

基于多尺度多路径集成网络的轴承故障诊断方法*

BEARING FAULT DIAGNOSIS METHOD BASED ON MULTI-SCALE AND MULTI-PATH ENSEMBLE NETWORK

戚博炜** 李媛媛*** 宋丽媛

(上海工程技术大学 电子电气工程学院, 上海 201620)

QI BoWei LI YuanYuan SONG LiYuan

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

摘要 针对卷积神经网络参数量大、层数深、泛化性能弱的问题,提出了一种基于多尺度多路径集成网络(Multi-scale Multi-path Convolutional Neural Network, MCS-CNN)的轴承故障诊断方法。首先,提出多尺度卷积块,利用深度卷积和逐点卷积来降低网络参数量、增加网络宽度,有效提取多尺度特征;其次,提出集成块,利用多路径的连接方式将低级和高级特征连接起来以增加网络深度,进而提高模型诊断精度;最后,在凯斯西储大学轴承数据集上进行验证。结果表明,在高噪声和跨负载场景下的故障诊断精度分别可达 97.5% 和 98.25%,在混合场景下相较于现有的诊断方法精度提高了 15 个百分点以上,有效表明了此方法的鲁棒性和泛化性。

关键词 轴承 卷积神经网络 多尺度多路径 故障诊断 集成网络

中图分类号 TH133

Abstract To address the problems such as complex convolutional neural network parameters, deep layers and weak generalization performance, a multi-scale multi-path convolutional neural network (MCS-CNN) based bearing fault diagnosis method was proposed. Firstly, a multi-scale convolution block was proposed to reduce the number of network parameters and increase the network width by using depthwise convolution and pointwise convolution, so as to extract multi-scale features effectively. Secondly, an ensemble block was proposed to increase the network depth by connecting low-level and high-level features through multiple paths, thereby improving the diagnostic accuracy of the model. Finally, the effectiveness of the method was verified on the Case Western Reserve University bearing dataset. The results show that the fault diagnosis accuracy of the proposed method can reach 97.5% and 98.25% in high-noise and cross-load scenarios, and the accuracy in mixed scenarios is improved by more than 15% compared to existing diagnosis methods, which prove the robustness and generalisability of the proposed method.

Key words Bearing; Convolution neural network; Multi-scale and multi-path; Fault diagnosis; Ensemble network

Corresponding author: LI YuanYuan, E-mail: liyuanyuanedu@163.com

The project supported by the National Key R&D Program of China (No. 2020AAA0109300).

Manuscript received 20220920, in revised form 20221208.

0 引言

在工业生产中,轴承作为重要的工业基础件在各类旋转机械中广泛使用,承担着旋转机械工作过程中的大部分载荷^[1]。轴承一旦失效,不仅增加设备的维护费用,还会造成更大的经济损失。

目前,常用的轴承故障诊断方法可以分为两种:信

号分析法和大数据驱动智能诊断法。振动信号分析法主要是将振动信号通过傅里叶分析、小波分析、经验模态分解等方式进行时频域分析,通过划定阈值的方式实现故障诊断。郭学卫等^[2]将样本熵与分数阶傅里叶变换相结合进行故障特征提取,通过划定阈值的形式进行不同故障模式的划分。这种方式通常对机械系统物理性能的变化特别敏感,尤其是在噪声环境下,模

* 20220920 收到初稿, 20221208 收到修改稿。科技部重大专项(2020AAA0109300)资助。

** 戚博炜,男,1998年生,浙江温州人,汉族,上海工程技术大学硕士,主要研究方向为智能控制。

*** 李媛媛(通信作者),女,1979年生,上海人,汉族,上海工程技术大学教授,主要研究方向为智能控制。

型的阈值很难划分,无法得到预期的结果。随着大数据时代的来临,许多基于数据驱动的方法相继出现并且应用到故障诊断中^[3-4]。在此背景下,深度学习中的非线性机制可以直接拟合高维数据,尤其是卷积神经网络(Convolutional Neural Network, CNN)中的权值共享和稀疏连接,能更好提取故障样本中的信息。王海龙等^[5]将传统经验模态分解与CNN相结合来实现轴承的故障诊断,利用经验模态分析将原始的轴承振动信号转换成二维特征图,再将特征图导入CNN中进行故障特征提取,取得了较好的诊断结果。刘红军等^[6]使用格拉姆角场将轴承振动信号转换为二维图像,并通过改进的CNN模型实现了对轴承的故障诊断。然而,无论是振动信号分析方法还是将信号转换为图像的方法,都需要工作人员具有一定的先验知识,不利于工程实际中的应用。对此,冯连强等^[7]将原始振动信号直接输入一维卷积神经网络中进行故障诊断,试验结果表明了所提方法的有效性。刘洋等^[8]提出一种将注意力模块与一维卷积神经网络相结合的滚动轴承故障诊断方法,该方法在强噪声环境下同样具有良好的表现,且相较于传统方法,收敛速度更快,泛化性能更好。

上述方法在单一场景中对轴承的故障诊断有一定的效果。然而在工程实际中,旋转机械由一个齿轮、轴承和其他机械部件组成。即使是简单的旋转机械,其复杂性也会导致测量的振动信号中包含多个内在振荡模式,由于机器组件之间存在相互作用和耦合效应,这就意味着单一尺度的卷积不足以描述机器振动信号^[9],所以充分提取故障特征可以有效提高模型的鲁棒性和泛化性。HUANG W Y等^[10]提出多尺度级联CNN,并试图利用具有不同大小的卷积核对轴承的故障特征进行提取,试验结果表明卷积核的大小与信号频率有着一定关联。在此基础上,LIU H H等^[11]发现轴承振动信号具有多尺度时间特性。CNN在故障诊断中有着广泛的应用,但因其是传统网络,故存在一定缺点,如模型层数较多,参数量较大,不能满足工程实际应用中计算机性能的要求。因此,轻量化网络登上历史舞台,CHOLLET F^[12]提出了深度可分离卷积的概念,并将其应用在Inception中,有效降低了Inception中参数的数量,并且在图像分类任务中取得了更好的效果。HOWARD A G等^[13]将MobileNet中的传统卷积替换为深度可分离卷积来实现网络轻量化的目的。后来,朱富等^[14]采用一维深度可分离卷积神经网络来改进MobileNet,可以直接将轴承振动信号作为模型输入,实现轻量化部署,在两组轴承试验数据集上都取得了较好的诊断效果。深度可分离卷积由深度卷积和逐点卷积组成。具体而言,深度卷积是一个完全稀疏的

