²

(1. Faculty of Printing Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an 710048, China;
2. College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: A siamese network model is proposed for fault diagnosis of rolling bearings under small samples and strong noise. First, a series of time-frequency images are obtained from fault signals by the continuous wavelet transform, and the convolutional neural network is introduced to realize the pattern recognition. Secondly, the small samples are recombined with each other to form new sample pairs through cross matching. Thus, the number of fault samples are increased dramatically. Thirdly, a siamese network model including two sub-models is formulated, which uses the new sample pairs. Finally, a new classifier is designed for the siamese network model to realize fault classification with small samples under strong noise. The proposed faulty diagnose method is evaluated by using fault samples from both noise database and experimental measurement. The accuracy values are 96.25% and 97.08%, respectively. Results show that one fault can be identified by the proposed siamese network model using only 20 samples, which is less than the samples required by CNN model to reach a similar accuracy.

Keywords: siamese network; convolutional neural network; continuous wavelet transform; fault diagnosis

0 引 言

近年来,随着现代测试技术和人工智能等领域的发展,以数据驱动和机器学习为代表的智能化机械故障诊断模型正在逐步替代人工经验知识,成为该领域重要研究与发展方向^[1-3]。

滚动轴承作为机械设备核心传动部件,其安全稳定运行极为重要,若要实现其智能监测及诊断目标,除了需要选择合理的机器学习理论方法外,还在很大程度上需要充足的故障样本以完成诊断模型的训练和建模^[4-5]。然而,实际工程中故障样本的获取通常存在成本高昂、数量有限、数量不均等问题,进而从很大程度上限制了智能模型的应用范围及行业推广^[6-8]。因此,国内外学者针对

收稿日期:2022-05-11 Received Date: 2022-05-11

* 基金项目:中国博士后基金(2019M663783)、陕西省自然科学基金基础研究计划(2021JM-334)、陕西省技术创新引导专项-区域创新能力引导计划(2020QFY03-06)项目资助

小样本条件下滚动轴承智能诊断开展了系列研究: Fan 等^[9]针对样本不均问题提出了一种模糊支持向量机 (support vector machine, SVM), 利用熵值构建隶属度判定函数, 实现了部分样本不足或不均条件下的模型构建, 发现 SVM 在小样本训练和预防过拟合方面具有一定优势; 陈超等^[10]提出一种改进最小二乘 SVM 的迁移学习方法, 运用辅助数据弥补不同工况下滚动轴承样本不足, 大幅提高了诊断精度。SVM 作为一种浅层网络模型, 在面对机械故障信号复杂时频信息时, 其识别能力和泛化能力有限。

随着深度学习技术的快速发展, 多种深层神经网络模型被应用于小样本诊断问题中: Mao 等^[11]采用生成对抗网络 (generative adversarial network, GAN) 与堆栈降噪自编码器结合的方法解决了轴承故障样本数据不平衡问题, 依靠 GAN 生成故障样本并提高了诊断精度; 戴俊等^[12]利用自编码器 (auto-encoders, AE) 对生成样本信息和正常样本信息进行重构学习, 通过生成网络与鉴别网络之间的博弈缩小学习误差, 在无故障样本情况下实现机械系统的异常监测; Zhou 等^[13]设计了一种 GAN 模型, 对发生器和鉴别器进行了改进, 其采用全局优化方案生成更多的判别故障样本, 弥补了小样本故障诊断引起的数据分布不均匀问题; Zhang 等^[14]利用注意力机制提升了 BiGRU 模型的效果, 所建立模型在小样本条件下有效捕捉了信号特性并实现了轴承和齿轮的故障诊断。GAN、AE 等模型能够根据已有样本信息特性进行学习和重构, 生成一定数量的模拟样本以实现样本数量的增加, 这类方法在图像处理等领域取得丰富成果, 但机械故障信号具有噪声干扰强、规律不明确等特点, 生成样本与真实样本之间可能会存在较大差异, 导致正确率大幅下降。

在深度学习模型中, 卷积神经网络 (convolutional neural network, CNN) 凭借其良好鲁棒性和泛化能力^[15-16], 在强噪声、小样本工况下的轴承故障诊断中得到广泛应用; Jia 等^[17]提出了一种深度标准化卷积神经网络模型, 用以解决经典 CNN 中样本分布不均现象, 所提出方法在特征学习和诊断效果方面均有明显提升; 朱会杰等^[18]针对轴承故障信号的冲击特性, 提出了权值求和和大尺度最大值池化等策略, 有效解决了特征的平移不变性, 提高了模型的泛化能力; Seungmin 等^[19]针对样本数量不足、分布不均等问题, 构建了一种集成化 CNN 模型, 通过多个 CNN 子网络组合提高了综合诊断正确率; 张西宁等^[20]提出了以全局均值池化层取代全连接层的改进型 CNN 模型, 通过预训练微调的方式, 在跨工况、跨型号的故障诊断中完成了迁移学习任务, 解决了小样本过拟合问题; 刘飞等^[21]提出了一种基于迁移学习与深度残差网络的滚动轴承快速故障诊断算法, 残差网络由经典 CNN 发展而来, 其通过构建残差块并使用残差连接的

方式, 有效降低了 CNN 映射学习难度, 提高了有限故障样本下的诊断速度和正确率; 何强等^[22]构建了一种梯度惩罚 Wasserstein 距离生成对抗网络并用于扩充故障样本, 构建了基于自注意力卷积神经网络的故障诊断模型, 该方法在轴承负载不平衡和小样本情况下的鲁棒性和泛化能力均优于传统 CNN; Zou 等^[23]在强噪声、小样本条件下的转子系统滚动轴承故障诊断中采用了对抗降噪卷积神经网络模型, 提高了抗噪能力并且利用对抗机制扩展了学习样本。以上方法从不同角度提升了 CNN 网络泛化能力和迁移能力, 大幅降低了建立模型所需的样本数量, 提高了诊断效率。通过对 CNN 结构的改进能够有效提升小样本处理能力, 但无法像 GAN、AE 等模型一样增加新的生成样本。

分析当前各类深层次网络模型在小样本故障诊断中的特点, 认为 CNN 及其衍生模型具备较强的抗噪声能力和泛化能力, 为了进一步加强 CNN 的小样本故障诊断能力, 本文提出了一种基于孪生网络的小样本故障诊断方法: 研究以两个 CNN 网络为基本单元, 将故障信号的时频联合图像作为输入信息, 以孪生网络结构为核心结构实现两个子模型的组合应用, 针对故障诊断设计了专用分类器, 实现了强噪声、小样本状态下的滚动轴承故障诊断, 通过西储大学滚动轴承数据库和真实机械设备实测数据证明了所提出方法的正确性和有效性。

