



REVIEW

滚动轴承故障诊断研究综述

金学波^{1,*}, 王继阳¹¹北京工商大学, 计算机与人工智能学院, 北京 100048

学术编辑: Shenglun Yi; 收稿日期: 2024-01-03; 录用日期: 2024-03-01; 发布日期: 2024-04-12

*通讯作者: 金学波, jinxuebo@btbu.edu.cn

文章引用

金学波, 王继阳. 滚动轴承故障诊断研究综述. 人工智能前沿与应用, 2024, 1(1): 1-15.

Citation

Jin, X., & Wang, J. (2024). Research Review of Rolling Bearing Fault Diagnosis. *Frontiers and Applications of Artificial Intelligence*, 1(1), 1-15.

© 2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 License.

摘要

滚动轴承作为旋转机械的核心部件, 保持对轴承健康状态的监测能保证整个机械设备的正常运转。对于轴承故障检测方法的研究已经有了长久的发展, 本文从传统故障检测方法到引入深度学习算法进行故障检测最后提出迁移学习在轴承故障诊断领域的应用进行了综述。传统的故障检测方法可以根据诊断步骤分为特征提取、故障识别两类, 阐述了不同方法的适用条件以及应用缺陷。基于深度学习的轴承故障检测, 尽管有了一定的发展, 但是因其在数据和标签上的局限性, 该类型的模型仍然有很大的发展空间和研究潜力。随着迁移学习方法的引入解决了在数据和标签上的限制, 为轴承故障检测提供了新的思路和方向, 并阐述了迁移学习方法目前面临的困境。

关键词: 滚动轴承, 故障诊断, 深度学习, 迁移学习

Research Review of Rolling Bearing Fault Diagnosis

Xuebo Jin^{1,*} and Jiyang Wang¹¹School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China

Academic Editor: Shenglun Yi; Submitted: 2024-01-03; Accepted: 2024-03-01; Published: 2024-04-12

*Correspondence Author: Xuebo Jin, jinxuebo@btbu.edu.cn

Abstract

Rolling bearings as the core components of rotating machinery, keep monitoring the health status of bearings can ensure the normal operation of the whole mechanical equipment. The research on bearing fault detection methods has been developed for a long time, and this paper reviews the traditional fault detection methods to the introduction of deep learning algorithms for fault detection and finally proposes the application of migration learning in the field of bearing fault diagnosis. Traditional fault detection methods can be divided into two categories according to the diagnostic steps: feature extraction and fault identification, and the applicable conditions and application defects of different methods are described. Bearing fault detection based on deep learning, in spite of some development, still has a lot of room for development and research potential for this type of model due to its limitations in data and labeling. With the introduction of the migration learning method to solve the limitations in data and labels, it provides new ideas and directions for bearing fault detection, and describes the dilemmas currently faced by the migration learning method.

Keywords: Rolling bearings, fault diagnosis, deep learning, transfer learning

1 引言

作为旋转机械的核心部件，滚动轴承在车辆、航空航天、数控机床、机电设备、精密仪器等机械设备中有广泛应用，其具有复杂结构、运行速度快、负载大的特点，被称为“工业的关节”[1]。滚动轴承其在旋转机械中的关键地位，直接决定着机械设备的健康状态、工作性能以及设备的安全性。当轴承出现故障时，往往会造成异常振动，对机械结构产生剧烈冲击，进而影响机械系统的其他部件，对机械设备的安全性和生产效率造成巨大威胁。

因滚动轴承对制造、安装和环境要求很高，同时常在高载荷的工况下运作，所以滚动轴承相对于旋转机械中的其他零件有着更高的故障率。根据文献指出[2]，滚动轴承故障率占风力涡轮机械设备故障的30%。此外，训练轴承的年度检查发现三分之一的轴承需要更换。2009年8月17日俄罗斯萨扬水电站机组水轮机轴承振动幅值超标而未按规定“卸荷并停机”，从而导致顶盖固定螺栓发生疲劳断裂，造成75人死亡，并造成巨大经济损失的重大事故。因此了解并监测轴承的运作状态，尽早发现轴承故障，实现精准诊断轴承故障对机械设备的安全性和可靠性十分重要[43-49]。

本文以滚动轴承故障诊断的三个阶段为主线分别展开进行论述，在第二部分中第一节内容为典型轴承故障诊断方法，第二节为基于深度学习的轴承故障诊断方法，基于迁移学习的轴承故障诊断方法为第三节，最后第三部分对整体内容进行总结，并提出了对于未来发展方向的论述。本文对每个阶段中被广泛使用的诊断方法分别进行了介绍和分析，并在每部分的末尾进行了对比和总结以供学者进行比对和参考。

2 滚动轴承故障诊断

滚动轴承作为精密机械零件的一种，其主要作用为将轴承与轴之间的滑动变为滚动，以减少轴与轴承座之间的摩擦力。滚动轴承的自身结构仅由外圈、内圈、滚子、保持架四部分组成，其结构如图1所示[3]。其中，外圈固定在轴承座上，内圈在轴上固定，滚子作为轴承的主要部件位于外圈和内圈中间，决定着轴承的承载能力，保持架用于保持滚子之间的距离，防止滚动体出现碰撞摩擦损伤[4]。

现阶段，在故障信号采集方面能够体现轴承状态的信号主要包括振动、声发射、温度、电流、油液成分等。其中振动信号包含十分丰富的轴承故障信息，以振动信号数据为基础的轴承故障诊断研究已经较为长久的发展，基

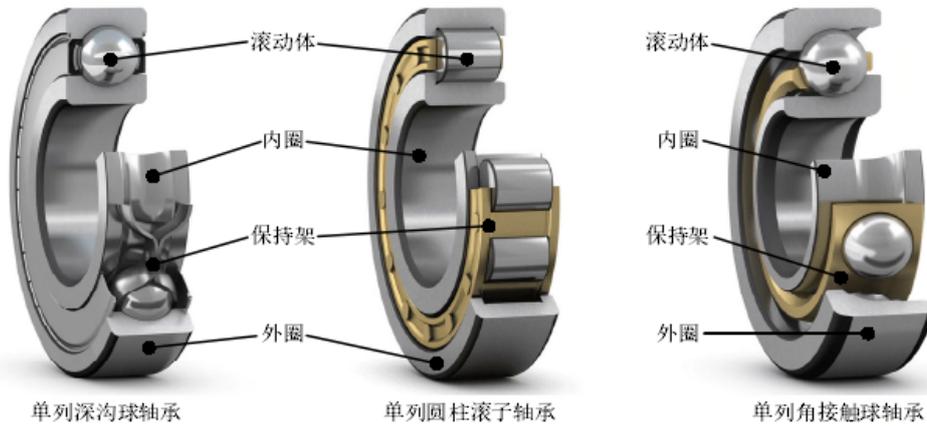


图 1. 轴承结构图

于振动信号分析的轴承故障诊断目前是故障诊断领域普遍应用且十分高效的方法，因此本文限定于基于振动信号的故障诊断研究。

在滚动轴承工作过程中，滚子与内圈或外圈碰撞产生振动，这种由轴承各部件之间碰撞所诱发的振动属于强迫振动，由强迫振动所产生的频率称为固有频率，这是由轴承结构本身所决定的。当滚动轴承周期性转动时，形成的故障冲击也呈周期性分布，容易引起在内外滚道或滚动体上产生磨损、断裂等情况，破坏轴承的内部运动平衡。因此当轴承旋转工作时，内部的振动十分复杂，其中不仅包括这自身的振动，也包括机械设备剩余零部件的振动。按照故障发生的位置可以将故障类型分为内圈故障、外圈故障和滚子故障 [5]。

