

# 基于 Laplace 小波卷积和 BiGRU 的少量样本故障诊断方法

罗浩<sup>1</sup>, 何超<sup>1</sup>, 陈彪<sup>1</sup>, 路颜萍<sup>1</sup>, 张欣<sup>2</sup>, 张利<sup>1</sup>

(1. 辽宁大学 信息学院, 沈阳 110036;

2. 东北大学 材料科学与工程学院, 沈阳 110819)

**摘要:** 针对滚动轴承通常在复杂条件下工作易发生故障以及训练样本较少等问题。提出一种具有全局平均池化(global average pooling, GAP)并融合双路 Laplace 小波卷积和双向门控循环单元(dual Laplace wavelet convolution bidirectional gated recurrent unit, DLWCB)的故障诊断方法。首先 Laplace 小波卷积将原始信号从时域转换为频域,接着利用双路卷积和 BiGRU 挖掘少量样本的多尺度和时空特征;然后设计 GAP 降低模型的参数量并全面融合各 GRU 细胞提取的时空特征。其中从优化算法和目标函数入手,引入标签平滑、AdamP 等提升 DLWCB 应对少量样本的能力,最后实现复杂工况下故障诊断。在两种轴承数据集、有限噪声样本下,50 s 内便可完成训练,达到 98% 以上准确率,所提方法具有良好的泛化性、鲁棒性和诊断效率。

**关键词:** 拉普拉斯小波卷积核;双向门控循环单元;标签平滑;故障诊断;少量样本

**中图分类号:** TP311;TP391

**文献标志码:** A

**DOI:**10.13465/j.cnki.jvs.2022.24.006

## Small sample fault diagnosis based on Laplace wavelet convolution and BiGRU

LUO Hao<sup>1</sup>, HE Chao<sup>1</sup>, CHEN Biao<sup>1</sup>, LU Yanping<sup>1</sup>, ZHANG Xin<sup>2</sup>, ZHANG Li<sup>1</sup>

(1. School of Information, Liaoning University, Shenyang 110036, China;

2. School of Materials Science and Engineering, Northeastern University, Shenyang 110819, China)

**Abstract:** Targeting the problems that rolling bearings usually work under complex conditions, causing breakdown easily and small training samples, a fault diagnosis method with global average pooling (GAP) and fusion of dual Laplace wavelet convolution and bidirectional gated recurrent unit (DLWCB) was proposed. Firstly, Laplace wavelet convolution was utilized to transform original signals from time to frequency domain, and then the multi-scale and spatiotemporal characteristics of small samples were mined by dual convolution and BiGRU. In addition, GAP was designed to reduce the number of parameters of the model and integrate the spatiotemporal characteristics of GRUs. From the optimization algorithms and objective functions, label smoothing and AdamP were introduced to improve the capacity of DLWCB to cope with small samples, and finally achieve fault diagnosis under complex conditions. In two rolling bearing data sets, training can be completed in 50 seconds under limited noise samples, and the accuracy is over 98%. The proposed method has the better capacities of generalization, robustness, and diagnosis efficiency.

**Key words:** Laplace wavelet convolution kernel; bidirectional gated recurrent unit; label smoothing; fault diagnosis; small sample

机械旋转设备故障诊断已步入智能化、自动化时代。其中,滚动轴承扮演着重要角色,其健康状态直接影响设备的稳定性、可靠性<sup>[1]</sup>。然而,滚动轴承往往在复杂条件下工作,受到材料退化、温湿度等多种要素影响,故障后轻则影响工厂效益,重则导致

人员伤亡。因此,对其健康状态监测是极具研究意义的课题。

过去基于信号分析、群智能进化、机器学习的故障诊断方法不断涌现<sup>[2-4]</sup>。然而相关算法依赖专家先验知识,难以处理高维数据;群智能优化的寻优结果难以稳定且时间复杂度高。随后,深度学习智能故障诊断如火如荼。或多角度分析旋转机械振动信号特征,或改进深度优化算法,或设计新颖高效的网络结构<sup>[5]</sup>。

近些年,少量样本故障诊断成为研究热点。有的

基金项目: 国家自然科学基金(51704138);辽宁省教育厅科学技术研究项目(LJKMZ20220450);辽宁省自然科学基金(2020-BS-082)

收稿日期: 2021-07-27 修改稿收到日期: 2021-10-14

第一作者 罗浩 男,博士,副教授,1987年生

通信作者 张利 男,博士,教授,1971年生

利用模型特征提取优势和正则化策略,例如 Han 等<sup>[6]</sup>提出基于双向长短时记忆(bidirectional long short-term memory, BiLSTM)和胶囊网络少量样本故障诊断方法。振动信号经卷积神经网络(convolutional neural network, CNN)和 BiLSTM 去噪融合后,胶囊网络对少量样本故障诊断有着良好性能。Saufi 等<sup>[7]</sup>提出基于谱峰度滤波和粒子群优化堆叠稀疏自编码器少量样本诊断方法,当每个故障类型训练数为 100 时,取得较高精度。有的依据数据分布生成多而高质量样本,例如 Li 等<sup>[8]</sup>提出条件 Wasserstein 生成对抗网络(conditional Wasserstein generative adversarial networks, CWGAN),利用源域大量标签数据训练 CWGAN 生成大量样本。针对有限标签目标域数据,加载预训练参数并微调 CWGAN 实现少量样本迁移,取得良好效果。有的应用元学习、迁移学习等新兴机器学习技术实现诊断,例如 Zhang 等<sup>[9]</sup>提出基于孪生神经网络的小样本故障诊断方法,输入相同或不同类的样本对,计算两者特征向量的  $L1$  距离,判断是否同属一类来训练模型,最后将支持集与查询集作为特征对并计算相似度来实现故障诊断。Wang 等<sup>[10]</sup>在此基础上,利用全连接来模拟特征对相似性度量,并添加正则化手段提升效果。Wu 等<sup>[11]</sup>比较了基于特征迁移、微调、元关系网络三者间小样本迁移性能,得出当样本较少或源域、目标域相似性较大时,元迁移占据主导地位;反之,特征迁移优势逐渐明显。

针对模型特征提取优势和正则化策略,除上述胶囊网络外,沈涛等<sup>[12]</sup>探究 CNN-LSTM(convolutional neural network-long short-term memor)对少量样本表达能力,选择训练集比例 0.3 来探讨模型应对各种复杂工况能力。然而单尺度单向的 CNN-LSTM,没有充分挖掘模型性能,并且批归一化和 Dropout 等都略常规。Yang 等<sup>[13]</sup>提出一种基于 1DCNN 和 BiGRU 框架损害检测方法,将多传感器信号组成信号矩阵输入模型中,融合两者提取的特征实现故障诊断,但未探究少量训练样本下 BiGRU 效果。另外,双路 Laplace 小波卷积和双向门控循环单元(dual Laplace wavelet convolution bidirectional gated recurrent unit, DL-WCB)每类训练数是 Saufi 等研究的 60% 为 60 个,泛化难度更大。

在上述分析基础上,研究内容总结如下:

(1) 从拓扑结构和正则化出发,提出一种少量样本故障诊断思路,设计一种端对端融合双路 Laplace 小波卷积核和 BiGRU 的故障诊断方法。首先提出

Laplace 小波卷积核,并和平均能量池化共同作用于所设计模型,提取多尺度特征,大卷积核具有强鲁棒性<sup>[14]</sup>。然后, BiGRU 学习隐藏信息进一步提取高级特征。设计全局平均池化(global average pooling, GAP)增强通道间联系并提高 BiGRU 的特征利用率。

(2) 引入故障诊断不常见的手段标签平滑正则化(label smoothing regularization, LSR)和流形正则化并结合批归一化(batch normalization, BN)、PreLU 动态激活、AdamP 等来改善 DLWCB 泛化性,缓解少量样本下模型的过拟合现象。

(3) 针对噪声样本,提出具有参数传递的故障诊断框架,经少量样本微调模型便具备更高的效率。探究了 BiGRU 和 GAP 在少量样本故障诊断应用中的性能。实验表明,该方法具有较高诊断效率。

## 1 相关理论基础

### 1.1 卷积神经网络

卷积神经网络(convolutional neural network, CNN)一般由滤波块和分类诊断块两个模块组成。一般故障诊断 CNN,如图 1 所示。

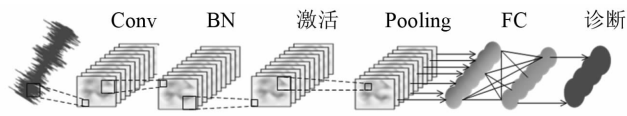


图 1 故障诊断 CNN 结构

Fig. 1 CNN for fault diagnosis

信号处理领域,一般应用具有相同内核的一维卷积计算信号的延迟累积。输出  $y$  如式(1)所示

$$y = \text{PReLU}(k * x) = \text{PReLU}\left(\sum_{w=1}^W k_w x_{t-w+1} + b_w\right) \quad (1)$$

式中:  $k_w$  和  $b_w$  为卷积核和偏置;  $x_{t-w+1}$  为输入振动信号信号;  $*$  为卷积;  $\text{PReLU}(\cdot)$  为神经网络激活函数。

池化对特征进行选择并降低模型参数数量防止过拟合,故障诊断中常用最大池化。激活函数可以增强神经网络的表示和学习能力,提高计算效率。BN 不仅可以提高优化效率,而且由于其随机选择批次,可以增强模型的泛化能力。全连接将振动信号分布式特征表示映射到样本标记空间,最后应用 SoftMax 激活函数进行故障诊断。

### 1.2 双向门控循环神经网络

门控循环神经网络(gate recurrent unit, GRU)由更新门  $z_t$  和重置门  $r_t$  组成。BiGRU 由双向多个 GRU 组成,如图 2、图 3 所示,信号通过不同且不共享参数的隐藏层将前后向输出连接到相同层以提取过去和未来的

特征。如式(2)~式(4)所示

$$\vec{h}_t = GRU(\vec{h}_{t-1}, x_t) \quad (2)$$

$$\tilde{h}_t = GRU(\tilde{h}_{t-1}, x_t) \quad (3)$$

$$h_t = w_t \vec{h}_t + v_t \tilde{h}_t + b_t \quad (4)$$

式中： $w_t, v_t$  为前向和后向的状态权重矩阵； $x_t$  为输入信号； $b_t$  为偏置。

然而，深度故障诊断中，一般取最后一个隐藏神经元细胞输出作为 BiGRU 学习到的振动信号特征并作为后续输入，忽略了其他 GRU 细胞，在 DLWCB 中，将在 BiGRU 后连接 GAP 解决。

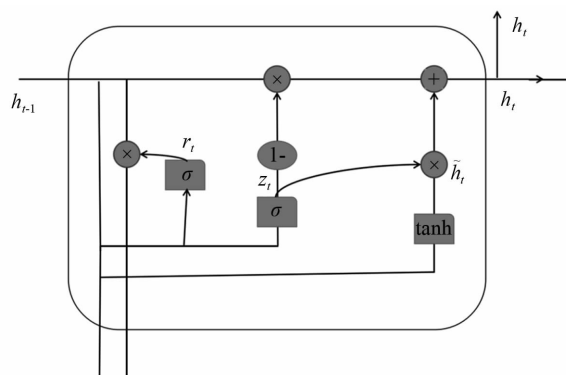


图 2 GRU 细胞

Fig. 2 GRU cell

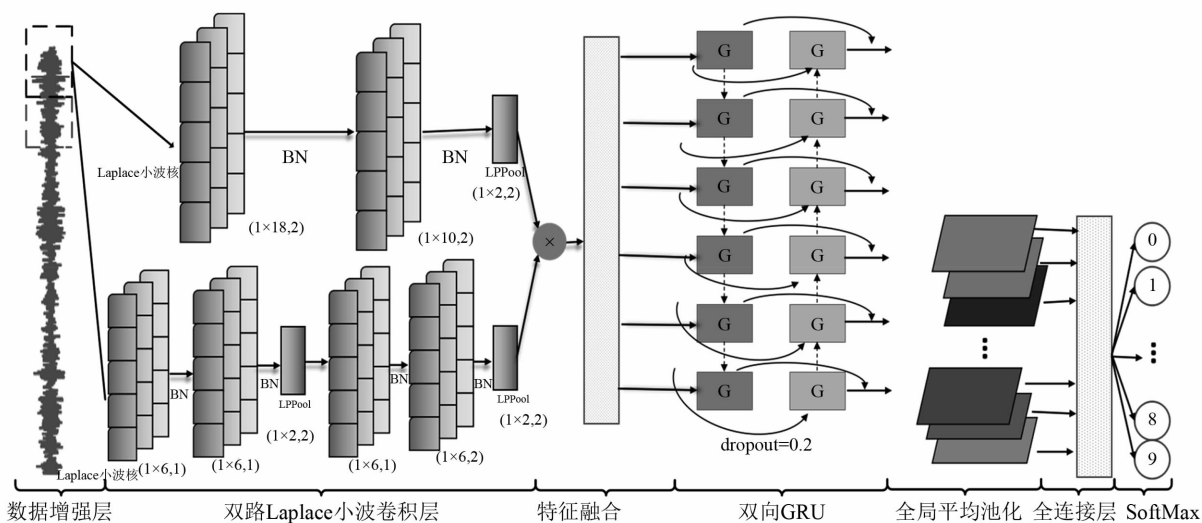


图 3 DLWCB 智能故障诊断模型

Fig. 3 DLWCB intelligent fault diagnosis model

## 2 Laplace 小波卷积

### 2.1 Laplace 小波卷积定义

受 Morlet 小波<sup>[15-16]</sup>启发，将 Laplace 小波思想融入卷积核中，提出并定义 Laplace 小波卷积核。

在时域中，小波基本字典  $\psi_{u,s}(t)$  定义如式(5)

$$\psi_{u,s}(t) = \psi\left(\frac{t-u}{s}\right) \quad (5)$$

式中： $\psi(\cdot)$  为小波基函数； $t$  为时间； $s$  为尺度因子， $u$  为平移因子； $s, u$  为自适应可调节参数。

考虑到滚动轴承的机械振动信号属于实信号，故采用实 Laplace 小波基函数分析信号，如式(6)所示

$$\psi(t) = A e^{\frac{-\xi}{\sqrt{1-\xi^2}} 2\pi f(t-\tau)} \times \sin[2\pi f(t-\tau)] \quad (6)$$

