

DOI:10.19651/j.cnki.emt.2105654

# 基于 Inception-ResNet 模型的轴承故障分类\*

孔子宇 王海瑞

(昆明理工大学 信息工程与自动化学院 昆明 650500)

**摘要:** 针对常用的故障诊断深度学习需要较高的设备成本与较长的训练时间,提出一种基于 Inception-ResNet 模型的轴承故障分类方法。通过使用 Inception 网络的并行结构使网络学习到不同尺度的特征,引入了残差结构来减少因网络加深所导致的退化现象,并加入了三维卷积,使不同通道间的信息相互交融。为验证本文方法的性能,使用凯斯西储大学数据集与 IMS 数据集进行验证,并与传统的浅层学习方法和深度学习方法进行了对比实验。结果表明,相较于其他方法,所提方法不仅拥有优良的诊断能力,在资源占用与训练效率上也更加优秀。

**关键词:** 卷积神经网络; Inception 网络结构; 轴承故障诊断

中图分类号: TP2 文献标识码: A 国家标准学科分类代码: 510.1050

## Fault classification of bearing based on the Inception-ResNet model

Kong Ziyu Wang Hairui

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** In view of the high equipment cost and long training time required by common deep learning methods for fault diagnosis, this paper proposes a bearing fault classification method based on the Inception-ResNet model. By using the Inception network's parallel structure, the network learns features of different scales, resizing structures are introduced to reduce degradation caused by network deepening, and three-dimensional convolution is added to allow information between different channels to blend. In order to verify the performance of this method, case Western Reserve University data set and IMS data set were used for verification, and compared with the traditional shallow learning method and deep learning method, experiments were conducted. The results show that, compared with other methods, the method presented not only has excellent diagnostic ability, but also is better in terms of resource utilization and training efficiency.

**Keywords:** convolution neural network; Inception network structure; bearing fault diagnosis

## 0 引言

从 20 世纪以来,机械设备都朝着向精密化、大型化的方向发展,轴承作为一种关键的机械零件,广泛使用在工业、运输、农业等诸多方面,与人们的生活息息相关。若机械中的轴承发生了损坏,可能会产生重大的经济损失与人员伤亡。因此,快速、低成本的轴承故障检测与诊断在减少机械设备的故障与经济损失中有重要意义<sup>[1]</sup>。

传统的轴承故障诊断的方法包括振动噪声分析、温度测定、磨屑分析等方法,需要丰富的专家经验才能判断是否发生了故障。分析方法包括经验模态分解(empirical mode decomposition, EMD)<sup>[2]</sup>, 变分模态分解(variational mode decomposition, VMD)<sup>[3]</sup>, 小波变换(wavelet transform,

WT)<sup>[4]</sup>等。随着机器学习的兴起,使用该方法的浅层学习技术进行轴承故障诊断成为主流,其流程可大体分为三步:首先是对提取到的轴承振动信号进行预处理,以突出信号的特征信息;其次选取合适的模型及方法进行训练及优化;最后将采集到的数据进行诊断。常用的诊断方法主要有支持向量机(support vector machine, SVM)<sup>[5]</sup>, 随机森林(random forest)<sup>[6]</sup>, K-最近邻算法(K-nearest neighbor, KNN)<sup>[7]</sup>等。但是,这种浅层学习的准确度受到信号的预处理的方法的影响十分严重<sup>[8-9]</sup>。

近年来,使用深度学习(deep learning)进行故障诊断的方法逐渐兴起。相较于传统的机器学习,深度学习在于随着数据规模的增加其性能也不断增长,并能从原始数据中自动提取特征信息,避免了繁琐的人工进行特征提取的操

收稿日期:2021-01-04

\* 基金项目:国家自然科学基金(61263023, 61863016)项目资助

作。常用的方法主要有以下几种:自动编码器(AutoEncoder),该方法是一种无监督学习算法,由编码器与解码器组成,可自动学习到输入数据的隐含特征,从而便于进行故障诊断;深度置信网络(deep belief network, DBN),由多个受限玻尔兹曼机(restricted boltzmann machines, RBM)堆叠而成,是一种概率生成模型,生成模型是建立一个观察数据和标签之间的联合分布,常用于预训练自动编码器的权值,使之达到更好的效果;深度残差收缩网络(deep residual shrinkage network, DRSN)<sup>[10]</sup>,为深度残差网络(deep residual network, ResNet)的一种改进,通过注意力机制发现样本中不重要的噪声信息,并通过软阈值函数置零,在故障诊断方面有较高的准确率;深度卷积网络(deep convolutional neural networks, DCNN),常用于计算机视觉领域,特点是可以通过深层次的网络结构提取更多的特征,且计算量较小,因此在故障诊断领域使用较多。曲建岭等<sup>[11]</sup>通过使用一维卷积神经网络,在不依赖信号处理技术与专家经验的情况下获得了较高的故障诊断准确率。赵小强等<sup>[12]</sup>使用改进的 AlexNet 在变负荷、强噪声条件下获得了更高的精度与更强的鲁棒性。汤芳等<sup>[13]</sup>提出了一种基于稀疏自编码的深度神经网络,使用其训练得到的参数和轴承振动信号频谱的频谱训练深度神经网络比 BP 神经网络有更好的效果。

