

doi: 10.7690/bgzdh.2026.02.007

基于注意力机制和轻量级自适应 CNN 模型的滚动轴承故障诊断

汤家辉¹, 孙 宁^{1,2}, 王松雷³(1. 南京信息工程大学自动化学院, 南京 210044; 2. 无锡学院自动化学院, 江苏 无锡 214105;
3. 江苏省特种设备安全监督检验研究院无锡分院, 江苏 无锡 214174)

摘要: 为解决滚动轴承故障在实际复杂环境中的诊断需要精准性、鲁棒性和泛化性等全面的性能, 提出一种融合注意力机制的轻量级自适应 CNN 网络(1D-LECA-Inception)。通过 1 维的深度可分离卷积重构 Inception 模块并拓宽卷积核的尺度, 由有效通道注意力(efficient channel attention, ECA)模块筛选出不重要的信息, 融入了残差结构、批量归一化层(batch normalization, BN)以及自适应激活函数 AdaptH_Swish 来提升整体网络模型的稳定性和泛化能力, 并通过帕德博恩和凯斯西储轴承数据集与其他分类模型进行对比试验。结果表明: 不论是同负荷、变负荷还是噪声干扰条件下, 该方法在与其他分类模型的对比中综合表现更优。

关键词: 故障诊断; 轻量化; 深度可分离卷积; 自适应激活函数

中图分类号: TP183 **文献标志码:** A

Rolling Bearing Fault Diagnosis Based on Attention Mechanism and Lightweight Adaptive CNN Model

Tang Jiahui¹, Sun Ning^{1,2}, Wang Songlei³(1. *Nanjing University of Information Science and Technology, Nanjing 210044, China*; 2. *Wuxi University, Wuxi 214105, China*;
3. *Wuxi Branch Institute, Special Equipment Safety Supervision and Inspection Research Institute, Wuxi 214174, China*)

Abstract: In order to solve the problem that the diagnosis of rolling bearing fault in the actual complex environment requires comprehensive performance such as accuracy, robustness and generalization, a lightweight adaptive CNN network (1D-LECA-Inception) is proposed, which integrates the attention mechanism. The Inception module is reconstructed by 1-D depth-separable convolution and the scale of the convolution kernel is broadened, and the unimportant information is screened out by the efficient channel attention (ECA) module. The residual structure, batch normalization (BN) and adaptive activation function AdaptH_Swish are integrated to improve the stability and generalization ability of the overall network model, and the Paderborn and Case Western Reserve bearing data sets are used to compare with other classification models. The results show that the method performs better than other classification models under the same load, variable load and noise interference conditions.

Keywords: fault diagnosis; lightweight; depth-separable convolution; adaptive activation function

0 引言

近年来, 基于深度学习的轴承故障方法, 如深度置信网络、深度自编码器和卷积神经网络等已成为研究的热点^[1]。与信号分析方法相比, 基于数据驱动的方法不需要依赖专业的信号处理知识, 避免了信号分析方法过程中可能造成的信息缺失。李益兵等^[2]提出将粒子群优化算法应用于改进的深度置信网络, 在降低故障检测时间成本的同时获得了较高的故障识别能力。Xiong 等^[3]将深度自编码器与生成对抗网络结合, 通过在生成对抗网络中加入梯度惩罚项, 改善了生成网络在训练过程中不稳定甚至无法收敛的问题, 从而实现了滚动轴承的智能故

障诊断。Zou 等^[4]提出了一种将信号预处理技术与 EMD 方法相结合的 EMD-LSTM 轴承故障诊断系统, 利用 RNN 层中的长短期记忆(long short term memory, LSTM)单元来区分故障。贺思艳等^[5]提出了一种增强的深度卷积神经网络, 利用格拉姆角场将振动信号重构为灰度图像, 然后采用卷积网络进行训练分类。

上述方法多采用图像处理中的一些经验, 主要针对滚动轴承诊断的某方面进行了优化, 但是综合性能并不全面。笔者结合复杂环境下滚动轴承诊断的需求, 设计一种融合注意力机制的轻量级自适应 CNN 网络。

