

基于 Transformer-CNN 的多模态融合模型在工业轴承故障预测中的应用研究

杨涛¹ 王海程² 赵帅业² (通讯作者) 李国新² 纵绚³

1 滨州市科技创新发展研究院, 山东滨州, 256600;

2 滨州职业学院, 山东滨州, 256600;

3 浙江交通职业技术学院, 浙江杭州, 311112;

摘要: 针对工业轴承故障预测中的数据融合难、边缘推理慢及仿真与工况脱节问题, 提出了 MT-CNN 多模态注意力模型与 HILPM 仿真系统。MT-CNN 通过多模态输入与多分支网络提取特征, 结合物理引导的注意力机制实现故障分类与 RUL 预测; HILPM 系统以 PLC 为核心构建边缘-云端协同架构, 集成硬件加速与闭环验证平台, 实现虚实结合的故障仿真与实时推理。实验显示, MT-CNN 在 CWRU 数据集上分类准确率达 98.67%, RUL 预测 RMSE 为 8.32, 以 2.23M 参数实现 3.9ms 推理延迟, 显著优于现有模型; 工业验证中系统故障识别率达 97.2%, 误报率 0.32%, 满足工业实时可靠性要求。

关键词: 滚动轴承; 故障诊断; 跨模态注意力; 边缘计算

DOI: 10.64216/3080-1516.25.02.064

1 绪论

随着工业智能制造向数字化、智能化加速转型^[1], 设备健康管理及故障预警已成为保障生产连续性和降低运营风险的核心。作为旋转机械的关键部件, 轴承的运行可靠性直接关系到生产效率和操作安全。然而, 轴承内圈、外圈等部件易发生多种故障, 是导致意外停机、生产损失和安全问题的主要原因^[2]。工业界正从被动和计划维护转向预测性维护^[3], 以规避停机与资源浪费风险, 亟需通过精准轴承故障预测实现这一目标。

深度学习 (DL)^[2] 在计算机视觉与自然语言处理等领域已取得显著进展^[4], 但仍面临多模态融合困难、物理关联性弱及实时性不足等工业挑战。因此, 亟需开发能够深度融合多模态信息、引入物理先验知识, 并支持端到端协同诊断与预测的新模型。主要研究内容如下:

(1) 提出多模态注意力融合策略, 动态融合多域特征并抑制噪声, 在 CWRU 数据集准确率达 98.67%, 噪声环境下性能优异;

(2) 设计轻量化时频 Transformer, 参数量仅 0.6 M, 特征提取精度 86.7%, 推理速度达 ViT-Tiny 的 3.2 倍, 支持边缘实时诊断;

(3) 构建硬件在环预测性维护系统, 融合边缘与云端能力, 通过专用 AI 扩展卡提升 PLC 算力, 实现高精度低延迟预警。

2 国内外现状

近年来, 随着工业物联网 (IIoT) 与人工智能技术的融合发展, 设备预测性维护已成为工业智能化转型的核心研究方向。

在智能算法层面, 深度学习模型 (如 CNN、LSTM) 因强大的特征学习能力被广泛应用于故障诊断。例如, 刘明和陈辉提出基于卷积神经网络 (CNN) 的机械设备故障诊断方法, 通过频域特征与时域统计量的融合, 显著提升了对轴承表面缺陷的识别精度^[5]; 高翔和王磊则构建了基于长短期记忆网络 (LSTM) 的旋转机械故障预测模型, 通过学习时间序列数据中的微小变化, 实现了对风力发电机状态的早期预警^[6]。

在仿真验证层面, 硬件在环^[7]技术已被用于控制系统测试, 但将其与预测性维护结合的研究较少。现有 HIL 平台缺乏精细化故障模拟与闭环验证能力, 且边缘-云端协同延迟高、反馈慢, 制约系统实时性与适应性^[8]。

