

DOI: 10.19650/j.cnki.cjsi.J2311247

# 齿轮箱故障边缘智能诊断方法及应用研究\*

吴启航<sup>1</sup>, 丁晓喜<sup>1,2</sup>, 何清波<sup>3</sup>, 黄文彬<sup>1,2</sup>

(1. 重庆大学机械与运载工程学院 重庆 400044; 2. 重庆大学高端装备机械传动全国重点实验室 重庆 400044;  
3. 上海交通大学机械系统与振动全国重点实验室 上海 200240)

**摘要:**针对齿轮箱运行状态监测数据量大而数据价值密度低导致的数据传输和存储困难、受到带宽影响导致的故障辨识实时性差以及大而深的深度学习模型难以有效部署至边缘端硬件等问题,本文提出了一种基于乘法-卷积网络(MCN)的齿轮箱故障边缘智能诊断方法。首先,综合考虑信号滤波在特征表征以及深度学习在特征提取的优势,设计了一种轻量化的MCN模型,同时在嵌入式微处理器搭建了一套端侧边缘智能处理原型与系统。该系统可以直接部署于齿轮箱边缘,通过云服务器训练和更新MCN模型参数并部署至边缘端,于边缘端完成数据采集、处理和故障状态辨识等功能,将大量传感器数据直接消耗在边缘端。实验结果显示MCN具有99.75%的平均识别精度,且部署MCN的齿轮箱故障边缘智能诊断系统可以在0.696 s内准确识别出故障状态。

**关键词:** 齿轮故障诊断; 边缘计算; 乘法-卷积; 深度学习; 嵌入式系统

**中图分类号:** TH165+.3 TH17 **文献标识码:** A **国家标准学科分类代码:** 510.40

## Edge intelligent fault diagnosis method in the application of gearbox

Wu Qihang<sup>1</sup>, Ding Xiaoxi<sup>1,2</sup>, He Qingbo<sup>3</sup>, Huang Wenbin<sup>1,2</sup>

(1. College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China;  
2. State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, China;  
3. State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract:** To address the problems such as difficult data transmission and storage due to the large amount of operational status monitoring low-value density data, poor real-time performance of fault identification due to bandwidth impact, and the difficulty of deploying effectively large and deep learning models to edge-side hardware, this study proposes a gearbox edge intelligent fault diagnosis method based on multiplicative-convolutional network (MCN). Firstly, motivated by the merits of feature representation in signal filtering and feature extraction in deep learning, a lightweight MCN model is formulated. Secondly, a set of end-side edge intelligent processing unit prototype is made by using the embedded microcontroller unit. The system can be deployed directly at the edge of the gearbox, where the parameters of the MCN-based edge model can be trained and updated on the cloud side and deployed to the edge. The edge-side completes data acquisition, processing, and fault status identification, which can consume a large amount of sensor data directly. The experimental results show that MCN has an average recognition accuracy of 99.75%, and the gearbox edge intelligent diagnosis system deployed with MCN can accurately identify the fault state at 0.696 s.

**Keywords:** gear fault diagnosis; edge computing; multiplication-convolution network; deep learning; embedded system

## 0 引言

齿轮箱作为机械设备中重要的传动部件被广泛应用

于风电、船舶、航空等领域<sup>[1]</sup>,但其低速重载的工作环境很容易导致齿轮等关键零件发生磨损甚至断裂<sup>[2]</sup>,这些故障往往会导致设备整体运行不畅乃至设备停机等问题,给工业生产带来了极大的经济损失<sup>[3]</sup>。当齿轮故障

发生后再对其进行维护是一件不经济且不及时的任务并且伴随着安全问题,因此,利用传感器收集一维振动信号对齿轮运行状态进行实时监测和故障诊断成为了解决此类问题的重要途径<sup>[4]</sup>。

对振动信号的时域特征如均方根值、峭度指标等以及频域特征如啮合频率等进行监测和识别是传统故障诊断主要方法<sup>[5-7]</sup>。然而,由于齿轮箱的结构复杂,多个齿轮啮合时产生的激励信号混叠,始终存在的轴承振动信号干扰,以及运行过程中复杂工作环境导致的强背景噪声都会将齿轮本身的故障信号淹没,进而导致故障诊断困难<sup>[8]</sup>。如何快速、智能化地监测齿轮运行状态及故障诊断是一个亟待解决的问题。另外,随着工程中对设备稳定运行要求的提升,被监测的齿轮箱数量也大幅增加,这导致了需要被监测的设备和用于采集振动信号的传感器数量上升,海量的传感器数据冗余在云服务器中,人工处理这些数据是一件费时且不经济的任务<sup>[9]</sup>。如何存储和处理这些数据对计算机的计算能力和数据储存能力都是一个巨大考验,而且加大了计算机的能耗。同时,由于数据的私密性,在数据传输过程中的安全问题也是值得关注的部分。

2016年,边缘计算的概念被正式提出<sup>[10]</sup>,机械设备边缘故障模式也应运而生。该模式结合深度学习技术,旨在设备边缘端部署云服务器训练的网络模型,利用边缘硬件有限的计算资源完成传感器数据处理、网络模型推理、输出故障诊断结果等功能。设备边缘故障诊断可以将健康状态时所监测的低价值密度传感器数据直接消耗在边缘端,仅输出故障特征和发生故障后采集得到的价值密度高的数据<sup>[11-13]</sup>,高效地解决了云端数据的冗余问题,降低云服务器的计算负担和能耗。同时,由于传输数据量减少且边缘设备靠近齿轮箱端,齿轮故障诊断的实时性以及数据传输的安全性也会得到进一步的提升。因此,机械设备边缘智能诊断吸引了国内外学者的广泛关注。Wang等<sup>[14]</sup>提出了工业物联网边缘的高效数据约简方法,用于永磁同步电机(permanent magnet synchronous motor, PMSM)轴承故障诊断。但该方法更多地将边缘计算作为一种数据预处理方式,而不能直接完成故障诊断。Qian等<sup>[15]</sup>提出了一种基于边缘计算的旋转机械实时故障诊断与动态控制方法,并设计了边缘计算节点(edge computing node, ECN)。该节点利用传统的BP网络(back propagation neural network)作为故障辨识模型,但BP网络的故障辨识精度低于卷积神经网络(convolutional neural network, CNN)。Park等<sup>[16]</sup>提出了一种使用LSTM递归神经网络的边缘计算轻量级实时故障检测系统。但该方法存在参数量较大、对边缘硬件计算能力要求较高等问题。另一方面,随着机器学习技术的发展,越来越多的先进的深度学习模型被提出和应用