式中： $f$  为信号频率； $\xi$  为黏性阻尼比； $\tau$  为时间参数； $A$  为小波归一化函数。

由式(5)、式(6)可得实 Laplace 小波卷积字典  $\psi_{L,u,s}(t)$  如式(7)所示

$$\psi_{L,u,s}(t) = A e^{\frac{-\xi}{\sqrt{1-\xi^2}} 2\pi f\left(\frac{t-u}{s}-\tau\right)} \times \sin\left[2\pi f\left(\frac{t-u}{s}-\tau\right)\right] \quad (7)$$

将式(7)代入式(1)，得到 Laplace 小波卷积核输出  $y_L$ ，如(8)式所示

$$y_L = \text{PReLU}[\psi_{L,u,s}(t) * x_t] =$$

$$\text{PReLU}\left[\sum_{t=1}^T \psi_{L,u,s}(t) x_t + b_t\right] \quad (8)$$

### 2.2 Laplace 小波卷积参数量

暂不考虑偏置，普通卷积参数量为卷积核尺寸与卷积核个数的乘积，而 Laplace 小波卷积核  $s, u$  需调整，参数量是卷积核个数的二倍。DLWCB 中，第一层参数量为  $50 \times 2 = 100$ ，而普通卷积核参数量为  $50 \times 18 = 900$ 。显然卷积核尺寸越大，参数减少越明显。

### 2.3 参数 $s$ 和 $u$ 更新

$s$  和  $u$  更新依据反向传播算法，在第一层 Laplace 小波卷积中，参数更新可以描述为

$$\begin{cases} \ell_{u_w} = \frac{\partial l}{\partial u_w} = \frac{\partial l}{\partial y_w} \frac{\partial y_w}{\partial \psi_{u,s}^w} \frac{\partial \psi_{u,s}^w}{\partial u_w} \\ u_w \leftarrow u_w - \alpha \ell_{u_w} \end{cases} \quad (9)$$

$$\begin{cases} \ell_{s_w} = \frac{\partial l}{\partial s_w} = \frac{\partial l}{\partial y_w} \frac{\partial y_w}{\partial \psi_{u,s}^w} \frac{\partial \psi_{u,s}^w}{\partial s_w} \\ s_w \leftarrow s_w - \alpha \ell_{s_w} \end{cases} \quad (10)$$

式中： $\ell$  为参数  $s$  和  $u$  的梯度； $l$  为 LSR 多重目标函数； $y_w$  为 Laplace 小波卷积核 ( $\psi_{u,s}^w$ ) 特征图； $\alpha$  为学习率。 $s$  和  $u$  更新规则为学习率  $\alpha$  和  $\ell$  的乘积作差。

同时,根据链式求导法则可以得到两个参数的偏导数如式(11)和式(12)所示,并将式(11)、式(12)代入式(9)、式(10)中对两个参数进行更新。

$$\frac{\partial \psi_{L,u,s}}{\partial u} = \frac{2A\pi f e^{\frac{-\xi}{\sqrt{1-\xi^2}} \cdot 2\pi f (\frac{t-u}{s} - \tau)}}{-s} \cdot \left\{ \frac{-\xi}{\sqrt{1-\xi^2}} \sin \left[ 2\pi f \left( \frac{t-u}{s} - \tau \right) \right] + \cos \left[ 2\pi f \left( \frac{t-u}{s} - \tau \right) \right] \right\} \quad (11)$$

$$\frac{\partial \psi_{L,u,s}}{\partial s} = 2A\pi f e^{\frac{-\xi}{\sqrt{1-\xi^2}} \cdot 2\pi f (\frac{t-u}{s} - \tau)} \times \frac{t-u}{-s^2} \cdot \left\{ \frac{-\xi}{\sqrt{1-\xi^2}} \sin \left[ 2\pi f \left( \frac{t-u}{s} - \tau \right) \right] + \cos \left[ 2\pi f \left( \frac{t-u}{s} - \tau \right) \right] \right\} \quad (12)$$

3 DLWCB 故障诊断模型

CNN – RNN( convolutional neural network – recurrent

neural network)已取得一定应用<sup>[17-18]</sup>。但少量样本下其性能表现却鲜有研究,而且优化算法和训练方式较为常规,潜在性能未能得到进一步挖掘。为此,提出一种少量样本智能故障诊断方法——DLWCB,具体结构见图3。DLWCB 由数据增强层、双路 Laplace 小波卷积层、特征融合层、BiGRU、GAP 和诊断层组成。

图3 结构中,将 Laplace 小波思想融入卷积过程中,并应用于首个卷积层。

GAP 则解决了 1.2 节所提问题,通过计算各个 GRU 提取的特征矩阵作全局平均池化,综合考虑各个 GRU 的特征输出,降低了诊断层的训练压力并提高特征利用率。DLWCB 相关结构细节参数,如表 1 所示。

表 1 DLWCB 结构细节  
Tab.1 Structure details of DLWCB 图层类型

| 图层类型       | 核大小/步长      | 隐藏数    | 激活函数           | BN 层         | 输入大小       | 输出大小       | 参数量            |
|------------|-------------|--------|----------------|--------------|------------|------------|----------------|
| Conv1d_1   | 18/2 & 10/2 |        | PReLU(50 & 30) | YES(50 & 30) | (1,1 024)  | (30,248)   | 15 820         |
| Maxpool_1  | 2/2         |        |                |              | (30,248)   | (30,124)   |                |
| Conv1d_21  | 6/1 & 6/1   |        | PReLU(50 & 40) | YES(50 & 40) | (1,1 024)  | (40,1 014) | 12 460         |
| Maxpool_21 | 2/2         |        |                |              | (40,1 014) | (40,507)   |                |
| Conv1d_22  | 6/1 & 6/2   |        | PReLU(30 & 30) | YES(30 & 30) | (40,507)   | (30,249)   | 12 840         |
| Maxpool_22 | 2/2         |        |                |              | (30,249)   | (30,124)   |                |
| BiGRU      |             | 150/30 | Tanh           |              | (30,124)   | (30,60)    | 248 400/59 760 |
| GAP        |             |        |                |              | (30,60)    | (30,1)     |                |
| FC         |             | 10     | SotfMax        |              | 30         | 10         | 310            |

4 实验结果与分析

4.1 数据描述与实验配置

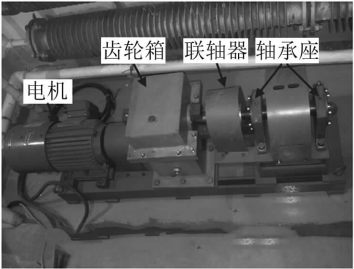
本节将结合两个轴承故障诊断案例,以证明所提方法性能、效率和应用价值。实验应用 pytorch 1.8.0,运行在 GTX970M GPU 上。采用平均功率池化,PReLU 动态激活函数,AdamP 反向传播算法<sup>[19]</sup>和带有流形正则项的标签平滑损失函数<sup>[20]</sup>来训练 DLWCB。数据集 S1 – 凯斯西储大学滚动轴承数据<sup>[21]</sup>。采样频率 12 kHz,负载 0 ~ 2.237 1 kW。电动烟花引起单点故障。加速度传感器位于电机外壳驱动端,收集加速度数据。据负载不同,划分 ABCD 4 组,如表 2 所示。

