

一种结合改进 Inception V2 模块和 CBAM 的 轴承故障诊断方法

姚齐水, 别帅帅, 余江鸿, 陈前旭

(湖南工业大学机械工程学院, 湖南 株洲 412007)

摘要: 传统深度学习的轴承故障诊断方法网络复杂, 训练参数多, 模型泛化性弱。针对上述问题, 在工业大数据背景下, 提出一种结合改进 Inception V2 模块和 CBAM 注意力机制的轴承故障诊断方法, 改进后的 Inception V2 模块通过增加平均池化层, 进一步拓宽分支网络结构, 从而提高网络表达能力。将轴承振动信号通过小波变换转换为时频图, 作为卷积神经网络的输入, 通过改进 Inception V2 模块对输入特征进行自适应特征提取, 跨通道对提取的特征进行信息组织; 通过 CBAM 注意力机制生成通道和空间的双重注意力权重, 增强相关度高的特征并抑制相关度不高的特征; 将生成的特征数据输入到全局平均池化层, 并输出故障诊断结果。实验结果表明: 该方法可以建立“浅层”卷积神经网络模型, 减少模型参数, 加快模型收敛速度, 实现 99.75% 的准确率; 同时在不同负载以及高噪声条件下, 模型有较好的泛化性, 更适合应用在工业大数据中。

关键词: 故障诊断; 滚动轴承; 卷积神经网络; CBAM 注意力机制; 故障特征

中图分类号: TH165⁺.3; TH133.33 **文献标志码:** **文章编号:** 1004-4523(2022)04-0949-09

DOI: 10.16385/j.cnki.issn.1004-4523.2022.04.019

引言

滚动轴承作为机械设备的關鍵基础部件, 其性能状态将决定机械设备能否安全可靠地运行。对轴承的运行状态进行有效地检测和故障预警, 可以有效避免事故的发生, 减少经济损失^[1]。

在《中国制造 2025》的蓝图下^[2], 中国的机械工业信息化进入了快速发展轨道, 机械设备更加复杂, 产生的数据海量^[3]。传统的故障诊断方法已经不能满足轴承类转动部件故障诊断的需求, 因此, 基于数据驱动的故障诊断方法具有更大的研究价值, 基于深度学习方法的故障诊断技术将更能满足需求。目前, 越来越多的卷积神经网络正逐步从巨型网络向轻型网络发展, 其模型精度也越来越高。高峰等^[4]通过卷积神经网络自适应提取特征的优点, 建立了基于卷积神经网络的诊断模型。当样本数据较少且分布不均衡时, 故障诊断率仍然很高。袁建虎等^[5]提出了一种基于小波时频图像和卷积神经网络的轴承诊断方法。利用小波变换得到轴承振动信号的时频图像, 以灰度图像作为输入, 利用卷积神经网络对轴承故障类型进行识别。庞俊等^[6]通过改进非线性映射函数, 同时在卷积神经网络中加入残差神经元,

把振动信号转换为灰度图的形式作为特征输入, 经过试验得到比较高的准确率。传统深度学习的轴承故障诊断方法存在网络复杂, 训练参数多, 模型泛化性弱的问题。

近些年, 网络模型中引入注意力机制能够提取到输入特征中更为关键的信息, 因此被广泛应用于各类深度学习任务当中。朱浩等^[7]提出了一种基于感知网络的结构, 通过注意机制对通道中的数据特征赋予不同的权重, 从而获得更关键的特征数据和更准确的识别率。邓佳林等^[8]构建了基于通道注意机制的一维卷积神经网络故障诊断模型 SECNN。通过对信道特征的学习, 增强了网络的适应性, 与传统的卷积神经网络相比有了显著的提高。相比上述注意力机制, CBAM (Convolutional Block Attention Module) 结合空间和通道的双重注意力机制, 能够获得更为全面的特征信息。

本文提出一种基于 Inception V2 模块进行改进的卷积神经网络模型, 引入 CBAM 注意力机制, 将凯斯西储大学轴承数据集作为输入, 通过改进的网络模型, 在不失轴承故障识别的准确率的同时, 简化网络, 减少训练参数, 提高训练速度, 在保证训练准确率的同时提高效率, 为轴承故障诊断提出了一种新方法。

收稿日期: 2021-09-23; **修订日期:** 2021-11-22

基金项目: 湖南省自然科学基金资助项目(2021JJ50054)。

1 卷积神经网络

1.1 卷积层与池化层

早期的卷积神经网络是一种受视觉系统神经机制启发的生物物理模型^[9]。多隐层人工神经网络是为模仿生物大脑皮层而专门设计的,广泛应用于计算机视觉和深度学习等相关领域。卷积神经网络通常具有深层结构,一般由输入层、卷积层、下采样层、全连接层和输出层组成。卷积层也称为“检测层”,下采样层也称为“池化层”。卷积层和池化层是神经网络特征提取的核心模块。在卷积层,卷积核执行卷积运算如下式所示:

$$y_j^l = f\left(\sum_i W_{ij}^l x_i^{l-1} + b_i^l\right) \quad (1)$$

式中 b_i^l 为偏置项; W_{ij}^l 为权值项; y_j^l 为是第 l 层卷积层特征; x_i^{l-1} 为 $l-1$ 层的第 i 个特征图; $f(\cdot)$ 为激活函数。

