

基于MTF和改进残差网络的轴承故障定量诊断方法

李凌轩^{1,2}, 马振玮^{1,2}, 于泽峻^{1,2}, 邢 壮^{1,2}

(1. 东北大学 机械工程与自动化学院, 辽宁 沈阳 110819;

2. 东北大学秦皇岛分校 控制工程学院, 河北 秦皇岛 066004)

摘 要: 有别于目前滚动轴承故障诊断多集中在定性分析阶段, 提出了一种使用图像分类的滚动轴承故障定量诊断方法. 采用重叠采样方法, 对一维时序数据进行数据增强, 使用马尔可夫转换场(Markov transition field, MTF)方法将一维时序数据转换成二维图像, 为输入到神经网络模型中提供二维图像样本并保留了时域信息, 搭建和训练基于迁移学习微调处理的ResNeXt和ResNeSt改进残差网络, 将故障图像进行分类并实现故障诊断. 采用混淆矩阵和t分布领域嵌入(t-distributed stochastic neighbor embedding, t-SNE)可视化方法进行实验, 结果表明, 该滚动轴承故障定量诊断方法能够实现多工况滚动轴承故障的定量诊断, 且具有诊断精度高和训练速度快的优点.

关键词: 轴承故障; 马尔可夫转换场; 残差网络; 迁移学习; 定量诊断

中图分类号: TP 277 文章编号: 1005-3026(2024)05-0697-10

Method for Bearing Fault Quantitative Diagnosis Based on MTF and Improved Residual Network

LI Ling-xuan^{1,2}, MA Zhen-wei^{1,2}, YU Ze-jun^{1,2}, XING Zhuang^{1,2}

(1. School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China;

2. School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China.

Corresponding author: LI Ling-xuan, E-mail: lingxuan_li@163.com)

Abstract: Different from the current rolling bearing fault diagnosis which mainly focuses on the qualitative analysis stage, a quantitative fault diagnosis method for rolling bearings based on image classification is proposed. The overlapping sampling method is used to enhance the one-dimensional time series data, and then the Markov transition field (MTF) method is used to convert the one-dimensional time series data into two-dimensional images, which provide two-dimensional image samples for inputting into the neural network model and retain the time-domain information. The ResNeXt and ResNeSt modified residual networks with fine-tuning processing based on transfer learning are built and trained to classify fault images and realize fault diagnosis. The confusion matrix method and t-distributed stochastic neighbor embedding (t-SNE) visualization method are used to carry out experiments. The results show that the proposed method for rolling bearing fault diagnosis can realize the quantitative diagnosis of multi-working condition rolling bearing fault, and has higher diagnosis accuracy and faster training speed.

Key words: bearing fault; Markov transition field (MTF); residual network; transfer learning; quantitative diagnosis

滚动轴承作为旋转机械设备的重要组成零件, 其健康状态通常决定着整个旋转机构是否可以正常运行并达到足够的旋转精度. 滚动轴承故

障可能导致严重的工程事故并造成巨大的经济损失, 及时地对滚动轴承的故障进行诊断是非常必要的. 滚动轴承的组成结构和其所处的不同工

况条件决定了其故障类型和故障程度的多样性,对故障类型和故障程度的诊断分类具有重要的工程意义.

使用滚动轴承振动信号对轴承故障诊断的传统方法多是采用信号分解和包络分析获取振动信号故障特征频率,然后与已知的故障特征频率分别进行对应比较从而对故障类型进行诊断. Cocconcelli 等^[1]使用短时傅里叶变换(short-time Fourier transform, STFT)对在非平稳条件下球轴承进行故障特征提取,在时频域中对每个周期的短时傅里叶变换进行平均处理以达到增强故障特征的目的. STFT 存在时域不准确和窗口大小不确定的缺点,小波变换(wavelet transform, WT)在一定程度上解决了这些问题. Sun 等^[2]使用离散小波变换与包络分析相结合的方法,提取滚动轴承振动数据的特征谱,应用频谱互相关系数对不同工况的滚动轴承进行故障诊断. 经验模态分解(empirical mode decomposition, EMD)是 Huang 等^[3]提出的一种时频信号分解工具,用于分析非平稳和非线性信号,它将信号分解为固有模态函数之和的形式,从而表示信号的固有振动模式. Yu 等^[4]通过对包络信号应用 EMD 方法和希尔伯特变换得到局部希尔伯特边缘谱,实现了对滚动轴承的外圈和内圈故障的识别. 该方法提取的特征明显优于传统包络谱方法. 然而 EMD 存在端点效应和模态混叠等明显问题,在此基础上,甄冬等^[5]使用经过改进的集合经验模态分解方法(ensemble EMD, EEMD)^[6],并结合调制信号双谱(modulation signal bispectrum, MSB)分析解调方法. 崔玲丽等^[7]使用改进的奇异值分解(singular value decomposition, SVD),均能有效提取滚动轴承微弱故障特征. Gu 等^[8]提出了一种用于滚动轴承早期故障诊断的自适应变分模态分解(variational modal decomposition, VMD)和 Teager 能量算子方法,有效降低了滚动轴承早期故障特征提取的信号噪声. 信号分解法的泛化能力不足,在处理不同故障数据时需要手动提取信号特征. 此外,该方法仅限于定性分析,无法做到定量分析轴承故障,这限制了对轴承故障的深入了解.

