

学号： 232040191
密级： _____

学校代码： 10336
是否延期公开： 否

杭州电子科技大学

硕士学位论文 (专业学位)



论文题目： 基于质量感知对比学习与人机协同的
科研仪器异常检测方法研究

作者姓名： 王语瑛
指导教师： 陈科明 教授
专业学位类别： 电子信息
专业领域： 新一代电子信息技术
(含量子技术)
所在学院： 电子信息学院
完成时间： 2026 年 5 月

杭州电子科技大学硕士学位论文

基于质量感知对比学习与人机协同的科研仪器
异常检测方法研究

研 究 生： 王语瑛

指导教师： 陈科明 教授

2026 年 5 月

**Dissertation Submitted to Hangzhou Dianzi University
for the Degree of Master**

**Quality-Aware Contrastive Learning and
Human-in-the-Loop Approach for Anomaly
Detection in Scientific Research Instruments**

Candidate: Yuying Wang

Supervisor: Prof. Chen Keming

May, 2026

杭州电子科技大学

学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明： 所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。申请学位论文与资料若有不实之处，本人承担一切相关责任。

论文作者签名：

王语谈

日期： 2026 年 5 月 12 日

学位论文使用授权说明

本人完全了解杭州电子科技大学关于保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属杭州电子科技大学。本人保证毕业离校后，发表论文或使用论文工作成果时署名单位仍然为杭州电子科技大学。学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公布论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存论文。（保密论文在解密后遵守此规定）

论文作者签名：

王语谈

日期： 2026 年 5 月 12 日

指导教师签名：

陈利明

日期： 2026 年 5 月 12 日

摘 要

科研仪器是高校及科研院所进行科学研究的物质基础。随着仪器共享平台的持续建设,如何实现万台级设备的智能运维已成为亟待解决的关键问题。在物联网的远程监测场景中,受分钟级低频采样率和不稳定网络传输的限制,监测数据常伴随细节缺失、随机噪声干扰以及丢包等问题。此外,纯数据驱动的异常检测模型往往难以弥补特征表征与实际异常机理之间的差异,也难以及时适应环境漂移带来的模型性能退化。针对上述问题,本文围绕质量感知对比学习与人机协同机制展开科研仪器异常检测研究。

在工程实现方面,本文搭建了一套面向万台级科研仪器的智能运维监测平台。平台以非侵入式电流感知为基础,围绕大规模设备接入、海量时序数据存储与业务计算解耦需求,构建了感知接入、混合存储、异构计算与多端交互协同的一体化架构。其中,采用 MongoDB 与 MySQL 相结合的混合存储机制分别支撑高并发时序数据与结构化业务数据管理,利用 Spring Boot 业务中台与 Python 算法引擎的异构部署实现业务流转与模型推理分离,并通过跨平台交互设计为海量设备并发监测、算法部署与专家反馈闭环提供了系统支撑。

针对低频采样导致的波形特征模糊与数据质量不均问题,本文提出一种质量感知时频双视图对比学习算法(QA-TF-CL)。该方法首先针对分钟级电流序列构建时域切片与统计频域特征,并利用数据缺失与波动信息生成质量掩码;随后设计时频双编码器对双视图特征进行独立建模,通过时频交互门控融合获得统一高维状态表征;在此基础上,提出时间质量加权对比损失函数(TQ-NCE Loss),并结合重构约束实现鲁棒学习。结果表明,QA-TF-CL 在 multiple 仪器的工况划分任务中表现优异,F1 分数达到 91.74%;即使在数据丢包率为 50%的极端场景下,其性能仍显著优于传统模型,表明该方法在低质数据环境下仍具有较强表征能力。

针对模型难以持续优化及无监督结果缺乏语义解释的问题,本文进一步构建了专家反馈驱动的人机协同演化框架。该框架首先利用 t-SNE 流形学习与密度峰值聚类技术对高维状态特征进行可视化映射,并结合原型传播策略将抽象聚类结果转化为具有物理语义的可解释状态标签,使标注成本降至 0.06%;随后建立基于分位数统计的自适应双阈值机制,实现异常样本的分级筛选与告警抑制;进一步结合相似案例检索、知识库更新以及冻结微调演化策略,将专家确权结果持续反馈至底层模型,实现对工况漂移的快速适应。实验结果表明,系统仅需约 10 次专家确权即可使误警下降率(FARR)提升至 92%以上。

综上,本文的研究覆盖了数据采集、平台支撑、鲁棒表征学习、专家确权及模型演化全过程。该体系显著提升了科研仪器智能监测的可靠性与长效可用性,具有重要理论意义与工程应用价值。

关键词: 科研仪器, 异常检测, 对比学习, 人机协同, 在线演化

ABSTRACT

Scientific instruments are the material foundation for scientific research in universities and research institutes. With the continuous development of instrument-sharing platforms, achieving intelligent operation and maintenance (O&M) for tens of thousands of devices has become a pressing challenge. In remote monitoring scenarios based on the Internet of Things (IoT), monitoring data frequently suffer from missing details, random noise interference, and packet loss due to minute-resolution low-frequency sampling and unstable network transmission. In addition, purely data-driven anomaly detection models often struggle to bridge the gap between learned representations and actual anomaly mechanisms, while also failing to adapt promptly to performance degradation caused by operating condition drift. To address these issues, this thesis investigates anomaly detection methods for scientific instruments based on quality-aware contrastive learning and human-in-the-loop (HITL) collaboration mechanisms.

From the perspective of engineering implementation, an intelligent O&M monitoring platform for tens of thousands of scientific instruments is developed. Based on non-intrusive current sensing, the platform constructs an integrated architecture consisting of perception access, hybrid storage, heterogeneous computing, and multi-terminal interaction to satisfy the requirements of large-scale device access, massive time-series data storage, and the decoupling of business logic from computation. Specifically, a hybrid storage mechanism combining MongoDB and MySQL is adopted to support high-concurrency time-series data and structured business data management, respectively. Meanwhile, heterogeneous deployment of the Spring Boot business middleware and Python-based algorithm engine is employed to separate business workflows from model inference, thereby providing system-level support for large-scale concurrent monitoring, algorithm deployment, and expert-feedback closed loops through cross-platform interaction design.

To address waveform feature ambiguity and data quality inconsistency caused by low-frequency sampling, this thesis proposes a Quality-Aware Time-Frequency Dual-View Contrastive Learning algorithm (QA-TF-CL). First, temporal slices and statistical frequency-domain features are constructed from minute-level current sequences, and quality masks are generated according to data missingness and fluctuation information. Then, a dual-encoder architecture is designed to independently model temporal and frequency-domain features, while a time-frequency interactive gating mechanism is introduced to integrate them into unified high-dimensional state representations. On this basis, a Time-Quality weighted contrastive loss function (TQ-NCE Loss) is proposed and combined with reconstruction constraints to achieve robust representation learning. Experimental

results demonstrate that QA-TF-CL achieves excellent performance in operating condition classification tasks across multiple types of instruments, with an F1-score of 91.74%. Even under extreme scenarios with a packet loss rate of 50%, the proposed method still significantly outperforms conventional models, indicating strong representation capability in low-quality data environments.

To address the difficulty of continuous model optimization and the lack of semantic interpretability in unsupervised results, a human-in-the-loop collaborative evolution framework driven by expert feedback is further constructed. The framework first employs t-SNE manifold learning and density peak clustering to visually map high-dimensional state features, and combines prototype propagation strategies to transform abstract clustering results into interpretable state labels with physical semantics, reducing the labeling cost to merely 0.06%. Subsequently, an adaptive dual-threshold mechanism based on quantile statistics is established to achieve hierarchical anomaly screening and alarm suppression. Furthermore, by integrating similar-case retrieval, knowledge base updating, and freeze-and-finetune evolution strategies, expert confirmation results are continuously fed back to the underlying model, enabling rapid adaptation to operating condition drift. Experimental results show that only about ten rounds of expert validation are required to improve the False Alarm Reduction Rate (FARR) to over 92%.

In summary, this thesis covers the complete process of data acquisition, platform support, robust representation learning, expert label verification, and model evolution. The proposed framework significantly improves the reliability and long-term usability of intelligent monitoring for scientific instruments, demonstrating important theoretical significance and engineering application value.

