

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

张龙, 肖逸文, 周神赐, 王朝兵, 柳和生, 黄聪, 钱童帅

(华东交通大学机电与车辆工程学院, 330013, 南昌)

摘要: 针对深度学习模型决策机理和分类依据无法得知以及 Morlet、Laplace 等小波不能很好地匹配轴承真实故障脉冲响应的问题, 提出一种可解释卷积神经网络模型——Bdw-ResNet。首先, 设计了一种 sigmoid 激活函数加权的双阻尼小波, 并构建小波卷积层, 同时嵌入即插即用的轻量级的局部注意力机制。然后, 结合现有的残差神经网络, 建立了既可物理解释又与神经网络兼容的 Bdw-ResNet 模型以及故障诊断的完整流程, 并在南昌铁路局机车轴承数据集和青岛四方轮对轴承两个数据集上进行验证。最后, 从先验赋能和归因解释两个方面阐述了 Bdw-ResNet 模型及其组件的映射关系, 并进行了可解释性分析。结果表明: 所提方法在两个数据集上分别取得了 98.73%、99.46% 的识别准确率, 与基准模型和其他的组合相比性能提升明显, 为可解释性深度学习的故障诊断提供了一种解决思路。

关键词: 深度学习; 双阻尼小波; 注意力机制; 故障诊断; 可解释性

中图分类号: TH17 **文献标志码:** A

DOI: 10.7652/xjtuxb202503019 **文章编号:** 0253-987X(2025)03-0210-12

Application of Interpretable Convolutional Neural Network Enabled by Bi-Damped Wavelets in Bearing Fault Diagnosis

ZHANG Long, XIAO Yiwu, ZHOU Shenci, WANG Chaobing,

LIU Hesheng, HUANG Cong, QIAN Tongshuai

(School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: To address the issues of opacity in decision-making mechanisms and classification criteria in deep learning models, as well as the mismatch between wavelets such as Morlet and Laplace and the real fault impulse response of bearings, an interpretable convolutional neural network model, Bdw-ResNet, is proposed. Firstly, a bi-damped wavelet, weighted by sigmoid activation function, is designed, and a wavelet convolution layer is constructed, incorporating a plug-and-play lightweight local attention mechanism. Then, by integrating existing residual neural networks, the Bdw-ResNet model is established, which is both physically interpretable and compatible with neural networks, along with a comprehensive fault diagnosis process. The model is validated using the Nanchang Railway Bureau locomotive bearing dataset and the Qingdao Sifang wheelset bearing dataset. Finally, the mapping relationships of the Bdw-ResNet model and its components are explored, discussing priori empowerment and attributional explanation, followed by an interpretability analysis. The results show that the proposed method achieves recognition accuracies of 98.73% and 99.46%

on the two datasets, respectively, representing a significant performance improvement over the benchmark model and other combinations. It provides a viable solution for interpretable deep learning in fault diagnosis.

Keywords: deep learning; bi-damped wavelets; attention mechanisms; fault diagnosis; explainability

随着我国轨道交通迅速发展,车辆运营速度和铁路运营里程不断增加。截至 2023 年底,铁路营业里程达到 15.9 万 km,其中高铁达到 4.5 万 km^[1]。高速发展的交通事业给运维检修带来巨大的挑战。近年来,深度学习凭借强大的非线性特征提取能力、深层数据挖掘能力以及良好的泛化能力^[2-3],在图像识别、自然语言处理、数据预测等领域获得了成功应用,并被逐步引入到机械设备故障诊断领域^[4-6]。

尽管深度学习方法已经被广泛应用于旋转机械的故障诊断中,但其采用端到端的方式,将特征提取和模式识别融合在一个过程中,属于“黑盒”模型,可信度与可解释性较差^[7]。一方面,对于开发者和用户来说,不易理解模型诊断的逻辑和依据,从而很难个性化定制满足一定需求的深度学习模型;另一方面,实际应用场景往往远比实验室环境复杂得多且难以预料,缺乏可解释性和潜在不稳定性的黑箱模型存在误判、漏判风险,限制了其在高可靠性如航空航天、核电等工业场景的进一步应用^[8]。

小波变换作为一种先进的信号处理方法,在数据压缩、信号降噪等领域发展较为成熟且有深厚的数学理论支撑^[9-10]。考虑到卷积神经网络(CNN)的卷积层和小波变换都是做内积操作,具有一定的共性,很多学者将小波变换等信号处理先验知识嵌入到神经网络中,设计具有明确物理意义的小波卷积层去替代 CNN 的第一个卷积层。Alekseev 等^[11]提出一种名为 GaborNet 的可解释网络模型,Gabor 小波函数作为其第一层的卷积核,可以从图像中提取空间频率特征。Liao 等^[12]基于 Daubechie 小波构造了一个包含一系列小波结构的卷积核,利用小波核来挖掘信号的时频域特征,从而获得有效的故障特征。Li 等^[13]将小波基的特征提取能力和卷积核的学习能力相结合,提出了一种可解释的 WPConvNet 小波包核约束网络,用于噪声鲁棒故障诊断。然而,轴承真实振动信号中的故障脉冲响应往往呈现双边非对称衰减的波形,上述小波并不能很好地匹配轴承真实故障脉冲响应,影响了小波核网络的性能。

在深度 CNN 领域,已有学者提出将注意力机制作为模拟人类认知行为的手段,这为可解释性学习提供了另一条全新的道路^[14-15]。这种机制使神

经网络能够专注于相关信息,而忽略不重要的细节,最终增强网络的学习能力。Hu 等^[16]开创性地提出压缩和激励(SE)注意力机制,其利用全局池化将空间维度压缩为通道维度,从而促进增强的特征学习。然而,SE 降维一定程度会影响通道之间的依赖关系,并增加计算复杂度。为了克服性能和复杂性无法平衡的问题,Wang 等^[17]提出了一种高效通道注意力(ECA)模块,用一维卷积实现了局部跨通道交互,该模块避免了降维,有效捕获了跨通道交互的信息。但是,ECA 在处理全局上下序列依赖性和通道空间关系方面存在一定的限制,在面对一些复杂序列数据或特定任务时表现乏力。

基于上述原因,本文设计了一种能更好地匹配实际故障冲击的双阻尼小波,并构建了一个小波卷积层来取代残差神经网络(ResNet)的第一个卷积层。将双阻尼小波和基准 ResNet 组合得到 Bdw-ResNet,使 CNN 第一层的输出具有可解释性。同时,引入即插即用的局部注意力(LA)机制,不需要减少通道维度并且具有轻量级的网络架构。Bdw-ResNet 在两个实验中均表现出优秀的故障诊断识别性能,并且通过先验赋能和归因解释两方面进行了可解释性分析。

