

基于GRM-IConvNeXt模型的滚动轴承故障诊断方法

罗亨发, 于天壮, 周世华
(东北大学 机械工程与自动化学院, 辽宁 沈阳 110819)

摘要: 针对复杂的轴承运行工况以及传统深度学习故障诊断方法泛化能力弱、模型识别准确率不高等问题,提出了一种基于GRM-IConvNeXt模型的滚动轴承故障诊断方法.首先,提出了一种全局关系矩阵(global relationship matrix, GRM)的编码方法,利用其保留原始信号特征的优点将一维振动信号转换为二维图像.然后,构造了一个针对轴承故障诊断小样本分类的改进ConvNeXt(improved ConvNeXt, IConvNeXt)模型,并选用大小为 5×5 的卷积核和多个BN层与Hardswish激活函数以强化特征提取性能,同时通过CBAM机制根据GRM图像特征自适应地生成权重.实验结果表明,GRM-IConvNeXt模型在变工况和小样本的情况下都具有良好的特征提取能力和泛化性.

关键词: 滚动轴承;全局关系矩阵;IConvNeXt模型;CBAM;故障诊断

中图分类号: U 666.1 **文献标志码:** A **文章编号:** 1005-3026(2025)05-0062-09

A Fault Diagnosis Method of Rolling Bearings Based on GRM-IConvNeXt Model

LUO Heng-fa, YU Tian-zhuang, ZHOU Shi-hua

(School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China. Corresponding author: ZHOU Shi-hua, E-mail: zhou_shihua@126.com)

Abstract: Aiming at the problems of complex bearing operating conditions, weak generalization ability and low accuracy of model recognition of traditional deep learning fault diagnosis methods, a rolling bearing fault diagnosis method based on the GRM-IConvNeXt model is established. Firstly, a coding method of global relationship matrix (GRM) is proposed, which can transform one-dimensional vibration signals into two-dimensional images by taking the advantage of preserving the original signal features. Then, an improved ConvNeXt (IConvNeXt) model for small sample classification of bearing fault diagnosis is constructed, and a convolution kernel with a size of 5×5 , multiple BN layers and Hardswish activation function are selected to enhance the feature extraction performance. At the same time, weights are adaptively generated according to the GRM image features through the CBAM(convolutional block attention module) mechanism. The experimental results show that the GRM-IConvNeXt model has good feature extraction ability and generalization under off-design conditions and small samples.

Key words: rolling bearing; global relationship matrix (GRM); IConvNeXt model; CBAM (convolutional block attention module); fault diagnosis

随着智能制造技术的快速发展,旋转机械设备正逐步向智能化、先进化、轻量化方向迈进.滚动轴承作为旋转机械中的关键部件之一,其运行

状态直接决定设备能否正常运行^[1].一旦滚动轴承发生故障,可能会引发巨大的经济损失,并威胁到操作人员的生命安全^[2].

收稿日期: 2023-11-01

基金项目: 中央高校基本科研业务费专项资金资助项目(N2303011); 辽宁省自然科学基金资助项目(2022-MS-125).

作者简介: 罗亨发(1998—),男,江西赣州人,东北大学硕士研究生;周世华(1987—),男,河北衡水人,东北大学副教授,博士生导师.

目前,滚动轴承的故障诊断主要依赖于对振动信号的处理和分析,基于信号处理的诊断方法和智能诊断方法是当下主流的故障诊断方法.经验模态分解^[3]、变分模态分解^[4]和小波变换^[5]等传统的信号处理方法可以有效地从原始振动信号中提取到故障特征,但是这些方法过于依赖于人工经验.随着技术的发展,Wu等^[6]采用了一种基于连续小波变换和支持向量机的滚动轴承故障诊断方法,首先提出了一种小波选择准则来选择合适的小波进行特征提取,然后利用支持向量机对特征进行分类,通过凯斯西储大学数据集验证其具有更高的诊断准确性和稳定性.这类机器学习模型可以在一定程度上对故障进行识别,但在自动提取深层特征方面仍然存在不足.

近年来,深度学习由于其强大的特征提取能力被广泛应用于故障诊断领域.Wang等^[7]根据轴承振动信号的一维特性提出了一种多注意力一维卷积神经网络(convolutional neural network, CNN)以增强故障相关特征.Guo等^[8]提出了一种多标签一维生成对抗网络,从而达到智能判别特征的目的.然而深度学习模型仅仅依靠单通道信息或一维特征作为输入可能无法充分提取特征,进而导致故障识别精度不高.为了解决这一问题,往往将一维振动信号转换为二维图像,从而把图像分类的方法引入到故障诊断领域.Kaya等^[9]采用连续小波变换方法获得轴承振动信号的时频图,并将时频图输入到CNN中进行训练.Liu等^[10]利用小波包分解获得的初始特征通过循环图编码转换为二维图像.He等^[11]将振动信号和声音信号融合,利用马尔可夫转移场编码生成二维图像并输入到改进的ResNet模型中,可以在复杂工况中有效提取故障特征.

上述方法已经被证明可以解决装备的智能诊断问题,但其采用的模型存在特征提取能力较差、参数量较多等问题,且采用的二维图像编码方法比较依赖专家经验.