本文根据 Inception 的网络结构提出一种卷积残差神经网络。该方法使用“端到端”的训练方法,即不使用复杂的数学方法进行信号特征的提取,直接使用原始的振动信号进行轴承的故障诊断。相较于其他的深度神经网络,减少了训练模型时的资源占用,降低了对高性能设备的依赖,提升了训练速度与检测精度,并解决了因网络加深所产生的退化问题。

## 1 相关理论简介

### 1.1 卷积层

卷积层是卷积神经网络的重要组成部分,通过反向传播算法进行参数学习,由多组平行的卷积核(kernel)构成。计算方式为卷积核在输入数据上按一定距离滑动,每滑动到一个位置上就会进行一次元素对应乘积并求和的运算,得到特征图(feature map)上的一个元素。卷积层的计算公式<sup>[14]</sup>为:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{i,j}^l + b_j^l\right) \quad (1)$$

式中: $k$ 为卷积核; $b$ 为偏置项; $l$ 为神经网络层数; $i$ 为输入特征图层数; $j$ 为输出特征图层数; $*$ 为卷积计算; $M$ 为输入特征图的集合; $f$ 为激活函数。

在卷积层中引入激活函数目的是为了加入非线性因素,使神经网络可以进行非线性拟合。在现代神经网络中,常用的激活函数为整流线性单元(rectified linear unit, ReLU),其公式为:

$$f(x) = \max(0, x) \quad (2)$$

相较于传统的 Sigmoid 函数,减少了计算量,并解决了其因层数加深所产生的梯度弥散问题。

### 1.2 池化层

池化层的作用是对输入数据的特征图进行下采样,降低图片的维度,以达到降低计算量,降低过拟合的目的。常用的池化方法为平均池化与最大池化,通过计算图像区域的平均值或最大值为输出结果。具体公式为(最大池化):

$$x_{ij}^l = \max(x_{im}^{l-1}), i-1 \leq n \leq i+1 \quad (3)$$

式中: $l$ 为网络层数; $i$ 为通道数; $n$ 为参数序号。

### 1.3 批标准化

批标准化(batch normalization)是一种规范化方法,通过将神经网络中神经元的输出值的分布调整和缩放到均值为 0 方差为 1 的标准正态分布当中,来改善神经网络的性能<sup>[15]</sup>,可以使输出值远离大部分激活函数的饱和区域,提高模型的鲁棒性,防止梯度爆炸。数学公式为:

$$\mu_\beta = \frac{1}{m} \sum_{i=1}^m x_i \quad (4)$$

$$\sigma_\beta^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_\beta)^2 \quad (5)$$

$$\hat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \quad (6)$$

$$y_i = \gamma \hat{x}_i + \beta \quad (7)$$

式中: $m$ 为每批次的大小; $\gamma, \beta$ 为可被学习的参数; $\epsilon$ 为增加数值稳定性的值。

### 1.4 残差块

当深度神经网络的层数增加时,会发生网络退化现象(degradation),即随着网络层数的增多,损失值会逐渐下降,到达一定层数后趋于饱和,如果网络深度继续加深,损失值反而会增大。为了解决这一问题,He 等<sup>[16]</sup>提出了残差块(residual block)结构。其核心是添加恒等映射(identity mapping),把网络底层的特征传递到高层。如图 1 所示,残差块分为恒等连接部分与残差部分,输出为两部分结果相加。若输入输出的通道数不同则在恒等映射层需要使用  $1 \times 1$  卷积改变维度。数学公式可表示为:

$$x_l = x_{l-1} + f(x_{l-1}) \quad (8)$$

式中: $x$ 为输入; $l$ 为神经网络层数; $f$ 为残差操作。

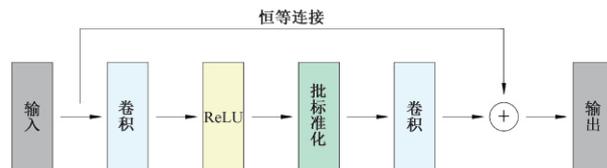


图 1 残差块

### 1.5 Inception 结构

Inception 结构是 GoogleLeNet<sup>[17]</sup>的核心组成单元,其结构由多种不同大小的卷积层和池化层组成,使之含有不

同大小的感受野,最后在通道上组合,使不同尺度的特征融合。Inception v1 结构如图 2 所示。由于使用大的卷积核会有较大的计算量,故采用  $1 \times 1$  卷积进行降维,以加快训练速度。从 Inception v2 模块开始引入了非对称卷积,并将大尺度卷积拆分成多个小尺度卷积,以减少计算量与计算参数。但是,当网络的深度加深时,网络不易训练,甚至无法收敛。