Keywords: Scientific Instruments, Anomaly Detection, Contrastive Learning, Human-in-the-loop, Online Evolution

目 录

摘 要.....	I
ABSTRACT.....	II
目 录.....	IV
1 绪论.....	1
1.1 课题来源.....	1
1.2 研究背景与意义.....	1
1.3 国内外研究现状.....	3
1.3.1 状态监测与故障诊断研究现状.....	3
1.3.2 无监督时序异常检测研究现状.....	5
1.3.3 人机协同研究现状.....	6
1.4 论文内容与结构安排.....	8
2 科学仪器智能运维监测平台.....	10
2.1 平台需求分析与总体架构.....	10
2.1.1 科学仪器运维需求分析.....	10
2.1.2 平台总体架构.....	12
2.2 边缘感知与数据采集.....	13
2.2.1 智能监测终端硬件设计.....	13
2.2.2 边缘侧数据聚合与传输策略.....	18
2.3 数据混合存储设计与实现.....	20
2.3.1 数据特征分类与存储选型.....	20
2.3.2 混合持久层实现.....	21
2.4 业务逻辑与算法引擎的集成.....	25
2.4.1 后端业务系统构建.....	26
2.4.2 算法引擎系统构建.....	27
2.4.3 跨语言通信与异步解耦机制.....	28

2.5 交互层设计与实现	29
2.5.1 网页端实现.....	29
2.5.2 微信小程序端实现.....	36
2.6 本章小结	37
3 基于质量感知对比学习的仪器状态辨识与异常检测方法.....	39
3.1 问题描述与总体架构	39
3.1.1 问题描述.....	39
3.1.2 总体架构.....	39
3.2 时频特征构建与质量感知建模	40
3.2.1 统计频域特征构建.....	41
3.2.2 数据质量掩码生成.....	41
3.2.3 多视图数据的标准化与输入实例构建.....	42
3.3 质量感知的双视图对比学习模型	44
3.3.1 双流编码器设计.....	44
3.3.2 时频交互门控融合.....	45
3.3.3 时质联合加权对比损失函数.....	46
3.3.4 联合优化目标.....	48
3.3.5 模型训练与实现细节.....	49
3.4 仪器状态聚类与异常判别	51
3.4.1 基于特征空间的无监督聚类.....	51
3.4.2 基于距离重构的异常评分机制.....	52
3.5 实验验证与分析	53
3.5.1 实验设置.....	53
3.5.2 预测结果对比分析.....	57
3.5.3 核心模块的消融实验.....	60
3.5.4 数据质量敏感性与鲁棒性分析.....	61
3.6 本章小结	62
4 基于人机协同的状态语义映射与知识演化机制.....	64

4.1 传统运维瓶颈与人机协同启示	64
4.1.1 传统科研仪器运维模式的局限性	64
4.1.2 跨领域人机协同的经验启示	64
4.1.3 人机协同闭环监测框架设计	65
4.2 无监督状态流形的语义重构与可视化	66
4.2.1 状态流形降维与拓扑可视化	67
4.2.2 基于原型的批量语义传播策略	68
4.3 专家反馈驱动异常确认机制	68
4.3.1 基于统计分位数的自适应双重阈值	69
4.3.2 相似案例检索机制	69
4.3.3 上下文增强的交互式反馈接口	71
4.4 动态故障知识库的构建与维护	71
4.4.1 故障知识单元	71
4.4.2 知识库自适应清洗	72
4.5 基于难例挖掘的对比微调策略	72
4.5.1 难例挖掘与有监督对比损失构建	73
4.5.2 冻结微调策略	74
4.6 实验验证与分析	74
4.6.1 语义标注效率实验	74
4.6.2 双重阈值参数选择实验	76
4.6.3 模型在线演化与误报抑制实验	77
4.7 本章小结	79
5 总结与展望	81
5.1 总结	81
5.2 未来展望	81
致谢	83
参考文献	84
附 录	91

1 绪论

1.1 课题来源

本课题依托于浙江省科技厅主导的“浙江省大型科研仪器设备管理服务协作平台”重点建设项目。该平台旨在响应国家关于深化科研设施与仪器开放共享的战略部署，通过构建覆盖全省的物联网感知网络，解决科研仪器资源利用率低、状态监管困难的行业痛点。

目前，该平台已成功接入超过万台大型科学仪器，积累了海量的分钟级电流时序数据与运行日志。本文的研究工作是该平台建设的重要组成部分，旨在挖掘平台海量低频监测数据背后的物理规律，为全省科研仪器的自动化考评、异常监测及共享服务提供精准的算法支撑与决策依据。

1.2 研究背景与意义

当前，全球科技竞争日益激烈，高水平科技自立自强已成为国家发展的重要驱动力。党的二十大报告明确提出，必须坚持“科技是第一生产力、人才是第一资源、创新是第一动力”，并强调要深入实施创新驱动发展战略^[1]。在这一宏观背景下，高校作为原始创新的重要阵地，其科研组织模式正由传统的单兵作战和零散化探索逐步转向建制化、跨学科、大协同的集群攻关模式。教育部印发的《关于加强高校有组织科研 推动高水平自立自强的若干意见》明确要求高校充分发挥新型举国体制优势，通过打破学科壁垒、优化资源配置和强化力量集成，提升支撑国家战略需求的能力^[2]。科研仪器作为科学研究的重要基础，其高效、稳定的运转是支撑这一战略转型的核心保障。

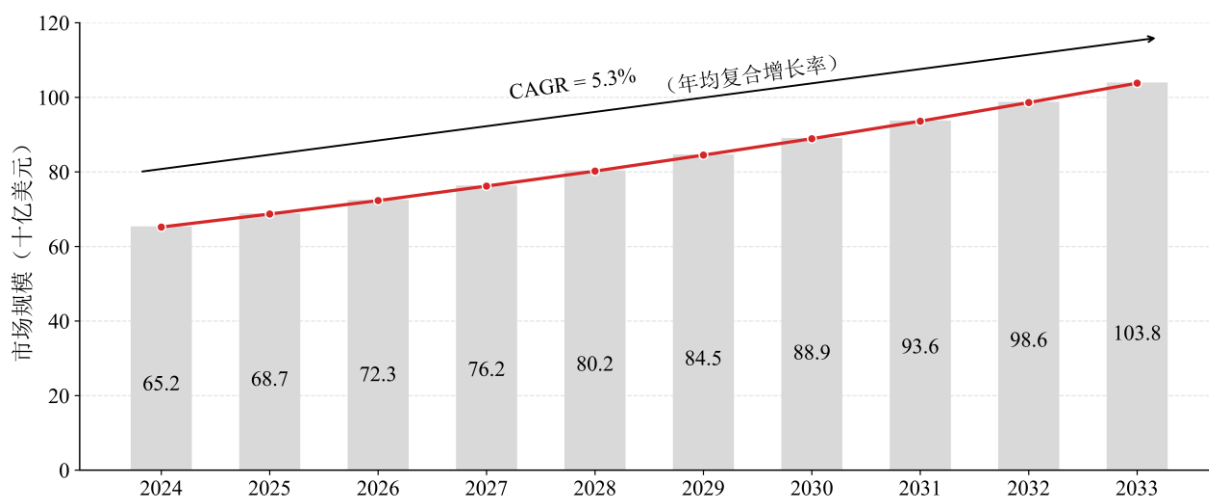


图 1.1 2024-2033 年全球科研仪器市场规模及预测趋势图

从全球范围来看，科研仪器市场正处于稳健增长阶段。据国际权威咨询机构 IMARC Group 发布的最新数据显示，2024 年全球科研仪器市场规模已达到 652 亿美元，约合人民币 4700 亿

元。随着生物医药、新材料及半导体等高新技术领域研发活动的日益频繁,预计到 2033 年,全球科研仪器市场规模将增长至 1038 亿美元左右,年均复合增长率保持在 4.48%至 5.3%之间[3][4]。

中国作为全球增长最快的二级市场,正经历从规模扩张向高质量发展转变的关键阶段。2024 年中国仪器市场规模约 1500 亿元,同比增长 18%。受国产替代政策驱动,预计 2025 年,中国科研仪器市场规模将达到 1700 亿元。随着企业研发投入的增加以及高端制造、医疗和环境监测等领域检测需求的持续增长,科学仪器的应用场景正由高校实验室不断向工业产线、医疗和环境监测等大规模商用领域渗透。这一变化将进一步拓展科学仪器产业的应用范围和市场空间,预计到 2029 年,包含软硬件服务在内的我国科学仪器产业整体规模有望突破 1.3 万亿元^[5]。

根据前瞻产业研究院发布的《2024 年中国科学仪器行业全景图谱》统计,仅 2023 年我国实验分析仪器的主营业务收入已超 380 亿元,电子测量仪器营收突破 460 亿元。中国与美国合计占据了全球超过 50%的科研仪器新增装机量,仅 2023 年亚太地区新增的 520 万台仪器中,中国市场就贡献了约 240 万台。如此庞大的资产规模意味着科研仪器的运维已不再是实验室中的辅助工作,而是直接关系到数千亿国有资产保值增值与运行效率的核心经济环节。面对如此庞大的增量与存量,传统的粗放式管理模式已难以适应现实需求,亟需向数字化、智能化的精细化运维转型。尽管硬件规模呈现爆发式增长,但在实际运行中,重购置、轻运维的结构性矛盾依然突出。从管理学角度看,大型科研仪器具有极强的资产专用性,因此对运行环境、操作规范和专业维护具有高度依赖性。于若溪^[6]等人的研究指出,由于核心技术高度依赖进口,我国高校在仪器运维层面长期存在被动受制的问题。一方面,国外供应商在维修服务中设置的技术壁垒导致维护成本上升,并延长了备件供应周期;另一方面,校内专业实验技术队伍建设滞后,缺乏能开展系统维护的复合型人才^[7]。

此外,随着跨校区办学及区域共享平台的推进,仪器运行负荷持续增加,设备老化、性能漂移以及环境耦合干扰等问题日益凸显^{[8][9]}。传统的基于人工经验的阈值报警和定期巡检模式,在处理海量、非平稳的实时监测数据时,存在响应滞后和误报率较高的问题。这种维护能力滞后与高频、高负荷需求之间的矛盾,已成为制约科研设施可持续运行的核心痛点。因此,亟需引入人工智能等先进技术,通过数字化手段融合专家经验,构建具备自适应和自演化能力的智能运维体系。

综上所述,本文聚焦于科研仪器智能运维的关键技术与系统实现。在理论层面,针对监测数据普遍存在的低采样率、弱标签和强噪声等特点,本文提出了基于质量感知的无监督表征学习方法。通过构建时频双视图对比学习架构与质量掩码机制,增强了模型对数据缺失与噪声干扰的鲁棒性,为复杂工业场景下的异常检测提供了新的研究视角。在应用层面,通过构建人机协同的闭环监测系统,有效利用专家反馈修正模型偏差,解决了无监督算法的语义缺失问题。利用专家确权标签实现模型在线演化,能够显著降低误报率,缓解报警疲劳。这