遍历输入特征图,自动提取特征;在卷积之后,使用 ReLu 激活函数,如下式所示:

$$\text{ReLu}(x) = \begin{cases} x & x > 0 \\ 0 & x < 0 \end{cases} \quad (2)$$

通过激活函数,将非线性因素引入神经元,使神经网络可以任意逼近任何非线性函数,并可应用于多种非线性模型。

池化层是一种降采样操作,保留输入数据中显著的特征,降低特征的维度,增大卷积核的感受野,使模型可以抽取更加全面的特征;同时,减少了下一层的输入,从而减少了计算量,加快了模型的优化。

1.2 Inception V2 模块

Inception 模块^[10]是卷积神经网络发展史上的一个重要节点。在 Inception 模块出现之前,大多数流行的卷积神经网络只是将卷积层越来越直接地堆叠起来,使网络越来越深,并增加卷积神经网络的规模以获得更好的性能,但这样网络就更容易过拟合,计算量也会成倍地增加。在此背景下,2014 年 ImageNet 竞赛冠军 GoogLeNet (Inception V1) 诞生了。GoogLeNet 的核心是它的基本单元 Inception 模块,同一层网络内使用不同尺寸的卷积核,提升了模型感知力,可以提取不同尺寸的特征,通过 1×1 卷积核作用到输入特征图的每个像素点,依靠减少 1×1 卷积核个数,减少输入输出特征图深度,从而达到降维的目的,减少了参数量和计算量。本文以 Inception V2 模块为基础,增加平均池化层,将每一层网络进行拓宽,分别设置 1×1 , 3×3 卷积核,以及

最大池化层,结构如图 1 所示。通过 1×1 卷积来降低通道数把信息聚集,将特征沿深度方向叠加输出,通过全局平均池和全连接层对数据进行分类。

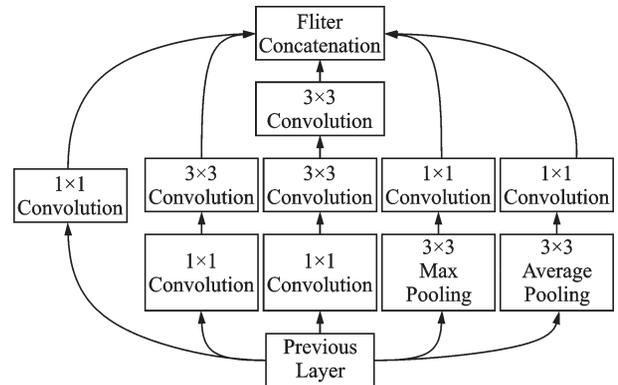


图 1 改进的 Inception V2 模块

Fig. 1 Improved Inception V2 module

1.3 CBAM 注意力机制模块

CBAM (Convolutional Block Attention Module) 双重注意力机制由 Woo 等^[11]首次提出,与只关注通道特征的 SENet 相比, CBAM 是一种结合了空间和通道的注意力模块,增强特征图中的有用特征,抑制无用特征,在实际应用当中可以取得更好的效果。图 2 为 CBAM 注意力模块的示意图。首先通过通道注意力机制模块对输入特征图进行处理,在通道注意力模块当中分别进行全局平均池化,全局最大池化提取更为丰富的高层次特征;接着分别通过 MLP (Multilayer Perceptron), 将通道数压缩为 C/r , 再扩张回 C , 其中 C 为通道数, r 为衰减比率, 取 $r=16$; 然后将 MLP 输出的特征经过 sigmoid 激活操作, 生成最终的通道注意力模块输出, 将该输出对应元素相乘。计算流程如下式所示:

$$M_c(F) = \sigma(MLP(\text{AvgPool}(F)) + MLP(\text{MaxPool}(F))) = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))) \quad (3)$$

式中 F 表示输入特征; σ 表示 sigmoid 激活函数。

空间注意力模块主要探讨在空间层面特征图的内在关系,即突出区域的重要性,与通道注意力模块相辅相成。空间注意力模块在算法上相对简单些,把通道注意力模块输出作为空间注意力模块所需要的输入,经过卷积核大小为 7×7 的标准卷积层后获得空间注意力模块的特征图。计算流程如下:

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (4)$$

式中 $f^{7 \times 7}$ 表示卷积核大小为 7×7 。

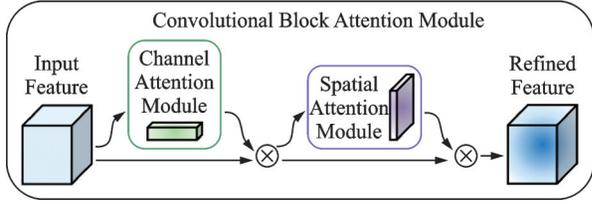


图2 CBAM注意力机制示意图

Fig. 2 Schematic diagram of CBAM attention mechanism

1.4 模型优化

在Inception V2模块中引入批量归一化(Batch Normalization)层^[11],解决在训练过程中中间层数据分布发生改变的问题,防止梯度消失或爆炸,加快训练速度。主要计算方法是:首先求出该批次数据的均值,再求出该批次数据的方差,对样本数据进行标准化处理,最后引入两个可学习的参数 β 和 γ ,使网络能够学习原始网络的特征分布。具体操作过程如下式所示:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (5)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (6)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (7)$$

$$y_i = \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i) \quad (8)$$

式中 σ_B^2 和 μ_B 分别表示每批次数据的方差和均值; \hat{x}_i 为每批次数据经过标准化处理后的数据; ϵ 为常数项; γ 和 β 为可学习参数。