收稿日期: 2024-11-11; 修回日期: 2024-12-12

基金项目: 国家自然科学基金(42275156); 教育部产学合作协同育人项目、江苏高校“青蓝工程”资助课题(202102224006), 江苏省大学生创新创业项目(202213982017Z)

第一作者: 汤家辉(2000—), 男, 江苏人, 硕士。

1 模型构建

1.1 DSC-Inception 模块

笔者参考了 InceptionV4 网络的结构,并针对 1 维振动信号的特征识别任务进行了改进,提出了融合深度可分离卷积的 DSC-Inception 模块,它的主要结构由 2 个组件构成:

1) 池化操作,特征通道数为 C 的 1 维信号将经过最大池化层提取最显著的特征,以减轻由微小扰动所带来的影响。在池化后连接的是瓶颈层,目的是通过减小特征的维度来减少模型的参数量和计算量,并且减轻由数据集少而导致的过拟合问题。

2) 多尺度深度可分离卷积操作,1 维信号将送入瓶颈层减小通道数,瓶颈层重构后的特征图将输入并联的 3 组通道数为 16 的滤波器,其中的卷积核分别是 10×1 、 20×1 和 40×1 大小的 1 维深度可分离卷积。最终每个并行深度可分离卷积将在通道方向上与池化操作的输出融合,形成更加全面和高级的特征,从而更适应 1 维振动信号的特征提取任务。

1.2 自适应激活函数 AdaptH_Swish

在训练深层网络时,激活函数中起着至关重要的作用,因为它增强了网络的非线性表达能力,但是增加了激活函数后同样会给模型带来额外的计算复杂度,影响网络运算速度,如何选取高效且轻量的激活函数成了一个难题。HardSwish 激活函数(H_Swish)的出现首次出现在 MobileNetV3 中,给出了一个答案,因为其使用巧妙的结构,克服了 Swish 激活函数的计算代价太大的问题,而广受关注。

虽然与 Swish 相比,H_Swish 的计算速度更快,但是 H_Swish 在梯度上存在一定的突变,所以在某些情况下并不是最优的。针对该问题,笔者提出自适应的 HardSwish 激活函数(AdaptH_Swish),通过引入一个可训练的参数 ∂ (初始值设置为 1.0) 来进一步优化网络的性能。AdaptH_Swish 的数学表达式为:

$$\text{AdaptH}_{\text{swish}(x)} = \partial x \frac{\text{ReLU6}(\partial x + 3)}{6} \quad (1)$$

1.3 模型网络架构

结合以上基础理论以及优化的模块,笔者专门为滚动轴承的故障诊断设计了一种融合高效通道注意力机制的 1 维轻量级 Inception 网络(1D-LECA-Inception),可以直接使用采集振动信号作为输入,实现端到端式的故障识别。1D-LECA-