卷积连接,而逐点卷积是一个 1×1 的卷积^[15]。虽然计算负载和参数量明显减少,但是也会带来精度的下降,尤其是在复杂情境下模型泛化性能会受到很大程度的影响。

针对上述算法模型参数量大、识别精度下降及泛化性能弱的问题,本文提出了一种新的网络结构——多尺度多路径集成网络(Multi-scale Multi-path Convolutional Neural Network, MCS-CNN)。首先,通过深度卷积和逐点卷积重新构建出一个多尺度卷积块扩大网络宽度,可以有效提取信号中的多尺度特征。该方式不仅可以降低模型的参数量,还提升了模型的特征提取能力。其次,通过多路径的方式把低级和高级特征连接起来,可以将梯度经过一条路径传播到网络的前层,可以避免网络深度造成的梯度消失,从而增加网络深度,进一步提高模型诊断性能。凯斯西储大学轴承数据集上的试验结果表明,所提方法在噪声和负载环境下,均能取得90%以上的准确率,且在混合负载及噪声的环境下准确率仍能达到83%。

1 卷积神经网络

CNN是由多层感知机(Multi-Layer Perceptron, MLP)变种而来的,它是模拟生物视觉皮层的一种前馈神经网络,目前被广泛应用在图像和视频等二维信号的识别中。1D-CNN是CNN的一种变型,它是一种单位宽度卷积核的深度反馈神经网络,常用在序列模型、自然语言处理领域。1D-CNN主要由三部分组成,分别为:卷积层、池化层和全连接层。

1.1 卷积层

卷积层用来减少输入的维数并且从中提取有用的特征,它最主要的两个特征是局部连接和权值共享。局部连接是指卷积层节点仅连接前一层的局部节点,只学习局部特征。这样可以大幅度减少参数量,加快模型学习的速度,减少过拟合的现象。权值共享是在一层网络中使用同一个卷积核来进行卷积计算。卷积运算是一种线性运算,但是神经网络需要对非线性函数进行拟合,因此需要用激活函数将其转化为非线性。具体的卷积运算为

$$a_i = f\left(\sum_{m=0}^n w_m x_m + w_b\right) \quad (1)$$

式中, a_i 为第 i 个序列段卷积计算的结果; w_m 为序列段中第 m 个元素的权重; x_m 为序列段的第 m 个元素; w_b 为该卷积核的偏置项; $f(\cdot)$ 为激活函数。

1.2 池化层

池化层是数据特征提取的一种方式,用于提取更具代表性的特征,主要的作用是同时减少参数量和计算量,防止模型过拟合,提高模型的泛化能力。特征提

取的误差主要来自两个方面:①领域大小受限造成估值方差增大;②卷积层参数误差造成估计均值偏移^[16]。池化层一般为最大池化和平均池化,它们分别对应减少这两种误差。两种池化运算如式(2)和式(3)所示。在分类应用中,最大池化的试验效果优于平均池化,因此本文将使用最大池化。

$$P_{\max}^{l(i,j)} = \max_{(j-1)w \leq t \leq jw} \{u^{l(i,t)}\}, \quad i = 1, 2, \dots, q \quad (2)$$

$$P_{\text{average}}^{l(i,j)} = \text{Average}_{(j-1)w \leq t \leq jw} \{u^{l(i,t)}\}, \quad i = 1, 2, \dots, q \quad (3)$$

式中, $u^{l(i,t)}$ 为第 l 层中的第 i 个特征张量的第 t 个神经元; w 为卷积核宽度; j 为第 j 个池化层。

1.3 全连接层

全连接层在卷积网络的模型中作为一个分类器,可以将特征中具有类判别性的局部信息集成起来。该层降低了特征定位对分类结果的影响,提高了整个网络的鲁棒性。最后,将全连接层的输出值传递到 Softmax 激活函数中,该函数映射了 $(0,1)$ 的数值,即标签的概率值。Softmax 激活函数如式(4)所示:

$$\sigma(Z)_j = \frac{e^{Z_j}}{\sum_{k=1}^m e^{Z_k}}, \quad j = 1, 2, \dots, m \quad (4)$$

式中, Z 为第 j 个神经元的输出; m 为神经元的个数。

2 多尺度多路径集成网络

现有的故障诊断方法具有参数量大、特征提取不充分和泛化性能弱的特点,无法充分提取轴承振动信号内各部件振动耦合导致的多尺度特性。为此,本文提出了一种新的端到端的网络模型,它主要由集成块、加权变换层、池化层和注意力机制四部分组成,具体结构如图1所示。集成块集成了多层特征,加权变换层进行通道加权和特征变换。在集成块内,批处理归一化和 ReLU 激活函数被认为是一个复合的非线性映射 $H(\cdot)$, 设输入为 x , 在第 l 个卷积层、BN 层和 ReLU 激活函数之后, x 的非线性映射为 $H_l(x)$, 即

