

doi:10.12068/j.issn.1005-3026.2025.20230296

齿轮故障机理嵌入的变负载智能故障诊断

于滨¹, 孙红春^{1,2}, 叶大勇^{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 航空动力装备振动及控制教育部重点实验室, 辽宁 沈阳 110819)

摘要: 在变负载条件下,基于机器学习的齿轮故障诊断模型面临着依赖特定目标工况样本训练的挑战.为了克服这一局限性,基于齿轮的故障机理,求解了信号中能够反映其健康状态且不随负载变化而改变的特征成分,以此构建了故障频率波形卷积模块,并将其内嵌于卷积神经网络中.此外,为增强网络的特征提取能力,引入多尺度注意力模块.基于上述模块,构建了变负载齿轮故障诊断模型(FWaveNet),将其应用于东北大学的齿轮故障数据集,结果显示其诊断精度相较于现有模型有显著提升.通过特定的信号处理技术和网络架构设计,在负载波动情况下实现了对齿轮健康状态的精确识别,为变负载齿轮故障诊断的工程应用提供了一种解决方案.

关键词: 深度学习;变负载;故障诊断;故障机理;齿轮

中图分类号: TH 165+3;TH 1.41 文献标志码: A 文章编号: 1005-3026(2025)04-0061-10

Intelligent Fault Diagnosis Under Variable Loads Based on the Embedded Gear Fault Mechanism

YU Bin¹, SUN Hong-chun^{1,2}, YE Da-yong^{1,2}

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

2. Key Laboratory of Vibration and Control of Aero-Propulsion Systems, Ministry of Education, Northeastern University, Shenyang 110819, China. Corresponding author: SUN Hong-chun, E-mail: hchsun@mail.neu.edu.cn)

Abstract: Under variable load conditions, machine learning-based gear fault diagnosis models face the challenge of relying on specific target condition samples for training. To overcome this limitation, the feature components in the signal that can reflect the health status of gears and remain invariant to load variations were solved based on the gear fault mechanism, thereby constructing a fault frequency waveform convolution module and embedding it into the convolutional neural network. Additionally, to enhance the network's feature extraction capability, a multi-scale attention module was introduced. Based on these modules, a variable load gear fault diagnosis model named FWaveNet was constructed and applied to the gear fault dataset from Northeastern University. The results showed that its diagnostic accuracy is significantly better than that of existing models. Through specific signal processing techniques and network architecture design, precise identification of gear health status under load fluctuations is achieved, and a solution for engineering applications in the fault diagnosis of variable load gears is provided.

Key words: deep learning; variable load; fault diagnosis; fault mechanism; gear

齿轮箱作为重要的动力传动部件被广泛应用于旋转机械中,但长期恶劣的工作环境易使其发生故障,从而导致巨大的经济损失甚至人员伤亡^[1].目前,基于深度学习的智能故障诊断技术正

在快速发展,而齿轮箱工况相对复杂,负载等参数时常变化,很难获得所有负载下的故障样本.因此,研究在变负载工况下准确而高效的齿轮故障诊断方法,对智能故障诊断技术的实际应用具

收稿日期: 2023-10-26

基金项目: 国家科技重大专项(J2019-I-0008-0008).

作者简介: 于滨(1996-),男,山东枣庄人,东北大学硕士研究生;孙红春(1974-),女,辽宁葫芦岛人,东北大学副教授,博士生导师.

有至关重要的作用。

目前,基于深度学习的故障诊断方法在处理大数据以及非线性关系拟合方面表现出了极大的优势,并得到了广泛的应用。例如, Li 等^[2]提出了基于空洞卷积神经网络(CNN)和空间 Dropout 的神经网络,实现了齿轮早期点蚀的检测。Emmanuel 等^[3]使用小波去噪函数对信号进行预处理,并将其转换为二维图像,并作为二维卷积神经网络的输入,实现了汽车行星齿轮的裂纹检测。张智禹等^[4]利用深度置信网络具有的贪心学习优势对特征集进行挖掘,同时结合注意力机制自适应地对描述齿轮箱状态的有效特征给予更多“注意”,从而提高齿轮箱故障诊断的精度。

然而,由于深度学习方法是通过训练集的输入-输出关系进行拟合,将模型在训练集上学到的知识应用于后续过程。因此,上述深度学习方法所取得的成就是基于同分布假设^[5]。即训练深度学习所使用的数据集分布需要尽可能地接近模型将被应用的实际场景。一旦二者分布不同,深度学习方法不仅失去了用于故障诊断的理论支撑,也难以取得良好的诊断效果^[6]。

目前,迁移学习方法常被用来解决变负载条件下的旋转机械故障诊断问题。如 Li 等^[7]使用粒子群优化算法和 ELU 激活函数, L2 正则等技巧优化深度学习网络,在使用原数据集完成网络训练后,使用少量目标工况下的数据集对网络进行微调,用于变工况齿轮的早期点蚀故障诊断。Zhang 等^[8]在模型训练过程中计算了训练集和少量目标工况样本在特征空间中的最小均方距离(MMD),并将其写入稀疏滤波模型的损失函数中,从而实现跨工况的齿轮故障诊断。赵桐等^[9]利用迁移学习解决当前由于训练数据少而导致识别率低的问题,设计了一种基于 CNN 的针对金属注射成形(MIM)工艺生产的小模数齿轮进行缺陷检测的方法。但基于迁移学习的方法需要预先收集和标注目标工况下的数据,增加了实际应用的难度。

针对上述问题,本文提出了一种新的变负载齿轮故障诊断网络,不依赖于目标工况数据和迁移学习。主要贡献包括:①提取了反映故障状态且不受负载变化影响的频域特征;②设计了故障频率波形卷积层(FWave Layer),实现了故障特征的自提取;③结合 FWave Layer 和多尺度注意力模块,构建了 FWaveNet 模型,无须迁移学习即可进行变负载故障诊断;④通过自制齿轮箱进行了实验,验证了模型的有效性。