Inception 的整体架构如图 1 所示。

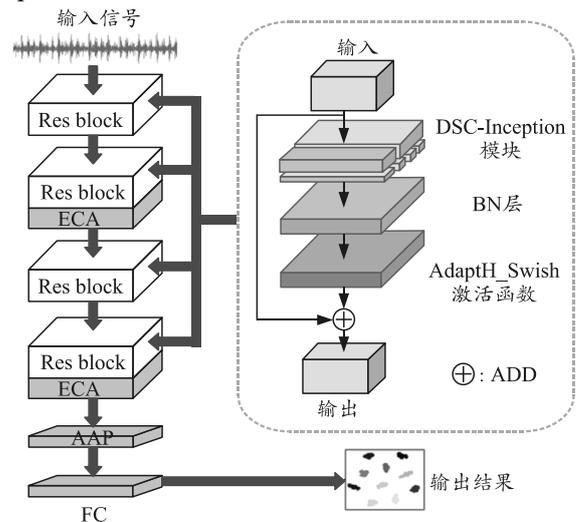


图 1 1D-LECA-Inception 网络整体架构

从图 1 中可以看出:Resblock 由 DSC-Inception 模块、批量归一化层(BN)和 AdaptH_Swish 激活函数组成,输入信号先会经过 DSC-Inception 来提取多尺度的特征,后经过 BN 和 AdaptH_Swish 来规范化特征分布和增强网络的非线性建模能力。为了进一步提升网络的鲁棒性,再进行残差融合。在每 2 个 Resblock 后插入一个 ECA 注意力模块的目的是:在提升网络有效特征提取能力的同时,尽量使用较少的参数,达到一个参数量和性能之间的平衡。此外,值得注意的是:由于网络的输入是 1 维振动序列,所以需要将 ECA 模块中的池化层和卷积层都修改为 1 维模式。AAP 为自适应平均池化,FC 为全连接层,ADD 为特征叠加。

1D-LECA-Inception 网络最终是通过 4 个 Resblock 和 2 个 ECA 所组成的主干网络来提取振动信号中的故障特征,随后将提取到的精细特征传入自适应平均池化层,再将池化压缩后的突出特征信息送入全连接层进行故障的诊断输出。

2 实验设置

2.1 实验数据

数据集的选取允许笔者全面评估模型在面对多样化故障情况时的性能。为了更深入探究模型在相同负荷下的故障诊断能力,选择帕德博恩大学轴承(paderborn university, PU)数据集^[6]。该数据集具有较高的分类难度。笔者选择在负载为转速 $n=1\ 500\ \text{r/min}$ 、负载扭矩 $0.7\ \text{N/m}$,轴承上的径向力 $1\ 000\ \text{N}$ 下获取的真实损伤振动数据进行分类评估。同时,这些数据根据损伤长度将故障严重程度分为

3 个级别, 最终数据一共分为 12 种故障状态和 1 种正常状态。

为了研究模型在不同负载下的鲁棒性和泛化性能, 采用凯斯西储大学 (case western reserve university, CWRU) 滚动轴承数据中心提供的数据集进行实验^[7]。在该数据集中, 所用的轴承型号为深沟球轴承 SKF6205, 振动信号的采集频率为 12 kHz。被诊断的轴承分为滚体损伤、外圈损伤、内圈损伤和正常状态 4 种状态。损伤的直径分别为 7、14 和 21 mils, 共计 9 种损伤状态和 1 种正常状态。

2.2 数据预处理

在深度学习中, 有效的数据预处理是避免模型由于训练集不足而产生过拟合的重要手段。笔者参考文献^[8]中采用重叠采样的手段进行数据预处理, 有助于避免由于分组采样导致的训练集不足问题。

试验时, 分别准备 4 个数据集, 如表 1 所示。数据集 A 来自 PU 数据集, 数据集 B、C 和 D 分别选自 CWRU 数据集中负载为 1、2、3 hp 下的滚动轴承故障数据。数据集的样本长度均设置为 1 024, 并按照 7:2:1 的比例将样本分为训练集、验证集和测试集, 其中训练样本采用重叠采样技术, 设置偏移量为 24。

表 1 试验数据集描述

数据集名	负载/hp	状态数量	训练集	验证集	测试集	总样本数
A	-	13	4 550	1 300	650	6 500
B	1	10	3 500	1 000	500	5 000
C	2	10	3 500	1 000	500	5 000

2.3 参数设置

为了让模型充分学习轴承振动信号中的故障特征, 将训练轮数 (Epoch) 设置为 100 次, 并且使用学习率 0.001 为 Adam 优化器来优化交叉熵损失函数。在前期的一些测试中发现, 128 的 Batch size 可以达到最高的准确率, 故将 Batch size 设为 128。试验中, 由于神经网络的初始权值是随机生成, 所以为了验证本方法的稳定性, 将取 10 次试验的平均值作为结果。

3 实验结果与分析

3.1 不同激活函数的性能对比分析

设计 5 组实验, 分别将使用 SiLU、Sigmoid、ReLU 和 H_Swish 激活函数替换网络中 AdaptH_Swish 激活函数, 以验证 1D-LECA-Inception 网络在使用不同激活函数下的故障识别性能, 使用数据