近年来,随着大数据技术和硬件性能的迅速发展,人工智能在多领域得到了广泛应用,包括深度学习模型在故障诊断和健康监测领域的发展^[9]. Fu 等^[10]使用支持向量机(support vector

machines, SVM)结合变分模态分解时频特征提取,实现轴承故障模式的识别. 路敦利等^[11]提出一种结合 k 近邻(k -nearest neighbor, KNN)算法与朴素贝叶斯算法的改进算法,显著提高了滚动轴承故障诊断的识别率. 然而,传统机器学习方法往往需要结合信号处理方法作为特征工程,因此故障诊断的结果很大程度上取决于信号特征提取方法的效果,仍然存在泛化能力较差等问题. 随着深度学习的兴起,其强大的特征提取能力被广泛应用于端到端的轴承故障诊断,从而省略了人工提取特征这一步骤. Eren 等^[12]采用一维卷积神经网络(convolutional neural networks, CNN)直接将轴承原始时序信号作为输入,无需人工数据处理方法参与,避免了数据处理对特征提取的影响. 该模型精简紧凑,与其他智能诊断方法对比验证了其可行性. Liu 等^[13]提出了一种基于物理的模拟来避免数据不足的方法,使用模拟数据作为源域的域对抗神经网络(domain-adversarial neural networks, DANN),将源域数据和目标域数据进行对抗训练,该方法在轴承故障诊断方面的性能表现优异. 王太勇等^[14]使用引入注意力机制的双向长短期记忆(long short-term memory, LSTM)网络,能够从原始信号中自适应提取多尺度特征,提高模型诊断精度. 尽管深度学习模型目前仍存在如提取特征和模型解释性差等问题,但其强大的特征提取能力使故障诊断达到非常高的诊断精度.

综上所述,滚动轴承故障诊断的主要工作是针对一维时序信号进行特征工程和识别分类. 为了充分利用深度学习模型在图像特征提取方面的优势,本文提出了一种结合一维信号转二维图像和深度学习图像分类的滚动轴承故障诊断的方法. 该方法采用了图像分类中深度学习处理的思路,并使用改进的残差网络模型来提高诊断精度,能够识别多工况条件数据的故障类型并实现故障程度的定量诊断,为故障诊断提供了新的解决思路.

1 模型与方法

首先采用重叠采样方法对一维时序数据进行数据增强;再使用马尔可夫转换场(Markov transition field, MTF)将一维时序数据转换成二维图像,随后搭建经迁移学习微调处理的 ResNeXt 和 ResNeSt 改进残差网络,并将图像数

据分别输入神经网络模型进行训练和验证,实现故障图像分类及故障诊断;最后采用可视化方法进行实验结果分析和评估.本文所提方法的流程如图 1 所示.

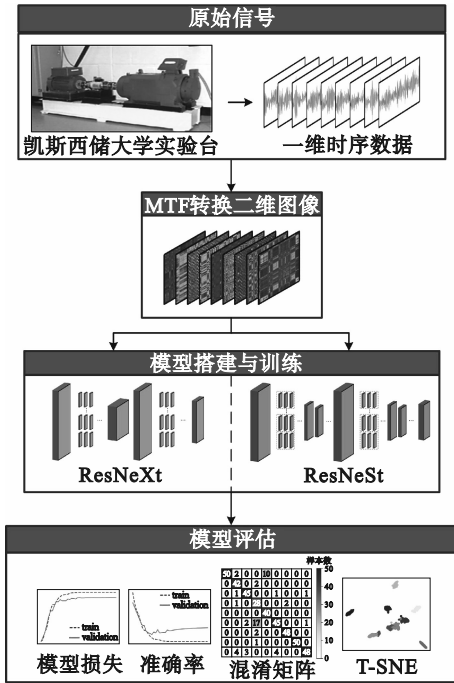


图 1 方法流程图

Fig. 1 Method flow chart

1.1 原始信号的重叠采样

使用数据增强方法对图像类训练数据进行处理,可以显著提高深度学习神经网络模型的泛化能力,同时减少模型对数据的过拟合现象,提高模型的鲁棒性.总体来说,数据增强方法包括颜色变化、缩放、旋转、镜像和裁剪等操作,这些操作可以增加数据所包含的信息量,从而提高图像中有效特征的比例.然而,本文所使用的图像数据是由一维信号转换得到,这些信号包含时域信息,与实物识别分类任务的图像存在较大差异,常规的数据增强方法在处理此类信号图片时并不适用.因此,本文采用直接对原始一维信号进行重叠采样方法^[15]作为数据增强的方法.该方法能增加采样信号的数据量,同时由于采样信号数量与转换后的图像数量相对应,因此该方法也能够增加对应训练集的数据图像.

重叠采样过程是选取一条数据所包含的信号点数作为窗口大小,根据一定的偏移量沿时间轴方向平移窗口,并截取数据以生成一个信号样本.不同数据样本之间存在信号的部分重叠,从而有效地增加了样本数量.重叠采样过程示意图如图 2 所示.其步骤如下:

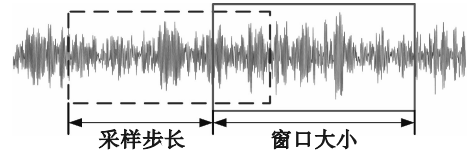


图 2 重叠采样示意图

Fig. 2 Schematic diagram of overlapping sampling

- 1) 根据所需图像尺寸选定窗口大小;
- 2) 根据所用总样本点数和样本数计算平移步长;
- 3) 选择窗口初始截取位置;
- 4) 按平移步长对窗口进行移动并截取信号样本.

