

基于时频图像和 CNN 的轴承故障诊断

王国柱^{1,2}, 周睿哲^{1,2}, 黄振³, 池晓航⁴, 张开拓^{1,2}

(1. 河南工学院 电缆工程学院, 河南 新乡 453003; 2. 河南省线缆结构与材料重点实验室, 河南 新乡 453003; 3. 广州恒星传导科技股份有限公司, 广东 广州 510385; 4. 吉林化工学院 信息与控制工程学院, 吉林 吉林 132022)

摘要: 为确保电机的正常运行, 保障企业的有序生产, 提出了一种基于时频图像和 CNN 的轴承故障诊断方法。首先, 对振动信号进行 CWT 处理, 获得时频特征图像; 然后, 对时频图像进行压缩, 并将其输入 CNN 进行训练, 使其能够准确地学习和识别不同类型的时频特征图像; 最后, 将测试时频图像输入 CNN 进行训练, 以验证训练结果的可靠性。结果表明: 所提方法能够有效地检测到轴承故障并识别故障类型, 识别率达到 99.7%。与传统的轴承故障诊断方法相比, 该方法具有更好的泛化能力。

关键词: 故障诊断; 卷积神经网络; 连续小波变换; 时频图像

中图分类号: TP271; TP277

文献标志码: A

文章编号: 2096-7772-(2025)05-0006-08

0 引言

电机作为工业生产中最常用的电器或各种机械设备的动力源, 出故障的情况较为常见, 电机滚动轴承故障占据了约 30% 至 40%^[1-4]。滚动轴承作为机器运转中的关键部件, 其故障不仅会导致生产中断, 还可能引发连锁反应, 影响整个生产过程的稳定。因此, 及时准确地检测和诊断滚动轴承故障, 不仅对提高生产效率至关重要, 还能有效预防事故的发生, 保障生产的安全性和稳定性。同时, 通过有效的故障检测和预测, 企业可以实现设备的合理维护和管理, 延长设备的使用寿命, 降低维修成本, 提高生产效率和经济效益。

传统的电机轴承故障诊断方法主要是通过提取电机的定子电流信号或机壳振动信号作为数据样本, 采用信号处理与分析的方法, 将轴承的故障分量表现在信号图谱中^[5-10]。随着科技的进步和工业自动化水平的提高, 研究人员不断探索实现轴承故障的自动化诊断和智能化分析的新方法和新技术, 主要包括开发先进的传感器技术、机器学习算法、深度学习算法和人工智能模型, 用于实现对轴承运行状态的实时监测和故障预测^[11,12]。随着这些方法的逐渐渗透,

也使故障诊断技术有了质的飞跃。小波变换 (Wavelet Transform, WT) 方法能够精确地捕捉信号中的瞬态和局部特征, 有效分析信号的时域和频域特性。连续小波变换 (Continuous Wavelet Transform, CWT) 以其随机选择小波基函数的特性而著称, 能够有效地捕捉信号的时频特性。卷积神经网络 (Convolutional Neural Networks, CNN) 作为一个多层感知器, 具有局部连接和权值共享的特性, 直接使用图像数据作为输入, 省去了特征提取和重构计算的步骤, 也能够自动提取图像的颜色、纹理、形状和结构等特征, 可以高效地实现对位移不变性、缩放不变性等特征的识别, 在图像识别领域得到了广泛的应用^[13-16]。

本文以凯斯西储大学 (CWRU) 滚动轴承实验室的轴承振动数据为研究对象, 通过 CWT 提取数据中的时频特性, 生成原彩时频图像, 然后将时频图压缩处理并输入 CNN 中进行训练, 最终实现对轴承故障的诊断并识别其故障类型。

1 CWT 与 CNN 理论

1.1 CWT 基本原理

CWT 是一种被广泛采用的数据处理技术, 在信

收稿日期: 2024-10-11

基金项目: 国家自然科学基金项目 (62273242); 河南省科技研发计划联合基金项目 (232103810028); 河南工学院高层次人才科研启动基金项目 (KQ1806)

第一作者简介: 王国柱 (1984—), 男, 河南焦作人, 副教授, 博士, 主要从事基于数据驱动的过程建模、优化及故障诊断研究。

号处理领域具有重要地位^[17], 其定义如下:

假设 $\Psi(t) \in L^2(R)$ ($L^2(R)$ 表示平方可积的实数空间, 即能量有限的信号空间), 其傅里叶变换为 $\Psi(\omega)$, 当 $\Psi(t)$ 满足式 (1) 条件时, 称 $\Psi(t)$ 为一个基本小波函数或者母小波函数。

$$C_\Psi = \int_R \frac{|\Psi(t)|^2}{|\omega|} d\omega < \infty \quad (1)$$

将母小波函数经过伸缩和平移操作可得到一个小波时间序列:

$$\Psi_{a,\tau} = \frac{1}{\sqrt{|a|}} \int \Psi\left(\frac{t-\tau}{a}\right) dt \quad (2)$$

对于任意函数 $f(t) \in L^2(R)$ 的连续小波函数, 其小波变换为:

$$W_f(a, \tau) = \langle f(t), \Psi_{a,\tau}(t) \rangle = \frac{1}{\sqrt{a}} \int f(t) \Psi\left(\frac{t-\tau}{a}\right) dt \quad (3)$$