于机械设备故障诊断领域<sup>[17-18]</sup>。王骁贤等<sup>[7]</sup>提出了一种基于多传感器信号的深度特征融合模型利用长短期记忆网络从多个传感器信号中提取和融合特征以诊断电机故障,并有望将该方法部署在物联网边缘计算节点中,但该方法在利用深度学习模型解决故障辨识问题上存在缺陷。胡葛庆等<sup>[19]</sup>提出了一种基于经验模态分解和深度卷积神经网络的行星齿轮箱故障诊断方法,将经验模态分解方法与深度卷积神经网络(deep convolutional neural network, DCNN)模型结合,利用DCNN自适应融合本征模态函数(intrinsic mode function, IMF)并提取故障特征以完成齿轮箱故障识别,Chen等<sup>[20]</sup>提出了一种自编码器深度卷积神经网络融合方法。这些方法都取得了很好的故障诊断效果,但是由于推理DCNN的成本巨大,具有高处理并行性和内存带宽的图形处理单元(graphics processing unit, GPU)一直是训练和推理DCNN的主要平台<sup>[11-13]</sup>。然而,一些深度学习模型能够达到10 000层甚至更多,具有数百万到数十亿个参数<sup>[12]</sup>,这使得它们难以部署在对硬件资源和能源要求严格、可移植性强的边缘设备中。为了缩小模型尺寸,Wang等<sup>[21]</sup>基于深度可分离卷积技术,提出了一种用于物联网环境下的轴承智能故障诊断轻量化模型(lightweight convolutional neural network, LCNN),但该方法会降低故障辨识精度且没有真正部署在边缘硬件设备中。因此,面向齿轮箱的边缘智能诊断,一方面对边缘端模型识别故障准确性提出了要求。另一方面,也对轻量化、可嵌入边缘端的深度学习算法本身提出了要求。

本文针对以上难点与需求,提出了一种基于乘法-卷积网络(multiplication-convolution network, MCN)的齿轮箱边缘智能故障诊断方法并将其部署于所设计的边缘智能处理单元中,共同组成了齿轮箱边缘智能故障诊断系统。该系统能够被直接部署于设备边缘端,实时完成数据采集、故障诊断及结果显示等功能。本研究通过实验定量描述了MCN的精度和边缘智能故障诊断特性,同时设计在线实验验证了齿轮箱边缘智能诊断系统的实时性和高效性。总结本文的主要贡献如下:

- 1) 综合考虑信号滤波在特征表征以及深度学习在特征提取的优势,提出了一种高精度、轻量化的MCN,用于齿轮箱故障识别。

- 2) 设计了一套由嵌入式硬件和算法共同组成的边缘智能诊断系统,同时制作出了边缘智能处理单元原型。该系统可在STM32F429芯片为核心的微控制器(microcontroller unit, MCU)上实现边缘端对齿轮运行状态实时监测和故障诊断,于齿轮箱边缘在线完成传感器数据采集、数据预处理、MCN网络模型推理、输出故障诊断结果以及重要故障数据上传云服务器等功能。

# 1 基于乘法-卷积网络模型的齿轮箱边缘智能诊断系统

## 1.1 轻量化的乘法-卷积网络模型

由于多分量耦合调制,被测信号的频谱存在多个调制频带(模态)。每个模态都相当于频谱与滤波核的点乘,它包含丰富的故障信息,具有确定的幅值、中心频率和带宽。同时,不同故障形式的频谱多样性和差异性对实际故障诊断具有指导意义。然而,传统的时域分析技术需要人工选择故障频带,这使处理海量数据成为非常费时的的工作。因此,如何利用频谱中的这

些判别性信息进行快速高效的智能故障诊断是一个富有挑战的任务。

为了解决这个问题,本文提出了一种 MCN 模型。如图 1 所示,它以振动信号的频谱作为输入样本,由特征分离器、特征提取器和分类器组成,总共 3 层。其中,特征分离器设计了多个可学习的滤波核,每个核作为一个乘法器来提取有区别的模态,并尽可能地学习一些判别性的故障知识,如模态的幅值、中心频率和带宽。随后,将得到的乘积堆叠成一个特征图,该特征图保留了丰富的故障信息,随后使用一个卷积层作为特征提取器,进一步将这些故障特征抽象到高维空间。最后,采用一个全连接层作为分类器进行故障识别。

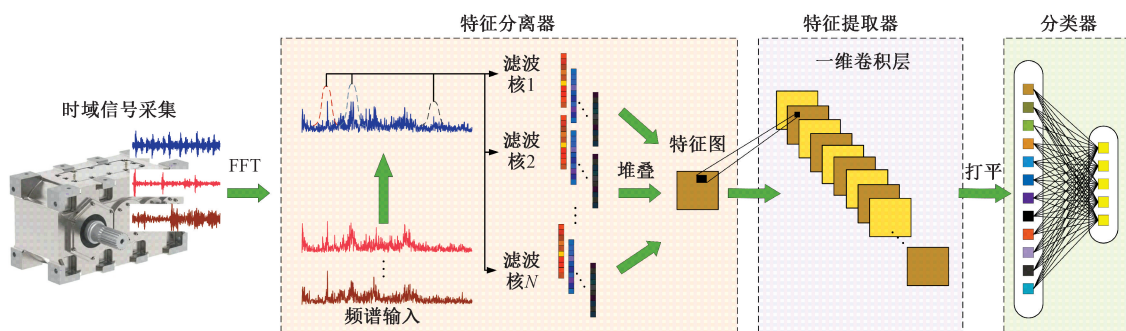


图 1 MCN 结构示意图

Fig. 1 The structure of MCN

由于模态具有特定的位置跟形状,对应丰富的故障信息,为了提取这些故障信息,本文提出了乘法滤波核,它有两个可学习参数,即中心频率和带宽惩罚系数。这两个参数主要用于沿频率轴滑动乘法滤波核,并控制滤波核的形状,通过这种方式可以很好地提取与故障相关的振幅和频率信息。模态提取的点乘运算可以表示为:

$$h_k = W(\omega_{ck}, \alpha_k) X(\omega) \quad (1)$$

其中,  $X(\omega)$  和  $W(\cdot)$  分别是谱样本和乘法滤波核。 $\omega_{ck}$  和  $\alpha_k$  分别是第  $k$  个滤波器核的中心频率和带宽惩罚系数。 $h_k$  是由第  $k$  个核得到的模态。本文中使用一系列乘法滤波核来提取存在于频谱中的多个模态并用以辨识故障。多尺度特征被由多个乘法滤波核组成的特征分离器分离。随后,这些模态被堆叠成特征图  $H$ 。

$$H = [h_1, h_2, \dots, h_N] \quad (2)$$

其中,  $N$  是滤波核的数量。本研究采用维纳滤波核(wiener filtering kernel, WFK)作为 MCN 第 1 层中的滤波核,其数学表达式如下:

$$W(\omega_c, \alpha) = \frac{1}{1 + 2\alpha(\omega - \omega_c)^2} \quad (3)$$

其中,  $\omega$  是频率,  $\alpha$  是 WFK 的带宽惩罚系数。在本研究中,频率的幅值经过了归一化处理。

为了防止提取到的模态可能发生重叠甚至相同而导

致模型收敛性和稳定性的性能下降,本文在 MCN 的特征分离器中引入了抗混叠约束,以使输出的特征映射更具区分性。具体而言,对于具有  $N$  个核的特征分隔符,第  $k$  次乘法输出可计算为:

$$h_k = \frac{X(\omega) - \sum_{i=1}^{k-1} h_i}{1 + 2\alpha_k(\omega - \omega_{ck})^2} \quad (4)$$

其中,分子项  $X(\omega) - \sum_{i=1}^{k-1} h_i$  表示第  $k$  个过滤内核的输入。在特征分离之后,1 个一维卷积层被用作特征提取器,进一步将特征映射抽象到高维特征空间。这里,采用整流线性单元(ReLU)作为激活函数。此外,还采用了自适应平均池化(AdpAvgpooling)来减少模型参数,抽象特征可以表示为:

$$Z = \text{AdpAvgpooling}(\text{ReLU}(w_{\text{conv}} \cdot H + b_{\text{conv}})) \quad (5)$$

