

引用:王庆楠,陈倩,李涛,等. 基于轻量化孪生网络的小样本轴承故障诊断[J]. 聊城大学学报(自然科学版),2026,39(1):11-23.

Citation: WANG Qingnan, CHEN Qian, LI Tao, TU Jihui. Few-shot fault diagnosis for bearing based on lightweight siamese networks[J]. Journal of Liaocheng University (Natural Science Edition), 2026, 39(1): 11-23.

文章编号 1672-6634(2026)01-0011-13

DOI 10.19728/j.issn1672-6634.2025010011

基于轻量化孪生网络的小样本轴承故障诊断

王庆楠¹, 陈倩², 李涛³, 涂继辉⁴

(1. 怀化学院 物电与智能制造学院, 湖南 怀化 418000; 2. 上海理工大学 管理学院, 上海 200093; 3. 湖北师范大学 电气与自动化学院, 湖北 黄石 435002; 4. 长江大学 电子信息与电气工程学院, 湖北 荆州 434023)

摘要 针对轴承状态实时监测中故障样本稀缺的问题,提出一种轻量化孪生网络框架,以实现小样本条件下的高效故障诊断。采用连续小波变换对一维振动信号进行预处理,有效抑制噪声并提取关键时频特征,将信号映射为二维时频图。基于度量学习构建正负样本对,实现小样本数据的增强与泛化能力提升。在模型设计上,采用 MobileNetV3S 作为主干网络构建孪生结构,引入 ECA 通道注意力机制以降低模型复杂度并强化关键特征提取;结合深层 Bneck 模块的跳跃连接,优化深层特征的信息传递路径;最终以全局平均池化替代全连接层,进一步压缩参数量与计算开销。实验结果表明,所提方法在样本极度有限的条件下仍可实现 91.67% 的故障识别精度,参数量仅为 1.16 M,计算量为 0.05 GFlops,在诊断准确性、模型轻量化和实时性方面均展现出优异性能。

关键词 轴承诊断;小样本;孪生网络;轻量化

中图分类号 TP18;TH133.3

文献标志码 A

开放科学(资源服务)标识码(OSID)



Few-shot fault diagnosis for bearing based on lightweight siamese networks

WANG Qingnan¹, CHEN Qian², LI Tao³, TU Jihui⁴

(1. School of Intelligent Manufacturing, Huaihua University, Huaihua 418000, China; 2. School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China; 3. School of Electrical and Automation, Hubei Normal University, Huangshi 435002, China; 4. School of Electronic Information and Electrical Engineering, Yangtze University, Jingzhou 434023, China)

Abstract To address the challenge of fault diagnosis under limited labeled data, this paper proposes a lightweight Siamese network tailored for few-shot bearing condition monitoring. Raw one-dimensional vibration signals are first converted into two-dimensional time-frequency images via Continuous Wavelet Transform, enabling effective denoising and salient feature representation. A metric learning paradigm constructs positive and negative pairs to enhance data diversity and model generalization. The network leverages MobileNetV3S as the backbone, augmented with an Efficient Channel Attention mechanism to selectively emphasize informative channels while maintaining low complexity. Additionally, skip connections in deep Bneck modules are incorporated to improve feature propagation, and global average pooling re-

收稿日期:2025-01-15

基金项目:国家自然科学基金项目(62473135);湖南省教育厅科学研究项目(21C0641);武陵山农业控制湖南省重点实验室课题(ZNKZN2020-2)资助

通信作者:涂继辉,男,汉族,博士,副教授,研究方向:深度学习与计算机视觉,E-mail:tujh@yangtzeu.edu.cn。

places fully connected layers to further reduce parameters. Experimental validation shows that the proposed approach achieves 91.67% accuracy with only 1.16 M parameters and 0.05 GFLOPs, demonstrating superior performance in accuracy, efficiency, and real-time applicability under extreme sample constraints.

Keywords bearing-diagnosis; few-shot; siamese-networks; lightweight

0 引言

轴承作为旋转机械中关键的支撑部件,既承担着传递动力和运动的重要功能,又因工况复杂、负载频繁变化而成为最易损坏的环节。其运行状态的准确评估对保障机械系统的稳定性与可靠性具有重要意义。传统的轴承故障诊断主要依赖人工经验对振动信号进行分析,受主观因素影响大,难以满足工业现场对诊断实时性、准确性和自动化水平不断提升的需求。随着人工智能技术的迅速发展,基于深度学习的特征自动提取与故障识别方法在工程诊断领域表现出显著优势,在测井诊断^[1]、遥感监测^[2]、瑕疵检测^[3]等领域已取得了良好成效,并逐步推广至轴承故障诊断中以推动其智能化发展。初期多采用传统卷积神经网络结构处理轴承信号,采用 LeNet-5 实现风电机组轴承故障识别^[4],AlexNet 增强对铁路货车轴承特征的提取能力^[5]。随后结构更深、表达能力更强的网络被引入,融合 RepVGG 前置卷积与 CapsNet 胶囊网络提取高维空间特征^[6],提升模型对复杂故障模式的识别能力。时频域分析方法与深度模型相结合也成为研究热点,将短时傅里叶变换得到的时频图输入 Transformer 模型,捕捉全局时序依赖特征,实现高精度故障分类^[7];进一步结合 Transformer 与 ResNet,构建联合特征提取通道,以增强时空信息表达^[8]。为提升模型对非平稳信号的适应性,可变形卷积与扰动样本训练机制被引入,构建改进的一维卷积神经网络结构^[9];ResNeXt 与 LSTM 的双通道架构也被提出,分别提取空间与时间序列特征,提升对复杂轴承模式的建模能力^[10]。尽管上述方法在诊断性能方面取得显著进展,但其有效性仍高度依赖于大规模、标注准确的训练数据,这在一定程度上限制了其在复杂工程场景中的推广与应用。

在实际工程应用中,轴承大多数时间处于正常运行状态,故障发生具有低概率、突发性强等特点,且一旦发生通常需立即更换部件,导致可获取的故障样本极为有限,从而造成正常与故障样本间严重的不平衡问题。这种样本分布的不均衡不仅加大了模型训练的难度,也容易引发过拟合,降低模型在实际场景中的泛化能力与诊断准确性。为应对小样本学习挑战,学术界提出了一系列方法,主要包括模型无关元学习 MAML 和迁移学习 TL。其中,MAML^[11]作为一种典型的元学习策略,通过优化模型初始化参数以提升其在新任务上的适应能力,从而显著增强模型的泛化能力。基于 MAML 的元学习用于故障诊断,以解决少样本条件下的故障识别问题^[12];并将多标签学习方法与 MAML 结合^[13],用于提升样本不平衡条件下的特征提取效果。然而,MAML 在任务分布与训练分布差异较大时,性能表现不稳定,仍存在一定局限性。迁移学习因其能够将源域知识迁移至目标域,缓解目标任务样本不足问题,也被广泛应用于轴承故障诊断。诸如构建基于 SE-ResNet 的迁移学习框架实现故障识别^[14],将残差注意力机制嵌入深度可分离卷积结构以提升模型迁移能力^[15],基于多尺度、不同频率的 IMF 分量实现跨域特征迁移^[16]。迁移学习在小样本条件下表现出一定优势,但其诊断效果高度依赖于源域与目标域之间的特征相似性以及迁移策略的选择,限制了其在复杂多变的工业环境中的普适性与稳定性。

针对上述挑战,本文提出一种基于轻量化设计的孪生网络模型,用于解决轴承故障诊断中因样本稀缺引发的过拟合及识别精度不足的问题。所提出模型以孪生网络架构为基础,采用引入 ECA 注意力机制的轻量级主干网络进行特征提取,并通过补全深层 Bneck 模块的跳跃连接增强深层信息的传播能力,同时使用全局平均池化替代传统全连接层以减少模型参数与计算复杂度。原始振动信号经小波变换转化为时频图后输入模型,通过度量学习方法构建样本对,实现小样本条件下的有效特征对比与分类。在公开轴承数据集上的实验结果表明,该方法在诊断准确率、模型轻量化和实时响应能力方面均展现出显著优势,为小样本故障诊断任务提供了一种有效的解决思路。