表 2 数据集 S1 描述  
Tab.2 Data set description of S1

| 数据集      | 负载/kW   | 位置  | 损伤直径/mm                 | 标签    |
|----------|---------|-----|-------------------------|-------|
| 0/       | 0/1/2/3 | 正常  |                         | 0     |
| 0.745 7/ |         | 内圈  | 0.177 8/0.355 6/0.533 4 | 1/2/3 |
| 1.491 4/ |         | 滚动体 | 0.177 8/0.355 6/0.533 4 | 4/5/6 |
| 2.237 1  |         | 外圈  | 0.177 8/0.355 6/0.533 4 | 7/8/9 |

数据集 S2 – 为验证算法的有效性和可靠性,搭建的轴承故障实验平台,如图 4 所示。该平台由电机、齿轮箱、联轴器、轴承座、操作柜和操作台等部分组成。

采集频率为 50 kHz,使用加速度计采集 1 000 r/min 下振动信号。采集单元为德国 Bruel & Kjaer Vibro 公司的一款用于振动分析的 VDAU – 6000,具有 16 个可以实现同步数据采集的输入通道。根据不同位置的不同状态将振动信号分为 4 类。如表 3 所示。



(a) 滚动轴承-转子-齿轮实验台



(b) 测试数据采集平台

图 4 机械故障模拟实验台

Fig.4 Mechanical fault simulation test bench

表 3 数据集 S2 描述  
Tab.3 Data set description of S2

| 状态 | 正常 | 内圈故障 | 外圈故障 | 滚动体故障 |
|----|----|------|------|-------|
| 标签 | 0  | 1    | 2    | 3     |

实验中,S1,S2 均采用滑动窗口采集样本,窗口大小为 400,样本长度为 1 024。两个数据集学习率分别为 0.001,0.000 4,最大迭代次数 epoch = 150,丢弃率 Dropout = 0.2,权重衰减系数为 0.000 1。采取早停策略(patience = 10)。训练样本比例是 0.1(20)~0.5(100)。

4.2 少量样本问题描述

少量样本的概念很难统一准确定义,这与数据和诊断模型复杂度有关。通常,机器学习中“少量样本”会导致模型过拟合。当训练数据不足时,网络学习到特征表示是有限的,只能很好地拟合训练数据,导致训练集精度高,测试集精度低。当训练数据充足时,算法可以有效提取整个数据集特征,在训练和测试集上都取得良好性能。因此,其在两个案例中描述如下:训练数据较少,以训练一个足够泛化的模型,可以在测试集上实现有效故障分类。

4.3 batch\_size 参数选择

批处理(batch\_size, b)会影响模型训练效率和泛化性。针对数据集 S1-B,训练集比例为 0.3(60),仅改变 b,实验结果如图 5 所示。

当  $b \leq 100$  时,  $b$  越大,DLWCB 收敛时需要的 epoch 越多,单个 epoch 较短的训练时间并不能使得整体收敛速度提高。至于收敛损失,  $l_{b=32,80,100} < 0.56$ ,相对其他情况收敛损失更低。三者验证集准确率基本相同。另外,从时间来看,  $t_{b=32} = 75.71$  s,然而  $t_{b=80,100} > 100$  s,

当  $b = 32$  时,达到与  $b = 80,100$  相似的准确率所需时间更少。综上,  $b = 32$  具有最佳诊断效率。

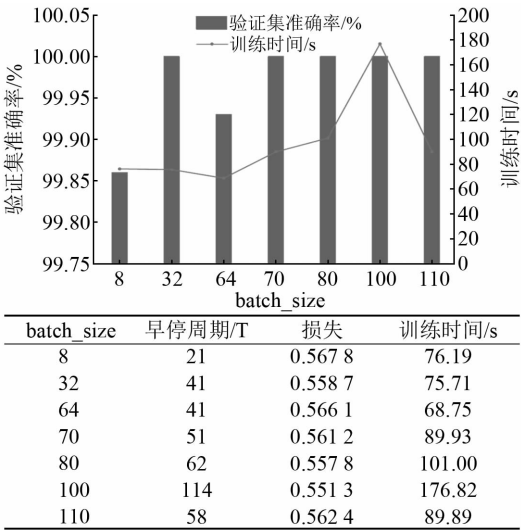


图 5 不同批次训练结果比较

Fig. 5 Comparison of training results of different batch\_size

4.4 Laplace 小波卷积核性能

CNN 的第一个卷积层影响着整个模型的性能<sup>[22]</sup>。因此将第一层普通卷积核替换为 Laplace 小波卷积核,与信号卷积达到小波分析的效果。这样使得时域信号转换到频域,更有效地识别轴承状态<sup>[23]</sup>。实验中,  $f = 100$ ,  $\xi = 0.03$ ,  $\tau = 0.1$ ,  $A = 0.08$ ,  $s \in [1, 100]$ ,  $u \in [0, 100]$ ,  $s, u$  向量维度为输出通道数大小。

采用数据集 S2,每类训练样本 20~100,对比 DLWCB 和 DCB(dual convolution bidirectional gated recurrent unit)(第一层未使用 Laplace 小波卷积)的性能表现,相关结果如表 4 所示。

表 4 数据集 S2 性能表现  
Tab.4 Performance of S2 data set

| 模型    | 训练比例(样本数) | 0.1(20)      | 0.2(40)      | 0.3(60)      | 0.4(80)      | 0.5(100)     |
|-------|-----------|--------------|--------------|--------------|--------------|--------------|
| DLWCB | 准确率/%     | 87.56 ± 0.52 | 94.15 ± 0.33 | 97.56 ± 0.65 | 98.36 ± 0.23 | 99.75 ± 0.01 |
|       | 时间/s      | 63.52        | 75.62        | 74.51        | 95.56        | 72.32        |
|       | 收敛损失      | 0.874 6      | 0.681 2      | 0.625 6      | 0.573 6      | 0.559 8      |
| DCB   | 准确率/%     | 70.23 ± 0.68 | 86.36 ± 0.96 | 95.98 ± 0.76 | 97.62 ± 0.26 | 99.75 ± 0.15 |
|       | 时间/s      | 47.32        | 68.23        | 62.35        | 93.28        | 94.56        |
|       | 收敛损失      | 1.352 4      | 0.853 2      | 0.682 3      | 0.613 6      | 0.568 8      |

随着训练样本数增加,模型性能逐渐提升。由于采用正则化策略,当训练样本数大于 60 时,取得较高精度。即使这样,DLWCB 相较于 DCB 准确率更高、方差更小、更稳定。当样本数少于 40 时,由于训练样本数太少导致正则化方法更难泛化网络。此时,DLWCB 在收敛损失上更低;当样本数为 20 或 40 时,DLWCB 达到 87.56% 和 94.15%,相较于 DCB 提升 17.33% 和 7.79%。但 DLWCB 的时间复杂度较高。

4.5 少量样本参数选择

在 S1-B,  $b = 32$  下,进行 5 次实验取平均值,得到训练时间和测试集准确率,如图 6 所示。

结合表 4 和图 6,轴承数据集 S2 由于信号特征不够明显,达到与 S1 相似性能需更多训练样本。

针对数据集 S2,据 4.2 节所述,同时考虑正则化的影响,采用训练集比例 0.3(60)建立少量样本,因为此时训练样本数对模型的影响还没达到最大,可以体现正则化对模型的作用;至于数据集 S1,当训练集比例大