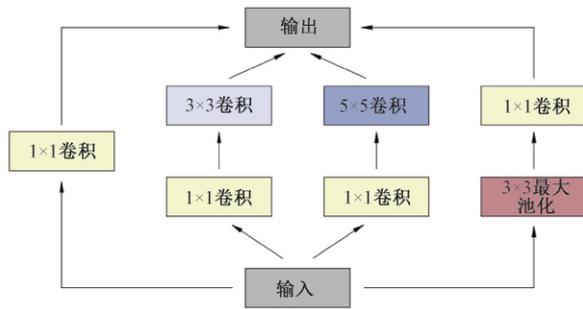


图 2 Inception v1 模块

## 2 基于 Inception-ResNet 模型的轴承故障分类方法

### 2.1 模型整体结构

本模型整体结构如图 3 所示。模型将一维振动信号转换为二维图片数据作为输入,然后使用卷积层提升通道数,提取初始的信号特征。在并行结构中使用多个小卷积核来

代替大卷积核来减少参数和计算量,从第一个并行结构之后,将卷积操作变为非对称卷积。常见的卷积核大小一般为  $N \times N$ ,而非对称卷积则是将  $N \times N$  大小的卷积核拆分为  $1 \times N$  和  $N \times 1$  大小的两个卷积核,由 Szegedy 等<sup>[18]</sup>提出。非对称卷积是把  $N \times N$  卷积计算变为了  $2 \times N$  卷积计算,在损失少量精度的前提下,降低了计算量与参数量,并且  $N \times N$  卷积与  $1 \times N$  和  $N \times 1$  卷积的结果等价。非线性激活层变为了两个,相比单一的非线性激活层网络的判别能力增强了不少。在 Inception 模块当中,通过使用不同大小的卷积核的并行计算来提取不同尺度大小的信息,由于使用的是通道连接,不同特征图之间没有通信,这样降低了网络的信息提取能力。为解决这一问题,在该模块中增加了一个三维卷积,将原始多通道二维数据变为单通道的三维数据,使通道数变为三维数据中的第三维度,三维卷积不仅可以感受原始二维数据的特征,并且可以在各种不同的特征图之间的信息相互交融。由于层数的加深会导致网络的退化问题,在通道连接后加入了残差链接,使网络有更强的拟合能力。直接将网络中的二维数据变为一维数据会使全连接层的参数过大,网络会占用过多的资源,不利于训练,在分类模块中使用全局平均池化(global average pooling, GAP)会改善这一问题。全局平均池化是对数据中的每一个特征图取平均值,特征图的数量即得到结果的参数量。在全局平均池化后加入了 Dropout 层,随机丢弃一定比例的信息,学习到具有更强的鲁棒性的特征。最后使用全连接层进行故障分类。

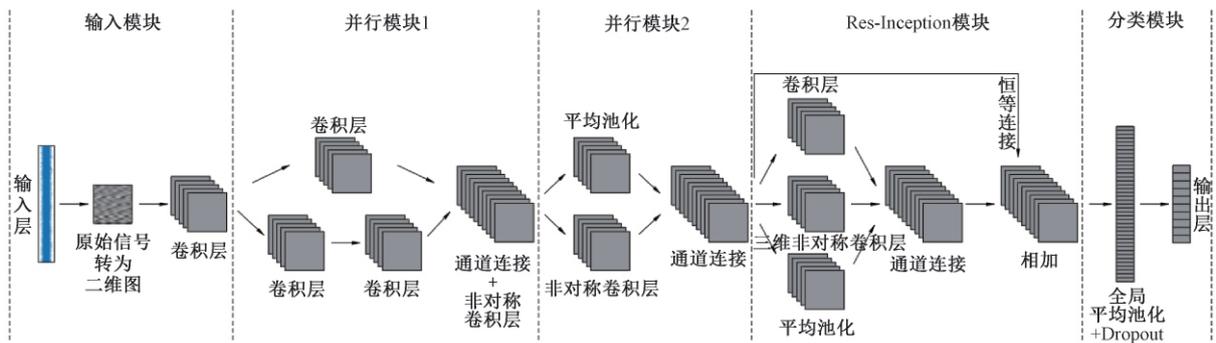


图 3 模型整体结构

### 2.2 模型训练

本模型使用的损失函数为交叉熵损失函数(cross entropy loss function)。该函数由 LogSoftmax 函数和负对数似然损失函数(negative log likelihood loss function)构成。假设神经网络输出为  $y_p$ ,真实值为  $y_t$ ,则该函数为:

$$L(y_t) = -\log y_t \quad (9)$$

$$L(y_p) = -\log y_p \quad (10)$$

$$\text{loss} = -\sum (L(y_t) \cdot \lg L(y_p) + (1 - L(y_t)) \cdot \lg(1 - L(y_p))) \quad (11)$$