1 理论分析

1.1 连续小波分析

对故障信号开展连续小波分析, 将获取的小波系数生成一幅时频图像, 并将其转存为数字图像作为故障模式分类的样本。连续小波变换与 Gabor 变换具有相同的优点, 二者都能够同时在时域和频域观测信号特性, 克服傅里叶变换全局性的缺陷。设待分析信号为 $f(t)$ 且其属于空间 $L^2(\mathbb{R})$, 则该信号的连续小波变换表述为:

$$W_f(a, b) = \int_{-\infty}^{\infty} f(t) \overline{\psi_{a,b}} dt \quad (1)$$

其变换结果 $W_f(a, b)$ 表示在一系列尺度因子 a 和偏移因子 b 下的小波系数, 在连续小波变换中二者均为连续变量; 小波函数系数反映了连续小波函数 $\psi_{a,b}$ 与待分析信号 $f(t)$ 之间的相似度。 $\psi_{a,b}$ 在式(1)中根据复频域内积定义取其复共轭, 具体形式如式(2)所示。

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

可知 $\psi_{a,b}$ 是母小波 ψ 在不同 a 和 b 取值下得到的一系列函数族^[24-25]。

在图像输入 CNN 前, 需要先对其进行预处理, CNN 输入层参数与图像尺寸密切相关, 合理的预处理将避免

过多无效参数设定,大幅提高模型训练效率并增强泛化能力。预处理过程中先要将小波系数进行图像化展示并将其压缩。压缩过程中包含对图像进行几何变换和重采样,参考同类研究常选取 32×32 至 256×256 像素范围内^[26-27],由于 CNN 内部大量参数采取的是稀疏连接方式,过多的参数不仅消耗资源且没有必要,本文将频谱图像压缩至较低的 64×64 像素。

1.2 卷积神经网络

本文所提出孪生网络模型建立在两个卷积神经网络子模型的基础上,这里对其核心网络结构层进行描述,以便于理解后续孪生模型结构。基本卷积神经网络包含输入层、多组卷积层和池化层、全连接层和输出层^[28]。

1) 卷积层

卷积层采用滑动窗口运算的方式对图像中像素进行遍历卷积运算,以实现提取图像局部特征。使用卷积核与输入图像进行卷积操作后,经过激活函数的非线性映射,即可得到一个相应输出图像,如式(3)所示。

$$x_j^c = f\left(\sum_{i \in M_j} x_i^{c-1} \times k_{ij}^c + b_j^c\right) \quad (3)$$

式中: c 为卷积层的层数; M_j 为第 $c-1$ 层的特征图个数; x_j^c 为第 c 层的第 j 个输出特征图; k_{ij}^c 为在第 c 层中与第 i 个特征图相关的第 j 个卷积核; b_j^c 为第 c 层第 j 个卷积核的偏置, f 为激活函数。

2) 池化层

CNN 中池化层与卷积层一般成对出现,卷积计算后立即开始池化计算。池化可以大幅度压缩特征图像,舍去冗余信息进而提高特征表达能力。池化计算方式本质是在某种规则下用一个像素对该区域多个像素进行表征,这里的某种规则包含平均值、最大值、随机选取等。其中最大值池化计算在 CNN 研究中适应性强且鲁棒性好,其数学表达式如式(4)所示。

$$X = f[\alpha S(x) + b] \quad (4)$$

式中: X 为输出, f 为激活函数, α 表示下采样系数, b 代表偏置项。

3) 输入层和输出层

按照本文图像化结果,单组 CNN 输入层包含 $64 \times 64 \times 3$ 共 12 288 个神经元;经过卷积和池化运算后,数据量将被大幅压缩,最后通过全连接层将输出降维至 128×1 维特征向量。

4) 激活函数

在卷积层选用了 Relu 激活函数,其表达式如(5)所示。Relu 是一种分段式激活函数,输入为负数时,输出结果为零,输入为正数时,输出与输入相同。

$$\text{ReLU}(x) = \max(0, x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (5)$$

ReLU 激活函数曲线如图 1(a) 所示,由于 ReLU 函数运算简单,其计算速度和模型收敛速度较快;其负半轴为 0,因此增加了稀疏性,可以有效减少过拟合。

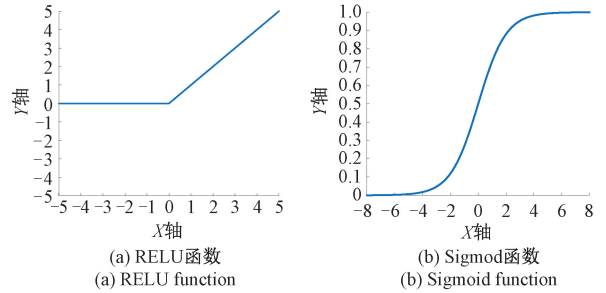


图 1 激活函数曲线图

Fig. 1 Activation function curves

全连接层通常采用 Sigmoid 激活函数,该函数应用广泛,当神经元所受到的激励大于阈值后会被激活,神经元就会产生输出信号,否则神经元处于抑制状态。Sigmoid 的公式为:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

Sigmoid 激活函数曲线如图 1(b) 所示,经过激活函数后输出值会被限制在 $0 \sim 1$,分别控制神经元处于抑制状态和激活状态。

2 孪生网络模型

通过原始故障样本进行交叉配对的形式,可以实现小样本扩容,原始样本将以新的样本对形式呈现;针对这种样本对,需要进行交叉训练与识别才能够完成分类任务。孪生网络通过两个完全相同的子模型,实现对样本对的交叉比对与模式识别。

2.1 小样本交叉扩充

样本对的扩充方案为:假设原始数据集中一共有 n 类故障,每类故障有 m 个样本,则原始数据集一共有 $n \times m$ 个样本,样本对本质是从其中抽取两个样本重新组队,这样的扩充样本对一共包含两类:第一类是同属于一类故障的同类样本相互配对,由于一共有 n 类样本,根据排列组合计算同类样本对数量为:

$$nA_m^2 = n \times m \times (m - 1) \quad (7)$$

第二类是不同故障类别之间相互配组成样本对,因此一共有 $n \times (n - 1)$ 种组合,则异类样本对总数量为 $A_n^2 C_m^1 C_m^1 = n \times (n - 1) \times m \times m$