重叠采样方法中的样本数 k 、所用信号点总数 z 、采样步长 c 和窗口大小 w 关系计算公式如下:

$$k = \frac{z-w}{c} + 1. \quad (1)$$

1.2 一维时域数据到二维图像的 MTF

本文使用的神经网络需要二维图像数据,而原始的时序信号是一维数据,因此需要一种方法将数据从一维转换为二维,同时需尽可能地保留原始一维数据中的时序信息.本文中所涉及的轴承故障时序信号满足马尔可夫模型的条件,表现出马尔可夫性,即该时序信号的下一个状态仅与当前的状态有关,与之前或者更早的历史状态无关.马尔可夫性的存在使得可以使用马尔可夫转移概率矩阵.通过采用 MTF,可以将一维时序信号有效地转换为具有时序信息的二维图像.运用图片分类神经网络对这些图像进行分类,从而达到故障分类的目的.

1.2.1 马尔可夫转移概率矩阵

马尔可夫转移概率矩阵方法可以将与时间相关的一维序列转换为二维矩阵,MTF 就是以此特性为基础将一维时序信号转换为包含时序信息的二维图片.在马尔可夫链中,假设系统随时间共有 n 种状态,用 S_k 表示可能的状态,所有状态构成的状态空间 S 为

$$S = \{S_1, S_2, \dots, S_n\}, n > 1. \quad (2)$$

用 T_t 表示马尔可夫链在时刻 t 时的状态,则时刻 t 下的状态概率用 P_t 表示为

$$P_t = P(T_t = S_i), t = 1, 2, \dots, n. \quad (3)$$

假设马尔可夫链当前 t 时刻状态 T_t 为 S_i ,则 $t+1$ 时刻状态 T_{t+1} 为 S_j 的概率即从状态 S_i 转移到状态 S_j 的转移概率,记为 P_{ij} ,那么状态转移概率可用条件概率表示为

$$P_{ij} = P(T_{t+1} = S_j | T_t = S_i), S_i \in S, S_j \in S. \quad (4)$$

由马尔可夫链中所有状态转移概率组成的二维矩阵形式为马尔可夫转移概率矩阵,表示如下:

$$P_{n \times n} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ P_{31} & P_{32} & \cdots & P_{3n} \\ \vdots & \vdots & \cdots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix} \quad (5)$$

1.2.2 MTF方法

尽管马尔可夫转移概率矩阵已经是二维数据并且可以转换为图像,但是由于它对时间序列的分布和状态沿时间发生顺序不敏感,仅使用马尔可夫转移概率矩阵将会导致数据丢失过多时间信息.因此,MTF方法^[16]的思路是在马尔可夫转移概率矩阵的基础上进行改进,原先按照不同状态进行排列的矩阵被转换为按照时间步长进行扩展的矩阵.MTF矩阵如公式(6)所示,通过将时间序列 $X = \{x_1, x_2, \dots, x_n\}$ 映射到 Q 个分位数空间 $[1, Q]$, 得到包含时间戳 i, j 的分位数 $q_i, q_j, q_i \in [1, Q], q_j \in [1, Q]$, 然后构建 $Q \times Q$ 的矩阵, P_{ij} 表示时间间隔为 $|i-j|=k$ 的 q_i 到 q_j 的转移概率,代入 q_i, q_j 值后得到 MTF 矩阵 M , 马尔可夫转换过程如图 3 所示.

$$M = \begin{bmatrix} P_{ij} | x_1 \in q_i, x_1 \in q_j & \cdots & P_{ij} | x_1 \in q_i, x_n \in q_j \\ P_{ij} | x_2 \in q_i, x_1 \in q_j & \cdots & P_{ij} | x_2 \in q_i, x_n \in q_j \\ \vdots & & \vdots \\ P_{ij} | x_n \in q_i, x_1 \in q_j & \cdots & P_{ij} | x_n \in q_i, x_n \in q_j \end{bmatrix} \quad (6)$$

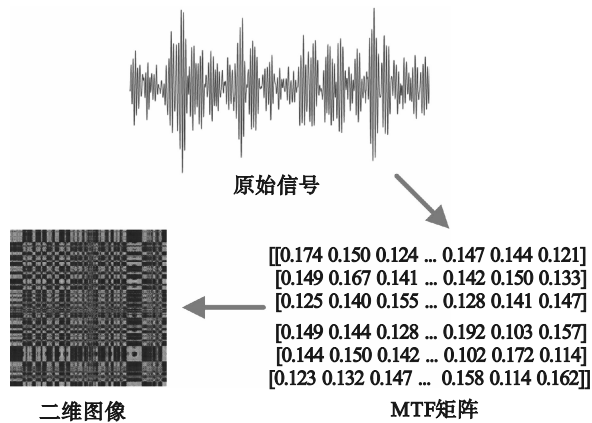


图 3 MTF 转换过程
Fig. 3 MTF transition process

1.3 改进残差网络的 ResNeXt 和 ResNeSt

本文采用了 ResNeXt 网络和 ResNeSt 网络进行模型训练.这两个网络都是在残差网络 ResNet 的基础上进行改进的深度学习网络,它们在原有的残差连接基础上引入了多分支网络结构,并且融入了注意力机制思想,以提高模型性能.

1.3.1 Inception 模块与残差网络

Inception 模块通常先使用 1×1 的卷积对数据进行压缩,然后通过将拥有不同大小卷积核的卷积层在宽度方向并联,网络能够在不同的感受野上对特征进行提取.这些特征随后被拼接起来输入到下一层中,从而在宽度方向提高网络的性能.如图 4 所示,通过在网络中反复堆叠 Inception 模块可以高效地加深网络,这种方法已被证明可以提升模型的准确率,同时有效避免过拟合.