其中,  $w_{\text{conv}}$  和  $b_{\text{conv}}$  是卷积核的权重和偏置。 $Z$  是高维特征空间中激活的特征映射。将这些特征映射展平并拼接成一个向量,使用一个全连通层作为分类器来进行故障辨识。分类器的数学表示为:

$$y = w_{fc} \cdot Z + b_{fc} \quad (6)$$

其中,  $w_{fc}$  和  $b_{fc}$  是分类器的权重和偏置。这里使用 softmax 函数将网络输出转换为故障类型的概率  $p$ ,交叉

熵损失计算如下:

$$p_i = \frac{e^{y_i}}{\sum_{j=1}^N e^{y_j}} \tag{7}$$

$$L(\theta; X) = \underset{|\omega_c, \alpha, w_{conv}, b_{conv}, w_{fc}, b_{fc}|}{\operatorname{argmin}} - \sum_{i=1}^k y_i \log(p_i) \tag{8}$$

其中,  $X$  是模型的训练谱样本,  $\theta = \{\omega_c, \alpha, w_{conv}, b_{conv}, w_{fc}, b_{fc}\}$  是模型参数,  $K$  是样本类的数量,  $y_i$  和  $y_j$  为第  $i$  和  $j$  个真实标签,  $p_i$  为第  $i$  个标签的预测概率。MCN 训练采用随机梯度下降方法更新模型参数。

在反向传播 (back propagation, BP) 过程中, 特征分离器作为 MCN 的第 1 层, 只需要更新中心频率  $\{\omega_{ci}\}$  ( $i=1, 2, \dots, N$ ) 和带宽惩罚参数  $\{\alpha_i\}$ 。根据链式推导规则, 其更新过程可以表示为:

$$\begin{cases} \omega_{ci} = \omega_{ci} - \eta \frac{\partial L}{\partial \omega_{ci}} = \\ \omega_{ci} - \eta \frac{\partial L}{\partial p_i} \frac{\partial p_i}{\partial y_i} \frac{\partial y_i}{\partial Z} \frac{\partial Z}{\partial h_i} \frac{\partial h_i}{\partial W} \frac{\partial W}{\partial \omega_{ci}} \\ \alpha_i = \alpha_i - \eta \frac{\partial L}{\partial \alpha_i} = \\ \alpha_i - \eta \frac{\partial L}{\partial p_i} \frac{\partial p_i}{\partial y_i} \frac{\partial y_i}{\partial Z} \frac{\partial Z}{\partial h_i} \frac{\partial h_i}{\partial W} \frac{\partial W}{\partial \alpha_i} \end{cases} \tag{9}$$

其中,  $\partial$  是微分算子,  $W$  是特征分离器的乘法滤波核,  $\eta$  是学习率。WFK 对参数  $\omega_{ci}$  和  $\alpha_i$  的偏导数值被给出:

$$\frac{\partial W}{\partial \omega_{ci}} = \frac{4\alpha(\omega - \omega_{ci})}{[1 + 2\alpha_i(\omega - \omega_{ci})^2]^2} \tag{10}$$

$$\frac{\partial W}{\partial \alpha_i} = \frac{2(\omega - \omega_{ci})^2}{[1 + 2\alpha_i(\omega - \omega_{ci})^2]^2} \tag{11}$$

表 1 给出了 MCN 的超参数说明。

表 1 MCN 的超参数说明

Table 1 Illustration of hyperparameters for the MCN

| 参数     | 符号                | 值                            |
|--------|-------------------|------------------------------|
| WFK 个数 | $N$               | 32                           |
| 中心频率   | $\{\omega_{ci}\}$ | 在 $[0, 1\ 280]$ Hz 下均匀初始化    |
| 带宽惩罚系数 | $\{\alpha_i\}$    | 在 $[25\ 000, 15\ 000]$ 下均匀分布 |
| 学习率    | $\eta$            | $1 \times 10^{-3}$           |

1.2 基于 MCN 的齿轮箱边缘智能诊断系统

与当前大量被采用的服务器离线推理深度智能网络算法进行齿轮箱故障诊断不同, 本文提出了一种基于 MCN 的齿轮边缘智能故障诊断方法, 并根据该方法设计了一套齿轮箱边缘端在线采集数据并实时故障诊断的边缘智能诊断系统, 该系统的方法流程如图 2 所示。系统

通过加速度计采集设备的振动加速度, 采集得到的加速度数据被传输至 MCU 中, 在 MCU 内完成数据清洗、数据处理以及 MCN 模型推理等任务, 根据 MCN 推理结果判断齿轮箱是否为正常运行状态。若状态正常, 则返回数据采集步骤, 若状态异常, 则向上位机可视化输出故障诊断结果以及向服务器上发重要的异常状态数据。当服务器接收到异常状态数据时, 服务器将完成对异常数据的存储功能, 同时当异常数据超出 MCN 训练数据范围导致 MCN 模型难以适用最新环境时, 服务器会重新训练新的 MCN。MCN 更新的部分为各层神经元之间链接的权重和偏置。训练结束后, 服务器将通过 ST-Link 将更新后的权重和偏置更新至齿轮箱边缘智能诊断系统中, 以利于系统适应新的工况。由于可以将大量价值密度低的数据消耗在边缘端, 因此所提系统大大降低了数据传输量和计算量, 且能实时、准确地输出故障类型。

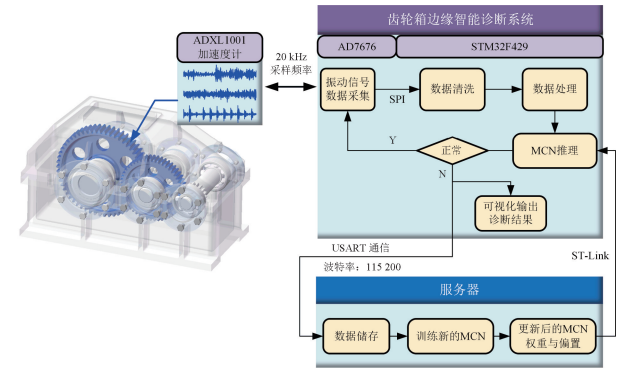


图 2 基于 MCN 的齿轮箱边缘智能诊断系统方法流程  
Fig. 2 The flow chart of gearbox edge intelligent diagnosis system based on MCN