1 齿轮故障机理嵌入卷积神经网络分析

1.1 齿轮故障机理

定轴齿轮箱在工作过程中,由于参与啮合的轮齿在单双数之间交替变化,其啮合刚度是随时间变化的周期性函数。因此,齿轮因正常工作而产生的信号是以啮合频率为基频的谐波叠加函数^[10-11],其可以简单表示为

$$x(t) = \sum_{m=1}^M A_m \cos(2\pi m f_{\text{mesh}} t + \theta_m). \quad (1)$$

式中: M 为谐波总阶数; f_{mesh} 为轮齿的啮合频率; θ_m 为谐波信号中第 m 阶谐波对应的初始相位; A_m 为其对应幅值。

假设某一齿轮的轮齿表面发生了损伤,当该损伤轮齿与相配合的齿轮啮合时,该损伤将导致齿轮振动信号发生变化,从而对齿轮的振动信号产生幅值调制和频率调制。这一结论已被许多研究^[12-14]证明,其数学模型表示为

$$x(t) = \sum_{m=1}^M a_m(t) \cos(2\pi m f_{\text{mesh}} t + b_m(t) + \theta_m). \quad (2)$$

其中, $a_m(t)$ 和 $b_m(t)$ 分别为调幅和调频函数,可表示为

$$a_m(t) = c \left[1 + \sum_{n=1}^N A_{mn} \cos(2\pi n f_{\text{fault}} t + \phi_{mn}) \right], \quad (3)$$

$$b_m(t) = \sum_{n=1}^N B_{mn} \sin(2\pi n f_{\text{fault}} t + \phi_{mn}). \quad (4)$$

其中: c 取决于故障信号振幅的常数; N 为故障频率总阶数; n 为分析故障频率时所考虑的信号阶次; ϕ , φ 为信号中各部分的初始相位; f_{fault} 为故障信号所出现的频率; B_{mn} 为第 n 阶调频信号的幅值。对于如齿根裂纹、齿面剥落等轮齿损伤, f_{fault} 等于故障轴的转频。

为简化计算过程,对信号中各分量仅取至一阶成分,将 $a_m(t)$, $b_m(t)$ 代入式(2),有

$$x(t) = [1 + A_{11} \cos(2\pi f_{\text{fault}} t + \phi_{11})] \times \cos[2\pi f_{\text{mesh}} t + B_{11} \sin(2\pi f_{\text{fault}} t + \phi_{11}) + \theta_1]. \quad (5)$$

将其展开:

$$x(t) = \cos[2\pi f_{\text{mesh}} t + B_{11} \sin(2\pi f_{\text{fault}} t + \phi_{11}) + \theta_1] + A_{11} \times \cos[(2\pi(f_{\text{mesh}} + f_{\text{fault}})t + B_{11} \sin(2\pi f_{\text{fault}} t + \phi_{11}) + \varphi_{11} + \theta_1)]/2 + A_{11} \cos[(2\pi(f_{\text{mesh}} - f_{\text{fault}})t - B_{11} \sin(2\pi f_{\text{fault}} t + \phi_{11}) - \varphi_{11} - \theta_1)]/2. \quad (6)$$

易知,式(6)中的第一项在频谱上以 f_{mesh} 为中心频率,并伴有以 f_{fault} 为间隔的边频族;而后两项则分别以 $f_{\text{mesh}} + f_{\text{fault}}$, $f_{\text{mesh}} - f_{\text{fault}}$ 为中心频率,伴有以 f_{fault} 为间隔的边频族。

因此,故障齿轮的振动信号在频域内将表现为以啮合频率 f_{mesh} 为中心,并以齿轮转频 f_{fault} 为间隔向两端逐级衰减的频谱成分,如图1所示.在齿轮的变负载故障诊断中,可以考虑提取信号频谱中的调制频谱成分,如图1所示,将其作为故障状态的判别依据.

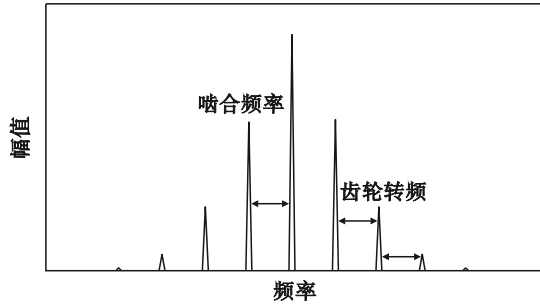


图1 故障齿轮频谱特征成分

Fig. 1 Spectral characteristic components of the fault gear

1.2 一维卷积层

目前,CNN的相关理论已足够完善,以一维和二维振动信号作为输入的神经网络^[15-16]已取得很好的诊断效果.本文将只对CNN中卷积核的计算过程进行说明,为后续的网络搭建奠定理论基础.

对于神经网络中的卷积层,其接收输入网络长度为 L 的序列 x ,使用宽度为 n 的卷积核,在该输入序列上依次进行点积运算,表示为

$$z(u) = \sum_{i=1}^n x(u+i)w(i) + b. \quad (7)$$

式中: z 为卷积层的输出特征值; w 为卷积核的权重参数; b 为偏置; $u=0, 1, \dots, L-n$.在CNN的前向传播过程中, z 将被输入到网络的后续层级,经由网络后续部分的计算得到网络的最终输出 y .在故障诊断任务中, y 为含有 c 个元素的向量,每个元素表示网络所判定的输入序列 x 对应的齿轮一类健康状态的概率.使用交叉熵函数将 y 与样本的实际标签 y^* 比较,从而得到网络的损失Loss:

$$\text{Loss} = - \sum_{i=1}^c y_i^* \lg(y_i). \quad (8)$$

在网络训练过程中,通过对网络各层参数计算损失的梯度来更新网络参数,使网络对输入样本的损失最小,从而能够计算出其对应的正确标签.

$$w^* = w - \Delta w = w - \frac{\partial \text{Loss}}{\partial w}. \quad (9)$$

更新后的权重 w^* 从网络的输入中提取特征

信息,从而用于后续的故障诊断.在式(7)中,不难得知^[17],与卷积核的 w 越相似的信号向量,其点积得到的值越大.因此,对于一段输入序列 x ,在与卷积核的所对应的波形越相似的位置上,输出特征图的值越大.利用这一特性,本文构造了FWave Layer,使用所关注的波形为卷积核赋值,并固定该处的权重,使之不会因网络迭代而改变其波形.此设计用于从网络的输入中提取该特征波形并将之增强,作为后续诊断网络的输入.