集 A 的检测实验对比结果如表 2 所示。

表 2 不同激活函数在数据集 A 上的性能对比

方法	Accuracy/%	Params/kB	FLOPs/M
SiLU	97.38	23.93	18.36
Sigmoid	95.69	23.93	18.36
ReLU	97.84	23.93	18.36
H_Swish	97.95	23.93	18.36
AdaptH_Swish	98.49	23.93	18.36

由表 2 可知: 采用 AdaptH_Swish 后, 准确率在使用 H_Swish 的基础上有了进一步的提升, 达到了 98.49%。为了更加直观地对比 AdaptH_Swish 激活函数性能的优越性, 图 2 展示模型分别使用 H_Swish 和 AdaptH_Swish 时的准确率和损失值曲线。

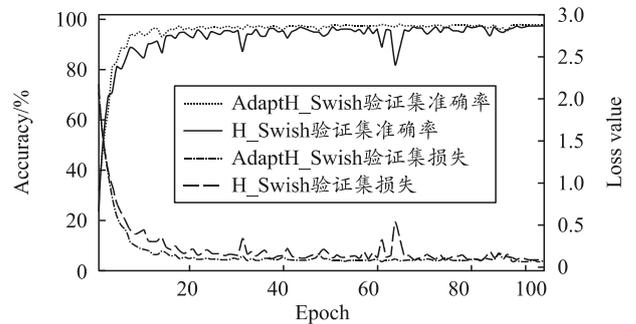


图 2 模型准确率和损失值曲线对比

从图 2 中可以看出: H_Swish 激活函数的训练效果并不稳定, 损失值在 10 轮训练后下降趋势并不明显, 同时产生了较大振荡, 鲁棒性不强; 反观本文中设计的自适应激活函数 AdaptH_Swish 的损失值在训练过程中没有出现明显不稳定点, 同时保持较好的下降趋势, 并最终趋于 0。

综合来看, AdaptH_Swish 激活函数可以更有效地提升网络模型的性能, 提升模型的训练速度和识别准确率, 在接下来的实验中都采用 AdaptH_Swish 作为 1D-LECA-Inception 网络中的激活函数。

3.2 同负载条件下的性能分析

为了充分评价本文中网络的性能, 选择不同 CNN 模型进行性能试验对比, 分别是 MobileNetV3-large、ShuffleNetV1^[9]、MCNN-LSTM 和 WDCNN。MobileNetV3-large 和 ShuffleNetV1 是 2 种被广泛应用的轻量级 CNN 模型; MCNN-LSTM^[10] 是一种多尺度 CNN 和 LSTM 所结合的轴承故障检测模型, 由 2 个不同尺度的 CNN 提取特征, 最终将特征融合输入 LSTM 进行故障分类; WDCNN^[11] 是 1 维 CNN 故障检测模型, 模型特征提取主干主要由一层大卷核和多层小卷积核组成, 并在 CWRU 数据集上

展现了很好的准确率。各模型在 A 数据集的结果如表 3 所示。

表 3 各模型在数据集 A 上的性能对比

方法	Accuracy/%	Params/kB	FLOPs/M
MobileNetV3-large	76.46	3 846.62	66.75
ShuffleNetV1	84.34	2 449.51	23.82
MCNN-LSTM	82.65	89.93	24.15
WDCNN	91.84	42.01	0.37
1D-LECA-Inception	98.49	23.93	18.36

从表 3 可知：本文中的 1D-LECA-Inception 模型在数据集 A 上的分类性能是最好的，10 次试验平均识别率达到 98.49%，对比属于轻量级 CNN 模型的 MobileNetV3-large 和 ShuffleNetV1，分别比其高 22.03 和 14.15 个百分点，取得了显著的优势。在模型参数量和复杂度方面，浮点计算量分别仅有其的 27.50% 和 77.07%，参数量则减少了至少 2 个数量级。与专门为滚动轴承故障检测而设计的 MCNN-LSTM 和 WDCNN 模型对比，1D-LECA-Inception 模型准确率分别超其 15.84 和 6.65 个百分点。参数