于 0.3 (60) 时,随着训练样本数增加,准确率提升不足 1% 且时间基本呈线性增加;当训练集比例为 0.3 (60) 时,DLWCB 达到 99% 以上。权衡两者,选择 0.3 (60) 为训练集比例,建立少量样本。

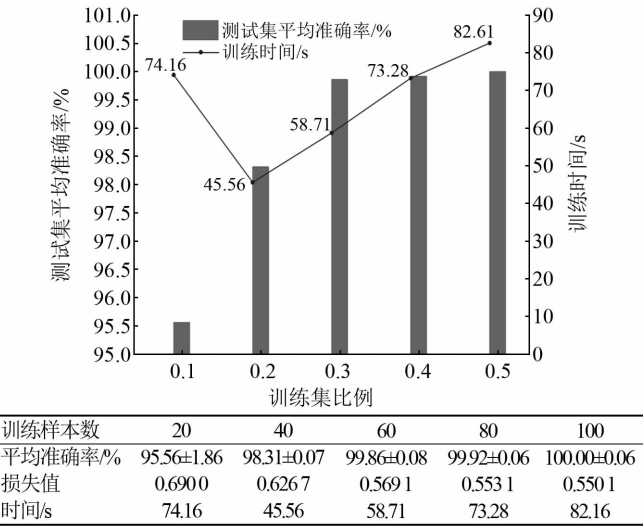


图6 S1 不同训练集比例下性能

Fig. 6 Performance of different training proportion in S1

同样,针对轴承数据集 S1,在 0.3 (60) 下,得到是否应用 Laplace 小波卷积的准确率,如表 5 所示。

| 表 5 Laplace 小波卷积      |       |        |       |       |
|-----------------------|-------|--------|-------|-------|
| Tab. 5 LW convolution |       |        |       |       |
| 模型                    | 测试集/% | 损失     | 时间/s  | epoch |
| DCB                   | 97.86 | 0.5927 | 49.05 | 29    |
| DLWCB                 | 99.86 | 0.5683 | 72.73 | 42    |

少量样本下,训练样本数对模型的影响较低,此时观察 Laplace 小波卷积的作用具有较大参考价值。从表 5 可以看到相较于原始卷积,Laplace 卷积最终收敛损失减少了约 0.02,准确率提升 2%。

综上,针对少量样本故障诊断,将信号处理领域中 Laplace 小波思想融入卷积核中也是一种应对思路。

4.6 正则化分析

采用数据集 S2,如 4.5 节分析,将训练集比例设置为 0.3 (60),建立少量样本。验证正则化方法的收敛性。

4.6.1 AdamP 收敛性分析

几种梯度下降算法在验证集上的表现,如图 7 所示。SGDM 收敛损失最大且收敛速度最慢。Adam, AdamP 有着较快的收敛速度。RMSprop 损失收敛曲线波动相对较大。反观 AdamP 则兼具平缓、较快收敛,在 epoch = 57 时收敛于 0.64,这使得 DLWCB 更具稳定性。

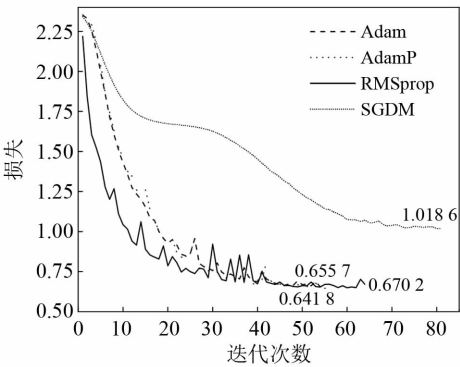


图7 不同算法损失值对比图

Fig. 7 Comparison of loss values of different algorithms

4.6.2 PReLU 和 ReLU 泛化性分析

ReLU 和 PReLU 两种激活函数在不同 epoch 下的收敛曲线,如图 8 所示。结果显示了在每个训练阶段后训练过程的训练准确率和误差。可以看出,与 ReLU 相比,PReLU - DLWCB 准确率更高并且收敛损失更低。由图 8(b)可知,PReLU 相较于 ReLU 训练过程更加平滑,在 epoch = 72 时,收敛到最低损失。表明 PReLU - DLWCB 更能学习到信号特征且更具稳定性。

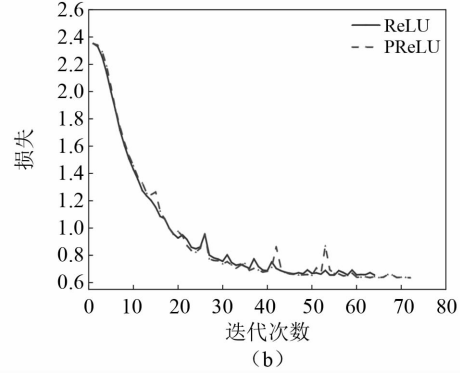
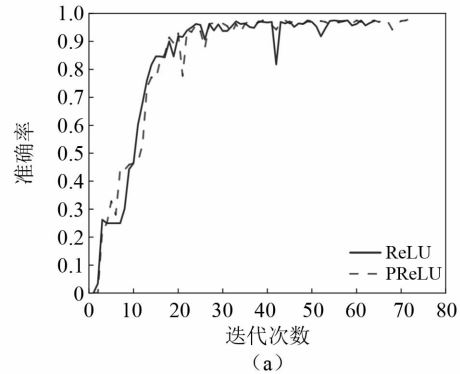


图8 不同激活函数下 DLWCB 性能对比

Fig. 8 Comparison of different activation functions

4.6.3 目标函数泛化性分析

针对少量样本,仅改变损失函数,LSR 与 CrossEntropy, Focal Loss<sup>[24]</sup> 和 GHMC (gradient harmonizing mechanism classification)<sup>[25]</sup> 损失收敛曲线,如图 9 所示。当训练集比例为 0.3 (60) 时,CrossEntropy 达到 97.32%,GHMC 达到 97.32%,LSR 达到 98.04%。在数据集 S2 下,LSR 在 epoch = 72 时最早完成收敛,收敛速度更快,准确率曲线波动较小。

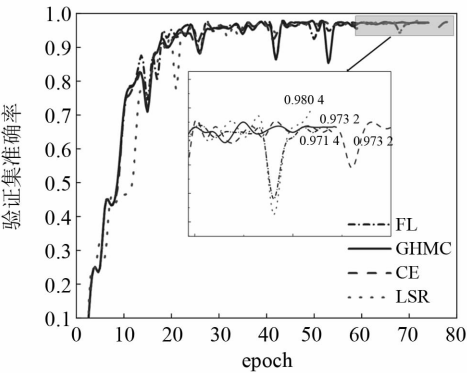


图 9 不同损失函数下 DLWCB 准确率对比图

Fig. 9 Accuracy comparison of under different loss functions

显然,少量训练样本下,LSR 促使模型更快、更好地学习数据特征,具备良好的稳定性和训练效率。

4.7 DLWCB 消融实验、通用性分析

为进一步分析 DLWCB 各部分作用,使用轴承数据集 S1-B,在训练样本比例 0.3 (60) 下与 DCNN, BiGRU, BiLSTM 进行性能对比。对比结果如表 6 所示。