1 方法介绍

1.1 双阻尼小波

轴承故障引起的脉冲响应通常被认为是单边衰减的振荡波形,而 Laplace 小波同样呈现单边振荡衰减波形,因此常被用于基于稀疏表征原理的轴承故障特征提取中。不同于 Laplace 小波等单边振荡衰减的波形,Morlet 小波呈现双边对称衰减的形状,被认为与齿轮故障引起的脉冲响应相似,因此常被用作齿轮故障特征提取的基函数^[18]。图 1 为常用的 3 种小波函数,其中 Laplace 和 Morlet 小波的时域数学表达式分别为

$$\psi(f,\xi,\tau,t) = A\exp\left(2\pi f(t-\tau) \frac{-\xi}{\sqrt{1-\xi^2}}\right)\sin[2\pi f(t-\tau)] \quad (1)$$

$$\phi(t) = Ce^{-t/2}\cos(5t) \quad (2)$$

式中: f 为采样频率; t 为时间; A 为 Laplace 小波归一化常数; ξ 为阻尼比; τ 为时间常数; C 为 Morlet 小

波归一化常数,计算公式为

$$C = \frac{\pi^{1/4}}{\sqrt{s + 0.01}} \quad (3)$$

其中 s 为尺度因子。

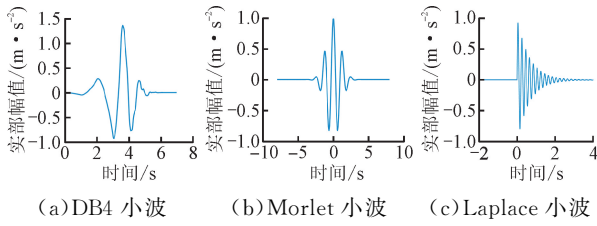


图1 3种常见小波的时域图

Fig. 1 Time-domain plots of three common wavelets

然而,由于故障机理、传递路径以及传感器的影响,轴承真实振动信号中的故障脉冲响应往往呈现双边非对称衰减的波形。图2所示为某真实数据集的内圈故障和外圈故障信号时域波形,可以发现,上述单阻尼控制的单边振荡衰减小波和双边对称衰减小波均不能很好地匹配轴承真实故障脉冲响应。

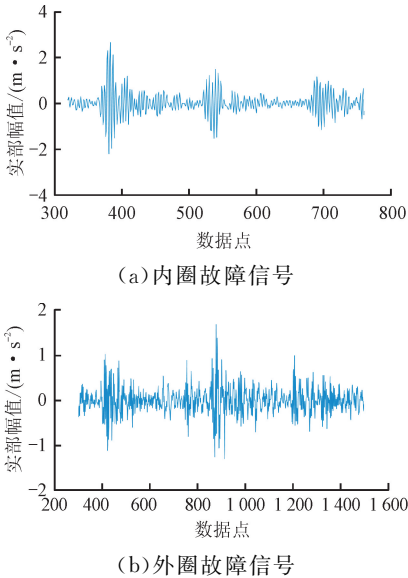


图2 故障信号时域波形

Fig. 2 Time-domain waveform of the fault signal

根据轴承故障冲击响应的特点以及现有常用小波函数的限制,本文在 Laplace 小波的基础上增加一个阻尼比^[19],提出一种双边非对称指数衰减的双阻尼小波。同时,为防止梯度爆炸,使小波卷积层与神经网络架构具有更好的兼容性,采用 sigmoid 激活函数进行加权。双阻尼小波的数学表达式如下

$$\text{sigmoid} \left[K_{\text{imp}}(t - \tau, 0) \exp \left[\omega(t - \tau) \frac{-\xi}{\sqrt{1 - \xi^2}} \right] \cdot \right.$$

$$\left. \cos(\omega(t - \tau)) - K_{\text{imp}}(\tau - t, 0) \exp \left[\omega(t - \tau) \cdot \frac{\zeta}{\sqrt{1 - \zeta^2}} \right] \cos(\omega(t - \tau)) \right] \quad (4)$$

$$K_{\text{imp}}(t, 0) = \begin{cases} 1, & t > 0 \\ 0, & \text{其他} \end{cases} \quad (5)$$

式中: $\omega = 2\pi f$, f 表示采样频率; ξ 和 ζ 为小波阻尼比。双阻尼比 ξ 和 ζ 可取不同值,从而控制小波左右两侧衰减快慢,以匹配轴承真实故障脉冲响应。图3从左到右依次给出了 φ_1 (500, 0.5, 0.2, 0.01)、 φ_2 (1 000, 0.15, 0.15, 0.03) 和 φ_3 (1 000, 0.2, 0.5, 0.05) 这3个模态参数的双阻尼小波时域波形。

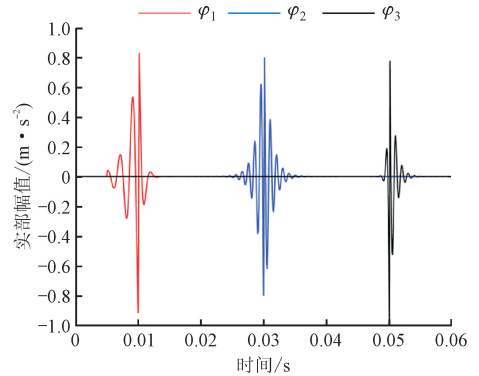


图3 不同参数组合下的双阻尼小波时域波形

Fig. 3 Bi-damped wavelets with different parameter combination

1.2 双阻尼小波卷积层

卷积神经网络的第一层非常重要,主要用来处理输入的高维数据,提取的特征映射会影响整个模型。通常,标准 CNN 的第一层会在输入信号和一组随机初始化的卷积核之间进行内积,但是这种随机初始化的卷积核不能从一维振动信号中提取出与故障冲击相关的冲击分量,从而使得到的特征映射失去可解释性。

在时域内^[20],小波字典 $\{\psi_{u,s}(t)\}$ 可以通过展开尺度因子 s , 并平移时移因子 u 依次得到,表达式如下

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi \left(\frac{t - u}{s} \right) \quad (6)$$

式中: $\psi(\cdot)$ 为具有时域表达式的小波基函数; s 为与频率成反比的尺度因子,主要控制小波函数的伸缩; u 为时移因子,主要控制小波的平移。

考虑到卷积操作与小波变换的相似性,基于双阻尼小波构造小波卷积层,用来取代第一个卷积层,将输入信号与一组参数化的小波字典进行卷积,卷

积核大小和输出通道数与参数 s, u, t 相关联,使卷积参数具有一定的小波物理意义,映射过程如图4所示。双阻尼小波卷积层定义为

$$y = \psi_{u,s}(t) * x \quad (7)$$

式中: $*$ 表示卷积运算; x 为输入信号。

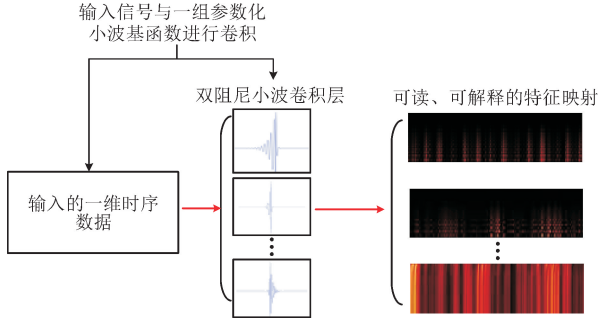


图4 双阻尼小波卷积层的映射过程

Fig. 4 Mapping of Bi-damped wavelet convolutional layer

在反向传播(BP)过程中^[21],双阻尼小波卷积层作为Bdw-ResNet的第一层,只需要更新 s 和 u 两个参数。更新过程如下

$$\begin{cases} \delta_{s_k} = \frac{\partial L}{\partial s_k} = \frac{\partial L}{\partial y_k} \frac{\partial y_k}{\partial \psi_{u,s}^k} \frac{\partial \psi_{u,s}^k}{\partial s_k} \\ \delta_{u_k} = \frac{\partial L}{\partial u_k} = \frac{\partial L}{\partial y_k} \frac{\partial y_k}{\partial \psi_{u,s}^k} \frac{\partial \psi_{u,s}^k}{\partial u_k} \\ u_k = u_k - \eta \delta_{u_k} \\ s_k = s_k - \eta \delta_{s_k} \end{cases} \quad (8)$$