1 轻量化孪生网络的轴承故障诊断方法

1.1 总体系统框架

针对轴承故障样本稀缺及诊断任务对实时性与模型轻量化的双重要求,本文提出一种融合孪生网络度

量学习机制与轻量化神经网络结构的轴承故障诊断方法,其整体结构如图1所示。该方法旨在利用孪生网络对小样本问题进行建模,并通过特征对比提升模型在小样本条件下的判别能力。首先,对采集到的原始一维轴承振动信号进行连续小波变换 CWT,将其转换为具有良好时频分辨能力的二维时频图,以增强非平稳信号特征的可表达性。随后,构建正负样本对作为孪生网络的输入,通过度量学习框架实现特征相似度的建模,从而间接实现对小样本数据的扩充。特征提取部分基于改进的 MobileNetV3S 架构,作为孪生子网络的共享特征提取器,用于提取判别性强的特征向量以支撑故障类型识别。

在网络设计方面,本文对 MobileNetV3S 主干结构进行了多项优化,引入高效通道注意力机制 ECA 以增强通道间信息的交互与重要特征的表达;通过优化 Bneck 模块结构并引入跳跃连接,缓解深层特征在传播过程中可能出现的信息衰减问题;采用全局平均池化替代传统全连接层,有效压缩模型参数规模并提升计算效率。与传统卷积神经网络相比,本方法通过样本对间的特征匹配操作,实现特征信息的高效压缩与传递,在保持模型轻量化的同时,提高了特征表达的准确性与鲁棒性。

在数据构建方面,本文利用孪生网络结构的并行性与相似性度量机制,设计了基于样本对的数据扩增策略,用以缓解轴承故障样本严重不足的问题。训练集中包含 C 类轴承状态,每类轴承状态均有 I 个样本,原始数据集的总样本数为 $C \times I$ 。随机选取两个样本重新组合,形成新的训练样本对,实现达到扩充数据的作用。属于同一种类别的两个样本组成同类样本对,随机选择两个进行组合可得 $C \times I \times (I-1)$ 组样本对;属于不同类别的两个样本组成异类样本对,随机选择两个进行组合可得 $C \times (C-1) \times I \times I$ 组样本对。通过该策略,模型训练样本数量实现大幅提升,相当于在不采集新数据的前提下实现 $O(C \times I^2)$ 级别的数据增强,有效提升模型的泛化能力与学习稳定性。

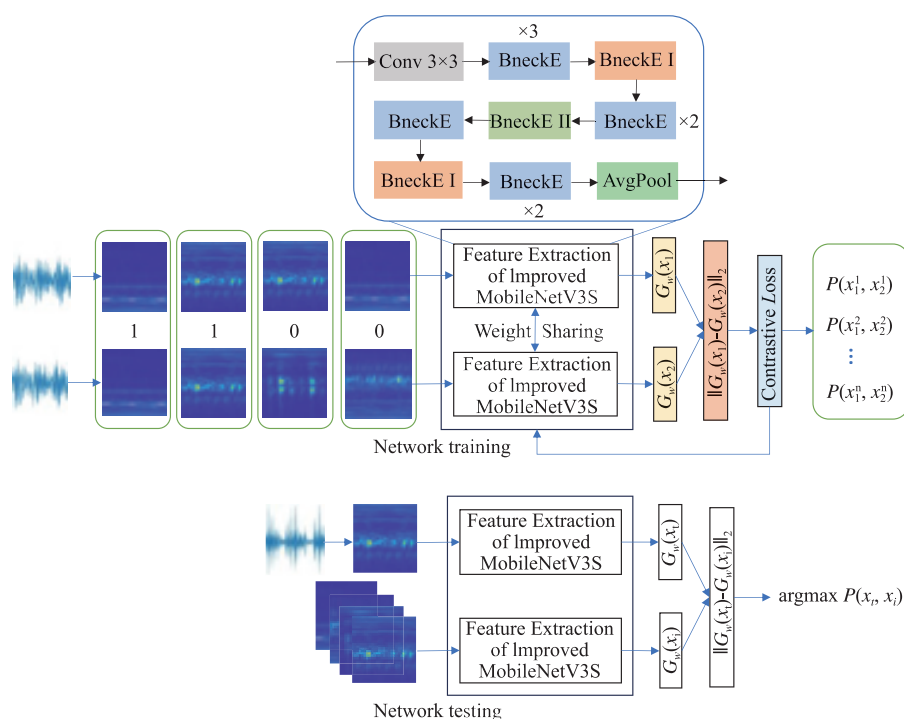


图1 轻量化孪生网络轴承故障诊断框架结构

Fig. 1 Framework structure of lightweight Siamese Networks bearing fault diagnosis

轴承故障诊断模型构造一个具有共享权重的并行神经网络结构,两个特征提取子网络共享相同的权重 w 和特征映射函数 $G_w(\cdot)$ 。在轴承故障诊断的过程中,子网络共享权重以相同的方式对输入样本特征提取,从而保证了特征映射的一致性和可比性。利用空间的语义距离来度量输入样本之间的相似性,相似度较高的样本会被映射到相近的位置,使得同类样本之间的距离较小,而不同类样本之间的距离较大。这种特性使得孪生网络在目标任务下,能够有效地对轴承样本对进行比较匹配。

在训练阶段,时频图构造的样本对根据类别标签划分为正样本对与负样本对,分别赋予标签 1 和 0。样本对的两个样本 x_1 和 x_2 成对输入提取高维特征向量,再映射到低维特征空间中,分别得到对应的特征表

示 $G_w(x_1)$ 和 $G_w(x_2)$ 。以对比损失函数作为训练目标,推动同类样本特征向量间的距离减小、异类样本间的距离增大。模型参数通过反向传播与梯度下降法迭代优化,直至损失函数收敛,完成模型训练。在测试阶段,对于待分类样本 x_t ,分别与每个类别的典型样本 x_i 组成测试样本对,并输入训练好的孪生网络中进行特征提取与相似度计算。以 sigmoid 函数转换为样本对之间的相似概率 $P(x_t, x_i)$,依据相似度最高的类别 $\operatorname{argmax} P(x_t, x_i)$ 判定该测试样本的轴承状态。

1.2 连续小波变换样本预处理

在轴承故障诊断中,原始采集的振动信号通常呈现为一维非平稳时域序列,虽能直接反映运行状态,但伴随大量噪声与冗余信息,严重制约了有效特征的提取与识别精度。为增强信号的可判别性与模型对关键特征的感知能力,本文采用连续小波变换 CWT 对原始信号进行预处理,将其从时域映射至二维时频域。CWT 具备优异的时频局部化能力,可通过多尺度与多时间位置的联合分解精细捕捉信号在不同时间-频率区域的瞬态变化,显著增强了非稳态振动信号的特征表达能力。相较于传统傅里叶变换,CWT 具备多分辨率分析优势,能在高频段提供优良的时间分辨率,在低频段保持良好的频率分辨率,尤其适用于非平稳信号的时频特征提取与表征。通过对基本小波函数 $\psi(\cdot)$ 进行尺度变换与偏移操作,连续小波变换可表示为:

$$W_f(a, \tau) = a^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-\tau}{a}\right) dt, \quad (1)$$

式中 $W_f(a, \tau)$ 表示信号 $f(t)$ 在尺度因子 a 和平移因子 τ 下的连续小波变换结果, $\psi(\cdot)$ 为选定的小波基函数。该变换通过多尺度分析方式在时频域内展开,有效提取信号中与故障相关的瞬时特征。连续小波变换 CWT 所生成的二维小波系数矩阵构成时频图,其横轴对应时间、纵轴对应频率,不同颜色深浅反映信号在特定时间和频率位置的能量强度。高亮区域表示局部能量显著,通常与振动信号的突发变化高度相关,可有效揭示轴承故障发生的时间与频率特征。在轴承状态监测中,冲击振动表现为幅值瞬时突变,对应于时频图中的高亮斑块。通过调节尺度因子 a 和偏移因子 τ , CWT 可实现多尺度、局部特征聚焦,有效放大信号细节并突出故障区域的纹理模式。故障信号常形成具有规律性的高能量结构,而背景噪声则呈现为随机分布的离散斑点。所获得的时频图在保留原始信号时域与频域结构的基础上,显著增强了微弱故障特征的可分辨性与抗干扰能力,不仅提升了对非平稳振动信号的可解释性与可视化效果,也为深度模型提供了更具判别力的高质量输入数据。