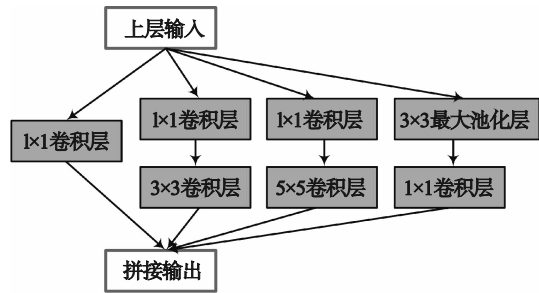


图 4 Inception 模块
Fig. 4 Inception block

在使用 Inception 模块以提升模型准确率时,需要特别注意卷积核的大小、数量等参数的设置,这些参数旨在实现特定的特征提取或变换.因此,在模型中引入 Inception 模块时,往往需要考虑更多超参数的调整.尽管这些结构的组合在某些任务上能够提高神经网络的性能,但由于缺乏明确的规律和设定规则,特别是在需要调整许多超参数时,模型结构的调试方式变得不确定,这是一个明显的缺点.

为了提取更多接近语义信息的特征,通常需要加深网络模型.然而,通过不断加深网络以提高网络性能的过程中,往往会面临训练准确率不增反降的退化问题,同时可能伴随着梯度爆炸等挑战.在图 5 中,表示了每一层对应的激活函数、卷积核大小和输出维数,ResNet 网络通过引入残差连接,成功地在保留上一层训练信息的基础上扩充模型的容量,一定程度上解决了退化问题.

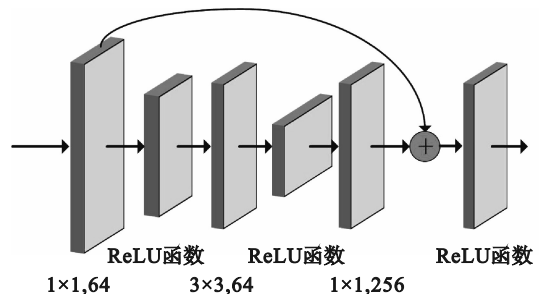


图 5 一般的残差连接结构
Fig. 5 General residual connection structure

残差连接与主线网络梯度的关系由乘法变为加法,有效地避免了梯度爆炸的问题.这一改进不仅解决了由于深层训练程度较好引发的梯度变小导致浅层无法充分训练问题,还为进一步提高模型精度提供了可能.这一改进正是残差网络能够让神经网络扩展至千层以上并达到很好精度的关键原因.

1.3.2 ResNeXt 网络与 ResNeSt 网络

在改善网络模型性能和结构方面,常见的做法是直接增加网络模型的深度或宽度,这种方式通常会增加模型的超参数数量.然而,ResNeXt 网络通过模块化的网络结构,简化了搭建过程,并聚合了具有相同拓扑结构的一组变换方式,同时减少了所需的超参数数量,提高了该网络的可调性,以便应用于不同的任务中.

ResNeXt 网络结合了 Inception 模块和残差网络的优点. Inception 模块的“拆分—变换—聚合”策略提高了网络模型的精度,而 ResNet 的残差连接使得模型能够大幅增加深度,并提取出更多特征以完成更复杂的任务.该模型将 Inception 模块中的多分支不同的卷积层改为使用具有相同结构的卷积层,并将这些分支的重复次数定义为“基数”.文献[17]验证了增加基数比增加深度或者宽度对提高模型精度更有效.尽管模块的基数设置较大,但通过分组卷积的方法,在提高模型精度的同时,减少了模型的参数数量,这在一定程度上减轻了过拟合问题,降低了计算开销,同时简化了模型的构建过程.在此基础上引入残差连接,构成了 ResNeXt 网络的残差模块,如图 6 所示.本文采用的 ResNeXt50-32×4d 网络的具体结构参数如表 1 所示,其中的 32 表示基数数目,4d 表示每一分支通道数是 4.

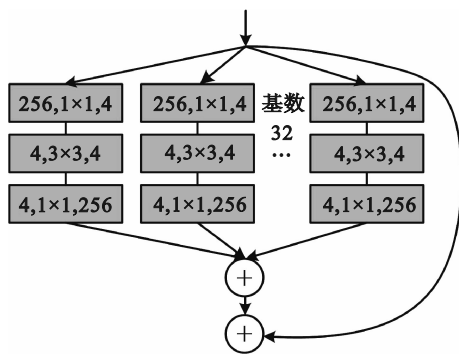


图 6 ResNeXt 残差模块结构

Fig. 6 ResNeXt residual block structure

注意力机制是让神经网络更加关注有用特征的学习方法.这种方法可以学习每个特征对目

标任务的重要程度,并将这些重要程度以不同的权重形式赋给对应的特征.这使得网络模型可以学习到更多对目标任务有用的特征,从而在一定程度上提升网络性能.受 SE-Net 和 SK-Net 注意力机制网络的启发,拆分注意力 ResNeSt 网络^[18]同样采用了将网络输入拆分成不同分支但相同结构的卷积层的方法.该方法关注不同通道特征信息,以确保在不同任务中能够方便适用,并在减少超参数的同时,通过应用注意力机制获取每个通道提取的特征与整体的关系.这使得网络更加重视学习对任务更为有效的特征,从而提高模型精度.以“基数”为单位拆分成 k 个分支后,对每个分支再以相同结构的卷积层拆分成 r 个不同子分支,并在当前“基数”分支中用 Softmax 加入每个子分支注意力权重.最后,将子分支的输出相乘再相加,得到该“基数”分支的输出.这样能够尽可能地提取到更多可以提升模型精度的特征.将这些分支的输出拼接后,得到每一分支的注意力输出,再加入残差连接就构成了 ResNeSt 网络基本的模块结构,如图 7 所示.