1.3 齿轮故障机理的嵌入

由1.1节可知,为提取齿轮信号中不随负载变化而变化的故障特征成分,关键在于检测并提取其频带中以故障轴转频为间隔分布的边频族.根据1.2节的分析,CNN中的卷积核可以有效提取输入特征图中与卷积核波形相似的成分.根据计算得出的波形,设计的齿轮故障频率波形(FWave)为

$$y = \left\{ \begin{array}{l} \frac{x}{f_{\text{fault}}} \left\lfloor \frac{x}{f_{\text{fault}}} \right\rfloor, x \in (0, 0.5f_{\text{fault}}] \cup (2f_{\text{fault}}, 2.5f_{\text{fault}}]; \\ \frac{x}{f_{\text{fault}}} \left\lfloor \frac{x}{f_{\text{fault}}} \right\rfloor, x \in (0.5f_{\text{fault}}, f_{\text{fault}}] \cup (2.5f_{\text{fault}}, 3f_{\text{fault}}]; \\ \frac{2}{f_{\text{fault}}} x - 2, x \in (f_{\text{fault}}, 1.5f_{\text{fault}}]; \\ 4 \frac{2}{f_{\text{fault}}} x, x \in (1.5f_{\text{fault}}, 2f_{\text{fault}}]. \end{array} \right. \quad (10)$$

其中, $\lfloor x \rfloor$ 为小于 x 的最大整数.该函数的波形如图2所示:

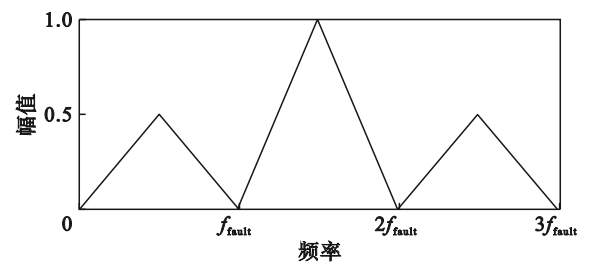


图2 FWave 波形

Fig. 2 FWave waveform

在该函数最大峰值的两侧 f_{fault} 处均存在较小幅值的峰值.在各峰值附近,函数的值以一定斜率下降,以避免因为故障轴转频的微小变化导致该波形无法提取输入中的有效特征.

根据1.2中所述,使用上述波形对CNN中的卷积核赋值,并固定其参数,使之不会因网络训练而改变.通过卷积核在输入特征图上的顺序点积操作,可以依次提取和加强输入频谱不同频段

上能够反映齿轮故障的调制边频族.以图 1 所示的频谱为例,在其时域信号中添加其他主频率、调制频率和噪声后,得到如图 3a 所示的频谱(图中 f_{mesh} 处为由于齿轮故障产生的以 f_{fault} 为间隔的调制频族, f_1 处为引入的与齿轮故障无关的定频分量, f_2 处为与齿轮故障无关的调制频族).使用 FWave Layer 对该频谱进行特征提取后,得到如图 3b 所示的输出特征图.

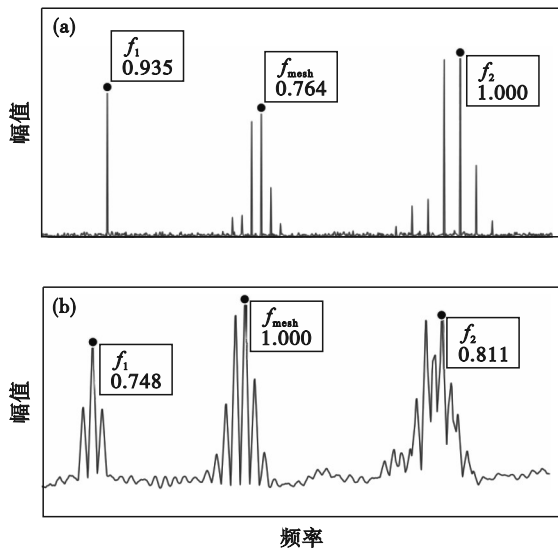


图 3 FWave Layer 处理前后的频谱变化

Fig. 3 Spectrum changes before and after processing by FWave Layer

(a)—FWave Layer 处理前; (b)—FWave Layer 处理后.

可以看出,经过 FWave Layer 的特征提取,在故障诊断中所关注的故障特征成分被有效放大,相对而言,其他与诊断无关的频率分量被抑制.本文将 FWave Layer 作为诊断网络的首层,用于对输入频谱进行预处理,提取并放大其中的故障特征.

2 基于故障机理嵌入的 FWaveNet 诊断模型

2.1 多尺度注意力模块

为增强诊断网络的特征提取能力,本文构建了多尺度卷积层,在同一网络层级中引入不同宽度的卷积核,将不同尺度的特征同时输入到网络的后续层级,以尽可能多地从输入中挖掘信息.多尺度卷积层的计算过程为

$$\left. \begin{aligned} \mathbf{Z} &= \text{concat}(\mathbf{z}^1, \mathbf{z}^2, \dots, \mathbf{z}^c), \\ \mathbf{z}^i &= [\mathbf{z}^i_1, \mathbf{z}^i_2, \dots, \mathbf{z}^i_l], \\ \mathbf{z}^i_j &= \sum_{m=0}^{n^i-1} w_m^i x_{j+m} + b_m^i. \end{aligned} \right\} \quad (11)$$

其中: $j=1, 2, \dots, l$; \mathbf{Z} 为多尺度卷积层输出; \mathbf{z}^i 为多尺度卷积层中卷积核宽度为 n^i 的卷积层输出的特征图; \mathbf{z}^i_j 为 \mathbf{z}^i 中第 j 个像素点; l 为特征图总长; x_j 为该层输入特征图中第 j 个像素点; w_m^i 为宽度为 n^i 的卷积核的第 m 个权重参数; b_m^i 为其偏置.