1.3 改进的 MobileNetV3S 子网络

在小样本轴承故障诊断任务中,孪生网络的诊断性能在很大程度上取决于其子网络的特征提取能力。浅层网络结构虽具有较低的计算复杂度,但其提取的主要为低级纹理或边缘信息,难以有效表征复杂振动信号中的高层语义特征,从而限制了模型对多样化故障模式的判别能力。尽管加深网络层数可提升特征表达能力,增强模型对非线性特征的建模能力,但随之而来的计算负担显著增加,不仅延长了模型训练和推理时间,也难以满足工业应用中对实时响应和资源约束的严苛要求。因此,在保持模型轻量化的同时,提升其对深层关键特征的感知能力,成为构建高效孪生网络结构的关键。

MobileNet 系列轻量化神经网络因其高度模块化设计与参数高效性,在资源受限场景中广泛应用。其中,MobileNetV1^[17]引入深度可分离卷积,将传统卷积操作取代为深度卷积与逐点卷积,在显著降低参数量与计算复杂度。MobileNetV2^[18]进一步在此基础上引入倒残差结构与线性瓶颈,有效提升了低维特征空间中的信息流通效率与梯度传播稳定性。MobileNetV3^[19]综合前两者优势,并融合了轻量级注意力机制 SE 及改进激活函数 h-swish,在保持模型紧凑性的同时增强了特征建模能力。MobileNetV3 提供两个版本,其中 MobileNetV3S 具备更浅的网络深度与更低的参数开销,适用于对计算资源和实时性要求较高的轻量级任务。本文选用 MobileNetV3S 作为孪生网络的主干结构,其核心 Bneck 模块融合了 1×1 降维卷积、 3×3 深度可分离卷积、SE 注意力机制及部分模块中的残差连接,在确保模型轻量化的同时实现了关键特征的有效保留与通道维度上的增强建模能力。

尽管 MobileNetV3S 在轻量化设计与特征提取能力之间取得了良好平衡,但其在小样本轴承故障诊断中的应用仍面临三方面亟需优化的问题:(1) 部分 Bneck 模块未配置跳跃连接,限制了深层特征的有效传递路径,易在网络加深时引发梯度消散与信息衰减;(2) 虽然 SE 注意力机制可增强通道间的特征建模能力,但

其依赖于全连接层进行通道维度的压缩与扩张,引发信息丢失并增加额外计算开销,与轻量化目标相悖;(3)在度量阶段继续采用传统全连接层作为判别结构,将显著提高参数规模,且在小样本条件下更易导致过拟合。针对上述问题,本文在以 MobileNetV3S 作为孪生网络特征提取子网络的基础上,引入跳跃连接优化 Bneck 结构、以 ECA 模块替代 SE 机制并采用全局平均池化替代全连接层,从而提升网络的特征表达能力与训练稳定性,满足小样本嵌入式故障诊断场景下的高效性与鲁棒性需求。

1.3.1 注意力机制优化。在 MobileNetV3S 的 Bneck 模块中,部分结构引入了 SE 注意力机制,以提升网络对通道间特征重要性的建模能力。该机制通过全局平均池化提取通道级全局信息,并依赖两层全连接网络对通道特征进行压缩与扩展,从而生成动态通道权重以实现特征重标定。该过程中通道维度的压缩可能导致关键信息的丢失,尤其在通道数量较少的情形下,难以有效捕捉跨通道的相关性。同时,全连接操作显著增加了模型参数量与计算开销,降低了网络的整体轻量化程度。为缓解上述问题并进一步增强模型的通道建模能力,本文引入 ECA^[20] 模块替代原结构中的 SE 模块,作为一种更为高效的通道注意力策略。ECA 机制摒弃了通道压缩与全连接操作,转而利用一维卷积直接在通道维度上建模局部通道间的交互关系,能够在保留高维语义信息的同时实现有效建模。具体而言,ECA 首先通过全局平均池化提取通道特征,再利用局部一维卷积核对相邻通道进行特征融合,并通过 Sigmoid 函数生成权重系数,实现通道维度上的特征重标定。本文在原 Bneck 模块基础上,将 SE 替换为 ECA 构建改进型 BneckE 结构,如图 2 所示。

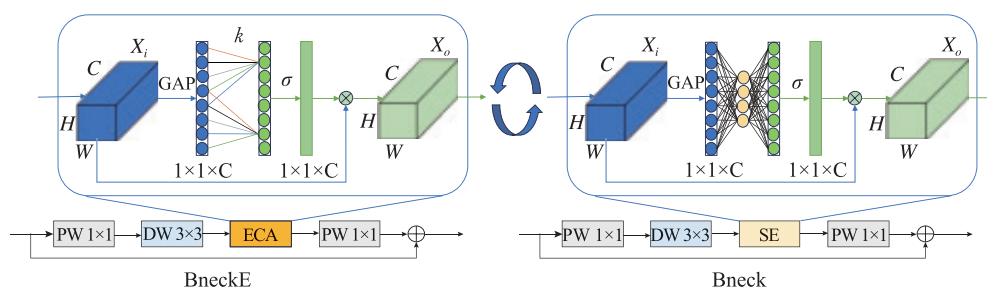


图2 引入 ECA 的 BneckE 结构图

Fig. 2 Architecture of BneckE module with ECA attention mechanism

ECA 模块对输入特征图 X_i 直接全局平均池化,然后通过自适应函数,得到与通道维度成比例的卷积核尺寸 k , 确定出通道交互的覆盖范围。卷积核尺寸 k 的自适应函数为 $k = \left\lceil \frac{\log_2 C}{\gamma} + \frac{\beta}{\gamma} \right\rceil_{\text{odd}}$, 式中 C 为给定通道的维数; γ 和 β 为超参量,控制调节 k 和 C 的比例; $\lceil \cdot \rceil_{\text{odd}}$ 表示取最接近函数的奇数。

该结构在无需显著增加参数的前提下,有效增强了模型对通道间局部依赖关系的建模能力,从而提升了对关键故障特征的响应效率。相比 SE 模块,ECA 通过 1×1 一维卷积实现跨通道注意力交互,显著降低模型复杂度与计算开销,减轻过拟合风险,更适用于小样本学习场景。ECA 仅包含全局平均池化与轻量卷积操作,具备高度参数效率与良好部署性能。因此,ECA 在保持网络轻量化的同时显著提升通道建模能力,为资源受限条件下的小样本轴承故障诊断提供了高效可行的解决方案。

1.3.2 深层 BneckE 跳跃连接补全。在 MobileNetV3S 网络中,BneckE 模块以倒残差结构为基本构建单元。当输入与输出特征图在空间尺寸与通道数上保持一致时,模块内部保留残差连接路径,以促进特征信息的直接传递与梯度的顺畅流动。然而,原始设计中在输入输出维度不一致的 BneckE 模块中省略了跳跃连接,未引入额外映射操作。尽管该策略保持了网络结构简洁,但在深层网络中省略跳跃路径会削弱语义特征的连续传输,限制深层信息的有效融合,进而影响梯度的稳定传播与模型整体的收敛性能。