$$H_l(x) = H_l(w_\lambda * x + b) \quad (5)$$

式中, w_λ 为卷积权值; b 为偏差; $*$ 为卷积运算; λ 为卷积核的大小。

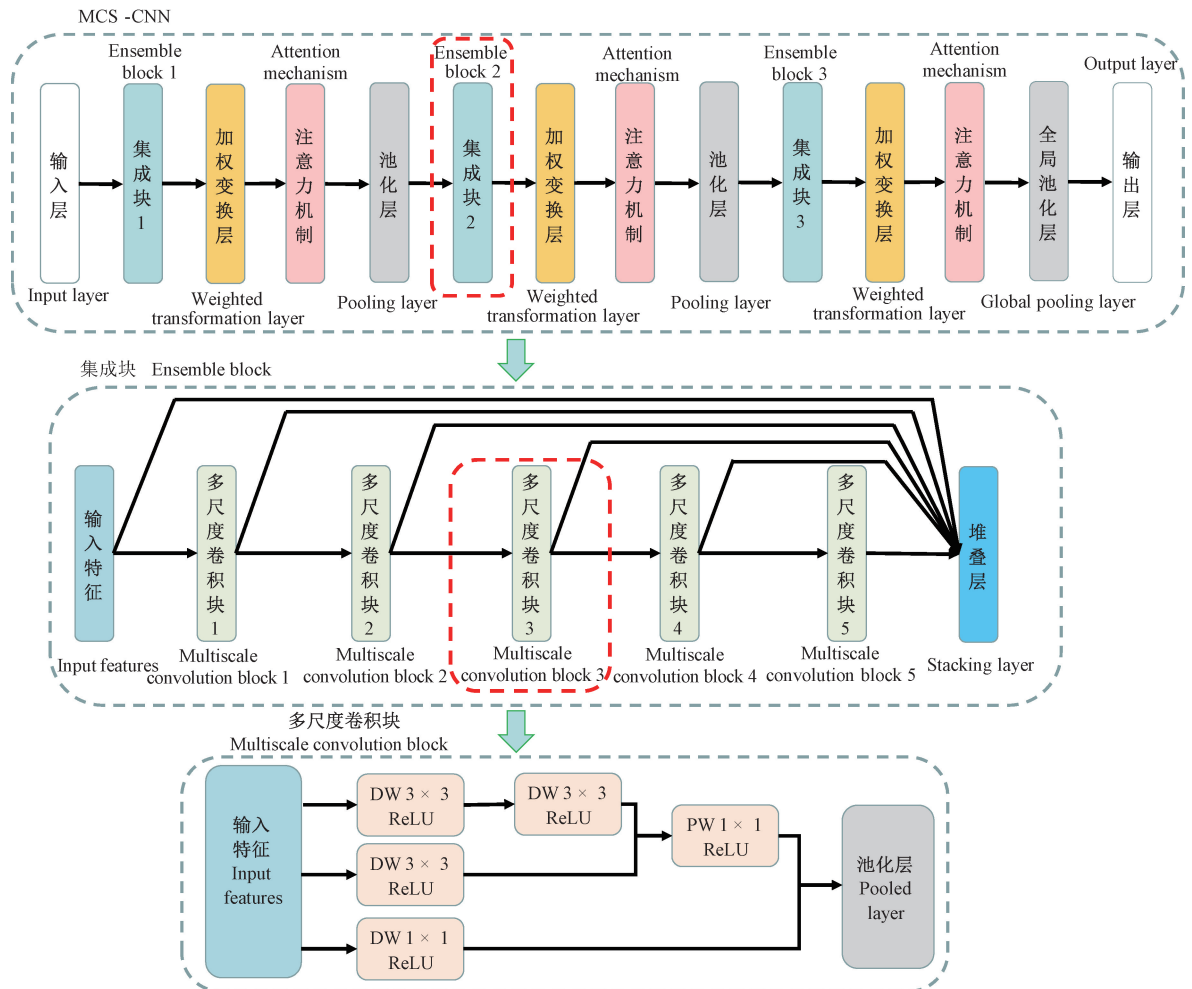


图1 MCS-CNN 结构

Fig. 1 Structure of the MCS-CNN

2.1 深度可分离卷积

深度可分离卷积由深度卷积 (Depthwise Convolution, DW) 和逐点卷积 (Pointwise Convolution, PW) 构成, 与常规卷积相比, 其运算成本和参数量会更低, 更适合工业场景的需求。

深度可分离卷积层与常规卷积的区别如图 2 所示。与常规卷积相比, 深度卷积只对应一个输入通道, 如图 2(b) 所示, 每个通道的卷积为

$$G_{h,n} = K_{i,n} F_{h+i-1,m} \quad (6)$$

式中, $K_{i,n}$ 为深度卷积的卷积核大小; $F_{h+i-1,m}$ 为输入特征; $G_{h,n}$ 为通道在 F 中产生滤波后的输出特征图的第 n 个通道。深度卷积的参数和计算代价近似表示为

$$P_{\text{conv}} = L_k C_{\text{in}} C_{\text{out}} \quad (7)$$

$$F_{\text{conv}} = 2L_k C_{\text{in}} C_{\text{out}} L_{\text{out}} \quad (8)$$

$$P_{\text{dw}} = L_k C_{\text{in}} \quad (9)$$

$$F_{\text{dw}} = 2L_k C_{\text{in}} L_{\text{out}} \quad (10)$$

式中, P_* 和 F_* 分别为卷积的参数和计算量; L_k 和 L_{out} 分别为输入输出卷积核层数; C_{in} 和 C_{out} 分别为卷积的输入和输出。与标准卷积相比, 深度卷积的参数和计算量为标准卷积的 $1/C_{\text{out}}$ 。逐点卷积使用长度为 1 的卷积来线性组合深度卷积的输出特征, 以形成新的特征。

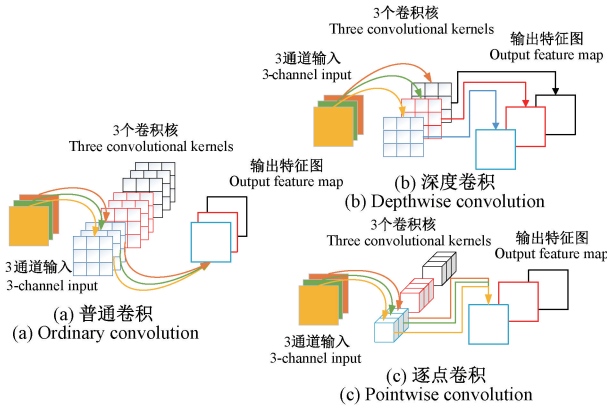


图 2 深度可分离卷积与普通卷积对比

Fig. 2 Comparison between depthwise separable convolution and ordinary convolution

如图 2(c) 所示, 对每个信道空间中的点进行加权求和, 以增加信道间的相关性。参数的数量和计算量可以写成

$$P_{\text{pw}} = C_{\text{in}} C_{\text{out}} \quad (11)$$

$$F_{\text{pw}} = 2C_{\text{in}} C_{\text{out}} L_{\text{out}} \quad (12)$$

$$\frac{L_k C_{\text{in}} + C_{\text{in}} C_{\text{out}}}{L_k C_{\text{in}} C_{\text{out}}} = \frac{1}{C_{\text{out}}} + \frac{1}{L_k} \quad (13)$$