齿轮箱边缘智能诊断系统主要由硬件和软件两部分组成, 系统的硬件部分其主要由 STM32F429 型微控制器作为数据处理模块, 由 AD7606 型模数转换芯片以及加速度计等共同组成数据采集模块, 同时, 由以上模块共同组成端侧边缘智能处理单元, 其实物如图 3 所示。数据处理模块的核心部件为 STM32F429 芯片, 采用  $24 \times 24 \times 1.4\text{ mm}$  的小型封装, 可移植性强, 便于工业现场部署。MCU 采用 Cortex M4 内核, 最高主频可达到 180 MHz, 具备浮点运算功能。其数据存储能力较强, 静态随机存取存储器 (static random-access memory, SRAM) 最大可用 256 KB, 快闪存储器 (FLASH) 最大可用 2 MB。值得注意的是从 Flash 执行时, STM32F429 能够提供 225 DMIPS/608 CoreMark 性能, 并且利用意法半导体的 ART 加速器实现了 FLASH 零等待状态, 这使极大降低了数据处理响应时长, 以上内存能够满足 MCN 的顺利部署。另外, 该芯片具有强大的数据通信能力, 包括 4 个通用同步/异步串行接收/发送器 (universal synchronous/asynchronous

receiver/transmitter, USART), 其具备独立的高精度波特率发生器, 不占用芯片定时器。所提齿轮箱边缘智能诊断系统与服务器使用 USART 通信, 波特率设置为 115 200 B。MCU 还包含 6 个速度达 45 Mbit/s 的 SPI, 可以使用 SPI 与 AD7606 通信。MCU 片上具备 3 个速度可达 2.4 MSPS 或 7.2 MSPS (交错模式下) 的 12 bit ADC, 然而为了实现更高精度的数据采集, 所提边缘智能诊断系统采用了外接 AD7606 芯片实现 16 bit 精度的数据采集。AD7606 型 A/D 模数转换芯片将模拟振动信号转化为数字信号, AD7606 芯片具有 8 个采集通道, 16 bit ADC, 全部通道可同时以最高 200 kHz 频率采集, 芯片采用 5 V 供电, 电压采集范围可选择  $\pm 5$  V 或者  $\pm 10$  V, 本文采用的采集范围为  $\pm 5$  V。MEMS 加速度计内部使用的是 ADXL1001 芯片, 该芯片具有  $30 \mu\text{g}/\sqrt{\text{Hz}}$  的典型噪声密度, 在本文中使用 3.3 V 供电, 可采集  $\pm 100$  g 以内的加速度值, 芯片外观采用  $5 \text{ mm} \times 5 \text{ mm} \times 1.80 \text{ mm}$  的 LFCSP 封装, 额定工作温度范围为  $-40^\circ\text{C} \sim +125^\circ\text{C}$ 。系统整体采用 12 V 直流供电。最终的故障诊断结果使用 USART 向上位机 (up computer) 可视化输出。特别地, 本文所提出的 MCN 离线训练是在 PyCharm 软件上实现, 用于训练 MCN 的服务器 CPU 为 Intel Core i9-9900K, GPU 为 NVIDIA GeForce RTX 2080 Ti, 内存为 64 GB。



图3 端侧边缘智能处理单元

Fig. 3 Edge intelligent processing unit

边缘智能诊断系统的软件部分主要采用 STM32CubeMX + Keil5 的编程方式, 以端侧边缘智能处理单元为载体, 实现系统初始化、传感器数据采集、数据预处理、MCN 边缘模型推理、结果输出等功能。程序采用 C 语言编程, 其流程图如图 4 所示。软件所用 STM32 外设 GPIO、定时器 TIM13 和 TIM14、SPI 以及 USART。为了降低功耗, 在本研究中 STM32 所用系统时钟频率设置为 84 MHz, 经过先进高性能总线 (advanced high-performance bus, AHB) 上分频器进行 2 分频后, HCLK 时钟频率为 42 MHz。由于 APB1 分频器设置为不分频, 因

此挂载在 APB1 上的定时器 TIM13 和 TIM14 时钟频率均为 42 MHz, 时钟周期为  $2.38 \times 10^{-8} \text{ s}$ 。其中 TIM13 作为计时器, 获取每个流程所需时间。TIM14 作为脉冲宽度调制 (pulse width modulation, PWM) 信号发生器, 产生 98% 占空比的 PWM 信号驱动 AD7606 以 20 kHz 采样频率进行加速度数据采集。 $t_0$ 、 $t_1$ 、 $t_2$ 、 $t_3$ 、 $t_4$  和  $t_5$  分别表示了初始化、数据采集、数据处理、MCN 模型推理、结果判断与输出以及休眠等程序步骤的延时。根据实验结果表明, 其中数据采集延时  $t_1$  为 0.32 s 左右, 数据处理延时  $t_2$  为 0.16 s 左右, MCN 模型推理延时  $t_3$  为 0.21 s 左右。对算法中部分程序的说明如下。

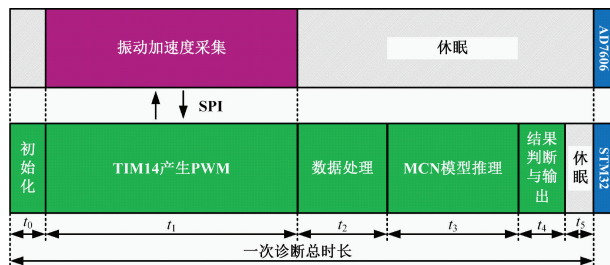


图4 端侧边缘智能处理单元软件算法流程

Fig. 4 Software algorithm flow chart of edge intelligent processing unit

### 1) 数据预处理

数据预处理包括两部分内容, 数据清洗和数据处理。数据清洗的作用是去除传感器数据中因为 AD7606 芯片硬件性能而导致的数据波动部分 (一般集中在开始采集和结束采集时), 以及去除齿轮箱停机或空转时采集的数据 (以均方根值作为判断是否空转的依据), 防止因为数据异常而导致的故障诊断错误。清洗后的数据通过快速傅里叶变换 (fast Fourier transform, FFT) 得到频谱数据, 处理后的频谱数据才能作为 MCN 的输入样本。数据预处理部分的伪代码由算法 1 给出。

### 算法1 数据预处理

**Input:** MEMS 传感器时域数据  $X[1][10\ 040]$ ; 均方根值阈值 RMS; 标志位 flag  
**Output:** 清洗后的传感器数据  $x[1][10\ 240]$ ; 频谱数据  $f[1][1\ 024]$   
 1: 声明  $\text{RMS} = 0.1$ ,  $\text{flag} = 0$ , 掐头去尾后的时域数据  $S[1][10\ 240]$ ;  
 2: **while**  $\text{flag} = 0$  **do**  
 3:   **for** ( $\text{int } i$ ;  $i < 10\ 240$ ;  $i++$ ) **do**  
 4:      $S[1][i] \leftarrow X[1][i+400]$ ; // 去除前后各 400 个点  
 5:   **end for**  
 6:    $X_{\text{rms}} \leftarrow \sqrt{\frac{1}{10\ 240} \sum S^2}$ ; // 计算均方根值  $X_{\text{rms}}$

```

7:  if  $X_{rms} > RMS$  then
8:       $x \leftarrow S$ ;
9:      flag  $\leftarrow 1$ ;
10: else
11: flag  $\leftarrow 0$ ;
12: end if
13: end while
14: if flag = 1 then
15:  $f \leftarrow FFT(x)$ ; // 得到频谱
16: return  $x, f$ 

```

## 2) MCN 模型推理

在 STM32 上实现 MCN 推理是本文所提出的齿轮箱故障边缘智能诊断应用的重要内容之一, 其中 MCN 模型由服务器预先离线训练完成并部署至边缘端在线完成边缘模型推理。部署至端侧边缘智能处理单元的 MCN 结构如图 5 所示, MCN 边缘网络模型推理部分的伪代码由算法 2 给出。