当前研究在算法精度、部署效率及仿真验证方面存在突出矛盾。亟需通过多模态融合、模型轻量化与深度硬件在环集成, 破解预测性维护的落地瓶颈, 已成为核心科学与工程挑战。

3 MT-CNN 模型算法设计

3.1 模型框架

针对 CWRU 数据集的特点, 提出一种融合多模态数据与注意力机制的 MT-CNN 模型, 通过多模态输入、多分支特征提取、多模态注意力融合和多任务输出四个模

块, 实现多域特征高效协同, 构建端到端的智能诊断闭环。模型结构如图 1 所示。

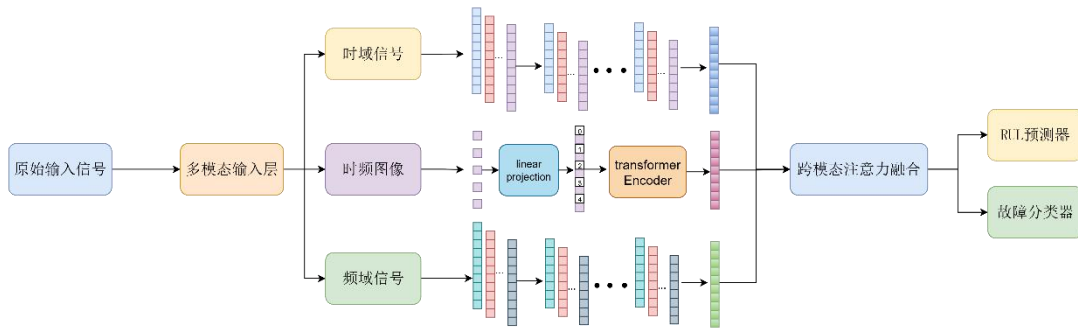


图 1 MT-CNN 模型架构图

3.2 多模态输入模块

3.2.1 数据预处理

CWRU 轴承数据集是故障诊断领域广泛使用的基准数据集。针对其特性, 对原始振动信号进行以下预处理:

(1) 滑动窗口分割

原始振动信号分割为 L 点的窗口, 切分相邻两个样本重叠率取 50%, 提升时序建模连续性。

$$X = [x_1, x_2, \dots, x_L] \in R^L \quad (1)$$

(2) 标准化处理

对每个窗口内的信号进行零均值化与单位方差归一化, 消除设备差异与噪声干扰。

$$\hat{x}_i = \frac{x_i - \mu}{\sigma}, \mu = \frac{1}{L} \sum_{i=1}^L x_i, \sigma = \sqrt{\frac{1}{L} \sum_{i=1}^L (x_i - \mu)^2} \quad (2)$$

3.2.2 多模态数据生成

为构建多视角特征表征, 生成以下三类模态数据:

(1) 时域波形

直接使用预处理后的信号, 该模态保留原始冲击脉冲时序特征, 有助于捕获瞬时故障特征。

$$M_{time} = \hat{X} \in R^{1 \times L} \quad (3)$$

(2) 频域波形

采用快速傅里叶变换、幅度谱提取及对数压缩, 增强对低幅值谐波的敏感性。具体步骤如下:

a. 计算快速傅里叶变换 (FFT)。

$$F(k) = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}, k = 0, 1, \dots, N-1 \quad (4)$$

b. 提取幅度谱并进行对数压缩。

$$A = |F| \in R^{L/2+1}, M_{freq} = \log(A + \phi) \in R^{D \times L'} \quad (5)$$