式中: u_k 和 s_k 分别为第 k 个通道的时移因子和尺度因子; δ_{s_k} 和 δ_{u_k} 为对应的梯度; ∂ 为导数算子; L 为交叉熵损失函数; y 表示双阻尼小波卷积层输出的结果; $\psi_{u,s}^k$ 为小波卷积层的第 k 个通道的小波核; η 为学习率。 u 和 s 通过减去链式求导法则得到的梯度与学习率的乘积,从而达到更新参数的目的。

1.3 改进的局部注意力机制

注意力机制作为一种权重参数的分配机制,目的是协助模型捕捉重要信息,从而有效增强神经网络的性能。然而,现有的一些用于处理一维数据的注意力机制,都不能很好的满足特定的任务需求。SE 只能处理通道数较少的情况并且计算量很大,ECA 在面对全局序列依赖性和通道空间关系方面存在一定的限制。根据文献[22-23]所述,将BN层放置在处理序列化数据的网络中并不是最优的,特别是注意力机制模块中。另外,当小批量过小时,由BN计算的平均值和方差可能不足以代表整个数据集,从而潜在地损害模型整体性能。

本文基于坐标注意力(CA)机制的设计思想,使用组归一化(GN)作为BN的替代,并采用核尺寸为 7×1 的一维卷积,提出了一种适用于一维数据并兼具轻量化特性的局部注意力(LA)机制,精确地捕获感兴趣区域的位置,同时保持了输入特征映射通道的维度,示意图如图5所示。

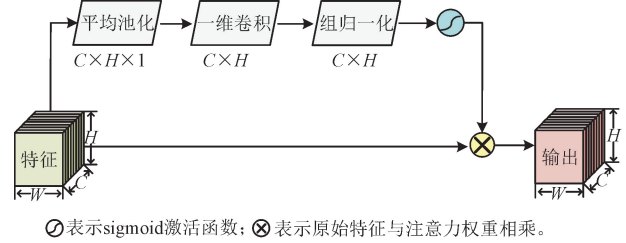


图5 局部注意力机制示意图

Fig. 5 Schematic diagram of the local attention mechanism

以下是局部注意力机制的完整推导过程,其中 C, H, W, G 分别表示通道数(即卷积核个数)、高度、宽度、组数。首先,为了应用条带池化,沿水平方向对每个通道进行平均池化,这将产生高度为 h 的第 c 个通道的输出表示,数学表达式如下

$$z_c^h(h) = \frac{1}{H} \sum_{0 \leq i < H} x_c(h, i) \quad (9)$$

式中: x_c 为原始特征。

由式(9)得到的输出 z_h 捕获了全局感受野。为了有效利用这些特性,应用核大小为 7×1 的一维卷积来增强位置信息。随后,采用GN对增强的位置信息进行处理,通过将通道方向分组,然后每个组内做归一化,计算 $(C//G) * H * W$ 的均值,来学习每通道的线性变换,以补偿可能失去的表示能力,进而得到位置注意力在水平方向的权重,表达式为

$$y^h = \sigma[G_n(F_h(z_h)))] \quad (10)$$

式中: σ 为非线性激活函数; F_h 为一维卷积过程; G_n 表示组归一化。将原始特征与局部注意力机制权重相乘,得到加权后的输出 Y

$$Y = x_c y^h \quad (11)$$

2 基于Bdw-ResNet模型的故障诊断

2.1 模型架构

将sigmoid激活函数加权并具有内在可解释性的双阻尼小波卷积层作为预处理层,同时引入即插即用的局部注意力机制模块,与现有的残差神经网络(ResNet18)相结合,得到的卷积神经网络称为Bdw-ResNet,模型的架构如图6所示。

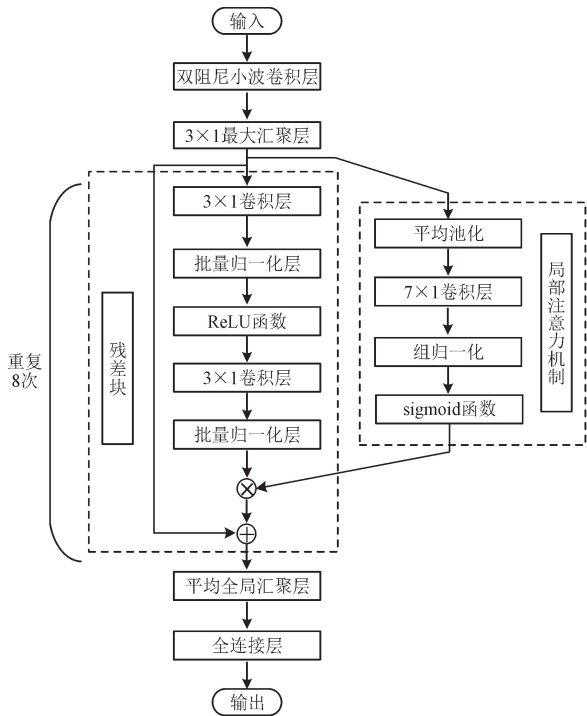


图 6 Bdw-ResNet 模型架构
Fig. 6 Bdw-ResNet model architecture

2.2 基于 Bdw-ResNet 模型的故障诊断

将嵌入双阻尼小波卷积层和局部注意力模块的 Bdw-ResNet 模型应用于滚动轴承故障诊断中, 流程如图 7 所示。

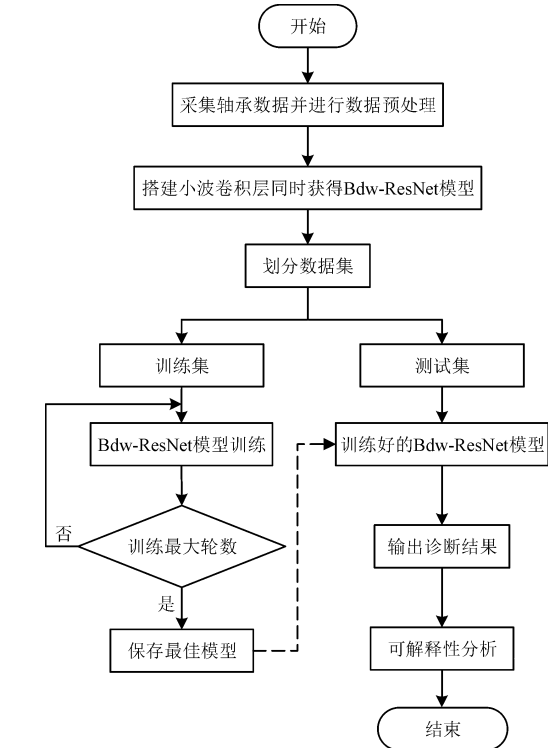


图 7 基于 Bdw-ResNet 的故障诊断流程图
Fig. 7 Fault diagnosis flowchart based on Bdw-ResNet

首先,利用安装在机械设备上的加速度传感器采集振动信号,并对数据进行归一化、降噪处理。然后,在时域内通过完全无重叠方式滑动窗口,对振动信号进行切割,从而生成一定数量的样本,作为后续网络模型的训练和测试。之后,利用 sigmoid 激活函数加权的双阻尼小波构建小波卷积层,替换基准 ResNet18 模型的首层,同时在残差块中插入局部注意力机制模块,得到 Bdw-ResNet。设置双阻尼小波、优化器、批次数和学习率等网络参数并通过反向传播进行迭代训练从而获得最佳的参数组合,通过测试集数据验证模型的有效性和优越性。最后,从热力谱图和累计频带图等角度进行可解释性分析。