其中:

$$\Psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \int f(t) \Psi\left(\frac{t-\tau}{a}\right) dt \quad (4)$$

式中: a 为与频率相关的伸缩尺度因子, τ 为平移因子, $\Psi_{a,\tau}(t)$ 为小波基函数, $W_f(a, \tau)$ 为小波变换系数。当 a 越大, 对信号中的低频信号的提取就更加有利, 相反, 当 a 越小时, 小波函数的频率越高, 能更有效地捕捉信号中的高频成分。

1.2 CNN 基本结构

CNN 的基本结构主要由以下几个部分构成: 输入层 (Input Layer)、卷积层 (Convolutional Layer)、池

化层 (Pooling Layer) 以及全连接层 (Fully Connected Layer), 每一层都承担着特定的功能, 共同完成对数据的处理和特征提取。输入层接收原始数据并传递给下一层, 卷积层主要负责提取数据特征, 池化层的作用则是通过减少数据维度和保留关键特征来简化模型, 全连接层完成分类或回归任务。这些部分相互连接, 组成深度学习模型。在 CNN 中, 卷积层、池化层和全连接层构成网络的隐含层, 是核心组成部分。

1.2.1 卷积层

卷积层包括多个特征图, 每个含有多个神经元。这些神经元通过卷积矩阵与上一层特征图的部分区域连接, 实现局部感知和特征提取。卷积矩阵表示卷积核的权重分布 (常见大小为 3×3 或 5×5)。卷积操作在神经网络中起到了至关重要的作用, 通过共享权重和局部连接的方式, 实现对图像等数据的特征提取和抽象。这种局部感知和权重共享的设计使得 CNN 在处理图像等数据时具有较高的效率和泛化能力。

对于彩色图像, 其图像矩阵是由 3 个通道 (RGB) 的 2 维矩阵构成, 对该图像进行卷积时, CNN 中的卷积核通常也对应着输入图像的 3 个通道 (如果输入是彩色图像), 输入图像的尺寸大小可灵活指定为任意值。卷积核的大小影响着提取的特征尺寸和复杂度。研究表明, 卷积核越大, 提取的特征越复杂。在卷积操作中, 需设定好填充 (Padding) 属性和步长 (Stride), 图像经过卷积后的结果如图 1 所示。

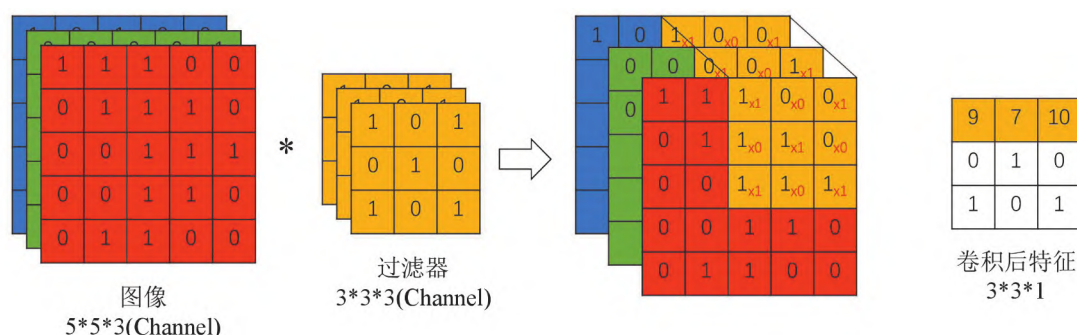


图 1 卷积计算过程

1.2.2 池化层

通过卷积层获得特征之后, 虽然该特征同样可以作为分类器的输入, 但是由于卷积计算之后的数据量巨大, 因此, 为了降低特征的维度和计算量, 需要通过池化层对得到的特征进行处理。

最常见的池化方法包括最大池化 (Max Pooling) 和平均池化 (Average Pooling)。在最大池化中, 通过事先设定的步长, 在每个池化区域内选择数值最大

的元素作为该区域的池化后特征值; 而平均池化则是对每个池化区域内的数值进行求平均, 得到该区域的池化后特征值。两种池化方法被广泛应用于 CNN 中, 旨在减小特征图的空间维度, 有效降低模型的参数数量和计算量, 同时保留关键特征信息。通过池化操作, CNN 能够更加高效地提取和保留数据中的重要特征, 进而提升模型的性能和效率, 使其在各种应用场景中展现出更为出色的泛化能力和鲁棒性。

CNN 利用局部感知区域和权值共享的方法,通过对空间或时间上的采样策略,全面地挖掘数据的局部特征。这种方法允许网络对数据中的重要信息进行有效提取,并且能够在处理大规模数据时保持高效性和准确性。这种设计可以优化网络结构,使得网络在一定程度上具备了平移和缩放的不变性,从而提升模型的泛化性能。

1.2.3 全连接层

在 CNN 中,全连接层起到特征分类器的作用。

卷积层和池化层将原始数据映射到隐层的特征空间内,全连接层则可以将隐层的特征映射到样本标记空间,即将前面经过多次卷积后的高度抽象化的特征进行整合归一化处理,对各种分类情况进行一个对应概率的输出。