通过轴承的结构参数、转速和故障类型等因素，能够确定出轴承故障特征的频率。假定轴承外圈固定，内圈同转轴做纯滚动，同时滚动体和滚道的质量均匀分布。在这样的情况下，轴承故障特征频率可由以下公式计算得出 [6]。

轴承的外圈故障特征频率：

$$f_o = \frac{Z}{2} \left(1 - \frac{d}{D} \cos \alpha \right) f_r \quad (1)$$

轴承的保持架故障特征频率：

$$f_e = \frac{1}{2} \left[1 - \frac{d}{D} \cos \alpha \right] f_r \quad (2)$$

轴承的滚动体故障特征频率：

$$f_b = \frac{D}{2d} \left[1 - \left(\frac{d}{D} \right)^2 \cos^2 \alpha \right] f_r \quad (3)$$

轴承的内圈故障特征频率：

$$f_b = \frac{Z}{2} \left[1 - \frac{d}{D} \cos \alpha \right] f_r \quad (4)$$

为了能更好的实现故障特征的提取，学者们通过联系轴承故障振动机理围绕该问题展开了大量的研究工作，提出了许多基于振动信号的故障检测方法，以实现更高效的轴承故障诊断 [7]。本节以轴承故障检测领域技术发展线为顺序对轴承故障检测方法进行概述，其中包含典型轴承故障诊断方法、基于深度学习的轴承故障诊断方法、基于迁移学习的轴承故障诊断方法三部分。

2.1 典型的滚动轴承诊断方法

当前对于基于传统轴承故障检测方法的研究主要分为故障信号采集、故障特征提取、故障类型识别三个步骤。典型的振动信号特征提取方法的发展同数字信号处理技术紧密联系，常见的信号处理方法主要分为时域分析、频域分析和时频分析方法。基于物联网的发展，我们能够轻松获取比以往都要丰富的数据量，在这样的背景下，传统机器学习方法有了丰富的发展，产生了多种不同类型的算法模型，在轴承故障诊断领域常用的方法有支持向量机、人工神经网络、k 近邻算法等。

2.1.1 振动信号特征提取方法

时域分析方法在轴承故障诊断领域应用较早，该方法是根据各类传感器采集到的振动信号，分析信号的特征和组成，将所有信息在时间坐标轴上显示。常用的信号指标包括峭度、均值、最大最小值、均值方根等。刘迎松等 [8] 针对滚动轴承故障信号在传输路径复杂和强噪声干扰条件下信噪比低、微弱故障特征难以提取的问题，提出了一种经过 SSA 改进的 VMD 算法与 MCKD 相结合的方法来提取微弱故障特征。王新刚等 [9] 通过选择有效模态分量、确定 VMD 中 K 值，选择有效模态分量，通过粒子群优化算法优化 MCKD 参数并进行信号增强，最后进行信号包络解调提取故障特征频率，实现滚动轴承故障特征频率的提取和故障诊断。陈远帆 [10] 等将高斯密度函数与网格搜索法相结合，选取均值、方差、峭度等时域特征参数进行信号特征提取特征提取，并将分析结果应用支持向量机模型获得了较高的轴承复杂故障诊断准确率。时域特征分析的过程针对了信息最完整的原始振动信号，简单明了方便计算，但是统计分析容易受到工况信息的影响，不能从特征信息中提取周期性，对噪声变化比较为敏感，不能获得任何频域信息，应用场景受限。

频域分析是将时域信号转变为频域信号进行分析的方法。在轴承故障检测频域分析方法具有更高的精度，该方法主要以傅里叶变换为基础，将信号分解为不同幅值和频率的正弦信号叠加。通过傅里叶变换得到的频谱、幅值谱、包络谱、功率谱、倒频谱以及细化谱都是常见的谱分析方法。这些谱分析法虽然提出时间较久，但是现在仍然被普遍应用于振动信号特征提取。刘敏等 [11] 将自适应噪声完备集合经验模态分解 (CEEMDAN) 与倒频谱相结合的方法对轴承的振动信号进行分析，经过 CEEMDAN 算法对振动信号降噪后，利用倒频谱算法对提取信号特征，得到精确的振动信号图谱。江志农等 [12] 提出利用最小熵卷积、自回归线性预测和小波去噪 3 种特征增强方法，逐步增强齿轮振动信号中的故障冲击特征，再对增强后的数据进行倒频谱分析，提取故障特征。赵克钦 [13] 等针对轴承转速变化会削弱快速谱峭度图方法特征提取性能的问题，通过对数包络阶次循环分量构造出对数包络阶次循环分量图得出最优调解频带，实现了在速度波动较大时，能够准确确定故障类型。在研究过程中，为了解决频域分析法对于非平稳振动信号无法解读频率在时间维度的信息的缺点，产生了小波变换、小波包变换、经验模态分解、短时傅里叶变换和维格纳-威尔分布等时频域分析方法。时频域分析方法能够描述信号的局部细节特征，对信号的频率与时间的相互关系进行反应，对于分析处理非平稳信号非常有利。小波变换相较于傅里叶变换是信号时间和频率的局部变换，从而有效地从信号中提取到信息，克服了傅里叶变换在处理非平稳信号时的局限性，是一种常用的时频域分析方法。从小波变换衍生出的方法有连续小波变换、离散小波变换、小波包变化等王茜等 [14] 提出将基于改进经验的小波变换的特征提取方法，结合了相关性阈值去噪的降噪能力，实现了在噪声影响下能准确提取振动信号特征。杨健等 [15] 在使用小波包变换方法时同时引入 Canopy 算法，使得模型算法具有更好的抗噪性和工作效率。陈代俊 [16] 通过结合 VMD—CWT 方法将振动信号转变为二维时频图，避免了人工提取故障特征不够充分的缺点，使模型识别精准度更高。

经验模态分解是一种自适应非线性非平稳信号时频分析方法。其原理是对非平稳信号分解为多个不同本征模函数 (IMF)，IMF 分量则包含了原始信号的不同时间尺度的局部特征信息。但是，经验模态分解也存在一些问题，针对强噪声背景，经验模态分解对噪声的敏感性会使分解结果发生模式混叠现象，造成结果失真。针对这一问题，有学者提出了集成经验模态分解方法 (EEMD)，该方法能有效针对随机噪声下的轴承故障，具有速度

快、误差小、性能稳定的特点。李军星等 [17] 用 EEMD 对振动数据进行预处理，降低了对数据特征提取难度，并削弱了噪声对数据的影响，为后续故障预测提高了预测精度。