表 6 DLWCB 消融实验  
Tab. 6 DLWCB ablation experiment

| 模型           | 测试平均<br>准确率/% | 模型          | 测试平均<br>准确率/% |
|--------------|---------------|-------------|---------------|
| PCA - SVM    | 78.42         | PCA - LSTM  | 88.33         |
| PCA - BiLSTM | 79.17         | DCNN        | 91.86         |
| PCA - GRU    | 88.50         | PCA - BiGRU | 89.83         |

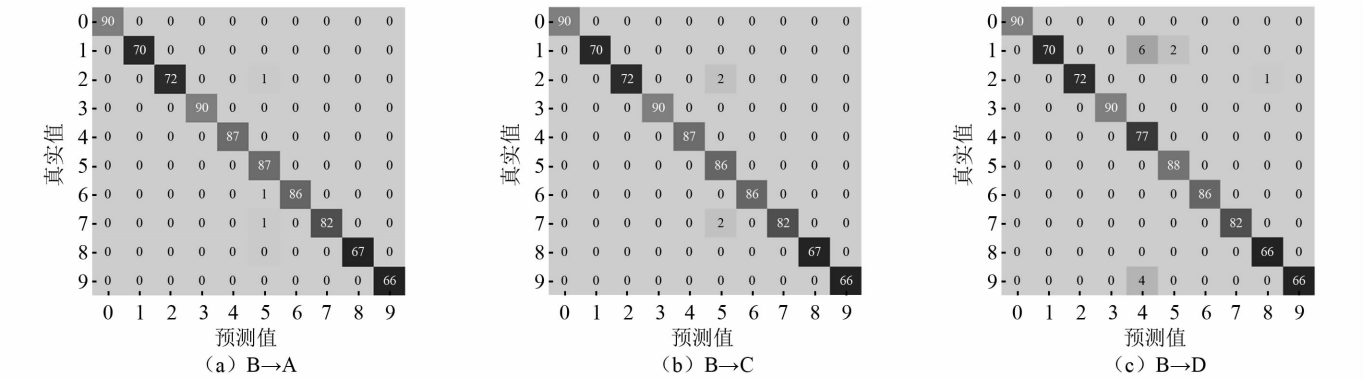


图 10 不同负载下域适应实验

Fig. 10 Domain adaptation under different loads

上述相关实验和分析表明,DLWCB 具有较高的故障识别率和定位表现,可以为轴承实际工作中的故障快速定位诊断和维护提供一定指导。

4.8 可视化分析

为进一步揭示所提方法的特征提取和识别过程,针对轴承数据集 S2,使用 T-SNE 将 DLWCB 提取的特征降至二维。相关结构特征表示,如图 11 所示,不同灰度描述不同状态。图 11(a)~图 11(e)依次为输入振动信号,大卷积核提取的低频特征,小卷积核提取的高频特征,BiGRU 只保存最后一个 GRU 提取的特征,

在对比实验过程中,单层 LSTM (long short-term memor) 和 GRU 处理少量的高维原始故障样本振动信号的效果不理想,故对原始振动信号采用主成分分析 (principal component analysis, PCA) 降维。图 6 中,DLWCB 精度达到 99.86%。表 6 表明,GRU 由于参数较少,相比于 LSTM 提升约 0.17%,BiGRU 相比于 GRU 提升约 1.3%。

另外,BiGRU 使得 DLWCB 学习到振动信号不同隐藏位置特征,准确率提升约 8%。DCNN 学习到少量样本高低频信号,提升约 10%。两者对于少量样本都有不错的性能表现。

同样,为验证 DLWCB 在不同负载下通用性,以数据集 S1-B 为训练集。将其迁移至数据集 S1 下 A, C, D 进行故障诊断,得到混淆矩阵如图 10 所示。

原始振动信号中,负载越低,信号中包含脉冲信息越少;负载越高,则信号包含噪声越多。由图 10 可知,DLWCB 在不同负载下仍具有较高识别率,在 A, C, D 下准确率分别是 99.62%, 99.50%, 98.35%, 说明 DLWCB 具有较好域适应能力。同样,观察到在不同负载下,对于 4, 5, 8 部分样本识别不敏感,还需进一步研究改善,但整体上对正常和故障样本都得到明确区分,各类诊断准确率在 98% 以上。

以及添加 GAP 后 BiGRU 保存各个 GRU 输出特征。

可以发现,图 11(b)、图 11(c) 对部分混乱样本进行初步学习,虽然大卷积核特征分离不明显,但可以有效应对噪声;小卷积核卷积越深,特征分离越明显。经特征融合后,相较于只关注最后一个隐藏 GRU 单元 (图 11(d)), 关注所有 GRU 单元使得 DLWCB 利用了更丰富的信息,提高了特征利用率,更有利于其从少量样本中分离出各类故障 (图 11(e))。这正是设计 GAP 的结果,反映出利用 GAP 针对少量样本故障诊断具备先进性能。