全连接层由它的层级本身和激活函数构成,激活函数的主要作用是增加模型的非线性表达能力、提高模型的鲁棒性、加速模型收敛等。全连接层的特征展开过程如图 2 所示。

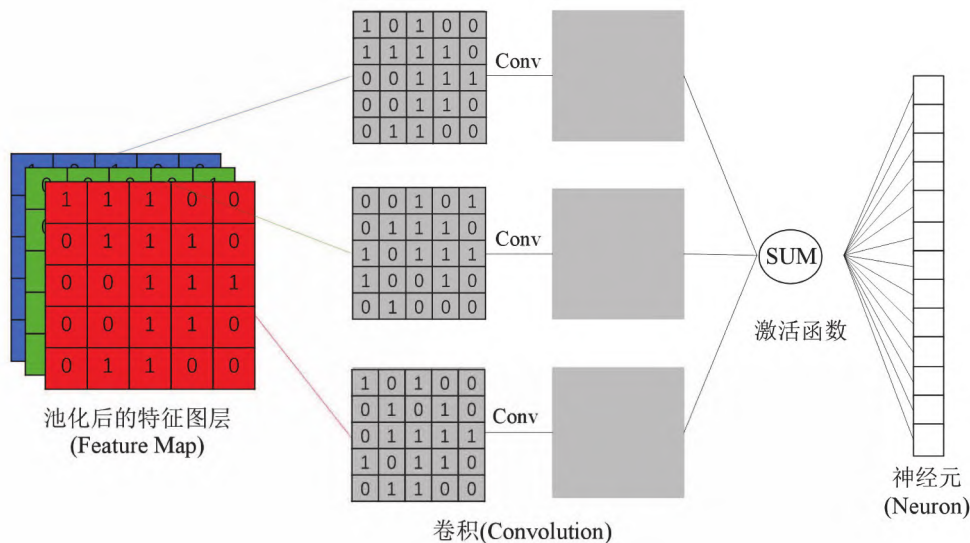


图 2 全连接层特征展开过程

2 基于时频图像和 CNN 的轴承故障诊断方法

2.1 故障诊断方法框架

本研究提出的故障诊断方法主要包含以下四个核心步骤,故障诊断流程如图 3 所示。

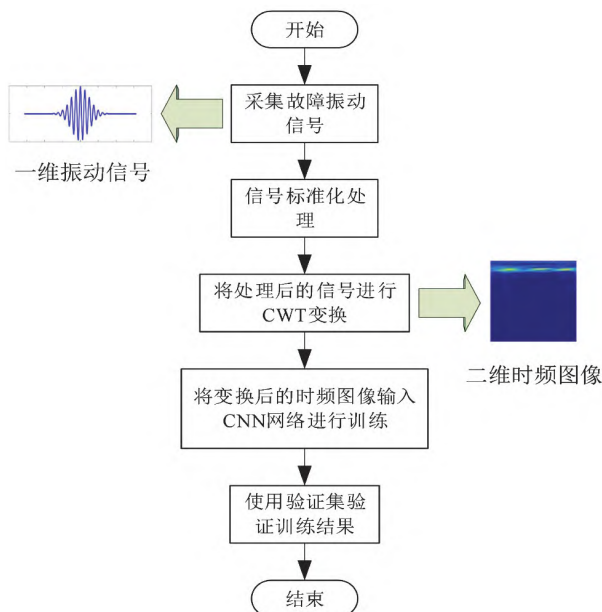


图 3 基于时频图像和 CNN 故障诊断算法流程图

(1) 采集振动信号,在电机滚动轴承试验台上采集不同故障类型、不同故障程度的数据集,并将数据进行标准化处理。(2) 对于样本数据中的故障信号,通过 CWT 生成原彩显示的时频图像。然后,对这些时频图像进行压缩处理,从而建立时频特征图像数据集。压缩的目的是为了降低输入 CNN 特征图像的大小,减小训练时间,提高训练速度。这种压缩处理有助于优化模型训练过程,使其更加高效而又不失准确性。(3) 建立 CNN 模型,根据图像大小和网络结构设置训练参数,选择一定数量的时频特征图像作为训练样本进行训练,使其能够准确地学习和识别不同类型的时频特征图像。(4) 对外部时频图像进行 CNN 模型测试,并对测试集进行多次反复测试。这一过程旨在验证模型的泛化能力和稳健性,确保其在新样本上的表现与训练时一致。通过多次测试,可以得到模型在不同数据集上的平均性能,进而评估其在实际应用中的可靠性和实用性。

