# 基于深度学习的工业故障诊断综述

刘昕晖[1],陈晫群[2],吕 岩[3]

(1. 吉林大学 机械与航空航天工程学院,长春 130022;2. 香港城市大学(东莞),广东 东莞 523770;3. 吉林大学 建设工程学院,长春 130015)

**摘 要**:系统梳理了深度学习应用于工业故障诊断领域的各类研究成果,以填补现有综述未完整梳理方法演进路线、关键技术方向和实际应用方面各类问题的空白。梳理过程中可观察到,深度学习在该领域的发展已经逐渐从搭建基础网络架构延伸到开发混合架构,当前研究方向集中在注意力机制、迁移学习、生成对抗网络等技术方向的创新开发。实际落地过程中,工业场景对模型提出了多重要求,包括数据获取难度大、模型逻辑难解释、运算速度需匹配工业生产节奏等。本文进一步展望了自动化机器学习、多模态融合、物理信息融合等未来方向,可方便后续研究者进行相关工作,促进智能诊断技术融入自主运维模式。

**关键词**:机械电子工程;工业故障诊断;注意力机制;迁移学习;自主机器学习;生成对抗网络
**中图分类号**:TP181 **文献标志码**:A **文章编号**:1671-5497(2026)07-1759-21
**DOI**:10.13229/j.cnki.jdxbgxb.20260088

## Review of industrial fault diagnosis based on deep learning

LIU Xin-hui[1],CHEN Zhuo-qun[2],LYU Yan[3]

(1. *School of Mechanical and Aerospace Engineering*, *Jilin University*, *Changchun* 130022, *China*; 2. *City University of Hong Kong*(*Dongguan*), *Dongguan* 523770, *China*; 3. *College of Construction Engineering*, *Jilin University*, *Changchun* 130015, *China*)

**Abstract**: This paper systematically reviews the various research results of in-depth learning applied in the field of industrial fault diagnosis, in order to fill the gaps in the evolution route, key technology direction and practical application of existing reviews. During the combing process, it can be observed that the development of deep learning in this field has gradually extended from the construction of basic network architecture to the development of hybrid architecture. The current research direction is focused on the innovative development of technology directions such as attention mechanism, transfer learning and generative adversarial networks. In the actual landing process, the industrial scene puts forward multiple requirements for the model, including the difficulty of data acquisition, the difficulty of model logic interpretation, and the need for the calculation speed to match the pace of industrial production. This paper further prospects the future directions of automated machine learning, multimodal fusion, physical information fusion, etc., which can facilitate the follow-up researchers to carry out related work and

**收稿日期**:2026-01-25.
**基金项目**:国家自然科学基金项目(52475143).
**作者简介**:刘昕晖(1962-),男,教授,博士.研究方向:机械电子工程.E-mail:liuxh@jlu.edu.cn
**通信作者**:吕岩(1983-)女,副教授,博士.研究方向:岩土体工程地质.E-mail:lvyy@jlu.edu.cn

promote the integration of intelligent diagnosis technology into the independent operation and maintenance mode.

**Key words**: mechanical electronics engineering; industrial fault diagnosis; attention mechanism; transfer learning; automated machine learning; generative adversarial networks

## 0 引 言

随着现代工业系统朝着集成化和智能化方向不断发展,设备故障变得日益复杂起来,传统诊断办法难以契合精准、快速、智能方面的要求,特别需要引入智能诊断的方式达成故障的自动提取以及识别。以卷积神经网络(Convolutional neural network,CNN)作为核心的混合模型,特别擅长从振动信号或时频图像当中去感知空间特征,与循环神经网络(Recurrent neural network,RNN)、长短期记忆网络(Long short-term memory,LSTM)等进行结合,可对时序动态进行建模,形成"感知-认知-适应"的技术路径[1,2]。为了应对数据稀缺、工况多变这类挑战,小样本学习、注意力机制、迁移学习,以及生成对抗网络等前沿技术得到了广泛研究,以此提高模型在复杂环境当中的泛化能力[3,4]。

尽管已有的研究在不同层面推动了故障诊断与深度学习的交叉融合[2,5],但仍存在以下不足:比较多的研究偏重单一网络或学习范式[6-9],没有对CNN、RNN、Transformer等主流架构以及其混合模型进行系统性的对比,关于注意力机制、迁移学习等关键技术的创新设计以及适配性方面,讨论并不到位,对于多类工业场景的应用成效,缺乏系统的总结[10,11],在数据稀缺、可解释性、实时部署等挑战方面[12,13],分析尚未达到体系化的认知。

为此,本文在对深度学习基础方法论及其发展演进过程进行梳理之后,系统地总结CNN、Transformer等模型,以及它们的混合模型在工业故障诊断中的发展历程,着重介绍了注意力机制、迁移学习、数据生成等关键技术的设计创新点,阐述了目前所面临的挑战,并且对未来研究方向进行了展望。

## 1 故障诊断中的深度学习方法

深度学习的发展起始于早期的多层神经网络以及反向传播算法,然而在过去曾由于理论方面的瓶颈以及算力的限制而陷入发展的低谷期[14],一直到大约2006年的时候,借助无监督预训练[15]、线性整流函数(Rectified linear unit,ReLU)激活函数、图形处理器(Graphics processing unit,GPU)计算以及大规模数据等关键方面的突破才得以实现复兴,并且由此催生了以卷积神经网络和循环神经网络作为代表的深度架构[16]。在现阶段,深度学习凭借包含数亿参数的深层模型,在图像识别、语音处理以及自然语言理解等众多领域已经达到了近乎人类的性能水平,并且在工业界得到了广泛的应用,不过在无监督学习、复杂推理以及与强化学习相结合的通用人工智能等更具挑战性的领域,目前仍然处于探索的阶段。

### 1.1 基础监督学习与深度学习

监督学习借助已标记的训练数据学习输入到输出的映射关系,这成为了深度学习在故障诊断里最为常见的范式[17]。它的核心流程,也就是前向传播计算预测、反向传播基于误差梯度调整网络参数,为端到端的智能诊断搭建了基础框架。不过,工业故障诊断场景对模型有着独特要求,信号通常具备强噪声、非平稳、多尺度等特性,同时故障样本非常稀缺。这使得研究者对基础学习范式展开针对性改进。基础的深度前馈网络通过堆叠非线性层,像采用ReLU激活函数,拥有强大的特征变换能力,能把原始工业信号的复杂分布"映射"到更易区分的特征空间(见图1)。这种自动特征学习能力取代了传统诊断里依赖专家知识的手工特征工程,成为深度学习助力故障诊断的核心优势。基础监督学习与深度学习框架给故障诊断提供了可学习的映射函数与优化工具,针对工业数据特点进行算法改进与结构设计,是其成功用于具体诊断任务的关键[18-21]。

在优化算法这一方面,标准的随机梯度下降(Stochastic gradient descent,SGD)及其变体,虽然是模型训练的基础,但在工业场景里容易受到噪声干扰,而且收敛速度慢。基于此,故障诊断领域引入了多种元启发式、贝叶斯优化方法,目的在于提高模型在复杂数据情况下的训练效率和稳定性。如采用粒子群优化(Particle swarm optimiza-

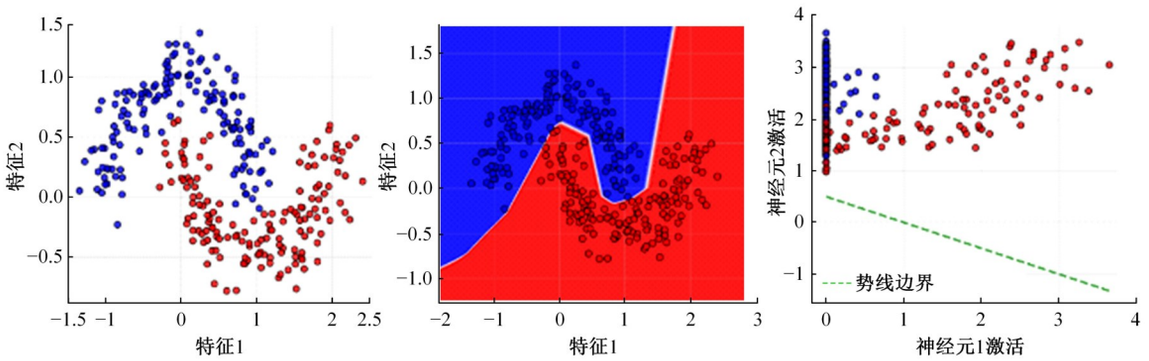



**图 1　多层神经网络空间扭曲示例**

**Fig. 1　Multi layer neural network space warp example**

tion，PSO）自动调整 CNN 的超参数，或是结合麻雀搜索算法对 CNN-LSTM 混合模型的参数进行优化，这些改进加快了模型收敛速度，并且提高了诊断精度[22-25]。

### 1.2　基于无监督与半监督学习的故障诊断

当监督学习面临标注数据稀缺且成本高昂、未标注数据大量存在，以及模型泛化能力不足等现实问题时，无监督和半监督学习便应运而生。它们在继承监督学习基础框架的同时，通过引入未标注数据提升模型性能。例如，无监督学习专注于从无标注数据中发现内在结构和特征表示，而半监督学习则通过在损失函数中结合监督损失与基于未标注数据的正则化损失，有效利用少量标注和大量未标注数据，从而显著提高数据利用效率、增强模型泛化能力，并更好地适应真实世界中海量未标注数据的应用场景。

在 2006 年左右，深度学习有过一次“复兴”，而这次复兴则很大程度上归功于无监督学习方法的突破。深度信念网络（Deep belief net，DBN）等无监督预训练方法通过逐层学习特征检测器，为深层网络的训练提供了有效的参数初始化策略。例如，文献[26]证明了无监督预训练在有限标注数据场景下的有效性，该方法特别适用于工业故障诊断中故障样本稀缺的实际情况，通过在大量无标签数据上预训练，模型能够学习到数据的基础分布特征。除此之外，生成对抗网络（Generative adversarial networks，GAN）和变分自编码器（Variational autoencoder，VAE）等生成模型为解决数据不平衡问题提供了新思路。一项研究提出的具有平衡串行 CNN 和 Transformer 的 GAN（Generative adversarial networks with balancing serial CNN and transformer，BCTGAN）通过串联式 CNN-Transformer 结构生成高质量故障样本，在极端不平衡条件下仍能保持极高的诊断准确率[27]。另外一种 VAE-CNN 混合模型在低信噪比的噪声环境下仍表现优异[28]。

非接触式诊断是深度学习的一个重要应用方向。基于红外热像仪和迁移学习的诊断方法，仅需少量标记样本即可实现转子-轴承系统故障的准确识别，有效解决了振动传感器安装受限场景的监测难题[26]。这种方法的核心创新在于从无监督的卷积自编码器向有监督的增强卷积神经网络进行参数迁移，使得模型初始参数已经包含了从红外图像中提取通用特征的强大能力。另一项创新性工作设计了无监督域共享 CNN 架构，引入 Cauchy 核诱导的最大均值差异（Maximum mean discrepancy，MMD）替代传统的高斯核函数。这种改进不仅提升了跨工况特征适配的计算效率，还增强了在从稳态速度到时变速度等复杂迁移场景中的鲁棒性，为处理非平稳信号提供了新的技术路径[29]。

### 1.3　基于小样本、零样本与迁移学习的故障诊断

工业故障诊断通常会遇到一个非常关键的挑战，那就是标注数据非常稀缺。针对这个情况，研究者们开发出了 3 种主要的技术范式。这 3 种技术范式依据可用数据、知识的多少，形成了一个从“数据驱动”到“知识引导”的连续谱系。其中，小样本学习是在每类只有少量标注样本的状况下进行工作的；零样本学习是在目标故障类别完全没有标注样本时，凭借先验知识或者属性描述进行推理；迁移学习是借助从相关源域迁移已经学到的知识，以此来辅助目标域的学习。这 3 种技术范式共同组成了一个用于解决数据稀缺问题的系统性工具箱。