在此之后一种新的信号模态分解方法被提出，即变分模态分解（VMD）。变分模态分解采用变分框架与交替方向乘子法对信号进行模态分解，同时提取出各模态的中心频率，克服了经验模态分解存在模态混叠的缺陷。并且相较于缺乏理论基础的经验模态方法，变分模态分解实质是多个适应维纳滤波器组，具有坚实的理论基础。黄晓诚等 [18] 基于优化处理的变分模态分解方法，对归一化处理对数据进行降噪重构，提取了故障特征并计算了归一化指标集合，通过多层感知器的神经网络对归一化数据进行训练，得到一种高精度电机轴承故障检测器。孟宗 [19] 等利用分量信号的能量之和占原信号能量的比值来判断变分模态分解的分解效果，从而找出最佳分解层数，并根据分量信号间的相关系数，判断中心频率相邻的分量信号与原信号的关系，通过主要分量的包络谱找出故障特征频率，判断故障类型。

短时傅里叶变换通过窗函数将原信号截取为多段信号进行傅里叶变换，得到信号的时间-频率联合分布。刘飞等 [20] 将短时傅里叶变换生成得到二维时频图，再通过伪色彩处理的方法能够将时域振动信号迅速转换为三通道图像数据，通过迁移学习和深度残差网络方法实现快速故障诊断。相对于短时傅里叶变换，维格纳-威尔分布具有良好的时频分辨性和时频聚集性，能够提取信号的整体信息特征，是分析非平稳信号的重要方法。王亚萍等 [21] 通过 SPVDM 时频分析法分析轴承故障振动信号，在时频图中提取出故障特征。当以多分量信号为研究对象时，维格纳-威尔分布的应用会受到限制，因此在分析多分量信号时会出现交叉干扰项，该缺陷会使其检测性能受到影响。

2.1.2 基于传统机器学习的故障分类方法

通过对轴承信号进行处理，可以基于提取的信号特征进行故障分类，最终诊断出轴承出现的故障类型，是轴承故障诊断的重要部分，选择合适的模式识别方法非常重要。在轴承故障诊断中机器学习算法 [50-57] 按照训练数据是否为标记数据可被分为监督学习和无监督学习 [66-72]，常用的算法包括支持向量机、人工神经网络 [58-65]、k 近邻算法、决策树与随机森林等算法。

支持向量机是一种典型的监督学习算法（如图 2 所示），被广泛的应用在轴承故障检测领域，并且对于支持向量机的模型优化也有了长足的发展。例如：网格优化算法、遗传算法等传统优化算法和粒子群优化、麻雀搜索算法等群智能优化算法，实现了自适应训练的支持向量机分类模型。

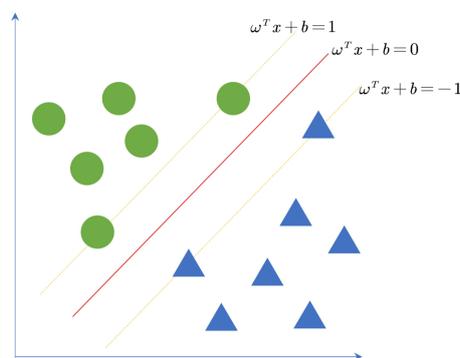


图 2. 支持向量机原理图

学者们提出的多核学习方法解决了支持向量机的核函数选择问题，使得该方法具有更大的灵活性和分类精度。陈司昱 [22] 根据对比试验分析得到最佳核函数，基于径向基核函数建立了支持向量机分类模型，采取包裹式选择特征进行两次特征选择，为支持向量机模型选择合适的故障特征。汤天宝 [23] 等先通过奇异谱分解方法

构建特征矩阵，通过两层结构支持向量机算法进行故障识别，提高了基于支持向量机算法等故障诊断准确率。神经网络是一种多层前馈神经网络（如图 3 所示），可以通过调整大量网络节点之间的关联关系来进行模式识别。因其能充分逼近复杂的非线性关系，是建立滚动轴承故障诊断模型的有效途径。基于人工神经网络的

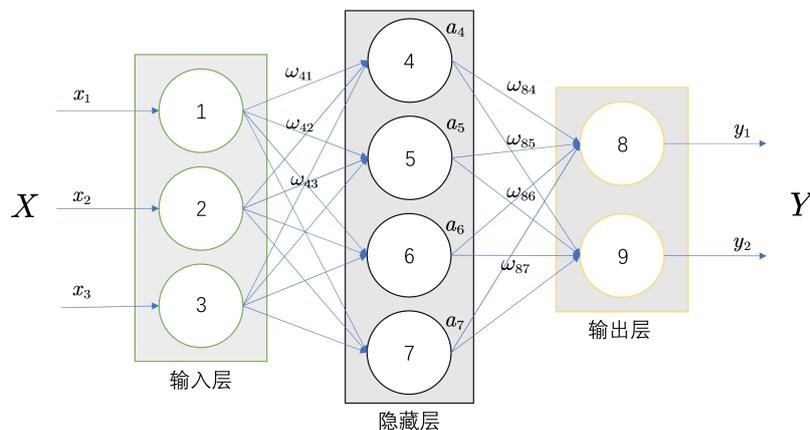


图 3. 神经网络原理图

故障诊断模型具有很强的自学能力，通过最小化经验风险，实现诊断模型的自动学习。易静姝 [24] 介绍了人工神经网络在轴承故障领域中的应用，分析了多种模型的优势和缺点。该模型存在随着模型复杂度提高，训练效率下降以及模型缺乏理论支持，可解释性较差的缺点。刘斌 [25] 等通过反向传播算法来训练双隐层前馈神经网络，将特征信号通过经验模态分析处理作为输入，最终则出故障诊断结果，使得该模型有较强等适应性。

k 近邻算法属于无监督学习类型算法（如图 4 所示），是一种原理简单的常用算法之一，能够根据已知类别的训练样本与待分类样本之间的最小距离差，搜索待分类样本的 K 个最近的邻居，然后根据最近的 K 个样本做出样本类别的判断，实现故障类别的分类。朱兴统 [26] 利用小波包分解原始数据，将构建好的特征向量输入 k 近邻算法进行轴承故障诊断。

决策树算法无需推理诊断模型和繁重的领域知识，是一种基于数据特征进行归纳推理从而实现分类的算法。丁明彬 [27] 等将小波变换和决策树模型进行结合，通过二次构建决策树模型来提高故障识别的准确率。随机森林算法是一种利用集成思想将决策树进行集成的算法。该方法具有可调参数少、过拟合风险低、分类精度高等优点。沙盟 [28] 等针对传统随机森林算法在故障诊断中会出现复杂情况下过拟合的现状，利用麻雀搜索算法优化随机森林中的树木个数和最小叶子节点数，改进传统随机森林算法，建立基于改进随机森林算法的故障诊断模型。

2.1.3 小结

以上是对典型故障诊断方法的概述，主要从特征提取和故障分类两方面展开。对于故障特征提取的方法中，时域分析方法具有涵盖信息全面完整、运算速度、快实时性强的优点。但是存在计算精度较差，对振动信号较敏感的特点，无法获取频域信息，使用场景有所限制；频域分析方法相对于时域分析具有较好的计算精度、能详细展示故障信息的优点。频域分析方法无法展示出时域信息、实时性较差，存在非平稳信号处理较差的情况。时频域分析方法以傅立叶变换为基础产生多种分析方法，其中 STFT 能够满足研究局部信号的要求，但无法自