深度可分离卷积实现了提取和组合特征的标准卷积的效果。与标准卷积相比, 深度可分离卷积的参数和计算量显著减少, 计算时间和存储空间也减少了一半。

深度可分离卷积虽然可以显著减少参数量和计算量, 但也带来了精度的下降。逐点卷积在进行通道融合时针对单一尺度, 并不能较好顾及信号中的多尺度特性。对此本文对深度卷积与逐点卷积进行重构, 如图 1 所示, 增加网络宽度, 提升模型多尺度特征提取的能力。通过不同卷积核大小的深度卷积进行多尺度特征提取后, 使用逐点卷积对原始特征进行通道融合, 后进行堆叠。最后通过加权变换层将原始特征与不同尺度特征进行融合, 可以有效增加模型的特征提取能力。

2.2 集成块

集成块内部采用通道级连接, 直接沿着通道方向连接低级和高级特征, 输出通道的数量等于所有输入通道的数量之和。通道级连接可以最大程度地保留不同层次的特征。因此, 使用通道连接来集成多级特征。为了更好地介绍本文方法, 定义 $H_l^i(x)$ 为卷积核大小为 λ 下的非线性映射。由于每个集成块中都是通过多个多尺度卷积块连接而成的, 每一个多尺度卷积块中都是由 1×1 和 3×3 的平行多尺度卷积并行连接而成的, 因此多尺度卷积块可以表示为

$$H_l(x) = H_l^1 \{ H_l^3 [H_l^3(x)] + H_l^1(x) \} + H_l^1(x) \quad (14)$$

第 n 个集成块 $f_n(x)$ 的输出是第 l 个多尺度卷积块的输出特征映射的连接, 可表示为

$$f_n = [H_{m+1}(x), H_{m+2}(x), \dots, H_{m+l}(x)] \quad (15)$$

式中, $[\cdot]$ 表示连接操作; m 表示集成块中第 $(n-1)$ 个多尺度卷积块池化层的输出特征。

对第 l 个卷积块的输出用多尺度块进行卷积, 经过 $H(\cdot)$ 线性映射得到特征映射。然后将特征映射连接起来以获得输出, 并将这些输出输入到下一层中。根据式 (14) 和式 (15) 可以得出一个基层块 f_n , 可以表示为

$$\begin{aligned} f_n = & [H_{(m+1)}^1 \{ H_{(m+1)}^3 [H_{(m+1)}^3(x)] + H_{(m+1)}^1(x) \} + \\ & H_{(m+1)}^1(x), H_{(m+2)}^1 \{ H_{(m+2)}^3 [H_{(m+2)}^3(x)] + \\ & H_{(m+2)}^1(x) \} + H_{(m+2)}^1(x), \dots, H_{(m+l)}^1 \{ \\ & H_{(m+l)}^3 [H_{(m+l)}^3(x)] + H_{(m+l)}^1(x) \} + \\ & H_{(m+l)}^1(x)] \end{aligned} \quad (16)$$

根据式 (16), 在多尺度卷积中, 将集成块的输入按渐进的方式进行卷积, 不仅可以解决梯度消失的问题, 还可以通过多尺度卷积层逐步地提取多尺度特征。这样, 就可以提高网络的学习性能。

2.3 加权变换层

在传统的卷积操作中, 一组卷积与一组输入特征映射紧密连接, 每个卷积都与所有输入特征映射进行卷积计算。由于特征映射与卷积之间这种密集的连接结构会导致较大的计算负荷, 特别是当输入特征映射或卷积核的数量较大时, 因此, MCS-CNN 通过深度可

分离卷积,进一步减少了计算负载和参数量。

在直接连接多层特征之后,如果采用池化层直接降低特征空间维数,就会忽略各层特征对最终分类任务贡献的影响与各层特征之间的关系。因此,添加了一个加权变换层来整合每个层的特征。在实际应用中,加权变换层由 1×1 的卷积核实现。卷积支持跨通道的信息交换和集成,在一个激活函数的前提下,输出所有输入通道的非线性组合。集成块输出特征通过加权变换层对特征进行变换,并通过池化层减少空间维度的一半,实现了集成块输出特征的通道级融合,降低后续的计算复杂度。

2.4 注意力机制

注意力机制是人类大脑所特有的信号处理机制,在某一场景下会将注意力聚焦在场景中的某一区域,从而获得需要被观察目标的细节信息,弱化其他不相干的内容。该机制能够有效提升信息处理的效率与准确性。

对轴承故障诊断模型而言,输入模型中的数据往往仅有部分包含故障信息,更多的是不包含故障信息的数据。因此,可采用注意力模块使模型重点关注包含故障信息的数据,忽略不包含故障信息的数据。图 3 所示为本文所使用的注意力模块(Efficient Channel Attention, ECA)。

注意力机制首先将输入的特征通过全局平均池化(Global Average Pooling, GAP)操作转化为一维特征向量,向量中每一个值都可以看作原始特征中每一通道的权重。ECA 模块通过对转换后的一维特征向量进行一维卷积来实现多通道交互,该操作相对于全连

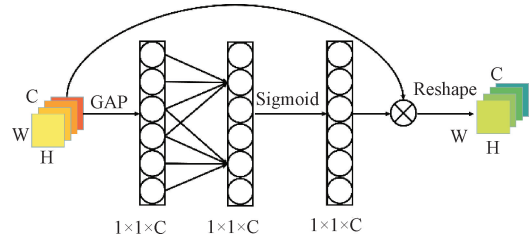


图 3 注意力机制

Fig. 3 Attention mechanism

接操作,可以有效减少训练参数。为避免降维操作所带来的信息损失,卷积操作前后的特征向量维度保持一致,然后经过 Sigmoid 激活函数获得通道注意力权重,最后与原始的特征进行相乘得到具有注意力的特征。通过引入 ECA 模块,可以将重要通道的特征强化,非重要通道的特征弱化,在一定程度上可以去除信号中噪声的干扰,提高诊断精度。

2.5 基于 MCS-CNN 的故障诊断系统

图 4 所示为基于 MCS-CNN 结构的智能故障诊断系统。该系统无须对输入信号进行任何处理,对 MCS-CNN 中的集成块和多尺度卷积块的数量进行调整,以保证模型具有最佳的故障诊断效果。