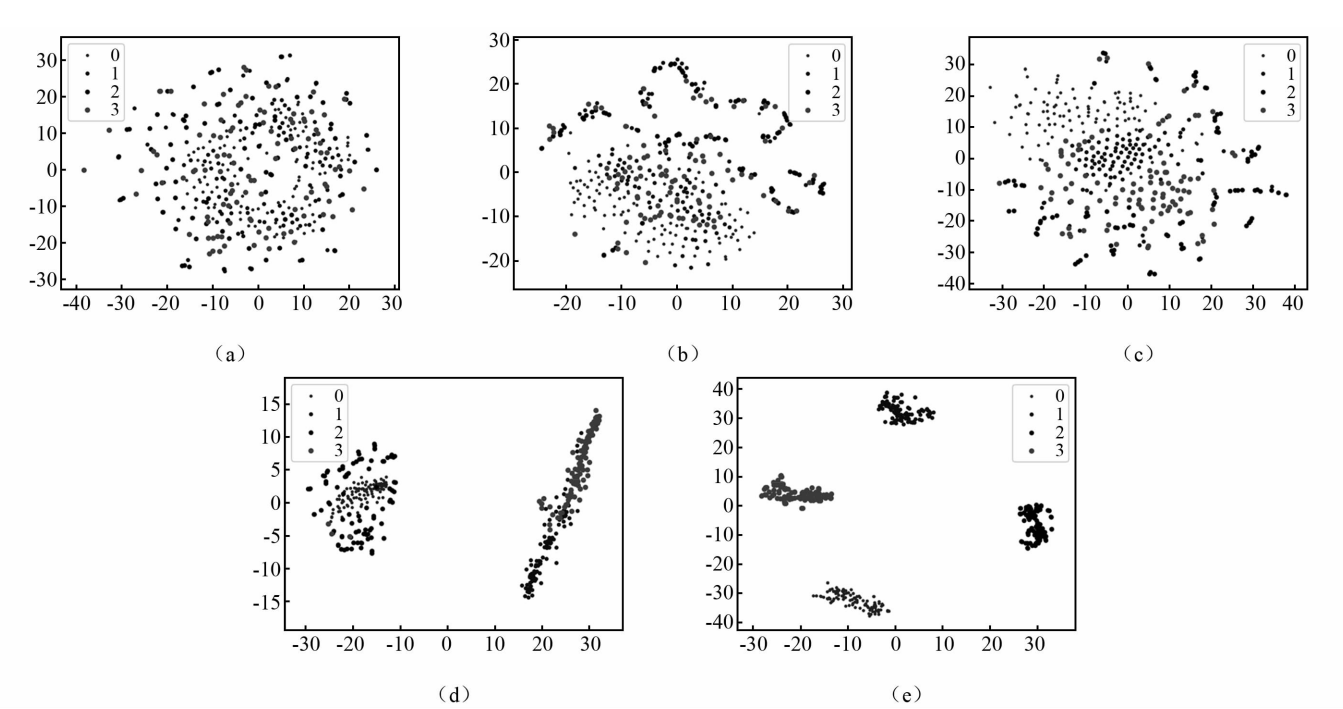


图 11 不同结构输出 T-SNE 可视化图

Fig. 11 Visualization of T-SNE with different structures

综上所述,DLWCB 可以更好地将特征从不同类中分离出来,也暗示 DLWCB 具有较强泛化性能。

4.9 抗噪鲁棒性分析

实际工厂信号多含有噪声,为此,分析不同信噪比条件下,DLWCB 抗噪能力。将高斯白噪声与原始样本混合形成新的含噪声复合样本,如式(13)所示。

$$R_{SN,dB} = 10\lg\left(\frac{P_{signal}}{P_{noise}}\right) \tag{13}$$

式中:  $P_{signal}$  为原始信号功率;  $P_{noise}$  为噪声功率。

与以往研究直接将含噪信号输入模型不同,DLWCB训练和测试均使用图 12 所示的故障诊断框架。此框架由离线预训练和在线测试组成。离线框架用于训练 DLWCB 所需预训练参数,在线框架主要用于泛化性测试和快速训练以应对噪声样本的故障诊断。

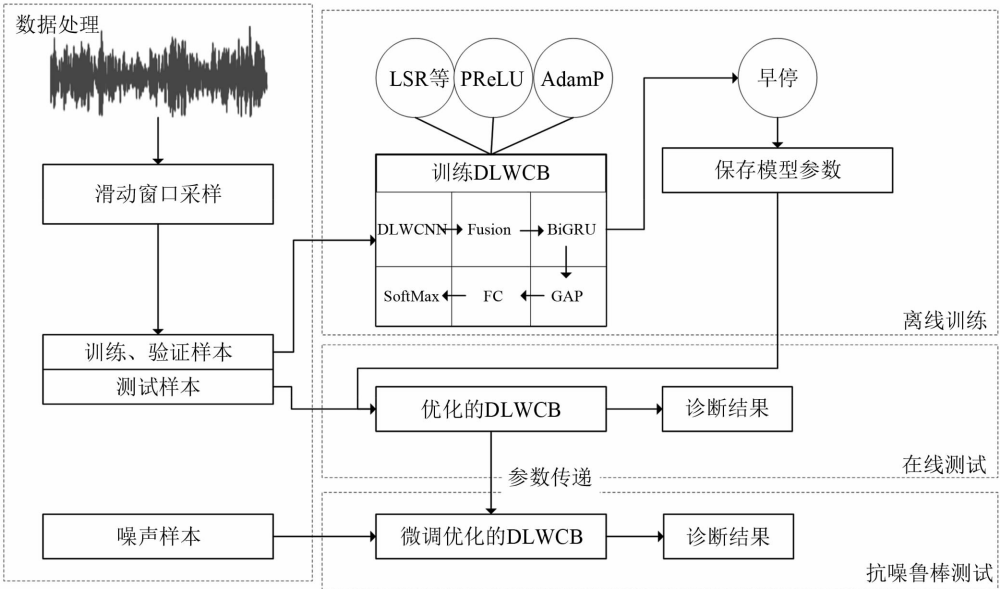


图 12 基于参数传递的 DLWCB 故障诊断框架

Fig. 12 DLWCB fault diagnosis framework based on parameter transfer

实验将对数据集 S1-B 和 S2 下的预训练参数进行微调。由于权重参数接近,这将减少 DLWCB 训练时间,提高诊断效率,以便快速应对各种噪声,实现故障

定位。

将信噪比 (signal-noise ratio, SNR) = -4 ~ 10 dB 的高斯白噪声添加到原始信号中,其他设置一致,加载训



练比例为 0.3(60)下预训练参数,使用相同比例含噪样本微调 DLWCB<sup>[26]</sup>。

针对轴承数据集 S1-B,表 7 中,当 SNR≥2 dB 时,DLWCB 取得较高的诊断精度,训练时间在 10~40 s。与图 7 比较,诊断时间缩短约 1/2。随着 SNR 降低,信号中噪声占比越来越高,对 DLWCB 的鲁棒性提出更

高的要求,诊断精度有所降低,但依然达到 90% 左右。一方面是由于 Laplace 小波卷积和大卷积核对噪声强抗性,过滤部分噪声并有效地提取信号特征;另一方面则是 BiGRU 提取少量噪声样本下不同位置隐藏信息,并通过 GAP 使得 DLWCB 关注到更多有用信息。

表 7 S1 下 DLWCB 不同 SNR 抗噪声能力  
Tab. 7 Anti-noise capability of different SNR of DLWCB in S1

| 预训练参数       |            | SNR/dB | -4    | -2    | 0     | 2     | 4      | 6      | 8      | 10     |
|-------------|------------|--------|-------|-------|-------|-------|--------|--------|--------|--------|
| 加载预训练参数(E)  | 平均测试集准确率/% |        | 89.83 | 94.71 | 97.57 | 99.14 | 99.21  | 99.57  | 99.43  | 99.79  |
|             | 训练时间/s     |        | 28.01 | 17.41 | 26.21 | 39.66 | 12.50  | 14.01  | 12.34  | 34.52  |
| 未加载预训练参数(F) | 平均测试集准确率/% |        | 84.36 | 90.29 | 93.57 | 96.14 | 98.14  | 98.93  | 99.00  | 99.00  |
|             | 训练时间/s     |        | 45.68 | 29.90 | 48.24 | 93.36 | 123.82 | 164.09 | 138.36 | 150.57 |

与 F 方案相比,加载预训练参数的 E 方案对含噪样本诊断效率取得提升。信噪比越高,F 所需诊断时间越长,但是 E 仅仅 35 s 就可以完成诊断;并且在各种 SNR 下,E 都取得约 1%~5% 领先,可以看出在噪声环境下,此方法可以取得较高的效率。

此外,针对数据集 S2 也做相同实验,数据处理方法与 S1 相同,也是利用含噪声样本微调。以进一步证明所提方法的泛化性和可靠性。除学习率变为 0.000 4 外,其他设置相同,得到相关结果如表 8 所示。

表 8 S2 下 DLWCB 不同 SNR 抗噪声能力  
Tab. 8 Anti-noise capability of different SNR of DLWCB in S2

| 预训练参数       |            | SNR/dB | -4    | -2    | 0     | 2     | 4     | 6     | 8     | 10     |
|-------------|------------|--------|-------|-------|-------|-------|-------|-------|-------|--------|
| 加载预训练参数(E)  | 平均测试集准确率/% |        | 81.59 | 87.89 | 93.02 | 93.89 | 95.91 | 96.24 | 97.56 | 98.21  |
|             | 训练时间/s     |        | 31.63 | 42.88 | 13.65 | 25.29 | 15.76 | 18.13 | 35.97 | 12.33  |
| 未加载预训练参数(F) | 平均测试集准确率/% |        | 59.31 | 63.15 | 75.45 | 81.92 | 92.57 | 93.81 | 94.99 | 96.31  |
|             | 训练时间/s     |        | 49.87 | 74.35 | 82.84 | 84.31 | 79.81 | 65.48 | 57.81 | 140.22 |