(3) 时频图像

通过连续小波变换生成。具体步骤如下:

a. 计算 CWT 系数, 输出尺度-时间矩阵。

$$Coef(a, b) = |CWT(a, b)|^2, CWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-b}{a}\right) dt$$

$$\psi(t) = \pi^{-1/4} e^{ia\omega t} e^{-t^2/2}$$

(6)

b. 对数压缩, 增强低能量区域可视化。

$$S(a, b) = \log_{10}(Coef(a, b) + \varepsilon) \quad (7)$$

c. 尺度转频率, 将尺度轴转为频率轴, 输出时频矩阵。

$$P(f, b) = S(a(f), b), f_a = \frac{f_c \cdot f_s}{a \cdot \Delta} \quad (8)$$

d. 图像标准化, 消除设备差异, 适配 Transformer 输入。

$$I(f, t) = 255 \times \frac{P(f, t) - P_{min}}{P_{max} - P_{min}} \quad (9)$$

3.3 多分支特征提取模块

3.3.1 时域/频域 CNN 分支特征提取

时域/频域分支采用 1D-CNN 处理输入信号, CNN 时域分支包含三个卷积层、两个池化层和一个全局自适应池化层。输入信号经卷积、ReLU 和池化操作重复提取, 最终通过全局池化输出特征。

3.3.2 时频图像 Transformer 分支特征提取

结合 CWRU 数据集特点, 简化计算流程, 在传统的 transform 模型基础上做以下优化简化, 具体步骤如下:

a. 简化多头自注意力机制。

$$\text{SimpleAttention}(Q, K) = \text{softmax}\left(\frac{QK^T}{\sqrt{dk}}\right)X, Q = XW^Q, K = XW^K \quad (10)$$

b. 前馈网络简化。

$$\text{SimpleFFN}(x) = \text{GELU}\left(xW_1 + b_1\right)\left(W_2 \in R^{d \times d}\right) \quad (11)$$

3.4 多模态注意力融合模块

多模态注意力融合模块首先对齐时域、频域和时频域特征，以时域为 Query，其余为 Key/Value，通过多头注意力与门控机制实现动态加权融合，生成判别性强的统一特征。关键技术包括：

(1) 特征投影实现维度对齐

采用全连接层将多模态特征映射至统一维度，并通过残差连接与层归一化确保特征一致性，为后续跨模态交互提供基础。

$$V'_t = \text{ReLU}(W_t V_t + b_t) \quad (12)$$

(2) 交叉注意力机制

Query 生成，以时域特征为 Query 向量，通过线性变换提取其全局时序模式：

$$Q = V'_t \cdot W_q \quad (13)$$

Key-Value 生成，将频域与时频域特征分别作为 Key 和 Value 向量，通过可学习参数矩阵进行特征解耦：

$$K = V_{f'} \cdot W_k, V = V_{f'} \cdot W_v \quad (14)$$

注意力权重，采用点积注意力计算跨模态相关性，权重公式为：

$$\alpha = \text{Softmax}\left(\frac{QK^T}{\sqrt{dk}}\right) \quad (15)$$

通过多头注意力聚合不同子空间的注意力分布，经拼接和线性映射生成融合特征，同时保留原始模态的语义信息。

$$F_{\text{fusion}} = \alpha \cdot V \quad (16)$$

(3) 动态加权

门控机制平衡模态贡献，引入门控单元，自适应调节各模态权重。

$$g = \sigma(W_g[V'_t; V'_f]) \quad (17)$$

$$F_{\text{out}} = g \cdot V'_t + (1 - g) \cdot F_{\text{fusion}} \quad (18)$$

3.5 多任务输出模块

本模块采用双分支结构，分别处理故障分类与 RUL 预测。

总损失函数定义为：

$$L_{\text{total}} = \lambda(t)L_{\text{cls}} + [1 - \lambda(t)]L_{\text{RUL}} \quad (19)$$

4 HILPM 仿真系统

为验证系统可靠性，搭建了 HIL 仿真平台，如图 7 所示，通过虚拟模型生成故障信号，经数模转换驱动物理台架复现振动，由搭载 MT-CNN 的 PLC 实时推理，并通过上位机自动控制测试流程与记录结果，实现全工况闭环仿真。