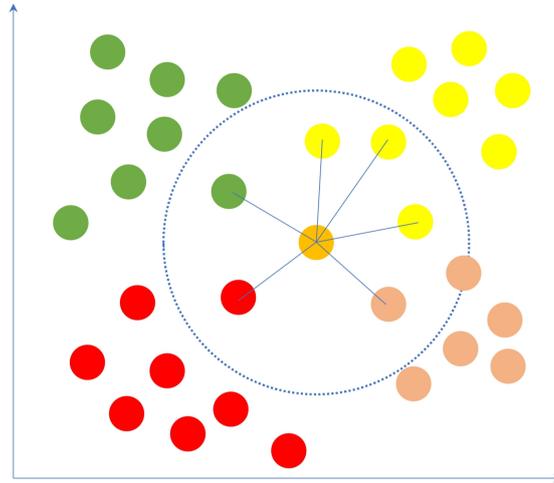


图 4. K 值最近邻原理图

适应调节时频分辨率；经验模态分解方法具有可改进的地方多、自适应强、时频聚集性好的优点和存在边界效应、对噪声相对敏感等缺点；维格纳-威尔分布方法具有适用于强噪声非平稳信号、自适应好的优点和冗余度过大、频率混叠的缺点。基于机器学习的故障分类方法类型较多，其中支持向量机算法具有分类精度高、能够处理高维线性特征的优点，但是算法本身受核函数影响，在大规模样本中表现较差；人工神经网络模型具有鲁棒性好、能充分逼近复杂线性关系的特点，但是计算效率差，容易出现过拟合的问题；对于 k 近邻算法其具有简单易实现、分类和回归问题都适用的优点，但是算法存在受 K 值的影响、依赖训练数据的缺点。不同的方法适用于不同的研究场景，在研究过程中可以针对实际情况选择合适的诊断方法，多个方法相互补短，能达到更好的诊断效果。

2.2 基于深度学习的轴承故障诊断检测

深度学习是机器学习的一个分支，深度学习源于人工神经网络，并受到人脑细胞功能的启发，它允许由多个处理层组成的计算模型学习具有多个抽象级别的数据表示。通过逐层堆叠非线性单元，具有更高的计算效率，针对大模型分析有显著优势。随着智能制造的发展，深度学习算法得到了广泛的应用，极大改善了语音识别、视觉对象识别、对象异常检测等领域的技术水平，在最近五年被引入到故障诊断的领域中。相较于传统的数据分析方法，它可以自动从原始数据中学习有效特征，无需手动提取和操作特征，具有很强的可学习性。应用在故障检测领域的深度学习算法包括堆栈自编码器、卷积神经网络、深度置信网络、循环神经网络等 [29]。

2.2.1 堆栈自编码器

堆栈自编码是一种以自编码器为基本模型构建的多层自编码器（如图 5 所示），具有深度全连接前向神经网络。自编码器是一种无监督的神经网络，自身由输入层、隐层、输出层 3 层网络结构组成，用于对输入信号进行变换。堆栈自编码器通过逐层堆叠多个自编码器来学习数据的多重表达方式。

对于使用自动编码器以及改进模型强大的特征提取能力来进行轴承故障诊断，已经进行了许多研究。但由于堆栈自编码器为非监督学习算法，通常需要结合归一化指数函数来辅助使用。贾美霞等 [30] 利用堆栈自编码器将处理好的数据集进行特征提取，并加入 Dropout 层和批标准化层，以提高特征的提取性能。苏靖涵 [31] 等针对存在轴承故障样本不足，训练样本与实际信号样本存在分布差异的问题，利用深度自编码器对迁移学习中源域的训练数据和目标域中的正常样本进行处理和微调，使得该模型在不同工况下的诊断精度更高。

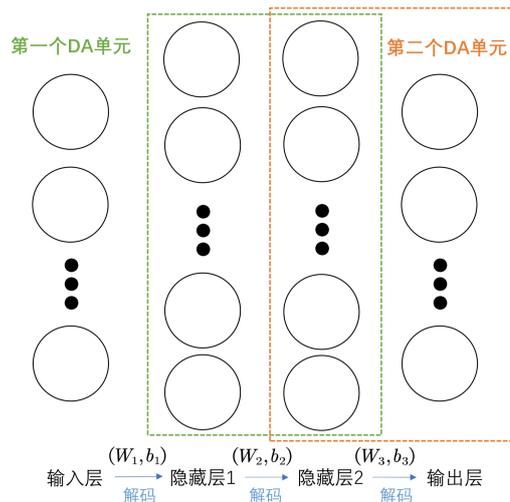


图 5. 堆栈自编码原理图

2.2.2 卷积神经网络

卷积神经网络是一种多层感知机神经网络（如图 6 所示），通过卷积运算代替一般神经网络的乘法运算，使用多数卷积核进行卷积处理，从而提取不同类型的特征。作为深度学习算法的典型模型，卷积神经网络通过反向传播算法更新网络参数，由卷积层、池化层和全连接层组成。卷积神经网络能够直接从原始数据中提取并学习特征，不依靠数据预处理，并且模型收敛速度快，不易产生过拟合，并且由于卷积神经网络的局部连接，权值参数共享的优点，可以减少模型计算的复杂度。其中，一维卷积神经网络是一个相对较浅的结构，其中的简单阵列

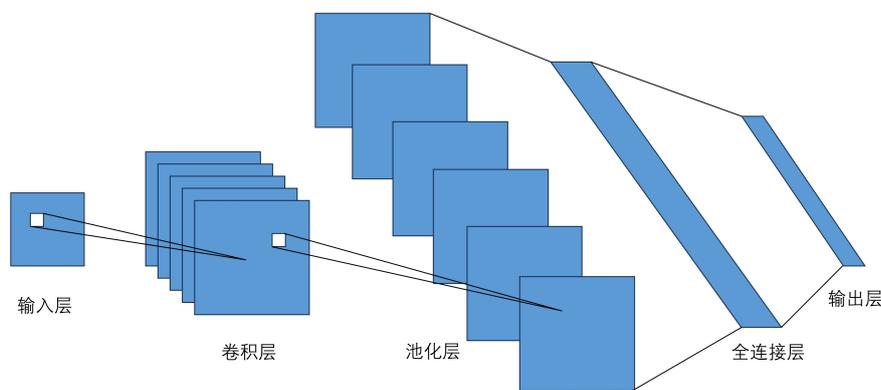


图 6. 卷积神经网络原理图

操作使得该模型在处理一维信号时更有效，因此该算法成功应用于轴承故障诊断领域。龚俊等 [32] 引入 SRM 模块到卷积层，加入风格融合，将各个风格信息纳入特征图，突出有用信息，忽略无效信息，将一维非平稳振动信号经过 SWT 转换为二维视频图像作为卷积神经网络的输入，提升了诊断精度。李辉 [33] 基于 CCSD 能有效一直高斯和非高斯噪声的特点，将一维信号转变为二维的循环相关熵谱密度，优化了卷积神经网络输入信号质量，解决了 CNN 在噪声干扰下故障识别精准度下降的问题。董绍江 [34] 等针对强噪声下故障难以诊断的问题，提出了结合奇异值分解、经验模态分解来对原始信号进行降噪，并最终与原始信号进行叠加的多层降噪方