2.2 CNN 模型架构设计

本研究设计了一种针对轴承故障诊断的 CNN 架构,如图 4 所示。

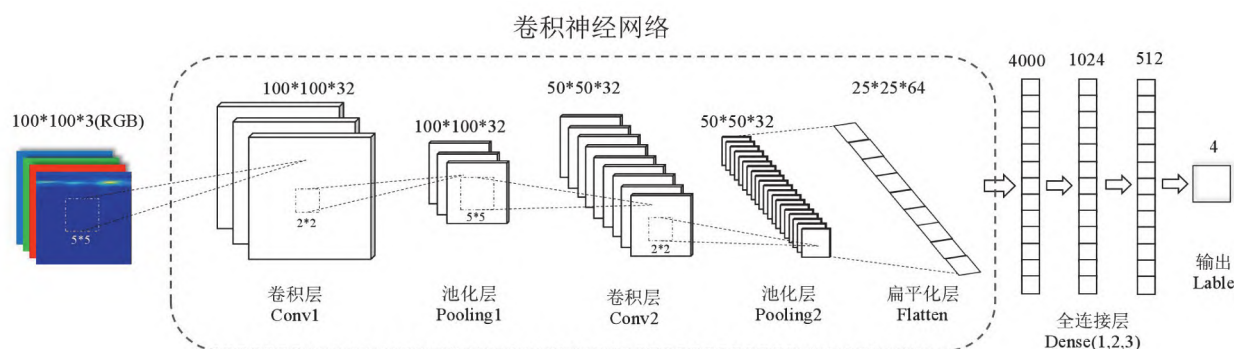


图 4 CNN 结构图

该网络模型旨在从轴承振动信号时频图像中自动提取故障特征, 并实现对轴承多种工况的精确分类。模型输入尺寸为 $100 \times 100 \times 3$ 的 RGB 图像, 该图像蕴含了丰富的故障表征信息。网络架构由卷积层、池化层、扁平化层和全连接层串联构成, 逐层递进地提取和抽象图像特征。具体而言, 模型首先通过卷积层 Conv1, 采用 2×2 卷积核及 32 个滤波器进行初步特征提取, 随后经由池化层 Pooling1 进行特征降维和鲁棒性增强。为进一步挖掘深层特征, 网络堆叠了卷积层 Conv2, 使用 2×2 卷积核和 32 个滤波器以学习更高级的特征表示, 并再次通过池化层 Pooling2 压缩特征维度。卷积和池化层后, 特征图经扁平化层 Flatten 转换为一维特征向量, 连接至全连接层 Dense(1,2,3)。全连接层采用多层结构, 神经元数量依次递减, 实现特征空间的高度非线性映射。

经过上述处理, 模型通过包含 4 个神经元的输出层输出轴承四种工况的概率分布, 分别为正常、外圈故障、内圈故障和滚动体故障工况。输出层采用 Softmax 激活函数, 确保输出值为各工况的概率值, 从而实现基于时频图像的轴承多工况智能诊断。

3 应用研究

3.1 时频数据集构造

电机轴承故障信号具有非平稳性, 在时域或频域中不同时刻的信号差异很大, 为了在信号中凸显故障特征, 采用时频图的数据代替一维振动数据。

利用连续小波算法提取振动信号的时频特性生成时频图像, 构造 CNN 的图像数据集。以 CWRU 滚动轴承实验室为例, 实验平台主要包括一个电动机驱动的测试系统, 用于模拟不同工况下的轴承振动。该平台由电动机、轴承、传感器和控制系统等组成, 如图 5 所示。电动机通过联轴器驱动测试轴, 该轴上

安装有不同类型的滚动轴承。通过调节不同的工作条件(如转速、负载等), 平台可以模拟各种工况下的轴承振动。

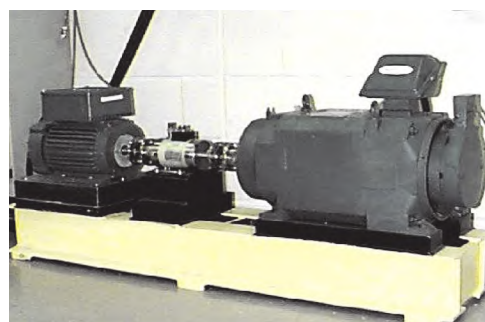


图 5 滚动轴承实验平台

该实验选用 CWRU 数据集中驱动端轴承的故障作为诊断对象, 包含正常、外圈故障、内圈故障和滚动体故障四种工况。故障样本分别为 7 英寸、14 英寸、21 英寸三种不同直径轴承的人为故障数据, 以体现不同的故障严重程度。每个采集的振动信号样本长度为 1024 点。图 6 展示了四种工况下的振动信号波形图。

对所有振动信号样本进行 CWT 处理, 采用 CMOR 小波基获取时频图并进行频谱分析。图 7 展示了不同故障工况下 CWT 后的时频图像。

本实验通过使用不同直径轴承的故障数据, 来全面评估诊断方法对不同故障严重程度的适应性和鲁棒性。最终, 对得到的时频图像进行预处理, 首先按 8 : 2 的比例随机划分训练集和测试集, 保证数据分布的均衡性。然后使用 PIL 库对图像进行压缩处理, 将原始时频图像统一重新采样至 100×100 像素大小。此外, 为提高模型性能, 对压缩后的图像进行标准化处理, 包括随机裁剪和像素值归一化等数据增强操作。这些预处理步骤不仅能显著减少数据维度, 提高计算效率, 同时也增强了模型的泛化能力。