孪生网络的输入是样本对,该样本对包含来自相同或不同类别的两个样本。上述配对扩充方法大幅增加了样本数量,总数量为式(7)和(8)的总和,相当于原始样

本扩充了 $(n \times m - 1)$ 倍, 相对于小样本故障识别, 扩充后样本对在建模和识别中难度将会大幅降低。

但是经过配对后, 原有的分类关系被彻底打乱, 若要识别某类故障, 需要从该类故障中选取一个原始样本, 将待检测样本与其共同构成一个样本对, 通过网络模型评估这个样本对中两个样本的相似度, 相似度越高, 则说明该样本越接近。因此, 样本对在多分类问题中使用时, 需要设计专用的分类器。

2.2 孪生网络的组成

孪生网络模型由 Chopra 等^[29]提出, 其利用两个子网络模型分别输入一个样本, 对两个子网络获取的特征向量进行相似度比较, 从而判断两个样本特征向量的相关性, 实现分类等功能。孪生网络不同于一般模型输出的分类结果, 其最后输出的是一个相似度指标, 该指标反映

了两个子网络输出的特征向量的相似性。孪生网络的原理与样本对形式可以很好匹配, 进而用于识别样本对中两个样本的差异。

依据样本扩充对和孪生网络的基本原理, 本文构建孪生网络基本结构如图 2 所示: 模型由两个子网络和一个决策网络部分组成, 子网络为两个完全相同的 CNN 模型; 子网络输入为 $64 \times 64 \times 3$ 的 RGB 图像, 经过两次卷积池化后展开成 $2\ 304 \times 1$ 的一维特征向量, 在全连接层压缩后成为 128×1 的输出向量; 决策层比较两个输出向量相似度, 这里以 L1 距离作为相似度评价标准, 得到一个 128×1 的距离特征向量, 这个向量反映了两个输入样本的相似程度; 但是以上结构仅能判定两个样本是否为同类样本, 需要进一步结合分类要求设计专用分类器以实现故障诊断。

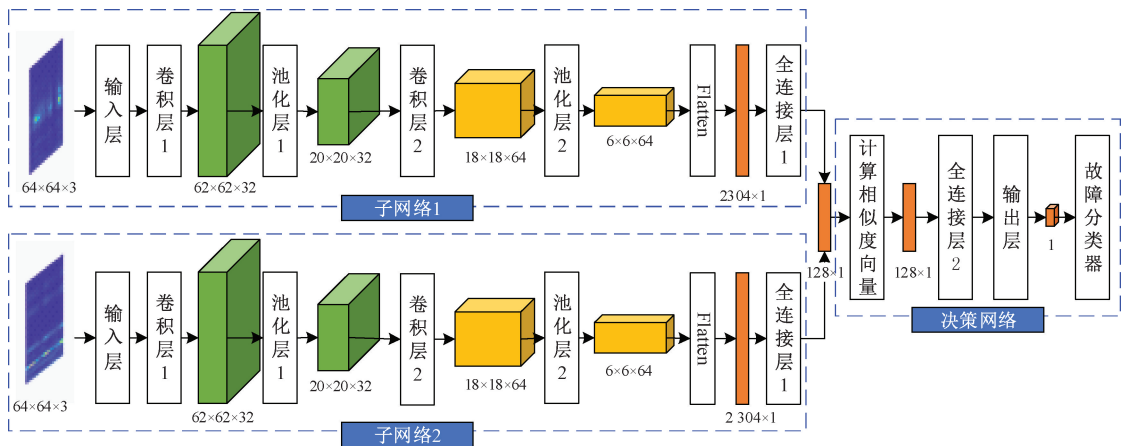


图 2 本文构建孪生网络

Fig. 2 Siamese network proposed in this article

模型训练选取了交叉熵作为损失函数, 交叉熵能够衡量同一个随机变量中的两个不同概率分布的差异程度, 表示为真实概率分布与预测概率分布之间的差异。所建立网络为每个样本对定义了范围为 0 至 1 的标签, 接近 1 则代表两个输入数据属于同一类概率较大, 反之亦然。损失函数如式(9), 交叉熵的值越小, 表示模型收敛越好。

$$\text{Loss} = -\frac{1}{N} \sum_{n=1}^N [T_n \log(P_n) - (1 - T_n) \log(1 - P_n)] \quad (9)$$

其中, N 为样本对数, x_1 和 x_2 分别代表两个子网络输出样本的特征向量; T_n 为第 n 个样本对的标签, 当 $T_n = 1$ 时 x_1 和 x_2 为同类样本, $T_n = 0$ 时为不同类样本; P_n 为经过 Sigmoid 激活函数得到的第 n 个样本对预测值, P_n 的大小由样本 x_1 和 x_2 之间的 L1 距离决定。

2.3 分类器设计

建立好模型后开始进行训练, 将孪生网络训练所

用样本对中的两张图像分别输入两个子网络, 样本对的组合方式参考 2.1 节。主要训练步骤包括: 1) 随机设置网络的初始化参数; 2) 以样本对的方式输入两个输入层后, 经过卷积和池化提取特征; 3) 计算网络在该次前馈传播过程中的目标值与实际值的差值; 4) 通过反向传播算法把网络中的误差从架构层次的最后一层依次向前推算反馈, 重新计算网络各层的误差参数后将其更新保留, 接着重复步骤 3) 和 4), 循环进行前馈、反馈的传播计算, 直到该网络的模型被收敛终止训练任务。

分类器采取了特征向量比较的方式进行判定, 其原理如图 3 所示。在训练完毕后, 将原始数据输入网络模型, 得到相应的 128×1 输出向量; 计算 n 类故障样本中每类故障的平均向量, 得到由 n 个大小为 128×1 的平均特征向量。将待检测样本输入 1 个子网络中, 依次与 n 个平均特征向量进行 L1 距离比较, 得到 n 个由 0~1 数值

组成的分类向量,其中,最大值对应的平均特征向量所属类别即为待测样本的判定类别。

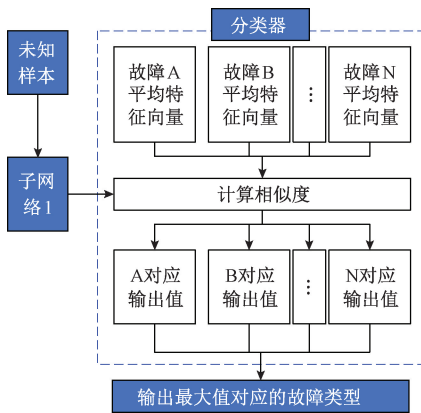


图 3 孪生网络分类器

Fig. 3 The classification for siamese network

3 实 验

3.1 建立模型

根据上述孪生网络的构成要素、设计方案和输入信息,在 Ubuntu 18.04 系统下实现程序调试运行,在 Python 3.0 环境下构建深度学习网络模型,主要参数记录在表 1 中,其中第 1~7 层表示两个子 CNN 网络的参数,第 8~10 层则是孪生网络的决策部分。