为加快模型训练时的收敛速度,采用 Adam 优化算

法<sup>[19]</sup>对本模型进行优化。该算法使用动量和自适应学习率来加快收敛速度,并且计算效率高,对内存需求小,可降低对设备的性能需求。在训练时使用了小批量样本集(mini-batch),Masters 等<sup>[20]</sup>的研究表明较小的批大小  $\leq 32$  时神经网络的性能会维持在较高水平,批大小过大会导致性能下降。

## 3 实验数据及分析

### 3.1 凯斯西储大学滚动轴承数据

此数据集来源于凯斯西储大学(Case Western Reserve

University, CWRU)滚动轴承数据中心<sup>[21]</sup>,是最常用的轴承故障诊断数据集,具有多种故障类别。图 4 所示为 CWRU 数据集所使用的实验台实物图。实验台使用 2 马力的电动机,驱动端轴承型号为 SKF6205,风扇端轴承型号为 SKF6203。模拟损伤为电火花加工单点损伤,损伤位置分别为外圈、内圈和滚动体,每个位置的损伤直径有 0.007,0.014 和 0.021 inch 三种,共计 9 种故障。

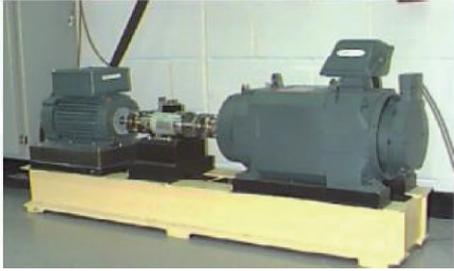


图 4 CWRU 轴承故障模拟实验台

本次实验选择位置为驱动端轴承,采样频率为 48 kHz,电机转速为 1 730 r/min 的故障数据。每条样本数据的采样长度为 5 184,转换为 72×72 大小的单通道二维图像,每种工作状态各构造 600 个样本,通过分层抽样的方法划分训练集与测试集。各故障样本数据如表 1 所示。

表 1 CWRU 数据集故障样本信息

标签编号	损伤直径/inch	损伤位置	训练集样本数	测试集样本数
1		滚动体	480	120
2	0.007	内圈	480	120
3		外圈	480	120
4		滚动体	480	120
5	0.014	内圈	480	120
6		外圈	480	120
7		滚动体	480	120
8	0.021	内圈	480	120
9		外圈	480	120
0	无	正常	480	120

1)模型参数的选取

参数的选取对模型的训练有很大影响,参数选择不当可能会导致模型训练速度过慢,过拟合,不收敛等现象,从而影响模型精度。

从图 5 可以看出,在较大的学习率下,模型收敛速度较快,训练批次在 5 轮以内即可达到 90% 以上的准确率,但是随着训练批次的增加,测试集准确率会出现大幅的波动,准确率有时会跌至 80% 以下,模型性能极不稳定,在 60~70 轮之后,准确率逐渐维持在 99.5% 以上。在较低的学习率下,模型收敛较慢,准确率波动较小,但是过低的学习率使模型的训练时间过长。基于以上结果,在训练前期

使用较高的学习率进行训练,使模型快速收敛达到较高的准确率,之后使用较低学习率,使模型平稳收敛。

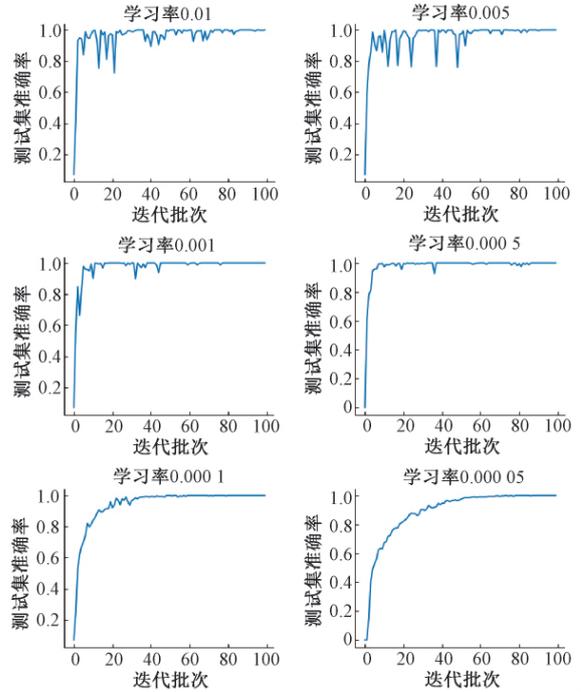


图 5 不同学习率的情况下准确率变化趋势

模型中各层权重值的初始化方式,对于网络的训练很重要,较差的权重初始化方法会导致梯度传播问题。Kaiming 初始化<sup>[22]</sup>是一种根据 ReLU 函数所设计的一种权重初始化方法,如图 6 所示,与传统的均匀初始化方法相比,网络训练效率有了极大地提升。