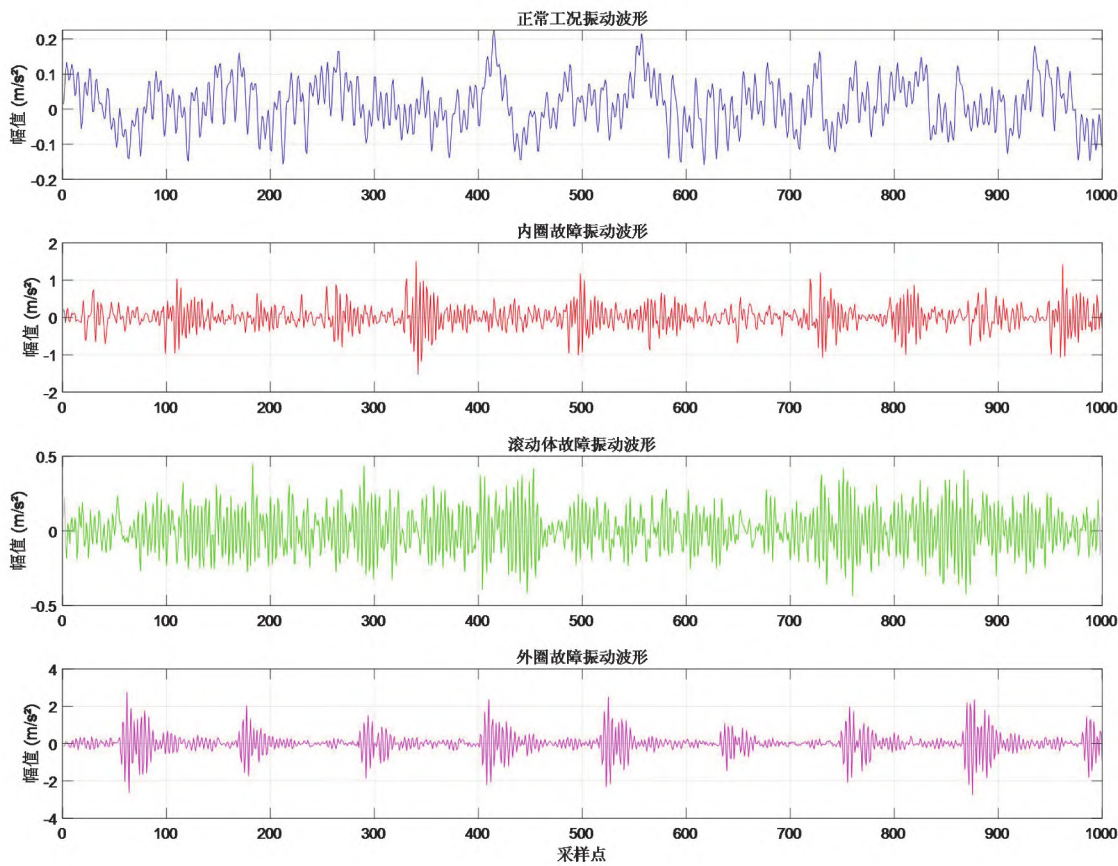


图 6 不同工况下的振动信号波形图

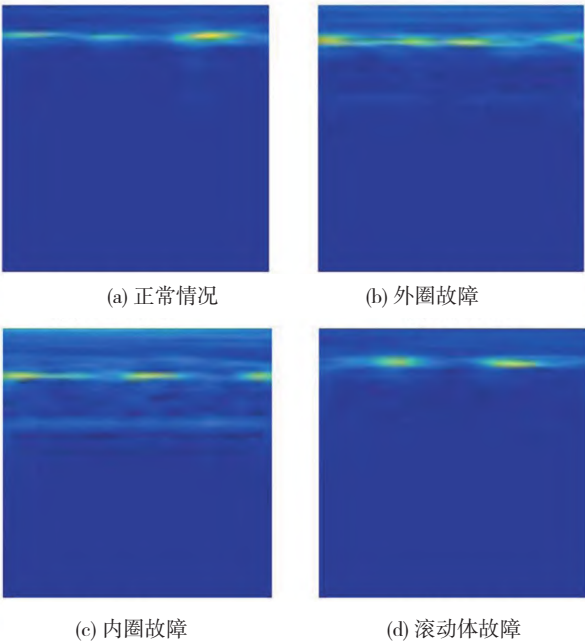


图 7 不同工况下的时频图

3.2 网络训练

3.2.1 基于时频图像的 CNN 训练

在确定了网络结构后，采用 Keras 和 Python 平

台加载训练集数据，并对 CNN 进行训练。所使用的 CNN 结构包括二维卷积层和最大池化层，具体的训练参数设置如表 1 所示。训练集和测试集包含四种工况的数据，其中正常数据共有 210 个图像样本，故障数据每类有 630 个图像样本。训练过程中，模型的训练精度和误差曲线如图 8 所示。从图 8 可以看出，随着迭代次数的增加，网络在第 20 至 30 次迭代时收敛到较好的结果，训练误差接近于零，训练精度保持在 0.99，准确率逐渐上升，损失率平滑下降，表明模型的鲁棒性不断增强。训练至第 20 次时，故障诊断的准确率就已趋于稳定。

准确率和模型损失是相互关联的反向指标：通常情况下，准确率越高，损失值越低，二者共同反映了模型的优化效果。表 2 中的实验结果充分证明了该模型在时频图像识别中的高准确率和较强的泛化能力，且准确率与损失值之间的反向关系进一步验证了模型的有效性。在 420 个测试样本中，仅有两个样本的预测结果与真实值不符，其他样本的预测值均与真实值一致。

表 1 CNN 训练参数设置

参数	设置值
训练轮数 (Epochs)	30
批量大小 (Batch_Size)	128
卷积核大小 (Conv2D_Kernel)	(5,5)
步长 (Strides)	(2,2)
填充方式 (Padding)	'same'
最大池化窗口大小 (MAX Pool_Size)	(2,2)
优化器 (Optimizer)	Adam
学习率 (Learning Rate)	0.001
损失函数 (Loss Function)	categorical_crossentropy
激活函数 (Activation function)	'relu'