表 1 ResNeXt 网络结构
Table 1 ResNeXt network structure

模块	ResNeXt-50(32×4d)	输出
conv1	$7 \times 7, 64$, 步长 2 3×3 , 最大池化层, 步长 2	112×112
conv2	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 & C=32 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$	56×56
conv3	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 & C=32 \\ 1 \times 1, & 512 \end{bmatrix} \times 4$	28×28
conv4	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 & C=32 \\ 1 \times 1, & 1024 \end{bmatrix} \times 6$	14×14
conv5	$\begin{bmatrix} 1 \times 1, & 1024 \\ 3 \times 3, & 1024 & C=32 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$	7×7
分类器	全局平均池化层, 9 维全连接层, Softmax	1×1

拆分注意力层是该网络的核心,这一过程将子分支输出相加输入到全局平均池化层中,以获取每一分支的全局信息.这些全局信息经过全连接层、批量归一化处理 and 激活函数 ReLu 的处理后,被输入到对应通道数目的全连接层.最后,这些信息被输入到各子分支的 r -Softmax 中,以获取每个子分支的权重.这些权重与对应的子分支相乘,从而得到每个子分支的输出,最后累加得

到该“基数”分支的加权权重,拆分注意力层结构如图8所示,可见网络在学习中可以重点学习有效的特征.

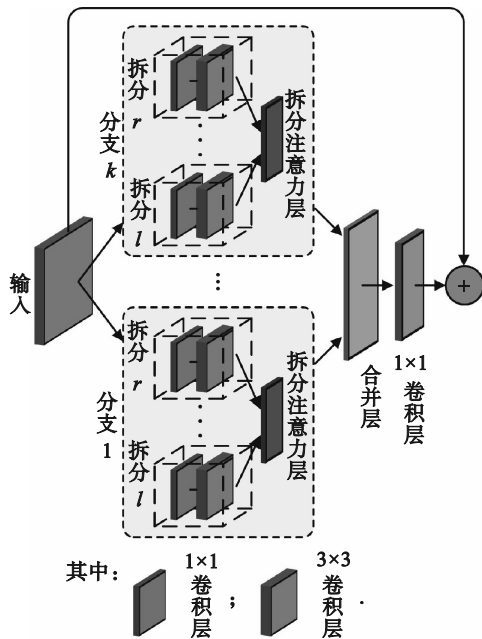


图7 ResNeSt残差模块结构

Fig. 7 ResNeSt residual block structure

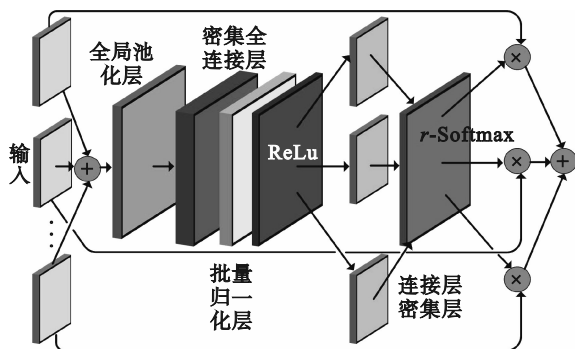


图8 拆分注意力层结构

Fig. 8 Split attention layer structure

1.3.3 迁移学习的微调策略

迁移学习是通过将从源域中所获得的知识或方法应用于相关的目标域,以完成目标域的任务或提升模型性能.在深度学习的神经网络层次化学习过程中,通过在类似但更广泛的数据集上进行训练,网络的浅层特征通常具备良好的泛化性,可以被用作在目标数据集的网络模型训练时的知识,以提高模型训练效率和性能.如图9所示,微调策略是深度学习的一种迁移学习方法,其过程包括在更广泛数据集上训练的模型作为预训练模型,直接将其权重作为目标数据集训练时的初始权重,经过模型训练后得到适用于目标数据集的最终模型.这一方法旨在实现将在更广

泛数据集上学习的知识应用于当前数据集的迁移学习目标.

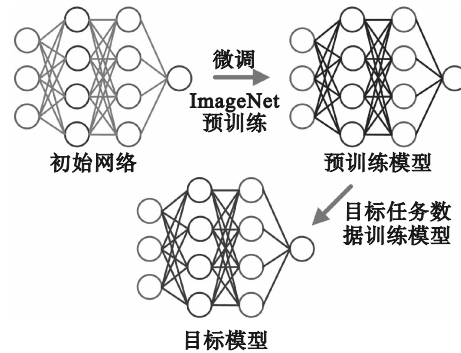


图9 微调策略示意图

Fig. 9 Schematic diagram of fine-tuning strategy

采用在 ImageNet 图片集上的预训练模型. ImageNet 是一个大型图片数据集,包含 14 197 122 张图片,分为 21 841 个类别,广泛应用于图像分类、目标检测、目标定位和场景分类等任务. Yosinski 等^[19]证实了网络浅层部分总是在学习数据集的通用特征,而深层部分则更专注于学习数据集任务的特定特征.因此,这些通用知识可以应用于多种数据集和任务.使用的预训练模型是在 ImageNet 的子集 ImageNet-1K 上训练得到的.在此过程中,随机初始化最后一个全连接层,并调整其维数以适配轴承的分类任务.通过在 ImageNet-1K 大量的图片数据上学习到的知识,将预训练模型应用于轴承数据集训练模型中.这一迁移可以在一定程度上提高模型分类精度,并减少所需的训练周期.

2 模型实验验证

2.1 实验数据集

轴承故障数据来自凯斯西储大学 (Case Western Reserve University, CWRU) 轴承数据中心^[20],轴承振动信号由轴承故障测试实验台采集获取,该数据集中轴承的故障均由人为在轴承不同位置加工产生,数据集包含传感器位置、电机转速、电机负荷、故障类型和故障程度等信息.