1.3.1 基于有限样本快速适配的小样本学习

小样本学习的主要目的是,在每类故障只有几个标注样本的情形下,训练出具备强大泛化能力的诊断模型。主流方法有3大类。基于数据增强的方法,借助生成模型或者基于物理机理的仿真,比如机构特征生成模型(Mechanism character generative model,MCGM)合成虚拟故障样本,有效增加训练集。基于度量学习的方法,如原型网络、匹配网络,学习一个嵌入空间,让同类故障样本的特征相互接近,异类样本彼此远离,在推理时通过计算和少数支持样本的距离进行分类。基于元学习的方法在大量相关任务上展开元训练,使模型拥有快速适应新故障类别的能力。这些方法让模型在只有轴承、齿轮等旋转机械的少量实测故障数据时,依然能够实现较高的诊断精度,特别适用于新型设备或罕见故障的初期诊断阶段[30]。

1.3.2 在无样本条件下进行知识推理的零样本学习

在目标故障类别完全不存在任何标注样本的情况下,零样本学习给出了可行的解决办法。它的核心想法是借助先验知识或者语义嵌入,搭建起可见故障类别与不可见故障类别之间的语义联系。一种典型的范式是"属性学习",也就是为每种故障确定一组能够解释的属性,训练模型学习从样本到属性向量的映射。在进行诊断时,就算遇到没见过的故障,模型也能够通过预测它的属性向量,并和故障-属性知识库做匹配来推断类别。另外一种范式是依靠生成模型,利用可见类别的样本、语义描述,生成不可见类别的合成特征,进而把零样本问题转变为常规分类问题。零样本学习为处理"未知的未知"故障、达成开集诊断开拓了途径,是建立具备认知扩展能力智能诊断系统的关键所在[31]。

1.3.3 迁移学习:跨域知识的复用与适配

迁移学习要把从一个相关领域(即源域,如某种稳定工况下的振动数据)学到的知识,迁移并适配到数据稀缺的新领域(即目标域,比如变工况下的数据),以此降低对目标域标注数据的依赖。依据源域与目标域的差异程度,迁移学习在故障诊断中呈现多种形式。特征迁移是较常见的方式,在源域上预训练特征提取器(如卷积自编码器[32]),再将其参数迁移到目标域模型作为初始化,使模型拥有从目标域少量数据中快速学习的能力。实例迁移方法是通过重加权源域样本,筛选出与目标域分布相似的样本参与训练。为应对更复杂的工况分布差异(即域偏移),域自适应方法被广泛研究,其核心是缩小源域与目标域在特征空间中的分布差异。例如,基于MMD的方法通过最小化域间距离来学习域不变特征[29],基于对抗学习的方法引入域判别器,驱动特征提取器产生混淆域判别器的特征,更精细的方法如分层交替迁移学习和基于局部中心矩差异(Local Central moment discrepancy,LCMD)的类别级对齐[33],致力于在更细粒度上实现特征分布的对齐,提高在变负载、变转速等复杂迁移场景下的诊断准确率和鲁棒性。

### 1.4 基于卷积神经网络的故障诊断

卷积神经网络凭借其强大的空间特征提取能力,已成为机械故障诊断领域的主导架构。在轴承与齿轮箱诊断领域,一维卷积神经网络和长短期记忆网络的混合架构展现出卓越性能。这种架构能够直接从原始振动信号中提取时空特征,无需依赖传统信号处理中的手动特征工程。在强噪声和变负载条件下,此类模型仍能保持极高的识别准确率。特别值得关注的是,基于自适应归一化卷积神经网络的方法在行星齿轮箱诊断中通过Teager能量算子预处理[34]有效抑制了转速波动干扰,解决了变工况下的诊断难题。在输配电网络保护领域,基于一维卷积神经网络的架构可直接处理同步相量测量单元采集的电压电流信号,在含分布式电源的主动配电网中实现较为准确的故障分类,且响应时间较短,完全满足继电保护速动性要求[35]。这种方法的创新之处在于其能够有效应对分布式发电机接入导致的故障电流水平变化,解决了传统保护方案面临的挑战。

然而,原始CNN在处理复杂工业信号时会遭遇信号非平稳性、噪声干扰以及故障特征多尺度等诸多挑战,鉴于此,研究人员开发出多种CNN变体与混合架构,对故障特征的感知与诊断能力有了一定提升。

一方面,CNN的架构不断演进以契合工业信号特性:一维卷积神经网络可直接处理原始振动信号,达成端到端诊断,例如借助引入残差学习在强噪声环境下保持了较高的准确率[36],借助多通道输入模拟图像处理,可充分运用故障信息的空间相关性[37];二维卷积神经网络借助短时傅里叶

变换[38]或者希尔伯特-黄变换[39]将时序信号转变为时频图像，利用CNN在图像识别方面的强大性能；多尺度卷积神经网络凭借并行运用不同尺寸卷积核来同时捕获故障的局部细节与全局趋势，比如基于选择性核块构造神经网络去噪的时间尺度自适应的卷积神经网络（Novel time-scale adaptive CNN with neural network denoiser constructed based on the selective kernel block，SKND-TSACNN）能自适应关注不同尺度特征[40]，多尺度CNN与LSTM的混合模型[41]依靠特征融合实现了较好的噪声鲁棒性。

另一方面，将CNN与其他网络模块相结合，混合架构突破了单一模型的限制。CNN-LSTM混合架构融合了CNN提取空间特征的能力以及LSTM进行时间序列建模的能力，适合处理动态时序信号。将其与迁移学习相结合可提升变工况下的泛化能力[42]，运用该架构可有效应对复杂海洋环境下的噪声干扰[43]，同时还可用于工业过程的故障预测。CNN-Transformer混合架构借助Transformer的全局自注意力机制，弥补CNN在长距离依赖建模方面的不足，比如凭借双头集成Transformer（Dual-head ensemble transformer，DHET）生成高质量故障样本来改善数据不平衡和特征不明显问题[44]，多尺度Transformer-CNN域适应网络凭借特征对齐解决了不同运行模式下的故障诊断问题[45]。另外，残差网络的引入有效解决了深层CNN训练时的梯度消失和网络退化问题，例如，依靠捷径连接构建的深层一维残差CNN[36]，在强噪声诊断中表现良好；在涡轮风扇发动机容错控制研究中采用残差思想[46]，可实现对执行器故障的高精度诊断，为后续控制提供可靠基础。

### 1.5 基于现代深度学习方法与技术的故障诊断

深度学习基础方法的演进呈现出明显的趋势：从完全监督到半监督/无监督，从独立学习到迁移学习，从单一模态到多模态融合。深度学习的此种演进则使得故障诊断系统能够更好地适应工业现场的数据稀缺、工况多变、噪声干扰等实际挑战。然而，基于深度学习的故障诊断在实际应用上仍存在着诸多问题，表1为对近些年基于深度学习的故障诊断综述的总结。

**表1 近年综述的贡献与局限**

**Table 1　Contributions and limitations of recent reviews**

| 文献 | 主要贡献 | 局限性 |
|---|---|---|
| [47] | 系统梳理了监督学习的基本原理及其在健康监测中的应用 | 未深入探讨数据稀缺、工况多变等工业现实问题下的模型适应性 |
| [48] | 指出了在半监督学习中标注不足时盲目引入无标签数据的风险 | 未充分扩展到深度学习在故障诊断中的实际应用场景 |
| [49] | 强调了迁移学习在跨工况、跨设备诊断中的潜力 | 对其与深度学习的深度融合机制缺乏系统性总结 |
| [50] | 提出的Transformer架构及其自注意力机制在自然语言处理领域取得突破 | 在故障诊断时序信号建模中的适用性与局限性有待探讨 |
| [51] | 为故障数据增强提供了新思路 | 实际效能与工程落地路径不明确 |

对于发动机以及涡轮机械这类高速旋转的设备而言，将残差网络和注意力机制相结合的深度学习模型呈现出了独特的优势，这些模型可精确识别叶片裂纹、转子不平衡等早期故障，在仿真数据里达成了较为准确的分类精度，在航空发动机故障诊断方面，基于气路数据的卷积神经网络模型针对作动器故障的诊断准确率有了显著的提高，比传统多层感知器模型要好很多。这种性能的提升主要是因为深度学习模型对复杂非线性关系有着强大的拟合能力，在注意力机制的技术发展过程中，混合域注意力架构呈现出了独特的优势：残差混合域注意力CNN[52]创新性地将时间域注意力和通道域注意力并行整合，这种双注意力机制在强噪声环境下的轴承诊断中表现出优越的鲁棒性，即便在低信噪比环境中依然可保持很高的诊断准确率。一项研究探寻了把先验知识融入注意力机制的新路径，该方法利用皮尔逊相关系数构建类别-属性相关矩阵，将其当作注意力生成的指导信息，使CNN模型在化工过程故障诊断中可自动聚焦于与特定故障模式最相关的工艺参数[53]，这种知识引导的注意力机制让诊断准确率有了一定程度的提升，而且还运用了可视化技术，清楚地展示了模型的决策依据，为破解深度学习“黑箱”难题提供了有价值的方向。注意力机制的发展正在促使故障诊断系统持续演进，在提升性能的同时还提高了可信度。

在生成模型这个方向上，有研究提出基于平衡串联卷积神经网络与变换器的生成对抗网络，

它代表了当下的技术前沿,该模型的生成器巧妙地把CNN提取局部特征的优势和Transformer捕捉长程依赖的能力融合在一起,能为极端不平衡的轴承数据集生成多样且真实的故障样本。实验显示,在故障样本与正常样本比例很高的极端情形下,该方法依然能维持较高的诊断准确率,比传统的过采样技术要好很多。另一条技术路线探索了变分推理与对抗训练的融合,基于变分自编码器提高的卷积神经网络模型借助VAE在数据生成和噪声鲁棒性方面的特性,在潜在空间里进行数据插值和扰动,有效增加了训练样本的多样性[28]。这种方法很适合处理早期微弱故障的诊断,针对小样本学习的特定要求,条件生成对抗网络依靠引入类别标签作为生成过程的控制条件,保证了生成样本的类别准确性,研究说明,GAN生成的轴承故障样本与真实样本在概率密度函数上呈现出高度相似性,为在有限数据条件下训练可靠的诊断模型奠定了坚实基础。

此外,自动化机器学习(Automated machine learning,AutoML)也是现代深度学习技术的重要方法之一。神经架构搜索(Neural Architecture search,NAS)作为AutoML的核心技术,在故障诊断领域展现出巨大潜力。一项研究探索了基于NAS的自动化流程用于设计反应堆冷却剂泵故障诊断的混合CNN模型[54]。该方法能够自动探索1D-CNN和2D-CNN的最佳组合方式,避免了依赖先验知识进行手动架构设计的局限性。最终搜索得到的混合模型在密封泄漏故障诊断中取得了非常高的最高准确率,同时将模型开发周期从数周缩短至数天。一种基于非支配排序遗传算法II(Non-dominated sorting genetic algorithm II,NSGA-II)的CNN结构自动搜索方法,可同时以诊断准确率和模型复杂度为优化目标[55],这种方法能够为特定的故障诊断任务生成一组在性能与效率之间达到最佳平衡的网络架构,可根据实际部署环境的计算资源约束选择最适合的模型。实验证明,自动搜索得到的网络在保持高精度的同时,计算复杂度可比人工设计的经典网络降低一个数量级。自动化机器学习的发展标志着故障诊断正在从“手艺活”转变为“标准化流程”,为大规模工业应用铺平了道路。

### 1.6 多类与多标签故障诊断

前面的章节重点是针对单故障模式的诊断进行探讨。不过,在实际的工业系统当中,会出现多种故障同时出现的情况,或者是单一故障呈现出来多种属性,这种状况促使了多类与多标签故障诊断方面的研究兴起。多类诊断属于经典的单标签多分类问题,意思就是要从多个相互排斥的故障类别里找出唯一正确的那一类。而多标签诊断则要复杂一些,它的目标是识别一个样本里同时存在的多种故障模式或者属性,这些标签之间通常不是相互排斥的,而且存在着依赖关系。深度学习方法给解决多标签故障诊断问题提供了有效的工具。它的核心挑战在于怎么能够对标签之间的相关性进行建模,并且学习到可以同时表征多种故障模式的复合特征。主要的技术路径有基于问题转换的方法、基于算法适配的方法、基于图神经网络与结构化预测的方法。