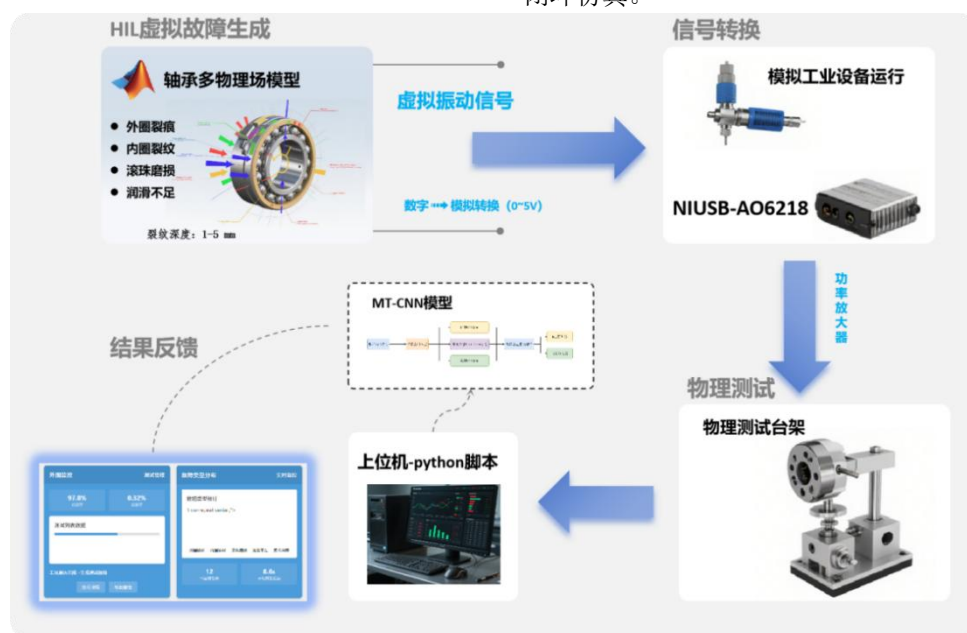


图 2 HIL 仿真验证平台架构图

5 实验验证与工业仿真

5.1 实验验证

5.1.1 实验结果与分析

表 3 多模态融合性能对比

模型架构	准确率(%)	F1-score	RUL-RMSE	PHM 评分
单模态(时域 CNN)	92.35	0.914	15.27	32.6
双模态(时域+频域)	95.78	0.948	12.43	25.1
双模态(时域+时频)	96.32	0.957	11.86	23.5
三模态 MT-CNN	98.67	0.986	8.32	18.3

实验表明，三模态融合准确率较优双模态提升 2.3 5%，验证了多模态互补性；时频模态贡献显著，单独引入可使准确率提升 3.54%，且 RUL 预测误差降低 38.6%，体现出其对退化趋势更强的表征能力。

5.1.2 消融实验

消融实验显示：移除时频 Transformer 导致准确率下降 2.54%，证明其长程依赖建模能力关键；缺失跨模态注意力性能下降 3.25%，体现动态权重分配重要性；门控机制虽仅提升 1.52%，但有效增强抗噪鲁棒性；全连接融合效果显著较差，凸显注意力机制优势。