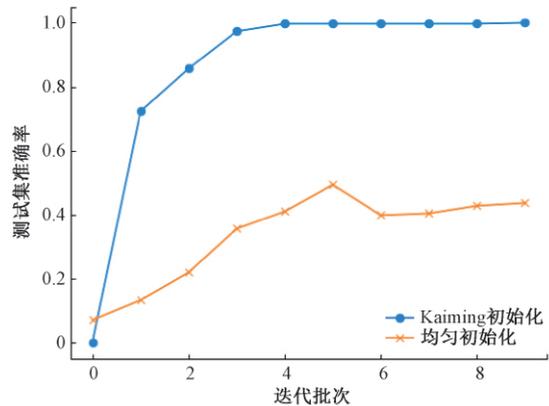


图 6 不同权重初始化方法下准确率变化趋势

如图 7 所示,在小批次的情况下,由于每轮训练次数的增加,模型训练速度较慢,但模型训练完成使用的轮次较少,准确率几乎不会波动。在较大的批次下训练,模型收敛速度较慢且准确率波动十分明显。相比之下,小批次的训练方法优势更加明显。

根据以上对比试验与反复调试,最终得到的超参数设置如表 2 所示。

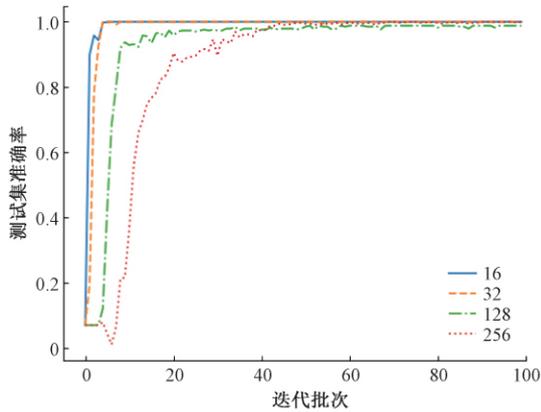


图 7 不同批大小的情况下准确率变化趋势

表 2 模型超参数取值

初始学习率	学习率衰减倍率	批大小	学习轮次
0.003	0.17	32	20

对于模型结构的参数的选择,改模型的目标是尽量使用较小的资源占用来达到较高的模型性能。对于初始部分,使用非对称卷积可以大幅减少参数与计算量,但是非对称卷积会在低层丢失大量初始特征,经反复实验较大的卷积核对本模型的性能提升无明显效果并且会使资源占用增加,所以选取  $3 \times 3$  的卷积核,在使特征图大小不变的同时提升维度,获取尽可能多的特征。在中部则选取较大的卷积核来减小特征图大小,减小模型的高层部分的计算量。Dropout 层的保留值设置为 0.5,以提高模型对噪音的鲁棒性。具体结构参数如表 3 所示。

表 3 模型结构参数

层数	类型及名称	输出大小	参数	参数量
0	输入层	1,72,72		
1	卷积层 (并行模块 1)	8,72,72	$k=(3 \times 3), s=1, p=1$	96
2	卷积层-支线 1	8,72,72	$k=(3 \times 3), s=1, p=1$	600
	卷积层-支线 2	8,72,72	$k=(3 \times 3), s=1, p=1$	600
	通道合并	16,72,72		
3	非对称卷积层 (并行模块 2)	16,70,70	$k=(3 \times 3), s=1, p=0$ $k=(3 \times 3), s=1, p=0$	1 600
4	池化层-支线 1	12,34,34	$k=(3 \times 3), s=2$	192
	非对称卷积层-支线 2	12,34,34	$k=(5 \times 1), s=(2,1), p=(1,0)$ $k=(1 \times 5), s=(1,5), p=(0,1)$	1 656
	通道合并 (Res-Inception 模块)			
5	卷积层-支线 1	8,34,34	$k=(1 \times 1), s=1, p=0$	192
	三维非对称卷积层-支线 2	1,8,34,34	$k=(3 \times 1 \times 1), s=1, p=(1,0,0)$ 8,34,34	16
	池化层-支线 3	8,34,34	$k=(1 \times 3 \times 1), s=1, p=(0,1,0)$ $k=(1 \times 1 \times 3), s=1, p=(0,0,1)$	192
	通道合并+恒等连接			
6	Dropout 层	24,34,34	$keep=0.5$	0
7	全局平均池化	24		0
8	全连接层	16	$in=24, out=16$	448

2)对比实验

为验证本模型的性能与优势,本文选用当前主流的 SVM、DBN+BPNN 方法与近年来热门的几种神经网络进行对比分析。由于神经网络的权重初始化的值是随机生成的,为避免随机性和偶然性的影响,在进行对比试