基于问题转换的方法把多标签问题转变为多个独立的二分类问题,或者采用标签组合策略。这种方法虽然直接,然而却忽略了标签之间的关联性,并且当标签空间较大时计算成本较高[56]。基于算法适配的方法直接设计或者适配神经网络来输出多个标签。常见的做法是在共享的特征提取网络之后,为每个标签设置独立的分类头,同时借助改进的损失函数进行联合优化。注意力机制在这类任务中非常有效,能够引导模型关注与不同故障标签相关的特征区域[57]。基于图神经网络与结构化预测的方法明确地对标签间的依赖关系进行建模。比如,把故障标签当作图节点,利用图神经网络来学习标签的联合表示,进而在推理的时候综合考量标签相关性,这对于存在强逻辑关联的复合故障诊断具有优势。多任务学习框架将不同类型标签的预测视为相关但不同的任务,通过共享底层特征、设计特定任务子网络的方式一同学习,达成知识的正向迁移,提高整体诊断效率[58]。

虽然多标签故障诊断研究有了一些成果,然而面临的挑战依旧严峻。在数据方面,同时标注多种故障的样本获取成本非常高,而且标签组合的长尾分布问题很突出。在模型方面,平衡不同标签预测任务之间的干扰、设计能捕获复杂标签关系的轻量级网络结构是关键所在。在评估方面,要超越单一准确率的指标模式,采用适合多标签场景的指标,比如汉明损失、子集精度、排序指标等。未来,融合物理先验知识来约束标签间的可能组合、开发小样本下的多标签诊断方法,还有

探索更具解释性的多标签预测模型,是推动该领域走向实际应用的重要方向。

## 2　深度学习关键技术创新

随着深度学习于机械故障诊断领域的运用,研究者们渐渐察觉到,仅仅凭借增加网络深度或者参数规模,已难以有效应对工业现场所面临的核心难题,真实的工业环境一般存在标注数据匮乏、故障模式繁杂多样、设备工况动态变动以及计算资源受限等棘手状况,为突破这些困境,近些年来故障诊断领域出现了一系列关键技术革新,在模型架构、训练策略以及数据利用等多个层面推动了该领域的实质进步。这些技术提高了诊断系统的性能指标,还提高了其在复杂工业场景中的实用性与可靠性。

### 2.1　迁移学习:破解小样本诊断困境

在工业实践的环境当中,要获取大量有高质量且带有标签的故障数据,一般会遭遇极大的挑战,一方面,关键设备在正常运行的阶段很少出现故障,这使得故障样本在自然状态下十分稀缺,另一方面,在特定工况之下收集到的数据,大多时候难以直接应用于其他的运行条件。迁移学习的核心价值在于,它可把从一个相关领域学到的知识,有效地转移到数据稀缺的目标领域,较大降低对目标域标注数据的依赖程度(见图2)[59]。

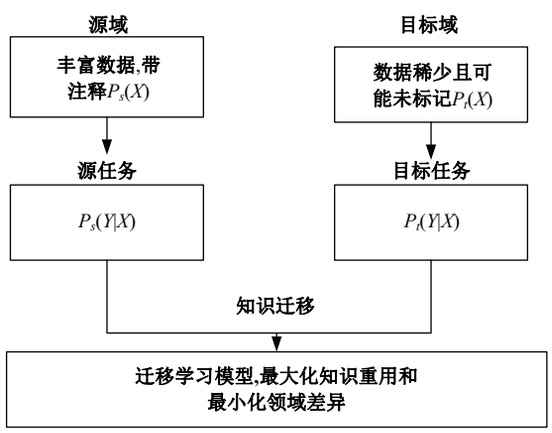



图2　迁移学习过程

**Fig. 2　Transfer learning process**

迁移学习的核心原理是风险上界理论:

$$R_\tau(h)\leqslant R_s(h)+\mathrm{dist}(D_s,D_\tau)+\lambda \quad (1)$$

式中:$R_\tau(h)$为目标域;$R_s(h)$为源域风险;$\mathrm{dist}(D_s,D_\tau)$为域间差异;$\lambda$则是理想常数。

为了在目标域$R_\tau(h)$上获得良好的性能,需要找到一个在源域上表现良好的模型,同时确保该模型学到的特征在源域和目标域之间是对齐的。为直观展示特征空间对齐原理,采用合成数据进行可视化示意[49](见图3)。

迁移学习借助知识迁移应对了工业故障诊断中标注数据稀缺这一关键挑战。当下主流方法能够归纳为基于差异度量、对抗学习、特征重构、直接微调等一些范式。基于差异度量的方式通过让源域与目标域特征分布差异最小化来达成对齐目的,其原理明了、计算高效,然而可能因全局对齐而忽视领域特有的特征,甚至导致负迁移现象出现,适用于工况变化幅度较小的情形[29]。基于对抗学习的方式引入域判别器,促使特征提取器去学习域不变表示,进而能够适应更为复杂的分布变化,不过存在训练不稳定、需要精细调参这样的挑战,适合从稳态到瞬变工况等差异明显的迁移任务。基于特征重构的方式明确地解耦域不变与域特定特征,从根源上避免了负迁移问题,只是网络结构复杂,对数据量、训练技巧有着较高要求,适用于异构传感器或跨设备的知识迁移[31]。直接微调预训练模型是一种简单有效的特征迁移策略,计算成本低。然而,在源域跟目标域差异过大时,可能会过拟合小样本目标数据,它适合具有相似数据模态的快速适配[26]。迁移学习的选择要权衡工况差异程度、数据相似性、可用计算资源、对模型可解释性的要求,它与物理机理深度融合是提高跨域泛化能力的关键。

迁移学习作为一种机器学习方法,借助源领域知识来提高目标领域模型性能,它的核心优势是凭借特征迁移大幅减少对目标领域标注数据的依赖,提升模型泛化能力,不过它仍存在负迁移现象、分布差异校正等关键问题有待优化,在处理异构特征空间和条件分布偏移时,更要保证其有效性与通用性。

### 2.2　注意力机制:从粗放感知到精准聚焦

注意力机制借助模拟人类选择性关注信息的能力,在深度学习中达成了对关键信息的动态加权,其代表性算法覆盖Transformer的自注意力、压缩与激励网络(Squeeze-and-excitation networks,SENet)的通道注意力以及分层注意力网络(Hierarchical attention networks,HAN)的分层注意力等,核心优势在于提高模型性能、强化可解释性以及处理长序列依赖,不过仍存在计算效率、解释可靠性以及跨领域适应性等问题,有待

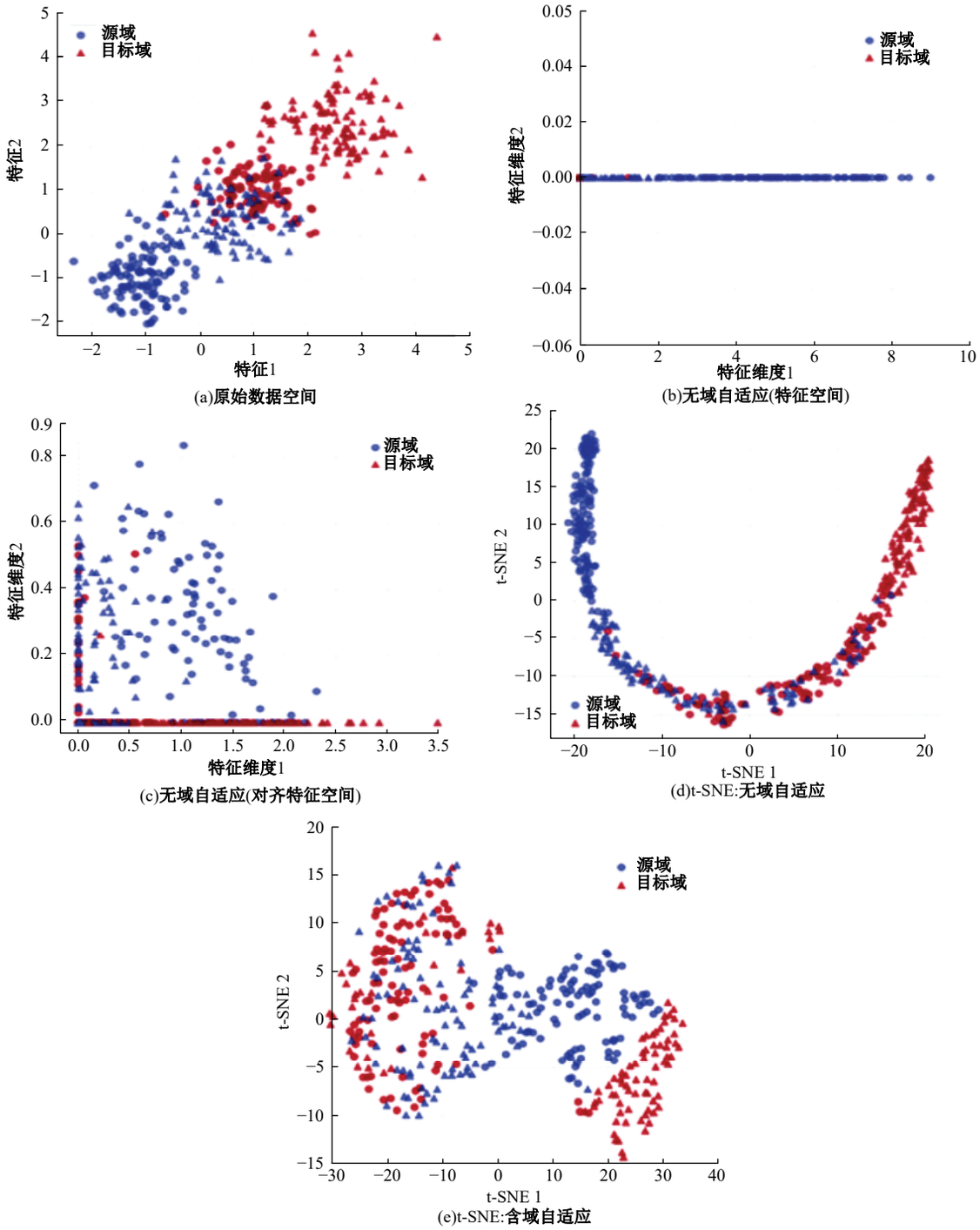



图3 迁移学习中的特征空间对齐

**Fig. 3 Feature space alignment in transfer learning**

改进[60]。

注意力机制有较强的可解释性,可借助生成权重矩阵直观呈现模型的关注点,并且还可以直接观察到哪些token之间存在强关联。另外,注意力机制拥有出色的并行性与适应性,可有效规避顺序依赖、特定长度或结构依赖,适配不同输入输出的任务。注意力机制还有较强的全局信息整合能力,可以从每个位置获取全局信息,同时对任意距离的依赖关系进行建模,可从大量信息中提取最为相关的部分。采用合成数据展示注意力机制的应用实例[61],如图4所示。