鉴于浅层 BneckE 模块主要提取局部细节与低级特征,且其结构计算成本低、参数量较小,本文对该部分结构保持原始设计,以最大程度保证网络的轻量化特性。为增强深层语义特征的信息流通效率,本文在深层 BneckE 模块中引入结构补全策略,通过重构残差路径提升模型的表达能力与训练稳定性。具体而言,当输入与输出通道数不一致时,采用 1×1 逐点卷积实现通道对齐,确保残差路径具备可加性;若存在空间尺寸不一致,则引入步长一致的卷积操作替代传统池化方式,以保持跳跃连接所需的维度一致性。对于第 4 与第 9 个 BneckE 模块,先通过平均池化调整空间尺寸,再以 1×1 卷积匹配通道数;对于第 6 个模块,直接

采用 1×1 卷积对通道维度进行对齐,如图 3 所示。结构补全后的 BneckE 模块强化了深层特征在网络中的流通路径,有效缓解了梯度消失与性能退化问题,有助于小样本条件下模型的收敛和结构稳定性。

1.3.3 全局平均池化替代全连接层。在原型 MobileNetV3S 结构中,特征提取阶段后接入全连接 FC 层以实现特征整合。FC 层在输入特征维度较高时通常伴随大量的参数,其参数规模随输入维度和输出维度的乘积线性增长,尤其在在小样本条件下极易导致模型过拟合,显著抑制其泛化能力,难以满足小样本轴承故障诊断任务对高效性与鲁棒性的需求。为缓解上述问题,本文在构建孪生网络的特征提取子网络时,于 MobileNetV3S 主干结构的末端引入全局平均池化 GAP,替代传统 FC 层以实现特征压缩与整合。GAP 通过对每个通道维度上的空间特征执行全局平均运算,直接将二维特征图映射为一维通道向量,用于后续相似性度量模块的输入表达,如图 4 所示。相较于 FC 层,GAP 无需引入额外参数,保留了全局语义信息,显著降低模型复杂度,并有效抑制小样本条件下的过拟合风险。

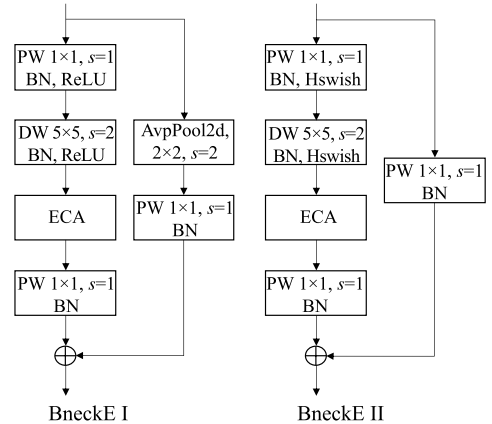


图 3 BneckE 模块的跳跃连接补全

Fig. 3 Skip connection of BneckE module

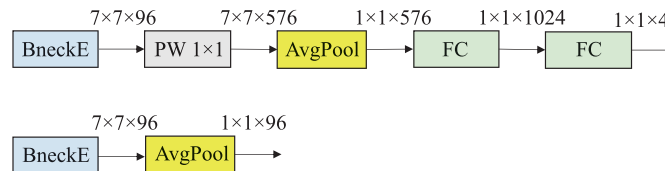


图 4 MobilenetV3S 末端的 GAP 改进

Fig. 4 Substitution of GAP in MobileNetV3S

GAP 的引入在结构设计上省略了 FC 层中的升维与非线性变换过程,进一步简化了网络前向推理路径,提升了模型的响应速度与部署效率。尤其在孪生网络中,特征提取子网络的输出主要用于表征样本间的语义相似度,而非直接进行显式分类,因此 GAP 替代 FC 的设计更契合任务需求,有助于实现高效、稳定的特征表达与信息传递,提升模型在资源受限环境中的实用性与适应性。

1.4 损失函数

在本研究提出的轴承故障诊断模型中,核心思想在于构建基于孪生网络(Siamese Network)架构的相似性度量机制,利用对比损失函数样本对间的特征表示进行监督学习,从而实现多类别轴承故障状态的有效区分。该模型以样本对作为输入,借助具有共享参数的特征提取子网络对每个样本进行高维嵌入表征,并通过计算所得特征向量之间的欧氏距离作为判别其相似性或差异性的度量依据。给定一个样本对 (x_1, x_2) ,孪生网络通过特征提取子网络 $G(\cdot)$ 得到特征向量 $G(x_1)$ 与 $G(x_2)$,计算其欧氏距离 $D(x_1, x_2) = \|G_w(x_1) - G_w(x_2)\|_2$,式中 x_1 和 x_2 是原始输入样本; $G_w(x_1)$ 和 $G_w(x_2)$ 是卷积网络提取的特征向量; $D(x_1, x_2)$ 是特征向量的欧氏距离; $\|\cdot\|_2$ 是范数 L2 用于计算欧氏距离。

通过引入对比损失函数进行端到端训练,网络能够在优化过程中有效学习到相似样本对应具有相近特征表示的嵌入映射。该损失函数旨在最小化同一类别样本对之间的特征距离,并在样本对属于不同类别时最大化其距离,从而引导模型在特征空间中形成具有判别性的类内紧聚与类间分离结构。训练过程中,模型参数通过梯度下降策略不断更新,使得网络输出的特征嵌入逐步符合上述优化目标,提升了模型在特征空间中的可分性与结构表达能力。

对比损失函数(Contrastive Loss) L 定义为 $L(Y, (x_1, x_2)) = \frac{1}{2}Y \cdot D^2 + \frac{1}{2}(1 - Y) \cdot \max(m - D, 0)^2$,

式中 Y 为样本对 (x_1, x_2) 的标签; $D = \|G(x_1) - G(x_2)\|$ 为样本对特征向量的欧氏距离; m 为样本对不相似的欧氏距离阈值。

当孪生网络接收来自同一类别的样本对作为输入时,其对应标签设定为 $Y = 1$,此时采用的对比损失函

数用于最小化两样本特征间的距离,对比损失函数为 $L(Y, (x_1, x_2)) = \frac{1}{2} D^2$ 。 L 函数值与特征向量的欧氏距离呈正比关系,特征向量越相似则欧氏距离越近, L 函数值趋 0,以学习相似样本的聚合特征。

当孪生网络接收来自不同类别的样本对作为输入时,其对应标签设定为 $Y=0$, 损失函数则旨在最大化其特征间的区分度,对比损失函数为 $L(Y, (x_1, x_2)) = \frac{1}{2} \max(m - D, 0)^2$ 。 L 函数值与特征向量的欧氏距离呈反比关系,特征向量越不相似则欧氏距离越近, L 函数值趋 0。当向量之间的距离甚至超过 m 设定阈值,说明二者已足够分离,此时 L 函数值为 0,参数停止更新。

通过对比损失驱动的孪生网络训练机制,模型可在小样本条件下构建清晰的特征判别边界,并实现对不同故障类型在嵌入空间中的有效聚类与区分。训练完成后,模型能够对任意测试样本与典型类别样本进行特征比对,依据最小特征距离完成故障类别判定,从而实现高效、精准的轴承故障诊断,在资源受限及数据稀缺场景中展现出良好的鲁棒性与泛化能力。

2 实验与分析

2.1 实验样本数据

本研究采用公认的西储大学轴承数据集作为实验数据来源。该数据集采集自轴承故障模拟平台,测试装置包括电机、扭矩传感器、功率测试仪以及被测的深沟球轴承。滚动轴承常见的故障类型主要集中在内圈、外圈和滚珠三个部位,为模拟典型故障状态,实验中采用放电加工 EDM 技术在轴承的上述三部位分别制造单个点坑,用以模拟正常状态、内圈故障、外圈故障和滚珠故障四种轴承运行状态。每类故障又进一步区分为三种损伤程度,分别对应点坑直径为 0.177 8 mm、0.355 6 mm 和 0.533 4 mm,以此反映不同程度的损坏情况。数据采集过程中,电机以 1 797 rpm、1 772 rpm、1 750 rpm 和 1 730 rpm 四种额定转速运行,从而形成多种运行工况。传感器采样频率设定为 12 kHz,采集得到的是一维的振动时域信号。由于原始数据量有限,为提升样本多样性与利用率,采用滑动窗口方式进行数据增强。训练样本使用长度为 1 024、步长为 256 的重叠窗口进行采样,不仅有效避免了等距采样可能造成的旋转周期信息遗漏,同时确保每个样本均封装了完整的故障特征信息,如图 5 所示。