如图 4 所示,首先,通过振动信号采集装置在旋转机械设备上获取轴承振动信号,利用数据重采样技术将采集到的信号分为训练样本和测试样本,如图 5 所示。其次,将训练样本送入 MCS-CNN 模型中进行权重训练,并和采集信号时轴承的状态相匹配。最后,保存 MCS-CNN 模型训练后的权重值,将测试样本送入含有训练后权重值的 MCS-CNN 中,输出故障诊断结果。

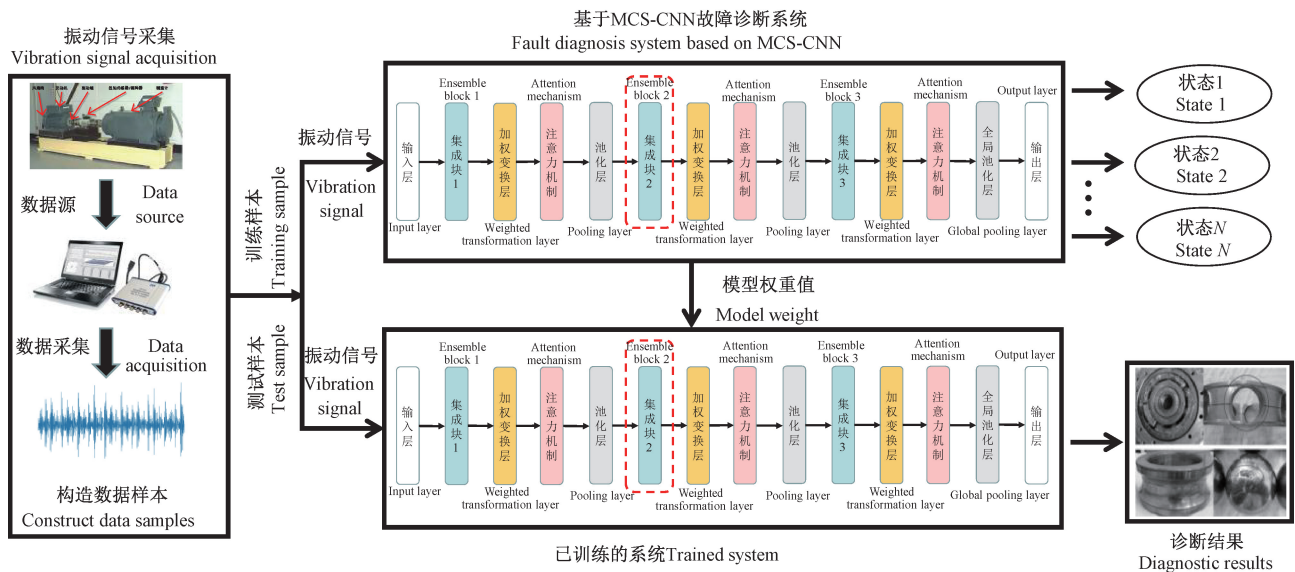


图 4 MCS-CNN 智能故障诊断系统

Fig. 4 Intelligent fault diagnosis system based on MCS-CNN

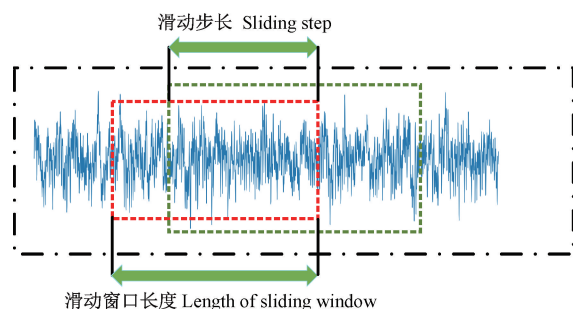


图 5 数据重采样

Fig. 5 Data resampling

3 试验验证

3.1 数据说明

本文数据来自凯斯西储大学 (Case Western Reserve University, CWRU) 轴承数据中心, 包含正常和故障轴承在四种工况下的滚动轴承测试数据。故障主要是用电火花技术在内圈、滚动体和外圈位置加工的单点损伤, 损伤直径为 $0.007 \sim 0.021$ in ($1 \text{ in} = 25.4 \text{ mm}$)。将出现故障的轴承重新安装到测试电动机中, 并记录 $0 \sim 3 \text{ hp}$ ($1 \text{ hp} = 745.699 \text{ W}$) 电动机负载下 (电动机转速为 $1720 \sim 1797 \text{ r/min}$) 的振动数据^[17]。

将上述的故障状态与健康状态共计 10 类数据作为本文试验数据, 数据采集用滑动窗口的方式进行数据增强。在负载为 $0, 1, 2, 3 \text{ hp}$ 下, 分别从每类故障状态的振动信号中分割出 200 个长度为 2 000 的子序列, 作为该类别的样本集, 如表 1 所示。

表 1 数据详细组成情况

Tab. 1 Detailed composition of data

故障类型 Fault type	故障直径 Fault diameter	采样数目 Number of samples				标签 Label
		0 hp	1 hp	2 hp	3 hp	
外圈 Outer ring	0.007	200	200	200	200	1
	0.014	200	200	200	200	2
	0.021	200	200	200	200	3
	0.007	200	200	200	200	4
内圈 Inner ring	0.014	200	200	200	200	5
	0.021	200	200	200	200	6
	0.007	200	200	200	200	7
滚动体 Rolling element	0.014	200	200	200	200	8
	0.021	200	200	200	200	9
正常 Normal	—	200	200	200	200	0

3.2 网络结构对模型性能影响

MCS-CNN 主要由多个集成块和多尺度卷积块组成, 它们的数量决定着模型的性能。因此在原始信号中加入信噪比 (Signal-Noise Ratio, SNR) 为 -4 dB 和 -6 dB 的振动信号验证 11 种不同组合的 MCS-CNN 模型性能, 结果如图 6 所示。其中 $\langle 3, 2 \rangle$ 表示 MCS-CNN 中包含 3 个集成块, 每个集成块中包含 2 个多尺度卷