注意力机制通过对关键信息进行动态加权,有效提高深度学习模型在故障诊断方面的特征选择能力与可解释性。依据其作用域和原理,主要

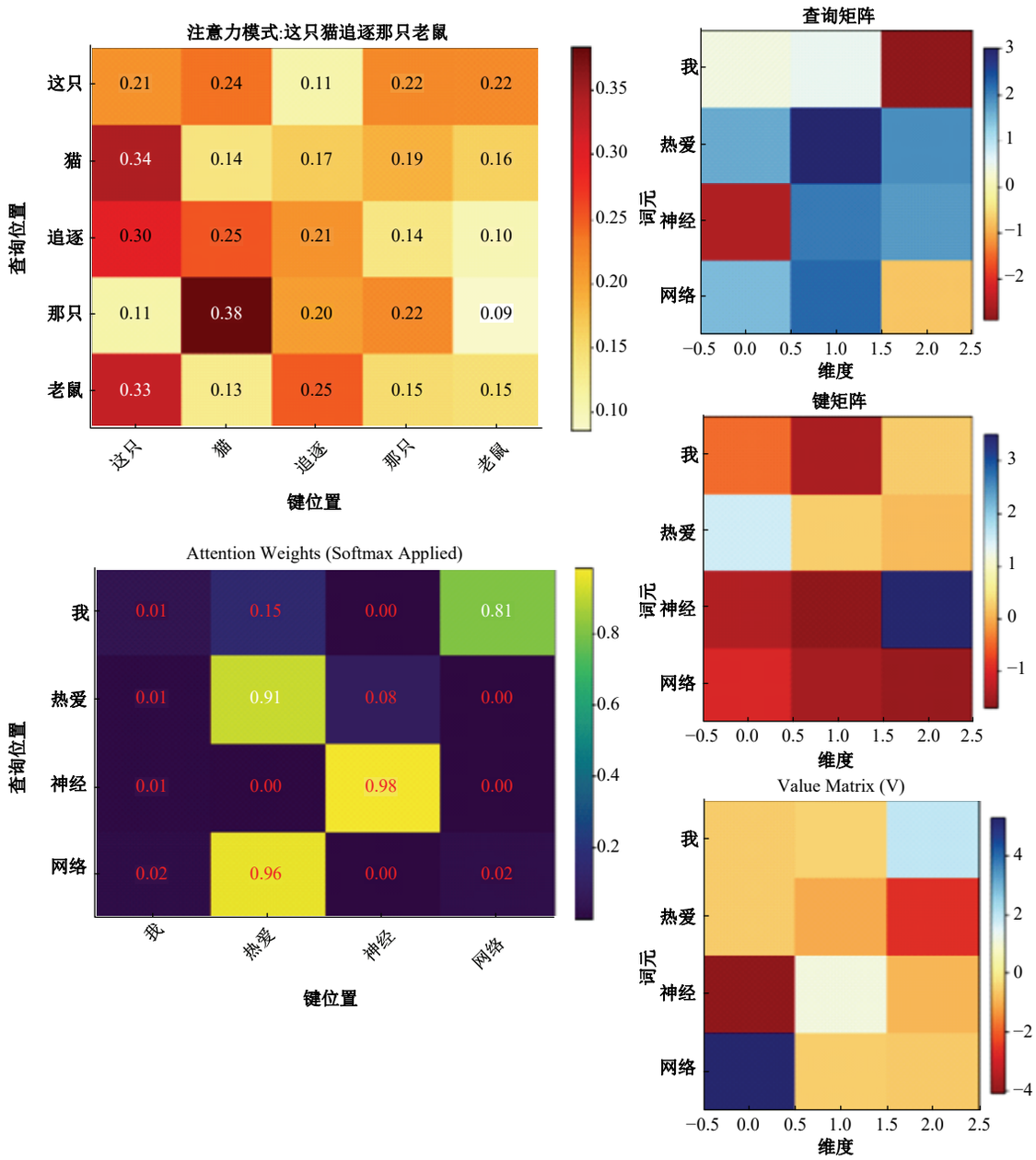



图 4　注意力机制应用示例

**Fig. 4　Application examples of attention mechanism**

可被分为自注意力、通道注意力、空间注意力、它们的混合变体。全局自注意力[50]借助计算序列内所有位置之间的关联权重捕捉长程依赖,其具备较强的建模能力、较高的并行效率,然而计算复杂度会随着序列长度呈平方增长,适用于对需要全局上下文建模的长序列振动信号展开分析。通道注意力会去学习各个特征通道的重要性,能够以较少的参数量增强判别性特征并抑制冗余,比较容易集成到现有的CNN当中,不过有可能会过度关注少数通道,适用于针对多通道传感器数据进行特征筛选、轻量级模型设计。空间注意力将关注点放在特征图里的关键区域,提高模型对于空间的感知能力,还能给出一定程度的可视化解释,不过它对噪声比较敏感,常被用于时频图像[38,39]或者红外热像图里的故障区域定位。混合域注意力[52]把通道注意力和空间注意力并行结合起来,能够更把握特征的重要性,在强噪声、复杂工况的情况下展现出出色的鲁棒性,只是其结构相对复杂一些。分层注意力机制适用于有层次化结构的信号(像是多尺度故障特征)。各类注意力机制共有的优势是模拟了人类专家的聚焦思维,让模型资源集中到最具判别性的信息片段上,

未来研究要进一步设计轻量化模块,还要探索注意力权重与物理故障机理之间的因果联系。

注意力机制的引入实际上模拟了人类专家的诊断思维——把有限的计算资源优先配置给最具判别性的信号片段和特征通道,实现从“粗放感知”到“精准聚焦”的范式转变。

### 2.3 生成式对抗网络:数据稀缺困境突破

GAN的核心优势是,即便没有显式分布假设,也可生成逼真的样本,同时还支持半监督学习。本节采用MNIST手写数字数据集进行训练,直观展示GAN的优势[62],图5为复杂分布的直接学习与半监督分类效果。

GAN借助生成器与判别器之间的对抗训练达成高质量数据生成(见图6~图8),其有代表性的算法囊括了WGAN(Wasserstein GAN)、CycleGAN以及StackGAN等。

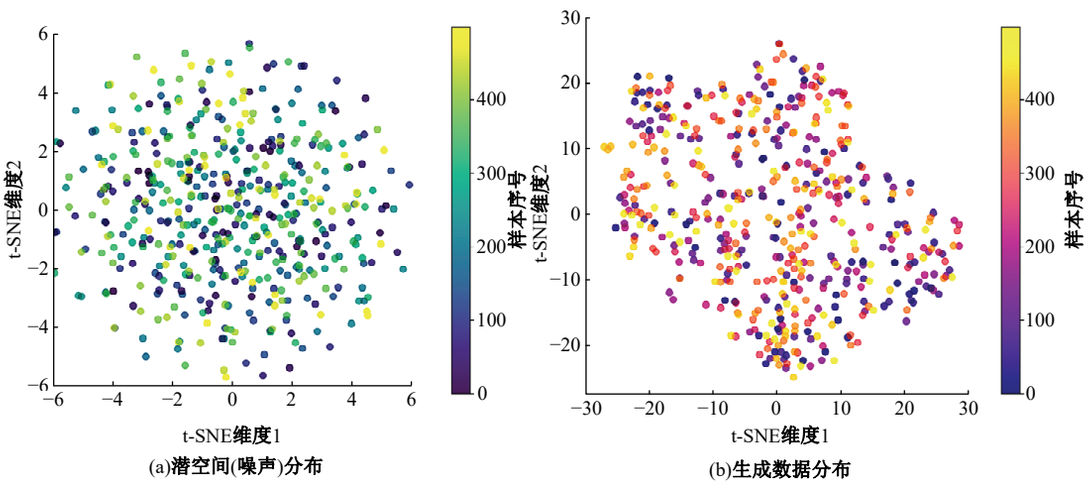



图5 复杂分布的直接学习与半监督分类效果

**Fig. 5 Direct learning and semi supervised classification effect of complex distribution**

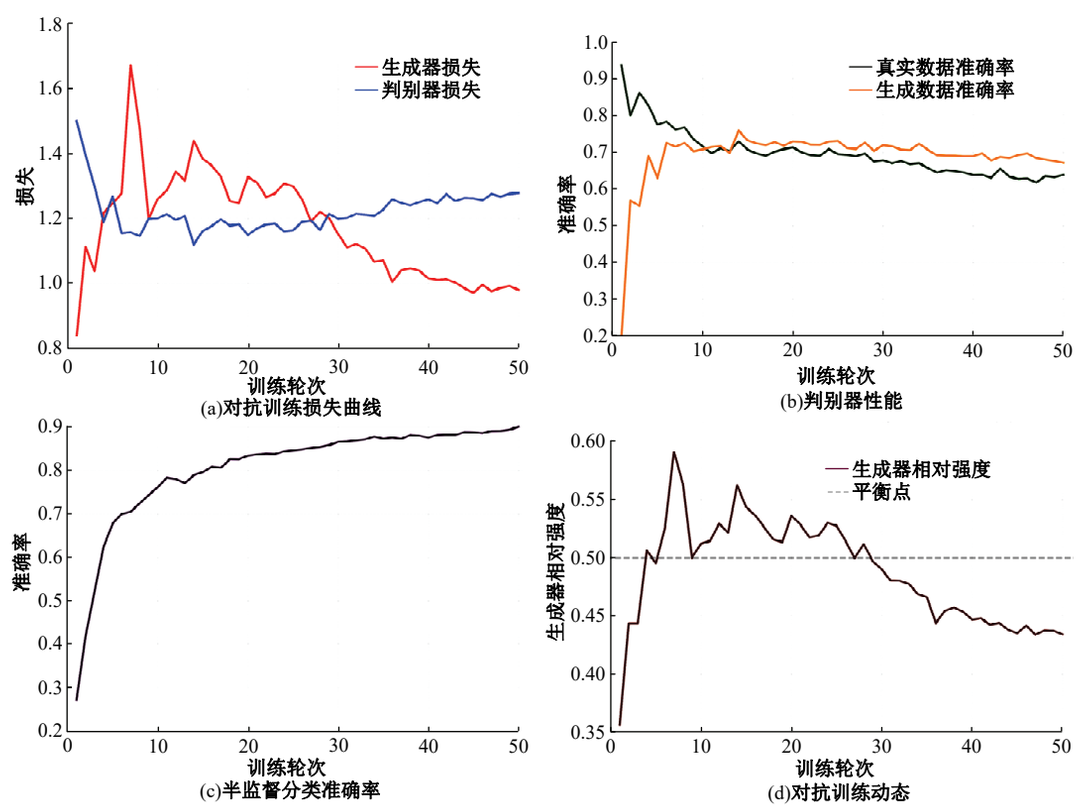
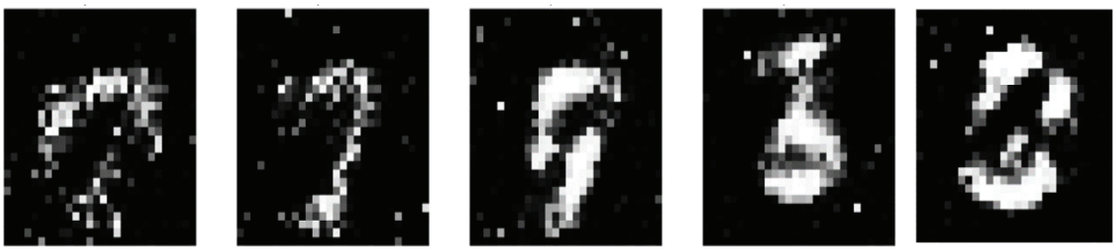



图6 生成器与判别器的对抗训练过程

**Fig. 6 Confrontation training process of generator and discriminator**



(a)训练轮次10　(b)训练轮次20　(c)训练轮次30　(d)训练轮次40　(e)训练轮次50

**图7　样本生成演进过程**

**Fig. 7　Sample generation evolution process**

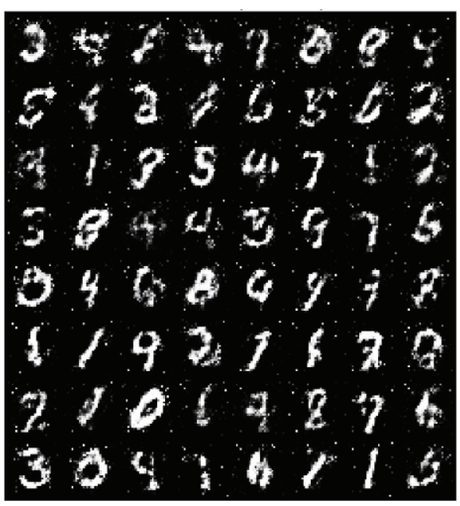


**图8　最终生成样本**

**Fig. 8　Final generated sample**

GAN的目标函数如下:

$$\min_G \max_D V(D,G)=E_{x\sim p_{\text{data}}(x)}\left[\log D(x)\right]+E_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \tag{2}$$

式中:$\min\limits_G \max\limits_D V(D,G)$为生成器$G$与判别器$D$之间的极大极小博弈;$E_{x\sim p_{\text{data}}(x)}$为对来自真实数据分布$P_{\text{data}}$的样本$x$取期望值;$E_{z\sim p_z(z)}$表示对来自先验噪声分布$P_z$的噪声向量$z$取期望值;$D(x)$为判别器对真实样本的估计概率;$D(G(z))$为判别器对生成样本的估计概率。