对于提升大型科研仪器的稳定运行能力、保障科研资产安全具有重要的工程实践意义。

1.3 国内外研究现状

1.3.1 状态监测与故障诊断研究现状

随着大型科学装置向高能级、高精密与长周期运行方向发展，其状态监测与故障诊断技术也逐步由早期的事后维修、基于经验的预防性维护，发展到如今数据驱动的预测性维护。这一过程体现了运维模式由依赖人工经验转向以数据智能驱动的自主决策^[10]。

(1) 基于机理模型与经验阈值的早期探索

在科研仪器运维的早期阶段，故障诊断主要依赖阈值监控与数据采集系统。此类方法通过设定上下限阈值监测电流、电压或真空度等关键参数，当参数超出阈值范围时即触发报警。随着对系统动态特性认识的不断加深，基于物理机理的模型诊断方法逐渐成为主流。该方法通过建立仪器组件的数学模型，如微分方程或状态空间模型，结合观测器或卡尔曼滤波生成残差信号，并通过分析残差的统计特性判断系统是否偏离正常状态^[11]。

然而，机理模型在大型科研场景中面临严峻挑战。首先，大型科学装置通常由成百上千个子系统耦合而成，其运行过程中伴随着复杂的非线性动态特征和热漂移现象，因而无法建立精确的全局解析模型^[12]。其次，物理模型对系统参数的准确性高度敏感，随着仪器老化，系统参数会发生变化，进而导致静态模型误报率显著上升，难以适应动态变化的实验环境。

(2) 统计特征与浅层机器学习的过渡

为了克服物理建模在处理复杂系统时的局限性，研究者开始转向由数据驱动的统计学习方法。主成分分析(Principal Component Analysis, PCA)等方法被引入以处理变量间的相关性。例如，Qin^[13]等人将高维监测数据投影到低维空间，利用 T^2 统计量和SPE统计量实时识别系统状态偏离，从而实现对系统运行趋势的初步监控。张成^[14]等提出了基于动态主元分析残差互异度的故障检测与诊断方法，通过改进残差特征的统计表征方式，提升了动态过程中的故障检测性能。在此基础上，研究者进一步引入浅层机器学习模型，旨在解决故障分类难题。支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)以及人工神经网络(Artificial Neural Network, ANN)等算法被广泛应用于该任务中^[15]。相比于单纯依赖统计指标的方法，这些算法能够借助非线性核函数或决策树结构，建立高维特征与具体异常模式之间的映射关系。此类方法标志着运维技术开始进入数据驱动阶段。

然而，这类方法的局限性同样较为明显。虽然这类方法在一定程度上降低了对精确机理建模的依赖，但模型性能仍在较大程度上取决于人工特征工程。运维人员通常需要结合领域知识，从小波能量谱、时域统计量和频域指标中手动构造并筛选特征。对于科研仪器海量、高维且非平稳的监测数据而言，这种特征构建方式不仅成本较高，而且难以有效表征复杂的时序关系和深层异常信息，从而限制了模型的泛化能力。

(3) 深度学习驱动的特征学习演进

随着深度学习技术的发展，研究者开始探索利用深度神经网络自动学习时序数据的高层特征表示，从而减少对人工特征工程的依赖。相比传统方法，深度学习能够通过多层非线性映射从原始数据中提取具有判别力的特征，在复杂工业数据分析中展现出显著优势^[16]。卷积神经网络（Convolutional Neural Network, CNN）是最早应用于工业信号分析的深度学习模型之一。CNN 通过局部感受野和参数共享机制，可以有效捕捉时间序列中的局部特征。Janssens^[17]等人提出将一维振动信号转换为二维特征图输入 CNN 进行故障识别，该方法能够有效提取局部异常模式，并显著提升故障分类精度。雷亚国^[18]等提出了基于深度学习的机械装备大数据健康监测方法，指出深度网络能够在大规模监测数据中实现从原始信号到故障表征的自动学习。吴定海^[19]等进一步对基于卷积神经网络的机械故障诊断方法进行了综述，系统总结了 CNN 在数据输入形式、网络结构设计和迁移学习等方面的应用进展。

尽管 CNN 擅长捕捉局部模式，但对科研仪器运行过程中的长期演化规律如设备退化和周期性趋势的建模能力有限。为此，具备记忆能力的循环神经网络（Recurrent Neural Network, RNN）被引入该领域。长短期记忆网络（Long Short-Term Memory, LSTM）通过门控机制，克服了传统 RNN 在长序列建模中易出现梯度消失的问题^[20]。在设备健康管理场景中，LSTM 能够挖掘电流或振动序列中的时序依赖关系，为趋势预测和状态切换判定提供动态上下文。

近年来，Transformer 架构的出现打破了 RNN 必须按序处理数据的限制。其引入的自注意力机制（Self-Attention）使模型能够直接计算序列中任意两个采样点之间的关联程度，从而实现全局建模^[21]。相比于 LSTM，Transformer 在处理科研监测场景中的长时段、多尺度数据时，不仅计算效率更高，还能更有效地捕捉跨越长周期的潜在相关特征。

为了更直观地展示上述状态监测技术的演进脉络及其在工业场景中的适用性，表 1.1 对不同诊断范式进行了系统性对比。

表 1.1 状态监测与故障诊断技术演进路线对比

技术发展阶段	主要技术特征	局限与挑战
传统机理与阈值	物理建模+固定阈值	难以精确建模，对设备老化极度敏感
浅层机器学习	人工特征提取+统计分类	依赖人工特征提取，泛化能力受限
深度学习	端到端的自动特征提取	强依赖高频采样数据与海量故障标签
本文方法	质量感知的时频对比学习	零标签依赖，自主挖掘高阶物理特征

综上所述，状态监测与故障诊断技术已由机理建模逐步发展到数据驱动特征学习。尽管以 CNN、LSTM 和 Transformer 为代表的深度模型在复杂时序建模方面展现出较强的能力，但在科研仪器监测场景中仍面临适配性不足的问题。在实际工况中，标签数据通常极度稀缺，监测信号的采集密度也远低于实验室环境。相较于依赖复杂模型结构或大规模标注数据的研究模式，本文更加关注方法与应用场景之间的适配性。针对科研仪器监测场景，本文直接从反映仪器物理特性的低频电流信号中提取深度表征，摆脱了对故障先验标签的依赖。本文采用基于无监督对比学习的表征学习方法，旨在解决科研仪器监测中数据充足但标注缺乏的真

实问题，为复杂环境下的异常识别提供一种更务实的解决方案。

1.3.2 无监督时序异常检测研究现状

在实际工业监测与设备运维场景中，由于异常样本往往稀缺甚至难以获取，传统监督学习方法难以直接应用。因此，无监督异常检测逐渐成为时间序列分析领域的重要研究方向。该类方法通过挖掘数据的内在结构特征，能够在缺乏人工标注的情况下实现异常识别，并在工业监测、设备运维及物联网系统中具有广泛的应用前景。

早期研究主要基于重构误差的生成式建模思想。该类方法假设模型在正常样本上训练后能够较好地重构正常数据，而对异常样本的重构能力较弱，因此可以根据重构误差的大小实现异常检测。自动编码器是最典型的重构类模型之一。Malhotra^[22]等人提出基于 LSTM 的自编码器结构，通过 LSTM 捕捉时间序列中的长期依赖关系，并根据重构误差分布实现异常判定。为了进一步增强模型对复杂数据分布的建模能力，研究者引入概率生成模型。例如，Su^[23]等人提出的 Omni Anomaly 框架利用变分自动编码器对多维时间序列进行概率建模，并通过引入随机隐变量刻画系统状态的不确定性，从而提升模型在多变量场景下的鲁棒性。Li^[24]等人提出的 MAD-GAN 通过生成器与判别器之间的对抗训练学习复杂的时间依赖关系，并结合生成误差与判别误差构建综合异常评分机制。然而，随着应用场景拓展至高噪声、强非平稳环境，生成式模型极易因过度拟合能力而同时重构正常与部分异常样本，降低了检测区分度^[25]。

为克服上述瓶颈，研究者开始转向以判别式表征学习为核心的对比学习范式。该方法通过构造正负样本对，在特征空间中缩小相似样本之间的距离，增大不相似样本之间的距离，从而学习具有判别力的高维表示。Chen^[26]等人提出的 SimCLR 框架与 He^[27]等人提出的 MoCo 框架验证了对比学习在特征表征能力方面的有效性。随着相关研究的发展，对比学习逐渐被引入时序建模领域。例如，Eldele^[28]等人提出的 TS-TCC 模型通过构建不同强度的数据增强视图，并利用跨视图预测任务学习时序上下文表征。进一步研究表明，仅依赖时间域信号构建对比视图，在周期性复杂或采样频率受限的场景中，可能难以充分刻画频谱层面的结构信息。为此，多视图学习策略逐渐受到关注，其中时频双视图融合成为重要发展方向。Zhang^[29]等人提出的 TF-C 框架认为，样本的时间域表示与频域表示在特征空间中应保持一致性。该方法利用傅里叶变换提取频域特征，并通过最大化时域嵌入与频域嵌入之间的互信息，实现对趋势变化与能量分布特征的联合建模，提升了特征表征能力。