Inception V2模块中使用了 1×1 的卷积核用来减少维度和修正线性激活(ReLU)。当网络模型输入为 $224 \times 224 \times 64$ 时,不经过 1×1 卷积核直接进入 64 通道 3×3 卷积核(使用全0填充),参数量为 $3 \times 3 \times 64 \times 64 = 36864$;经过 48 通道卷积核 1×1 后再进入 3×3 卷积核时,参数量为 $1 \times 1 \times 64 \times 48 + 3 \times 3 \times 48 \times 48 = 23808$,参数大约减少到原来的三分之二。因此通过 1×1 卷积核,能够对特征图进行降维,减少模型参数,能够有效加速模型收敛。

和全连接层相比,通过全局平均池化的方法^[12],赋予每一个通道实际的类别意义,全局平均池化的出现让卷积结构更简单,大大减少参数数量,达到优化网络结构,防止过拟合的目的,是一种更朴素的卷积结构选择。全局平均池化操作原理如下:假设要预测 K 个类别,在卷积特征提取的最后一层卷积层会生成 K 个特征图,通过全局平均池化就可以得到 K 个 1×1 的特征图,将这些 1×1 的特征图输入到Softmax分类器之后,每一个输出结果代表着这 K 个类别的概率,从而起到取代全连接层的效果。

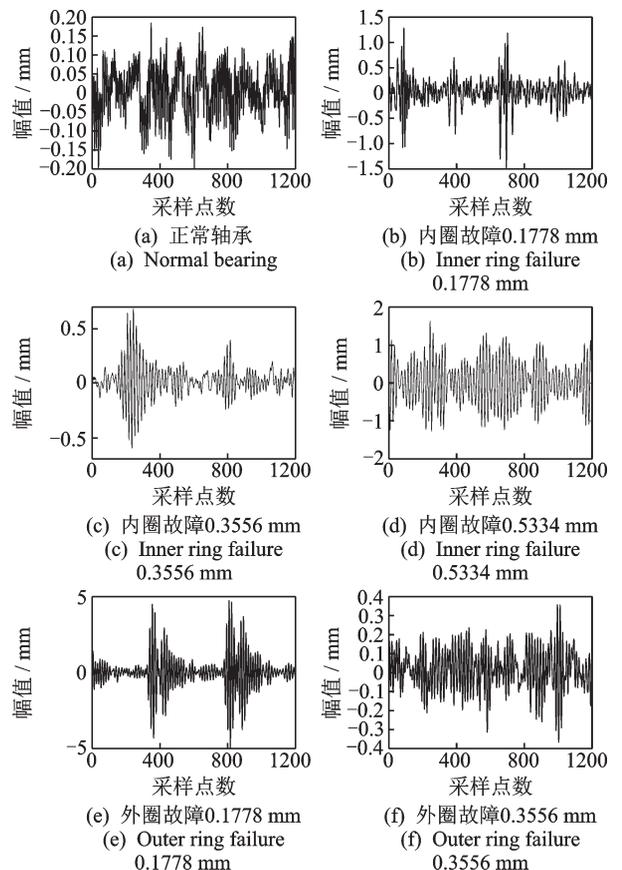
2 实验验证和性能分析

2019年Google推出了TensorFlow 2版本,改进的TensorFlow 2相对于1版本将Keras作为核心的高级API,更容易入手,因此本次实验采用使用TensorFlow 2.3搭建模型。

2.1 轴承故障诊断

2.1.1 实验数据

本文采用凯斯西储大学轴承数据集中采样频率为 48 kHz 驱动端轴承数据,对CBAM-InceptionNet卷积神经网络进行验证。轴承试验台由 1.5 kW 电机、转矩传感器、功率试验机和电子控制器组成。待测轴承支撑电机转轴。驱动端轴承为SKF6025。采用电火花加工技术,分别在轴承外圈、内圈和滚动体三个部位人为加工产生直径为 0.1778 mm , 0.3556 mm 以及 0.5334 mm 的单个故障。故障轴承安装在试验电机上,分别在 0 马力(1797 r/min)、 1 马力(1772 r/min)、 2 马力(1750 r/min)、 3 马力(1730 r/min)四种负载条件下进行试验。加速度传感器采集故障轴承的振动加速度信号,本次实验取 $0, 1, 2$ 马力三种负载工作条件下外圈、内圈、滚动体故障数据,其中外圈选择损伤点在 6 点钟方向位置的故障。图3为 1 马力工作条件下正常轴承和故障轴承时频信号。