法，同时引入注意力机制改进卷积神经网络模型来提高模型诊断精度。

2.2.3 深度置信网络

深度置信网络是由受限玻尔兹曼机的许多中间层组成的网络架构（如图 7 所示），这些网络层又分为显性层和隐性层，其中每个受限玻尔兹曼机层与前边的和后边的层进行通信，层内没有通信，最后一层用于分类。除了第一层和最后一层，深度置信网络中的每一层即可以充当早期节点的隐藏层，也可以在后续作为输入层和可见层。

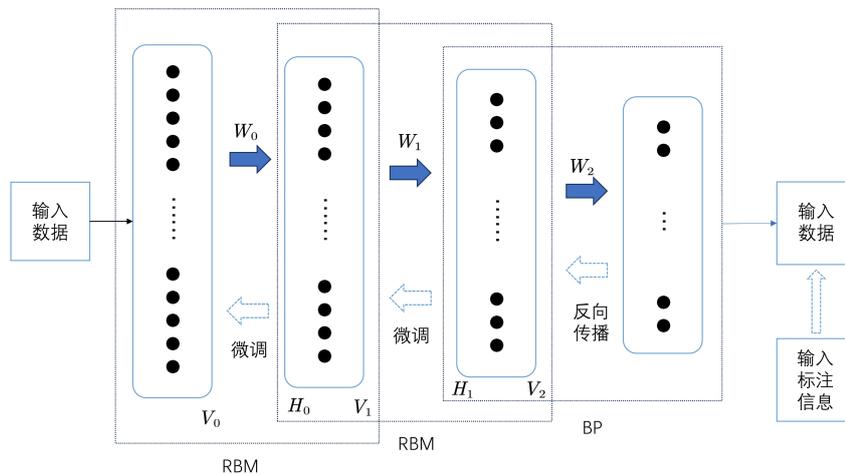


图 7. 深度置信网络原理图

深度置信网络既可以用于非监督学习，也可以用于监督学习，这些网络可以解决深度学习中的低速和过拟合的问题，通过训练神经元间的权重，能够让整个网络结构按照最大概率来生成训练数据，能够用于特征提取、故障分类和预测。刘鹏 [35] 利用 DBN 能处理高维非线性数据和有效提取故障特征的特点，建立了基于 DBN 网络的滚动轴承故障模型，针对浅层神经网络难以从振动信号中提取故障特征的难点提出了解决方案。沈长青 [36] 等针对传统动量法引起的错过最优点问题，提出一种基于 Nesterov 动量法的独立自适应学习率优化的深度信念网络，引入 Nesterov 动量法代替传统动量法预测参数下降的位置，控制参数达到最优点的速度，提高了模型的泛化能力。

2.2.4 循环神经网络

循环神经网络是一种具有短期记忆能力的神经网络（如图 8 所示）。它允许在具有隐藏层的同时将早期输出作为输入进行处理，能得到数据点间的时间关联性，因此循环神经网络能够反复处理数据，是一种可以用于处理时间序列的神经网络。

循环神经网络模型中的长短时记忆网络被广泛应用于滚动轴承振动信号的时间序列分析。邵良杉 [37] 采用改进的 HHO 算法对 LSTM 的输入参数进行优化，代替了人工选取参数，提高了一定的诊断精度。郑直 [38] 根据长短时记忆神经网络会因适应性动量估计优化器陷入局部极小值，导致故障诊断精度低，以及鲸鱼算法因寻优区域过大，导致寻优效率低的问题，提出了通过改进鲸鱼算法来优化长短时记忆神经网络的方法，来提高故障诊断精度。

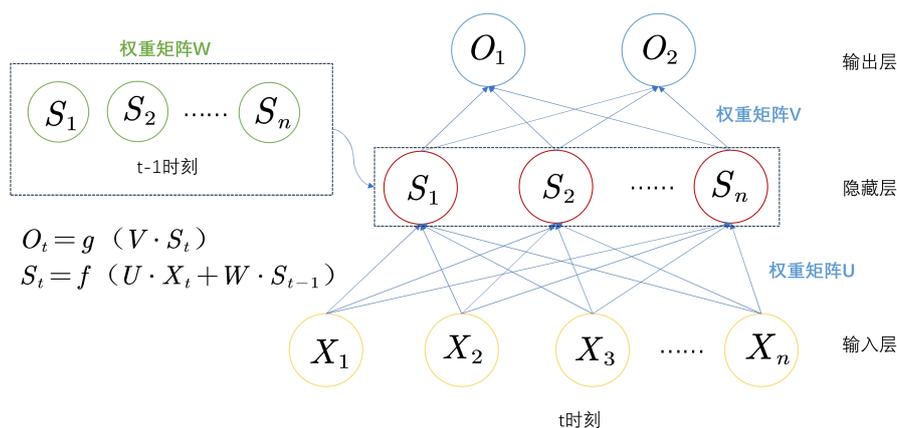


图 8. 循环神经网络原理图

2.2.5 小结

本节所提到的模型为滚动轴承故障诊断领域的常用算法，其中堆栈自编码算法具有能够处理复杂信号、逐层学习数据特征的优点，但存在模型训练困难、受参数选择的限制；卷积神经网络模型具有在卷积层所用权重和偏置参数较少、隐式学习训练数据的特点以及模型深度有限和训练成本较高的限制；深度置信网络算法具有较好的泛化能力，通过层级表示学习能够找出潜在故障，同样的该算法训练所诉时间久、容易出现过拟合的现象，模型的可解释性较差；循环神经网络适合处理时间序列数据，能够自动提取特征，但是在模型训练上容易出现梯度消失或梯度爆炸问题，且泛化能力有限。

总的来说，深度学习相较于传统的故障诊断方法更加出色，自学习能力强，且覆盖范围广、算法类型丰富，但是目前深度学习还存在一些局限性，例如对训练样本的需求量很大、训练模型的时间成本高、且模型复杂可解释性较差，这些仍是目前深度学习领域需要面对的难题。

2.3 基于迁移学习的轴承故障诊断方法

尽管当前深度学习在轴承故障检测领域有着很好的检测效率和检测精度，但这些能力的实现都是基于充足的故障数据和标签。然而在实际工程场景中，轴承通常在健康的状态下运作，故障样本量非常少；同时采集到带有标签的数据成本较高。如果对于数据的要求无法满足，那么深度学习方法的效率会下降，影响了机器学习算法在该领域的应用。而迁移学习能够针对故障样本较少这一问题，充分利用正常样本或关联数据集的信息，以解决目标工况中样本不足以及无标签的问题。

迁移学习方法属于深度学习算法的一种，其核心思想为找到两个或多个相关但是不同的学习任务共同特征，并应用共享知识来解决新领域的问题。迁移学习中的两个基本概念为领域 (Domain) 与任务 (Task)，其中迁移学习中的领域又包含源域 (Source domain) 和目标域 (Targetdomain)。源域指包含具有充足标签和数据的领域，目标域是指无标签或数据不足的领域，任务指源域和目标域的标签或标签函数。迁移学习领域可以按学习方法归分为归纳迁移学习、直推迁移学习和无监督迁移学习三种。不同的迁移学习方法有不同的特点和适用领域，如表所示 [39]。