为了直观展示无监督表征方法在对比学习、多视图融合和质量感知等方面的差异，表 1.2 对该领域的代表性方法进行了特性对比。结合表 1.2 可以看出，时序异常检测技术已由早期的重构生成式模型逐步发展到具备多视图建模能力的对比学习方法，但现有方法仍难以直接应用于大型科研仪器的实际运维场景。受限于存储资源与通信带宽，监测数据往往存在低采样率、非均匀采样及插值补全等情况。上述因素可能引入额外噪声或结构失真，从而影响特征空间的稳定性。现有对比学习损失函数，如 InfoNCE，通常对不同样本对采用相同权重，

缺乏针对数据质量差异的自适应调节机制。在复杂采样条件下，如何在保持判别能力的同时提升模型对低质量数据的鲁棒性，仍是时序异常检测领域有待进一步研究的重要方向^[30]。

表 1.2 无监督时序表征方法的关键特性对比

代表性方法	核心建模机制	是否对比学习	是否多视图融合	是否质量感知
Omni Anomaly	概率生成与对抗重构	×	×	×
SimCLR	单时域对比学习	✓	×	×
TF-C	时频特征一致性对齐	✓	✓	×
本文方法	时频质量联合加权的多视图对比学习	✓	✓	✓

针对上述问题，本文在时频双视图融合的对比学习框架基础上，提出了一种基于质量感知的自适应加权对比学习方法。不同于传统 InfoNCE 等损失函数对样本对赋予相同权重的处理方式，本文引入了面向数据质量差异的自适应加权机制。该机制能够动态评估不同样本对的数据质量，并抑制低质量样本对特征空间学习的不利影响，从而在复杂采样条件下提取更加稳健的时序表征，提升无监督异常检测模型的鲁棒性。

1.3.3 人机协同研究现状

除了算法层面的特征表征与鲁棒性挑战外，实际工程落地还面临模型输出与专家认知之间难以有效衔接的问题。随着智能运维系统逐渐由自动检测工具发展为辅助决策平台，如何有效融合数据驱动的分析能力与领域专家经验，已成为当前智能监测研究的重要方向。近年来，人机协同（Human-in-the-Loop, HITL）范式逐渐受到广泛关注，其核心思想是在模型决策过程中引入专家交互反馈，使算法系统能够不断融入领域知识并逐步实现自适应优化^[31]。

在实际的科研仪器运维场景中，单纯依靠数据驱动模型往往面临明显的语义解释不足问题。现有无监督学习模型的输出通常表现为抽象的聚类标签，如 Cluster 0、Cluster 1，或归一化异常评分，缺乏与实际物理工况相对应的语义解释^[32]。Holzinger^[33]等人在医疗诊断与复杂系统监测研究中指出，在高风险决策领域，仅依赖黑盒模型输出难以取得工程人员的信任，因此需要借助交互机制将专家经验融入模型决策过程，从而提升系统的可解释性与可信度。

在 HITL 框架下，如何高效利用专家标注资源成为关键问题。为此，主动学习与标签传播技术被广泛引入。主动学习通过不确定性采样等策略，自动筛选信息量较高的样本交由专家审核，从而有效降低人工标注成本^[34]。Wu^[35]等人提出了一种基于聚类假设的查询策略，结果表明，仅需标注少于 5% 的关键样本，即可训练出接近全监督精度的诊断模型，从而提升专家标注的利用效率。同时，Iscen^[36]等人提出的图传播模型基于数据流形假设，将少量专家确认的种子标签在高密度特征空间中传播至整个数据簇，实现了从点到面的知识泛化。这一机制在系统冷启动阶段具有重要的工程应用价值。

自 2023 年以来，随着大语言模型（Large Language Models, LLMs）与工业智能代理（Agents）

的快速发展，智能运维的人机协作模式也呈现出新的发展趋势。多模态大模型的引入提升了故障诊断交互的智能化水平，但在处理科研仪器底层的电流波形数据时，其通用建模能力并不一定能够适应具体场景需求。通用大模型在缺乏领域物理知识支撑的情况下，往往难以准确解读细微的波形畸变，容易给出偏离实际工况的判断。相关实践证明，大模型难以替代领域专家基于长期实践所形成的物理判断能力。因此，在实际工程应用中，仍需依托高保真的底层模型提取关键特征，并结合专家确权机制完成结果校验与语义赋予。这种人机协同的闭环机制为兼顾高精度与可解释性提供了可行路径。

此外，现有协同模式大多建立在静态分布假设之上，难以充分考虑长期运行过程中数据的非平稳特征。设备性能的自然退化或外部环境的周期性波动，都会导致监测数据产生概念漂移^[37]。传统的离线训练、在线部署模式在出现基线漂移时，其判别边界会逐渐失效，进而导致系统误报率显著上升。针对此问题，Gama^[38]等人指出需引入增量学习策略以缓解模型老化。但现有方案往往停留在算法层面的自动更新，缺乏系统性的知识管理机制。这种局限性在专家介入时尤为明显。由于缺乏有效融合，专家反馈往往仅被当作零散的修正信息。如果模型在学习这些新特征时处理不当，不仅无法实现持续优化，反而会陷入灾难性遗忘的困境^[39]。

为直观展示从传统运维到智能化人机交互的演进脉络，表 1.3 系统对比了不同运维方式的核心特性。

表 1.3 人机协同与智能运维交互范式对比

运维方式	交互方式	领域知识利用模式	概念漂移适应能力
传统静态部署	被动接收告警并线下处理	经验纯隐性，缺乏反馈	无（模型固化）
主动学习与标签传播	少量标注实现局部知识扩散	侧重系统冷启动阶段的训练	弱（缺乏在线纠偏）
大模型辅助运维	自然语言交互与日志问答	依赖预训练通识	弱（缺乏物理机理约束易产生偏差）
本文方法	基于上下文界面的语义确权	专家反馈的结构化知识更新 与模型微调	强（难例挖掘与对比微调）

综上所述，人机协同研究已经从最初的简单人工校验逐渐发展为融合主动学习、标签传播与持续学习等机制的协同框架。但在复杂科研仪器监测场景中，如何进一步实现专家知识与数据驱动模型之间的深度融合，并构建具备持续演化能力的运维知识体系，仍是有待突破的关键问题。

为此，本文针对离线训练与在线静态部署的传统模式，构建一种基于人机协同的动态知识演化与闭环诊断机制。本文工作的切入点在于：通过语义映射与标签传播策略，将模型输出的抽象高维空间距离转化为可被专家理解的物理语义，并通过自适应双重阈值与难例挖掘的冻结微调策略，将专家的离散反馈转化为驱动对比学习模型在线演化的动能，从而赋予系统持续对抗工况漂移与长期自适应生长的能力。

1.4 论文内容与结构安排

本研究立足于大型科研仪器在物联网运维中的实际需求，重点针对低频强噪环境下特征提取困难与静态模型难以适应动态环境这两类关键问题，构建了一套闭环监测体系。在硬件层面，利用非侵入式感知手段实现了仪器底层电流时序信号的合规、高效采集；算法层面，针对低频监测数据中的信息失真问题，提出了 QA-TF-CL 算法，通过时域与频域的双视图对比，提升了无监督状态辨识的精准度；应用层面，本研究不再单纯依赖静态模型，而是通过人机协同架构引入专家反馈，利用增量学习技术驱动模型参数持续更新，从而解决了智能算法在实际工程落地中的可解释性与性能衰减问题。

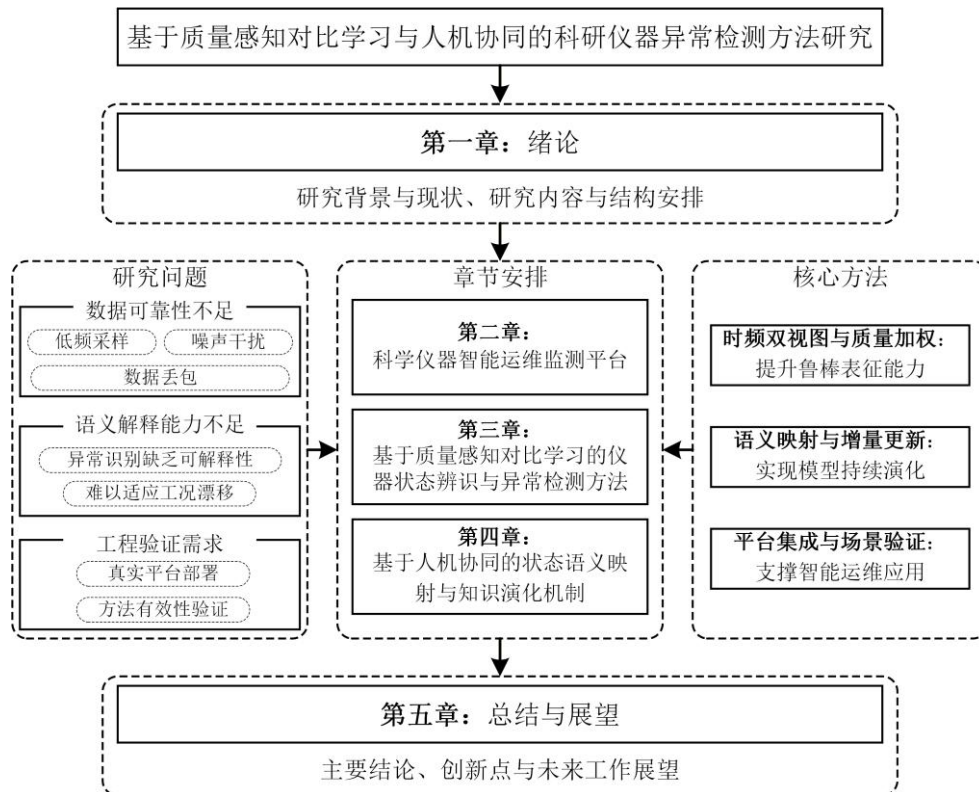


图 1.2 全文结构图

全文共分为五章，结构如图 1.2 所示，具体内容与结构安排如下：

第 1 章：绪论。阐述本课题的来源、研究背景及其在国家科研资产保值增值和高水平科技自立自强中的重要意义。系统梳理状态监测、无监督时序异常检测以及人机协同领域的国内外研究现状，并由此引出本文在低采样率和弱标签场景下的研究切入点。