本文提出了一种基于GRM-ICovNeXt模型的智能故障诊断方法.首先,提出的GRM编码方法能够在不依赖专家知识的前提下更好地捕捉到不同时间点之间的振动变化和全局关系.其次,提出的ICovNeXt模型利用了深度可分离卷积、先进的归一化层与激活函数的设置方法^[12],具有较好的特征提取能力和更少的参数.最后,

通过凯斯西储大学的轴承故障数据集^[13]和旋转机械故障模拟实验平台数据集^[14]对所提出的方法进行验证.试验结果证明了所提方法的可行性和优越性.

1 理论介绍

1.1 全局关系矩阵

本文提出了一种全局关系矩阵的时序数据特征表示方法,包含了原始时间序列的冗特征,更好地捕捉到不同时间点之间的振动变化和全局关系,同时提高了信号的稳定性.对于一个振动信号 $\mathbf{X}=(x_t, t=1, 2, \dots, N)$,可以通过以下步骤得到GRM图.

1) 在对振动信号的实际应用中通常采样点数较大,需要在进行GRM编码之前将振动信号的维数降低.分段聚合近似(PAA)的方法可以在保持振动信号基本趋势的同时也能保持较低的时间复杂度,所以将选择PAA算法对轴承振动信号进行预处理.PAA算法使用平均值来平滑数据具有一定的降噪效果,其中的缩减因子 r 的选择很重要, r 过大会令平滑后的结果丢失原始结构信息,而 r 太小则会导致降噪效果不明显.所以需要选择一个合适的缩减因子 r ,生成一个新的平滑时间序列,将维度 n 减少到 m ,公式如下:

$$\tilde{x}_i = \begin{cases} \frac{1}{r} \sum_{j=k \cdot (i-1)+1}^{k \cdot i} z_j, i=1, 2, \dots, m; \\ \frac{1}{r} \sum_{j=k \cdot (i-1)+1}^{k \cdot i} z_j, i=1, 2, \dots, m-1; \\ \frac{1}{N-r \cdot (m-1)} \sum_{j=k \cdot (i-1)+1}^{k \cdot i} z_j, i=m. \end{cases} \quad (1)$$

$$m = \left\lceil \frac{N}{r} \right\rceil. \quad (2)$$

其中 $\lceil \cdot \rceil$ 表示取整.通过计算分段常数的平均值进行降维,可以保持原始振动信号的近似趋势,最终新的平滑时间序列 \mathbf{X} 的长度为 m .

2) 构造一个 $m \times m$ 矩阵,捕捉到振动信号中的对应关系,将预处理后的振动信号转换为二维矩阵,依次将时序数据 \mathbf{X} 上的第 i 个时间戳对应的元素 x_i 作为参考点进行转换.矩阵 \mathbf{M} 中的每一个元素相互关联,其中的每一行每一列都是以某个时间戳为参考点,对角线元素值保留了原有的正负关系,即包含了时间序列数据内在相关关系.计算矩阵如下:

$$M = \begin{bmatrix} \text{sign}(\tilde{x}_1) & \tilde{x}_2 \times \tilde{x}_1 & \cdots & \tilde{x}_m \times \tilde{x}_1 \\ \tilde{x}_1 \times \tilde{x}_2 & \text{sign}(\tilde{x}_2) & \cdots & \tilde{x}_m \times \tilde{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_1 \times \tilde{x}_m & \tilde{x}_2 \times \tilde{x}_m & \cdots & \text{sign}(\tilde{x}_m) \end{bmatrix} \quad (3)$$

其中 sign 函数表示当 $x > 0$ 时, $\text{sign}(x)$ 为 1, 当 $x = 0$ 时, $\text{sign}(x)$ 为 0, 当 $x < 0$ 时, $\text{sign}(x)$ 为 -1. 全局关系矩阵可以从一维振动信号中提取更多的关键特征信息, 通过将信号转换为全局关系矩阵图, 可以更好地捕捉到不同时间点之间的振动变化和全局关系, 这些信息可能包含了轴承故障的特征.

3) 最后利用 \min - \max 归一化将矩阵 M 转换为灰度值矩阵, 最终得到灰度值矩阵 F , 得到最终的特征表示:

$$F = \frac{M - \min(M)}{\max(M) - \min(M)} \times 255. \quad (4)$$

通过上述的步骤将轴承振动信号编码为二维图像的一系列操作, 对数据集进行特征表示从而获得生成的图像.

将原始振动信号直接输入到 PAA 算法中以保留原始时间序列的尺度和分布信息, 有助于更准确地诊断故障. 用符号函数去表达全局关系矩阵的对角线元素并保留了其对应的正负关系, 准确地表达了每个时间戳相对于自身的全局位置关系, 同时保持了原始数据的符号信息. 这对于捕捉时间序列中的周期性或振荡特征非常有用, 特别是在轴承故障诊断中可能出现的情况.

在轴承故障诊断中, 将各个元素相乘, 强调了时间序列数据中每对时间戳之间的相对关系, 这有助于更准确地捕捉振动信号中的关键模式和振动特征. 元素相乘的方式可以更好地突出时间序列中的重要变化点. 当相邻时间戳之间的振动特征发生显著变化时, 元素相乘会产生更大的变化, 从而使这些关键点更容易被识别, 这对于检测轴承故障的早期迹象非常有用. 通过使用元素相乘的方式构建全局关系矩阵, 强调了时间序列数据中的相对关系, 突出了变化点、振动幅度并增强了非线性特征, 同时提高了信号的稳定性, 可以更好地捕捉和分析复杂的振动模式, 提高故障检测的精确性和可靠性.