表 1 网络模型的结构参数

Table 1 Structural parameters of the siamese network

层数	模型组成	算子尺寸	输出特征	
1	Input	Input	64×64×3	
2	Conv1	Conv1	3×3	62×62×32
3	MaxPool1	MaxPool1	3×3	20×20×32
4	Conv2	Conv2	3×3	18×18×64
5	MaxPool2	MaxPool2	3×3	6×6×64
6	Flatten	Flatten	-	2 304
7	Dense1	Dense1	-	128
8	L1 距离	-	-	128
9	Dense2	-	-	128
10	Dense3	-	-	1

为了对比所提出方法的有效性,研究与经典 CNN 模型进行了对比,CNN 参数设置与子网络参数相同,为了使 CNN 具备分类功能,在其输出层末端设计了专用的故障分类器。

构建模型后开始滚动轴承故障诊断实验,一共对两组样本进行分析:第一组为西储大学滚动轴承故障数据库,第二组实验为印刷机滚动轴承故障诊断实验。以上两组实验均以轴承附近固定机架测点的振动信号为信息来源开展,在小样本条件下训练诊断模型。

3.2 滚动轴承数据库诊断

1) 实验描述

西储大学滚动轴承故障数据库中采用的故障轴承型号为 SKF6205,该数据实验环境相对单一,研究中对数据加入了随机噪声以提高故障诊断难度。从数据库选取 10 类电机驱动端故障轴承,包括健康轴承数据、滚体故障、内圈故障和外圈故障,其中每类故障包含 0.178、0.356 和 0.533 mm 三种故障尺寸,采样频率 12 kHz。

从每类故障样本中选取 60 组数据,每组数据长度 6 000 个点,一共有 10 类样本,合计 600 组数据;为了保证小样本和强噪声环境,仅选取其中 1/3 用于模型参数调节,2/3 作为未知样本进行诊断测试集,这样每类故障仅有 20 个样本,其中 16 个作为训练集,4 个作为验证集。为了验证所提出方法对噪声的抑制能力,对所选故障信号添加不同信噪比的高斯白噪声,信噪比计算如式(10)所示。

$$\text{SNR} = 10 \lg \left(\frac{P_s}{P_n} \right) \quad (10)$$

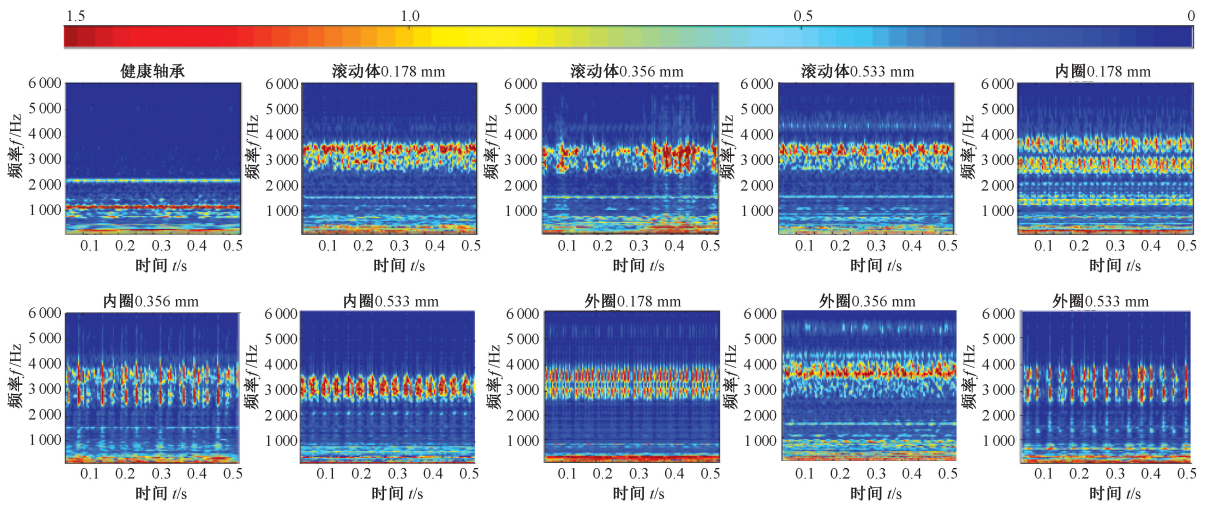
P_s 和 P_n 分别代表振动信号功率和所添加噪声功率,增加噪声由强到弱对应 5 种信噪比: -4、-2、0、2 和 4 dB。

图 4 为输入孪生网络模型原始时频图像信息,该图像由原始信号经过 CWT 分析获取,其横轴代表了不同时间尺度,纵轴代表了不同频域尺度,图像的强弱表示了该尺度条件下的小波系数。其中图 4(a)为原始故障信号,图 4(b)为增加了噪声后的图像,信噪比为 -4 dB。从图中发现:原始信号时频特性在 CWT 分析后,不同样本间差异的表现形式较为直观,对其开展识别的难度较低;添加噪声后信号图像信息复杂度大幅提升,难以直观区分,在小样本工况下进行准确识别具有相当难度。