3 实验验证

本节通过南昌铁路局机务段的 JL-501 机车轴承数据集和青岛四方轮对轴承数据集,对模型的有效性进行验证。同时,通过几组消融实验,对比体现本文方法的优越性。实验运行在 PyTorch 深度学习框架并在 Cuda11.8 上进行加速运算,硬件配置为 Intel 13600k CPU,Nvidia 4070 GPU 显卡。其他训练参数设置如下:①数据预处理,设置 mean-std 归一化方法,损失函数为交叉熵损失函数;②使用 Adam 优化器,默认学习率为 0.001;每训练 9 代,用乘以 0.9 的衰减系数来调整学习率;③为避免实验偶然性,每组实验进行 10 次取平均值。保留最佳组实验参数和模型,作为后期工程应用的预训练模型。

3.1 案例 1:南昌铁路局机车轴承数据集

案例 1 是对来自南昌铁路局机务段的 JL-501 型机车轴承数据集进行试验,轴承试验台如图 8 所示,主要由传感器、液压系统、主轴箱、采集卡、待测轴承、控制台等组成。电源为三相,四线,380 V, 50 Hz,采样频率为 20 kHz,电机输出功率为 5 kW。试验待测的机车轴承为内径 160 mm、外径 290 mm 的 NJ2232WB 型圆柱滚子轴承,并且可以根据需要采集 A、B、C 多个传感器信号。

如表 1 所示,本试验使用了 6 种故障状态和 1 种正常状态总共 7 种类型,且来自试验台正上方传感器 B 收集的机车轴承数据。为了防止信号重叠影响模型测试的性能,通过滑移时域信号的窗并完全无重叠切割的方式获得训练和测试的样本,每组样本数为 390,并按顺序分别编号为 L0~L6。

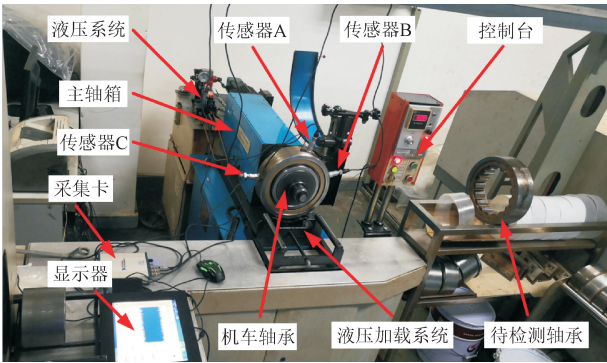


图 8 JL-501 机车轴承故障诊断试验台

Fig. 8 JL-501 locomotive bearing fault diagnosis test bench

表 1 机车轴承试验样本构造

Table 1 Structure of locomotive bearing test sample			
故障类型	故障深度/mm	样本数	标签
内圈故障	4	390	L0
	7	390	L1
	19	390	L2
外圈故障	7	390	L3
	19	390	L4
滚动体故障	4	390	L5
正常		390	L6

3.1.1 基于 Bdw-ResNet 模型的实验结果

图 9、图 10 所示为将表 1 中的机车轴承数据输入本文所提出的 Bdw-ResNet 模型中,经过 100 轮迭代训练得到的结果。由图 9 可知,模型在前 20 代训练速度较快,损失值减小至 0.2 附近;在第 58 代识别准确率最高,达到 98.73%。之后,准确率和 Loss 曲线均放缓,模型逐渐达到收敛状态,说明本文方法在本次实验中取得了满意的分类效果。

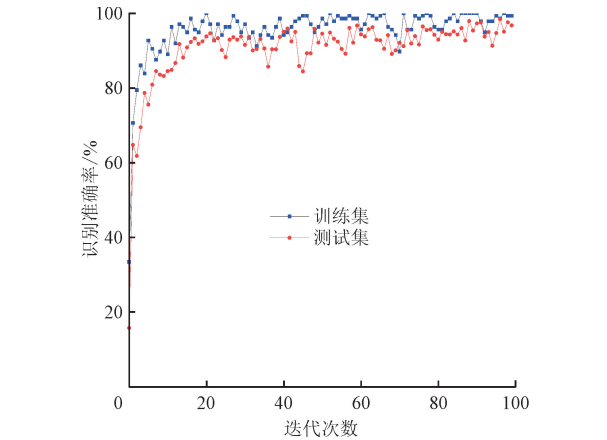


图 9 案例 1 识别准确率变化曲线

Fig. 9 Recognition accuracy curve of case 1

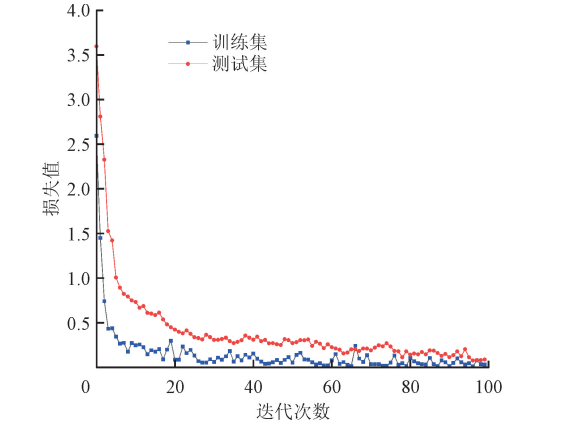


图 10 Loss 函数变化曲线

Fig. 10 Curve of loss function

图 11 为使用混淆矩阵将测试集分类的可视化结果。使用表 1 实验数据进行了 49 次重复实验,总共 2 730 个样本。实验结果表明,有 1 个 4 mm 内圈故障的样本被误判为 7 mm 内圈故障和 19 mm 内圈故障,另有一个 19 mm 外圈故障样本被误判为 7 mm 外圈故障。考虑到所制备的样本量比较小,存在部分类型故障样本特征不明显,与其他样本故障特征相似的情况,从而造成了误判的发生。但实验结果仍在可控范围之内。此外,在其他诊断的类别中,也均取得了较高的识别准确率。

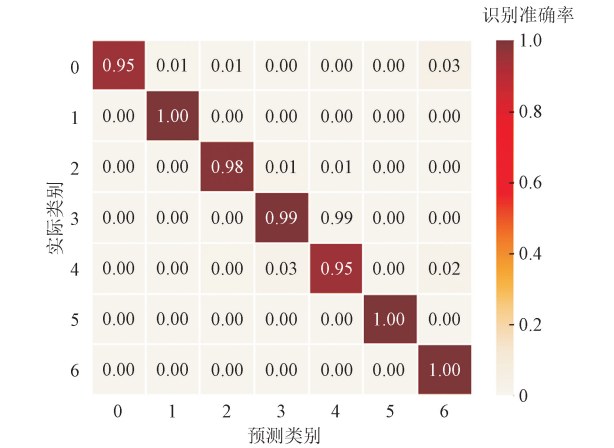


图 11 案例 1 混淆矩阵

Fig. 11 Case 1 confusion matrix

为进一步展示模型的特征提取能力和分类效果,使用 t 分布-随机邻近嵌入 (t -SNE) 算法进行特征可视化,如图 12 所示。由图可知,训练之后,模型自适应特征提取所获得的特征经过降维,各种故障类型没有重叠部分,区分度明显、清晰,再次直观地展示了本文所提方法的有效性和较好的分类效果。