第 2 章：科学仪器智能运维监测平台。作为本文研究的系统支撑平台，本章详细论述其总体设计需求与实现架构。针对科研仪器禁止侵入式改装的运维要求，设计了基于非侵入式感知的监测终端，并详细论述其硬件选型与工作机理。介绍了适用于 NB-IoT 窄带通信环境的低频数据传输策略，以及支撑千万级时序数据高效存储的异构架构，为后续模型研究奠定真实可靠的数据基础。

第 3 章：基于质量感知对比学习的仪器状态辨识与异常检测方法。本章针对分钟级低频

采样导致的信息缺失与插值噪声问题，提出了 QA-TF-CL 算法框架。该方法构建了包含波动能量与趋势强度的统计频域视图，并引入数据质量掩码来动态量化数据的可靠性；在此基础上，设计了时频双流编码器与门控融合机制，并提出时间质量联合加权对比损失函数（TQ-NCE），从而在高噪声环境下实现特征解耦与无监督状态聚类。

第 4 章：基于人机协同的状态语义映射与知识演化机制。为弥补无监督算法与物理工况之间的语义断层，本章探索了专家知识参与下的反馈演化机制。首先，利用 t-SNE 降维与密度峰值聚类（Density Peak Clustering, DPC）构建可视化状态地图，并通过原型批量传播策略实现低成本的语义标注。其次，建立了基于统计分位数的筛选机制，以缓解报警疲劳并提升预警信噪比。最后，提出基于难例挖掘的增量微调策略，使底层模型能够根据专家反馈进行持续优化，并推动系统由单一算法输出向结构化故障知识库演化。

第 5 章：总结与展望。回顾并总结本文在算法架构设计与系统工程实现方面的成果，并对未来科研仪器全生命周期管理的数字化演进方向进行展望。

2 科学仪器智能运维监测平台

本章主要论述科学仪器智能运维监测平台的设计与实现过程。该平台是支撑整个研究的硬件基础和软件支撑。作为后续算法模型与运行机制的支撑平台，本平台不仅需要解决多源异构数据的采集与存储问题，还直接决定了研究数据的质量与系统的交互方式。针对本文提出的无监督状态辨识方法，平台需提供稳定的低频时序数据流；针对人机协同演化机制，平台需提供高效的专家反馈通道。因此，本章将从需求分析、硬件感知、混合存储、异构架构以及交互设计五个方面，详细论述平台的构建逻辑与技术实现。

2.1 平台需求分析与总体架构

2.1.1 科学仪器运维需求分析

在构建监测平台前，需明确科研仪器运维场景与传统工业物联网的本质差异。这些差异不仅影响平台的架构设计边界，也直接关系到后续算法研究的输入特征与性能表现。

(1) 机理封闭性与状态识别

不同于工业生产线中通过 PLC 或标准协议直接读取状态码的白盒模式，大型科研仪器如透射电镜、核磁共振仪等，通常具有较强的结构封闭性。受限厂商的技术限制与协议壁垒，外部监测者通常无法直接获取核心控制指令。因此，本平台须采用侧信道分析方法，即利用外接式传感器采集电源线电流这一外部物理量，并通过其时序波形推断设备内部运行工况^[40]。然而，科研仪器的状态切换并非简单的二元切换过程，而是包含了预热、抽真空、校准、扫描和清洗等多个高度耦合的中间状态。图 2.1 展示了典型工况下的电流变化特征。由于不同工况下的电流特征存在重叠，传统固定阈值判定方法容易产生误报，因此要求平台算法具备较强的特征分析与解耦能力。

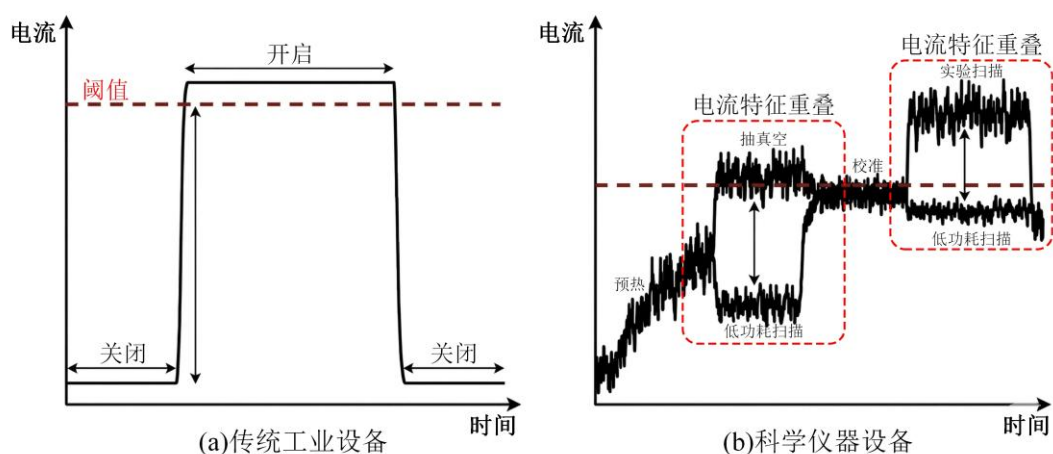


图 2.1 电流波动对比图

(2) 物理接入受限与感知数据的低频稀疏性

在实际部署环境中，科研仪器对供电稳定性和信号纯净度要求较高，严禁任何形式的侵入式线路改装。这一零干扰底线决定了感知端需采用非侵入式互感器。同时，考虑到实验室内部复杂的电磁环境以及全省范围内万级设备的规模化接入需求，本平台在通信链路设计上选择了具备深穿透、低功耗特性的 NB-IoT 技术。受限于 NB-IoT 的窄带特性，平台无法承载高频原始波形的实时回传。因此，监测终端采用边缘聚合、云端映射的处理策略，即在端侧完成瞬态特征的平滑与初步筛选，仅向上层平台回传分钟级的聚合数据。图 2.2 展示了采样阶段存在的特征丢失现象，这种时域稀疏性意味着后续研究需要在有限观测信息下尽可能提取有效的状态特征。

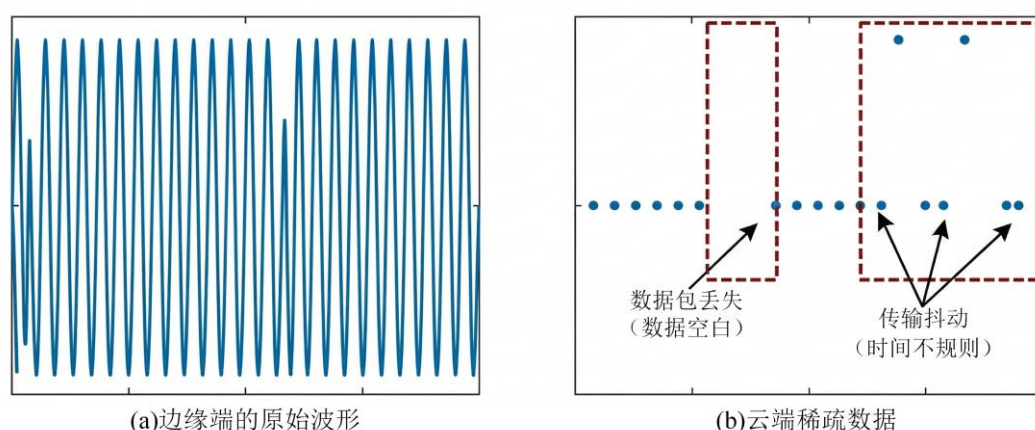


图 2.2 边缘原始波形与云端稀疏数据对比示意图

(3) 数据质量的非理想性

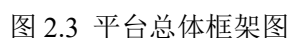
实际部署环境中的 NB-IoT 链路并非理想信道。实测数据及相关仿真研究^[41]表明，在地下室等弱信号覆盖区域以及大规模并发接入场景下，无线资源竞争与重传机制容易引发数据包丢失。此外，由于协议层重传机制具有随机性，平台侧接收到的数据流在时间轴上表现出明显的到达时间抖动和非等间隔特性。这种由物理环境与通信协议共同作用产生的数据质量缺陷，构成了本平台必须面对的原始数据条件，也为第三章中数据质量感知机制的研究提供了现实依据。

(4) 业务负载的非均衡性与交互式功能需求

科研仪器的负载特征具有明显的波动性，设备在大多数时间内处于低功耗待机状态，而在特定科研任务开展期间则表现出高频、高负荷的突发式运行。这种非均衡的负载模式要求平台架构同时满足海量历史数据存储需求和突发工况下的高强度计算需求。在架构设计上，本系统摒弃了传统的单体架构，转而采用了 Spring Boot 业务平台和 Python 算法平台解耦的异构模式，以实现业务逻辑与计算密集型任务之间的资源隔离与动态调度。