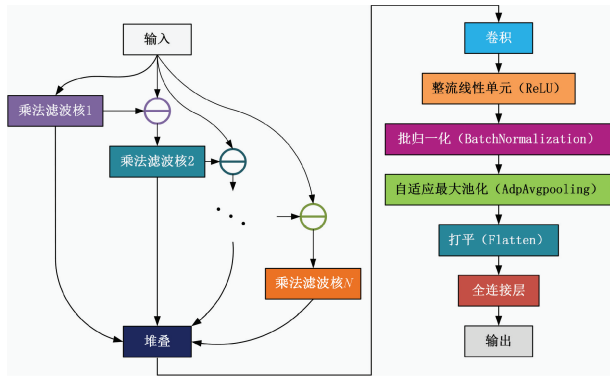


图 5 MCN 边缘网络模型结构

Fig. 5 Edge network model structure of MCN

## 算法 2 MCN 边缘网络模型推理

**Input:** 频谱  $f[1][1024]$ , 学习后的维纳滤波核  $W = [n][1024]$

**Output:** 故障诊断结果  $R$

```

1:  $N \leftarrow n$ ; // 获得维纳滤波核的个数
2: for (int  $i$ ;  $i < N$ ;  $i++$ ) do
3:    $h_i \leftarrow f \cdot M[i]$ ; // 获得乘法后的特征图
4:    $f \leftarrow f - h_i$ ; // 抗混叠约束
5: end for
6:  $H \leftarrow \text{Concat} \{h_1, h_2, \dots, h_i\}$ ; // 堆叠成特征图
7:  $C \leftarrow \text{Conv1D}(H)$ ; // 卷积
8:  $C \leftarrow \text{ReLU}(C)$ ;
9:  $C \leftarrow \text{AdpMaxpool}(C)$ ;
10:  $C \leftarrow \text{Flatten}(C)$ ; // 多维的数组打平成一维
11:  $R \leftarrow \text{Softmax}(C)$ ;
12: return  $R$ ;

```

## 2 实验验证

为了验证所提出的齿轮箱边缘智能诊断系统的准确性、实时性和高效性, 本文设计了以下两个实验。实验 1: MCN 模型识别精度验证。对比 MCN、传统深度学习模型、先进的深度学习模型以及轻量化的深度学习模型之间的性能差异, 验证 MCN 模型在故障诊断方面的优势; 实验 2: 基于 MCN 的边缘智能诊断系统精度验证。采用试验数据, 离线验证了边缘智能诊断系统的精度。同时搭建试验台实时采集数据, 在线校验了边缘智能诊断系统的实时故障诊断能力。

### 2.1 实验 1: MCN 模型识别云端精度验证

用于 MCN 模型训练的数据集采用重庆大学齿轮故障数据集, 该数据集来自于自制两级齿轮箱, 齿轮箱转速可调节, 调节范围在 200~1 100 rpm 之间。为了验证边缘智能诊断系统在各个工况条件下的识别精度, 本实验采用多种工况进行验证, 工况分别为 200、250、300、500、800、1 100 rpm。齿轮箱中间轴的齿轮可以手动更换为 5 种健康状态的齿轮, 5 种健康状态分别为: 健康、齿根裂纹、齿面剥落、齿面点蚀和断齿。振动信号采集使用单通道采集模式, 振动加速度传感器布置在中间轴端盖的 12 点方向, 采样频率为 20 480 Hz, 每一种健康状态的齿轮运行时采集 150 s 振动信号, 将得到的时域信号无重叠地平均分成 300 段, 即得到 300 个时域样本, 每个样本包含 10 240 个采样点。对每个样本通过 FFT 得到其频谱, 取频谱的前 1 024 个点作为频域样本。因此, 在每种工况下都得到了  $5 \times 300$  个样本。其中对每类齿轮运行状态的 300 个样本分别打上标签, 并采用 100 个样本用于训练模型, 100 个样本用于验证模型, 100 个样本用于测试模型的方式对 MCN 及其 8 种对比模型进行训练, 得到服务器端状态辨识精度, 以验证 MCN 在诊断齿轮运行状态方面的性能。对比模型主要可分为 3 类, 首先是传统的深度学习模型如 CNN、ResNet18 和 MLP<sup>[22]</sup>。其中 CNN 为最常用的卷积神经网络模型, ResNet18 为大体量的深度残差网络模型, MLP 是以 BP 神经网络思想为核心的多层感知网络模型。其次是较先进的深度学习模型如 VGG16, 它包含 13 个卷积层和 3 个全连接层, 依靠庞大的体积获得高识别精度<sup>[22]</sup>。最后是轻量化的网络模型如 LeNet<sup>[23]</sup>、MoblieNet<sup>[24]</sup>、LCNN<sup>[21]</sup>。其中 LeNet 和 MoblieNet 是著名的轻量化深度模型, LCNN 为利用深度可分离卷积方法改造过的轻量化 CNN 模型专用于设备故障诊断。使用测试样本进行测试后得到分类准确度, 每种网络在每种工况下循环 20 次, 求得其平均准确度和标准差, 模型分类准确度对比如图 6 所示。

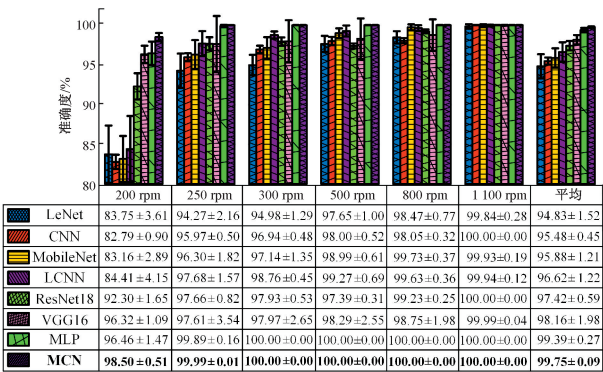


图 6 不同方法分类精度结果对比

Fig.6 Classification accuracy comparison for different methods

可以发现,MCN 在 8 种网络模型中有最高的平均分类准确度,高达 99.75%。并且在低转速时 MCN 的状态识别精度显著高于其他模型。在 200 rpm 工况下,MCN 的分类精度比传统的深度学习模型 CNN、ResNet18 和 MLP 分别高出 15.71%、6.2% 和 1.57%,比先进的深度学习模型 VGG16 高出 2.18%,比轻量化的深度学习模型 LeNet、MoblieNet 和 LCNN 分别高出 14.75%、15.34% 和 14.09%。在高转速条件下,MCN 的识别精度与参数量大的模型 VGG16 和 MLP 持平,但仍然高于传统的 CNN 以及轻量化的 LCNN 和 LeNet。另一方面,轻量化的 LCNN 和 LeNet 在故障状态辨识的平均精度方面显著低于大体量的模型 VGG16 和 MLP,这证明了模型的深度能够影响状态辨识精度。因此可以认为,MCN 的特征分离器能够有效分离出指导故障辨识的故障特征,这使得 MCN 在轻量化的条件下具有最高的故障辨识精度。

2.2 实验 2:边缘端智能诊断系统精度验证

将训练好的 MCN 模型部署至端侧边缘智能处理单元,分两部分实验对基于 MCN 的齿轮箱边缘智能诊断系统的精度、有效性和实时性进行验证,实验过程如下。

1) 齿轮箱边缘智能模型离线训练及精度验证。从每类工况条件下的测试集中随机抽取 30 组数据以验证齿轮箱边缘智能诊断系统的离线识别精度,记每类工况中识别正确的个数为 Nr,识别错误的个数为 Ne,记验证精度(Acc)计算式为:

$$Acc = \frac{Nr}{Ne + Nr} \times 100\%$$
 (12)

其验证结果如表 2 所示。可以发现,除了在 200 rpm 工况下验证精度为 98.6%,其余工况下验证精度均为 100%,该结果与云端 MCN 精度验证结果较相符合。通过本实验,可以认为齿轮箱边缘智能模型在各种工况下均具有较高的识别精度。