所用训练验证数据选用位于驱动端的加速度计以 48 kHz 的采样频率获取的振动数据,轴承型号为 SKF6205,选取轴承振动数据中的 3 种电机转速和 3 种电机负荷条件下的 4 种轴承状态.按故障类型和故障程度对数据集分类,每种类别均包含不同转速和负载数据,以实现在多种工况下对不同故障类型和故障程度的识别.电动机负

载与电动机转速对应的 3 种工况如表 2 所示. 选取的数据可分为正常轴承 1 种, 3 种滚动体故障和 3 种内圈故障, 外圈故障选用 2 种 3 点钟方向的故障, 共 9 个类别.

表 2 负载和对应电机转速对应的 3 种工况
Table 2 Three working conditions corresponding to load and corresponding motor speed

电动机负载/kW	电动机转速/(r·min ⁻¹)
0.735	1 772
1.470	1 750
2.206	1 730

由于试验台采样频率为 4 800 个/s, 且工况转速都接近 30 r/s, 轴承转一圈大约平均可产生近 160 个采样点. 考虑到需要足够多的样本点以确保学习效果, 同时为了减小信号样本初始点一致对样本的影响, 以及选用 2 的幂次方提高计算机

运算效率, 选择 2 048 个采样点作为一个样本. 为使各值均为整数, 选择 150 005 作为总样本点数, 平移步长为 993, 使得每种工况下的样本数量为 150 条, 最终每种故障程度的样本总数为 450 条. 表 3 为数据集划分情况, 图 10 则选取每个类别中的一张图像进行展示.

表 3 轴承故障数据集划分
Table 3 Division of bearing fault data set

状态	故障程度/mm	训练集	验证集	类别标签
正常	—	400	50	0
滚动体	0.177 8	400	50	1
	0.355 6	400	50	2
	0.533 4	400	50	3
内圈	0.177 8	400	50	4
	0.355 6	400	50	5
	0.533 4	400	50	6
外圈	0.177 8	400	50	7
	0.355 6	400	50	8

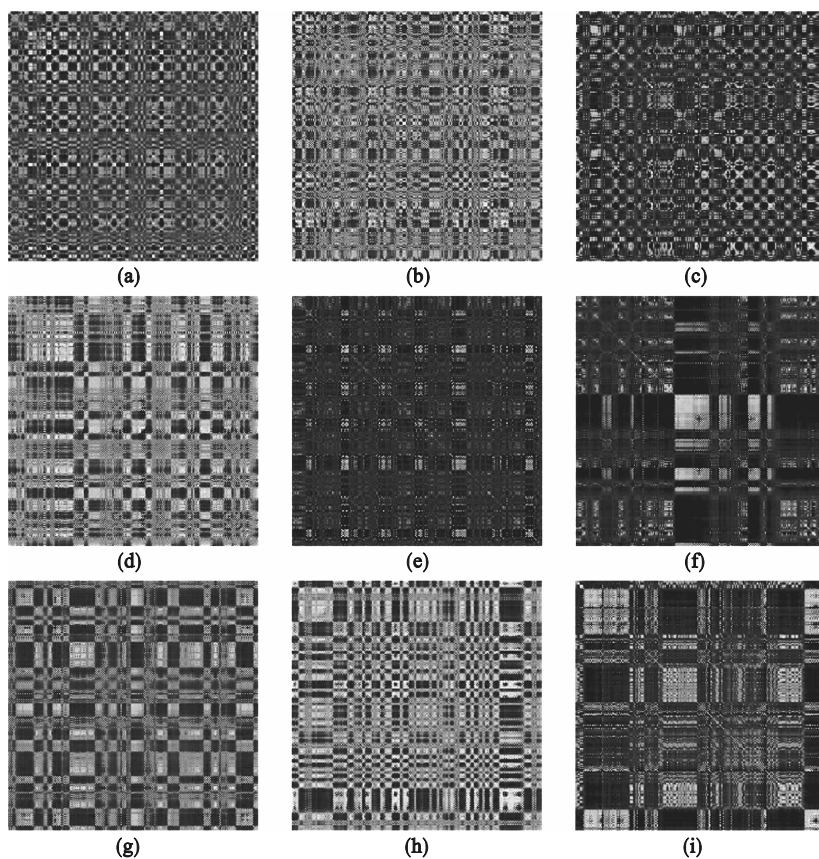


图 10 MTF 转换的图例

Fig. 10 Legends for MTF transition

(a)—正常轴承; (b)—0.177 8 mm 滚动体故障轴承; (c)—0.355 6 mm 滚动体故障轴承; (d)—0.533 4 mm 滚动体故障轴承; (e)—0.177 8 mm 内圈故障轴承; (f)—0.355 6 mm 内圈故障轴承; (g)—0.533 4 mm 内圈故障轴承; (h)—0.177 8 mm 外圈故障轴承; (i)—0.355 6 mm 外圈故障轴承.

2.2 实验结果

实验模型使用显存为 8 GB 的英伟达 RTX3070 显卡 GPU 进行训练, 训练轮数设置为

40, 考虑到要充分利用 GPU 显存和图片数据占用显存的容量, 设置每个训练批量为 50 个样本, 验证批量设置为 30 个样本, 选择使用 Adam 优化器.