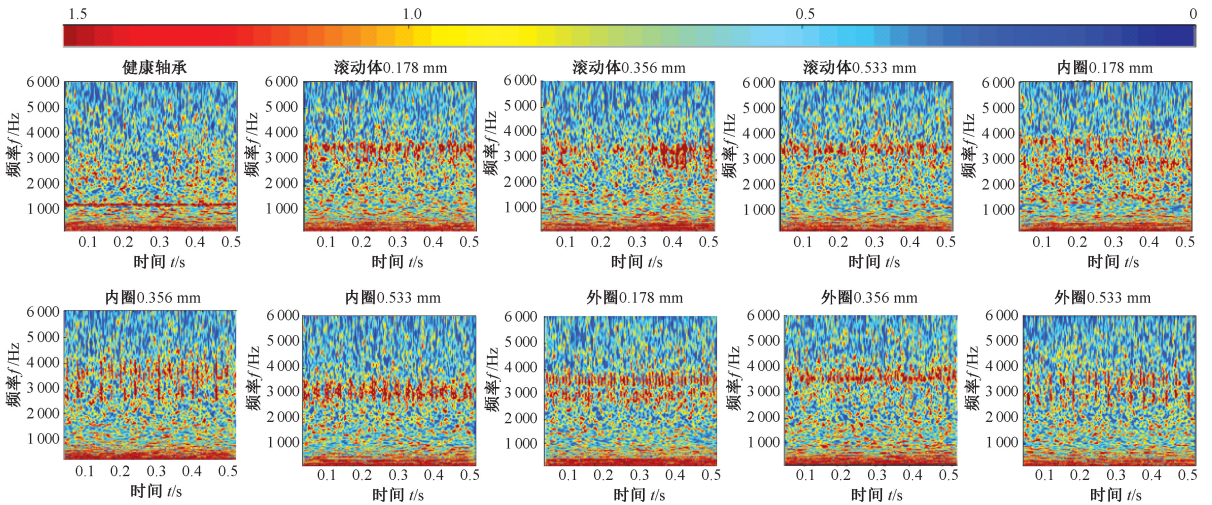
2) 模型参数调节

为了确保模型能够充分发挥性能,研究针对训练过程中的关键参数进行了调优分析,主要包括批处理大小、学习率和训练次数。

批处理大小(batch size):该参数决定了单次传递给模型用以训练的参数个数,合理设置该参数能够有效地提高训练速度,但在小样本学习中,会因为数据量较少造成训练梯度值出现较大波动。由于实验中每类故障样本仅有 16 个直接参与训练,故将 batch size 设为 16 以内,以防止参数过大导致泛化能力下降。



(a) 时频图像
(a) Time frequency images



(b) 加噪时频图像 (SNR=-4 dB)
(b) Time feueney images with noise (SNR=-4 dB)

图4 孪生网络模型的输入图像

Fig. 4 Inputting pictures for the siamese network model

研究以-4 dB 噪声样本为诊断对象,图5 描述了不同参数下的训练测试效果,当 batch size 增加到 6 以上时,数据正确率稳定于 99.5% 以上;经过对测试集验证,所训练模型在 batch size 为 8 时,达到了最高故障诊断正确率 97%,因此后续实验 3.3 节中设定参数为 8。

学习率 (learning rate): 该参数可确定每次迭代中的步长,使损失函数收敛到最小值。学习率对成功训练模型具有重要影响:学习率过低会陷入局部最优,过高则又会错过损失函数最小值从而无法收敛。

研究以-4 dB 噪声样本为诊断对象,图 6 为不同学习率下的模型训练测试效果:学习率高于 0.1 后正确率出现大幅下降;在 0.000 1~0.01 时,训练集正确率高于 99%,训练效果理想,最优故障诊断效果出现在 0.000 1~

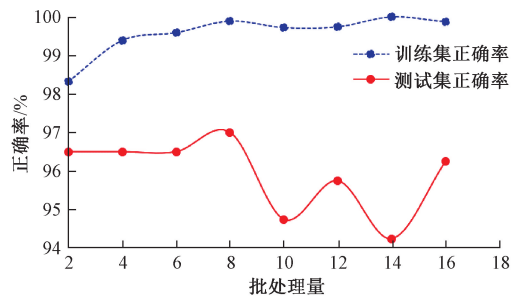


图5 批处理量对模型的影响

Fig. 5 Influence on model from batch size

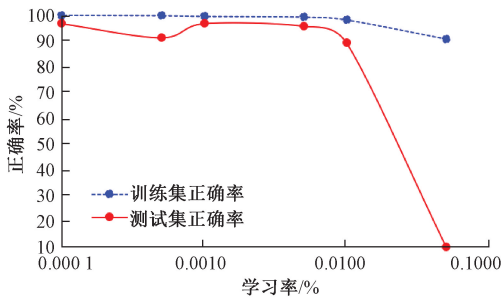


图 6 学习率对模型的影响

Fig. 6 Influence on model from learning rate

0.005 附近, 测试正确率高于 95%。综合评估后, 本文设定学习率为 0.001, 此时训练集正确率 99.56%, 测试集正确率 96.25%。

迭代次数 (epoch): 由于本次采用训练集数量较少, 因此训练次数在较小时就可以使模型完全收敛, 取得较高的训练正确率。

SN 模型与 CNN 模型训练正确率与迭代次数的关系如图 7 所示。在不同噪声情况下两种模型都能够快速收敛, 但 SN 模型在经过 2~3 次迭代后就能够完成训练并达到近 100% 准确率, CNN 需要 12~14 个周期才能实现, 反映出 SN 对小样本故障数据建模能力更强。

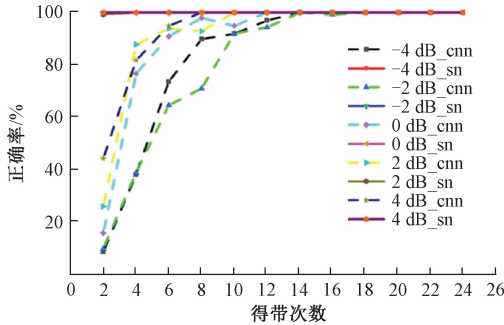


图 7 迭代次数对模型的影响

Fig. 7 Influence on model from epoch

3) 结果对比分析

在建立模型结构并调节优化参数后, 对不同信噪比下的滚动轴承故障信号分别进行建模与诊断研究, 记录测试集准确率在表 2 中。

表 2 所添加噪声的信噪比

Table 2 Diagnosis results of different SNR

SNR/dB	-4	-2	0	2	4
SN/%	96.25	98.5	99.5	99.75	99.75
CNN/%	81.49	93.06	96.11	98.89	97.78

在各类信噪比条件下 SN 模型的故障诊断正确率均处于较高水平, 说明所提出方法在小样本工况下, 仍具有

较强的噪声抑制能力; 在同样的工况下对比 CNN 模型可见: 随着噪声比例的增加, CNN 正确率表现出明显下降趋势, 诊断正确率大幅下降。

研究以 -4 dB 噪声样本为例, 进一步对混淆矩阵进行了对比分析, 如图 8 所示, 其中 b、i、o 分别代表滚动体故障、内圈故障、外圈故障, n 代表正常轴承。1、2、3 分别表示 0.178、0.356 和 0.533 mm 三种故障尺寸。可见 SN 模型正确率接近 100%, 其错误主要集中在左上方区域, 对滚动体故障程度识别略弱; CNN 正确率呈现全面下降趋势, 仅对外圈故障效果较好。

30	0	10	0	0	0	0	0	0	0	b1
0	39	1	0	0	0	0	0	0	0	b2
1	0	39	0	0	0	0	0	0	0	b3
0	0	0	40	0	0	0	0	0	0	i1
0	0	0	0	37	0	0	0	0	3	i2
0	0	0	0	0	40	0	0	0	0	i3
0	0	0	0	0	0	40	0	0	0	n
0	0	0	0	0	0	0	40	0	0	o1
0	0	0	0	0	0	0	0	40	0	o2
0	0	0	0	2	0	0	0	0	38	o3
b1	b2	b3	i1	i2	i3	n	o1	o2	o3	/