1.2 卷积神经网络模型 IConvNeXt

本文构建的 IConvNeXt 模型是以 ConvNeXt 网络为主体结构进行改进的卷积神经网络模型. ConvNeXt 网络是一种基于 ResNet 结构的网络模型, 从宏观设计、大卷积核、逆瓶颈层 (ResNeXt)^[15]、深度可分离卷积 (MobileNet V2)^[16] 和细节设计这

5 个角度依次借鉴了自注意力机制模型 Transformer 的架构和策略, 提供了比 Swin Transformer 更好的性能和更直接的模型结构.

ConvNeXt 模型在模型架构方面把 ResNet50 各个阶段堆叠 block 的次数由 (3, 4, 6, 3) 改为 (3, 3, 9, 3), 同时用卷积核大小为 4×4 步距为 4 的卷积层来代替原来 ResNet50 的下采样模块 stem 层. 另外, 为了降低计算量和获得更大的感受野, 采用 MobileNet 中的深度卷积并将其卷积核大小从 3×3 改为 7×7 . 使用 Transformer 中的反向瓶颈设计替代了 ResNet50 网络模型的瓶颈结构, 并在反向瓶颈结构中将深度卷积的位置向上移动一层以适应更大的卷积核. 为了加速神经网络的训练以及提升收敛效果, ConvNeXt 模型使用 Transformer 模型中的 LayerNorm 层替代 BatchNorm 层, 减少了网络中归一化层和激活函数的数量, 同时将采样层从网络结构中分离出来.

为强化现有网络模型的特征提取性能和降低计算量, IConvNeXt 模型中选择将 block 的堆叠次数由 (3, 3, 9, 3) 调整为 (1, 1, 3, 1), 并选用深度可分离卷积同时将 5×5 的卷积核替代原有的 7×7 卷积核, IConvNeXt block 的结构如图 1 所示, 其中, k 为卷积核尺寸, s 为步长, p 为填充.

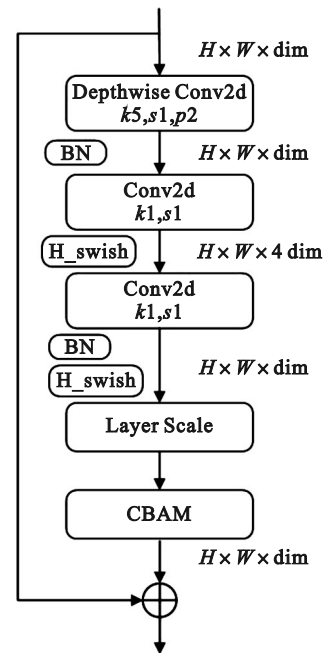


图 1 IConvNeXt 块

Fig. 1 IConvNeXt block

在百万级图像数据 ImageNet 分类任务中, 使用大卷积核如 7×7 可以扩大感受野并减少计算量和提高计算速度, 但在如滚动轴承故障诊断的这类小样本分类任务中, 使用 7×7 卷积核会导致网

络训练过程的参数增多和网络模型深度降低,从而影响深度卷积层的特征提取性能.因此,本模型选用 5×5 的卷积核来增强特征提取效率,在减少计算量和参数量的同时提升深度卷积层的深度和非线性表达能力.

将多个批量标准化(batch normalization)和激活函数 Hardswish 加入到 block 中,激活函数 Hardswish 公式如下:

$$\text{Hardswish}(x) = \begin{cases} 0 & x \leq -3; \\ x & x \geq 3; \\ x \cdot (x+3)/6 & \text{其他.} \end{cases} \quad (5)$$

Swish 是一种具有平滑、非单调等特点的激活函数,可使神经网络层具有更丰富的表现能力. Hardswish 是在 MobileNet V3 中提出的具有

Swish 函数的特性且具有数值稳定性好、计算速度快等优点的激活函数.通过引入 BN 层,可以有效地加速神经网络的训练和收敛过程并提高网络特征的提取效率.此外,与原有单一的 GELU 激活函数相比,引入多个 Hardswish 激活函数可以更好地防止模型在训练过程中发生梯度弥散问题,进一步增强卷积神经网络的非线性表达能力,从而更准确地从故障特征中提取有效信息.

同时在模型中引入 CBAM (convolutional block attention module), CBAM 是一种结合通道和空间的注意力模块,可以增强特征图中的有用特征并抑制无用的特征,从而提高网络对隐式重要特征的提取能力.同时 CBAM 是一种即插即用的简单注意力模块,图 2 为 CBAM 的示意图.