积块。信号样本中加入不同信噪比的高斯白噪声, 如图 7 所示。

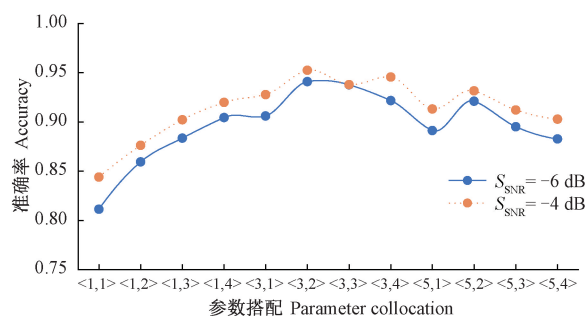


图 6 MCS-CNN 不同结构对比

Fig. 6 Comparison of different structures of MCS-CNN

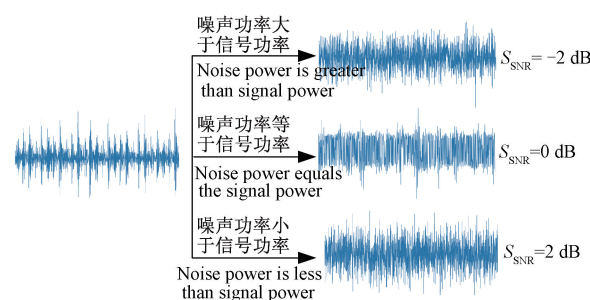


图 7 加入噪声的振动信号对比

Fig. 7 Comparison of vibration signals with noise

为了分析 MCS-CNN 中集成块数和集成块内多尺度分离卷积块数对轴承故障诊断的影响, 本文采用集成块个数为 $1, 3, 5$ 和多尺度卷积块为 $1, 2, 3, 4$ 的组合在 S_{SNR} 为 -6 dB 和 -4 dB 时故障诊断的结果。由图 7 可知, 集成块数为 3 且多尺度卷积块数为 2 时模型性能达到最佳, 识别准确率可以达到 94.12% 和 95.25% 。由图 7 可以发现, 集成块数目和多尺度卷积块数目过多会导致特征过度压缩, 从而降低网络的故障识别精度。本文接下来的试验都会以此结构开展, 模型具体结构如表 2 所示, 其中 $[1 \times 1, 3 \times 3]$ 表示多尺度卷积中卷积核大小, Avg 表示平均池化, GAP 表示全局平均池化。

表 2 MCS-CNN 的网络结构

Tab. 2 Network structure of MCS-CNN

网络层 Network layer	输入维度 Input dimension	输出维度 Output dimension	分组数 Groups	多尺度卷积块 Multiscale convolution block
Conv1_x	1×2000	48×1750	1	$[1 \times 1, 3 \times 3] * 2$
Weighted_1	48×1750	48×1750	48	1×1
Pooling 1	48×1750	48×875	—	$2 \times 2, \text{Avg}$
Conv2_x	48×875	48×764	48	$[1 \times 1, 3 \times 3] * 2$
Weighted_2	48×764	48×764	48	1×1
Pooling 2	48×764	48×382	—	$2 \times 2, \text{Avg}$
Conv3_x	48×382	48×333	48	$[1 \times 1, 3 \times 3] * 2$
Weighted_3	48×333	48×333	48	1×1
Pooling 3	48×333	48×2	—	GAP
Classification	—	—	—	Softmax

3.3 噪声测试

振动信号的采集过程中会混杂周边环境的信号,会有强噪声的干扰,影响故障诊断的识别准确率。为了验证本文算法在噪声环境中的优势,分别与支持向量机(Support Vector Machine, SVM)^[18]、CNN、ResNet、WT-CNN^[19]、MobileNet 进行对比,噪声验证的结果如表 3 所示。

表 3 不同方法性能对比

Tab. 3 Performance comparison of different methods

模型 Model	准确率 Accuracy/%				参数量 Parameters/M
	-4 dB	0 dB	4 dB	6 dB	
SVM	50.81 ±2.21	65.38 ±2.71	75.95 ±1.82	78.36 ±0.61	—
CNN	88.73 ±0.68	95.98 ±1.27	97.83 ±2.33	97.98 ±0.91	0.188
ResNet	85.86 ±0.82	94.35 ±0.17	98.93 ±0.27	99.93 ±0.12	4.309
WT-CNN	75.88 ±1.28	80.5 ±0.95	83.19 ±0.52	91.19 ±0.43	3.482
MobileNet	86.81 ±0.46	93.27 ±0.22	97.51 ±0.23	99.41 ±0.11	3.259
MCS-CNN	93.29 ±0.31	98.05 ±0.23	98.83 ±0.03	99.84 ±0.01	0.121

表 3 中,当 $S_{\text{SNR}} = -4$ dB 时, SVM 针对噪声强度大于原始信号强度时诊断效果极其差,仅为 50.81%,可见环境因素对时频域特征的影响极为明显。WT-CNN 是将轴承信号转化为时序图作为输入,无法对噪声数据进行滤波,因此环境噪声较大时诊断准确率仅为 75.88%。CNN、ResNet 和 MobileNet 都具有自适应特征提取能力,其中 ResNet 和 MobileNet 都具有一定的网络深度,但是 MCS-CNN 集成模式的设计不仅增加了网络深度,通过多尺度卷积块还提升了网络宽度,具备多尺度特征提取的能力。MCS-CNN 具有更佳准确和稳定的性能,在高噪声环境下识别精度仍能达到 93.29%,比最优算法约有 7 个百分点的提升。并且使用分离卷积替换原始卷积后明显降低模型参数量,仅为 0.121 M,仅为当前最优算法的 1/26。因此,本文提出的 MCS-CNN 模型更适用于工业高噪声环境下的轴承故障诊断。

3.4 负载测试

为验证模型在时变负载场景下仍保持准确可靠的诊断能力,基于 CWRU 试验数据进行不同负载构建。将使用某一负载下的数据进行训练,以此诊断其他负载下的轴承故障,来体现模型的跨负载故障诊断能力。

如图 8 所示,在注意力机制的加持下, MCS-CNN 模型能够更好地关注振动信号中的多尺度特征,并且每一个集合块的输出都会对信号数据进行随机舍弃,使模型具备较好的鲁棒性。

在 4 种跨负载故障预测中, MCS-CNN 依旧保持最佳的准确率,均超过了 94%,相较于 ResNet 和 MobileNet 约有 3~5 百分点提升。此外,与传统方法 SVM 和 CNN 相比,有着 5~14 百分点的提升,这是因为传统方法不能充分表示轴承的故障特征。WT-CNN 将信号转换成时频图,该方式不能充分覆盖跨负载间信号的特征,因此识别准确率均低于 85%,与本文方法有较大差距。故 MCS-CNN 在面对工业环境中变负载下的轴承故障诊断时具有更好的识别精度。