在样本划分策略上,本文针对四类轴承运行状态(正常、内圈故障、外圈故障、滚珠故障),在四种转速工况及三种故障损伤程度的组合条件下进行均匀采样。具体而言,每类状态在每种工况与损伤等级下分别提取 150 个训练样本,训练集共计包含 600 个样本,并确保在不同运行条件与损伤程度间样本数量保持均衡分布,以增强模型对多样工况下故障模式的泛化能力。为确保测试集与训练集之间的严格独立性,测试样本的提取采用与训练样本间隔一定距离的方式进行,并统一使用长度为 1 024 的滑动窗口,避免样本重叠或信息泄露,有效防止数据混淆导致的评估偏差。最终,每类故障状态对应的测试集中包含 30 个样本,总计构成 120 个测试样本,具体分布情况详见表 1。该样本划分与采样策略不仅在结构上实现了训练与测试数据的完全独立控制,而且通过精细的滑动窗口参数设计提升了原始数据的利用效率。在小样本条件下,有效缓解了模型训练中因样本稀缺带来的过拟合风险,为后续的特征提取与故障识别任务提供了稳定、充足且分布合理的数据基础。

表 1 轴承类型分析数据集

Tab. 1 Data set of bearing type analysis

Category	Training Samples	Test Samples	Labels
Normal	150	30	N
Inner Fault	150	30	I
Rolling Fault	150	30	B
Outer Fault	150	30	O

2.2 实验环境

实验环境硬件包括 Xeon E5-2650 2.20 GHz 中央处理器 CPU, NVIDIA GeForce RTX 2080Ti 图形处

理器 GPU, 64 GB 内存与 1 TB 存储空间。软件环境采用 Python 3.9 编程语言与 PyTorch 1.10.1 深度学习框架进行模型开发与训练, 学习率设为 0.001, 批处理大小 batch size 为 32, 优化器选择 Adam, 训练轮数设置为 100 次 epochs。上述配置在保障训练效率的同时, 有效平衡了模型收敛速度与性能表现, 为实验结果的稳定性与可靠性提供了技术保障。

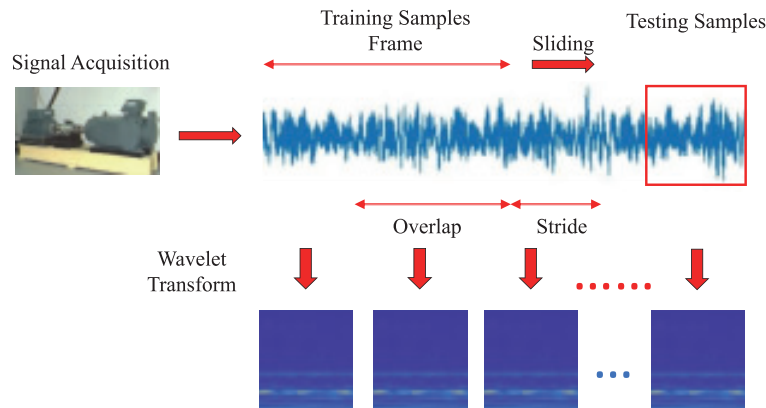
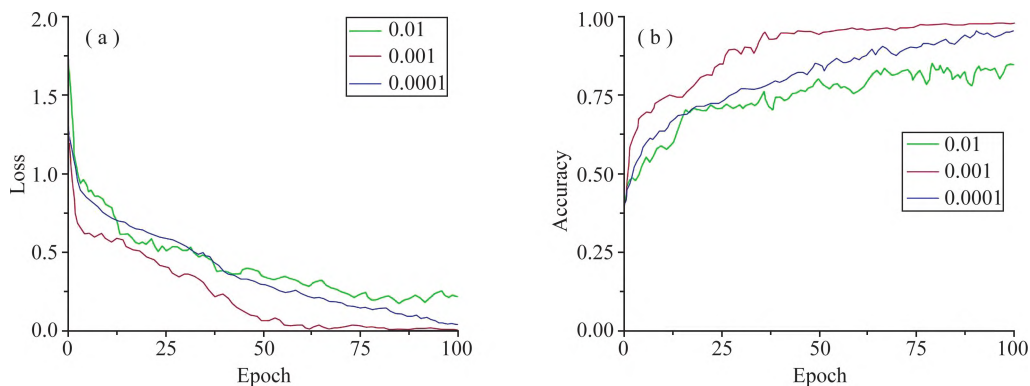


图 5 滑动窗口样本处理

Fig. 5 Sliding Window Sample Processing

本研究针对孪生网络在轴承故障诊断任务中的性能优化, 开展了以学习率为核心的超参敏感性分析。实验选取 0.01、0.001 和 0.000 1 三种学习率进行对比, 结合训练过程中的准确率变化与收敛稳定性进行系统评估, 结果如图 6 所示。当学习率设为 0.01 时, 尽管模型初期收敛速度较快, 但训练过程中波动显著, 存在跳过最优解的风险, 稳定性较差; 将学习率调整为 0.001 后, 模型在保证较快收敛的同时, 训练过程更加平稳, 识别准确率达到最优, 有效规避了梯度爆炸与欠拟合现象, 展现出良好的泛化能力; 进一步减小至 0.000 1 虽提升了训练稳定性, 但收敛速度显著降低, 且最终准确率有所下降。综合收敛效率、训练稳定性与模型性能表现, 确定 0.001 为最优学习率配置, 有效提升了模型在小样本轴承故障诊断场景下的识别能力与鲁棒性。



注: (a) Loss curves; (b) Accuracy curves.

图 6 不同学习率的模型性能曲线

Fig. 6 Model performance curves at various learning rates

2.3 实验综合分析

2.3.1 不同预处理方法的对比实验。在轴承故障诊断任务中, 不同小波基函数在提取振动信号时频特征方面存在显著差异, 直接影响孪生网络的故障识别性能。为系统评估各类小波基对模型诊断效果的影响, 本研究选取五种具有代表性的小波基函数进行对比实验, 包括实值小波 Mexican Hat (Mexh) 与 Gaussian (Gaus), 复值小波 Morlet (Morl) 与 Complex Gaussian (Cgau), 以及复合特性小波 Complex Morlet (Cmor)。在每类样本仅 15 张的小样本设定下, Mexh 与 Gaus 作为实值小波, 频域聚焦能力相对较弱, 导致模型提取的故障特征不够显著, Mexh 的诊断准确率是 80.16%, Gaus 的诊断准确率是 81.32%。Morl 与 Cgau 为复值小波, 具备更优的时频分辨能力, Morl 的诊断准确率是 84.36%, Cgau 的诊断准确率是 85.92%。其中,

Cmor 小波融合了 Morlet 小波的时频平衡特性与复小波的相位和幅度协同表达优势,能够在突出高频瞬态特征的同时,有效抑制模糊冗余信息,生成清晰度更高的时频图,从而显著提升模型的故障判别能力。在小样本条件下,Cmor 所获得的准确率最高,达到 91.67%,显示出最优的诊断性能。Cmor 小波在特征表达能力与诊断精度方面均优于其他候选小波,本研究采用 Cmor 小波作为故障信号预处理阶段的小波基函数,为后续的特征提取与模型训练提供高质量的输入表达。