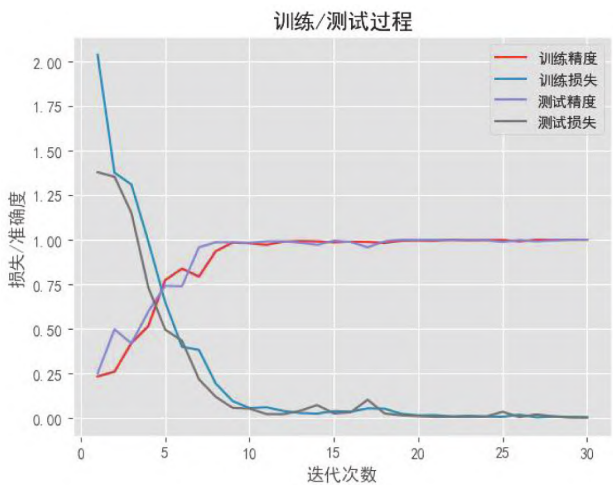


图 8 时频图像 CNN 训练结果

表 2 时频图像 CNN 多组实验结果

序号	训练准确率	模型损失	平均用时
1	99.23%	0.0091	7.12 秒
2	99.61%	0.0085	7.12 秒
3	98.52%	0.0145	7.12 秒
4	99.74%	0.0072	7.12 秒
5	97.89%	0.0186	7.12 秒

3.2.2 CNN 训练对比实验

本研究采用一维振动信号作为 CNN 的输入，且将网络的超参数（如学习率、Epoch、优化器等）设置与时频 CNN 相同。这些超参数控制了网络的训练过程，并在一定程度上影响模型的性能。具体来说，学习率决定了每次参数更新的幅度，Epoch 指定了整个训练集通过网络的次数，而优化器则用于调整模型权重以最小化损失函数。在网络结构方面，经过卷积层和池化层的特征提取后，一维振动信号被转化为高维特征向量，这些特征能够有效地代表振动信号中的潜在模式和故障特征。然后，这些特征向量会通过全连接层进行进一步的分类处理，最终实现对故障类型的识别与分类。该一维振动信号的 CNN 结构见图 9。

振动信号作为输入的 CNN 训练结果见图 10，更多的训练结果详见表 3。

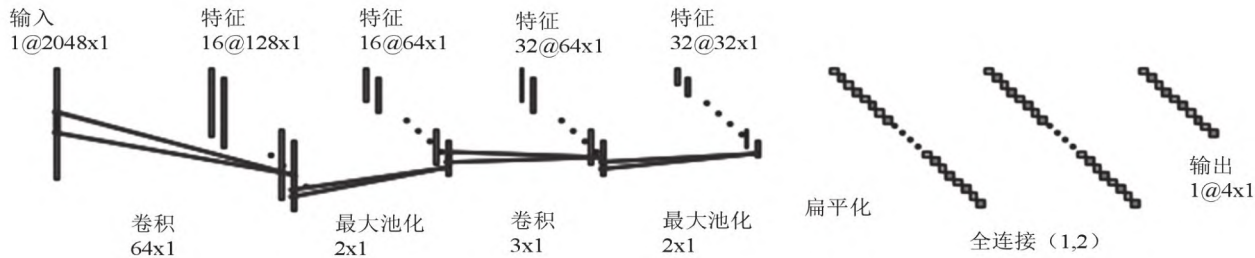


图 9 一维振动信号 CNN 结构

表 3 原始振动信号 CNN 多组实验结果

序号	训练准确率	模型损失	平均用时
1	98.33%	0.0469	0.324 秒
2	98.96%	0.0266	0.329 秒
3	95.82%	0.0847	0.318 秒
4	98.67%	0.0311	0.303 秒
5	98.81%	0.0405	0.305 秒

由图 10 和表 3 可以看出，一维振动信号的 CNN

训练结果与时频图像的 CNN 训练结果存在差异，五次随机试验中最高训练精度为 98.96%，低于时频图像 CNN 的最高训练精度 99.74%。这是因为时频图像 CNN 的抗干扰性更强，一维的时域振动信号转换为图像时频信号之后，时域振动的干扰影响更小，整体误差函数迭代下降更加平滑。