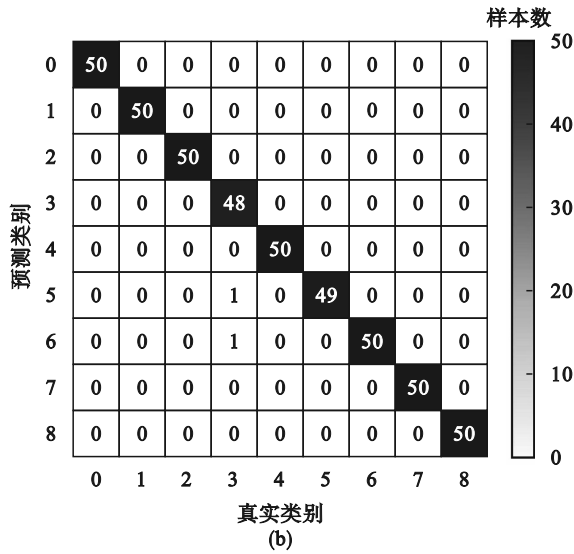


图 13 混淆矩阵

Fig. 13 Confusion matrix

- (a)—MTF-ResNeXt混淆矩阵;
- (b)—MTF-ResNeSt混淆矩阵.

使用 t 分布邻域嵌入 (t-distributed stochastic neighbor embedding, t-SNE) 降维可视化方法对模型分类结果进行评估. 图 14 和图 15 分别展示了两种模型的 t-SNE 可视化结果. 图 14a 和图 15a 显示的是所有种类的图像样本初始特征在二维坐标下的分布, 可见不同种类的特征随机分布于坐标中, 难以区分. 图 14b 和图 15b 则展示了经过模型学习后的图像样本特征再次投影到二维坐标轴上的分布. 显然, 不同类别的样本表现出明显的聚类效果, 表明模型学习到的特征能够有效地对故障的类型和故障程度进行诊断分类. 通过对比发现 ResNeSt 网络相较于 ResNeXt 网络学习到的不同类别特征在分布上更为分散, 这表明 MTF-ResNeSt 方法具有更优越的鲁棒性, 这可能与 ResNeSt 的注意力机制有关.

为了更好地展示所提模型方法的性能优势, 对比在 MTF 方法下神经网络结构的故障类型和故障程度分类性能. 所使用的数据集均为同一个 9 类别数据集. 为了充分发挥模型性能, 均采用迁移学习微调策略, 并与本文所提模型在相同数据集上进行微调预训练, 同时在相同设备环境进行目标模型训练. 结果为 ResNet18, ResNet34 和 ResNet50 的分类准确率分别是 98.44%, 98.89% 和 98.67%, 而 VGG19 分类准确率是 88.0%. 对比结果表明, 本文提出的 MTF-ResNeXt 和 MTF-ResNeSt 两种模型方法在模型精度上表现更为出色.

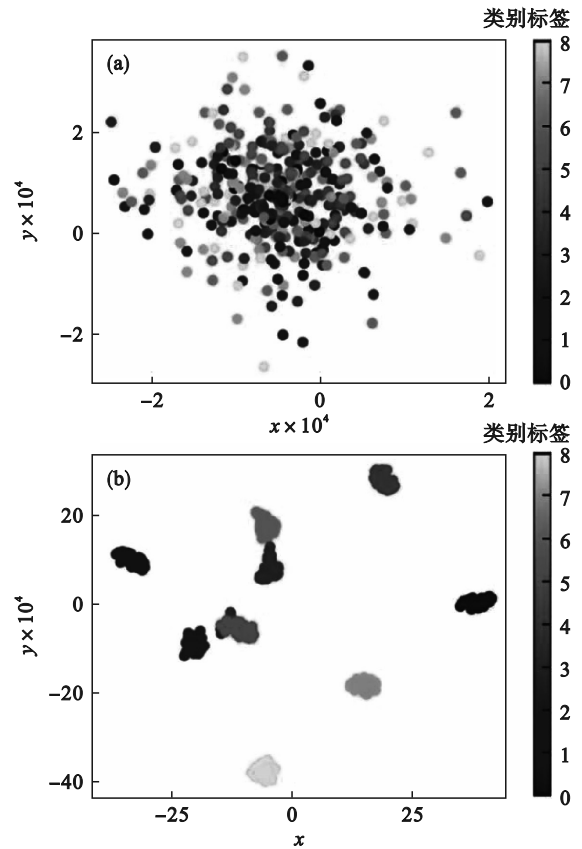


图 14 MTF-ResNeXt 的 T-SNE 可视化

Fig. 14 T-SNE visualization of MTF-ResNeXt

- (a)—初始分布;
- (b)—训练后分布.

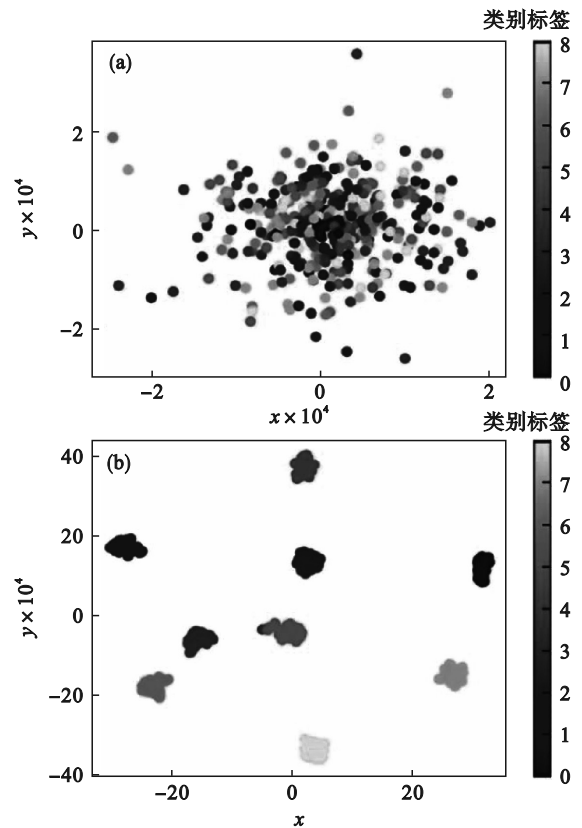


图 15 MTF-ResNeSt 的 T-SNE 可视化

Fig. 15 T-SNE visualization of MTF-ResNeSt

- (a)—初始分布;
- (b)—训练后分布.