为系统评估不同信号预处理方法在轴承故障诊断中的性能差异,尤其在小样本条件下的适应能力,本研究对比分析了三种典型的预处理方式,时域图 TDD、短时傅里叶变换 STFT 与基于 Cmor 小波的连续小波变换 CWT,如图 7 所示。实验在改进的 MobileNetV3S 孪生网络框架下开展,实施样本数量逐级递减的压力测试实验,分别在训练集样本数量从 150 至 15 的范围内进行模型训练,并评估各方法对识别准确率的影响,结果如图 8 所示。当样本量充足每类 150 个样本时,三种方法均可提取出具有良好判别能力的特征,其中 TDD 的准确率为 93.33%,STFT 的准确率为 98.75%,CWT 的准确率最高达 99.58%。然而在小样本场景(每类 15 个样本)下,识别性能呈现明显差异。TDD 的准确率急剧下降至 60.83%,STFT 的准确率为 85.42%,而 CWT 依然维持相对稳定水平,准确率达到 91.67%。这一结果充分表明,CWT 在小样本条件下具备更强的特征保持能力与抗噪鲁棒性。从方法机制上看,TDD 直接处理原始时域信号,无法有效表征非平稳信号中的局部瞬态特征,易受背景噪声干扰,导致特征提取效果显著下降;STFT 虽能提供一定的时频信息,但其固定窗口设计造成分辨率折中限制,在关键故障特征的表达上仍存在不足,尤其在样本稀缺条件下更为显著;相比之下,CWT 通过多尺度的连续变换,在时间与频率分辨率间实现灵活调节,能够突出非平稳信号中的局部关键细节,显著缓解 STFT 存在的时频模糊问题。所采用的 Cmor 小波具有优异的时间定位能力与频率聚焦特性,能够有效分离高频瞬态故障信息与低频运行状态特征,增强特征的判别性与模型的泛化能力。基于 Cmor 小波的 CWT 方法在特征提取效果、抗干扰能力与小样本适应性等方面均表现最优,最终被本研究选定为振动信号预处理的标准方法,为构建高性能轴承故障诊断模型提供坚实的数据支撑。

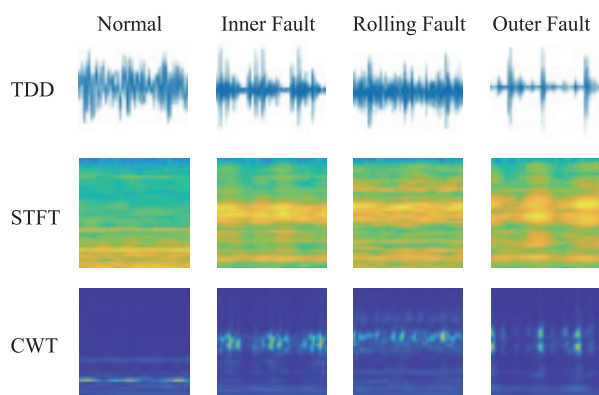


图 7 不同预处理样本

Fig. 7 Various pre-processed samples

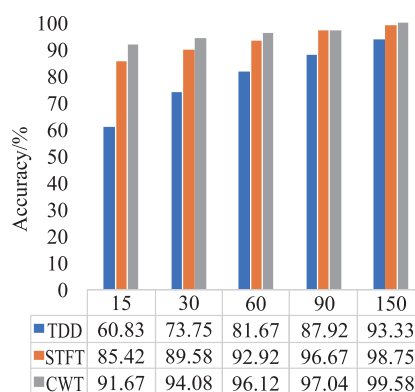


图 8 不同样本预处理方法的诊断准确率

Fig. 8 Diagnostic Accuracy Rates of Various Sample Pretreatment Methods

2.3.2 不同注意力机制的对比实验。在小样本轴承故障诊断任务中,以小波时频图作为模型输入,在每类仅 15 个样本的极小样本条件下,系统比较了四种主流注意力机制(SE、CA、CBAM、ECA)在改进的 MobileNetV3S 孪生网络框架下的性能表现,实验结果如表 3 所示。以 SE 模块为基准,网络识别准确率为 90.52%,参数量为 2.52 M。引入 CA 注意力机制后,准确率下降至 89.98%,但参数量减少至 2.28 M,降低幅度为 10.53%,显示其在参数优化方面具有一定优势。然而,CA 所依赖的坐标编码更适用于自然图像中的复杂空间结构,在结构规则、空间变化较小的小波时频图中难以充分发挥其建模能力。相比之下,CBAM 模块融合通道与空间注意力,在准确率上提升至 91.96%,但代价是参数量上升至 3.06 M,增加了 17.65%,在轻量网络中引入了较大开销,性能增益与复杂度提升不成正比。ECA 注意力机制在本研究中展现出优异的性能表现,识别准确率达 91.58%,相较 SE 提升了 1.06%,同时参数量降低至 2.06 M,压缩率达到 22.33%。ECA 通过引入局部一维卷积实现通道间的信息交互,避免了 SE 中全连接操作带来的信息压缩与

计算冗余,有效增强了对关键微弱特征的响应能力。其高效且轻量的设计使其与 MobileNetV3 中 Bneck 结构结合后,能够更精准地捕捉轴承故障中表现微弱、局部性强的特征模式,尤其适用于小样本、微故障等数据受限的场景。综合分析准确率、参数量以及实际应用适配性,ECA 注意力机制在小波时频图驱动的轴承故障诊断任务中实现了性能与效率的最优平衡,充分证明其在轻量级诊断模型中的应用价值。

2.3.3 改进 MobileNetV3 孪生网络的消融试验。为全面验证所提出三项结构优化策略在小样本轴承故障诊断任务中的有效性,本文以小波时频图作为输入,在改进 MobileNetV3S 构建的孪生网络框架下设计并开展系统性的消融实验。优化策略包括将原始 SE 通道注意力模块替换为 ECA,以提升通道间信息交互效率并降低计算冗余;在深层 Bneck 模块中引入跳跃连接,增强深层特征的传递与稳定性;以及用全局平均池化 GAP 替代末端全连接层 FC,以减少参数规模并提升模型的泛化能力。在每类仅 15 个样本的极小样本条件下,实验结果表 2 所示,ECA 模块相较于 SE 在准确率提升 1.06%的同时,参数量减少 0.46 MB,压缩 18.25%,显著增强了对微弱故障特征的建模能力;跳跃连接引入后,准确率进一步提升 0.12%,参数仅增加 0.01 MB,有效缓解了深层网络中的梯度消失与特征退化问题;GAP 替代 FC 结构则使参数量减少 0.91 MB,压缩 36.11%,并保持优良的判别性能。当三项优化联合应用时,改进 MobileNetV3S 孪生网络在总体准确率提升 1.15%的同时,模型参数总量减少 1.36 MB,压缩幅度达 53.97%,在实现高精度识别的同时兼顾了轻量化需求。上述结果充分证明了该优化策略在小波时频图驱动的小样本故障诊断任务中的适用性和有效性,展现出优异的性能与效率平衡,尤其适用于资源受限的工业场景的需求。

2.3.4 与传统网络的性能对比实验。进一步验证改进 MobileNetV3S 孪生网络相较于传统卷积网络,在极端小样本条件下的鲁棒性与泛化能力,设计并实施了样本数量逐级递减的压力测试实验。通过将每类训练样本数由 150 依次递减至 15,系统评估不同数据稀缺程度下各模型的故障识别性能变化趋势。实验结果表明,在样本数量充足每类 150 个时,所有模型均可实现较高的诊断准确率,其中深层网络 ResNet50 依托其强大的特征提取能力取得 99.78%的最高准确率,表现最优;本文提出的改进 MobileNetV3S 孪生网络紧随其后,准确率达到 99.58%,高于原始 MobileNetV3S 的准确率 99.42%及其他浅层网络模型 AlexNet 的准确率 99.36%、VGG16 的准确率 99.48%。当训练样本数递减至极限水平每类 15 个时,传统卷积神经网络诊断准确率骤降下降,AlexNet 的准确率是 58.48%、VGG16 的准确率是 57.27%、ResNet50 的准确率是 59.12%,MobileNetV3S 的准确率是 59.36%,显现明显的过拟合现象及泛化性能衰退。改进 MobileNetV3S 孪生网络在同等条件下依然保持了 91.67%的高准确率,展现出显著优于其他模型的稳健性与抗过拟合能力。该性能优势主要得益于孪生网络架构中的权重共享机制,使得模型能够有效捕捉类别间的特征相似性,在样本极度稀缺时依然实现高效的特征判别与故障识别。综合分析表明,改进 MobileNetV3S 孪生网络在小样本学习场景下具备卓越的鲁棒性与推广能力,具有良好的工程实用前景与应用价值。