(a) 孪生网络

(a) SN

38	1	1	0	0	0	0	0	0	0	b1
0	11	8	0	0	0	4	0	0	17	b2
16	3	21	0	0	0	0	0	0	0	b3
0	0	0	40	0	0	0	0	0	0	i1
0	0	0	0	27	0	0	0	0	13	i2
0	0	0	0	0	40	0	0	0	0	i3
0	0	0	0	0	0	40	0	0	0	n
3	0	0	0	0	0	0	37	0	0	o1
0	0	0	0	0	0	0	0	40	0	o2
0	0	0	0	8	0	0	0	0	32	o3
b1	b2	b3	i1	i2	i3	n	o1	o2	o3	/

(b) 卷积神经网络

(b) CNN

图 8 混淆矩阵对比

Fig. 8 Comparison of confusion matrix

3.3 印刷机滚动轴承诊断

采用与 3.2 节中相同结构的孪生网络模型和调节参数, 对实际机械设备开展滚动轴承故障研究。

1) 实验描述

故障诊断实验针对辊子轴承开展, 印刷机中包含大量的墨辊和水辊, 其通常采用相同的轴承型号, 这些辊子轴承性能的退化及故障, 会导致印刷品整体质量下降且难以排查。本次选用了墨辊一侧的辊子轴承为对象, 利用电火花加工了保持架、内圈、外圈等 3 种故障类型进行替换实验并采集故障信号, 由于印

刷机包含关联机构众多导致运行时噪声干扰严重,增大了诊断难度。

该组实验在 GUANGHUA PZ1650 四开单色胶印机上开展实验,印刷设备运行中处于长期高速负载状态,加上各类化学溶剂、油墨、粉尘等污染,其轴承工作强度大、损耗快。实验装置如图 9 所示:实验轴承型号 NSK6001Z 滚动轴承,采集设备为西门子 LMS 振动测试系统,采集卡型号 LMS-SCM-V8,加速度传感器型号 PCB-333B30,灵敏度为 100 mV/g,测试其单向振动信号。

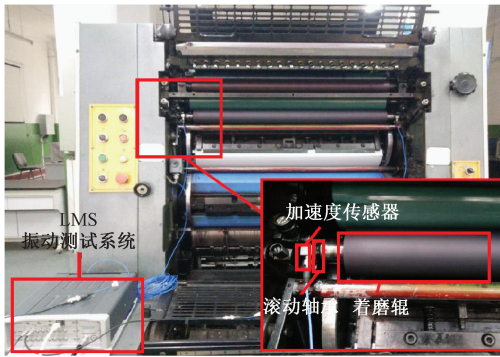


图 9 故障诊断实验装置

Fig. 9 Experimental device for fault diagnosis

所采用故障轴承如图 10 所示,实验包含 6 类样本:健康轴承、内圈故障、外圈故障、保持架故障,其中内圈和外圈各包含 0.3 mm 和 0.5 mm 两种故障尺寸;轴承转速为 4 r/s,采样频率 3 200 Hz。从每类故障样本中仅选取 60 组数据,每组数据长度 20 000 个点,一共有 6 类样本,合计 360 组数据。与第一组实验采取同样的数据选择方式,选取其中每类故障样本数量仍为 20 个,合计 120 组数据进行模型训练,其余 240 组数据进行模型测试。

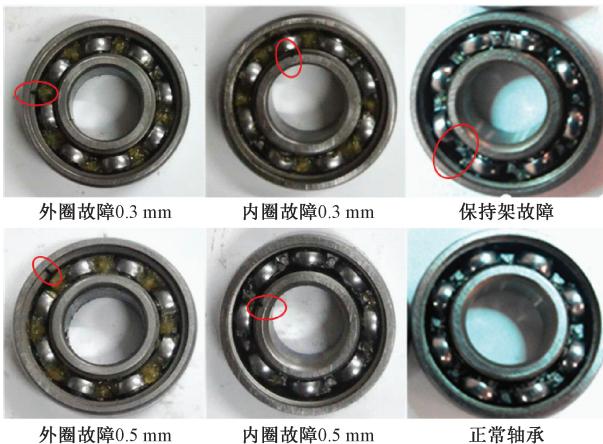


图 10 实验用故障轴承

Fig. 10 Rolling bearings with different faults in experiment

图 11 为输入孪生网络模型原始时频图像信息,由于设备运行中会引入大量倍频干扰和噪声干扰,可以发现 6 组图像中噪声干扰非常严重,其包含模式复杂,难以通过的特征或者经验进行判断,因此需要发挥深层网络模型在图像识别方面的优势。

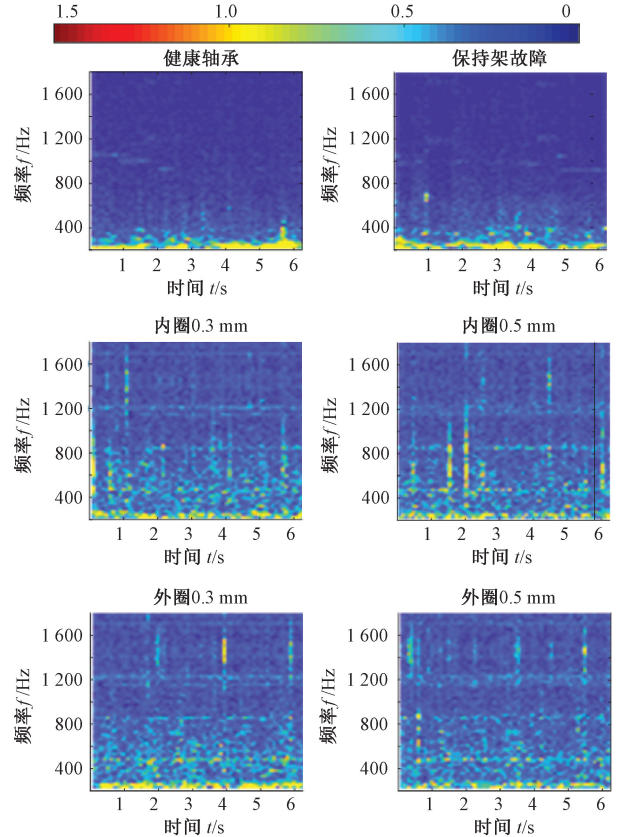


图 11 模型输入图像

Fig. 11 Inputting pictures for the siamese network model

2) 结果对比分析

采用表 1 中孪生网络模型进行训练和诊断,参数调优设定与 3.2 节一致,实验表明:所提出方法正确率达到 97.08%,成功实现了小样本条件下多种类型故障的诊断;采用表 1 中 CNN 模型对比,其准确率仅为 60.83%,无法诊断印刷机滚动轴承故障。