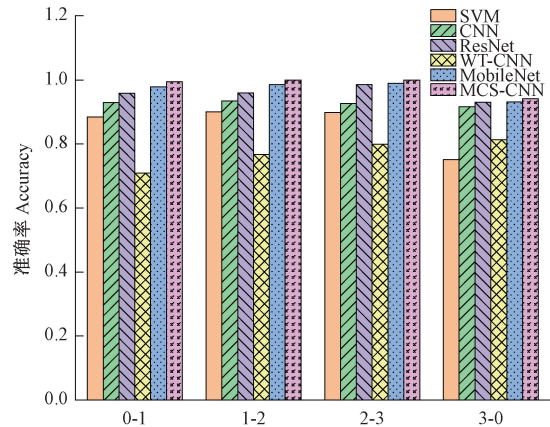


图 8 跨负载故障诊断准确率

Fig. 8 Accuracy of cross load fault diagnosis

3.5 负载及噪声混合测试

在 CWRU 试验数据的基础上构造了负载和噪声混合场景,进一步验证模型的可靠性和优越性。在这种情况下,随机地从 0、1、2、3 hp 中选择数据来构建训练样本。在保证与训练样本不重复的情况下,从 0、1、2、3 hp 中再次随机抽取,并加入不同信噪比的噪声来构建测试样本,测试结果如图 9 所示。

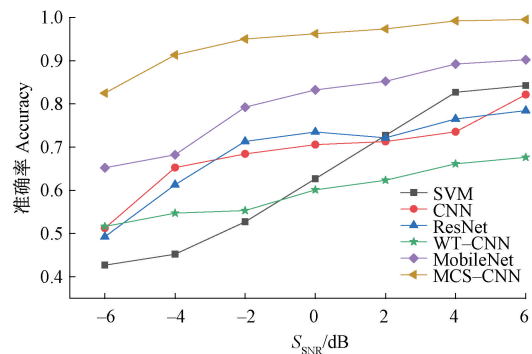


图 9 负载与噪声混合情况下准确率对比

Fig. 9 Comparison of accuracy under mixed load and noise

由图 9 可知,由于混合场景下的数据差异性较大,模型识别准确率均低于先前单一噪声场景,但是更贴合现实场景。其中, SVM、CNN、ResNet 以及 WT-CNN 在该场景下的准确率明显低于 MobileNet 和 MCS-CNN。MCS-CNN 在强噪声环境中仍能达到 82% 的准确率,比现有最佳模型 MobileNet 高出了 15 个百分点左

右。在低噪声环境中,本文模型识别准确率达到100%,相较于 MobileNet 提升了10 百分点。试验结果表明,MCS-CNN 对高噪声、多负载情况下的故障诊断是有效的,同时也表明了本文模型具有更好的泛化性能。

3.6 可视化

深度学习模型的内部机制复杂,为了更好地描述其内部结构,图10所示为模型处理后的不同故障特征。

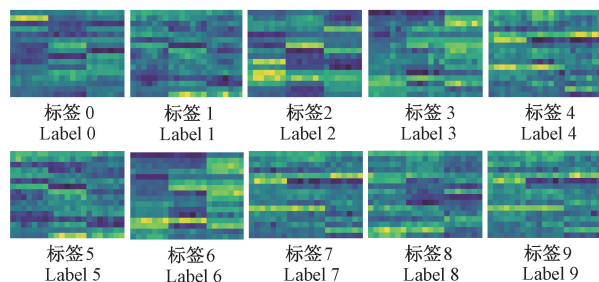


图10 不同故障的特征图

Fig. 10 Feature map of different faults

图10中,标签0表示正常轴承信号的特征图,标签1~标签9分别对应故障标签1~9的轴承信号的特征图。特征图是由一个轴承振动信号经过卷积特征提取后输出的一个12×8的矩阵,经过激活函数的作用后,亮色代表被激活,暗色表示未被激活。由图10中可以看出,不同的故障类型具有不同位置的激活区域,也证实本文方法从原始信号中提取特征具有较强的能力。

为进一步解释模型的有效性,采用 t -SNE^[20] 可视化模型中每个集成块提取的故障特征,可视化结果如图11所示。由图11可以看出,模型故障特征提取的过程具有递进关系,从输入数据时完全交错在一起到第一个集成块逐渐剥离出来直到第二个可以形成明显的点簇。通过第三个集成块后可以基本识别轴承不同故障损伤程度或位置损伤程度,图11中明显可以看出不同故障状态下提取的故障特征有着类内小、类间大的特点。综上,MCS-CNN 模型在提取时序信号特征的过程中具有递进性,从而使网络具备良好的鲁棒性和泛化性能。

4 总结

本文从振动信号内含有尺度特征、工业环境硬件配置低和深层网络导致梯度消失的角度出发,提出了一种新的多尺度多路径集成网络,并验证了本文方法的有效性。本文结论如下:

1)引入多尺度多路径的结构设计,有效增加了网络深度和宽度,使得网络能够学习轴承振动信号内的多尺度特征,具有较好的泛化性和抗噪性。

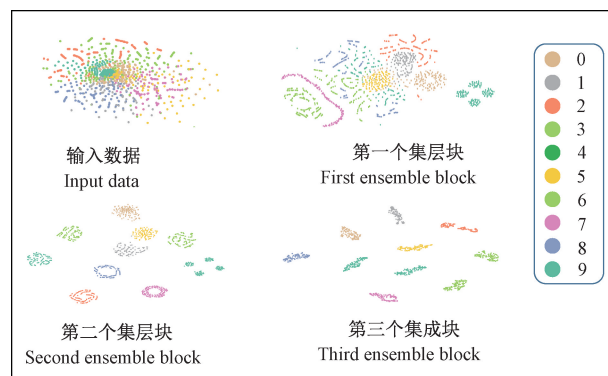


图11 不同集成块输出特征点簇

Fig. 11 Output feature point clusters of different ensemble blocks

2)引入深度可分离卷积,对深度可分离卷积内的深度卷积和逐点卷积进行重构,在确保网络性能的同时降低模型参数,更适用于工业场景应用。

3)与现有方法相比,MCS-CNN 在噪声场景和变负载场景中保持着93%以上的准确率,在混合场景中仍能保持82%以上,体现模型的鲁棒性。

4)对模型内部机制的可视化,本文模型在不同尺度下提取的故障特征具有递进性,使得模型能够表现出较好的泛化性能。

参考文献 (References)