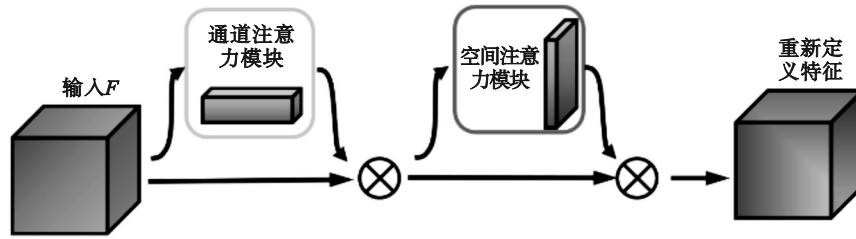


图2 CBAM

Fig. 2 CBAM

如图 2 所示,双通道注意力模块通过通道和空间 2 个注意力模块来计算注意力权重图并对特征图进行重新定义.计算公式如下:

$$\begin{aligned} M_c(F) &= \sigma(\text{MLP}(\text{AvgPool}(F)) + \\ &\quad \text{MLP}(\text{MaxPool}(F))) = \\ &\quad \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))). \end{aligned} \quad (6)$$

其中: F 表示输入特征; σ 表示 sigmoid 激活函数; MLP 表示多层感知机; AvgPool 和 MaxPool 分别表示平均池化操作和最大池化操作; W_1 和 W_0 表示 2 个分支的全连接层权重矩阵; F_{avg}^c 和 F_{max}^c 分别表示在特定尺度 c 下经过平均池化和最大池化后的特征向量.输入的特征图首先通过通道注意力模块进行处理,利用模块中的全局最大池化和全局平均池化提取信号特征图中的高层次特征,然后通过多层感知器 MLP 把 2 个池化层输出的高级特征组合到一起,再将输出的特征经过 sigmoid 函数激活操作生成最终输出.

空间注意力模块是更加细化的注意力机制,把通道注意力模块的输出作为其所需要的输入,然后利用最大池化操作和平均池化操作生成最大池化特征和平均池化特征,而后将其拼接再经过卷积层后得到空间注意力特征图,计算公式如下:

$$\begin{aligned} \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \\ \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])). \end{aligned} \quad (7)$$

其中: $f^{7 \times 7}$ 表示卷积核为 7 的卷积运算; F_{avg}^s 和 F_{max}^s 分别表示在特定尺度 s 下经过平均池化和最大池化后的特征向量.

2 方法流程

基于上述分析,本文提出了一种基于 GRM-ICovNeXt 的滚动轴承故障方法,其流程图如图 3 所示,具体步骤如下:

- 步骤 1 采集滚动轴承各工况下的振动信号;
- 步骤 2 通过 GRM 编码技术将截取的振动信号二维图像化;
- 步骤 3 按照 8:2 的比例划分训练集与测试集,并将训练集作为输入训练 ICovNeXt 网络模型;
- 步骤 4 通过测试集验证该预训练模型的性能,由分类器输出滚动轴承故障状态分类结果,从而完成故障诊断任务.

3 实验验证

实验均是基于 PyTorch 深度学习框架,在运行内存为 16 GB 的 NVIDIA-RTX-3060 显卡和英

特尔 i7-12700H 处理器上完成的.通过凯斯西储大学轴承故障数据集和旋转机械故障模拟实验平台数据集的实验数据来验证所提方法的效果.