其中,判别器$D$试图最大化$V(D,G)$,即准确区分真实数据$x$与生成数据$G(z)$;而生成器$G$则试图最小化$V(D,G)$,即让$D(G(z))$接近1,从而最小化$\log(1-D(G(z)))$。实践中常采用梯度更加友好的等效目标:$\max_G E_{z\sim p_z}[\log D(G(z))]$,即$-\log D$技巧:它的目标不再是让$\log(1-D(G(z)))$最小,而是让$\log D(G(z))$最大。在训练初期的时候,$D(G(z))$是非常小的,这就导致了原目标的梯度极其平缓,难以学习,而这个新目标则能够提供更加强烈的梯度信号[63]。

工业设备故障诊断一直以来都面临着数据不平衡的问题,生成对抗网络技术的出现为这一挑战给予了有变革性的解决办法,它借助博弈学习的方式生成在统计分布上与真实故障样本高度一致的合成数据,达成了对稀缺故障模式的数据提高(见图9)。

生成对抗网络及其变体借助博弈学习来生成逼真样本,为故障诊断里的数据不平衡、稀缺问题的解决给出了创新办法。在标准GAN框架下,生成器和判别器展开极小极大博弈,其概念简单并且能够产生多样样本,不过却被训练不稳定、模式崩溃等问题所困扰。WGAN运用Wasserstein距离、梯度惩罚机制,改善了训练稳定性,让损失函数和生成质量相关联,然而计算成本较高,适用于对生成样本保真度要求严格的情形。条件GAN[27]通过把故障类别标签当作控制条件引入,达成了针对特定故障类型的可控生成,能够有效处理极端数据不平衡问题,不过其生成多样性可能会受条件信息的限制。循环一致GAN借助循环一致性损失达成无配对数据的域间转换,这为跨工况数据模拟创造了条件,不过也有可能带来无关失真。再者,把GAN和变分自编码器相结合的VAE-GAN[28],融合了VAE的规整潜在空间、GAN的高质量生成能力,适宜在噪声环境里生成早期微弱故障数据,只是训练难度偏高。这些方法不但能够直接扩充训练集,还能够和诊断模型进行联合或者对抗性训练,进而提高小样本情形下的模型泛化能力。未来,要让生成样本契合物理故障演化机理、建立可靠的生成质量评估标准、研发更高效的训练策略,这是把GAN可靠应用于工业现场的关键所在。

### 2.4　自动化机器学习:从经验依赖到自主优化

深度学习模型所呈现出的效果,在相当程度上取决于网络架构的设计以及超参数的配置情

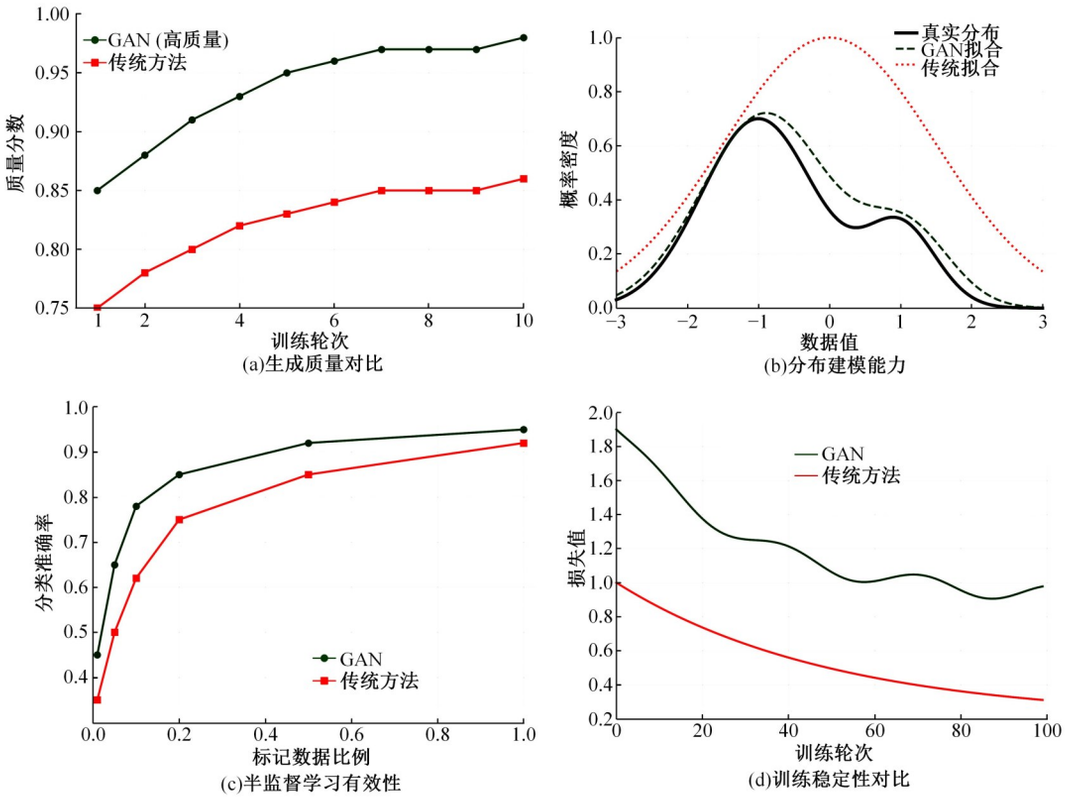



**图9 GAN与传统方法比较**

**Fig. 9 GAN vs traditional methods comparison**

况,在过往,这个过程往往需要借助大量专家的经验以及计算资源,依靠不断试错来实现优化,然而自动化机器学习,也就是AutoML技术,其目的在于把这个过程转变为系统化且自动化的模式,降低技术方面的门槛,还可发现那些超越人工设计性能的模型结构。

AutoML把构建机器学习流水线的整个过程,转化成为一个针对混合搜索空间的优化问题,以便找寻到一个最优配置$C^*$,该配置覆盖了算法选择,这属于离散范畴,以及超参数,超参数包含连续和离散两种类型,以此来让模型在验证集上的损失达到最小:

$$C^* = \arg\min_{C \in CS} L\left(P(C), D_{\text{train}}, D_{\text{val}}\right) \tag{3}$$

式中:$CS$为所有可能配置所构成的组合空间;$P(C)$为依据配置$C$在训练集上进行拟合得到的流水线;$L$为用于评估性能的损失函数。

此空间有高维的特性,呈现出非凸的状态,并且评估代价颇为高昂。要解决评估代价高昂这一问题,AutoML的核心在于贝叶斯优化,它会构建一个代理模型用以近似真实的损失函数$L(C)$,还会定义一个采集函数来平衡探索和利用,可智能地挑选下一个待评估的配置,最常被使用的采集函数是期望改进,假设当前最佳观测损失是$L^*$,对于新配置$C$,其改进量的期望值如下:

$$EI_t(C) = E\left[\max\left(0, L^* - L(C)\right)\right] = \left(L^* - \mu_t(C)\right)\Phi(Z) + \sigma_t(C)\phi(Z) \tag{4}$$

式中:$\mu_t$和$\sigma_t$分别为代理模型所给出的预测均值与标准差;$Z = \dfrac{L^* - \mu_t(C)}{\sigma_t(C)}$;$\Phi$和$\phi$分别为标准正态分布的累积分布函数以及概率密度函数。

系统借助迭代的方式实现最大化期望改进,逐步接近最优配置。

鉴于穷举搜索并不具有可行性,自动机器学习运用动态资源分配策略以便迅速淘汰质量欠佳的配置,把计算资源集中于有潜力的候选者之上。逐次减半便是这一理念的典型呈现:在每一轮$i$中,针对$n_i$个存活的配置分配预算$b_i$展开评估,仅仅保留性能最为优良的$1/\eta$比例进入到下一轮,并且增加其预算:

$$\begin{cases} n_{i+1} = \left\lfloor \dfrac{n_i}{\eta} \right\rfloor \\ b_{i+1} = \eta \cdot b_i \end{cases} \tag{5}$$

式中：$\eta$为削减因子，其数值一般设定为3。

整个过程会持续进行，直至预算全部用完或者仅剩下一个配置，以此在有限的总预算条件下，尽可能提高找到良好配置的概率[64]。

自动化机器学习致力于把模型设计、超参数调优等经验性流程予以系统化、自动化，使得深度学习在工业故障诊断里的应用门槛得以降低。神经架构搜索[54]能够于预设搜索空间内自动探寻最优网络结构，有可能找出超越人工设计的架构，不过其需要极高的计算成本，适用于为复杂或新型故障模式探寻定制化模型。超参数优化方法借助建立代理模型来智能地探寻超参数组合，能够有效提高模型性能并且减少人工调参时间，然而有可能陷入局部最优，适宜在新设备或新工况下快速优化模型配置。动态资源分配策略[64]通过早期淘汰表现欠佳的配置，把有限计算资源集中于有潜力的候选者，能够大幅缩短自动搜索的总时间，尤其适合在计算资源受限的边缘侧进行快速模型开发。元学习方法致力于“学会学习”，借助在先验任务上获取的经验，达成对新诊断任务的少样本快速适配，但其效果取决于元训练阶段任务分布的相似性。这些自动化技术标志着故障诊断模型开发从“手艺活”朝着“标准化流程”的转变。未来的趋势是开发计算效率更高的搜索算法、把领域知识融合到搜索空间设计中，并且推动轻量级 AutoML 在边缘设备上的部署，最终达成诊断模型的自适应演进与调整。

AutoML 能够将人工的机器学习设计经验，转化为一个由“代理模型”和“采集函数”驱动的、具备“动态资源分配”能力的自动化元优化系统。正是由于它的此种特点，AutoML 流程中的算法选择、超参数优化和架构设计等使得非专家也能够构建高性能模型。除此之外，通过搜索空间定义、搜索策略和性能评估等组件，AutoML 也可以使机器学习使用门槛显著降低并提升开发效率，同时能够更好地发现人类专家可能忽略的优化方案。

上述提到的种种关键技术并非孤立存在，而是呈现出深度融合的发展趋势。迁移学习为小样本场景提供了知识基础，注意力机制确保了有限计算资源的高效利用，生成对抗网络解决了数据稀缺的根本问题，而自动化机器学习则优化了整体解决方案的设计过程。未来的智能诊断系统可构建于这样的技术生态之中：先是借助自动化机器学习为特定设备类型迅速设计基础模型架构，接着运用迁移学习让其适配至具体工况，随后结合生成对抗网络扩充稀缺故障样本，最后借助注意力机制提高在强噪声环境下的特征选择能力，这种技术融合会推动诊断精度提高，还可以让诊断系统拥有自我演进的能力，也就是可依据设备运行数据的积累持续优化自身性能，适应设备老化以及工况变化带来的挑战，最终达成真正意义上的自主智能运维。

### 2.5　综合性能对比

为了评估深度学习在工业故障诊断里关键技术的性能差异，本节设计了一个综合性对比实验，该实验是以光伏系统故障图像数据为基础进行的。实验将光伏系统故障数据集当作测试平台，针对迁移学习、注意力机制、生成对抗网络这3种具有代表性的技术进行多维性能评估。这个数据集涵盖了4种光伏故障类型，分别是线间故障（LL）、短路故障（SC）、局部阴影叠加线间故障（PS＋LL）、局部阴影叠加短路故障（PS＋SC），一共有480个二维图像样本，每一种类型有120个样本，所有样本都是一维故障信号经过线性小波变换之后所形成的时频图像。所有方法均采用相同的预处理流程和超参数设置。图像统一调整为128×128像素，采用ImageNet标准化参数进行预处理。数据集划分遵循6∶2∶2的比例（60%训练、20%验证、20%测试）。迁移学习实验中，将LL和SC两类故障（240个样本）作为源域，PS＋LL和PS＋SC两类故障（每类仅使用12个样本，共24个样本）作为目标域，模拟小样本跨故障类型迁移场景；注意力机制实验中，在标准CNN架构基础上集成压缩与激励（Squeeze-and-excitation，SE）注意力模块；GAN实验中，人为构造10∶1的数据不平衡比例，模拟工业中罕见故障的实际分布。3种方法的基干网络均采用具有可比性的CNN架构：3个卷积块（通道数分别为32、64、128）、全局平均池化层和全连接分类层（4个输出单元）。迁移学习采用两阶段训练策略（源域预训练80轮次，目标域微调80轮次）；注意力CNN在卷积块后插入SE注意力模块；GAN采用条件生成架构，生成器和判别器均基于卷积网络设计。