- [1] 徐海铭,夏乔阳,李勇,等. 基于深度可分离卷积神经网络轴承剩余寿命预测[J]. 机械强度, 2022, 44(4): 763-771.
XU HaiMing, XIA QiaoYang, LI Yong, et al. Bearing remaining life prediction based on deep separable convolutional neural network [J]. Journal of Mechanical Strength, 2022, 44(4): 763-771 (In Chinese).
- [2] 郭学卫,申永军,杨绍普. 基于样本熵和分数阶傅里叶变换的滚动轴承故障特征提取[J]. 振动与冲击, 2017, 36(18): 65-69.
GUO XueWei, SHEN YongJun, YANG ShaoPu. Application of sample entropy and Fractional Fourier transform in the fault diagnosis of rolling bearings [J]. Journal of Vibration and Shock, 2017, 36(18): 65-69 (In Chinese).
- [3] 朱冠霖,王兆强,王异凡,等. 基于神经网络和证据融合的液压泵故障诊断研究[J]. 机电工程, 2020, 37(12): 1498-1503.
ZHU GuanLin, WANG ZhaoQiang, WANG YiFan, et al. Hydraulic pump fault diagnosis based on neural network and evidence fusion [J]. Journal of Mechanical & Electrical Engineering, 2020, 37(12): 1498-1503 (In Chinese).
- [4] 曲建岭,余路,袁涛,等. 基于一维卷积神经网络的滚动轴承自适应故障诊断算法[J]. 仪器仪表学报, 2018, 39(7): 134-143.
QU JianLing, YU Lu, YUAN Tao, et al. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network [J]. Chinese Journal of Scientific Instrument, 2018, 39(7): 134-143 (In Chinese).
- [5] 王海龙,夏筱筠,孙维堂. 基于 EMD 与卷积神经网络的滚动轴承故障诊断[J]. 组合机床与自动化加工技术, 2019(10): 46-48.

- WANG HaiLong, XIA XiaoJun, SUN WeiTang. Rolling bearing fault diagnosis based on EMD and convolutional neural network[J]. Modular Machine Tool & Automatic Manufacturing Technique, 2019(10): 46-48 (In Chinese).
- [6] 刘红军, 魏旭阳. 基于 GADF 与卷积神经网络的滚动轴承故障诊断研究[J]. 机电工程, 2021, 38(5): 587-591.
- LIU HongJun, WEI XuYang. Rollingbearing fault diagnosis based on GADF and convolutional neural network [J]. Journal of Mechanical & Electrical Engineering, 2021, 38(5): 587-591 (In Chinese).
- [7] 冯连强, 徐江, 田瑞明, 等. 基于一维卷积神经网络的轴承故障诊断方法研究[J]. 重型机械, 2021(1): 57-62.
- FENG LianQiang, XU Jiang, TIAN RuiMing, et al. Research on bearing fault diagnosis based on one-dimensional convolutional neural network [J]. Heavy Machinery, 2021(1): 57-62 (In Chinese).
- [8] 刘洋, 程强, 史曜炜, 等. 基于注意力模块及 1D-CNN 的滚动轴承故障诊断[J]. 太阳能学报, 2022, 43(3): 462-468.
- LIU Yang, CHENG Qiang, SHI YaoWei, et al. Fault diagnosis of rolling bearings based on attention module and 1D-CNN [J]. Acta Energaie Solaris Sinica, 2022, 43(3): 462-468 (In Chinese).
- [9] ZHANG L, XIONG G L, LIU H S, et al. Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference [J]. Expert Systems with Applications, 2010, 37(8): 6077-6085.
- [10] HUANG W Y, CHENG J S, YANG Y, et al. An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis[J]. Neurocomputing, 2019, 359: 77-92.
- [11] LIU H H, HAN M H. A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings [J]. Mechanism and Machine Theory, 2014, 75: 67-78.
- [12] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 1800-1807.
- [13] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. 2017: 1704. 04861. <http://arxiv.org/abs/1704.04861v/>.
- [14] 朱富, 刘畅, 王贵勇, 等. 改进 MoblieNet 网络在轴承轻量化诊断中的应用[J]. 机械科学与技术, 2024, 43(1): 31-38.
- ZHU Fu, LIU Chang, WANG GuiYong, et al. Application of improved MobileNet network in bearing lightweight diagnosis [J]. Mechanical Science and Technology for Aerospace Engineering, 2024, 43(1): 31-38 (In Chinese).
- [15] 鲍文霞, 吴德钊, 胡根生, 等. 基于轻量型残差网络的自然场景水稻害虫识别[J]. 农业工程学报, 2021, 37(16): 145-152.
- BAO WenXia, WU DeZhao, HU GenSheng, et al. Rice pest identification in natural scene based on lightweight residual network [J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(16): 145-152 (In Chinese).
- [16] 于龙蛟, 于博, 李春庚, 等. 优化卷积网络及低分辨率热成像的夜间人车检测与识别[J]. 红外技术, 2020, 42(7): 651-659.
- YU LongJiao, YU Bo, LI ChunGeng, et al. Detection and recognition of persons and vehicles in low-resolution nighttime thermal images based on optimized convolutional neural network [J]. Infrared Technology, 2020, 42(7): 651-659 (In Chinese).
- [17] 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.
- [18] CHEN X J, YANG Y M, CUI Z X, et al. Vibration fault diagnosis of wind turbines based on variational mode decomposition and energy entropy[J]. Energy, 2019, 174: 1100-1109.
- [19] LIANG P F, DENG C, WU J, et al. Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform [J]. Computers in Industry, 2019, 113: 103132.
- [20] 邓佳林, 邹益胜, 张笑璐, 等. 一种改进 CNN 在轴承故障诊断中的应用[J]. 现代制造工程, 2020(4): 142-147.
- DENG JiaLin, ZOU YiSheng, ZHANG XiaoLu, et al. Application of an improved CNN in fault diagnosis of bearings [J]. Modern Manufacturing Engineering, 2020(4): 142-147 (In Chinese).