为了在网络诊断中赋予重要通道更高的权重,并减少对任务贡献较小的特征对网络后续部分的影响,引入如式(12)所示的通道注意力机制,以对多尺度卷积层输出的特征进行加权.

$$\left. \begin{aligned} \mathbf{Y} &= [\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^c] = [\text{att}^1 \mathbf{Z}^1, \text{att}^2 \mathbf{Z}^2, \dots, \text{att}^c \mathbf{Z}^c], \\ \text{s.t. } \sum_{i=1}^c \text{att}^i &= 1. \end{aligned} \right\} \quad (12)$$

其中, att^i 用于衡量卷积核宽度为 n^i 的卷积层输出中每个通道的重要程度,通过对 \mathbf{Z} 进行全局平均池化,并由 2 个全连接层和 softmax 函数构成的小型神经网络计算出各通道对应的 att 值,从而根据不同通道的重要程度对其进行加权.这一计算过程如式(13)所示.

$$\left. \begin{aligned} \text{att} &= \text{Sigmoid}(w_2 \text{ReLU}(w_1 \text{GAP}(\mathbf{Y}) + b_1) + b_2), \\ \text{GAP}(\mathbf{Y}) &= \left[\frac{1}{l} \sum_j y_j^1, \frac{1}{l} \sum_j y_j^2, \dots, \frac{1}{l} \sum_j y_j^c \right]. \end{aligned} \right\} \quad (13)$$

多尺度卷积层与通道注意力机制共同构成了多尺度注意力模块,其计算过程如图 4 所示.

2.2 FWaveNet 架构

本文所提出的 FWaveNet 架构如表 1 所示.网络的输入为对振动信号进行快速傅里叶变换 (FFT) 后得到的频率幅值谱.网络首层为 FWave Layer,用于提取和增强输入频谱中的故障频率特征,经由 FWave Layer 增强后的特征图再与原频域信号进行通道合并,共同作为后续层级的输入.第 2, 3 层为多尺度注意力模块,通过多尺度卷积和通道注意力机制自主学习输入特征图的信息.随后,经由 2 个卷积层和全连接层进行特征的进一步浓缩和样本分类,并最终输出诊断结果.

2.3 诊断流程

在变负载情况下齿轮故障诊断的流程如图 5 所示.首先通过加速度传感器采集齿轮在工作过程中产生的振动信号,并依据齿轮所处的负载将其划分为不同的数据集,依据齿轮的转速、传感器的采样频率和样本长度计算该任务下的 FWave 波形,生成相应的诊断网络.随后使用某一工况下的数据集进行网络训练,使用其他工况下的数据集测试诊断效果,从而将该网络应用于实际故障诊断.

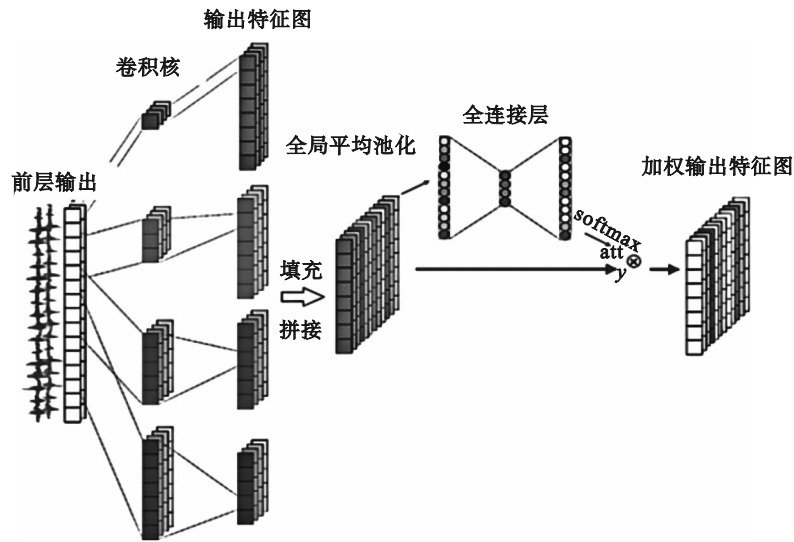


图 4 多尺度注意力模块计算过程
Fig. 4 Multi-scale attention module calculation process

表 1 FWaveNet 架构
Table 1 Architecture of FWaveNet

层	层的类型	核尺寸	步幅	输出
1_FWave Layer	卷积核	1×9	1	2×1024
2_Multi-scale attention	卷积核	1×1, 1×3, 1×5, 1×7	1	24×1024
2_Pool	最大池化	1×2	2	24×512
3_Multi-scale attention	卷积核	1×1, 1×3, 1×5, 1×7	1	96×512
3_Pool	最大池化	1×2	2	96×256
4_Conv	卷积核	1×3	1	192×256
4_Pool	最大池化	1×2	2	192×128
5_Conv	卷积核	1×3	2	192×64
GAP	全局平均池化	—	—	192×1
Full connected_1	全连接层	—	—	128
Full connected_2	全连接层	—	—	32
Full connected_3	全连接层	—	—	类别