此外，科学仪器的异常判定具有较高的专业壁垒，尤其是性能漂移等轻微故障往往依赖资深人员的领域经验^{[42][43]}。传统的物联网平台多采用单向的监控模式，难以处理算法误报或判断模糊的情况。因此，平台功能设计必须具备双向交互能力，即支持专家反馈接口。这不

基于科学仪器运维的特点与实际工程约束，本平台采用分层解耦的架构，构建了从边缘感知到云端决策的完整技术闭环。



(1) 感知接入层。系统底层通过自主研发的非侵入式监测终端提取电流特征, 通过 NB-IoT 链路接入电信运营商的 AEP 平台。

(3) 核心业务逻辑层。平台将常规业务处理模块与算法计算模块解耦，通过部署异构服务实现系统架构的分离。Spring Boot 中台负责常规业务流与流程管理；Python 算法中心负责模型的在线推理。两者基于底层核心监控表的异步轮询实现资源隔离，确保重计算任务不阻

塞系统响应。

(4) 网络层。通过 API Gateway 统一封装后端接口，集成鉴权、协议转换与流量削峰功能。网关根据请求类型将流量分发至业务中台或算法中心，确保系统跨域访问的安全性与可扩展性。

(5) 应用交互层。面向不同角色提供多端服务与可视化界面。

2.2 边缘感知与数据采集

本节介绍了基于 NB-IoT 的非侵入式电流监测终端。该终端通过非接触方式捕获仪器的动态电流，并将其转化为数字时序特征上传至平台。具体外观如图 2.4 所示，尾部采用插拔式转接架构。设备串接于市电插座与仪器电源线之间，内部集成电流互感器，利用电磁感应原理采集回路电流。通过侧信道采集方式避免对仪器原有电路的改动，确保监测过程对精密仪器供电质量的零干扰。对于被测设备而言，监测终端的接入仅相当于无损延长的电源线，既满足即插即用的部署需求，又保障仪器运行的合规性与安全性。



图 2.4 终端检测仪实物示意图

2.2.1 智能监测终端硬件设计

本终端硬件系统采用模块化架构，在实现高精度电流采集的同时，满足分布式部署所需的低功耗与宽电压适应能力。系统以 MCU 处理模块为核心控制单元，外围集成电流采集模块、NB-IoT 通信模块、GPS 定位模块、LED 指示模块及多级电源转换模块，整体硬件架构如

图 2.5 所示。

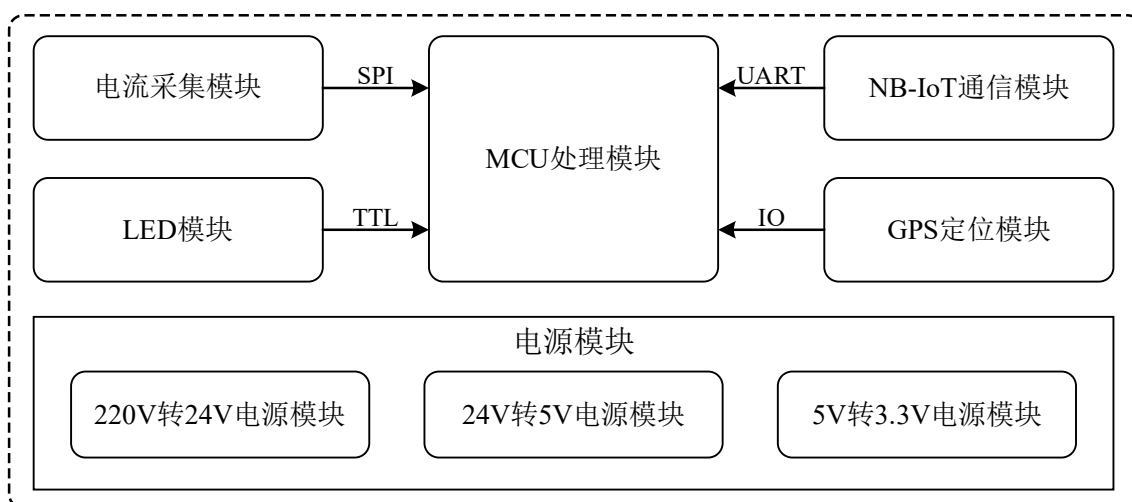


图 2.5 硬件结构图

(1) 总体架构与信号链路设计

终端硬件基于采集驱动和通信同步的逻辑构建。系统以低功耗 MCU 为中枢，通过模块化设计实现强弱电信号的物理与逻辑解耦。其信号流向逻辑如图 2.6 所示，主要步骤如下：

电流采集模块完成对科研仪器运行电流的感知与模拟信号调理，并由专用计量芯片将其转化为高精度的数字特征。MCU 随即通过串行总线读取这些特征数据，在本地完成边缘侧的二次计算与数据封装，处理后的数据通过串口封装发送至 NB-IoT 通信模块并上传至远程服务器；与此同时，GPS 模块提供设备的空间位置信息，LED 模块实时指示系统状态，而电源模块则为全系统提供稳定的多级直流电压供给。

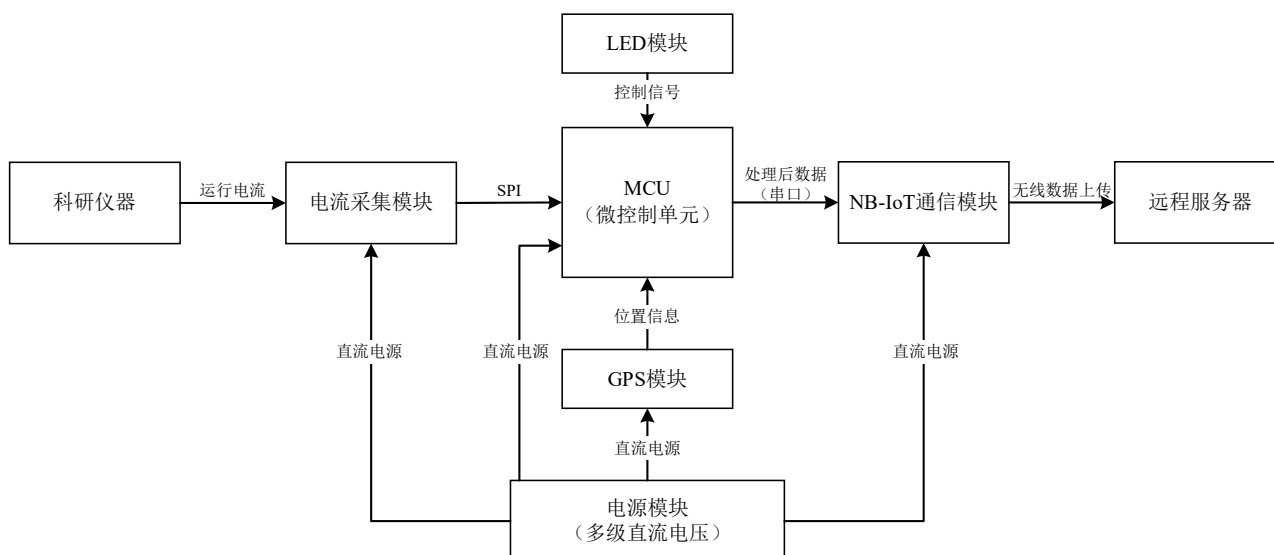


图 2.6 智能监测终端信号流向图

(2) MCU 处理模块

MCU 处理模块是智能监测终端的核心控制单元，负责数据总线调度、边缘特征聚合和外设通信管理等核心任务。系统采用基于 ARM Cortex-M3 内核的 STM32F103C8T6 作为主控芯

片。该芯片最高主频可达 72MHz，片上集成了丰富的 GPIO、通用定时器以及 USART/SPI 串行通信接口，提供足够的计算能力以支持边缘端的降频算法和统计量计算需求。

(3) 电流采集模块

电流采集模块将科研仪器的交流运行电流转化为供 MCU 处理的数字特征。考虑到实验场景对电磁兼容性与电气隔离的严格要求，本模块采用基于开合式电流互感器的非侵入式采集方案，如图 2.7 所示。

在前端感知层面，选用变比为 2000:1 的精密开合式电流互感器，在实现强弱电物理隔离的同时，有效覆盖了 0.1A 至 100A 的宽量程需求，确保了监测系统在极端异常电流下的安全性^[44]。互感器输出的微弱电流信号通过高精度低漂移的采样电阻进行 I-V 变换，将电流幅值映射为差分电压信号，为后续调理链路提供输入基准。

在信号调理层面，针对实验室环境中的复杂电磁干扰，电路设计了差分输入结构，并引入了二阶 RC 低通滤波器（截止频率设定为 2kHz）。该滤波器能有效滤除高频毛刺与工频谐波干扰，显著提高信噪比，确保信号满足奈奎斯特采样定理要求^[45]。

在核心计量层面，选用 V9260S 单相专用计量芯片。该芯片内部集成了两路高精度 Sigma-Delta ADC，支持 6400Hz 的过采样率^[46]。电路通过 SPI 总线与主控 MCU 通信，V9260S 内部集成的 DSP 引擎可实时计算电流有效值、有功功率等关键指标，并通过全量程校准逻辑将测量误差控制在 0.1% 以内。