轴承故障诊断中的迁移学习方法通常以故障样本丰富的实验数据集为源域，挖掘源域中的知识，找到源域与目标域间数据分布的相似性，找到源域与目标域之间数据分布的相似性，将源域中学习到的知识应用于目标域实现故障诊断。

表 1. 迁移学习类型及特征

迁移学习类型	相关领域	源域标签	目标域标签	目标域标签
归纳迁移学习	多任务学习	有	有	回归和分类
	自学习	无	有	回归和分类
直推迁移学习	自学习	无	有	回归和分类
无监督迁移学习	领域适应、采样选择偏置等	无	无	聚类和降维

王鹏 [40] 提出改进的交替迁移学习方法，分层计算不同的损失函数，避免了源域数据与目标域数据特征知识空间内的分布差异大的问题，降低了网络的计算难度，获得了更好的训练效果。高丽鹏 [41] 构建基于深度多模态迁移诊断模型，利用大量带标签的源域数据预训练模型，通过域适应方法来最小化源域和目标域之间的分布差异，提高了模型精度。

现有的迁移学习方法为机械故障诊断领域做出了很大的贡献，该方法克服了样本不足、获取故障样本成本过高的情况，并且该方法解决了没有先验知识的建模问题。但是仍应该指出的问题是目前大多数迁移学习方法都在很大程度上依赖数据预处理，要求源域和目标域的数据维度一致，这需要一定的数据预处理方法，如时频域分析等。迁移学习方法同样在提高诊断性能方面面临挑战，当面对在不同工况下进行转换是，变化的特征会出现转移误差，并且由于实验室和工业实际情况的尺度不同，尺度变化的问题会导致机械系统之间的转移误差。这些都是迁移学习方法当前需要面对和解决的问题。

3 结论

尽管轴承故障检测领域已经有了久远的发展，从多种角度提出了丰富的诊断方法，并获得优秀的成果。但是这些方法仍存在一些局限性，较多数的方法注重于提高诊断性能以及诊断精度上，反而对于诊断算法的智能型以及对多种工况的适应性没有作较深的研究发展。虽然基于深度学习的轴承故障检测方法能够实现自动提取特征、故障识别和分类的功能，但是这些方法大多基于神经网络的诊断模型，需要人工进行结构架构并调整模型参数，这些构建和调整的过程将消耗大量时间和资源。而且，深度学习模型都要基于大量的数据和标签样本进行训练，但在实际工程中，轴承往往在健康状态下运作，故障样本十分匮乏，这都将限制深度学习在该领域的使用和发展 [42]。

当前，滚动轴承故障检测领域对于迁移学习方法的使用越来越多，因其不依赖大量数据和标签样本的特点，为轴承故障检测提供了新的解决方案。迁移学习的原理是能够从一个或多个相关场景中提取知识，以提高在目标域中的学习性能。基于迁移学习的识别模型能从源域数据和目标域数据中学习共同特征空间，减小不同域之间的分布差异。但该方法无法针对目标域自适应调整模型参数，继而会影响模型的适应性和识别精度。因此基于迁移学习方法的故障识别方法可以将提高在多工况下的适应性和更高的识别精度作为未来发展方向。

创新说明

本文的内容概述了当前在滚动轴承故障诊断领域基于振动信号的轴承故障诊断方法。本文按照滚动轴承故障领域的三个发展阶段为线，将本文内容分为传统故障诊断方法、基于深度学习的轴承故障诊断方法以及基于迁移学习的轴承故障诊断方法三部分内容。

本文的创新点和意义在于每部分内容都罗列了当前最常用的、不同类型的故障诊断方法。针对每种诊断分析方法提出了对应的适用情景和模型的局限性，并在每部分的最后对同一类的诊断方法进行总结分析和比较，以供学者进行比对和参考，根据实际情况选择合适的诊断模型，以实现更好的诊断效果。同时分析了迁移学习方法

在滚动轴承故障诊断领域的提供了新的思路和发展方向以及在该领域应用存在的困境和提升空间。

参考文献

- [1] Jiangquan Zang, Yi Sun, Liang Guo, et al. A new bearing fault diagnosis method based on modified convolutional neural networks[J]. Chinese Journal of Aeronautics, 2020, 33: 439-447.
- [2] Duy-Tang Hoang, Hee-Jun Kang. A survey on Deep Learning based bearing fault diagnosis[J]. Neurocomputing, 2019, 335: 327-335.
- [3] 焦静. 基于同轴振动特征融合的滚动轴承故障诊断研究 [D]. 北京交通大学, 2022
- [4] 刘永志. 基于深度迁移学习的滚动轴承故障诊断方法研究 [D]. 西南交通大学, 2022
- [5] 金国强. 基于深度学习的复杂工况下端到端的滚动轴承故障诊断算法研究 [D]. 中国科学技术大学, 2020
- [6] Rai Akhand, Upadhyay S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings[J]. Tribology International, 2016.
- [7] Bodi Cui, ang Weng, Ning Zhang. A feature extraction and machine learning framework for bearing fault diagnosis[J]. Renewable Energy, 2022, 191: 987-997.
- [8] 刘迎松, 魏志刚, 束海星, 等. 基于参数自适应 VMD 和 MCKD 的滚动轴承微弱故障特征提取 [J]. 噪声与振动控制, 2023, 43(03): 102-109.
- [9] 王新刚, 王超, 韩凯忠. 基于优化 VMD 和 MCKD 的滚动轴承早期故障诊断方法 [J]. 东北大学学报 (自然科学版), 2021, 42(03): 373-380, 388.
- [10] 陈远帆, 李舜酩. 基于高斯混合模型与改进网格搜索法的轴承故障诊断 [J]. 重庆理工大学学报 (自然科学), 2016, 30(03): 34-39.
- [11] 刘敏, 叶艳媛, 杨清清, 等. 基于 CEEMDAN 和倒频谱方法的圆锥滚子轴承振动信号分析 [J]. 机电工程技术, 2023, 52(08): 165-170.
- [12] 江志农, 张永申, 冯坤, 等. 基于特征增强倒频谱分析的齿轮故障诊断方法 [J]. 机械传动, 2019, 43(10): 13-17, 55.
- [13] 赵克钦, 程峰, 杨世飞. 变转速下对数平方包络谱在滚动轴承故障诊断中的应用 [J]. 噪声与振动控制, 2023, 43(02): 132-138.
- [14] 王茜, 田慕琴, 宋建成, 等. 基于经验小波变换的振动信号特征量提取 [J]. 振动与冲击, 2021, 40(16): 261-266.
- [15] 杨健, 张永平. 基于小波包和聚类算法的滚动轴承故障检测研究 [J]. 盐城工学院学报 (自然科学版), 2023, 36(01): 67-73.
- [16] 陈代俊, 陈里里, 董绍江. 基于 VMD-CWT-CNN 的滚动轴承故障诊断 [J]. 机械强度, 2023, 45(06): 1280-1285.
- [17] 李军星, 徐行, 贾现召, 等. 基于 EEMD 与 CNN-BiLSTM 噪声环境下滚动轴承故障预测方法研究 [J]. 轴承, 2023.
- [18] 黄晓诚, 贺青川, 陈文华. 基于 VMD 与 MLP 的电机轴承故障检测方法 [J]. 机电工程, 2022, 39(07): 911-918.
- [19] 孟宗, 吕蒙, 殷娜, 等. 基于改进变分模态分解的滚动轴承故障诊断方法 [J]. 计量学报, 2020, 41(06): 717-723.
- [20] 刘飞, 陈仁文, 邢凯玲, 等. 基于迁移学习与深度残差网络的滚动轴承快速故障诊断算法 [J]. 振动与冲击, 2022, 41(03): 154-164.
- [21] 王亚萍, 许迪, 葛江华, 等. 基于 SPWVD 时频图纹理特征的滚动轴承故障诊断 [J]. 振动. 测试与诊断, 2017, 37(01): 115-119, 203.
- [22] 陈司昱. 基于支持向量机的滚动轴承故障特征选择和诊断方法研究 [D]. 长春工业大学, 2023
- [23] 汤天宝, 周志健, 张涛, 等. 基于奇异谱分解和两层支持向量机轴承故障诊断方法 [J]. 噪声与振动控制, 2022, 42(01): 100-105.
- [24] 易静姝. 神经网络在滚动轴承故障诊断中的应用与发展 [J]. 价值工程, 2019, 38(24): 274-276.