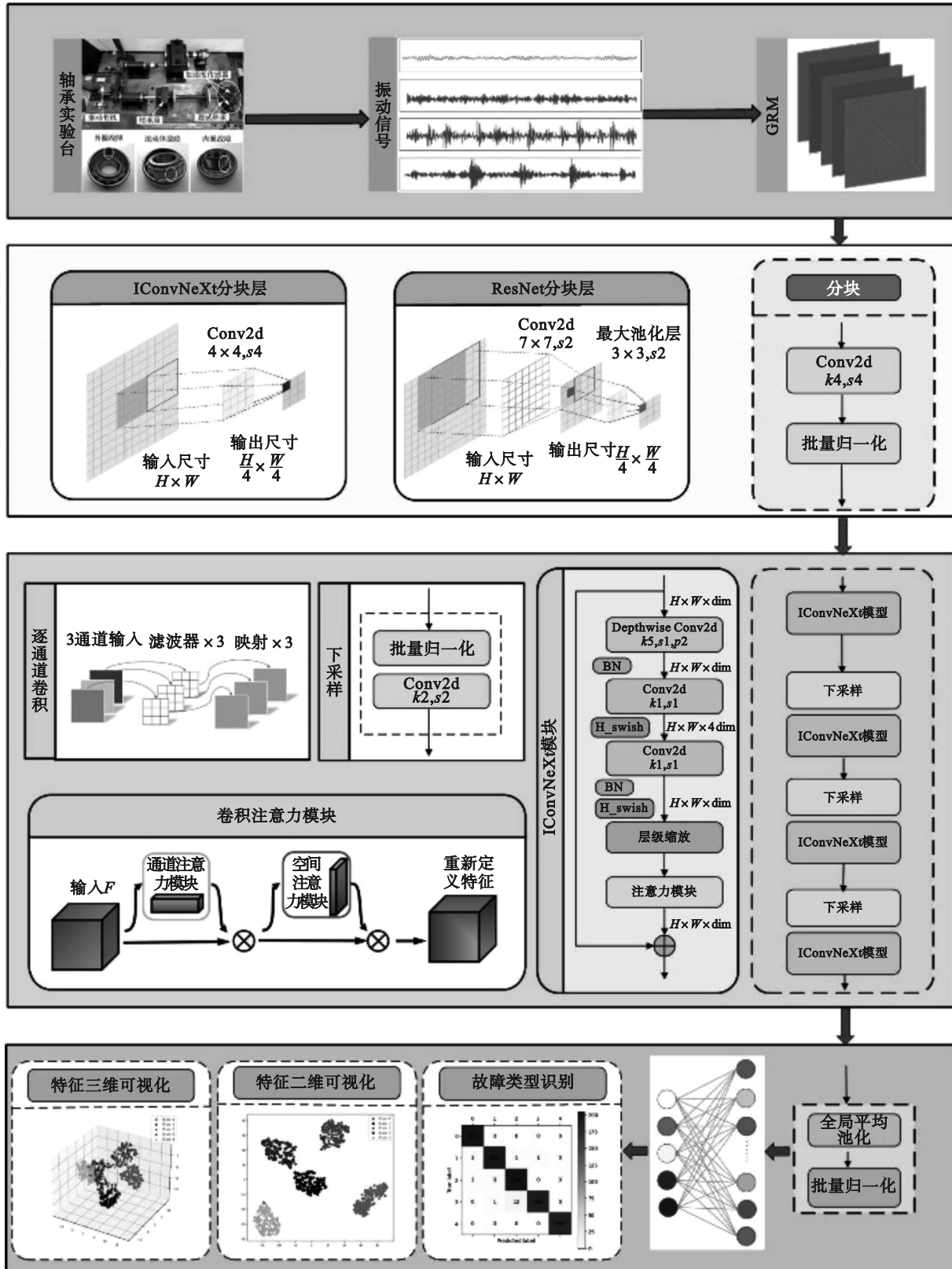


图3 GRM-IConvNeXt故障诊断方法流程图

Fig. 3 Flowchart of GRM-IConvNeXt fault diagnosis method

3.1 凯斯西储大学轴承数据集实验

实验采用 CWRU 滚动轴承数据对本文

IConvNeXt模型进行验证.轴承实验台由电机、功率

试验机、扭转编码器和电子控制器组成.选用驱

动端轴承 SKF6025 的数据进行实验,在轴承内圈、外圈和滚动体分别有人为加工的直径为 0.177 8, 0.355 6, 0.533 4 mm 的单点故障.对 2.24 kW 负载条件下采样频率为 12 kHz,电机转速为 1 730 r/min 的振动数据进行处理,数据包含 9 种故障状态和 1 种健康状态.

为了能够令 CBAM 模块最大程度地提取样本在时序上的相关特征,且保证单个样本中尽可能地包含有一个振动周期内的数据,对每个轴承样本截取 1 000 个样本点进行 GRM 转换,并通过重叠采样的方法将每个状态的数据划分为 500 个样本,10 种状态共计 5 000 个样本,按照 8:2 的比例划分训练集和测试集,样本构造如表 1 所示.

表 1 CWRU 10 种故障类别的数据集状态
Table1 Data set status of ten CWRU fault categories

标签	故障类型	损伤尺寸	训练 样本	测试 样本
		mm		
0	无故障(N)	0	400	100
1	内圈故障(IR)	0.177 8	400	100
2	内圈故障(IR)	0.355 6	400	100
3	内圈故障(IR)	0.533 4	400	100
4	滚动体故障(BR)	0.177 8	400	100
5	滚动体故障(BR)	0.355 6	400	100
6	滚动体故障(BR)	0.533 4	400	100
7	外圈故障(OR)	0.177 8	400	100
8	外圈故障(OR)	0.355 6	400	100
9	外圈故障(OR)	0.533 4	400	100

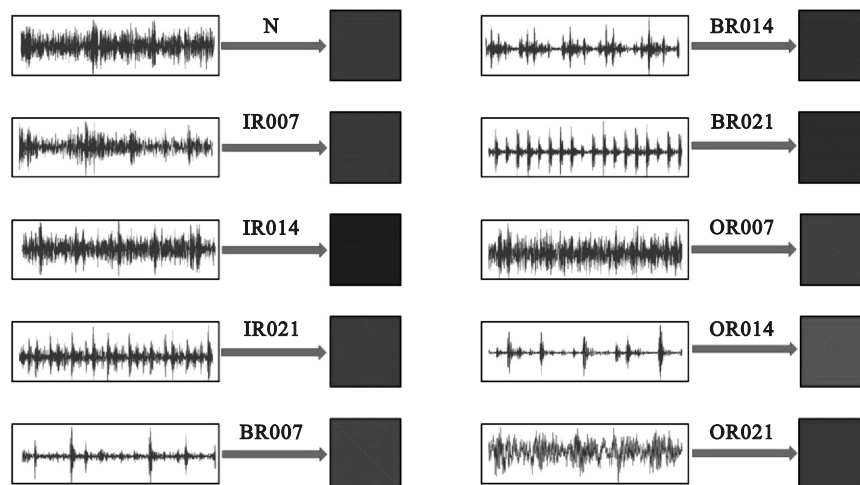


图 4 原始时域信号与 GRM 图像

Fig. 4 Original time domain signals and GRM images

为验证本文方法的有效性,将其与 ConvNeXt, ResNet, MobileNet 进行实验对比分析.为了评估模型的稳定性,将每个模型进行了 10 次实验以避

图 4 所示为 10 种状态下的振动信号通过全局关系矩阵编码后的二维图像,图像分辨率大小为 224×224.在训练阶段模型采用 Adam 算法进行优化, batchsize 设置为 16,学习率为 0.001,训练 epoch 为 30,损失函数为交叉熵损失函数.