表2呈现出来3种方法于光伏故障数据集上

的综合性能表现情况。注意力CNN在具备充足数据的条件时表现比较稳定,其训练准确率能够达到93.40%,验证准确率为76.04%,这显示出它在标准数据状况下具有一定可靠性。迁移学习在目标域数据极其稀缺,也就是仅使用10%的原始样本这种情形下,依旧取得了95.83%的验证准确率,明显比从零开始训练的基线模型要优异,这验证了它在小样本场景下具备卓越的有效性。基线CNN在数据不平衡条件,即不平衡比例为10:1的状况下,最终达到了96.15%的验证准确率,不过在训练过程中呈现出来较大的波动性,这反映出它对数据分布敏感的固有特性。

**表2 综合性能对比**

**Table 2 Comprehensive performance comparison**

| 评估指标 | 基线CNN(不平衡) | 注意力CNN | 迁移学习 |
|---|---|---|---|
| 最终训练准确率 | 0.974 7 | 0.934 0 | 0.995 8(源域)/0.875 0(目标域) |
| 最终验证准确率 | 0.961 5 | 0.760 4 | 1.000 0(源域)/0.958 3(目标域) |
| 训练轮次 | 80 | 80 | 80+80 |
| 训练稳定性 | 中等(波动较大) | 良好 | 优秀 |
| 小样本适应性 | 差 | 中等 | 优秀 |

能够发现,不同方法在性能和稳定性之间呈现出来比较明显的权衡趋势。迁移学习于跨域适应能力方面变得非常突出,在源域预训练阶段就达到了接近完美的100%准确率,并且在仅仅使用少量目标域样本的情形下能够迅速适应,最后达成了95.83%的验证准确率。注意力机制在标准数据条件时展现出不错的训练稳定性,不过验证准确率相对较低,为76.04%,这或许说明其在当前数据集上存在一定程度的过拟合倾向。基线CNN尽管最终达到了较高的准确率,但其在训练过程当中出现的剧烈波动,以及验证准确率从88.46%下降至44.23%之后又回升到96.15%,体现出它对超参数、数据分布具有高度的敏感性。

由图10能够看出,GAN训练呈现出来典型的对抗性振荡特性。生成器损失从初始的0.887 6开始逐步上升,一直攀升到1.709 3,而判别器损失则从0.636 6下降至0.412 3,在150轮训练周期里,二者一直在进行对抗博弈。损失函数这种反向变化的趋势是GAN训练本身具有的特性,也就意味着,随着生成器能力不断提高,判别器进行区分任务时难度会变得更大,进而导致其损失上升,生成器为了生成更逼真的样本去欺骗逐渐强大的判别器,其损失也可能会相应增加。训练在第120轮之后趋向于相对稳定的状态,这表明生成器和判别器达到了一种动态平衡。

大量实验结果显示,迁移学习在小样本跨域场景里表现出色,源域预训练准确率能够达到

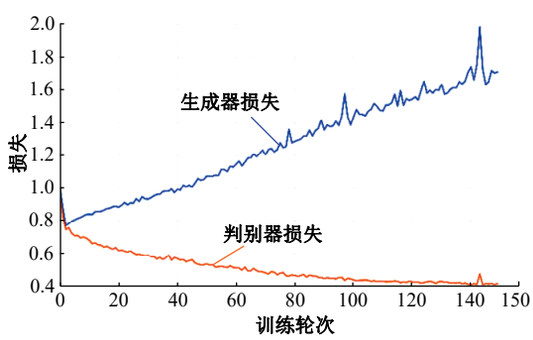



**图10 GAN训练损失曲线**

**Fig. 10 GAN training losses curve**

100%,目标域微调(只有24个样本)时准确率是95.83%,这证实其拥有强大的知识迁移能力,它的核心优势是极大降低对目标域标注数据的需求,从而实现快速域适应,不过它依赖源域与目标域的分布相似性,并且存在负迁移风险,双阶段训练还增加了计算复杂度。注意力机制借助SE模块提供了有价值的模型可解释性,特征权重可视化显示出模型对高频故障特征的关注,增强了诊断决策的透明度,然而在此次实验中其验证准确率仅为76.04%,并且训练过程出现异常波动,这表明注意力机制在提高模型鲁棒性和泛化能力方面仍需要优化,额外的参数开销可能也会影响部署效率。生成对抗网络为解决数据不平衡问题提供了创新方式,对抗训练动态(生成器损失1.709 3,判别器损失0.412 3)展现出其博弈学习特性,GAN能够生成逼真的故障样本以扩充少数类数据(见图11),但是训练不稳定、模式崩溃、生成质量评估困难等技术挑战限制了其在工业中的

应用,150轮训练周期也带来了较大的计算负担。从整体情况来看,迁移学习在数据稀缺的场景当中具有比较突出的优势,不过它对域相关性存在一定的依赖,注意力机制能够增强可解释性,然而在性能提高方面相对有限,GAN则可以拓展数据生成的能力,可是其训练稳定性欠佳。

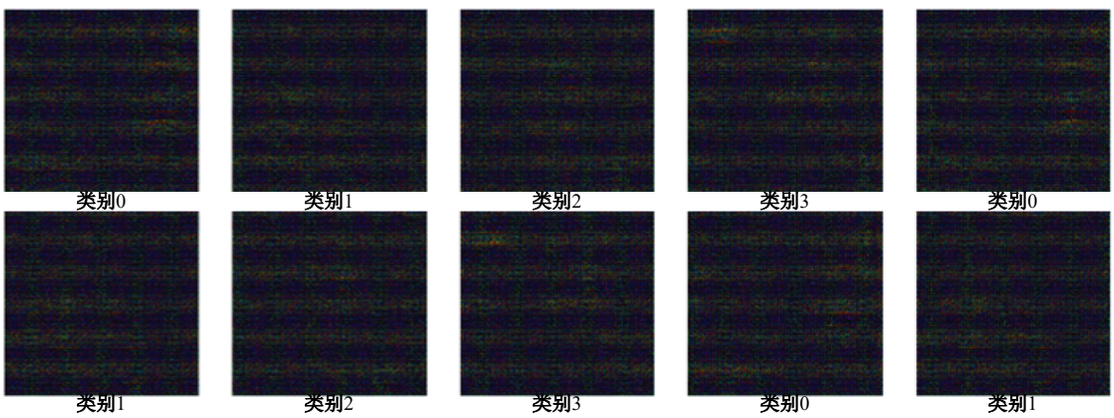



**图 11　GAN 生成样本**

**Fig. 11　GAN generated samples**

## 3　当前深度学习面临的问题

虽然深度学习于机械故障诊断领域的理论研究收获了丰富成果,各种各样的创新算法在实验室环境里呈现出引人注目的性能表现,然而当这些技术实际被部署到复杂且多变的工业现场时,就会发觉存在诸多问题,这些问题不光是源于算法自身的技术局限性,更在深层次上体现了数据驱动方法与工业系统工程之间的复杂关系。工业环境的严苛状况、设备运行的持续状态、安全要求的严格程度以及维护决策的复杂情形,共同构成了深度学习技术在故障诊断领域应用时不得不面对的系统性挑战。核心问题如表3所示。

**表 3　核心问题**

**Table 3　Core issues**

| 类别 | 具体表现 | 影响 |
|---|---|---|
| 数据稀缺与不平衡 | 故障样本极少,不同故障发生率差异较大 | 模型对罕见故障的识别能力差,整体性能下降 |
| 模型可解释性较差 | 深度学习作为“黑箱”,决策依据不明确 | 在关键场景难以被信任与采纳 |
| 实时性与边缘部署困难 | 模型计算复杂,难以满足毫秒级别的响应 | 可能导致故障的漏报、误报 |
| 跨域泛化能力不足 | 在训练分布外的工况下的性能有所下降 | 在新设备、新环境中需要对模型进行重新训练 |
| 多模态数据融合复杂 | 异构传感器数据在时序、语义上存在一定差异 | 信息未能做到有效互补,系统的鲁棒性较低 |

### 3.1　数据稀缺与不平衡

工业环境里故障数据获取面临的困境成为深度学习应用的首要妨碍,此问题的复杂程度远远超过一般机器学习场景,在理想状况下,深度学习模型开展训练需要大量高质量且带有标注的数据,可是工业现场的实际情形却与理想状况截然不同,关键设备一般在严格监控下运行,一旦出现异常迹象就会马上进行维护干预,这致使严重故障的实际场景数据非常稀少。更麻烦的是,不同故障类型的发生频率有很大差别,某些罕见故障在整个设备生命周期中或许只出现几次,致使正常状态样本与特定故障样本的比例可能达到数百比一的极端情形,这种极端的数据不平衡影响模型的整体准确率,还会让模型对罕见故障的识别能力严重欠缺,而这正是工业诊断中最需关注的关键所在[27,28]。

### 3.2　模型可解释性挑战

深度学习模型缺乏可解释性已成为限制其在工业现场部署的关键妨碍,在安全关键场景中这一问题的严峻性非常凸显,当诊断系统给出故障预警或维护建议时,现场工程师所需的并非仅仅是一个简单的分类结果,更在于理解模型做出该决策的深层缘由、置信水平以及相关的推理流程,在涉及高价值设备或连续生产流程的情形下,基于“黑箱”模型的诊断结果很难直接被用作停机检修或重大维修决策的依据。

当前可解释性研究方法大致分成3类:事后解释方法、内在可解释模型以及归因分析技术。

事后解释方法,如局部可解释模型不可知解释(Local interpretable model-agnostic explanations,LIME)和沙普利加法解释(SHapley additive exPlanations,SHAP)依靠局部近似来阐释单个预测的决策依据,能提供特征关键性的量化指标,不过其解释的稳定性和一致性仍存在疑问[65]。内在可解释模型借助注意力机制、稀疏编码等技术在设计阶段就引入可解释性元素(比如视觉注意),以此来展示模型关注的特征区域,为理解模型行为提供了直观线索,然而注意力机制所展示的关注区域本质上仍是统计相关性的反映,而不是真正的因果解释,这难以直接揭示故障背后的物理机理,并且不同初始化或训练条件下注意力分布的稳定性仍然存在疑虑[50]。归因分析技术依靠计算输入特征对输出结果的贡献度来揭示因果关系,然而在处理复杂非线性系统时面临计算复杂度和准确性两方面的挑战,且提供的解释往往停留在统计相关性层面,难以建立与物理故障机制的直接联系,在工业现场的实际应用中仍存在一定的局限性[52,53]。

### 3.3 实时性与边缘部署问题

工业故障诊断对于实时性有着近乎苛刻的要求,这种特性与深度学习模型的计算复杂性形成了强烈反差,成为技术落地的重大妨碍,在旋转机械监测等典型场景里,早期故障的特征大多时候瞬间消失,诊断系统要在毫秒级别的时间内完成信号采集、特征提取以及状态识别,任何十分突出的延迟都有可能致使故障漏报或者误报。这种实时性要求体现在推理速度方面,还覆盖模型更新频率、系统响应时间以及决策执行延迟等多个维度,共同构建起复杂的时序约束体系。

当前的研究在效率优化上取得了一定进展,模型压缩技术借助剪枝、量化以及知识蒸馏等方法大幅降低参数量和计算量,轻量级网络设计从架构角度优化计算效率,硬件加速技术依靠专用芯片和并行计算提高处理速度。这些技术的综合运用让深度学习模型在特定场景中已能接近实时性要求,比如图像类故障诊断可达到每秒数十帧的处理速度,振动信号分析可以在百毫秒内完成状态识别,然而这种效率提升往往随着性能损耗,压缩后的模型在复杂故障识别、噪声鲁棒性以及泛化能力方面出现较大衰减,在多故障并发、早期微弱故障检测等有挑战性的场景中表现欠佳[66]。

### 3.4 跨域泛化能力问题

工业设备的运行工况一直处于动态变化的状态,其本身有的时变性致使在特定条件下训练的诊断模型在新环境中遭遇诸多问题,而该问题的实质是这些模型在独立同分布假设下进行训练,却要应用于分布外的环境中。