图 12(a)和(b)对印刷装备滚动轴承故障诊断结果进行了混淆矩阵分析:SN 模型对于各类故障识别率均处于较高水平,但会把个别内圈故障与正常样本误判为外圈故障,后续可针对性增加外圈样本进一步提高正确率;CNN 在识别保持架故障方面性能优于 SN,但对内、外圈故障和正常样本无法区分,极易把正常轴承误判为外圈故障。

40						保持架故障	真实标签
	40					内圈0.3 mm	
		37			3	内圈0.5 mm	
			38		2	健康轴承	
				40		外圈0.3 mm	
			2		38	外圈0.5 mm	
保持架故障	内圈0.3 mm	内圈0.5 mm	健康轴承	外圈0.3 mm	外圈0.5 mm	预测标签	

(a) 孪生网络
(a) SN

40						保持架故障	真实标签
	29	7		3	1	内圈0.3 mm	
	4	28	2	2	4	内圈0.5 mm	
			5	10	25	健康轴承	
		1		27	12	外圈0.3 mm	
			4	19	17	外圈0.5 mm	
保持架故障	内圈0.3 mm	内圈0.5 mm	健康轴承	外圈0.3 mm	外圈0.5 mm	预测标签	

(b) 卷积神经网络
(b) CNN

图 12 混淆矩阵对比

Fig. 12 Comparison of confusion matrix

4 结 论

本文针对小样本条件下的滚动轴承故障诊断问题开展研究,以孪生网络结构为核心架构,以连续小波分析为信号解析方法,实现了多种工况下的滚动轴承故障诊断目标,主要结论有:

1) 连续小波变换所获取频谱图像对于故障信号有着一定的表征能力,但是在小样本、强噪声的条件下,其识别难度极大,一般 CNN 模型难以有效区分。

2) 基于孪生网络架构的诊断模型则可以充分发挥 CWT 图像信息的优势,通过交叉样本对大幅度扩充样本数据集,在少数样本下完成模型训练。

3) 研究设计了专用的孪生网络模型分类器,将相似度判定问题转换为根据概率值分类问题;同时,针对 SN 模型关键参数进行了对比分析,确定了滚动轴承诊断中的调优方案。

4) 所提出方法在实验 1 中各类加噪声条件下正确率达到 95% 以上,高于 CNN 模型;在实验 2 印刷机故障诊

断中达到 97.08% 正确率,大幅高于 CNN 模型 60.83% 的正确率。

本文所提出模型通过每类故障状态的 16 个训练样本和 4 个验证样本,成功在噪声干扰条件下完成了小样本模型训练,并对不同类型、不同尺寸的多组滚动轴承故障实现了诊断。

参考文献

[1] SHAO H, JIANG H, LIN Y, et al. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders [J]. Mechanical Systems and Signal Processing, 2018, 102(1): 278-297.

[2] EREN L, INCE T, KIRANYAZ S. A generic intelligent bearing fault diagnosis system using compact adaptive 1d CNN classifier [J]. Journal of Signal Processing Systems, 2019, 91: 179-189.

[3] 张西宁, 郭清林, 刘书语. 深度学习技术及其故障诊断应用分析与展望 [J]. 西安交通大学学报, 2020, 54(12): 1-13.

ZHANG X N, GUO Q L, LIU SH Y, et al. Analysis and prospect of deep learning technology and its fault diagnosis application [J]. Journal of Xi'an Jiaotong University, 2020, 54(12): 1-13.

[4] 曾大懿, 蒋雨良, 邹益胜, 等. 一种新的轴承寿命预测特征评价指标构建与验证 [J]. 振动与冲击, 2021, 40(21): 18-27.

ZENG D Y, JIANG Y L, ZOU Y SH, et al. Construction and verification of a new evaluation index for bearing life prediction characteristics [J]. Vibration and Shock, 2018, 54(5): 94-104.

[5] 雷亚国, 贾峰, 孔德同, 等. 大数据下机械智能故障诊断的机遇与挑战 [J]. 机械工程学报, 2018, 54(5): 94-104.

LEI Y G, JIA F, KONG D T, et al. Opportunities and challenges of mechanical intelligent fault diagnosis in big data era [J]. Journal of Mechanical Engineering, 2018, 54(5): 94-104.

[6] HU T, TANG T, LIN R, et al. A simple data augmentation algorithm and a self-adaptive convolutional architecture for few-shot fault diagnosis under different working conditions [J]. Measurement, 2020, 156: 107539.

[7] ZHOU F, YANG S, FUJITA H, et al. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data [J]. Knowledge-Based Systems, 2020, 187: 104837.

[8] CHENG F, ZHANG J, WEN C, et al. Large cost-sensitive margin distribution machine for imbalanced data