3.3 时频图像 CNN 故障诊断

将训练好的时频图像 CNN 模型导入故障诊断程序，针对不同工况下的测试集，利用事先经过训练的

模型进行故障识别验证,不同工况下的识别结果如图 11 所示。

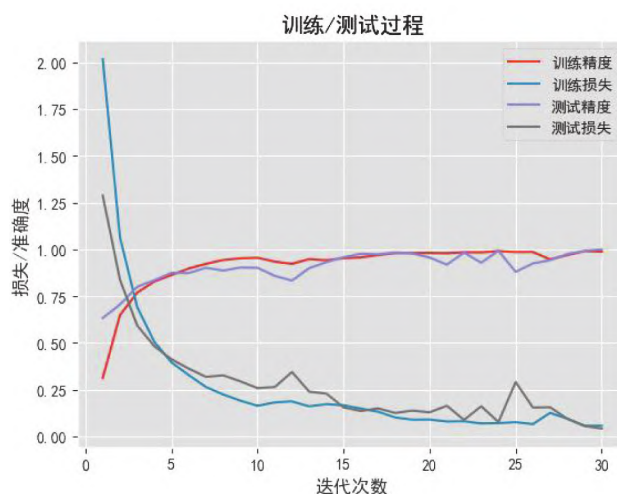


图 10 原始振动信号 CNN 训练结果

由图 11 可见,正常和外圈故障的识别效果非常出色。对于内圈故障和滚动体故障的识别率相对较低。这是因为内圈故障与滚动体故障的时频图像具有一定的相似性。

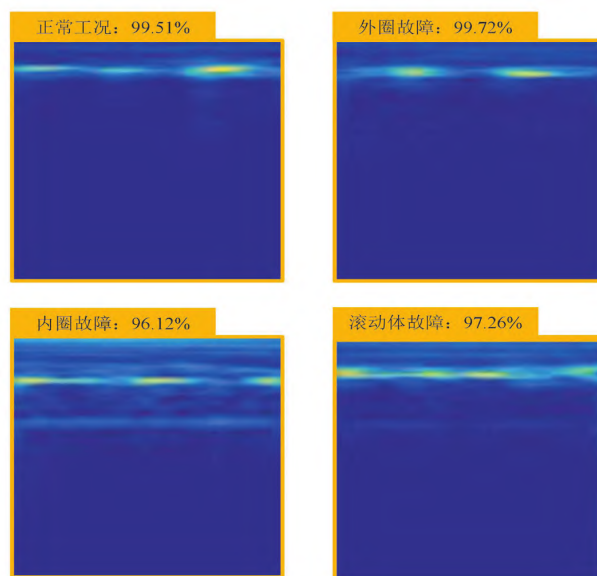


图 11 轴承故障识别率

如图 12 所示,基于 CNN 的轴承故障诊断方法不仅能准确识别同一故障类型下形态各异的时频图像(如内圈故障类型下差异较大的时频图像,红框所示),而且能有效区分具有相似时频特征的不同故障类型(如绿框所示外圈故障与内圈故障)。这充分说明 CNN 模型能成功捕捉到故障信号的本质特征,可实现复杂工况下可靠的智能诊断。

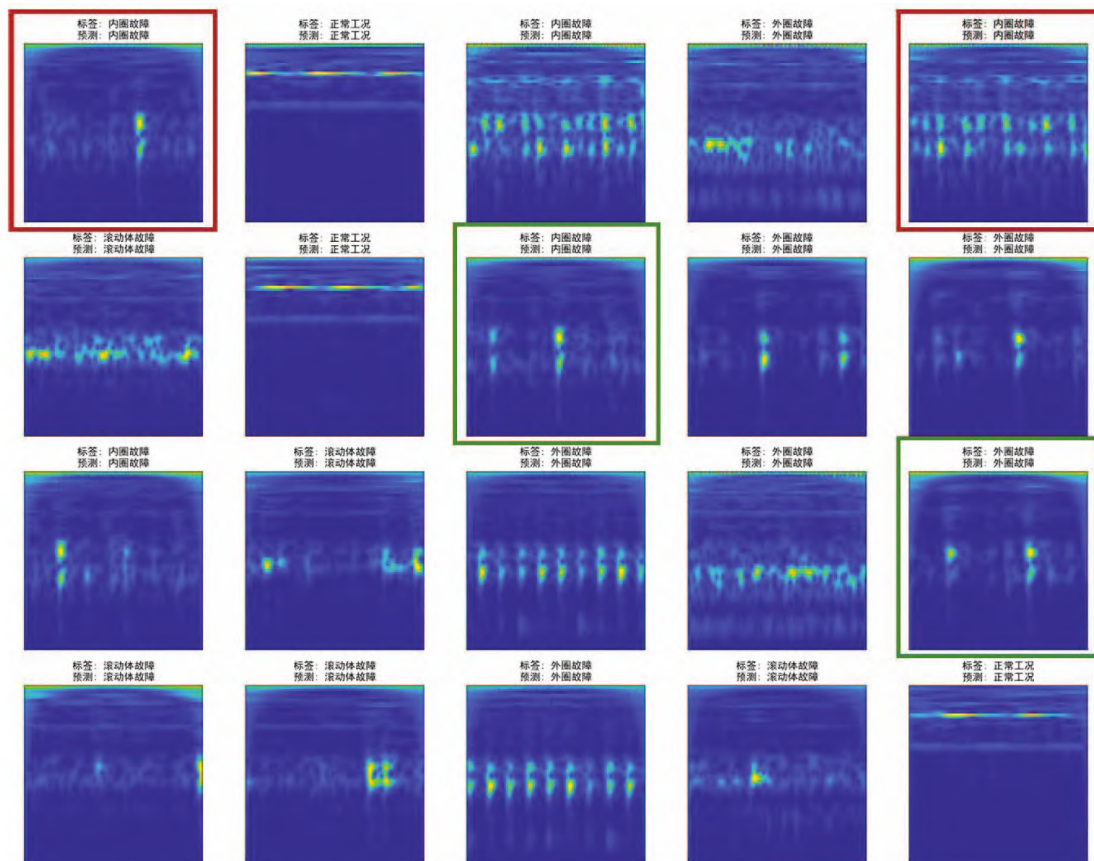


图 12 轴承四类工况识别结果

4 结论

为了解决滚动轴承故障诊断这个难题, 运用了 CWT 来精确分析轴承故障振动信号, 生成对应的时频特征原彩图像, 并将其作为 CNN 的输入数据。最终, 得出了以下结论:

(1) 基于时频图像与 CNN 的轴承故障诊断方法是可行且有效的, CNN 通过提取时频图像中的特征信息, 能够准确地表达故障的真实状况。相对于传统信号处理方法, 该方法具有更高的准确率, 且能够精准识别出故障类型。

(2) 时频图像 CNN 方法展现了卓越的识别能力。该方法能够精准地识别同一种故障下差异较大的时频图像, 同时能有效区分出时频特征相似但故障类型不同的图像, 展现了在处理不同故障时的强大泛化能力, 为故障诊断和预测提供了可靠的技术支持。

参考文献:

- [1] 余萍, 曹洁. 深度学习在故障诊断与预测中的应用[J]. 计算机工程与应用, 2020, 56(3): 1-18.
- [2] 龚俊, 张月义, 陈思聪, 等. 基于 SWT 与改进卷积神经网络的轴承故障诊断[J]. 现代电子技术, 2024, 47(6): 68-74.
- [3] 段毅博, 黄民, 王帅, 等. 基于 CNN-Transformer 模型的自动扶梯轴承故障诊断方法研究[J]. 现代电子技术, 2024, 47(16): 1-6.
- [4] 秦凯, 边莉, 张宁. 基于智能方法的电机故障诊断技术综述[J]. 工业仪表与自动化装置, 2016(1): 19-22.
- [5] 刘朦月. 基于振动信号的电机轴承故障诊断方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2016.
- [6] 刘长良, 武英杰, 甄成刚. 基于变分模态分解和模糊 C 均值聚类的滚动轴承故障诊断[J]. 中国电机工程学报, 2015, 35(13): 3358-3365.
- [7] 赵超阳, 陈亮, 韦隆, 等. 基于改进 EMD-Kurtogram 法的滚动轴承早期故障诊断研究[J]. 现代电子技术, 2024, 47(4): 159-163.
- [8] 胡永涛, 周强, 范峥, 等. 基于 DTCWT-MAP 去噪及 Hilbert 包络的滚动轴承故障诊断[J]. 河南工学院学报, 2020, 28(1): 8-14.
- [9] NIU F L, HUANG J, YANG J Q, CHEN L Y, JIN H. Rotor broken-bar fault diagnosis of induction motor based on HHT of the startup electromagnetic torque[J]. Frontiers of Electrical and Electronic Engineering in China, 2006, 1(2): 188-193.
- [10] 许允之, 龚乃玮, 葛焱, 等. 基于信息融合方法的电机故障诊断研究[J]. 华北电力大学学报(自然科学版), 2012, 39(3): 53-58.
- [11] 李恒, 张氢, 秦仙蓉, 等. 基于短时傅里叶变换和卷积神经网络的轴承故障诊断方法[J]. 振动与冲击, 2018, 37(19): 124-131.
- [12] 李巍华, 单外平, 曾雪琼. 基于深度信念网络的轴承故障分类识别[J]. 振动工程学报, 2016, 29(2): 340-347.
- [13] 王永乐, 续婷, 杜敦伟, 等. 基于多特征与卷积神经网络的 SAR 图像分类方法[J]. 数学的实践与认识, 2020, 50(6): 140-147.
- [14] 刘乙萱, 苏鑫. 基于两阶段迁移学习的 Multi-scale SE-ResNet50 深度卷积神经网络的多标签航空图像分类问题研究[J]. 数学的实践与认识, 2023, 53(6): 174-186.
- [15] 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络[J]. 自动化学报, 2016, 42(9): 1300-1312.
- [16] 毛小玲, 向往, 欧阳明昆, 等. 基于改进卷积神经网络的脑电信号焦虑情绪量化识别[J]. 广西科学, 2022, 29(2): 269-276.
- [17] 李从跃, 胡以怀, 沈威, 等. 基于小波变换和 CNN 的船用机械故障诊断[J]. 中国测试, 2024, 50(3): 183-192.

Bearing Fault Diagnosis Based on Time-Frequency Image and Convolutional Neural Network

WANG Guozhu^{1,2}, ZHOU Ruizhe^{1,2}, HUANG Zhen³, CHI Xiaohang⁴, ZHANG Kaituo^{1,2}

(1. School of Cable Engineering, Henan Institute of Technology, Xinxiang 453003, China;

2.Key Laboratory of Cable Structure and Materials in Henan Province, Xinxiang 453003, China;

3.Guangzhou Hengxing Conductive Technology Co., Ltd, Guangzhou 510385, China;

4.School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China)

Abstract: To ensure the normal operation of the motor, improve the safety and reliability of industrial production, and ensure the orderly production of enterprises, a bearing fault diagnosis method based on time-frequency images and CNN is proposed. Firstly, perform CWT processing on the vibration signal to obtain time-frequency characteristic images; Then, the time-frequency images are compressed and fed into a CNN for training, enabling it to accurately learn and recognize different types of time-frequency feature images; Finally, the test time-frequency images are input into the CNN for training to verify the reliability of the training results. The results show that the proposed method can detect bearing faults and identify fault types effectively, with a recognition rate of 99.7%. Compared with traditional bearing fault diagnosis methods, this method has better generalization ability.

Key words: fault diagnosis; CNN; CWT; time-frequency image

(责任编辑 王 磊)