目前应对域适应挑战的技术体系主要有3个流派:基于差异度量的方法借助最小化源域与目标域之间的分布差异来学习域不变特征,如最大均值差异、相关对齐等度量函数被广泛用于特征空间的对齐优化[29];基于对抗学习的方法依靠引入域判别器促使特征提取器学习域不变表示,这种生成式对抗的理念在多个领域呈现出较强潜力[27];基于重构的方法利用共享编码-私有编码的架构分离域不变特征与域特定特征,保证学习到的核心特征有跨域稳定性[31]。这些方法在实验室环境中表现良好,在一些基准测试里甚至能达到近乎完美的域适应效果,然而当面对真实工业场景中的复杂分布变化时,模型性能可能会出现明显下滑。

深入分析泛化困境的成因,可以识别出多个层次的技术挑战。在数据层面,不同工况下的信号特征存在系统性差异,包括幅值缩放、频率偏移、相位变化等线性变换,以及信噪比变化、调制特征改变等非线性变换。在特征层面,域不变特征与任务相关特征深度纠缠,简单的特征对齐可能导致有用信息的损失。在算法层面,现有方法对超参数极为敏感,在不同迁移任务中需要复杂的重新调优,严重影响了其实用性。在理论层面,域适应的泛化误差边界和性能极限尚不明确,算法设计缺乏坚实的理论指导[25,33,34]。

### 3.5 多模态数据融合问题

现代工业装备配备的多样化传感器构成了一个丰富的信息生态系统,振动、声学、温度、压力、电流、油液等多源数据共同描绘着设备的健康状态。这种多模态特性既带来了信息冗余的便利,也引入了数据融合的挑战:如何从异构信息源中提取一致性知识,如何平衡不同模态的贡献权重,如何实现跨模态的协同推理,有无此类问题直接决定了该诊断系统是否准确可靠。多模态融合的深层价值在于其能够突破单一传感器的感知局限,通过信息互补和交叉验证构建更加全面、鲁棒的设备状态认知体系。

然而,理想中的多模态融合系统与现实可行状况之间依旧存在着差距,不同传感器的采样频率有所不同,这致使时序对齐面临险阻,比如振动信号可能按照千赫兹的频率进行采样,而温度信号却仅以赫兹频率更新,这样的多速率特性给实时融合造成了严峻挑战,模态之间的质量差异同样不能被忽视,某些传感器在特定工况下可能会出现失效或者性能降低的情况,融合系统要拥有故障容错以及能力退化感知机制。异质数据处于不同的数学空间,振动信号呈现为时频分布,温度信号是标量序列,电流信号含有谐波成分,这种语义鸿沟让直接比较与融合变得极为险阻,而且多传感器系统的成本、复杂度以及维护要求呈现出指数级增长态势,对其在工业现场的规模化部署形成了制约[12,35,37]。

## 4　未来展望

深度学习通过其强大的端到端特征学习能力,成功将工业故障诊断从依赖手工特征工程的范式转变为数据驱动的智能新范式。从卷积神经网络对空间模式的感知,到循环神经网络对时序动态的捕捉,再到注意力机制与Transformer对关键信息的聚焦,模型的训练始终围绕着更精准的识别故障特征这一核心目标。但将这些高性能模型部署于情况多变的工业现场时,便会发现其在数据稀缺下的泛化能力、复杂噪声中的鲁棒性及作为“黑箱”的决策可信度等方面依然面临着严峻的挑战。因此,未来的深度学习将不能再是孤立的数据分析工具,而应是能够将感知、认知、决策与行动深度融合的自主智能体。展望未来,故障诊断领域将呈现以下几个重要发展趋势(见表4)。这些趋势不仅将解决当前面临的挑战,还将推动诊断系统向更高层次的智能化演进。

### 4.1　AutoML驱动的自适应诊断系统

当前深度学习模型的设计严重依赖专家经验,从网络架构选择到超参数调优都需要大量人工干预。一综述系统总结了自动化机器学习的最新进展,指出AutoML通过NAS、超参数优化(Hyperparameter optimization,HPO)等技术,能为特定故障诊断任务自动设计最优模型架构[67]。未来,基于AutoML的自适应诊断系统将具备以下能力:①根据设备类型和工况特征自动生成定制化诊断模型;②在运行过程中持续优化模型参数以适应设备老化;③实现跨设备、跨工况的零样本或少样本快速适配。这种自动化能力将大幅降低深度学习在工业现场的应用门槛,使非专家也能构建高性能诊断系统[68-70]。

**表4 未来发展趋势**

**Table 4　Future trends**

| 发展趋势 | 核心技术 | 可解决的问题 |
|---|---|---|
| AutoML驱动的自适应诊断 | NAS、超参数优化、动态资源分配 | 跨域泛化能力不足 |
| 多模态混合深度学习 | 跨模态注意力、知识迁移、多尺度融合 | 多模态数据融合复杂 |
| 物理信息融合的因果诊断 | 物理约束网络、因果推理、数字孪生 | 模型可解释性较差 |
| 边缘-云协同的分布式智能诊断 | 轻量化模型、联邦学习、5G/IoT | 实时性与边缘部署困难 |
| 生成式人工智能驱动的数据增强与故障预测 | 扩散模型、条件GAN、时序生成模型 | 数据稀缺与不平衡 |

### 4.2　多模态混合深度学习的深度融合

现代工业装备所配备的各类多样化传感器会产出振动、声学、温度、压力、电流等多种不同类型的异构数据,文献[71]对多模态混合深度学习的最新发展情况做了较为详细的分析,说明借助有效的跨模态特征融合方式,可突破单一传感器在感知方面存在的局限。未来的多模态诊断系统会达成以下几点:①基于注意力机制来进行动态特征权重分配,依照工况的变化对各模态的贡献度做自适应调整;②实现跨模态知识迁移,借助丰富模态的知识来辅助稀缺模态的学习;③进行多尺度时空特征融合,同步捕捉故障的局部细节以及全局趋势。这种深度融合可提升诊断的精度,而且依靠信息互补可提高系统在传感器出现失效或者性能退化状况时的鲁棒性。

### 4.3　物理信息融合的因果诊断模型

当下数据驱动的诊断模型欠缺对物理故障机制的透彻理解,致使可解释性欠佳,泛化能力也受限,未来的研究方向会着重于把物理先验知识深度融入深度学习框架:①开展基于物理约束的神经网络设计,以此保证模型输出契合物理规律;②将因果推理与深度学习相结合,从相关性分析转变为因果性诊断;③实现数字孪生与深度学习的协同,于虚拟空间里模拟故障演化进程。一项在森林应用中的研究显示[72],融合领域知识的深度学习模型在复杂环境下的表现要比纯数据驱动方法出色,在故障诊断领域,这种物理信息融合可破解“黑箱”难题,让诊断决策精准又可靠。

### 4.4 边缘-云协同的分布式智能诊断架构

工业现场对于实时性有着极为严苛的要求,然而深度学习模型却存在计算复杂性的问题,这两者之间形成了矛盾,未来的解决办法会基于边缘-云协同架构,具体如下:①在边缘端部署轻量化模型,以此实现毫秒级的快速检测;②在云端开展复杂分析以及模型的持续优化;③借助联邦学习来支持隐私保护型的协同训练[73,74]。一项针对离心泵故障诊断的研究显示[75],经过精心设计的轻量化模型,在维持高精度状况下,可大幅度降低计算复杂度,把5G和物联网技术相结合,这种分布式架构可达成广域设备群的协同监测以及智能运维。

### 4.5 生成式人工智能驱动的数据增强与故障预测

数据稀缺以及不平衡属于工业故障诊断的关键瓶颈所在,生成式人工智能技术针对这一问题给出了新的思路:①借助扩散模型生成高质量的故障样本,模拟从正常状态至故障状态的连续演化进程;②运用条件生成对抗网络进行少样本学习,在极端不平衡的状况下维持诊断性能[76,77];③利用时序生成模型开展故障预测,提前对潜在风险发出预警[78]。这些技术可扩充训练数据,而且还可模拟罕见故障场景,为应急演练以及预防性维护给予支持。

上述技术的发展也呈现出深度融合的态势,未来智能诊断系统可构建于这样的技术生态之上:先是凭借AutoML为特定设备类型迅速设计基础模型架构;接着,利用自监督学习从海量无标签数据里预训练通用特征提取器;再联合多模态融合技术整合异构传感器信息,借助物理信息融合保证模型契合故障机理;最后,在边缘-云协同架构里达成实时监测和持续优化。这种技术融合把诊断精度提升到新的高度,还让诊断系统拥有了自我演进的能力,可依据设备运行数据的积累持续优化自身性能,以适应设备老化以及工况变化带来的挑战。

深度学习在故障诊断领域的发展正从“感知智能”迈向“认知智能”,未来的诊断系统不再只是简单的模式识别工具,而是成为有物理理解、因果推理、自主学习以及协同决策能力的智能体。这种演进可解决当前工业现场面临的实际难题,还会为构建自主智能运维体系提供有力的技术支撑,随着计算能力的不断提升、算法理论的持续突破以及工业数据的日益增多,深度学习有望在故障诊断领域产生更为深远的影响,推动工业智能化步入新的发展阶段。

## 5 结束语

本文系统梳理了深度学习在工业故障诊断领域的研究进展,搭建起了一个囊括方法体系、关键技术以及应用场景的完整分析架构,分析结果显示,深度学习的技术突破大多时候源自不同范式的交叉融合,如从监督学习朝着无监督/半监督学习的拓展,从单一模型向着混合架构的演进,以及从数据驱动到物理信息融合的探索等情况。类似这样的融合创新提升了诊断性能,而且提高了系统在复杂工业环境下的适应性和鲁棒性,该领域的发展正呈现出从追求单一模型性能,转向注重技术可解释性、跨场景泛化能力以及工程落地可行性的深刻变化,虽然深度学习在故障诊断领域已经取得了一定进展,然而其在工业现场的实际部署依旧面临数据稀缺、模型可解释性、实时性要求等诸多挑战。未来的发展道路将在很大程度上取决于能否把数据驱动的方法与物理机理、领域知识进行更深层次地融合,在不确定的工业环境中达成稳定、可信的自主智能。

**参考文献:**

[1] Minh N Q, Khiem N T, Giang V H. Fault classification and localization in power transmission line based on machine learning and combined CNN-LSTM models[J]. Energy Reports, 2024, 12: 5610-5622.

[2] Wang Z, Liu Q, Chen H, et al. A deformable CNN-DLSTM based transfer learning method for fault diagnosis of rolling bearing under multiple working conditions[J]. International Journal of Production Research, 2021, 59(16): 4811-4825.

[3] Ye L, Ma X, Wen C. Rotating machinery fault diagnosis method by combining time-frequency domain features and CNN knowledge transfer[J]. Sensors, 2021, 21(24): No. 8168.

[4] Wang H, Xu J, Yan R, et al. A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(5): 2377-2389.

[5] Dibaj A, Ettefagh M M, Hassannejad R, et al. A hybrid fine-tuned VMD and CNN scheme for untrained compound fault diagnosis of rotating machinery with unequal-severity faults[J]. Expert Systems with Applications, 2021, 167: No. 114094.

[6] Bu C, Liu T, Wang T, et al. A CNN-architecture-based photovoltaic cell fault classification method using thermographic images[J]. Energies, 2023, 16(9): No. 3749.

[7] Pham M T, Kim J M, Kim C H. 2D CNN-based multi-output diagnosis for compound bearing faults under variable rotational speeds[J]. Machines, 2021, 9(9): No. 199.

[8] Hu P, Zhao C, Huang J, et al. Intelligent and small samples gear fault detection based on wavelet analysis and improved CNN[J]. Processes, 2023, 11(10): No. 2969.

[9] Song X, Cong Y, Song Y, et al. A bearing fault diagnosis model based on CNN with wide convolution kernels[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(8): 4041-4056.

[10] Zhao K, Xiao J, Li C, et al. Fault diagnosis of rolling bearing using CNN and PCA fractal based feature extraction[J]. Measurement, 2023, 223: 113754.

[11] Wang X, Zhou P, Peng X, et al. Fault location of transmission line based on CNN-LSTM double-ended combined model[J]. Energy Reports, 2022, 8: 781-791.

[12] Zhang Y, He L, Cheng G. MLPC-CNN: a multi-sensor vibration signal fault diagnosis method under less computing resources[J]. Measurement, 2022, 188: No. 110407.

[13] Cheng Y, Wang D, Zhou P, et al. A survey of model compression and acceleration for deep neural networks[EB/OL]. [2025-12-19]. https: //doi. org/10. 48550/arXiv. 1710. 09282.