- classification[J]. *Neurocomputing*, 2017, 224: 45-57.
- [9] FAN Q, WANG Z, LI D, et al. Entropy-based fuzzy support vector machine for imbalanced datasets [J]. *Knowledge-Based Systems*, 2017, 115:87-99.
- [10] 陈超, 沈飞, 严如强. 改进 LSSVM 迁移学习方法的轴承故障诊断[J]. *仪器仪表学报*, 2017, 38(1): 33-40.
- CHEN CH, SHEN F, YAN R Q, et al. Enhanced least squares support vector machine-based transfer learning strategy for bearing fault diagnosis[J]. *Chinese Journal of Scientific Instrument*, 2017, 38(1):33-40.
- [11] MAO W, LIU Y, DING L, et al. Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study [J]. *IEEE Access*, 2019, 7: 9515-9530.
- [12] 戴俊, 王俊, 朱忠奎, 等. 基于生成对抗网络和自动编码器的机械系统异常检测[J]. *仪器仪表学报*, 2019, 40(9): 16-26.
- DAI J, WANG J, ZHU ZH G, et al. Anomaly detection of mechanical systems based on generative adversarial network and auto-encoder [J]. *Chinese Journal of Scientific Instrument*, 2019, 40(9): 16-26.
- [13] ZHOU F, YANG S, FUJITA H, et al. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data [J]. *Knowledge-Based Systems*, 2020, 187: 104837.
- [14] ZHANG X, HE C, LU Y, et al. Fault diagnosis for small samples based on attention mechanism [J]. *Measurement*, 2022, 187: 110242.
- [15] 许子非, 金江涛, 李春. 基于多尺度卷积神经网络的滚动轴承故障诊断方法[J]. *振动与冲击*, 2021, 40(18): 212-220.
- XU Z F, JIN J T, LI CH, et al. New method for the fault diagnosis of rolling bearings based on a multiscale convolutional neural network[J]. *Vibration and Shock*, 2021, 40(18): 212-220.
- [16] 宫文峰, 陈辉, 张泽辉, 等. 基于改进卷积神经网络的滚动轴承智能故障诊断研究[J]. *振动工程学报*, 2020, 33(2): 400-413.
- GONG W F, CHEN H, ZHANG Z H, et al. Intelligent fault diagnosis for rolling bearing based on improved convolutional neural network [J]. *Journal of Vibration Engineering*, 2020, 33(2): 400-413.
- [17] JIA F, LEI Y, LU N, et al. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization [J]. *Mechanical Systems and Signal Processing*, 2018, 110: 349-367.
- [18] 朱会杰, 王新晴, 芮挺, 等. 基于平移不变 CNN 的机械故障诊断研究[J]. *振动与冲击*, 2019, 38(5): 45-52.
- ZHU H J, WANG X Q, RUI T, et al. Machinery fault diagnosis based on shift invariant CNN [J]. *Vibration and Shock*, 2019, 38(5): 45-52.
- [19] SEUNGMIN H, JONGPIL J. A weighted CNN ensemble model with small amount of data for bearing fault diagnosis[J]. *Procedia Computer Science*, 2020, 175: 88-95.
- [20] 张西宁, 余迪, 刘书语. 基于迁移学习的小样本轴承故障诊断方法研究[J]. *西安交通大学学报*, 2021, 55(10): 30-37.
- ZHANG X N, YU D, LIU SH Y, et al. Fault diagnosis method for small sample bearing based on transfer learning[J]. *Journal of Xi'an Jiaotong University*, 2021, 55(10): 30-37.
- [21] 刘飞, 陈仁文, 邢凯玲, 等. 基于迁移学习与深度残差网络的滚动轴承快速故障诊断算法[J]. *振动与冲击*, 2022, 41(3): 154-164.
- LIU F, CHEN R W, XING K L, et al. Fast fault diagnosis algorithm for rolling bearing based on transfer learning and deep residual network [J]. *Vibration and Shock*, 2022, 41(3): 154-164.
- [22] 何强, 唐向红, 李传江, 等. 负载不平衡下小样本数据的轴承故障诊断 [J]. *中国机械工程*, 2021, 32(10): 1164-1171.
- HE Q, TANG X H, LI CH J, et al. Bearing fault diagnosis method based on small sample data under unbalanced loads [J]. *China Mechanical Engineering*, 2021, 32(10): 1164-1171.
- [23] ZOU L, LI Y, XU F. An adversarial denoising convolutional neural network for fault diagnosis of rotating machinery under noisy environment and limited sample size case [J]. *Neurocomputing*, 2020, 407(24): 105-120.
- [24] ZHAO R, AN L, SONG D, et al. Detection of chlorophyll fluorescence parameters of potato leaves based on continuous wavelet transform and spectral analysis[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2021, 259(5):119768.
- [25] MOORTEL I D, MUNDAY S A, HOOD A W. Wavelet analysis: The effect of varying basic wavelet parameters[J]. *Solar Physics*, 2004, 222: 203-228.
- [26] 鄢仁武, 林穿, 高硕勋. 基于小波时频图和卷积神经网络的断路器故障诊断分析[J]. *振动与冲击*, 2020, 39(10): 198-205.
- YAN R W, LIN C, GAO SH X. Fault diagnosis and

analysis of circuit breaker based on wavelet time-frequency representations and convolution neural network[J]. *Vibration and Shock*, 2020, 39(10): 198-205.

- [27] 赵小强, 罗维兰. 改进卷积 Lenet-5 神经网络的轴承故障诊断方法[J]. *电子测量与仪器学报*, 2022, 36(6): 113-125.

ZHAO X Q, LUO W L. Improved convolutional Lenet-5 neural network for bearing fault diagnosis[J]. *Journal of Electronic Measurement and Instrumentation*, 2022, 36(6): 113-125.

- [28] 刘文彪, 段礼祥, 耿帆, 等. 基于 CNN-BLSTM 网络的轴承性能退化预测[J]. *电子测量与仪器学报*, 2021, 35(2): 80-86.

LIU W B, DUAN L X, GENG F, et al. Bearing performance degradation prognosis based on CNN-BLSTM network[J]. *Journal of Electronic Measurement and Instrumentation*, 2021, 35(2): 80-86.

- [29] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification[C]. *IEEE Computer Society Conference on CVPR*, 2005, 1: 539-546.

作者简介



徐卓飞(通信作者), 2009 年于西安理工大学获得学士学位, 2012 年于西安理工大学获得硕士学位, 2015 年于西安理工大学获得博士学位, 现为西安理工大学讲师, 主要研究方向为机械设备状态监测与故障诊断, 数字信号处理及模式识别。

E-mail: xuzhuofei@xaut.edu.cn.

Xu Zhuofei (Corresponding author) received his B. Sc. and M. Sc. degrees both from Xi'an University of Technology in 2009 and 2012, and received his Ph. D. degree from Xi'an University of Technology in 2015. He is currently a lecturer at Xi'an University of Technology. His main research interests include condition monitoring and fault diagnosis of mechanical systems, vibration signal analysis & processing and pattern recognition.



李旭东, 2021 年于西安理工大学获得学士学位, 现为西安理工大学硕士研究生, 主要研究方向为机械设备状态检测与故障诊断、深度学习。

E-mail: bianjie2985@163.com。

Li Xudong received his B. Sc. degree from Xi'an University of Technology in 2021. He is currently a master student at Xi'an University of Technology. His main research interests include condition monitoring and fault diagnosis of mechanical systems, signal processing and deep learning.



张武, 2006 年于湖北文理学院获得学士学位, 2009 年于西安理工大学获得硕士学位, 2012 年于西安理工大学获得博士学位, 现为西安科技大学副教授、硕士生导师, 主要研究方向为车辆传动设计与理论分析、机械系统微动磨损分析。

E-mail: zw818@xust.edu.cn.

Zhang Wu received his B. Sc. degree from Hubei University of Arts and Science in 2006, and received his M. Sc. and Ph. D. degrees both from Xi'an University of Technology in 2009 and 2012, respectively. He is currently an associate professor and a master advisor at Xi'an University of Science & Technology. His main research interests include design and theoretical analysis of vehicle transmission, fretting wear analysis in mechanical system.