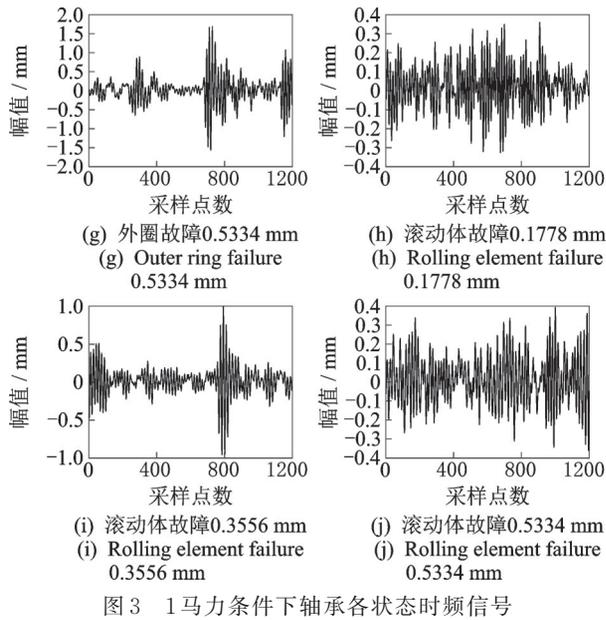


图 3 1 马力条件下轴承各状态时频信号
Fig. 3 Time-frequency signal of each state of the bearing under the condition of 1 horsepower

2.1.2 数据处理

传统的信号分析方法大多适应于工况稳定的情况下,如包络谱法等,而滚动轴承振动信号是一种典型的非线性、非平稳的振动信号,单一的时域分析或者频域分析在分析非平稳振动信号时能力有限,而通过时频分析,提供了时域与频域的联合信息分布,能够在时频域中识别出该信号的真实频率,提取更为有效的信号特征。

短时傅里叶变换在对非平稳振动信号进行分析时,窗函数太宽会导致时间分辨率低,窗函数太窄又会导致频率分辨率低,如何设置窗口的宽度成为了该方法一个明显的缺陷。小波变换能够有效弥补短时傅里叶变换的不足,自动适应不同频率信号的分析要求,对信号进行多尺度细化分析,因此更适合应用在滚动轴承振动信号分析中。通过连续小波变换(CWT)将振动信号转化为时频图,既能够生成详细反映信号特征的二维图,又能够充分发挥卷积神经网络在图像处理问题中的优势,更适合应用在本文的方法中。由于Morlet复小波在时域和频域均有较好的分辨率,更适合应用在工程实践中,因此本文采用Morlet复小波进行信号处理。

2.1.3 数据集构建

本文选取凯斯西储大学 48 kHz 采样频率下的驱动端轴承故障数据进行分析。首先用cmor3-3小波对 0 马力条件下轴承振动数据进行小波信号处理,将一维振动信号转换为二维的时频图(224×224),如图 4 所示。通过 TensorFlow 2.3 制作数据集,设置每种故障的训练集为 800,测试集为 200,将训练集分别输入到卷积神经网络中,最后通过测试集验证模型故障识别准确率。

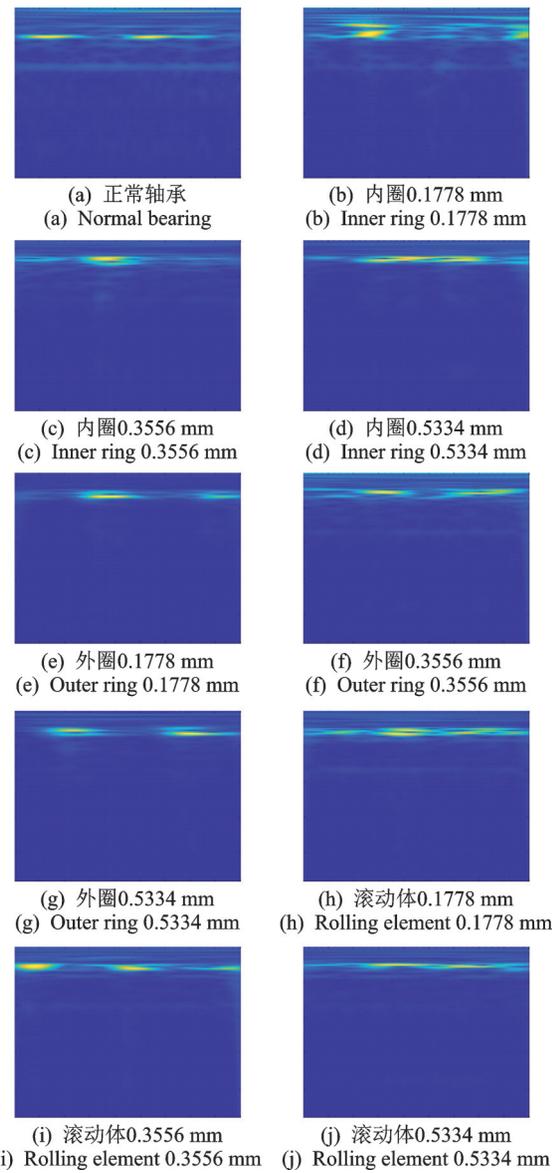


图 4 0 马力条件下轴承各状态时频图
Fig. 4 Time-frequency diagram of each state of the bearing under the condition of 0 horsepower