- [25] 刘斌, 刘佳, 张海鹏. 基于经验模态分析的机床主轴轴承外圈非接触式故障检测方法 [J]. 制造技术与机床, 2023, No.727(01): 21-28.
- [26] 朱兴统. 基于小波包分解和 K 最近邻算法的轴承故障诊断方法 [J]. 装备制造技术, 2020, (02): 24-27, 45.
- [27] 丁明彬. 基于小波变换和决策树的电机滚动轴承故障诊断 [J]. 内燃机与配件, 2023, No.395(23): 54-57.
- [28] 沙盟. 基于改进随机森林的电机轴承故障诊断研究 [D]. 太原科技大学, 2023
- [29] Bo Peng, Ying Bi, Bing Xue, et al. A Survey on Fault Diagnosis of Rolling Bearings[J]. Algorithms, 2022, 15: 347.
- [30] 贾美霞, 韩宝坤, 王金瑞, 等. 基于迁移堆栈自编码器的轴承故障诊断方法 [J]. 噪声与振动控制, 2021, 41(06): 84-89, 125.
- [31] 苏靖涵, 张潇. 基于深度迁移自编码器的变工况下滚动轴承故障诊断方法 [J]. 计算机测量与控制, 2021, 29(07): 85-90, 99.
- [32] 龚俊, 张月义, 陈思戢, 等. 基于 SWT 与改进卷积神经网络的轴承故障诊断 [J]. 现代电子技术, 2024, 47(06): 68-74.
- [33] 李辉, 徐伟杰. 噪声干扰下的 CCSD-CNN 轴承故障诊断方法 [J]. 轴承, 2023, (10): 93-100.
- [34] 董绍江, 裴雪武, 吴文亮, 等. 基于多层降噪技术及改进卷积神经网络的滚动轴承故障诊断方法 [J]. 机械工程学报, 2021, 57(01): 148-156.
- [35] 刘鹏, 皮骏, 胡超. 基于 DBN 网络的滚动轴承故障诊断 [J]. 组合机床与自动化加工技术, 2024, (01): 140-144.
- [36] 沈长青, 汤盛浩, 江星星, 等. 独立自适应学习率优化深度信念网络在轴承故障诊断中的应用研究 [J]. 机械工程学报, 2019, 55(07): 81-88.
- [37] 邵良杉, 朱思佳. 基于改进 HHO-LSTM 的滚动轴承故障诊断研究 [J]. 机械强度, 2024, 46(01): 17-23.
- [38] 郑直, 张华钦, 潘月. 基于改进鲸鱼算法优化 LSTM 的滚动轴承故障诊断 [J]. 振动与冲击, 2021, 40(07): 274-280.
- [39] Zhang Shen, Zhang Shibo, Wang Bingnan. Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review[J]. IEEE Access, 2020, 8: 29857-29881.
- [40] 王鹏, 李丹青, 王恒. 基于改进交替迁移学习的滚动轴承故障诊断算法 [J]. 振动与冲击, 2024, 43(05): 239-249.
- [41] 高丽鹏, 雷文平, 曹亚磊, 等. 深度多模态迁移学习在轴承故障诊断中的研究 [J]. 机械设计与制造, 2023: 1-5.
- [42] Neupane Dhiraj, Seok Jongwon. Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset With Deep Learning Approaches: A Review[J]. IEEE Access, 2020, 8: 93155-93178.
- [43] Liu, J., Wang, Q., Song, Z., & Fang, F. (2021). Bottlenecks and countermeasures of high-penetration renewable energy development in China. Engineering, 7(11), 1611-1622.
- [44] Zhang, J., Feng, J., Zhou, Y., Fang, F., & Yue, H. (2012). Linear active disturbance rejection control of waste heat recovery systems with organic Rankine cycles. Energies, 5(12), 5111-5125.
- [45] Wang, W., Liu, J., Zeng, D., Fang, F., & Niu, Y. (2020). Modeling and flexible load control of combined heat and power units. Applied Thermal Engineering, 166, 114624.
- [46] Liu, J., Song, D., Li, Q., Yang, J., Hu, Y., Fang, F., & Joo, Y. H. (2023). Life cycle cost modelling and economic analysis of wind power: A state of art review. Energy Conversion and Management, 277, 116628.
- [47] Fang, F., & Xiong, Y. (2014). Event-driven-based water level control for nuclear steam generators. IEEE Transactions on Industrial electronics, 61(10), 5480-5489.
- [48] Fang, F. A. N. G., Tan, W., & Liu, J. Z. (2005). Tuning of coordinated controllers for boiler-turbine units. Acta Automatica Sinica, 31(2), 291-296.
- [49] Lv, Y., Lv, X., Fang, F., Yang, T., & Romero, C. E. (2020). Adaptive selective catalytic reduction model development using typical operating data in coal-fired power plants. Energy, 192, 116589.
- [50] Wang, N., Fang, F., & Feng, M. (2014, May). Multi-objective optimal analysis of comfort and energy management