由表 8 可知,随着信噪比增加,DLWCB 的准确率逐渐提升,其在干扰场景下也具有相当高的性能表现,最高达到 98.21%,表明了该方法具备良好的鲁棒性。同样可以看到,加载预训练参数在时间和准确率上都取得提升。以上两种数据集的实验充分说明 E 方案具备较高诊断效率。

5 结 论

针对少量样本的故障诊断问题,提出一种端到端融合 Laplace 小波卷积的 DLWCB 故障诊断模型,无需额外降噪算法,便可实现高效故障诊断。首先,提出 Laplace 小波卷积核并评估其性能。其次,通过 LSR 多重目标函数、引入 AdamP、PReLU 等在故障诊断领域不常用正则化进一步提高泛化性。最后,使用可视化手段初步理解 DLWCB。实验主要从准确率和时间复杂度等方面验证该方法的可靠性,表明在少量样本下,DLWCB 取得优势,具备一定应用价值。

然而,各个故障类别训练样本数量平衡,在实践中,可能会遇到类别不平衡的数据集。针对少量不平衡数据集,在未来的工作中,考虑从生成对抗网络、时

间卷积网络、注意力机制等拓扑结构或者元学习、小样本学习、迁移学习、集成学习等机器学习技术方面入手扩展该方法的应用,探究它们的性能表现。

参 考 文 献

[ 1 ] JIAO J Y, ZHAO M, LIN J, et al. A comprehensive review on convolutional neural network in machine fault diagnosis [J]. Neurocomputing, 2020,417: 36-63.

[ 2 ] LUO H, He C, ZHOU J N, et al. Rolling bearing sub-health recognition via extreme learning machine based on deep belief network optimized by improved fireworks[J]. IEEE Access, 2021,9: 42013-42026.

[ 3 ] KE Y, YAO C, SONG E Z, et al. An early fault diagnosis method of common-rail injector based on improved CYCBD and hierarchical fluctuation dispersion entropy [J]. Digital Signal Processing, 2021,114: 103049.

[ 4 ] 何勇,王红,谷穗. 一种基于遗传算法的 VMD 参数优化轴承故障诊断新方法[J]. 振动与冲击, 2021,40(6): 184-189.

HE Yong, WANG Hong, GU Sui. New fault diagnosis approach for bearings based on parameter optimized VMD and genetic algorithm[J]. Journal of Vibration and Shock, 2021, 40(6): 184-189.

[ 5 ] ZHANG S, ZHANG S B, WANG B N, et al. Deep learning

- algorithms for bearing fault diagnostics: a comprehensive review[J]. IEEE Access, 2020,8: 29857 – 29881.
- [6] HAN T, MA R Y, ZHENG J G. Combination bidirectional long short-term memory and capsule network for rotating machinery fault diagnosis [J]. Measurement, 2021, 176: 109208.
- [7] SAUFI S R, AHMAD Z A B, LEONG M S, et al. Gearbox fault diagnosis using a deep learning model with limited data sample[J]. IEEE Transactions on Industrial Informatics, 2020,16(10): 6263 – 6271.
- [8] LI C Z, YANG K H, TANG H C, et al. Fault diagnosis for rolling bearings of a freight train under limited fault data: few-shot learning method [J]. Journal of Transportation Engineering, Part A: Systems, 2021,147(8): 04021041.
- [9] ZHANG A S, LI S B, CUI Y X, et al. Limited data rolling bearing fault diagnosis with few-shot learning [J]. IEEE Access, 2019,7: 110895 – 110904.
- [10] WANG C J, XU Z L. An intelligent fault diagnosis model based on deep neural network for few-shot fault diagnosis[J]. Neurocomputing, 2021,456: 550 – 562.
- [11] WU J Y, ZHAO Z B, SUN C, et al. Few-shot transfer learning for intelligent fault diagnosis of machine [J]. Measurement, 2020,166: 108202.
- [12] 沈涛,李舜酩. 针对滚动轴承故障的批标准化 CNN – LSTM 诊断方法题名[EB/OL]. (2021 – 03 – 31). <http://kns.cnki.net/kcms/detail/11.5946.TP.20210330.1708.003.html>.
- [13] YANG J X, YANG F, ZHOU Y X, et al. A data-driven structural damage detection framework based on parallel convolutional neural network and bidirectional gated recurrent unit[J]. Information Sciences, 2021,566: 103 – 117.
- [14] GU J D, TRESP V, HU H. Capsule network is not more robust than convolutional network [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [s.l.]: CVPR, 2021.
- [15] LI T F, ZHAO Z B, SUN C, et al. WaveletKernelNet: an interpretable deep neural network for industrial intelligent diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics:Systems, 2022,52(4): 2302 – 2312.
- [16] 邓飞跃,强亚文,郝如江,等. 基于自适应 Morlet 小波参数字典设计的微弱故障检测方法研究[J]. 振动与冲击, 2021,40(8): 187 – 193.
- DENG Feiyue, QIANG Yawen, HAO Rujiang, et al. A study on a weak fault detection method based on adaptive parametric dictionary design using the Morlet wavelet [J]. Journal of Vibration and Shock, 2021,40(8): 187 – 193.
- [17] CHEN X H, ZHANG B K, GAO D. Bearing fault diagnosis base on multi-scale CNN and LSTM model [J]. Journal of Intelligent Manufacturing, 2021,32(4): 971 – 987.
- [18] QIAO M Y, YAN S H, TANG X X, et al. Deep convolutional and LSTM recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads [J]. IEEE Access, 2020,8: 66257 – 66269.
- [19] HEO B, SANGHYUK C, SEONG J O, et al. AdamP: slowing down the slowdown for momentum optimizers on scale-invariant weights [C]//9th International Conference on Learning Representations. [s.l.]: ICLR, 2021.
- [20] SHEN Z Q, LIU Z C, XU D J, et al. Is label smoothing truly incompatible with knowledge distillation: an empirical study [C]//9th International Conference on Learning Representations. [s.l.]: ICLR, 2021.
- [21] Case Western Reserve University (CWRU). Bearing data center[EB/OL]. [2022 – 09 – 27]. <https://csegroups.case.edu/bearingdatacenter/pages/download-data-file/>.
- [22] ZHANG W, PENG G L, LI C H, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals [J]. Sensors, 2017,17(2): 425.
- [23] ZHAO Z B, LI T F, WU J Y, et al. Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study[J]. ISA Transactions, 2020, 107: 224 – 255.
- [24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 42 ( 2 ): 2999 – 3007.
- [25] LI B Y, LIU Y, WANG X G, et al. Gradient harmonized single-stage detector [C]//AAAI Conference on Artificial Intelligence. Honolulu: AAAI, 2019.
- [26] DONG Y J, LI Y Q, ZHENG H L, et al. A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: solving the small sample problem [J]. ISA Transactions, 2021, 121: 321 – 348.