图 5 是本文所提出的 GRM-ICovNeXt 模型经过 30 轮训练后的准确率与损失曲线图、混淆矩阵图和 T-SNE 降维算法图.从图 5a 中可以看到 ICovNeXt 模型对滚动轴承故障状态的测试分类结果,分类准确率随着迭代次数的增加在不断提升,当模型训练到第 12 次时训练集和测试集的准确率已经达到 100%,此时测试集的损失函数 Loss 值已降至 0.002 20,模型很快就达到了稳定的收敛状态.

混淆矩阵可以很直观地表示出模型的预测性能,从图 5b 表示的混淆矩阵中可以很清晰地看出所有的测试样本都能被正确分类.为了更加直观地表示本文方法在特征表达上的优势,使用 T-SNE 算法可视化方法将从轴承振动信号中提取到的高维深层抽象特征映射到二维和三维平面进行可视化.如图 5c 特征二维可视化图和图 5d 特征三维可视化图所示,提取出来的 10 种特征分类非常明显,不同故障类别很好地聚拢在一起,不存在误分的情况,分类结果与混淆矩阵得出的结论相符.这充分证明了本文所提的 ICovNeXt 模型可以正确地将各种轴承故障特征分离,具有良好的诊断效果.

免诊断过程中的偶然性,用平均预测准确率和准确率方差对这些模型进行比较和评价,其结果如表 2 所示.

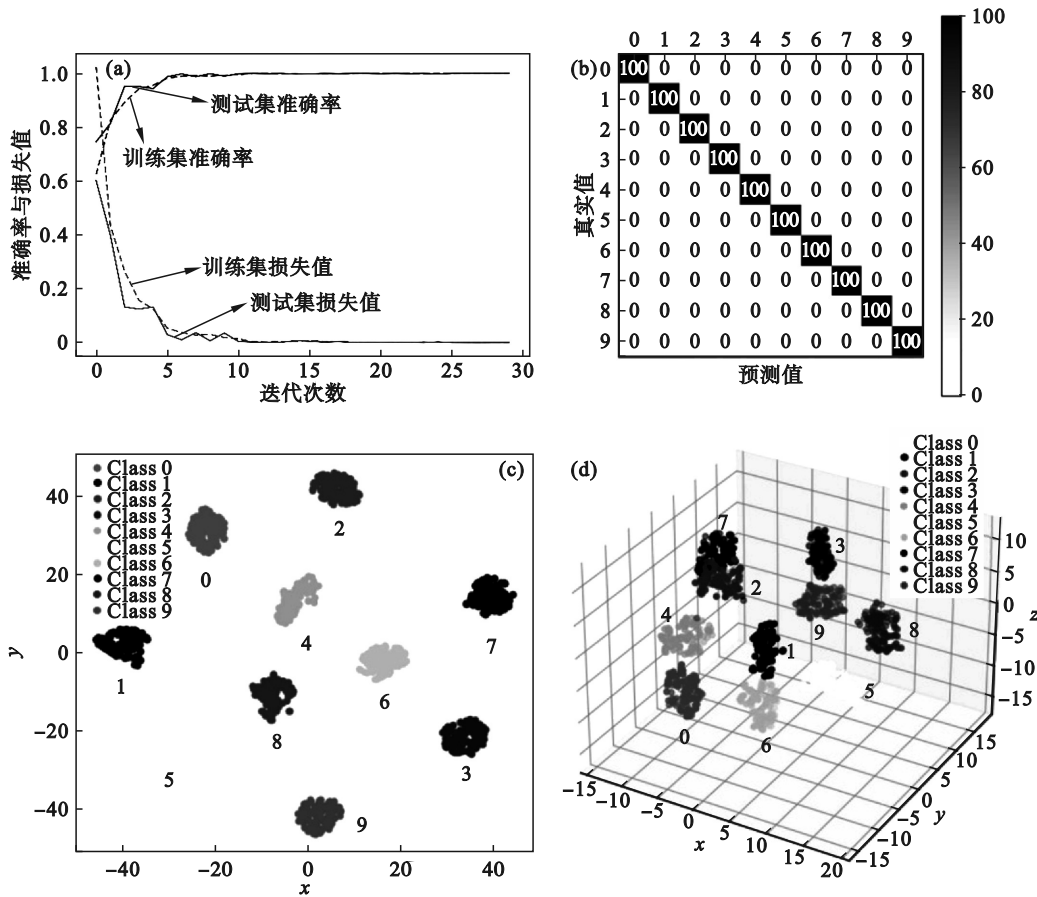


图 5 自适应特征后的输出效果图

Fig. 5 Output renderings after adaptive features

(a)—准确率与损失曲线图; (b)—混淆矩阵图; (c)—特征二维可视化图; (d)—特征三维可视化图.

表 2 准确率对比
Table 2 Comparison of accuracy

网络	准确率/%	方差
GRM+ICovNeXt	100	0.004 2
GRM+ConvNeXt	97.55	0.098 7
GRM+ResNet	95.62	0.206 2
GRM+MobileNet	92.46	0.586 4

ICovNeXt, ConvNeXt, ResNet 和 MobileNet 对 GRM 编码的图像识别准确率分别为 100%, 97.55%, 95.62% 和 92.46%. ICovNeXt 网络模型具有最高的平均识别准确率(100%)和最低的平均预测准确率方差(0.004 2), 表现了其良好的图像特征提取能力和稳定性.