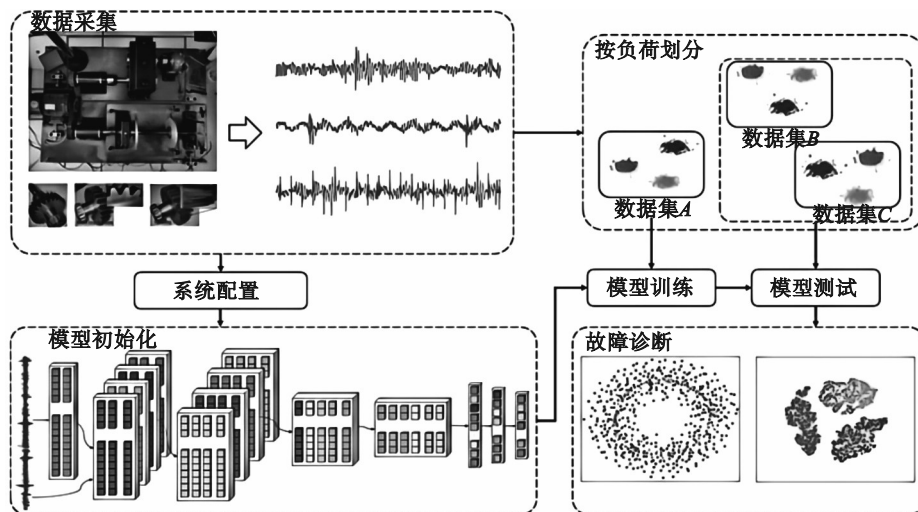


图 5 故障诊断流程
Fig. 5 Fault diagnosis process

3 实验研究

3.1 实验设置和数据描述

实验中使用的数据集来自如图 6 所示的齿轮故障试验台.系统动力由电机提供,其转速为 900 r/min,并通过传动带和联轴器将动力输入至齿轮箱中.齿轮箱的输出端连接磁粉制动器,用于模拟系统的工作负载,本文中设置了 0, 0.2, 0.4 N·m 3 组工作负载.如图 7 所示,本实验在齿轮箱驱动轮齿根处加工了 3 mm 深的缺陷,以模拟齿根裂纹;在齿面加工了 15 mm 长的凹坑,以模拟齿面剥落,同时用健康齿轮作为对比组.

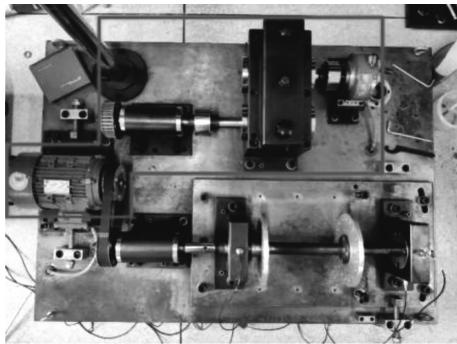


图 6 东北大学齿轮故障试验台

Fig. 6 Test bench for the gear failure at Northeastern University

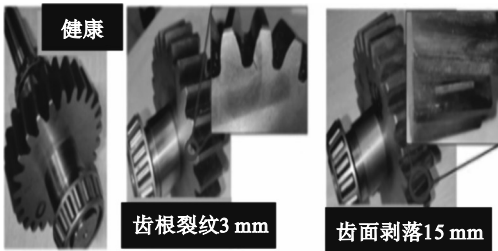


图 7 齿轮健康状态说明

Fig. 7 Description of the gear's health status

通过放置在箱体外壳上从动轮上方的加速度传感器以 10 kHz 的采样频率收集齿轮运行过程中的振动信号,构成本实验数据集.依据系统所处的负载情况将数据集分为 A, B, C 3 组,分别对应 0, 0.2, 0.4 N·m 这 3 种工作负载.数据集的具体构成如表 2 所示,每个样本包含 1 024 个采样点的振动信号.

在本文的实验部分,为得到待测模型的跨负载故障诊断性能,分别使用某一负载下的数据集训练模型,并用另外 2 种负载下的数据集测试该模型,共计 6 组诊断任务.每个任务重复 50 次实

验,并取 95% 置信区间的诊断准确率作为模型性能的评价指标.

表 2 数据集构成

Table 2 Composition of the dataset

健康状态	样本数		
	A	B	C
健康	1 000	1 000	1 000
齿根裂纹	1 000	1 000	1 000
齿面剥落	1 000	1 000	1 000

3.2 消融实验

为解决变负载条件下的齿轮故障诊断问题,基于齿轮故障机理求解出齿轮振动信号在频域上不受负载情况影响的共性特征,并设计 FWave Layer 用于从输入频谱中提取该特征.为自适应地从 FWave Layer 的输出和齿轮原频谱中提取合适的特征成分,本文设计了多尺度注意力模块,从输入特征图中学习多尺度的特征并进行自适应加权.为验证本文主要创新结构在故障诊断中的有效性,本节设计了消融实验,实验对比项的细节如下:

1) Baseline.

本文的 Baseline 为包含 5 个卷积层的神经网络架构. Baseline 的首层是一个核宽为 1×9 的普通卷积层,第 2, 3 层的核宽均为 1×5,以保证网络具有较大的感受野.其余结构与本文所提网络相同.

2) Baseline+FWave Layer(B+FWave).

为探究本文方法中使用多尺度注意力模块在 FWave Layer 的输出和信号频谱中自适应学习多尺度特征的思想对诊断任务的有效性,在实验中将多尺度注意力模块改为核宽为 1×5 的普通卷积层,并保持每个层级的输出通道数与原 FWaveNet 相同,以验证该结构的有效性.

3) Baseline+Multi-scale attention module(B+MA).

为探究 FWaveNet 中利用齿轮故障频域波形为卷积核赋值的思想对跨工况故障诊断的有效性,在实验中保留原 FWave Layer 的结构,但不再使用故障波形为之赋值,并使其能够随着网络训练而更新参数.

在齿轮变负载故障数据集中分别测试了上述对比方法和所提方法的性能,得到消融实验的结果如表 3 所示.