表 4 各模型在变负荷条件下的性能对比

方法	$B \rightarrow C$	$B \rightarrow D$	$C \rightarrow B$	$C \rightarrow D$	$D \rightarrow B$	$D \rightarrow C$	AVG
MobileNetV3-large	97.4	90.4	94.6	92.2	85.6	82.6	90.50
ShuffleNetV1	99.4	97.0	95.8	90.0	93.6	84.4	93.40
MCNN-LSTM	71.4	74.0	84.7	84.5	73.8	77.5	77.65
WDCNN	99.2	90.6	95.5	93.8	88.3	89.4	93.00
1D-LECA-Inception	100.0	98.2	97.8	99.5	96.7	98.2	98.40

从表 4 可知：MCNN-LSTM 模型在变工况场景下的识别稳定性并不是很好，模型的平均准确率低于 80%，这说明该方法过度提取了源域数据的分布，导致迁移至目标领域后的泛化性能严重下降。MobileNetV3-large、ShuffleNetV1 和 WDCNN 的性能稍好，分别能够实现 90.5%、93.4% 和 93.0% 的准确率，但是其稳定性并不好，准确率在一些迁移场景下产生了较大的波动，说明这些方法提取的特征的分布具有一定的偶然性，没有提取出最显著的特征。笔者提出 1D-LECA-Inception 模型的波动性最小，可以达到 94.5% 的平均识别率。

此外，1D-LECA-Inception 模型表现出对源域数据分布依赖性最小的特点，如在使用训练集 D 迁移至测试集 C 进行识别时，各对比模型分类准确率普遍下降。然而，1D-LECA-Inception 在这种情况下的准确率下降较小，且超过其他模型至少 9 个百分点。综合以上结果，1D-LECA-Inception 模型在变负荷条件下展现出较高的泛化性能和鲁棒性。

3.4 变噪声条件下的性能分析

在工业生产中，轴承故障诊断一直面临着噪音

量和浮点计算量分别为 MCNN-LSTM 的 26.60% 和 76.02%，虽然浮点计算量比 WDCNN 模型高，但是参数量减少了 43.03%。综上所述，笔者提出的 1D-LECA-Inception 相较于其他轻量化 CNN 模型，在同负荷的任务中展示出了更好的分类性能，且参数量和计算量也具有一定的优势。

3.3 变负荷条件下的性能分析

滚动轴承的振动特性会随工作负荷的变化而发生显著变化，这种变化导致传统分类器在提取特征时难以正确归类振动信号；因此，需要具有较高泛化性能的模型来应对。

笔者选用 CWRU 中 3 种不同负荷状态的 B 、 C 和 D 数据集进行模型训练，通过将训练集和验证集选取其中一种负荷，而测试集将使用其余 2 种负荷下的信号进行测试，以更全面地验证模型的鲁棒性和泛化能力。各模型在变负荷条件下的试验结果如表 4 所示。

干扰的挑战，进而影响了故障诊断的准确率。为了更真实地模拟工业实际情况，通过在原始振动信号中引入不同信噪比 (signal noise ratio, SNR) 的高斯白噪声进行实验。

在前期的实验中，各个模型在 CWRU 数据集的子集 B 上均能达到 100% 的识别率，所以为公平判断各个模型的抗噪性能，避免其他因素的干扰，选取子集 B 作为试验数据集，并在测试集中加入 SNR 值为 5~25 dB 的高斯白噪声。每种模型在变噪声环境下对比结果如图 3 所示。