[14] Schmidhuber J. Deep learning in neural networks: an overview[J]. Neural Networks, 2015, 61: 85-117.

[15] Hinton G E, Osindero S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[16] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[17] Zachariades C, Xavier V. A review of artificial intelligence techniques in fault diagnosis of electric machines [J]. Sensors, 2025, 25: No. 5128.

[18] Choi D J, Han J H, Park S U, et al. Comparative study of CNN and RNN for motor fault diagnosis using deep learning[C]//2020 IEEE 7th International Conference on Industrial Engineering and Applications, Bangkok, 2020: 693-696.

[19] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.

[20] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323: 533-536.

[21] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[22] Wang J, Joshi G. Cooperative SGD: a unified framework for the design and analysis of local-update SGD algorithms[J]. JMLR, 2021, 22(213): 1-50.

[23] Zhu Y, Li G, Tang S, et al. Acoustic signal-based fault detection of hydraulic piston pump using a particle swarm optimization enhancement CNN[J]. Applied Acoustics, 2022, 192: No. 108718.

[24] Tang S, Zhu Y, Yuan S. Intelligent fault identification of hydraulic pump using deep adaptive normalized CNN and synchrosqueezed wavelet transform[J]. Reliability Engineering & System Safety, 2022, 224: No. 108560.

[25] Zhang C, Chen P, Jiang F, et al. Fault diagnosis of nuclear power plant based on sparrow search algorithm optimized CNN-LSTM neural network[J]. Energies, 2023, 16(6): No. 2934.

[26] Zhiyi H, Haidong S, Xiang Z, et al. An intelligent fault diagnosis method for rotor-bearing system using small labeled infrared thermal images and enhanced CNN transferred from CAE[J]. Advanced Engineering Informatics, 2020, 46: No. 101150.

[27] Chen H, Wei J, Huang H, et al. Novel imbalanced fault diagnosis method based on generative adversarial networks with balancing serial CNN and Transformer (BCTGAN) [J]. Expert Systems with Applications, 2024, 258: No. 125171.

[28] Wang Y, Li D, Li L, et al. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model[J]. Heliyon, 2024, 10(15): No. e35407.

[29] Cao H, Shao H, Zhong X, et al. Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds[J]. Journal of Manufacturing Systems, 2022, 62: 186-198.

[30] Yu G, Wu P, Lv Z, et al. Few-shot fault diagnosis method of rotating machinery using novel MCGM based CNN[J]. IEEE Transactions on Industrial Informatics, 2023, 19(11): 10944-10955.

[31] Jia W, Dong Z, Shi H, et al. A zero-shot bearing fault diagnosis framework utilizing spatial relation-

ships among primary features under variable working conditions[J]. Mechanical Systems and Signal Processing, 2025, 238: No. 113103.

[32] Lv Y, Qi X, Zheng X, et al. Unsupervised quantitative judgment of furnace combustion state with CBAM-SCAE-based flame feature extraction[J]. Journal of the Energy Institute, 2024, 116: No. 101733.

[33] Meng Z, Cao W, Sun D, et al. Research on fault diagnosis method of MS-CNN rolling bearing based on local central moment discrepancy[J]. Advanced Engineering Informatics, 2022, 54: No. 101797.

[34] Wang C, Li H, Zhang K, et al. Intelligent fault diagnosis of planetary gearbox based on adaptive normalized CNN under complex variable working conditions and data imbalance[J]. Measurement, 2021, 180: No. 109565.

[35] Gu Y, Zhang Y, Yang M, et al. Motor on-line fault diagnosis method research based on 1D-CNN and multi-sensor information[J]. Applied Sciences, 2023, 13(7): No. 4192.

[36] Peng D, Liu Z, Wang H, et al. A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains[J]. IEEE Access, 2019, 7: 10278-10293.

[37] Li X, Li P, Zhang Z, et al. CNN-LSTM-based fault diagnosis and adaptive multichannel fusion calibration of filament current sensor for mass spectrometer[J]. IEEE Sensors Journal, 2024, 24(2): 2255-2269.

[38] Chen J, Jiang J, Guo X, et al. An efficient CNN with tunable input-size for bearing fault diagnosis[J]. International Journal of Computational Intelligence Systems, 2021, 14(1): No. 625.

[39] Yuan Z, Zhang L, Duan L, et al. Intelligent fault diagnosis of rolling element bearings based on HHT and CNN[C]//IEEE 2018 Prognostics and System Health Management Conference, Chongqing, China, 2018: 292-296.

[40] Yu Z, Zhang C, Liu J, et al. SKND-TSACNN: a novel time-scale adaptive CNN framework for fault diagnosis of rotating machinery[J]. Knowledge-Based Systems, 2023, 275: No. 110682.

[41] Chen X, Zhang B, Gao D. Bearing fault diagnosis base on multi-scale CNN and LSTM model[J]. Journal of Intelligent Manufacturing, 2021, 32(4): 971-987.

[42] Luo R, Li Y, Guo H, et al. Cross-operating-condition fault diagnosis of a small module reactor based on CNN-LSTM transfer learning with limited data[J]. Energy, 2024, 313: No. 133901.

[43] Kang J, Zhu X, Shen L, et al. Fault diagnosis of a wave energy converter gearbox based on an Adam optimized CNN-LSTM algorithm[J]. Renewable Energy, 2024, 231: No. 121022.

[44] Snyder Q, Jiang Q, TRIPP E. Integrating self-attention mechanisms in deep learning: a novel dual-head ensemble transformer with its application to bearing fault diagnosis[J]. Signal Processing, 2025, 227: No. 109683.

[45] Zhu Q X, Qian Y S, Zhang N, et al. Multi-scale Transformer-CNN domain adaptation network for complex processes fault diagnosis[J]. Journal of Process Control, 2023, 130: No. 103069.

[46] Cheng D, Liu L, Yu Z. CNN-based intelligent fault-tolerant control design for turbofan engines with actuator faults[J]. IEEE Access, 2021, 9: 28122-28139.

[47] Jiang T, Gradus J L, Rosellini A J. Supervised machine learning: a brief primer[J]. Behavior Therapy, 2020, 51(5): 675-687.

[48] Li Y F, Liang D M. Safe semi-supervised learning: a brief introduction[J]. Frontiers of Computer Science, 2019, 13(4): 669-676.

[49] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[50] 冯志刚,王颖,王宇. 基于多通道融合的齿轮箱故障诊断方法[J/OL]. [2025-12-19]. https://doi.org/10.13229/j.cnki.jdxbgxb.20241377.

[51] Wang K, Gou C, Duan Y, et al. Generative adversarial networks: introduction and outlook[J]. IEEE/CAA Journal of Automatica Sinica, 2017, 4(4): 588-598.

[52] Peng D, Wang H, Desmet W, et al. RMA-CNN: a residual mixed-domain attention CNN for bearings fault diagnosis and its time-frequency domain interpretability[J]. Journal of Dynamics, Monitoring and Diagnostics, 2023, 2: 115-132.

[53] Huang Y, Zhang J, Liu R, et al. Improving accuracy and interpretability of CNN-based fault diagnosis through an attention mechanism[J]. Processes, 2023, 11(11): No. 3233.

[54] Zhang J, Liang J, Liu J. Fault diagnosis in reactor coolant pump with an automatic CNN-based mixed model[J]. Progress in Nuclear Energy, 2024, 175: No. 105294.

[55] He C, Ge D, Yang M, et al. A data-driven adaptive fault diagnosis methodology for nuclear power systems based on NSGAII-CNN[J]. Annals of Nuclear

Energy, 2021, 159: No. 108326.

[56] Dembczyński K, Waegeman W, Cheng W, et al. On label dependence and loss minimization in multi-label classification[J]. Machine Learning, 2012, 88(1/2): 5-45.

[57] Nam J, Kim J, Mencía E L, et al. Large-scale multi-label text classification-revisiting neural networks[J]. Lecture Notes in Computer Science, 2014, 8725: 437-452.

[58] Zhang Z. Graph neural network-based bearing fault diagnosis using Granger causality test[J]. Expert Systems with Applications, 2024, 242: No. 122827.

[59] Weiss K, Khoshgoftaar T M, Wang D. A survey of transfer learning[J]. Journal of Big Data, 2016, 3(1): No. 9.

[60] CHEN Q, ZHANG F, WANG Y, et al. Bearing fault diagnosis based on efficient cross space multi-scale CNN transformer parallelism[J]. Scientific Reports, 2025, 15(1): No. 12344.

[61] Zhao X, Yin F, Durrett G. Understanding synthetic context extension via retrieval heads[EB/OL]. [2025-12-19]. https://doi.org/10.48550/arXiv.2410.22316.

[62] Liao Y L, Smidt T. EQUIFORMER: Equivariant graph attention transformer for 3D atomistic graphs [EB/OL]. [2025-12-19]. https://doi.org/10.48550/arXiv.2206.11990.

[63] Ali M, Ali M, Hussain M, et al. Generative adversarial networks (GANs) for medical image processing: recent advancements[J]. Archives of Computational Methods in Engineering, 2025, 32(2): 1185-1198.

[64] Baratchi M, Wang C, Limmer S, et al. Automated machine learning: past, present and future[J]. Artificial Intelligence Review, 2024, 57(5): No. 122.

[65] Alvarez-melis D, Jaakkola T S. On the robustness of interpretability methods[EB/OL]. [2025-12-19]. https://doi.org/10.48550/arXiv.1806.08049.

[66] Siddique N I. Fault classification and location of a PMU-equipped active distribution network using deep convolution neural network (CNN)[J]. Electric Power Systems Research, 2024, 229: No. 110178.

[67] He X, Zhao K, Chu X. AutoML: a survey of the state-of-the-art[J]. Knowledge-Based Systems, 2021, 212: No. 106622.

[68] Zhao X, Zhu X, Zhang R, et al. Imbalanced fault diagnosis of a conditional variational auto-encoder with transfer and adversarial structures[J]. Advanced Engineering Informatics, 2025, 64: No. 103032.

[69] Shang Z, Wang Z, Pan C, et al. Physics-informed auto-encoder based on digital twin for rolling bearing fault diagnosis under imbalanced sample conditions [J]. Engineering Applications of Artificial Intelligence, 2026, 163: No. 113017.

[70] Lu Z, Chu Z, Zhu M, et al. Unsupervised feature learning using locality-preserved auto-encoder with complexity-invariant distance for intelligent fault diagnosis of machinery[J]. Applied Intelligence, 2025, 55(6): 379.

[71] Bayoudh K. A survey of multimodal hybrid deep learning for computer vision: architectures, applications, trends, and challenges[J]. Information Fusion, 2024, 105: No. 102217.

[72] Yun T, Li J, Ma L, et al. Status, advancements and prospects of deep learning methods applied in forest studies[J]. International Journal of Applied Earth Observation and Geoinformation, 2024, 131: 103938.

[73] Chen C, Wang T, Mao D, et al. A multi-scale graph pyramid attention network with knowledge distillation towards edge computing robotic fault diagnosis[J]. Expert Systems with Applications, 2025, 260: No. 125469.

[74] Yu Z, Wang Y, Zong X, et al. GRUDMU-DSCNN: an edge computing method for fault diagnosis with missing data[J]. Applied Intelligence, 2025, 55(2): No. 140.

[75] Zaman W, Siddique M F, Ullah S, et al. Hybrid deep learning model for fault diagnosis in centrifugal pumps: a comparative study of VGG16, ResNet50, and wavelet coherence analysis[J]. Machines, 2024, 12(12): No. 905.

[76] Wang X, Jiang H, Mu M, et al. A trackable multi-domain collaborative generative adversarial network for rotating machinery fault diagnosis[J]. Mechanical Systems and Signal Processing, 2025, 224: No. 111950.

[77] Wang X, Jiang H, Zeng T, et al. An adaptive fused domain-cycling variational generative adversarial network for machine fault diagnosis under data scarcity [J]. Information Fusion, 2026, 126: No. 103616.

[78] Fan C, Zhang Y, Ma H, et al. A novel lightweight DDPM-based data augmentation method for rotating machinery fault diagnosis with small sample[J]. Mechanical Systems and Signal Processing, 2025, 232: No. 112741.