2.2 模型构建

改进后的 Inception V2 模块增加平均池化层拓宽网络的宽度,在跨通道组织信息时,获得更为全面的特征信息,提高网络表达能力,在每个改进后的 Inception V2 模块后加入 CBAM 注意力模块,进一步提高故障识别能力。通过分析卷积神经网络在滚动轴承故障诊断任务中的表现,可以发现深层的网络虽然能够提取到更为深层的特征图像,但同样也会带来更多的计算量,模型也会增大;而浅层的网络又会经常面临过拟合的问题,模型不够稳定且影响故障诊断精度。因此,合理的模型结构可以获得比较丰富的特征图,提取到更多的有效特征,既能够准确地对滚动轴承故障进行诊断,又能够减少冗余的计算,减小模型大小。

首先对输入数据进行标准卷积运算,经过最大

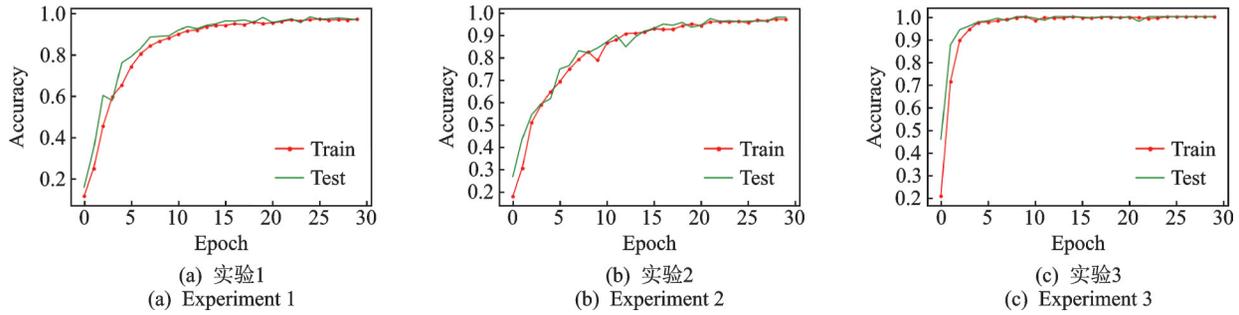


图7 训练准确率

Fig. 7 Training accuracy

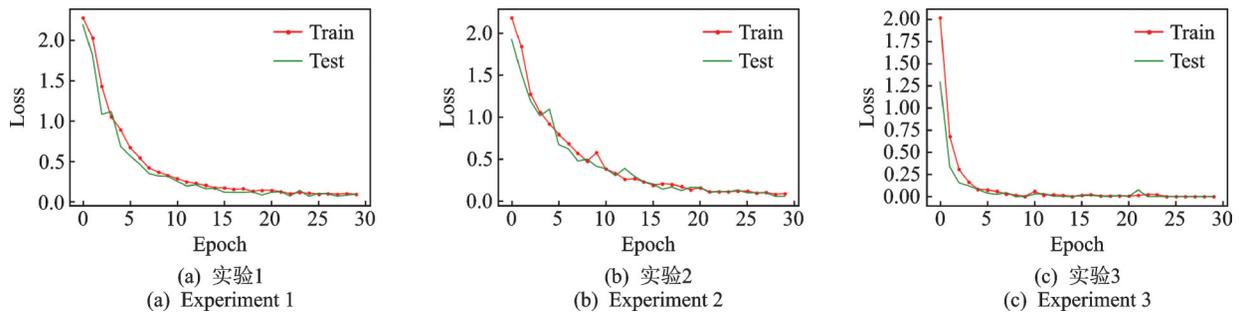


图8 训练损失值

Fig. 8 Training loss

训练至第21个epoch时模型收敛,但模型不够稳定。实验三在经过9次训练后,模型收敛,训练集准确率达到100%,Loss值从2.0降低至0.0008,测试集准确率达到100%,Loss值从1.3降低至0.0008。实验结果表明,在对输入特征进行镜像、翻转、随机裁剪等变化后,改进的Inception V2模块有助于提高故障识别准确率,加入CBAM注意力机制后,模型收敛速度、图像分类的准确率、模型的稳定性都得到了很好的提升。为进一步说明改进后的CBAM-InceptionNet模型的优越性,分别记录三组实验训练结束后的总参数量如表2所示。从表中可以清晰地看到,增加Inception V2模块宽度能够减少训练参数,提升训练准确率,但相对模型稳定性有所下降,引入CBAM注意力机制后,模型总参数仅增加6950,相比较而言,参数增加0.004%,准确率可以提升100%。

表2 各实验训练总参数

组别	训练次数/epoch	总参数量
1	30	1580602
2	30	1482762
3	30	1587552

2.3.2 变负载条件下模型性能验证

实际工作中的轴承会在不同负载的条件下运转,故障振动频率改变,故障特征也会随之改变。为

验证CBAM-InceptionNet模型在变载负载环境下自适应性,分别用1,2,3马力负载下的实验数据训练三组实验模型,而后用另外两种负载数据集作为测试集以验证CBAM-InceptionNet模型故障诊断性能。同时引入传统两层CNN模型、文献[8]中人工特征+SVM模型以及文献[13]中的WDCNN作为对比实验,实验结果如图9所示。