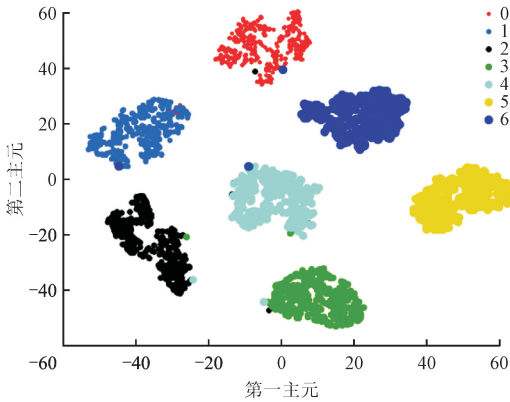


图 12 案例 1 特征降维图

Fig. 12 Case 1 dimensionality reduction of features

3.1.2 对比实验

为了验证 Bdw-ResNet 的模型性能以及本文所提方法的优越性,进一步使用未经任何处理的残差神经网络基准骨架 Backbone、Morlet 小波、Laplace 小波进行实验,对比本文所提双阻尼小波相对其他小波函数的优势。设置 Laplace 小波参数 A 为 $1/e$, ξ 为 0.3 , τ 为 0.1 。Morlet 小波参数 C 与尺度因子 s 有关,随着 s 初始化自动设置。

在每一组实验中再分别使用当前主流的适用于一维时序数据的 SE、ECA 注意力机制进行实验,对比本文所提的 LA 相对其他注意力机制的优势,共进行 16 组实验。为了避免干扰,取 10 次诊断的平均结果作为最终的准确率,结果如图 13、表 2 所示。

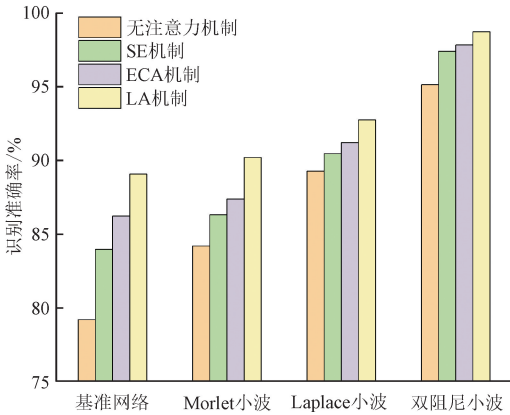


图 13 不同组合对比实验结果柱状图

Fig. 13 Comparison of experimental histograms of different combinations

(1)从图 13 横坐标看,将卷积神经网络的预处理层替换为小波卷积层的确显著提高了模型性能,

Morlet 小波和 Laplace 小波的识别准确率分别提升了 5%和10.08%,证明本文提出的双阻尼小波效果最显著,相对于基准网络提升了 15.9%。

表 2 不同组合模型的识别准确率

Table 2 Model accuracy for different combinations

类型	识别准确率/%			
	无注意力机制	SE 机制	ECA 机制	LA 机制
基准网络	79.20	83.97	86.23	89.08
Morlet 小波	84.20	86.32	87.39	90.20
Laplace 小波	89.28	90.46	91.21	92.75
双阻尼小波	95.14	97.41	97.84	98.73

注:加粗数据表示最优值。

(2)对于同一种小波函数而言(包括基准网络),引入注意力机制能给模型识别准确率带来一定程度的提升。以 Morlet 小波为例,SE 和 ECA 机制相对于未引入注意力机制的模型分别提升了 2.12%、3.19%,而本文所引入的轻量级 LA 机制准确率 98.73%,比无注意力机制提升 8%,效果最佳。

综上,本文提出的以 ResNet18 作为基准骨架,将首个预处理层替换为双阻尼小波卷积层,并引入 LA 机制的 Bdw-ResNet 模型,在南昌铁路局的机车轴承数据集中表现出最佳的效果。

3.2 案例 2:青岛四方轮对轴承数据集

青岛四方轮对轴承试验台如图 14 所示,主要由风扇电机、驱动电机、车轮和车轴、皮带传动系统、待测轴承、径向和垂向加载装置等组成。采样频率为 20 kHz,轴承转速为 589 r/min,垂向载荷设置为 146 kN,车辆的运行速度为 90 km/h。



图 14 青岛四方轮对轴承试验台

Fig. 14 Qingdao Sifang wheelset bearing test bench

本次实验同样通过完全无重叠方式滑移时域窗口进行采样,每组样本设置为 1 024 个样本点,得到了 1 种正常和 7 种故障状态总计 8 种类型的数据样本,其中每种数据类型包含 195 个样本,总共 1 560 个样本,依次编号为 D0~D7,如表 3 所示。

表 3 轮对轴承样本

Table 3 Sample structure of wheelset bearings			
故障类型	故障尺寸/(mm×mm)	样本数	标签
正常		195	D0
滚动体故障	3×35	195	D1
内圈故障	3×45	195	D2
滚动体故障	5×30	195	D3
外圈故障	10×30	195	D4
外圈故障	10×45	195	D5
保持架故障	10×35	195	D6
内圈故障	10×45	195	D7

3.2.1 基于 Bdw-ResNet 模型的实验结果

同理,将表 3 的轴承数据输入到 Bdw-ResNet 模型中训练 100 代,得到的结果如图 15、图 16 所示。由图可知,模型训练到第 25 代准确率达到 99.46%,损失值降低到 0.15 以下,之后准确率曲线和 Loss 曲线都趋于平稳,只在小范围内波动,说明模型已达收敛状态,在本次实验中取得了较好的表现。

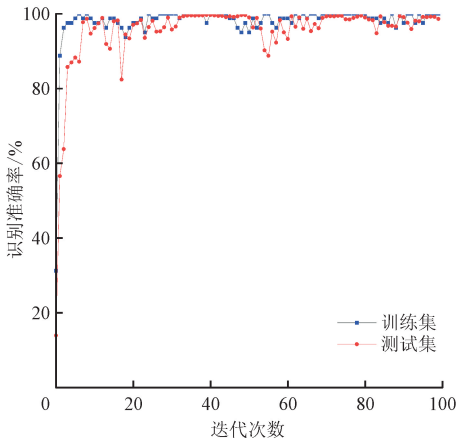


图 15 案例 2 识别准确率变化曲线

Fig. 15 Recognition accuracy curve of case 2

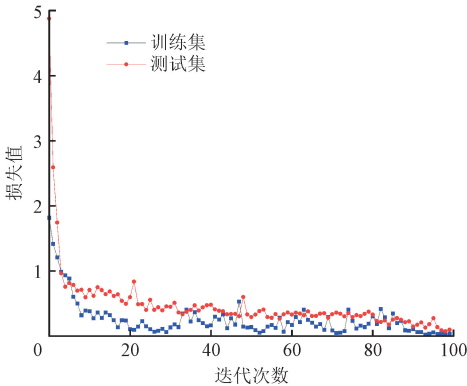


图 16 案例 2 的 Loss 函数变化曲线

Fig. 16 Curve of loss function of case 2

图 17 为使用混淆矩阵的测试集分类的可视化结果。使用表 3 数据进行 64 次重复实验,总共 1 560 个样本。结果表明,有一个 3 mm×35 mm 的滚动体故障被误判成了 10 mm×30 mm 的外圈故障类型,其他类型都保持着 100% 的识别准确率,说明本次故障诊断任务取得了较好的分类结果。