表3 不同模型的消融实验准确率
Table 3 Accuracy of the ablation experiments for different models

模型	A→B	A→C	B→A	B→C	C→A	C→B	平均值
Baseline	98.96±0.12	78.41±1.33	99.47±0.1	85.38±0.74	87.91±1.05	91.64±1.55	90.30±0.37
B+FWave	99.96±0.01	87.72±1.48	99.95±0.01	88.93±1.07	95.02±0.70	98.83±0.18	95.07±0.33
B+MA	99.07±0.13	79.22±0.91	99.68±0.05	86.93±0.47	89.25±1.34	92.36±0.66	91.09±0.33
本文方法	99.98±0.01	94.82±0.34	99.97±0.01	98.32±0.27	97.62±0.18	99.45±0.05	98.36±0.14

由表3可知,本实验中的A→B和B→A任务相较于其他任务较为简单,因此所有实验项均取得了较好的诊断结果.在Baseline的基础上加入多尺度注意力模块后,网络的准确率虽有所提升,但提升幅度较小.而本文引入FWave Layer方法后,网络性能有了较大提升,说明本文利用齿轮故障机理进行跨工况故障诊断的方法是有效的.最终,结合了FWave Layer和MA的方法取得了最高的诊断准确率.

3.3 对比实验

为了验证该方法相比于同类方法具有更好的效果,本实验采用了如下对比模型:

1) AlexNet模型.本文采用了与AlexNet^[18]相似的5层网络架构,目前有大量旋转机械故障诊断工作采用AlexNet设计的网络,为此本文也将AlexNet作为对比模型,分别以信号的原始时间序列和频谱作为输入,记作AlexNet-T和AlexNet-F.

2) ResNet模型.He等^[19]通过添加额外的连接来解决梯度消失和梯度爆炸等问题,从而允许构建非常深的神经网络.选择网络层数较少的ResNet18网络作为对比模型,分别以信号的原始时间序列和频谱作为输入,记作ResNet-T和ResNet-F.

3) WKCNN模型.Song等^[20]提出使用宽卷积核构造神经网络,从而增大网络的感受野,提高其泛化能力,用于噪声下的跨工况故障诊断.同样,本文分别以信号的原始时间序列和频谱作为输入,记作WKCNN-T和WKCNN-F.

4) TICNN模型.Zhang等^[21]使用变Dropout率的宽卷积核、小Batch-size等技巧训练网络,使网络在较高噪声下具有较强的变负载故障诊断能力.本文分别以信号的原始时间序列和频谱作为输入,记作TICNN-T和TICNN-F.

5) CNLSTM模型.Hao等^[22]结合CNN与LSTM的性能设计网络,使网络可以提取时间序列中的空间和时间特征.该方法对旋转机械工作负载和信噪比具有较强的适应性.由于该方法是

基于时间序列设计的,因此在对比实验中仅用时域信号作为网络输入,记为CNLSTM-T.

6) 1D-MSDAN模型.Wang等^[23]提出了一种基于领域适应的迁移学习方法,构造了多尺度卷积神经网络,能够从输入中进行特征提取,并使用最大均值差异(MK-MMD)作为分布度量,从而使网络获得变工况条件下的故障诊断能力.本文分别以信号的原始时间序列和频谱作为该网络的输入,记作MSDAN-T和MSDAN-F.

7) DANN模型.Ganin等^[24]将领域自适应嵌入特征学习过程中,使分类决策基于具有领域不变性或领域相似性的特征.分别以信号的原始时间序列和频谱作为输入,记作DANN-T和DANN-F.

使用上述对比模型,在本文数据集中重复前述实验,并取50次实验得到的95%置信区间作为评价指标,实验结果如表4所示.

观察上述实验结果,可得出结论如下:

1) 相比于其他对比方法,本文方法在齿轮变负载下的故障诊断中取得最佳效果.尤其相比于WKCNN, CNLSTM, TICNN和DANN等泛化网络,在各种任务下本文方法均能取得优于这些方法的效果.这表明其具有极佳的泛化能力,足以胜任变负载下的故障诊断.

2) 在齿轮的变负载故障诊断任务中,相比于以时域信号作为网络输入,使用频谱信号作为输入可以获得更高的准确率.在对比模型如AlexNet, ResNet, WKCNN, TICNN, 1D-MSDAN的测试中,以时域信号作为输入难以取得较好的诊断效果,而以频谱作为输入时,虽这些方法较FWaveNet性能欠佳,但都取得了高于90%的平均准确率.

3) 本文方法能够在不使用目标域数据参与训练的条件下获得较好的跨负载诊断效果.相比于目前典型的域泛化方法,本文方法摆脱了训练时对目标域样本的需求,依靠嵌入的变负载故障机理实现了跨域诊断.在对比项中,MSDAN-F取得了仅次于FWaveNet的诊断效果.该方法使用

领域适应方法,在网络的训练过程中引入了目标域信息,从而获得了跨负载诊断的效果.在 $A \rightarrow C$ 任务中,该方法的诊断准确率甚至高于FWaveNet.

但本文的FWaveNet不需要使用迁移学习方法,在训练过程中不会引入目标域的信息,即可取得与MSDAN相似,甚至优于其的诊断效果.