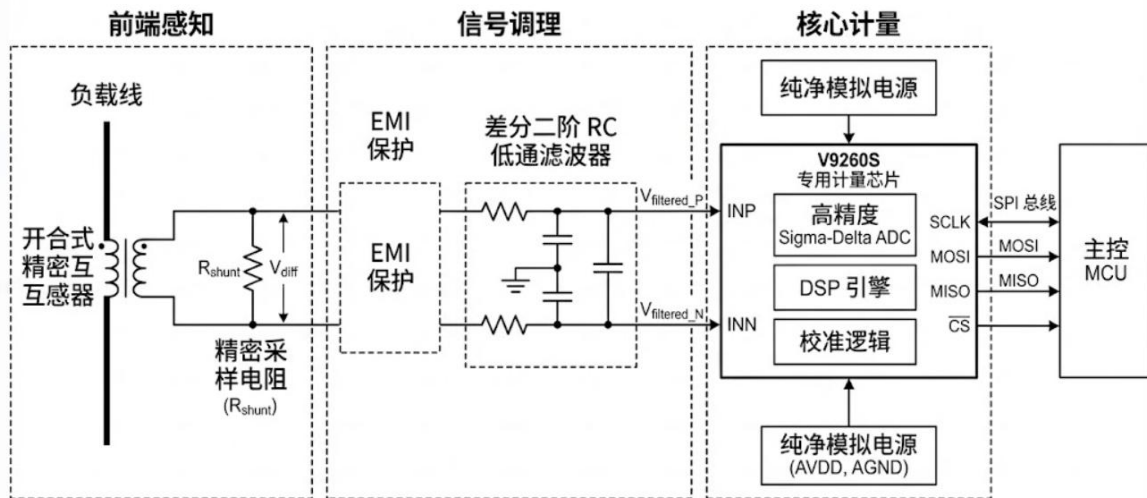


图 2.7 电流采集模块设计图

(4) NB-IoT 模块

针对高校实验室位于地下室或受精密仪器电磁屏蔽影响而导致信号覆盖较弱的问题，本系统选用广和通 L620 工业级 NB-IoT 模组。该模组支持运营商广泛覆盖的 850 MHz (Band 5) 与 900 MHz (Band 8) 频段，最大增益可达 +20 dB，能够有效穿透实验室墙体与金属屏蔽层，保障数据上行链路的可靠性^[47]。

如图 2.8 所示，硬件电路实现主要围绕通信接口逻辑、射频阻抗匹配和电源瞬态响应三个方面展开。

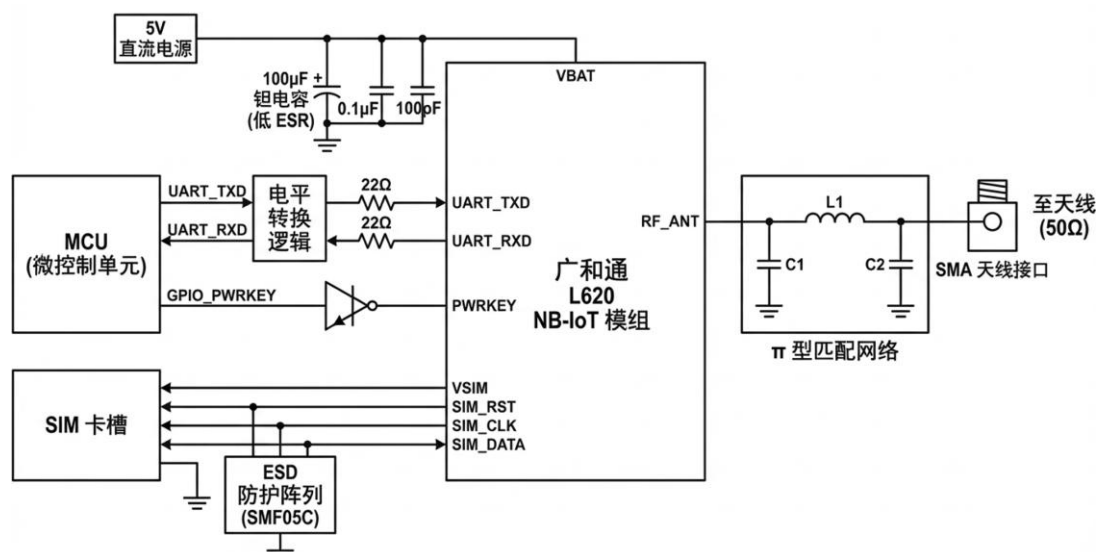


图 2.8 NB-IoT 通信模块设计图

1) MCU 通过 UART 串行接口 (TXD/RXD) 与 L620 模组连接, 利用 3GPP AT 指令集实现网络附着、数据透传及低功耗模式 (PSM) 切换^[48]。考虑到系统电磁兼容性, 在 UART 链路中串联匹配电阻以抑制信号振铃, 并在 MCU 与模组之间预留电平转换逻辑, 以确保通信链路在不同电压下的稳定性。此外, 设计了由 GPIO 控制的 PWRKEY 引脚驱动电路, 使 MCU 在系统死机或网络异常时, 通过硬件方式拉低引脚实现模组复位从而提高系统的自动恢复能力, 确保系统在异常情况下的连续运行。

2) 射频性能对弱信号环境下的丢包率有重要影响。为了最大化射频信号功率的传输, 在模组 RF_ANT 引脚与 SMA 天线接口之间设计了 Π 型阻抗匹配网络^[49]。该网络由两个接地电容和一个串联电感构成, 允许在 PCB 打样后根据网络分析仪的实测结果微调元件参数, 将传输线阻抗严格匹配至 50Ω , 从而将电压驻波比控制在 1.5 以下, 最大限度减少信号反射损耗。

3) NB-IoT 模组虽然平均功耗较低, 但在发射瞬间, 其峰值电流可达 500mA 以上。若电源供电不足, 会导致电压跌落引起的模组重启或 MCU 复位。为此, 电源输入端采用了大容量低 ESR 钽电容与多层陶瓷电容并联的组合滤波方案。钽电容用于提供瞬态电流补偿, 防止电压骤降, 陶瓷电容则用于滤除射频噪声。同时, 在 SIM 卡接口电路中集成了专用的 ESD 防护阵列, 防止插拔 SIM 卡时的人体静电损坏模组内部敏感电路。

(5) LED 模块

LED 模块是智能监测终端唯一的本地状态指示接口。鉴于监测设备通常部署于隐蔽位置, 该模块主要用于向现场运维人员提供网络链路状态的实时指示, 从而降低安装调试与故障排查的难度。

LED 驱动电路由 MCU 的 GPIO 引脚直接控制。电路中串联限流电阻以控制工作电流, 在保证 LED 在昏暗实验室环境下具备良好可视性的同时, 尽可能降低系统静态功耗。MCU 通过输出高低电平控制 LED 的通断, 实现不同的闪烁逻辑, 其控制流程如图 2.9 所示。

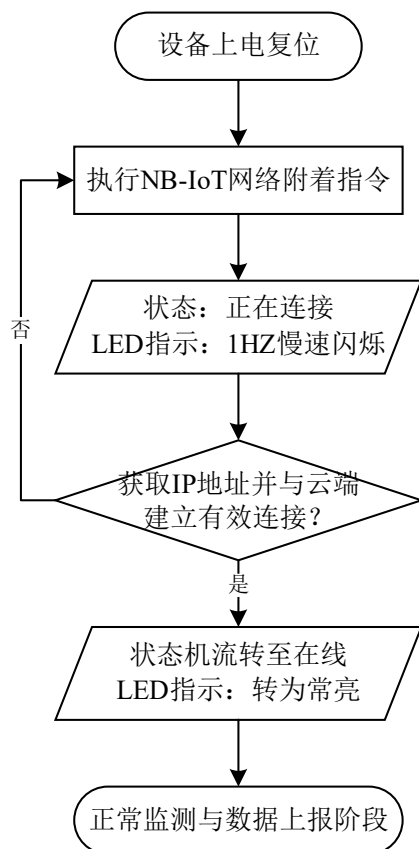


图 2.9 LED 状态反馈流程图

(6) GPS 模块

为满足高校资产管理中仪器位置监测与时空溯源的需求，本系统选用 Air530 双模卫星导航模块^[50]，该模块支持 GPS L1 与 BDS B1 双星系联合定位，能够在复杂电磁环境中提供精确的授时与定位服务。

Air530 通过 UART 串行接口与 STM32F103C8T6 连接，通信波特率配置为 9600bps。模块输出遵循标准 NMEA-0183 协议，MCU 通过解析协议帧提取经度、纬度及 UTC 时间信息。为缩短首次定位时间，电路设计中引入了后备电源接口并外接纽扣电池，以在主系统断电时保持星历数据，从而实现快速热启动。同时，为降低金属机柜对信号的屏蔽影响，硬件预留了 IPEX 天线接口，支持外接高增益有源陶瓷天线，以提高搜星速度和定位稳定性。

通过引入该模块，平台不仅能够实现设备物理位置的静态记录，还可以结合电子围栏逻辑，在设备发生未经授权的位移时即刻触发告警，从而提升实验室精密仪器的安全管理水平。

(7) 电源模块

电源模块负责为终端各功能单元提供稳定、低噪声的供电电压。针对科研仪器复杂的电磁环境，本系统设计了 AC-DC→DC-DC→LDO 的三级降压架构。220V 交流电首先经由 EMI 滤波网络与隔离型开关电源转换，形成 24V 直流母线，在满足外围传感器供电需求的同时降低传输损耗；随后利用高效率 DC-DC 转换芯片降压至 5V，为 NB-IoT 通信模组及 GPS 模块提供足以应对射频发射瞬间大电流脉冲的能量储备；最后，通过低噪声线性稳压器将电压进