验时,对每个方法进行 10 次实验,并记录每个方法的稳定性。选用试验方法的数据处理方法与参数设置如下。

(1)SVM

由于 SVM 为浅层学习方法,使用 VMD+样本熵来提取数据特征。核函数使用径向基函数(radial basis

function, RBF), 并使用粒子群算法对惩罚系数  $C$ 、核函数系数  $\gamma$  进行优化, 种群数量 100, 迭代次数 300, 惯性因子 0.8, 加速常数均取 2。

(2) DBN+BPNN

数据处理方法采用 EMD, 数据采样数量改为 2 500。训练方法为使用 DBN 对 BPNN 的权值进行预训练, 网络结构设置为 2 500-1 536-768-384-96-10。DBN 由高斯-伯努利受限玻尔兹曼机构成, 学习率 0.001, 批大小 32, 动量值 0.95, 每层迭代次数 100。BPNN 参数为: 学习率 0.003, 批大小 32, 迭代次数 30。

(3) AlexNet<sup>[23]</sup>

由于本文训练数据与 AlexNet 的原训练数据大小不同, 将输入卷积层参数调整为  $5 \times 5$  卷积核, 1 步长, 2 填充, 并把原

模型中使用的局部响应归一化(local response normalization, LRN)替换为效果更加优秀的批标准化。不对数据进行特征提取。学习率 0.000 1, 批大小 32, 迭代次数 60。

(4) ResNeXt<sup>[24]</sup>

网络结构选用 ResNeXt-14( $32 \times 4d$ )。不对数据进行特征提取。学习率 0.000 05, 批大小 32, 迭代次数 35。

(5) DenseNet<sup>[25]</sup>

网络结构选用 DenseNet-121( $k=32$ )。不对数据进行特征提取。学习率 0.000 05, 批大小 32, 迭代次数 60。

以上选取的神经网络均使用 Adam 优化算法和交叉熵损失函数。实验环境使用 Debian 10 操作系统, CPU 型号为 AMD Ryzen 7 3700X, GPU 型号为 NVIDIA GeForce RTX2060。实验对比结果如表 4 所示。

表 4 各方法的性能数据

方法	参数量	前向/反向传播 内存消耗量/MB	最高准确度/ %	最低准确度/ %	平均训练耗时/s	
					GPU	CPU
SVM			99.38	96.88	251.50	1 532.52
DBN+BPNN	5 355 760	20.92	99.88	97.84	(DBN)	(DBN)
		(仅 BPNN)			26.13	134.84
AlexNet	28 714 512	372.80	100	99.40	183.68	2 939.22
ResNeXt	7 395 600	883.18	100	100	210.20	3 497.90
DenseNet	5 063 440	600.99	100	100	1 237.14	5 623.08
本文方法	6 476	183.30	100	100	78.64	318.84

通过表 4 可以得知, 选用的全部方法都具有很高的分类准确率。在对数据使用了特征提取后, SVM 与 DBN+BPNN 这两种浅层学习的方法也取得了优秀的性能, 但是由于每次实验测试集的划分是随机选取的, 导致准确率的稳定性较于神经网络有明显的差距。对比其他的实验方法, 本文方法不仅有最高的准确率, 在资源占用, 训练时间上都有明显的优势。虽然上节实验表明增加批大小会加快神经网络的训练速度, 但是会降低其稳定性, 本文方法在使用较小的批大小后仍具有较快的训练速度。在使用 CPU 进行训练时, 其他深度学习方法训练时间为数十分钟, 本文方法仅需要不足 6 min 的时间即可完成训练。该结果表明, 本模型不需要高性能 GPU 也可以在较短的时间内完成训练, 并有极高的准确度, 可节省大量的模型训练成本。

由图 8 可知, DenseNet 由于网络深度较大, 该模型在训练的前 25 轮准确度出现了很大的波动, 训练速度较慢, AlexNet 和 ResNeXt 在 25 轮之后趋于稳定在最高准确度附近, 本文方法在 10 轮以内达到了最高的准确度。此实验结果表明, 本文方法比另外 3 种深度学习方法的训练速度更快, 稳定性更强。

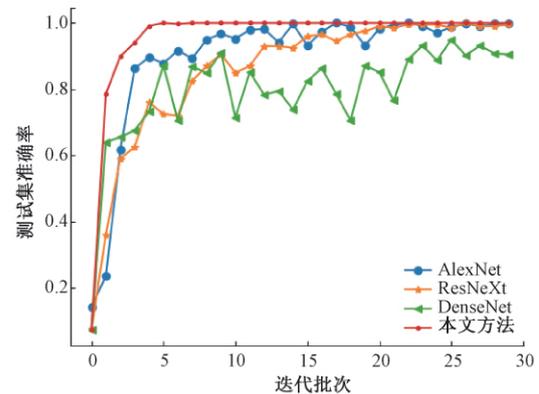


图 8 不同深度学习方法的稳定性对比

为了直观地了解本文方法进行特征学习的能力, 使用流形学习方法 t-分布随机邻域嵌入 (t-distributed stochastic neighbor embedding, t-SNE) 对输入模块和 Res-Inception 模块的输出值进行降维, 并进行了可视化表达。从图 9 可以清晰地了解到, 数据在刚进入网络时各种类别分布混乱, 无法分辨。而经过 Res-Inception 模块之后, 输出值的各种类别在位置上被进行了划分, 同类数据聚集