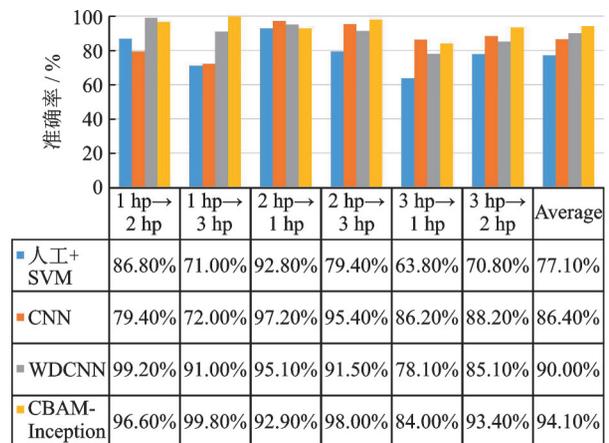


图9 不同模型在变负载条件下故障识别准确率

Fig. 9 Different models' fault recognition accuracy rate under variable load conditions

分析图9可知,人工提取特征配合支持向量机的方法准确率低于另外三种网络模型,主要原因在于人工提取故障特征自适应性不强,支持向量机的非线性表达能力不足导致了整体故障识别率低的结果;传统的卷积神经网络在激活函数的激励下,有比

较强的非线性表达能力,但是浅层的模型识别能力不强,深层的模型收敛慢,且模型的泛化性有所欠缺;WDCNN使用 116×1 的大卷积核获取输入特征的全局信息,故障识别率能达到90%,有较大的提升;CBAM-InceptionNet通过Inception模块获得输入特征的不同信息,引入双重注意力机制,进一步突出有用特征,弱化无用特征,故障识别率整体高于WDCNN,达到了94.1%。整体来看,CBAM-InceptionNet在变负载环境下有较好的自适应性,故障识别准确率保持在一个较高的水平,且浮动较小,模型的稳定性明显优于另外三种模型。

2.3.3 模型自适应性验证

在实验一0马力负载训练集中随机加入少量1,2,3马力负载数据,令模型在新构建的训练集中进行一定次数的微调训练,并进行测试。实验结果如图10所示,其中模型在对1马力负载测试集进行试验时,故障识别率能达到100%,平均故障识别准确率达到97.8%。按照均等数量分别从四组实验测试集中随机选取,构建包含四种复杂环境下的混合测试集,测试准确率达到99.11%,引入混淆矩阵(如图11所示)进一步说明模型在不同负载条件下有较好的自适应性,能够准确识别故障类型。

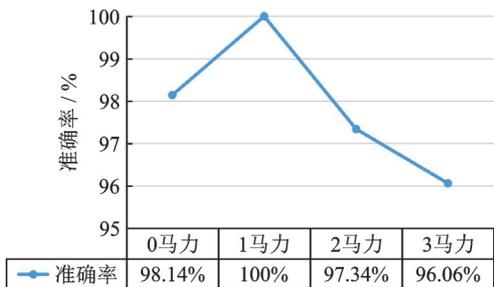


图10 不同负载条件实验结果图

Fig. 10 Experimental results diagram of different load conditions

Label	B_1	104	0	0	0	0	0	0	2	0	
	B_2	0	108	0	0	0	0	0	0	0	
	B_3	0	0	92	0	0	0	0	0	0	
	I_1	0	0	0	101	0	0	0	0	0	
	I_2	0	0	0	0	97	0	0	0	0	
	I_3	0	0	0	0	0	102	0	0	0	
	Nor	0	0	0	0	0	0	102	0	0	
	O_1	0	0	0	0	0	0	0	87	0	
	O_2	0	0	0	0	0	0	0	0	101	
	O_3	0	0	0	0	0	0	0	0	0	104
		B_1	B_2	B_3	I_1	I_2	I_3	Nor	O_1	O_2	O_3
		Predict									

图11 混淆矩阵

Fig. 11 Confusion matrix

2.3.4 高噪声条件下模型性能验证

在实际运行过程中,轴承所处的工作环境往往比较复杂,采集到的轴承振动信号会夹杂各种噪声。因此,为了进一步验证模型在不同噪声环境下的故障诊断性能,根据噪声数据与正常数据的3:8比例构造训练集,并用加入噪声后的测试集来评价模型的性能。

以内圈0.1778 mm故障为例,加入信噪比 $SNR=10,12,14,16,18,20$ 的噪声信号。图12为原始信号与添加信噪比 $SNR=10,20$ 的时域信号,通过对比可知信噪比 SNR 数值越大,噪声越小,信号失真越小。