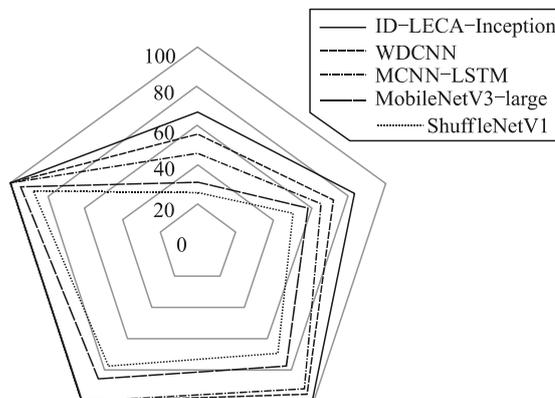


图 3 各模型的抗噪性能雷达

由图 3 可知: ShuffleNetV1 和 MobileNetV3-large 的抗噪性最弱, 在 SNR=5 dB 的噪声环境中准确率低于 40%, 可能是这 2 种模型大量采用小卷积结构而导致在提取特征信息时容易受到高频信号的干扰; MCNN_LSTM 的抗噪性能要比 ShuffleNetV1 和 MobileNetV3-large 强, 虽然在噪声强时识别率很差, 但是噪声逐渐减弱至 SNR=20 dB, 准确率识别已经趋于 100%; 得益于 WDCNN 的大卷积核与小卷积结合的设计可以有效筛选出中低频的噪声, 其在抗噪性能上的表现更优于 MCNN_LSTM。笔者提出的 1D-LECA-Inception 模型使用多尺度卷积层和自适应激活函数等多种技术的融合, 可以更好提取多样化的信息并使得特征的分布更加稳定, 避免随机干扰, 在不同信噪比下的故障诊断准确率均高于其他方法, 当 SNR \geq 15 dB 时的准确率就可以与诊断原始信号无异, 这也进一步说明本模型在噪声环境下也有很强的泛化和鲁棒性。

4 结论

笔者提出了一种新型轻量级深度学习模型 1D-LECA-Inception, 用于滚动轴承故障诊断。该模型直接使用 1 维原始振动信号, 适用于复杂环境下高效识别故障。主要结论如下:

- 1) 调整了 Inception 模块中的卷积大小与维度, 通过深度可分离卷积重构了 DSC-Inception 模块, 大幅减少了参数量和计算复杂度, 而不损失性能。
- 2) 设计了自适应的激活函数 AdaptH_Swish, 证明在本任务中优于原始的 H_Swish 和其他常见激活函数, 使网络训练更平稳高效。
- 3) 实验结果表明: 1D-LECA-Inception 即使在少样本训练条件下也能保持较高识别准确率, 并且稳定性优于对比网络模型。在 6 种可变负载场景下, 平均识别率达 98.4%, 并且在 SNR 超过 15 dB 时不

受噪声干扰, 证明其跨领域泛化性和鲁棒性优异。

参考文献:

- [1] HOANG D T, KANG H J. A survey on deep learning based bearing fault diagnosis[J]. Neurocomputing, 2019, 335: 327-335.
- [2] 李益兵, 王磊, 江丽. 基于 PSO 改进深度置信网络的滚动轴承故障诊断[J]. 振动与冲击, 2020, 39(5): 89-96.
- [3] XIONG X, JIANG H K, LI X, et al. A Wasserstein gradient-penalty generative adversarial network with deep auto-encoder for bearing intelligent fault diagnosis[J]. Measurement Science and Technology, 2020, 31(4): 045006.
- [4] ZOU P, HOU B, LEI J, et al. Bearing fault diagnosis method based on EEMD and LSTM[J]. Computers Communications & Control, 2020, 15(1): 1-14.
- [5] 贺思艳, 任利娟, 田新诚. 基于卷积神经网络的滚动轴承故障诊断[J]. 兵工自动化, 2019, 38(3): 39-41, 54.
- [6] LI F, WANG L, WANG D, et al. An adaptive multiscale fully convolutional network for bearing fault diagnosis under noisy environments[J]. Measurement, 2023, 216: 112993.
- [7] 全钰, 庞新宇, 魏子涵. 基于 GADF-CNN 的滚动轴承故障诊断方法[J]. 振动与冲击, 2021, 40(5): 247-253, 260.
- [8] 段洁利, 于世伟, 解明坤, 等. 基于一维轻量化 CNN 的山地索道轴承故障诊断[J]. 农业工程学报, 2023, 39(14): 70-79.
- [9] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//the IEEE conference on computer vision and pattern recognition. IEEE, 2018: 6848-6856.
- [10] CHEN X, ZHANG B, GAO D. Bearing fault diagnosis base on multi-scale CNN and LSTM model[J]. Intelligent Manufacturing, 2021, 32: 971-987.
- [11] ZHANG W, PENG G, LI C, 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.