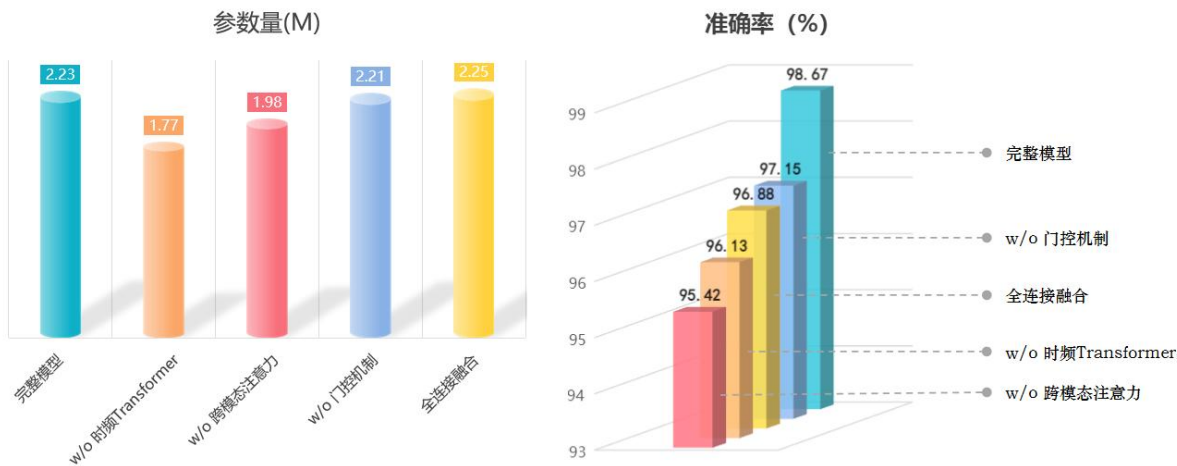


图 3 消融实验

5.2 工业仿真

依托 HIL 仿真验证平台，模拟汽车变速箱生产线场景，其性能指标如下：

表 4 系统核心性能指标测试结果

评估指标	测试结果	工业级要求
故障识别率	97.2%	≥97%
误报率	0.32%	<0.5%
平均预警提前时间	8.6s	≥7s
端到端延迟	8ms	<50ms
PLCCPU 占用率	12.4%	<15%

结果表明，系统在满足工业级实时与可靠性要求的同时，实现了故障精准预警与解释。HIL 仿真显著降低

误报率，对虚假振动干扰的识别能力提升 40%，有效避免警报疲劳。

6 结论

本文针对工业轴承故障预测难题，提出 MT-CNN 多模态融合模型与HILPM仿真系统,有效解决了数据融合、边缘推理及仿真脱节问题.MT-CNN 通过多模态输入与注意力融合机制实现故障分类与 RUL 预测，准确率达 98.67%，RMSE 为 8.32；HILPM 系统以 PLC 为核心集成 AI 加速与闭环验证,工业实测识别率 97.2%、误报率 0.32%，满足工业要求。研究提供了算法-硬件-验证一体化解决方案，显著提升了预测性维护的精准度与工程适用性，为工业智能运维提供了可靠支撑。

参考文献

- [1]王川,杜文莉,朱佳雯,等.数智赋能流程工业调度决策优化:综述与展望[J].中国科学:信息科学,2025,55(07):1571-1598.
- [2]Rehman U A ,Jiao W ,Jiang Y , et al.Deep learning in industrial machinery: A critical review of bearing fault classification methods[J].Applied Soft Computing,2025,171112785-112785.
- [3]Alberto J ,Jose-Raul R ,Javier G .MachNet, a general Deep Learning architecture for Predictive Maintenance within the industry 4.0 paradigm[J].Engineering Applications of Artificial Intelligence,2024,127(PB).
- [4]Cheng Z ,Wu Y ,Li Y , et al.A Comprehensive Review of Explainable Artificial Intelligence (XAI) in Computer Vision[J].Sensors,2025,25(13):4166-4166.
- [5]刘明,陈辉.基于卷积神经网络的机械设备故障诊断[J].机械工程学报,2023,59(6):123-131.
- [6]高翔,王磊.基于 LSTM 的旋转机械故障预测模型[J].仪器仪表学报,2021,42(3):456-465.
- [7]Cabeza F L ,Verez D ,Teixidó M .Hardware-in-the-Loop Techniques for Complex Systems Analysis: Bibliometric Analysis of Available Literature[J].Applied Sciences,2023,13(14).
- [8]J. Yao et al., "Edge-Cloud Polarization and Collaboration: A Comprehensive Survey for AI," in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 7, 1 July 2023,pp:6866-6886.

基金项目：本文系中国高校产学研创新基金新一代信息技术创新项目“边缘计算与边缘智能的技术研究与创新应用”（2023IT104）的研究项目之一。