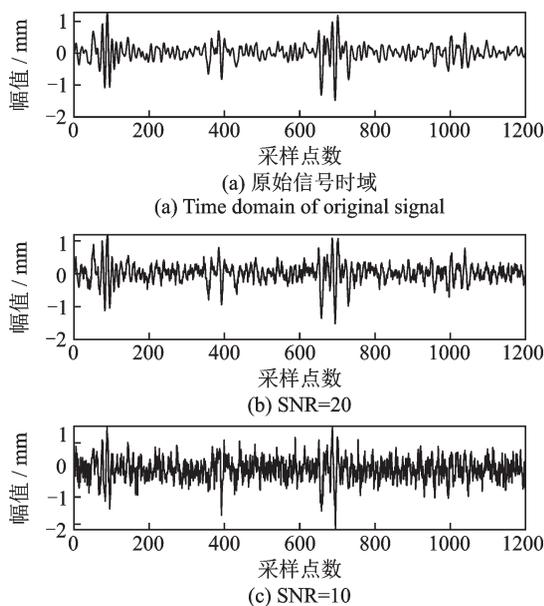


图12 轴承内圈0.1778 mm故障信号与加噪信号时域图

Fig. 12 Time domain diagram of 0.1778 mm fault signal and noise signal of bearing inner ring

为了对比实验,引用文献[14]中实验数据作为对照,实验结果如表3所示,通过分析可知,CBAM-Inception网络模型平均故障诊断准确率可以达到94.92%,在噪声环境下故障识别准确率最高,模型自适应性强;Inception V1模型使用多尺度卷积增加

表3 模型抗噪声实验结果

Tab. 3 Anti-noise experiment results of models

	模型		
	Inception V1	VGG16	CBAM-Inception
$SNR=10$	58.43%	73.68%	92.12%
$SNR=12$	66.77%	82.30%	94.25%
$SNR=14$	72.73%	87.52%	93.68%
$SNR=16$	76.42%	90.30%	96.35%
$SNR=18$	77.67%	91.34%	96.40%
$SNR=20$	78.74%	91.81%	96.75%
平均值	71.79%	86.15%	94.92%

网络宽度,因为其缺乏通道以及空间的特征标定,模型学习深度不够从而导致准确率较低;VGG16模型通过不断叠加网络深度,提高了模型的故障诊断能力。然而,随着网络深度的增加,训练参数的数量也随之增加,精度也有所提高,但模型的实际效益相对较低。

2.3.5 模型泛化性验证

为了充分验证本文方法的优越性以及泛化性,采用西安交通大学轴承数据集^[15]进行实验,其中西安交通大学轴承数据集采用的是 LDK UER204 滚动轴承,不同于凯斯西储大学数据集中的 SKF6205 滚动轴承。

首先选取数据集中内圈故障、外圈故障、保持架故障以及正常四类状态数据,按照相同的数据处理方法将振动信号转换为时频图,然后分别组成训练集以及测试集以供模型实验验证。通过分析图 13 可知,本文方法能够有效诊断不同型号的滚动轴承故障,有较为广泛的应用场景。

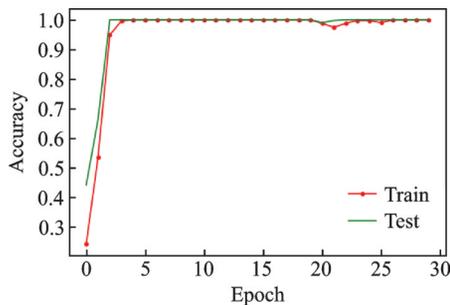


图 13 西安交通大学轴承实验结果

Fig. 13 Bearing test results of Xi'an Jiaotong University

3 结 论

针对传统卷积神经网络忽略通道特征和空间特征所造成的故障识别率低、模型参数多、以及自适应性差等问题,提出了改进后的 CBAM-InceptionNet 故障诊断模型,简化网络模型的结构,弥补了通道特征和空间特征的缺失。实验结果表明,该模型在轴承故障诊断中具有良好的性能。

(1)提出的 CBAM-InceptionNet 卷积神经网络,能够快速有效地进行轴承故障诊断。

(2)改进后的 Inception V2 模块把不同卷积核大小的输出沿深度方向上堆叠在一起,提取不同层次特征,借助这种自动提取特征的功能,可以直接从原始图像像素中提取更有效的特征数据,并学习良好的特征表示。

(3)CBAM-InceptionNet 在变负载条件以及高噪声下表现稳定,能够在不同型号的滚动轴承故障

识别中保持一个较高故障诊断水平。

(4)构建以 Inception 模块为主引入注意力机制的卷积神经网络,模型训练更为快速,可以占用更少的计算机资源,为简化神经网络提供了一种方法。

参考文献:

- [1] 张神林. 基于卷积神经网络的滚动轴承及行星齿轮箱故障诊断方法[D]. 马鞍山: 安徽工业大学, 2018.
Zhang Shenlin. Fault diagnosis method of rolling bearing and planetary gearbox based on convolutional neural network[D]. Ma'anshan: Anhui University of Technology, 2018.
- [2] 黄乐安. 德国工业 4.0 对中国制造 2025 战略的启示[J]. 中外企业家, 2016(34): 265-267.
Huang Lean. The enlightenment of German Industry 4.0 to China's 2025 Strategy[J]. Chinese and Foreign Entrepreneurs, 2016(34): 265-267.
- [3] 雷亚国, 贾峰, 孔德同, 等. 大数据下机械智能故障诊断的机遇与挑战[J]. 机械工程学报, 2018, 54(5): 94-104.
Lei Yaguo, Jia Feng, Kong Detong, et al. Opportunities and challenges of mechanical intelligent fault diagnosis under big data[J]. Chinese Journal of Mechanical Engineering, 2018, 54(5): 94-104.
- [4] 高峰, 曲建岭, 余路, 等. 基于卷积神经网络的滚动轴承故障诊断算法[J]. 信息技术, 2019, 43(4): 68-72.
Gao Feng, Qu Jianling, Yu Lu, et al. Rolling bearing fault diagnosis algorithm based on convolutional neural network[J]. Information Technology, 2019, 43(4): 68-72.
- [5] 袁建虎, 韩涛, 唐建, 等. 基于小波时频图和 CNN 的滚动轴承智能故障诊断方法[J]. 机械设计与研究, 2017, 33(2): 93-97.
Yuan Jianhu, Han Tao, Tang Jian, et al. Intelligent fault diagnosis method for rolling bearings based on wavelet time-frequency diagram and CNN[J]. Machinery Design and Research, 2017, 33(2): 93-97.
- [6] 庞俊, 刘鑫, 段敏霞, 等. 基于改进卷积神经网络轴承故障诊断[J]. 组合机床与自动化加工技术, 2021(3): 66-69.
Pang Jun, Liu Xin, Duan Minxia, et al. Bearing fault diagnosis based on improved convolutional neural network[J]. Modular Machine Tool and Automatic Manufacturing Technology, 2021(3): 66-69.
- [7] 朱浩, 宁芊, 雷印杰, 等. 基于注意力机制-Inception-CNN 模型的滚动轴承故障分类[J]. 振动与冲击, 2020, 39(19): 84-93.
Zhu Hao, Ning Qian, Lei Yinjie, et al. Fault classifica-

- tion of rolling bearing based on attention mechanism-Inception-CNN model [J]. *Journal of Vibration and Shock*, 2020, 39(19): 84-93.
- [8] 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. An improved CNN application in bearing fault diagnosis [J]. *Modern Manufacturing Engineering*, 2020 (4) : 142-147.
- [9] Goodfellow I, Bengio Y, Courville A. *Deep Learning* [M]. Cambridge: the MIT Press, 2016.
- [10] LIN M, CHEN Q, YAN S. Network in network [J]. arXiv preprint arXiv: 1312.4400, 2013.
- [11] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3-19.
- [12] 胡晓依, 荆云建, 宋志坤, 等. 基于CNN-SVM的深度卷积神经网络轴承故障识别研究[J]. *振动与冲击*, 2019, 38(18): 173-178.
- Hu Xiaoyi, Jing Yunjian, Song Zhikun, et al. Research on bearing fault recognition of deep convolutional neural network based on CNN-SVM [J]. *Journal of Vibration and Shock*, 2019, 38(18): 173-178.
- [13] Zhang Wei, Peng Gaoliang, Li Chuanhao, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals [J]. *Sensors*, 2017, 17(2): 425.
- [14] 胡向东, 梁川. 基于SE-ResNeXt的滚动轴承故障诊断方法[J]. *计算机测量与控制*, 2021, 29(7): 46-51.
- Hu Xiangdong, Liang Chuan. The fault diagnosis method of rolling bearing based on SE-ResNeXt [J]. *Computer Measurement and Control*, 2021, 29(7): 46-51.
- [15] Wang Biao, LEI Yaguo, LI Naipeng, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearing [J]. *IEEE Transactions on Reliability*, 2020: 69(1): 401-412.

A bearing fault diagnosis method combining improved inception V2 module and CBAM

YAO Qi-shui, BIE Shuai-shuai, YU Jiang-hong, CHEN Qian-xu

(School of Mechanical Engineering, Hunan University of Technology, Zhuzhou 412007, China)

Abstract: The traditional deep learning bearing fault diagnosis method has a complex network, many training parameters, and weak model generalization. In response to the above problems, under the background of industrial big data, a bearing fault diagnosis method combining the improved Inception V2 module and the CBAM attention mechanism is proposed. The improved Inception V2 module further broadens the branch network structure by adding the average pooling layer, thereby improves network expression ability. The bearing vibration signal is converted into a time-frequency image through wavelet transform, which is used as the input of the convolutional neural network. The features of the input are adaptively extracted through the improved Inception V2 module, and the extracted features are organized across channels. Through CBAM attention mechanism, dual attention weights of channel and space are generated, enhancing the features with high correlation and suppressing the features with low correlation. The generated feature data is input to the global average pooling layer and the fault diagnosis result is outputted. Experimental results show that this method can establish a "shallow" convolutional neural network model, reduce model parameters, speed up model convergence, and achieve an accuracy of 99.75%. At the same time, under different loads and high noise conditions, the model has good generalization. It is more suitable for application in industrial big data.

Key words: fault diagnosis; rolling bearing; convolutional neural network; CBAM attention mechanism; fault characteristics

作者简介: 姚齐水(1967—),男,博士,教授。电话:13873383978;E-mail:yaoqishui@126.com。