在了一起,不同类别的数据在位置上有了明显的间隔。

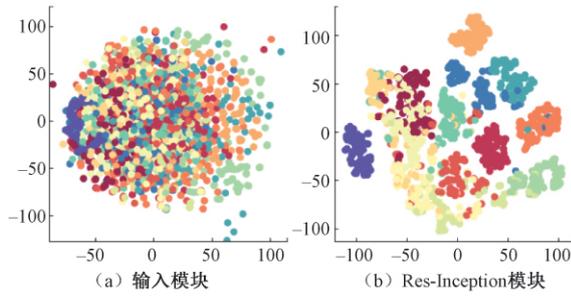


图 9 CWRU 数据特征可视化

### 3.2 IMS 轴承数据集

为了继续验证该模型的有效性,使用 IMS 轴承数据集<sup>[26]</sup>再次进行验证。该数据集是辛辛那提大学智能维护系统中心(center for intelligent maintenance systems, IMS)提供的轴承的全生命周期数据集。实验台如图 10 所示。该实验台是将 4 个型号为 Rexnord ZA-2115 的轴承安装在同一个轴上,使用转速为 2 000 r/min 的交流电机驱动,以 20 kHz 的频率进行采样,每次采样 20 480 个数据点,每个数据点包含轴向与垂直的两个方向的振动信息。

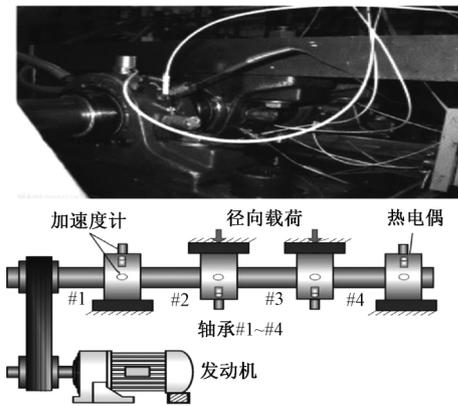


图 10 IMS 数据集实验台实物与结构

IMS 数据集根据不同情况共记录了 3 组数据,本文以第 1 组实验数据来构建数据集。在该组数据中,共记录了 2 156 次实验数据,其中,轴承 1 和轴承 2 正常,轴承 3 为内圈故障,轴承 4 为滚动体故障,共 3 种数据类别。在每个所采集样本中,轴向与垂直振动信号各采集 576 个,转换为双通道  $24 \times 24$  的二维图像,每种数据采集 2 400 次,并使用分层抽样划分测试集与训练集。样本数据如表 5 所示。

表 5 IMS 数据集故障样本信息

标签编号	损伤位置	训练集样本数	测试集样本数
0	正常	2 000	400
1	内圈	2 000	400
2	滚动体	2 000	400

在该数据集的对比试验中,所有模型将输出类别更改为 3 类,DBN+BPNN 结构改为 1 152-1 024-768-384-96-3,其余参数设置和数据处理方式与 CWRU 数据集部分相同。每个方法进行 10 次实验,结果如表 6 所示。

表 6 IMS 数据集中各方法的性能数据

方法	最高 准确度/%	最低 准确度/%	平均训练耗时/s	
			GPU	CPU
SVM	100	99.37	281.36	
			250.54	1 374.50
DBN+BPNN	100	99.70	(DBN) 21.70	(DBN) 79.29
			(BPNN)	(BPNN)
AlexNet	100	100	163.69	2 536.14
ResNeXt	100	100	183.43	1 304.20
DenseNet	100	100	1 176.66	5 400.96
本文方法	100	100	56.25	112.63

表 6 的数据表明,所有的方法都可以在本数据集上获得很好的准确度,原因可能是 IMS 数据集的故障分类较少,并且信号的故障特征比 CWRU 数据集更加明显。SVM 与 DBN+BPNN 两种方法仍然会因为随机划分测试集导致的准确度波动,这继续表明了深度学习相较于浅层学习具有更高的稳定性。由于此数据集的特征数量比 CWRU 数据集小,所有方法的训练时间都有所减少,本文方法在训练效率上仍然好于其他方法,具有最低的训练时间。

## 4 结 论

本文提出了一种新型的卷积残差神经网络用于处理轴承的故障诊断方面的问题,它是基于 Inception 结构搭建而成的,并通过与多种诊断方法进行对比试验,得出如下结论:

- 1) 本模型使用原始的故障信号直接转换为二维图片进行故障诊断,避免了人工进行故障特征提取的过程,提升了本模型的使用效率。
- 2) 与传统的浅层学习方法相比,本模型具有更好的稳定性,避免了因随机初始化权值等因素对诊断准确率的影响。
- 3) 由两种轴承数据集的对比实验可知,本模型不仅可以一直保持最高的准确率,而且在资源占用和训练时间上也远低于其他的深度学习方法,极大地降低了对高性能设备的依赖。

### 参考文献

[1] GAN M, WANG C, ZHU C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of

- rolling element bearings[J]. *Mechanical Systems & Signal Processing*, 2016, 72-73(5):92-104.
- [2] KOSTOPOULOS S V. Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition[J]. *Mechanical Systems and Signal Processing*, 2013, DOI: 10.1016/j.ymssp.2013.02.020.
- [3] 许爱华, 闫俊泉, 伍旭灿, 等. 基于 VMD 和 MP 算法的旋转机械故障特征提取[J]. *国外电子测量技术*, 2017, 36(8):11-17.
- [4] 刘军, 刘诚诚, 王洋洋. 基于小波变换的输电线路保护研究[J]. *电子测量技术*, 2018, 41(6):103-108.
- [5] 郑国强, 汪玉, 高博, 等. 基于 SOLPP 维数约简的电机轴承故障特征提取[J]. *电子测量与仪器学报*, 2018, 32(6):186-192.
- [6] HAN T, JIANG D. Rolling bearing fault diagnostic method based on VMD-AR model and random forest classifier[J]. *Shock and Vibration*, 2016, 2016(6):1-11.
- [7] WANG Q, LIU Y B, HE X, et al. Fault diagnosis of bearing based on KPCA and KNN method[J]. *Advanced Materials Research*, 2014, 986-987:1491-1496.
- [8] SHAO H D, JIANG H K, ZHAO H W, et al. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis[J]. *Mechanical Systems and Signal Processing*, 2017, 95:187-204.
- [9] SHAO H D, JIANG H K, ZHAO H W, et al. An enhancement deep feature fusion method for rotating machinery fault diagnosis [J]. *Knowledge-Based Systems*, 2017, 119:200-220.
- [10] ZHAO M, ZHONG S, FU X, et al. Deep residual shrinkage networks for fault diagnosis [J]. *IEEE Transactions on Industrial Informatics*, 2020 (99):1-1.
- [11] 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.
- [12] 赵小强, 张青青, 陈鹏, 等. 基于 PSO-BFA 和改进 Alexnet 的滚动轴承故障诊断方法[J]. *振动与冲击*, 2020, 39(7):21-28.
- [13] 汤芳, 刘义伦, 龙慧. 稀疏自编码深度神经网络及其在滚动轴承故障诊断中的应用[J]. *机械科学与技术*, 2018, 37(3):352-357.
- [14] BOUVRIE J. Notes on convolutional neural networks[R]. Massachusetts: Center for Biological and Computational Learning, 2006, DOI:10.1145/238061.238084.
- [15] LI Y, WANG N, SHI J, et al. Revisiting batch normalization for practical domain adaptation [J]. *Pattern Recognition*, 2016, DOI: 10.1016/j.patcog.2018.03.005.
- [16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [J]. *IEEE Computer Society*, 2016, DOI: 10.1109/CVPR.2016.90.
- [17] LIU P, LIU X, LIU M, et al. Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network [J]. *Remote Sensing*, 2019, 11(7), DOI:10.3390/rs11070830.
- [18] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [J]. *IEEE Computer Society*, 2015, DOI: 10.1109/CVPR.2016.308.
- [19] CHEN Y, WEN Z, TANG K, et al. SSE composite index forecasting model via BP neural network with ADAM optimizer [J]. *International Journal of Computer Applications Technology and Research*, 2020, 9(1):8-14.
- [20] MASTERS D, LUSCHI C. Revisiting small batch training for deep neural networks [J]. *Computer Science*, 2018, 20(6):1-18.
- [21] The Case Western Reserve University Bearing Data Center. Bearing data center fault test data[EB/OL]. [2020-11-15]. <http://cseyroups.case.edu/bearingdatacenter/pages/download-data-file>.
- [22] HE K M, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imageNet classification[J]. *IEEE Computer Society*, 2015, DOI: 10.1109/ICCV.2015.123.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6):84-90.
- [24] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks[J]. *IEEE Computer Society*, 2016, DOI: 10.1109/CVPR.2017.634.
- [25] 徐文学, 夏懿, 张德祥. 基于稠密卷积网络的轴承故障诊断[J]. *噪声与振动控制*, 2020, 40(4):80-86.
- [26] QIU H, LEE J, LIN J, et al. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics [J]. *Journal of Sound & Vibration*, 2006, 289(4-5):1066-1090.

#### 作者简介

孔子宇, 硕士研究生, 主要研究方向为故障诊断。

E-mail: 844514004@qq.com

王海瑞, 硕士生导师, 教授, 主要研究方向为计算机应用基础和工程技术。

E-mail: hrwang88@163.com