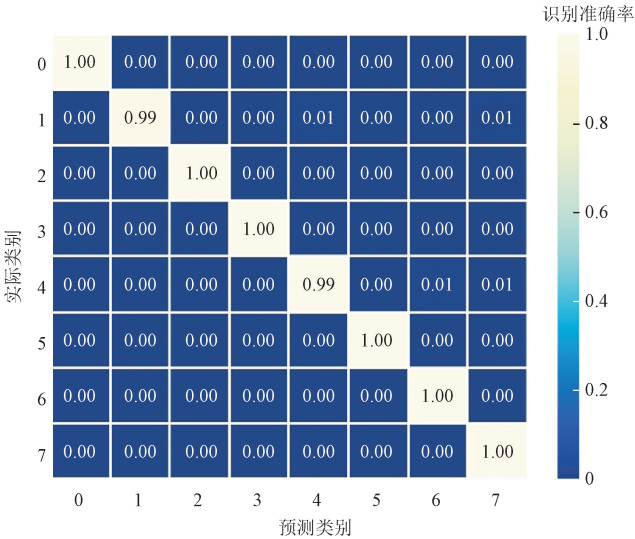


图 17 案例 2 混淆矩阵

Fig. 17 Case 2 confusion matrix

为了更加直观地展现模型的特征提取能力,再次使用 *t*-SNE 算法进行高维数据降维及可视化,结果如图 18 所示。由图可知,对于经过训练后的模型,各种故障类型没有重叠区域,且特征分类明显。

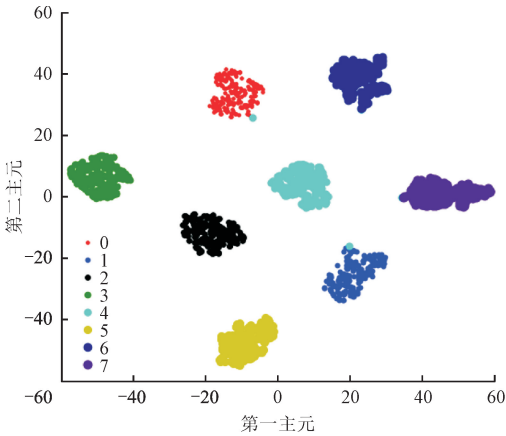


图 18 案例 2 特征降维图

Fig. 18 Case 2 dimensionality reduction of features

3.2.2 对比实验

同样地,将基准网络的预处理层依次替换为 Morlet 小波卷积层、Laplace 小波卷积层、双阻尼小波卷积层,小波参数设置同案例 1,以及将 SE、

ECA、LA 机制插入残差块中,验证 Bdw-ResNet 模型优越性,对比实验结果如表 4 和图 19 所示。

表 4 不同组合模型的识别准确率

Table 4 Model accuracy for different combinations

类型	识别准确率/%			
	无注意力机制	SE 机制	ECA 机制	LA 机制
基准网络	86.00	89.91	90.18	91.00
Morlet 小波	87.30	90.12	88.95	92.03
Laplace 小波	90.42	92.09	91.70	93.85
双阻尼小波	94.46	96.16	96.80	99.46

注:加粗数据表示最优值。

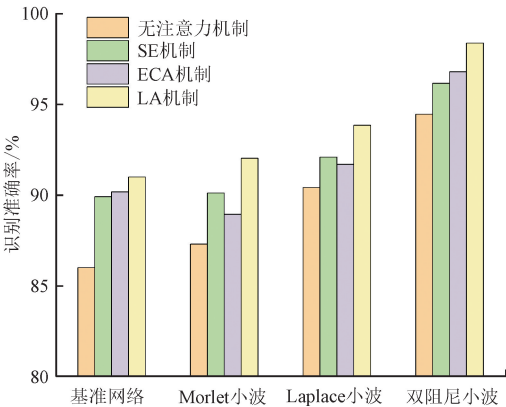


图 19 对比实验结果柱状图

Fig. 19 Comparison of experimental histograms

从表 4 和图 19 可知,在青岛四方轮对轴承故障诊断实验中,将预处理层替换为小波卷积层和嵌入注意力机制的确增加了模型性能,并且本文所提出的双阻尼小波和 LA 机制的组合取得了最优异的表现,准确率 99.46%,再一次验证了本文所提出基于 Bdw-ResNet 模型方法的有效性和高效性。

4 可解释性分析

可解释性的关键在于构建模型中映射关系,并与使用者先验知识的直接联系,从这一角度出发,出现了两种类型的可解释路径。一种是基于使用者的先验知识设计透明可理解的算法映射关系,提高模型的内在可解释性。另一种是将已有算法的非线性映射关系解构成符合先验知识的可理解的线性低维映射。将这两种分别简称为先验赋能和归因解释^[24-25],下面采用这两种类型对本文所提的 Bdw-ResNet 模型及其相关组件进行可解释性分析。

4.1 先验赋能

(1)小波变换作为一种经典的时频分析方法,其

核心是滤波器与信号的内积运算。而卷积实际上也是在 进行内积操作,如图 20 所示。鉴于两者在处理高维数据方面具有相似性,同时由于故障机理、传递路径以及传感器的影响,现有的小波不能很好地匹配轴承真实振动信号中的故障脉冲响应,故本文提出了一种双边非对称衰减的小波,并构建了小波卷积层替换基准 CNN 的第一层卷积层,利用成熟的小波理论先验知识嵌入网络,提高了模型内在可解释性。

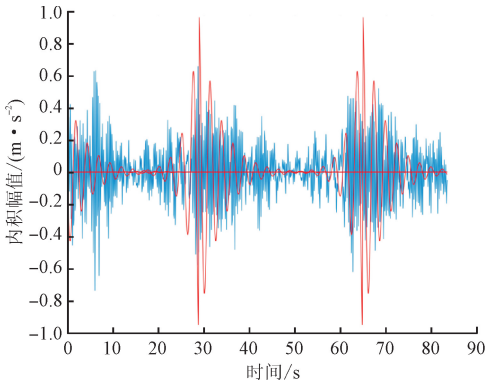


图 20 双阻尼小波与信号内积过程

Fig. 20 Process of bi-damped wavelets and signal inner product

(2)激活函数凭借其能为神经网络引入非线性能力,以便网络能够学习和表征各种数据的复杂模式和特征,从而获得了广泛的应用。有研究表明,波动较大的小波值不利于优化,需要用激活函数来处理指数分量。鉴于此,本文使用 sigmoid 激活函数来调整双阻尼小波基函数的以 e 为基底的指数分量,以将其限制在(0, 1) 范围内,同时让两个可学习的参数 u 、 s 范围限制在 $(-1, 1)$ 之间,sigmoid 激活函数为

$$s(x) = \frac{1}{1 + e^{-x}}$$

(12)

这种操作类似于数据归一化,引入激活函数来调节比例因子 u 和平移因子 s 的一致性。作为小波权重,激活函数解决了由于数值过大引起的模型梯度爆炸的问题,从而设计出既可物理解释又与神经网络相兼容的双阻尼小波卷积层。

4.2 归因解释

(1)热力谱图作为一种可视化工具,通常以特殊的高亮形式展示特征的分布和聚集程度。为更好地理解本文提出的 LA 机制如何发挥作用,将局部注意力权重映射到输入信号生成热力谱图,解释模型的决策依据。以预处理层为双阻尼小波卷积层和 Laplace 小波卷积层为例,输出模型最后一层的权重特征,得