表4 不同模型的准确率
Table 4 Accuracy of different models

模型	$A \rightarrow B$	$A \rightarrow C$	$B \rightarrow A$	$B \rightarrow C$	$C \rightarrow A$	$C \rightarrow B$	平均值
AlexNet-T	92.98±3.00	63.50±2.71	98.63±0.56	68.78±0.96	65.63±4.06	65.82±3.41	75.89±1.05
AlexNet-F	99.97±0.01	76.69±0.89	99.98±0.01	78.78±1.66	90.28±2.00	97.51±1.61	90.54±0.56
ResNet-T	94.46±1.04	77.83±0.44	97.22±0.34	77.36±0.49	75.05±1.37	90.14±0.64	85.34±0.72
ResNet-F	99.35±0.62	71.81±0.54	99.90±0.10	72.80±0.59	96.62±0.48	95.54±0.60	89.33±0.48
WKCNN-T	96.11±1.28	54.32±1.10	87.14±0.75	52.94±1.13	48.54±3.09	65.07±6.75	67.35±1.20
WKCNN-F	99.75±0.09	85.43±1.46	99.94±0.02	80.68±1.41	90.34±0.86	99.05±0.13	92.53±0.39
TICNN-T	86.51±3.25	63.86±1.49	66.65±0.02	51.36±2.58	41.96±3.74	62.84±3.83	62.20±1.45
TICNN-F	99.50±0.24	79.84±1.54	99.52±0.21	83.01±1.56	87.65±1.60	95.71±1.03	90.87±0.60
CNNLSTM-T	99.73±0.17	63.67±2.31	86.89±1.67	59.55±1.78	54.87±1.72	41.77±1.76	67.75±0.72
MSDAN-T	99.74±0.24	96.57±2.88	99.90±0.04	78.50±7.33	95.56±2.45	82.36±7.71	92.11±1.97
MSDAN-F	99.82±0.09	94.43±0.69	99.83±0.06	96.32±0.51	99.26±0.24	99.39±0.16	98.18±0.15
DANN-T	86.11±0.12	74.84±0.41	90.68±0.26	68.69±0.31	64.20±0.18	54.70±0.33	73.20±0.27
DANN-F	87.84±0.64	64.96±0.84	80.72±0.22	77.50±0.37	69.10±0.47	71.50±0.24	89.91±0.46
FWaveNet	99.98±0.01	94.82±0.34	99.97±0.01	98.32±0.27	97.62±0.18	99.45±0.05	98.36±0.14

4) 使用复杂的网络结构可能会提高诊断准确率,但是会带来降低诊断实时性的问题.通过对比发现,具有18层网络结构的ResNet18在变负载诊断上的准确率会比仅有5层结构的AlexNet网络更高,但是仍低于嵌入了故障机理的本文网络.实验发现,在相同设备下,ResNet18每次运行所用的时间是本文网络的5倍,说明结构复杂的网络会导致更长的运算时间,影响诊断的实时性.

3.4 诊断性能的可解释性探讨

为了证明所提方法具有更佳的分性能 and 更好的诊断效果,对各模型从训练集和测试集样本提取的特征进行了 t 分布随机邻域嵌入(t -SNE)可视化分析,以 $C \rightarrow A$ 任务为例,如图8所示.从图8可以看出,在经数据集 C 训练的诸多对比网络中,只有MSDAN和FWaveNet能够将数据集 A 中的样本映射到特征空间中与数据集 C 相同的位置.也就是说,不同负载下的样本对于这两种网络而言没有显著区别.可以推论,2种网络能够有效提取信号中反映齿轮故障状态的特征,而忽略负载不同对诊断结果的干扰.而对于其他网络,虽然能够将数据集 C 中的样本予以区分,但将数据集 A 输入到网络中,这些网络无法将其映射到特征空间中与 C 完全相同的位置,导致其在

跨负载任务中诊断准确率较低.

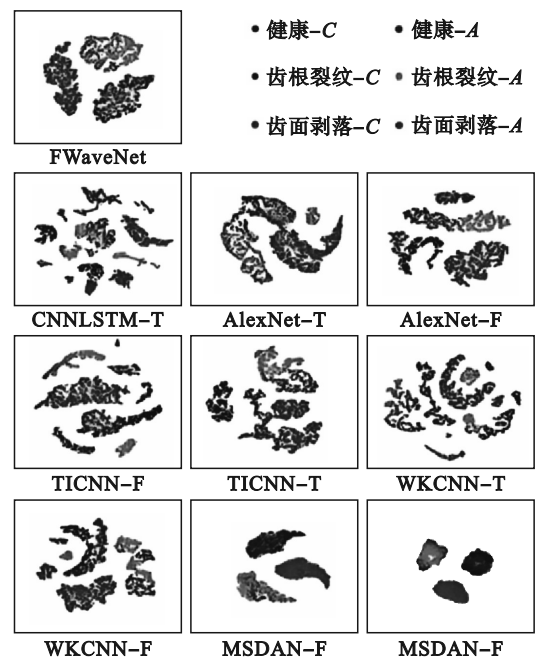


图8 $C \rightarrow A$ 任务下的各对比模型特征可视化

Fig. 8 Visualization of features for various comparison models under $C \rightarrow A$ tasks

为进一步探究本文FWaveNet中各关键层级对输入信号的作用,以 $C \rightarrow A$ 任务为例,对模型中各关键层级(FWave Layer, Multi-scale attention 2, Convolution Layer 5)从训练集和测试集中提取

的特征进行 t -SNE 可视化分析, 如图9所示。

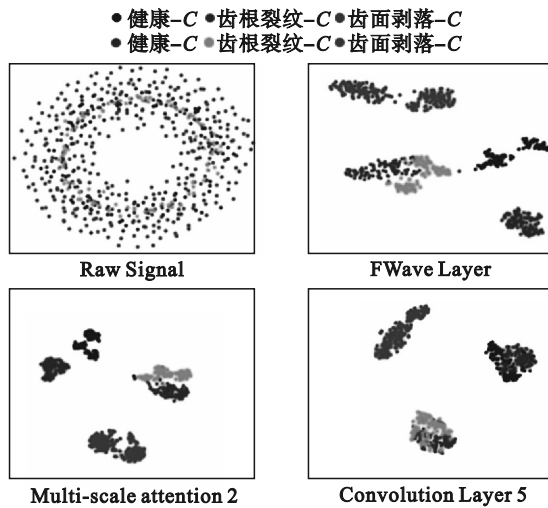


图9 $C \rightarrow A$ 任务下的FWaveNet关键层级特征可视化
Fig. 9 Visualization of key levels of FWaveNet under $C \rightarrow A$ tasks

观察图9可以看出,在FWaveNet中,经过FWave Layer后,输入频谱中的关键信息得到了增强,使不同类别的样本能够被区分开,同时,相同故障类型的样本近似分布在特征空间内的相近位置.经过后续的多尺度注意力层和卷积层后,不同类型样本间的差异更加明显,在不同负载下相同故障类型样本的距离进一步缩小.在Convolution Layer 5中,不同负载下相同故障类型的样本被映射到特征空间中的相同位置.由此可以推断所提出的方法在特征嵌入阶段可以学习到有效的关键分类特征,并有效排除变负载条件对诊断造成的不利影响。