3 结 语

提出了一种基于 MTF 和改进残差网络的滚动轴承智能故障诊断方法. 通过 MTF 方法, 将一维时间序列信号编码为二维图像, 以保留原始信号的时域信息. 在此基础上, 应用改进的残差网络对图像进行分类识别, 实现了 CWRU 轴承数据集故障类型和故障程度的定量诊断. 提出的 MTF-ResNeXt 和 MTF-ResNeSt 方法在故障诊断中表现出良好的特征提取和分类性能. 该模型还可以进一步简化诊断流程, 并与实时在线诊断平台整合, 从而实现对滚动轴承故障的实时在线诊断.

参考文献:

- [1] Cocconcelli M, Zimroz R, Rubini R, et al. STFT based approach for ball bearing fault detection in a varying speed motor [C]//2nd International Conference on Condition Monitoring of Machinery in Non-stationary Operations. Hammamet: Springer, 2012: 41-50.
- [2] Sun W, Yang G A, Chen Q, et al. Fault diagnosis of rolling bearing based on wavelet transform and envelope spectrum correlation [J]. *Journal of Vibration and Control*, 2013, 19(6): 924-941.
- [3] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 1998, 454(1971): 903-995.
- [4] Yu D J, Cheng J S, Yang Y. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings [J]. *Mechanical Systems and Signal Processing*, 2005, 19(2): 259-270.
- [5] 甄冬, 田少宁, 郭俊超, 等. 改进型 EEMD 和 MSB 解调方法及其在轴承故障特征提取中的应用 [J]. *振动工程学报*, 2023, 36(5): 1447-1456.
(Zhen Dong, Tian Shao-ning, Guo Jun-chao, et al. An improved decomposition method using EEMD and MSB and its application in rolling bearing fault feature extraction [J]. *Journal of Vibration Engineering*, 2023, 36(5): 1447-1456.)
- [6] Wu Z H, Huang N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method [J]. *Advances in Adaptive Data Analysis*, 2009, 1(1): 1-41.
- [7] 崔玲丽, 刘银行, 王鑫. 基于改进奇异值分解的滚动轴承微弱故障特征提取方法 [J]. *机械工程学报*, 2022, 58(17): 156-169.
(Cui Ling-li, Liu Yin-hang, Wang Xin. Feature extraction of weak fault for rolling bearing based on improved singular value decomposition [J]. *Journal of Mechanical Engineering*, 2022, 58(17): 156-169.)
- [8] Gu R, Chen J, Hong R J, et al. Incipient fault diagnosis of rolling bearings based on adaptive variational mode decomposition and Teager energy operator [J]. *Measurement*, 2020, 149: 106941.
- [9] Shenfield A, Howarth M. A novel deep learning model for the detection and identification of rolling element-bearing fault [J]. *Sensors*, 2020, 20(18): 5112.
- [10] Fu W L, Zhou J Z, Zhang Y C. Fault diagnosis for rolling element bearings with VMD time-frequency analysis and SVM [C]//2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC). Qinhuangdao, 2015: 69-72.
- [11] 路敦利, 宁芊, 杨晓敏. KNN-朴素贝叶斯算法的滚动轴承故障诊断 [J]. *计算机测量与控制*, 2018, 26(6): 21-23, 27.
(Lu Dun-li, Ning Qian, Yang Xiao-min. Fault diagnosis of rolling bearing based on KNN-naive Bayesian algorithm [J]. *Computer Measurement and Control*, 2018, 26(6): 21-23, 27.)
- [12] 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(2): 179-189.
- [13] Liu C Y, Gryllias K. Simulation-driven domain adaptation for rolling element bearing fault diagnosis [J]. *IEEE Transactions on Industrial Informatics*, 2022, 18(9): 5760-5770.
- [14] 王太勇, 王廷虎, 王鹏, 等. 基于注意力机制 BiLSTM 的设备智能故障诊断方法 [J]. *天津大学学报(自然科学与工程技术版)*, 2020, 53(6): 601-608.
(Wang Tai-yong, Wang Ting-hu, Wang Peng, et al. An intelligent fault diagnosis method based on attention-based bidirectional LSTM network [J]. *Journal of Tianjin University (Science and Technology)*, 2020, 53(6): 601-608)
- [15] 张伟. 基于卷积神经网络的轴承故障诊断算法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2017.
(Zhang Wei. Study on bearing fault diagnosis algorithm based on convolutional neural network [D]. Harbin: Harbin Institute of Technology, 2017.)
- [16] Wang Z G, Oates T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks [C]//Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, 2015: 275970614.
- [17] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 5987-5995.
- [18] Zhang H, Wu C R, Zhang Z Y, et al. ResNeSt: split-attention networks [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop (CVPR). New Orleans: IEEE, 2022: 2735-2745.
- [19] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks [C]//28th Conference on Neural Information Processing Systems (NIPS). Montreal, 2014: 8-13.
- [20] Case Western Reserve University Bearing Data Center. Data file [EB/OL]. (2018-06-10)[2021-03-10]. https://cse.groups.case.edu/bearing_datacenter/pages/download-data-file.