表 2 消融实验对比

Tab. 2 Comparison of ablation experiments

ECA	Skip Connections	GAP	Accuracy/%	Parameters/M
			90.52	2.52
✓			91.58	2.06
	✓		90.64	2.53
		✓	90.54	1.61
✓	✓	✓	91.67	1.16

2.3.5 不同小样本学习方法的性能对比实验。为全面评估本文所提改进型 MobileNetV3S 孪生网络在小样本轴承故障诊断中的性能与应用潜力,设计与主流小样本学习方法的对比实验,选取基于 MobileNetV3S 的迁移学习模型及代表性元学习方法 MAML 作为对照,确保模型结构、参数规模及输入一致,基于轴承故障小波时频图输入,训练样本量限定为每类 15 个,统一测试集下对识别准确率进行客观比较。迁移学习通过预训练模型微调实现,准确率为 84.86%,但受限于源域与目标域分布差异及样本匮乏,迁移效果及泛化能力受损。MAML 作为典型优化型元学习方法,通过多任务二阶梯度优化实现快速适应,准确率提升至 87.54%,显示较强小样本适应性,但其浅层结构限制了对低纹理、细节丰富时频图的特征表达能力。

相比之下,本文提出的改进 MobileNetV3S 孪生网络在相同条件下达 91.67% 准确率,显著优于前两者。该模型在轻量化设计基础上融合高效 ECA 通道注意力机制以强化通道特征选择并减少参数,引入深层 Bneck 模块跳跃连接以优化特征流通和梯度传递,有效缓解深层网络训练难题;通过孪生网络权重共享及对比学习机制,提升微弱故障特征判别能力与泛化性能;此外,末端全局平均池化替代全连接层,进一步降低模型复杂度及过拟合风险。尽管迁移学习与元学习在小样本领域具一定优势,面对轴承小波时频图中稀疏且局部差异显著的特征,仍存在明显局限。改进型 MobileNetV3S 孪生网络通过结构优化与度量学习的协同设计,显著提升了故障特征捕捉能力、模型鲁棒性及泛化性能。

2.3.6 不同特征提取子网络的孪生网络性能对比实验。在小样本轴承故障诊断任务中,孪生网络中特征提取子网络的结构设计对整体识别性能具有决定性影响。为系统评估改进 MobileNetV3S 作为特征提取子网络的有效性,本文将其与浅层网络 AlexNet、VGG16,深层网络 ResNet50 及多种轻量化网络 ShuffleNetV2、EfficientNet-B0、SqueezeNet、Ghost、原型 MobileNetV3S 嵌入权重共享的孪生网络框架中,基于每类仅 15 张小波时频图的极小样本条件,开展系统性对比实验。实验结果表明,各类网络在特征提取能力、泛化性能及参数效率方面存在显著差。浅层网络 AlexNet 与 VGG16 因结构浅显、语义表达能力不足,主要提取低层次纹理与边缘信息,难以充分表征轴承小波时频图中的关键冲击特征。AlexNet 孪生网络的诊断准确率是 87.63%,VGG16 孪生网络的诊断准确率是 88.72%,AlexNet 孪生网络参数 57.10 M 和计算量 0.71 GFLOPs,VGG16 孪生网络参数 138.31 M 和计算量 15.52 GFLOPs,模型参数规模庞大且计算复杂度较高,制约了其在资源受限的工业现场的实部署能力。深层网络 ResNet50 虽具备强大的特征学习能力,但考虑到轴承时频图信息密度较低,深层网络容易产生冗余表征及特征损失,ResNet50 孪生网络准确率是 90.25%、参数量 25.56 M、计算量 3.53 GFLOPs,准确率提升有限而模型规模与计算负担依旧较重,不利于现场轻量化应用。

轻量化网络因其低计算复杂度与模型尺寸,在工业环境中表现出良好应用前景。实验结果显示,在极小样本条件下,SqueezeNet 孪生网络的准确率是 87.86%、ShuffleNetV2 孪生网络的准确率是 89.57% 及 Ghost 孪生网络的准确率是 90.12%,参数量和计算量均控制在较低水平,但其对小波时频图中细粒度冲击特征的捕捉能力仍受限,识别性能存在提升空间。原型 MobileNetV3S 孪生网络通过引入 SE 注意力机制实现了 90.56% 的准确率,参数量为 2.52 M,计算量 0.06 GFLOPs,性能表现较为均衡。EfficientNet 孪生网络虽达到了最高的准确率 91.56%,但其较大的参数规模 5.30 M 及计算开销 0.39 GFLOPs,限制了其在现场应用能力。

表 3 不同主干孪生网络的小样本轴承故障诊断结果对比

Tab. 3 Comparison of fault diagnosis results of Few-shot bearings with various Siamese Networks

Model	Accuracy/%	Parameters/M	GFLOPs
SamAlexNet	87.63	57.10	0.71
SamVgg16	88.72	138.31	15.52
SamResNet50	90.25	25.56	3.53
SamSqueezeNet	87.86	4.80	0.34
SamShuffleNetV2	89.57	2.28	0.15
SamGhost	90.12	5.18	0.14
SamEfficientNet	91.84	5.30	0.39
SamMobilenetV3S	90.52	2.52	0.06
Ours	91.67	1.16	0.05

本文提出的改进 MobileNetV3S 孪生网络通过集成高效 ECA 注意力机制替代传统 SE 模块、引入深层 Bneck 跳跃连接优化特征传递路径、以及采用全局平均池化替代冗余全连接层,实现了在极小样本条件下的优异表现。具体而言,本文模型达成了 91.67% 的识别准确率,参数量 1.69 M,计算量 0.05 GFLOPs。与原型 MobileNetV3S 孪生网络相比准确率提升 1.15%,参数规模减少 53.97%、计算量减少 16.67%;相较于 SqueezeNet 孪生网络,准确率提升 3.81%,参数量减少 75.83%,计算量降低 85.29%;本文模型虽较 Effi-

cientNet 孪生网络准确率低 0.17%，但参数量压缩 78.11%、计算量压缩 87.18%。综合权衡准确率、模型复杂度与计算效率，本文模型提出的改进 MobileNetV3S 孪生网络实现了轻量化与高效性的高度融合，显著提升了小样本轴承故障诊断的实用性与推广价值，体现出其在资源受限工业环境应用的显著优势。

2.3.7 改进 MobilenetV3S 孪生网络轴承诊断能力的可视化分析。为直观评估和展示改进 MobilenetV3S 孪生网络的轴承特征分布特性，本文利用 t-SNE 方法对小波时频图输入及网络提取的深层特征进行降维映射，结果如图 9 所示，直观反映了各类别特征在低维空间中的聚类效果与可分性。结果表明，原始输入特征在低维空间中呈现高度混杂状态，类别间边界模糊，难以支持有效的分类判别。相比之下，经改进 MobileNetV3S 孪生网络提取的特征在二维嵌入空间中表现出显著的类别区分性，不同故障类别形成了紧密且分离明显的聚类结构，类内样本高度集中，类间距离明显增大。该现象表明网络有效构建了区分性强且鲁棒的特征表示，显著增强了轴承状态的判别能力及模型的泛化性能。