3.2 旋转机械故障模拟实验台数据实验

在旋转机械故障模拟实验台上完成了滚动轴承正常状态和故障状态下在不同转速条件下的故障实验, 采集到包括内圈故障(IR)、外圈故障(OR)、滚动体故障(BR)、正常状态(N)以及外圈和滚动体混合故障(OBR)共 5 种状态类型的实验数据, 为后续故障识别提供数据支持, 其中每种状态都包含了水平和垂直方向的振动信号.

如图 6 所示, 实验平台采用的是 QPZZ- II 旋转机械振动及故障模拟实验台, 由加速度传感器、变速驱动电机、齿轮箱、轴和轴承等组成, 可以快速模拟故障状态.

实验所用的轴承结构参数如表 3 所示, 型号为 NU205EM 圆柱滚子轴承. 通过在内圈、外圈和滚动体上人工切割出裂纹缺陷来模拟相关故障, 以便于分析不同故障的振动特征.

设置好实验台故障模拟系统的参数, 利用信号采集系统和加速度传感器采集振动信号, 采样频率为 20 kHz, 调节电机转速至 600, 1 200, 1 800 r/min 完成轴承各状态下的振动信号数据的采集, 将采集的垂直方向的振动信号按转速分别标记为工况 0, 工况 1 和工况 2 三个数据集.

模型参数设置和样本构造方法与实验一相同, 将 3 个转速下生成的振动信号分别按窗口大小为 1 000 的信号长度进行截取, 而后通过 GRM 编码成大小为 224×224 的二维图像. 生成 5 种状态共计 5 000 个样本, 按照 8:2 的比例划分训练集和测试集, 样本构造如表 4 所示.

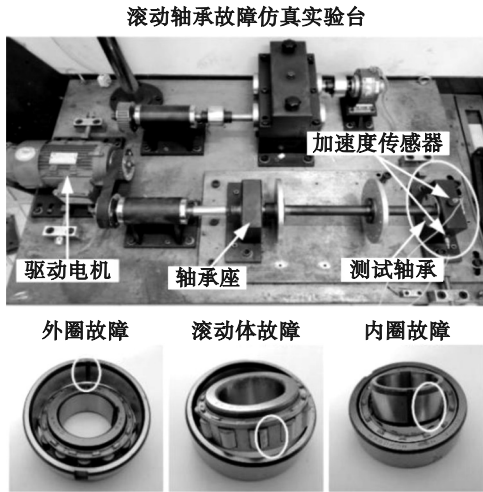


图 6 QPZZ-II 旋转机械振动及故障模拟实验台及滚动轴承部分故障实际效果图

Fig. 6 QPZZ-II rotating machinery vibration and fault simulation test bench and actual effect picture of rolling bearing faults

表 3 滚动轴承结构参数

Table 3 Rolling bearing structure parameters

参数	数值
内圈直径 d_i /mm	25
外圈直径 d_o /mm	52
滚动体直径 d /mm	8
滚动体数目 Z	12
轴承节圆直径 D_m /mm	39.5
轴承宽度/mm	15
接触角/ $^\circ$	0
滚动体有效长度 l_w /mm	10

将训练样本和测试样本输入设计好的

ICovNeXt 模型进行学习训练,每个工况的数据分别进行 10 次实验,每次实验进行 30 轮迭代训练,其结果如图 7 所示.

表 4 实验台 5 种故障数据集

Table 4 Five kinds of fault data sets of the test bench

标签	故障类型	训练样本	测试样本
0	N	800	200
1	IR	800	200
2	OR	800	200
3	BR	800	200
4	OBR	800	200

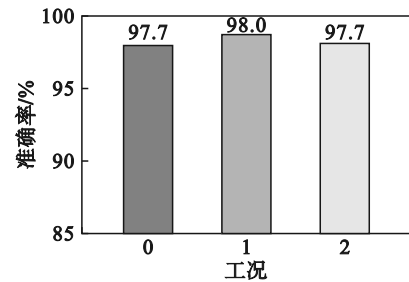


图 7 不同工况下的平均分类准确率

Fig. 7 Accuracy under different working conditions

ICovNeXt 模型对工况 0 测试集的平均分类准确率为 97.7%, 工况 1 测试集的平均分类准确率为 98.0%, 工况 2 测试集的平均分类准确率为 97.7%. 图 8 所示为 3 种工况下 10 次实验中出现结果次数最多的混淆矩阵图, 有少数样本存在误分类的情况, 但整体上分类效果比较稳定, 准确率基本能保持在 97% 以上. 结果表明, 本文所提方法对不同工况下的轴承故障数据都有较高的平均预测精度.