然而从表 2 中可以看出,在 200 rpm 时,有 2 次故障状态发生了误判。检查结果发现,齿面剥落和断齿有

表 2 离线识别精度验证结果  
Table 2 Offline accuracy verification results

| 工况        | 验证样本<br>总数 | 正确个数<br>(Nr) | 错误个数<br>(Ne) | 验证精度<br>(Acc)/% |
|-----------|------------|--------------|--------------|-----------------|
| 200 rpm   | 150        | 148          | 2            | 98.6            |
| 300 rpm   | 150        | 150          | 0            | 100             |
| 500 rpm   | 150        | 150          | 0            | 100             |
| 700 rpm   | 150        | 150          | 0            | 100             |
| 900 rpm   | 150        | 150          | 0            | 100             |
| 1 100 rpm | 150        | 150          | 0            | 100             |

小概率会被识别为齿面点蚀。Tang 等<sup>[25]</sup>的研究说明了数据集的分布一致性是决定网络模型故障识别精度的重要原因之一。因此,为了分析在 200 rpm 下精度较低的原因,本研究计算了齿面剥落和断齿所用数据在 200 和 1 100 rpm 时的频谱幅值分布情况,结果如图 7 所示。可以发现,在低转速如 200 rpm 时,这两种故障类型的频谱幅值分布对比高转速如 1 100 rpm 一致性较差,这种分布一致性差异加重了 MCN 的识别难度,进而导致了低转速时 MCN 的识别精度略低。另外,分析数据在低转速时分布一致性差的原因,在 200 rpm 时,放置故障齿轮的中间轴转频仅为 1.92 Hz,数据的周期性差致使在相同采样和数据长度下,低速时部分采集数据的频谱存在失真。

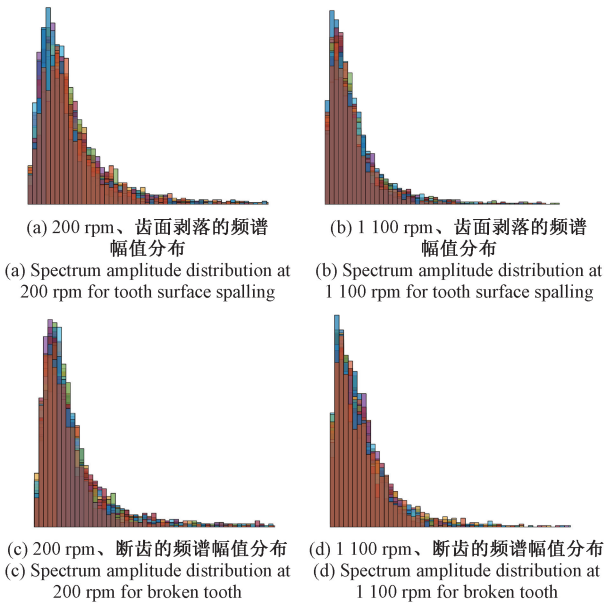


图 7 两种转速下的频谱幅值分布对比

Fig.7 Comparison of spectrum amplitude distribution at two speeds

2) 齿轮箱边缘智能诊断系统在线诊断验证。基于前述实验,为了验证边缘智能诊断系统在实时数据处理与

在线故障诊断方面的性能,本文设置了齿轮箱边缘智能诊断系统在线诊断实验。实验设备由自制故障齿轮实验台、端侧边缘智能处理单元和上位机共同构成,其中自制故障齿轮实验台主要由电机和安装有齿根裂纹齿轮的故障齿轮箱组成。实验设置电机转速 300 rpm, MEMS 加速度计安装在中间轴靠近故障齿轮的端盖上方 12 点钟方向,实验设备如图 8 所示。

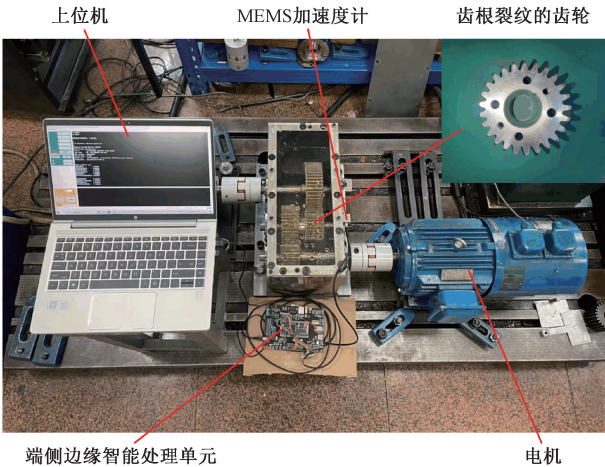


图 8 实时故障识别实验

Fig. 8 Real-time fault identification experiment

本实验设置两组对照实验以验证边缘智能诊断系统的时效性:向上位机发送采集数据和不向上位机发送采集数据。实验结果如表 3 所示。可以发现,两组实验均输出最有可能的故障为齿根裂纹,且两组实验中齿轮箱边缘智能诊断系统识别结果显示齿根裂纹故障的概率均接近 100%。

表 3 两组实验故障辨识精度实验

Table 3 Comparison of fault classification accuracy between two groups

| 故障状态 | 发送             | 不发送            |
|------|----------------|----------------|
|      | 采集数据下的<br>辨识概率 | 采集数据下的<br>辨识概率 |
| 健康   | 0.000 050      | 0.000 027      |
| 齿根裂纹 | 0.999 947      | 0.999 972      |
| 齿面剥落 | 0.000 003      | 0.000 001      |
| 齿面点蚀 | 0.000 000      | 0.000 000      |
| 断齿   | 0.000 000      | 0.000 000      |

此外,可以发现,向上位机发送采集数据实验组所用总时长为 2.766 s 左右,其中数据预处理与发送时间约为 2.231 s, 占总时长的 80.6%。不向上位机发送采集数据实验组所用总时长仅为 0.696 s 左右。两组实验实时性对比结果如表 4 所示。

表 4 两组实验实时性对比

Table 4 Comparison of real-time performance between two groups

| 时间         | 发送<br>采集数据组 | 不发送<br>采集数据组 | 时间占比        |
|------------|-------------|--------------|-------------|
| 数据采集       | 0.324 7 s   | 0.324 7 s    | 11.7%/46.6% |
| 数据预处理(与发送) | 2.230 7 s   | 0.160 9 s    | 80.6%/23.1% |
| 模型推理       | 0.207 6 s   | 0.207 6 s    | 7.5%/29.8%  |
| 总用时        | 2.766 2 s   | 0.696 4 s    | —           |

本部分实验证明了齿轮箱边缘智能诊断系统可以高效地实现边缘端故障诊断,耗时短,精度高。对比上发采集数据和不上发采集数据两组实验,不上发采集数据组总用时 0.696 4 s, 仅占发送采集数据组总用时的 25.2%, 当传感器数量增加时,该占比值将进一步缩小,由此判断齿轮箱边缘智能诊断系统具有良好的实时性。另外,当选择不上发价值密度低的数据时,齿轮箱边缘智能诊断系统可以将数据大量消耗在边缘端。