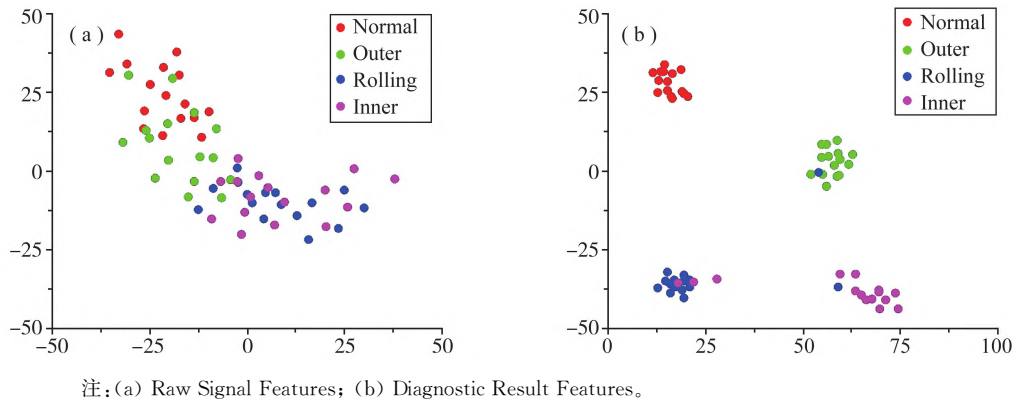


图 9 t-SNE 可视化结果

Fig. 9 t-SNE visualization

为全面评估改进后的 MobileNetV3S 孪生网络在轴承故障诊断任务中的分类效果，本文采用混淆矩阵对模型的识别结果进行了详细的性能解析，如图 10 所示。实验结果表明，该模型对“正常状态”与“外圈故障”两类的识别准确率均达到 100%，不存在误判现象，充分体现其在健康状态监测及典型冲击故障检测中的稳定性与判别能力。相较之下，“滚动体故障”类别存在少量混淆，约 6.67% 的样本被误分类为“外圈故障”或“内圈故障”，反映该类故障信号特征的复杂性及与其他类别间的相似性对识别性能的影响显著。尤其是“内圈故障”类别，识别准确率相对较低，约 20% 的样本被误判为“滚动体故障”，其主要原因在于故障信号传播过程中经多层介质衰减和干扰，导致信号强度减弱且特征模糊，难以被模型充分捕获。滚动体故障由于其动态接触特性，时频特征表现复杂多变，进一步加剧了两类故障间的特征重叠。小波时频图中两者的高度相似性反映了其物理结构上的耦合关系，成为模型精确区分的主要挑战。鉴于现场轴承诊断的核心任务侧重于故障状态的有效识别，该模型已能充分满足在线监测的基本应用需求。混淆矩阵分析不仅揭示了物理信号传播机制对诊断性能的深远影响，也为后续特征增强与模型优化提供了理论支持。总体而言，改进 MobileNetV3S 孪生网络在小样本条件下表现出卓越的识别能力，在样本稀缺的条件下能够准确区分正常与故障状态，满足轴承故障早期预警和及时维护的实际需求，并兼顾故障类别的判别。

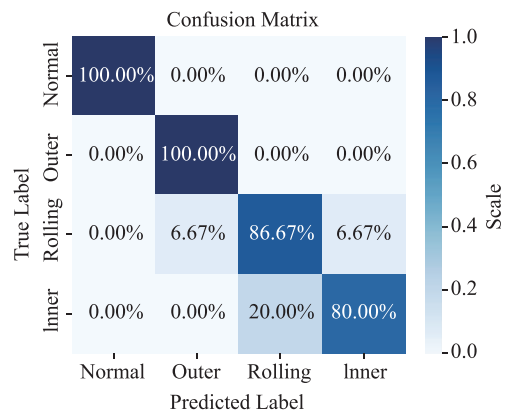


图 10 混淆矩阵结果

Fig. 10 Confusion matrix accuracy results

3 结论

本文提出了一种基于轻量化设计的孪生神经网络用于轴承故障诊断，旨在应对轴承故障样本稀缺与诊断精度之间的矛盾，兼顾模型的准确性、鲁棒性。所提出方法通过多层次结构优化和特征增强策略，实现了

在小样本条件下的高效故障识别与强泛化性能,其优越表现主要归因于以下关键改进:(1)小波时频预处理增强特征可辨性,对原始振动信号进行连续小波变换,显著突出关键的非平稳故障特征,抑制背景噪声干扰,为后续网络提取更具区分度的特征提供了良好的输入基础,提升了模型对弱故障信号的敏感性。(2)构建基于 MobileNetV3S 的孪生网络架构,所设计的孪生结构不仅能够提取单个样本的深层语义特征,还通过度量学习机制对样本对之间的相似性进行建模,有效缓解了因小样本而导致的过拟合问题,增强了模型在故障类别判别上的泛化能力。(3)多项结构优化提升性能与效率,引入 ECA 通道注意力模块,实现了对局部重要通道特征的动态加权,提高了对关键故障信息的响应能力;跳跃连接机制增强了深层特征的传递路径与表达稳定性;全局平均池化替代传统全连接层,在显著压缩模型参数规模的同时,有效降低了过拟合风险,并提升了网络推理效率与部署可行性。

本文所提出的改进 MobileNetV3S 孪生网络通过融合小波特征预处理、度量学习与轻量结构设计,在保持模型高精度诊断能力的同时,有效降低了对大规模样本数据的依赖,展现出良好的实时性、鲁棒性与工程应用潜力,为面向实际工业场景的轴承故障智能诊断提供了切实可行的技术方案。

参 考 文 献

- [1] LI S, FU B, WEI J D, et al. Ultrasonic logging image denoising based on CNN and feature attention[J]. IEEE Access, 2021, 9: 116845-116856.
- [2] WANG Y L, FENG W Q, JIANG K, et al. Real-time damaged building region detection based on improved YOLOv5s and embedded system from UAV images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16(5): 4205-4217.
- [3] 李辉,吕祥聪,徐凌伟,等. 融合形变卷积和自注意力的色素布匹瑕疵检测[J]. 聊城大学学报(自然科学版), 2022, 35(6): 1-10.
- [4] 郭俊锋,孙磊,王森生,等. 基于 1D-LeNet-5 模型的滚动轴承故障诊断方法[J]. 兰州理工大学学报, 2023, 49(5): 34-41.
- [5] 汤武初,刘佳彬,吕亚博,等. 基于一维 AlexNet 卷积神经网络的铁路货车轴承故障识别[J]. 黑龙江交通科技, 2024, 47(10): 125-130+136.
- [6] 尹爱军,吕明阳,杨敏英,等. RepVGG 与 CapsNet 融合的轴承故障诊断[J]. 振动与冲击, 2024, 43(14): 301-307.
- [7] 齐萌,王国强,石念峰,等. 基于时频图与视觉 Transformer 的滚动轴承智能故障诊断方法[J]. 轴承, 2024 (10): 115-123.
- [8] 陈文强,安迪,张磊,等. 融合 Transformer 与 ResNet 的轴承故障诊断[J]. 自动化与仪器仪表, 2024 (12): 111-114.
- [9] 王体春,解缙,咸玉贝,等. 滚动轴承故障诊断的 TD-DCCNN 方法研究[J]. 重庆理工大学学报(自然科学), 2023, 37(7): 135-143.
- [10] 杜浩飞,张超,李建军. 基于 SENet-ResNext-LSTM 的风机轴承故障诊断[J]. 机械强度, 2023, 45(6): 1271-1279.
- [11] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]// Proceedings of the 34th International Conference on Machine Learning, 2017: 1126-1135.
- [12] SU H, XIANG L, HU A J, et al. A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions[J]. Mechanical Systems and Signal Processing, 2022, 169(4): 1058-1065.
- [13] YU C C, NING Y Q, QIN Y, et al. Multi-label fault diagnosis of rolling bearing based on meta-learning[J]. Neural Computing & Applications, 2021, 33(9): 5393-5407.
- [14] 唐宇峰,杨泽林,胡光忠,等. 基于 SE-ResNet 和迁移学习的轴承故障诊断方法研究[J]. 四川轻化工大学学报(自然科学版), 2024, 37(4): 19-27.
- [15] 邓明星,王严智,敖金艳,等. 基于深度迁移学习的少样本变工况轴承故障诊断[J/OL]. 武汉大学学报(工学版), 2025: 1-20. <https://link.cnki.net/doi/10.14188/j.1671-8844.2024.0419>.
- [16] 尹洪申,刘文峰,俞啸,等. 多尺度迁移学习的轴承故障诊断[J]. 机械设计与制造, 2025(1): 10-14.
- [17] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. 2017, arXiv:1704.04861.
- [18] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [19] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [20] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.

(责任编辑:赵海军)