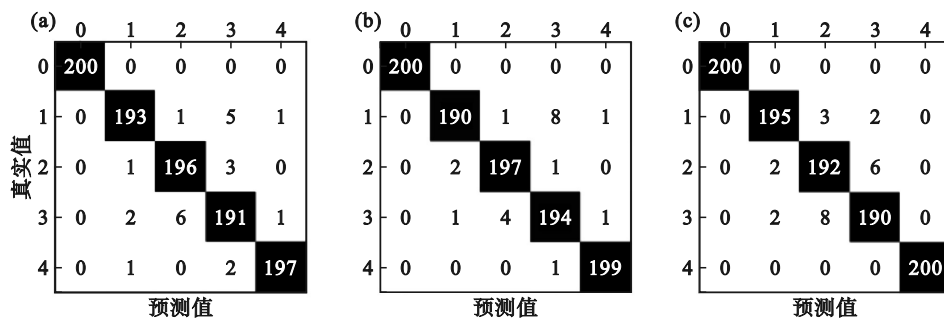


图 8 各工况混淆矩阵图

Fig. 8 Confusion matrix diagram of each working condition

(a) — 工况 0; (b) — 工况 1; (c) — 工况 2.

4 结 论

1) 提出了一种全局关系矩阵的时序数据特征表示方法, 能够快速有效地对滚动轴承的振动

信号进行图像化, 并更好地捕捉到不同时间点之间的振动变化和全局关系, 令 ICovNeXt 模型很好地提取到故障信号间的关联特征.

2) 对 ConvNeXt 模型进行了改进以强化对轴承故障信号的特征提取性能和降低计算量, 引入

CBAM 模块,根据 GRM 编码的图像特征从不同角度和空间维度自适应地生成权重,抑制干扰信息,保留和突出关键故障特征,极大地提高了模型的泛化能力。

3) 通过凯斯西储大学和模拟实验台两个数据集对本文方法进行实验验证,实验结果表明所提方法在变工况和小样本的情况下都具有良好的特征提取能力和泛化性。

参考文献:

- [1] 刘磊,李舜酩,陆建涛.基于卷积神经网络的旋转传动部件故障诊断综述[J].机械设计,2022,39(10):1-8.
(Liu Lei, Li Shun-ming, Lu Jian-tao. Overview of fault diagnosis of rotary transmission parts based on convolution neural network[J]. *Mechanical Design*, 2022, 39(10): 1-8.)
- [2] Guo S, Yang T, Gao W, et al. A novel fault diagnosis method for rotating machinery based on a convolutional neural network[J]. *Sensors*, 2018, 18(5): 1429-1445.
- [3] 杨建华,韩帅,张帅,等.强噪声背景下滚动轴承微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 Jian-hua, Han Shuai, Zhang Shuai, et al. Empirical mode decomposition of weak fault characteristic signal of rolling bearing under strong noise background [J]. *Acta Vibration Engineering*, 2020, 33(3): 582-589.)
- [4] 贾亚飞,朱永利,王刘旺.基于VMD和Wigner-Ville分布的局放信号时频分析[J].系统仿真学报,2018,30(2):569-578.
(Jia Ya-fei, Zhu Yong-li, Wang Liu-wang. Time-frequency analysis of partial discharge signal based on VMD and Wigner-Ville distribution [J]. *Journal of System Simulation*, 2018, 30(2): 569-578.)
- [5] 尹爱军,李海珠,李江,等.Wigner-Ville分布复小波相似性评价及应用[J].振动测试与诊断,2020,40(1):7-11.
(Yin Ai-jun, Li Hai-zhu, Li Jiang, et al. Similarity evaluation and application of complex wavelet of Wigner-Ville distribution [J]. *Vibration Test and Diagnosis*, 2020, 40(1): 7-11.)
- [6] Wu C X, Chen T F, Jiang R, et al. A novel approach to wavelet selection and tree kernel construction for diagnosis of rolling element bearing fault [J]. *Journal of Intelligent Manufacturing*, 2017, 28(8): 1847-1858.
- [7] Wang H, Liu Z L, Peng D D, et al. Understanding and learning discriminant features based on multi-attention 1DCNN for wheelset bearing fault diagnosis [J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(9): 5735-5745.
- [8] Guo Q W, Li Y B, Song Y, et al. Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network [J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(3): 2044-2053.
- [9] Kaya Y, Kuncan F, Ertunc H M. A new automatic bearing fault size diagnosis using time-frequency images of CWT and deep transfer learning methods [J]. *Turkish Journal of Electrical Engineering and Computer Sciences*, 2022, 30(5): 1851-1867.
- [10] Liu X P, Xia L J, Shi J, et al. A fault diagnosis method of rolling bearing based on improved recurrence plot and convolutional neural network [J]. *IEEE Sensors Journal*, 2023, 23(10): 10767-10775.
- [11] He K, Xu Y, Wang Y, et al. Intelligent diagnosis of rolling bearings fault based on multisignal fusion and MTF-ResNet [J]. *Sensors*, 2023, 23(14): 62-81.
- [12] Liu Z, Mao H, Wu C Y, et al. A ConvNet for the 2020s [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, 2022: 11976-11986.
- [13] 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.
- [14] 任朝晖,于天壮,丁东,等.基于VMD-DBN的滚动轴承故障诊断方法[J].东北大学学报(自然科学版),2021,42(8):1-6.
(Ren Zhao-hui, Yu Tian-zhuang, Ding Dong, et al. Fault diagnosis method of rolling bearing based on VMD-DBN [J]. *Journal of Northeastern University(Natural Science)*, 2021, 42(8): 1-6.)
- [15] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks [C]// 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 1492-1500.
- [16] Chen Y H, Yange T J, Emer J S, et al. Eyeriss v2: a flexible accelerator for emerging deep neural networks on mobile devices [J]. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2019, 9(2): 292-308.