3 齿轮箱智能诊断模型边缘特性分析

受到边缘硬件资源的限制(例如 STM32F429 型微控制器可供嵌入神经网络算法的 FLASH 为 1 024 kB、RAM 为 192 kB),部署于边缘端的模型须兼具高性能和轻量化的特性。因此,本文通过对比前述 8 种网络模型的轻量化指标,以验证 MCN 的轻量化。常用的轻量化指标有模型层数、模型参数量以及每秒浮点运算次数(floating-point operations per second, FLOPS)<sup>[11-13,21]</sup>。其中层数和参数量大小可以直观地反映出模型的尺寸,FLOPS 可以反映模型的计算速度。由于要部署在边缘硬件中,因此在这些常用的指标之外,本文还选取了反映边缘模型大小的 FLASH 和 RAM 作为重要的模型轻量化性能指标。对比结果如表 5 所示。

表 5 模型轻量化对比

Table 5 Lightweight comparison of models

| 模型        | 层数 | 参数量        | FLOPS<br>/MB | FLASH<br>/kB | RAM<br>/kB |
|-----------|----|------------|--------------|--------------|------------|
| CNN       | 5  | 181 317    | 17.74        | 728.68       | 385.25     |
| ResNet18  | 18 | 3 851 269  | 175.35       | 15 001.60    | 196.02     |
| MLP       | 7  | 1 751 173  | 1.75         | 6 840.32     | 10.02      |
| LeNet     | 5  | 3 327      | 0.29         | 13.00        | 48.12      |
| MobileNet | 54 | 2 229 701  | 97.46        | 8 642.56     | 316.02     |
| LCNN      | 9  | 65 726     | 22.26        | 249.88       | 409.49     |
| VGG16     | 16 | 37 965 701 | 6 965.69     | 90 767.36    | 644.02     |
| MCN(4核)   | 3  | 5 197      | 1.31         | 19.82        | 36.04      |

对比结果可以发现,只有 4 个滤波核的 MCN 的轻量化性能显著优于 VGG16、ResNet18 等层数深、参数量大的模型,4 核 MCN 的参数量仅为 VGG16 的 0.01%。在计算速度方面,4 核 MCN 的 FLOPS 也显著小于这两个模型。另外,MobileNet 虽然比 VGG16 和 ResNet18 有更深的层数,但是其参数量和 FLOPS 却比它们显著降低,这说明 MobileNet 的轻量化性能更加优异。同理,改进后的 LCNN 比 CNN 也更加轻量。然而,4 核 MCN 比 LCNN 和 MobileNet 有着更少的参数量和更小的 FLOPS,这证明 MCN 比这些以轻量化著称的模型有着更好的轻量化性能。在全部模型中,只有 LeNet 和 4 核 MCN 的尺寸相近。观察体现模型边缘硬件轻量化性能的指标,即对 FLASH 和 RAM 的占用量,只有 LeNet 和 MCN 能被顺利部署在以 STM32F429 型芯片为和核心的端侧边缘智能处理单元中,其他模型均超出该范围。然而,在前述的实验中,MCN 的平均故障识别精度比 LeNet 高出 4.92%。

鉴于 MCN 为简单的 3 层网络结构,在卷积核参数不变的条件下,使网络模型参数量发生改变的因素主要在于乘法滤波核的个数。本文选取 200 rpm 工况下的数据集采用默认数据类型 Float32 分别对带有不同个数乘法滤波核的 MCN 进行训练,这些 MCN 所带有的乘法滤波核个数分别为 4、8、16 和 32。对比测试模型对端侧边缘智能处理单元的 FLASH 和 RAM 占用情况。其结果如表 6 所示。可以看到当乘法滤波核个数增加时,其模型训练后的精度损失不大,但参数个数却成倍数的增长,且对 STM32F429 芯片中 FLASH 和 RAM 的占用量也相应增加,甚至超过了芯片本身的储存空间,例如当有 32 个滤波核时,RAM 占用量高达 260.02 KB。因此,选择 4 个滤波核的 MCN 部署至边缘端硬件最为合适。

表 6 模型参数量对模型轻量化的影响验证结果  
Table 6 The influence of the model parameter number on the model lightweight

| 滤波核<br>个数 | 测试平均<br>准确度% | 参数个数   | FLASH<br>占用量 | RAM<br>占用量 |
|-----------|--------------|--------|--------------|------------|
| 4         | 96.38        | 5 197  | 19.82 kB     | 36.04 kB   |
| 8         | 96.81        | 6 997  | 37.13 kB     | 68.02 kB   |
| 16        | 97.01        | 10 597 | 71.75 kB     | 132.02 kB  |
| 32        | 98.50        | 17 797 | 141.00 kB    | 260.02 kB  |

另一方面,本文设计实验讨论了模型训练和推理过程中数据类型量化对模型精度的影响。将输入数据和训练好的 MCN 通过多种数据类型进行量化,发现不同数据类型能直接影响模型精度。由于输入数据所在范围为 0~0.647 5 之间,且大量数据在  $2^{-4}=0.062 5$  以下,当小数位不足 4 位时,大量输入数据被置 0,从而导致模型精

度损失较大。因此,使用数据量化的方式减少数据传输量时,至少保证小数位具有 4 位精度。

4 结 论

针对由于设备对象多、工况复杂、传感器数据量海量增长而造成的智能诊断链路长、数据传输量大、服务器端数据冗余和故障诊断不及时等问题,本文提出了一种基于 MCN 的齿轮故障边缘智能诊断方法,并将其部署于齿轮箱边缘智能诊断系统中。综合考虑信号滤波在特征表征以及深度学习在特征提取的优势,本文设计了一种轻量化的 MCN,用于齿轮箱复杂特征提取与辨识。同时,设计了一种基于 MCN 的端侧边缘智能处理原型与系统,在云端完成 MCN 模型训练并将参数下发至边缘端,边缘端快速完成振动传感器实时数据采集、数据预处理、MCN 模型推理、故障结果输出以及将重要故障数据上发至服务器端等功能。通过对比实验验证了 MCN 的故障辨识性能和轻量化性能,其平均故障辨识准确度高达 99.75%。另外,部署 MCN 的齿轮箱故障边缘诊断系统可以被直接部署在齿轮箱边缘,在线、实时、准确、快速地推理出当前齿轮运行状态。随后,本文进一步讨论了 MCN 模型的边缘特性,给出了最适合部署的 MCN 模型细节。本文所提方法为智能轴承和智能齿轮诊断领域提供了新的发展思路,展望未来,虽然本研究实现了轻量化的 MCN 至边缘端处理设备的部署,但是 MCN 受到恒定工况的局限,泛化能力较差,在实际应用中面临变速变载等复杂非平稳系统影响。因此,泛化能力更强的深度网络模型如何轻量化并部署于边缘端处理设备是未来值得研究的课题之一。

参考文献

[ 1 ] 古莹奎,曾磊,张敏,等. 基于 CEEMDAN-SQI-SVD 的齿轮箱局部故障特征提取[J]. 仪器仪表学报,2019,40(5):78-88.  
GU Y K,ZENG L,ZHANG M, et al. Feature extraction method for gearbox local fault based on CEEMDAN-SQI-SVD [ J ]. Chinese Journal of Scientific Instrument, 2019,40(5):78-88.  
[ 2 ] 赵丽娟,刘晓东,李苗. 齿轮故障诊断方法研究进展[J]. 机械强度,2016,38(5):951-956.  
ZHAO L J, LIU X D, LI M. Research progress of methods of gear fault diagnosis[J]. Mechanical Strength, 2016,38(5):951-956.  
[ 3 ] 邵毅敏,王新龙,刘静,等. 基于边缘接触时变刚度的轮齿表面剥落动力学模型与响应特征[J]. 振动与冲击,2014,33(15):8-14.