4 结 论

1) 为了提高变负载情况下基于深度学习的齿轮故障诊断模型的性能,本文基于齿轮故障机理,求解了其频谱中不受负载影响的特征成分,并构造了故障频率波形卷积层(FWave Layer)。

2) 基于该结构和多尺度注意力模块,设计了不依赖目标工况数据即可实现变负载下齿轮故障诊断的模型FWaveNet.通过消融实验证明了所设计模块的必要性,同时对比实验证明了本文方法相较于其他泛化方法乃至迁移学习方法具有更好的诊断效果。

3) 使用 t -SNE 可视化方法,说明了本文方法能够将不同工况下相同健康状态的样本映射到

特征空间相同位置,使之具备较好的诊断能力。

参考文献:

- [1] 张佳雄,魏静,张春鹏,等.高阶调谐齿轮参数设计及动态响应研究[J].振动工程学报,2020,35(2):369-378.
(Zhang Jia-xiong, Wei Jing, Zhang Chun-peng, et al. Parameter design and dynamic response study of a high-order tuning gear [J]. *Journal of Vibration Engineering*, 2020,35(2):369-378.)
- [2] Li X Y, Kong X W, Liu Z D, et al. A novel framework for early pitting fault diagnosis of rotating machinery based on dilated CNN combined with spatial dropout [J]. *IEEE Access*, 2021, 9: 29243-29252.
- [3] Emmanuel S, Yihun Y, Nili A Z, et al. Planetary gear train microcrack detection using vibration data and convolutional neural networks [J]. *Neural Computing and Applications*, 2021, 33(24): 17223-17243.
- [4] 张智禹,尹爱军,谭建.融合注意力机制的改进DBN变工况齿轮箱故障诊断方法[J].振动与冲击,2021,40(14):47-52.
(Zhang Zhi-yu, Yin Ai-jun, Tan Jian. Improved DBN method with attention mechanism for the fault diagnosis of gearboxes under varying working condition [J]. *Journal of Vibration and Shock*, 2021, 40(14): 47-52.)
- [5] Zhang R, Tao H Y, Wu L F, et al. Transfer learning with neural networks for bearing fault diagnosis in changing working conditions [J]. *IEEE Access*, 2017, 5: 14347-14357.
- [6] Stacke K, Eilertsen G, Unger J, et al. Measuring domain shift for deep learning in histopathology [J]. *IEEE Journal of Biomedical and Health Informatics*, 2021, 25(2): 325-336.
- [7] Li J L, Li X Y, He D, et al. A domain adaptation model for early gear pitting fault diagnosis based on deep transfer learning network [J]. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2020, 234(1): 168-182.
- [8] Zhang Z W, Chen H H, Li S M, et al. Sparse filtering based domain adaptation for mechanical fault diagnosis [J]. *Neurocomputing*, 2020, 393: 101-111.
- [9] 赵桐,雷保珍,王训伟,等.基于迁移学习和特征重用的MIM齿轮缺陷检测[J].传感器与微系统,2021,40(10):129-131,135.
(Zhao Tong, Lei Bao-zhen, Wang Xun-wei, et al. MIM gear defect detection based on transfer learning and feature reuse [J]. *Transducer and Microsystem Technologies*, 2021, 40(10): 129-131, 135.)
- [10] Meng Z, Shi G X, Wang F L. Vibration response and fault characteristics analysis of gear based on time-varying mesh stiffness [J]. *Mechanism and Machine Theory*, 2020, 148: 103786.
- [11] Zghal B, Graja O, Dziedzic K, et al. A new modeling of planetary gear set to predict modulation phenomenon [J]. *Mechanical Systems and Signal Processing*, 2019, 127:

- 234–261.
- [12] Luo W, Qiao B J, Shen Z X, et al. A dynamic model to predict modulation sidebands of planetary gear sets with localized planet faults [C]//2018 Prognostics and System Health Management Conference (PHM–Chongqing). Chongqing, 2018: 1269–1273.
- [13] Luo H G, Hatch C, Hanna J, et al. Amplitude modulations in planetary gears [J]. *Wind Energy*, 2014, 17(4): 505–517.
- [14] Feng Z P, Zuo M J. Vibration signal models for fault diagnosis of planetary gearboxes [J]. *Journal of Sound and Vibration*, 2012, 331(22): 4919–4939.
- [15] Wang C D, Sun H C, Cao X. Construction of the efficient attention prototypical net based on the time-frequency characterization of vibration signals under noisy small sample [J]. *Measurement*, 2021, 179: 109412.
- [16] Zhang D, Chen Y Y, Guo F H, et al. A new interpretable learning method for fault diagnosis of rolling bearings [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 70: 3507010.
- [17] Zhang Q S, Wu Y N, Zhu S C. Interpretable convolutional neural networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 8827–8836.
- [18] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84–90.
- [19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016: 770–778.
- [20] Song X D, Cong Y Y, Song Y F, et al. A bearing fault diagnosis model based on CNN with wide convolution kernels [J]. *Journal of Ambient Intelligence and Humanized Computing*, 2022, 13(8): 4041–4056.
- [21] Zhang W, Li C H, Peng G L, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load [J]. *Mechanical Systems and Signal Processing*, 2018, 100: 439–453.
- [22] Hao S J, Ge F X, Li Y M, et al. Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks [J]. *Measurement*, 2020, 159: 107802.
- [23] Wang K, Zhao W, Xu A D, et al. One-dimensional multi-scale domain adaptive network for bearing-fault diagnosis under varying working conditions [J]. *Sensors*, 2020, 20(21): 6039.
- [24] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks [J]. *Journal of Machine Learning Research*, 2017, 17(1): 2096–2030.