到的热力谱图见图 21、图 22。

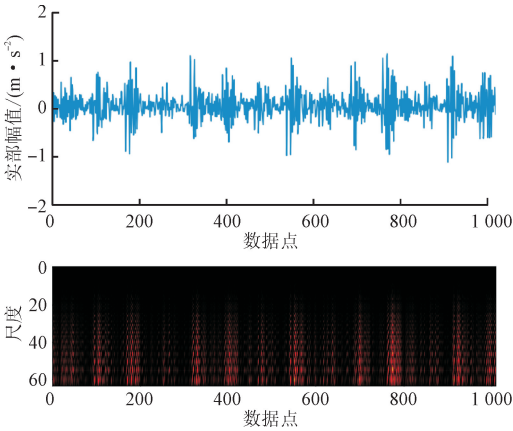


图 21 双阻尼小波-热力谱图

Fig. 21 Bi-damped wavelet-thermodynamic spectrum

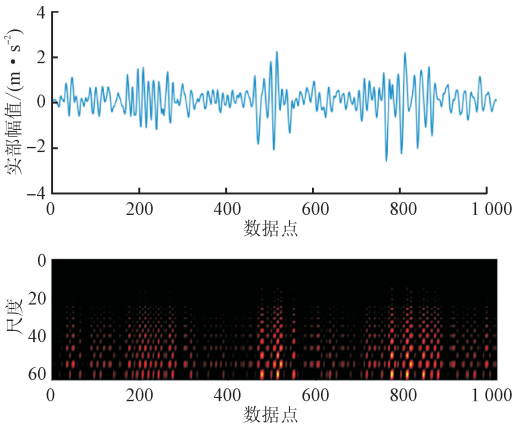


图 22 Laplace 小波-热力谱图

Fig. 22 Laplace wavelet-thermodynamic spectrum

由图 21、22 可知,随着模型训练并反向传播更新,LA 机制会自动学习原始信号中最突出的部分,从输入信息中筛选出利于模型预测的特征并赋予其更高的权重。原始信号冲击部分越尖锐,映射在热力谱图的高亮部分越显著,说明权重越高,表明相应特征对于模型决策贡献越多。每种故障类型对应一种特征映射关系,网络自动学习并强化这种关系,最后通过分类层将多种故障类型识别出来。

(2) 在故障诊断模型的决策中,如何将现有模型的非线性映射关系解构成符合先验知识的可理解的线性低维映射,是一个巨大的挑战。基于此,利用 Kurtogram 函数得到 4 个状态信号的谱峭度及其相关频带^[26]:正常状态,最优中心频率为 7.5 kHz,带宽为 1.67 kHz;内圈故障,最优中心频率为 7.92 kHz,带宽为 0.83 kHz;外圈故障,最优中心频率为 8.57 kHz,带宽为 0.83 kHz;滚动体故障,最优中心频率为 9.5 kHz,带宽为 0.625 kHz。4 种状态下的故障频带图见图 23。

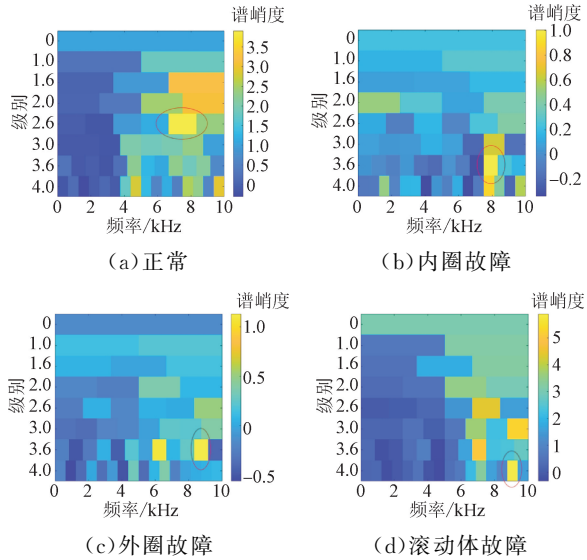


图 23 4 种状态下的故障频带

Fig. 23 Frequency bands of faults in four states

累计频带作为一种理解小波核网络组件如何决策的有效手段^[27],包含所有通道的小波卷积核,其通过计算每个小波核权重的归一化傅里叶变换值和故障频率,来识别小波卷积层重点关注的部分。累计频带图中,纵坐标为 1 对应的频率即为识别到的故障频率。

为了直观地感受本文提出的双阻尼小波卷积层自适应调整权重,进行了故障区域的定位。对训练前后的累计频带进行统计,中心频率范围为 2.86~8.57 kHz。如图 24 所示,累计频带从 2.86 kHz(训练前)过渡到 8.57 kHz(训练后)时,计算得到的中心频率为 8.57 kHz,位于正常、内圈故障、外圈故障和滚动体故障带宽范围内。由此可见,本文构建的双阻尼小波卷积层可以看作是一个最优频带调整的过程,能够解释为集中在包含更多故障信息的频率上,善于提取优越的故障频带,进而识别出最优频段。

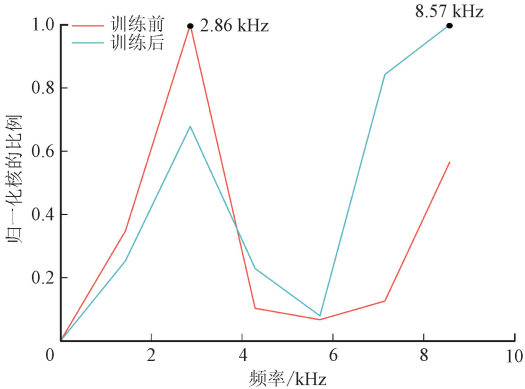


图 24 训练前后累计频带变化

Fig. 24 Changes in the cumulative frequency band before and after training

5 结 论

针对 Morlet、Laplace 等小波与真实轴承故障脉冲响应不匹配的问题,本文提出一种既可物理解释又与神经网络兼容的 Bdw-ResNet 模型以及故障诊断的完整流程,主要结论如下。

(1) 提出了一种 sigmoid 激活函数加权的双阻尼小波,同时引入即插即用的轻量级的局部注意力机制,利用组归一化和一维卷积等技术,组建了 Bdw-ResNet 模型,结合模型优秀的自适应特征提取能力,对滚动轴承进行故障诊断。在南昌铁路局机车轴承数据集和青岛四方轮对轴承数据集上开展试验,分别取得了 98.73% 和 99.46% 的识别准确率。

(2) 与融合不同类型的 Morlet、Laplace 小波函数和主流 SE、ECA 注意力机制算法组合进行对比,结果表明,双阻尼小波和 LA 机制的组合取得了最高的识别准确度。热力谱图和累计频带图等结果均表明本文所提的 Bdw-ResNet 模型及组件能够调整最优频带,从而自动定位故障区域,具有较好的可解释性。

本文主要聚焦的是单一工况下的任务,在以后的研究中,我们将探索变工况、跨机器设备等更复杂的迁移目标,在智能诊断的鲁棒性、信号降噪和可解释性方面带来更多的成果。

参考文献:

- [1] 顾晓辉,杨绍普,刘文朋,等. 高速列车轴箱轴承健康监测与故障诊断研究综述 [J]. 力学学报, 2022, 54(7): 1780-1796.
GU Xiaohui, YANG Shaopu, LIU Wenpeng, et al. Review of health monitoring and fault diagnosis of axle-box bearing of high-speed train [J]. Chinese Journal of Theoretical and Applied Mechanics, 2022, 54(7): 1780-1796.
- [2] 黄梓幸,宋冬利,罗亚萍,等. 车辆服役环境对高速列车轴箱轴承温度的影响分析 [J]. 中南大学学报(自然科学版), 2024, 55(3): 1252-1262.
HUANG Zixing, SONG Dongli, LUO Yaping, et al. Analysis of influence of vehicle environment on temperature of high speed railways axle box bearing [J]. Journal of Central South University(Science and Technology), 2024, 55(3): 1252-1262.
- [3] 张龙,刘皓阳,张号,等. 改进共空间模式与多源特征融合的轴承智能诊断方法 [J]. 西安交通大学学报, 2023, 57(8): 127-137.