- for intelligent buildings. In The 26th Chinese control and decision conference (2014 CCDC) (pp. 2783-2788). IEEE.
- [51] Zhang, X., Fang, F., & Liu, J. (2019). Weather-classification-MARS-based photovoltaic power forecasting for energy imbalance market. *IEEE Transactions on Industrial Electronics*, 66(11), 8692-8702.
- [52] Wei, L., & Fang, F. (2016). H_∞ -LQR-Based Coordinated Control for Large Coal-Fired Boiler-Turbine Generation Units. *IEEE Transactions on Industrial Electronics*, 64(6), 5212-5221.
- [53] Fang, F., Zhu, Z., Jin, S., & Hu, S. (2020). Two-layer game theoretic microgrid capacity optimization considering uncertainty of renewable energy. *IEEE Systems Journal*, 15(3), 4260-4271.
- [54] Liu, J., Zeng, D., Tian, L., Gao, M., Wang, W., Niu, Y., & Fang, F. (2015). Control strategy for operating flexibility of coal-fired power plants in alternate electrical power systems. *Proceedings of the CSEE*, 35(21), 5385-5394.
- [55] Fang, F., & Wu, X. (2020). A win-win mode: The complementary and coexistence of 5G networks and edge computing. *IEEE Internet of Things Journal*, 8(6), 3983-4003.
- [56] Lv, Y., Fang, F. A. N. G., Yang, T., & Romero, C. E. (2020). An early fault detection method for induced draft fans based on MSET with informative memory matrix selection. *ISA transactions*, 102, 325-334.
- [57] Fang, F., Jizhen, L., & Wen, T. (2004). Nonlinear internal model control for the boiler-turbine coordinate systems of power unit. *PROCEEDINGS-CHINESE SOCIETY OF ELECTRICAL ENGINEERING*, 24(4), 195-199.
- [58] Xu, D., Zhu, Z., Liu, C., Wang, Y., Zhao, S., Zhang, L., ... & Cheng, K. T. (2021). Reliability evaluation and analysis of FPGA-based neural network acceleration system. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 29(3), 472-484.
- [59] Li, W., Wang, Y., Li, H., & Li, X. (2019, January). P3M: a PIM-based neural network model protection scheme for deep learning accelerator. In *Proceedings of the 24th Asia and South Pacific Design Automation Conference* (pp. 633-638).
- [60] Wang, Y., Deng, J., Fang, Y., Li, H., & Li, X. (2017). Resilience-aware frequency tuning for neural-network-based approximate computing chips. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 25(10), 2736-2748.
- [61] Qu, S., Li, B., Wang, Y., Xu, D., Zhao, X., & Zhang, L. (2020, July). RaQu: An automatic high-utilization CNN quantization and mapping framework for general-purpose RRAM accelerator. In *2020 57th ACM/IEEE Design Automation Conference (DAC)* (pp. 1-6). IEEE.
- [62] Wang, C., Wang, Y., Han, Y., Song, L., Quan, Z., Li, J., & Li, X. (2017, January). CNN-based object detection solutions for embedded heterogeneous multicore SoCs. In *2017 22nd Asia and South Pacific design automation conference (ASP-DAC)* (pp. 105-110). IEEE.
- [63] Xu, D., Chu, C., Wang, Q., Liu, C., Wang, Y., Zhang, L., ... & Cheng, K. T. (2020, October). A hybrid computing architecture for fault-tolerant deep learning accelerators. In *2020 IEEE 38th International Conference on Computer Design (ICCD)* (pp. 478-485). IEEE.
- [64] Liu, C., Chu, C., Xu, D., Wang, Y., Wang, Q., Li, H., ... & Cheng, K. T. (2021). HyCA: A hybrid computing architecture for fault-tolerant deep learning. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 41(10), 3400-3413.
- [65] Wang, Y., Li, H., & Li, X. (2017). A case of on-chip memory subsystem design for low-power CNN accelerators. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(10), 1971-1984.
- [66] Lian, S., Han, Y., Chen, X., Wang, Y., & Xiao, H. (2018, June). Dadu-p: A scalable accelerator for robot motion planning in a dynamic environment. In *Proceedings of the 55th Annual Design Automation Conference* (pp. 1-6).
- [67] Chang, K., Wang, Y., Ren, H., Wang, M., Liang, S., Han, Y., ... & Li, X. (2023). Chipgpt: How far are we from natural language hardware design. *arXiv preprint arXiv:2305.14019*.

- [68] Wang, Y., Han, Y., Zhang, L., Li, H., & Li, X. (2015, June). ProPRAM: Exploiting the transparent logic resources in non-volatile memory for near data computing. In Proceedings of the 52nd Annual Design Automation Conference (pp. 1-6).
- [69] Li, C., Wang, Y., Liu, C., Liang, S., Li, H., & Li, X. (2021). GLIST: Towards in-storage graph learning. In 2021 USENIX Annual Technical Conference (USENIX ATC 21) (pp. 225-238).
- [70] Han, Y., Wang, Y., Li, H., & Li, X. (2014, November). Data-aware DRAM refresh to squeeze the margin of retention time in hybrid memory cube. In 2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (pp. 295-300). IEEE.
- [71] Wu, B., Wang, C., Wang, Z., Wang, Y., Zhang, D., Liu, D., ... & Hu, X. S. (2020). Field-free 3T2SOT MRAM for non-volatile cache memories. IEEE Transactions on Circuits and Systems I: Regular Papers, 67(12), 4660-4669.
- [72] Ma, X., Wang, Y., Wang, Y., Cai, X., & Han, Y. (2022). Survey on chiplets: interface, interconnect and integration methodology. CCF Transactions on High Performance Computing, 4(1), 43-52.



金学波 教授，博士生导师。1994年毕业于吉林大学（原吉林工业大学）获学士学位，1997年毕业于吉林大学（原吉林工业大学）获硕士学位，2004年获得浙江大学控制科学与工程博士学位，导师为孙优贤院士。研究方向为信息融合、模式识别与预测、大数据分析、深度学习等。近年来在相关领域主持了1项国家科技支撑计划课题、4项国家自然科学基金面上项目等多项研究课题。获2021年度中国粮油学会科学技术奖一等奖。在时序信号模式识别、图像目标检测与识别等研究领域，已发表SCI、EI收录等高水平学术论文159篇，其中7篇为ESI高被引论文（前1%）、3篇ESI热点论文（前0.1%），已授权国家发明专利20余项，出版关于传感器信号识别与状态估计、多传感器信息融合的学术专著3部。担任SCI收录期刊Sensors编委，为IEEE/CAA Journal of Automatica Sinica、Knowledge-Based Systems等中科院一区SCI期刊审稿人。

Xuebo Jin (Fellow, ASP) received the B.S. and M.S. degrees in control theory and control engineering from Jilin University, Changchun, China, in 1994 and 1997, and the Ph.D. degree in control theory and control engineering from the University of Zhejiang, Zhejiang, China, in 2004. She was a Senior Visiting Scholar with the University of Illinois at Chicago, Chicago, IL, USA, in 2007. From 2009 to 2012, she was an Assistant Professor with Zhejiang Sci-tech University. Since 2012, she has been a Professor with Beijing Technology and Business University, Beijing, China. Her research includes a variety of areas in information fusion, big data analysis, condition estimation, and video tracking.

Xuebo Jin (Fellow, ASP) received the B.S. and M.S. degrees in control theory and control engineering from Jilin University, Changchun, China, in 1994 and 1997, and the Ph.D. degree in control theory and control engineering from the University of Zhejiang, Zhejiang, China, in 2004. She was a Senior Visiting Scholar with the University of Illinois at Chicago, Chicago, IL, USA, in 2007. From 2009 to 2012, she was an Assistant Professor with Zhejiang Sci-tech University. Since 2012, she has been a Professor with Beijing Technology and Business University, Beijing, China. Her research includes a variety of areas in information fusion, big data analysis, condition estimation, and video tracking.



王继阳 2022 入学北京工商大学智能制造工程专业。研究方向为轴承故障诊断、轴承寿命预测等。

Jiyang Wang enrolled in Beijing Technology and Business University in 2022, majoring in intelligent manufacturing engineering, focusing on bearing fault diagnosis and bearing life prediction.