- SHAO Y M, WANG X L, LIU J, et al. Time-varying stiffness model and dynamic response characteristics of gears with tooth surface spalling and edge contact[J]. *Vibration and Impact*, 2014, 33(15): 8-14.
- [4] 李泉昌, 何清波, 邵毅敏, 等. 移不变时频流形自学习的旋转机械故障特征增强[J]. *振动工程学报*, 2020, 33(3): 622-628.
- LI Q CH, HE Q B, SHAO Y M, et al. Fault feature enhancement of rotating machinery via shift invariant time-frequency manifold self-learning[J]. *Journal of Vibration Engineering*, 2020, 33(3): 622-628.
- [5] DING X, LIN L, HE D, et al. A weight multinet architecture for bearing fault classification under complex speed conditions[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70.
- [6] 李舜酩, 郭海东, 李殿荣. 振动信号处理方法综述[J]. *仪器仪表学报*, 2013, 34(8): 1907-1915.
- LI SH M, GUO H D, LI D R. Review of vibration signal processing methods[J]. *Chinese Journal of Scientific Instrument*, 2013, 34(8): 1907-1915.
- [7] 王骁贤, 陆思良, 何清波, 等. 变转速工况下基于多传感器信号深度特征融合的电机故障诊断研究[J]. *仪器仪表学报*, 2022, 43(3): 59-67.
- WANG X X, LU S L, HE Q B, et al. Motor fault diagnosis based on deep feature fusion of multi-sensor data under variable speed condition[J]. *Chinese Journal of Scientific Instrument*, 2022, 43(3): 59-67.
- [8] 余晓霞, 汤宝平, 魏静, 等. 强背景噪声条件下自适应图卷积神经网络的航空发动机附件机匣故障诊断方法[J]. *仪器仪表学报*, 2021, 41(8): 78-86.
- YU X X, TANG B P, WEI J, et al. Fault diagnosis for aero-engine accessory gearbox by adaptive graph convolutional networks under intense background noise conditions[J]. *Chinese Journal of Scientific Instrument*, 2021, 41(8): 78-86.
- [9] 孟玲霞, 徐小力, 蒋章雷, 等. 风电机组齿轮箱早期故障预警方法研究[J]. *仪器仪表学报*, 2016, 37(12): 2758-2765.
- MENG L X, XU X L, JIANG ZH L, et al. Study on the early fault warning method for wind turbine gearbox[J]. *Chinese Journal of Scientific Instrument*, 2016, 37(12): 2758-2765.
- [10] SHI W, CAO J, ZHANG Q, et al. Edge computing: Vision and challenges[J]. *IEEE Internet of Things Journal*, 2016, 3(5): 637-646.
- [11] CAO J, ZHOU Y, ZHU X, et al. A lightweight deep learning algorithm for WiFi-based identity recognition[J]. *IEEE Internet of Things Journal*, 2021, 8(24): 17449-17459.
- [12] DENG L, LI Q, HAN S, et al. Model compression and hardware acceleration for neural networks: A comprehensive survey[J]. *Proceedings of the IEEE*, 2020, 108(4): 485-532.
- [13] MOLANES R F, AMARASINGHE K, RODRIGUEZ-ANDINA J J, et al. Deep learning and reconfigurable platforms in the Internet of Things: Challenges and opportunities in algorithms and hardware[J]. *IEEE Industrial Electronics Magazine*, 2018, 12(2): 36-49.
- [14] WANG X, LU S, HUANG W, et al. Efficient data reduction at the edge of industrial Internet of Things for PMSM bearing fault diagnosis[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70.
- [15] QIAN G, LU S, PAN D, et al. Edge computing: A promising framework for real-time fault diagnosis and dynamic control of rotating machines using multi-sensor data[J]. *IEEE Sensors Journal*, 2019, 19(11): 4211-4220.
- [16] PARK D, KIM S, AN Y, et al. LiReD: A light-weight real-time fault detection system for edge computing using LSTM recurrent neural networks[J]. *Sensors*, 2018, 18(7): 2110.
- [17] DING X, HE Q. Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis[J]. *IEEE Transactions on Instrumentation and Measurement*, 2017, 66(8): 1926-1935.
- [18] LI X, LI J, QU Y. Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning[J]. *Chinese Journal of Aeronautics*, 2020, 33(2): 418-426.
- [19] 胡芑庆, 陈徽鹏, 程哲, 等. 基于经验模态分解和深度卷积神经网络的行星齿轮箱故障诊断方法[J]. *机械工程学报*, 2019, 55(7): 9-18.
- HU N Q, CHEN H P, CHENG ZH, et al. Fault diagnosis for planetary gearbox based on EMD and deep convolutional neural networks[J]. *Journal of Mechanical Engineering*, 2019, 55(7): 9-18.
- [20] CHEN F, LIU L, TANG B, et al. A novel fusion

approach of deep convolution neural network with auto-encoder and its application in planetary gearbox fault diagnosis [J]. Proceedings of the Institution of Mechanical Engineers, 2021, 235(1): 3-16.

- [21] WANG Y, YAN J, SUN Q, et al. Bearing intelligent fault diagnosis in the industrial Internet of Things context: A lightweight convolutional neural network[J]. IEEE Access, 2020, 8: 87329-87340.
- [22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. [J]. CoRR, 2014, abs/1409.1556.
- [23] ZHAO Z, LI T, WU J, et al. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study [J]. ISA Transactions, 2020, 107.
- [24] HOWARD A, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [J]. arXiv preprint arXiv: 1704.04861, 2017.
- [25] TANG J, WU Q, LI X, et al. An efficient sequential embedding ConvNet for rotating machinery intelligent fault diagnosis[J]. IEEE Transactions on Instrumentation and Measurement, 2023, DOI: 10.1109/PTC.2019.8810946.

## 作者简介



吴启航, 2021 年于湖北工业大学获学士学位, 现为重庆大学博士研究生, 主要研究方向为智能故障诊断、边缘计算。

E-mail: wuqh@cqu.edu.cn

**Wu Qihang** received his B. Sc. degree from Hubei University of Technology in 2021. He is currently pursuing

his Ph. D. degree at Chongqing University. His main research interests include intelligent fault diagnosis and edge computing.



E-mail: dxxu@cqu.edu.cn

**Ding Xiaoxi** (Corresponding author) received his B. Sc. and Ph. D. degrees both from the University of Science and Technology of China in 2012 and 2017, respectively. He is currently a lecturer at Chongqing University, China. His main research interests include signal processing, data mining, and intelligent monitoring system for health monitoring and fault diagnosis in machines.



何清波, 分别于 2002 年和 2007 年在中国科学技术大学合肥获得机械工程学士学位和博士学位。现任上海交通大学机械系统与振动国家重点实验室教授, 主要研究方向为振动分析、信号处理和复杂机器智能监测、诊断和控制的智能结构。

E-mail: qbhe@sjtu.edu.cn

**He Qingbo** received his B. Sc. and Ph. D. degrees in mechanical engineering from the University of Science and Technology of China, Hefei, China, in 2002 and 2007, respectively. He is currently a professor with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, China. His main research interests include a combination of vibration analysis, signal processing, and smart structure for intelligent monitoring, diagnosis, and control in complex machines.