- ZHANG Long, LIU Haoyang, ZHANG Hao, et al. Intelligent bearing fault diagnosis using modified common spatial pattern and multi-source feature fusion [J]. Journal of Xi'an Jiaotong University, 2023, 57(8): 127-137.
- [4] ZHAO Zhibin, LI Tianfu, WU Jingyao, et al. Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study [J]. ISA Transactions, 2020, 107: 224-255.
- [5] ZHANG Yu, TINO P, LEONARDIS A, et al. A survey on neural network interpretability [J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2021, 5(5): 726-742.
- [6] 李响,付春霖,雷亚国,等. 保证数据隐私的装备协同智能故障诊断联邦迁移学习方法 [J]. 机械工程学报, 2023, 59(6): 1-9.
LI Xiang, FU Chunlin, LEI Yaguo, et al. Federated transfer learning method for privacy-preserving collaborative intelligent machinery fault diagnostics [J]. Journal of Mechanical Engineering, 2023, 59(6): 1-9.
- [7] HUANG Wenyi, CHENG Junsheng, YANG Yu. Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection [J]. Mechanical Systems and Signal Processing, 2019, 114: 165-188.
- [8] WU Hao, LI Jimeng, ZHANG Qingyu, et al. Intelligent fault diagnosis of rolling bearings under varying operating conditions based on domain-adversarial neural network and attention mechanism [J]. ISA Transactions, 2022, 130: 477-489.
- [9] 李霖,张西宁,刘书语,等. 基于二维经验小波纹理域特征自适应提取的轴承故障诊断方法 [J]. 西安交通大学学报, 2021, 55(12): 79-86.
LI Lin, ZHANG Xining, LIU Shuyu, et al. Bearing fault diagnosis method based on 2D empirical wavelet transform texture domain feature adaptive extraction [J]. Journal of Xi'an Jiaotong University, 2021, 55(12): 79-86.
- [10] 李晋斐,赵冬青,王栋民,等. 一种基于组合赋权法的小波去噪质量评价方法 [J]. 北京航空航天大学学报, 2023, 49(3): 718-725.
LI Jinfei, ZHAO Dongqing, WANG Dongmin, et al. A quality evaluation method for wavelet denoising based on combinatorial weighting method [J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(3): 718-725.
- [11] ALEKSEEV A, BOBE A. GaborNet: Gabor filters with learnable parameters in deep convolutional neural network [C]//2019 International Conference on Engi-

- neering and Telecommunication (EnT). Piscataway, NJ, USA: IEEE, 2019: 1-4.
- [12] LIAO Mengliang, LIU Chang, WANG Cong, et al. Research on a rolling bearing fault detection method with wavelet convolution deep transfer learning [J]. IEEE Access, 2021, 9: 45175-45188.
- [13] LI Sinan, LI Tianfu, SUN Chuang, et al. WPCConvNet: an interpretable wavelet packet kernel-constrained convolutional network for noise-robust fault diagnosis [J/OL]. IEEE Transactions on Neural Networks and Learning Systems (2023-06-15) [2024-06-01]. <https://doi.org/10.1109/TNNLS.2023.3282599>.
- [14] 张德祥, 王俊, 袁培成. 基于注意力机制的多尺度全场景监控目标检测方法 [J]. 电子与信息学报, 2022, 44(9): 3249-3257.
- ZHANG Dexiang, WANG Jun, YUAN Peicheng. Object detection method for multi-scale full-scene surveillance based on attention mechanism [J]. Journal of Electronics & Information Technology, 2022, 44(9): 3249-3257.
- [15] 施健聪, 王兴龙, 张俊. 基于双重注意力机制的异步电机故障诊断方法 [J]. 振动与冲击, 2023, 42(21): 110-118.
- SHI Jiancong, WANG Xinglong, ZHANG Jun. Induction motor fault diagnosis method based on dual-attention mechanism [J]. Journal of Vibration and Shock, 2023, 42(21): 110-118.
- [16] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [17] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECANet: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 11531-11539.
- [18] 刘方, 翟涛涛, 侯超强, 等. 基于多普勒调制时移 Laplace 小波的列车轴承故障声信号瞬态成分快速提取方法 [J]. 仪器仪表学报, 2022, 43(3): 40-48.
- LIU Fang, ZHAI Taotao, HOU Chaoqiang, et al. A fast transient component extraction method of train bearing fault acoustic signal based on Doppler modulated time-shifting Laplace wavelet [J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 40-48.
- [19] LI Jimeng, TAO Jinxin, DING Wanmeng, et al. Period-assisted adaptive parameterized wavelet dictionary and its sparse representation for periodic transient features of rolling bearing faults [J]. Mechanical Systems and Signal Processing, 2022, 169: 108796.
- [20] 王天宇, 陈晗, 王刚, 等. 采用小波变换和双向长短期记忆网络的脑电睡眠分期模型 [J]. 西安交通大学学报, 2022, 56(9): 104-111.
- WANG Tianyu, CHEN Han, WANG Gang, et al. EEG sleep staging model using wavelet transform and bidirectional long short-term memory network [J]. Journal of Xi'an Jiaotong University, 2022, 56(9): 104-111.
- [21] LI Tianfu, ZHAO Zhibin, SUN Chuang, et al. Wavelet KernelNet: an interpretable deep neural network for industrial intelligent diagnosis [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, 52(4): 2302-2312.
- [22] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 13708-13717.
- [23] WU Yuxin, HE Kaiming. Group normalization [C]//Computer Vision-ECCV 2018. Cham, Switzerland: Springer International Publishing, 2018: 3-19.
- [24] 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 Ruqiang, SHANG Zuogang, WANG Zhiying, et al. Challenges and opportunities of XAI in industrial intelligent diagnosis: priori-empowered [J]. Journal of Mechanical Engineering, 2024, 60(12): 1-20.
- [25] 严如强, 周峥, 杨远贵, 等. 可解释人工智能在工业智能诊断中的挑战和机遇: 归因解释 [J]. 机械工程学报, 2024, 60(12): 21-40.
- YAN Ruqiang, ZHOU Zheng, YANG Yuanguai, et al. Challenges and opportunities of XAI in industrial intelligent diagnosis: attribution interpretation [J]. Journal of Mechanical Engineering, 2024, 60(12): 21-40.
- [26] HE Chao, SHI Hongmei, LIU Xiaorong, et al. Interpretable physics-informed domain adaptation paradigm for cross-machine transfer diagnosis [J]. Knowledge-Based Systems, 2024, 288: 111499.
- [27] SHANG Zuogang, ZHAO Zhibin, YAN Ruqiang. Denoising fault-aware wavelet network: a signal processing informed neural network for fault diagnosis [J]. Chinese Journal of Mechanical Engineering, 2023, 36(1): 9.

(编辑 亢列梅)