一步稳压至 3.3V，为 MCU 主控单元与 V9260S 计量芯片供电。这种结合开关电源与线性稳压器的混合供电策略，有效抑制了电源纹波对模拟采样精度的影响，并在系统功耗与信号质量之间取得了较好的平衡。

2.2.2 边缘侧数据聚合与传输策略

智能监测终端的软件系统负责实现原始电信号的特征提取与数据转发。针对 NB-IoT 窄带宽、高时延的传输特性，本系统在边缘侧摒弃了高频透传模式，采用本地高频采样、边缘特征聚合和稀疏周期上报的策略，以提升链路能效并降低云端并发负载。

(1) 嵌入式软件调度架构

终端软件设计遵循模块化与低耦合原则，核心任务调度由 MCU 主控单元统一调度。系统运行流程采用有限状态机（FSM）模型，主要包含系统初始化、网络附着、数据采集循环与异常处理四个状态。

MCU 调度流程如图 2.10 所示。系统上电后首先进行硬件自检与外设初始化。MCU 通过 UART 串口向 NB-IoT 模组发送 AT 指令序列以请求入网，同时唤醒 GPS 模块获取地理位置与 UTC 时间。完成握手检测、SIM 卡鉴权及网络注册后，系统进入主循环状态：

- 1) 采集任务：MCU 通过 SPI 总线定时读取 V9260S 计量芯片寄存器中的电流有效值。
- 2) 缓存任务：将读取的瞬时数据写入 MCU 内部的环形缓冲区，等待数据量达到阈值。
- 3) 通信任务：当缓冲区满或达到定时上报周期时，触发数据封装与无线发送中断。

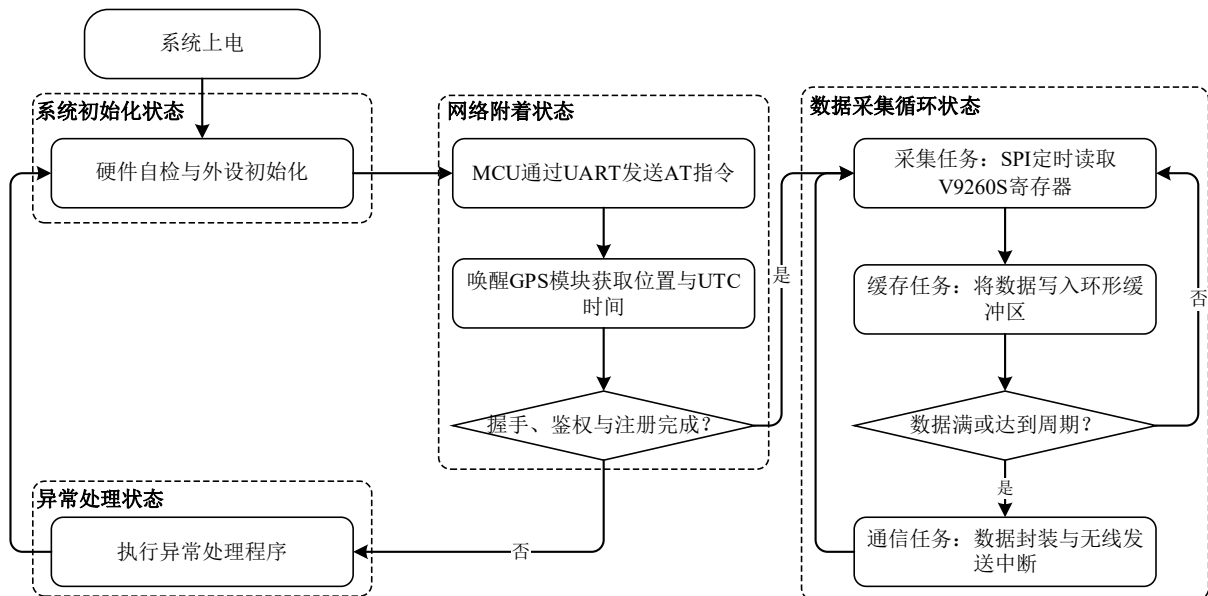


图 2.10 MCU 调度流程图

(2) 基于 LwM2M 的协议栈设计

考虑到实验室环境中可能存在信号盲区，以及 NB-IoT 的窄带特性，网络层采用了由开放移动联盟提出的 LwM2M（Lightweight Machine to Machine）协议。相比 HTTP 和 MQTT，LwM2M 底层基于 UDP 协议，头部开销极小，能够更好地适应资源受限设备的通信需求。

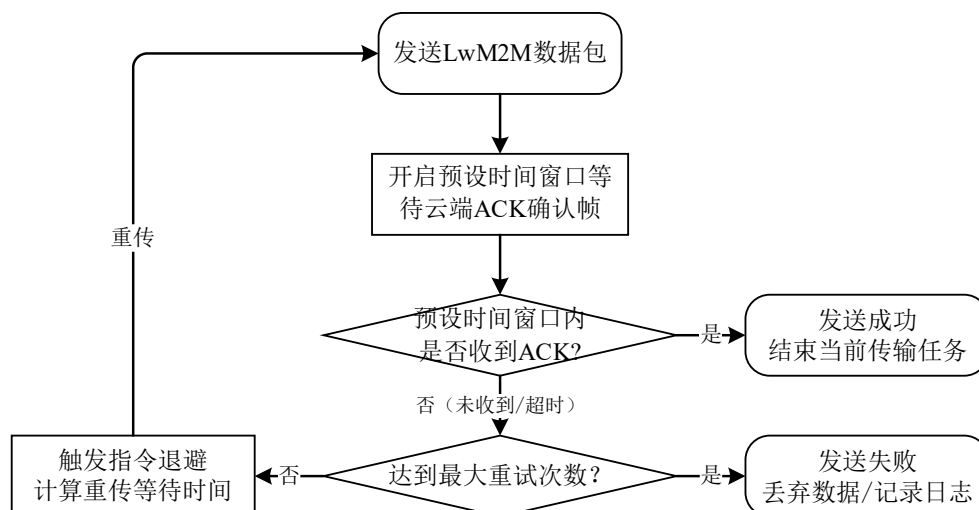


图 2.11 应用层传输流程图

系统定义了符合 LwM2M 规范的资源对象结构，将电流数据、设备状态、信号强度等映射为标准化的资源 ID。考虑到 UDP 协议本身的不可靠性，软件层设计了图 2.11 所示的应用层重传机制。当 MCU 发送数据包后，若在预设时间窗口内未收到云端的 ACK 确认帧，系统将采用指数退避方式延长等待时间并触发重传，直至发送成功或达到最大重试次数。该机制有效缓解了实验室弱信号环境下的瞬时丢包问题，提高了关键业务数据的到达率。

(3) 数据聚合与帧结构定义

为平衡数据实时性与传输功耗，终端摒弃了传统的即采即发的透传模式，转而采用时间分片聚合策略。考虑到长期监测产生的巨大数据量，且科研仪器的运行状态通常较为平稳，系统将边缘侧的记录频率设定为每分钟 1 次。MCU 在本地设置数据缓冲区，每采集一个电流点即存入 Flash，当累计达到 30 个数据点后，将这组数据作为一个数据包进行封装。这种批处理方式将网络请求频率降低为原来的 1/30，大幅减少了频繁唤醒带来的额外流量消耗，同时也让 NB-IoT 模组有更多时间处于休眠状态，显著延长了设备续航。

系统定义了两种核心数据帧格式：初始化帧与周期性上报帧。

1) 初始化帧

设备每次重启或重连成功后，首先发送初始化帧。该帧包含功能码、设备唯一标识(IMEI)及信号强度，用于云端注册设备上线状态并评估当前网络质量。

表 2.1 初始化帧结构

序号	参数项	数据类型	说明	内容示例
1	功能码	Hex	固定值	INIT_ID
2	设备标识	String	15 位 IMEI 码	866497066682841
3	信号强度	Byte	反映当前基站信号质量(>12 为优)	13

2) 周期性上报帧

周期性上报帧是平台获取设备电流信息的核心数据帧，其帧结构如下表 2.2 所示，包含

功能码、IMEI、时间戳以及聚合后的电流数据块。其中，时间戳记录了该组数据打包时的 Unix 时间，确保数据的时序可追溯，电流数据块则包含 30 个连续的电流有效值。

表 2.2 周期性上报帧结构

序号	参数项	数据类型	说明	内容示例
1	功能码	Hex	固定值	M_NORMAL_REPORT_ID
2	设备标识	String	15 位 IMEI 码	866497066682841
3	运行电流集合	Array[Float]	包含 30 个浮点型电流数据 (每点间隔 1 分钟)	{0.57, 1.36, ..., 1.88}
4	上报时间戳	Long	打包时刻的 Unix 时间戳(ms)	1761208338000

通过上述策略，系统在边缘侧完成了数据的初步清洗与聚合，在保留科研仪器运行状态可观测的前提下，通过大幅降低无线链路的激活频次，有效缓解了 NB-IoT 的带宽压力与功耗开销，实现了监测精度与系统能效的平衡。

2.3 数据混合存储设计与实现

科研仪器监测平台在运行过程中会产生多种类型的异构数据，包括分钟级的电流采样序列、低频更新的设备元数据以及瞬态的系统运行状态。若采用单一的关系型数据库进行存储，面对海量时序数据的写入压力时会出现性能瓶颈，且难以满足非结构化数据的灵活查询需求。因此，本系统采用混合存储策略，根据数据特征进行分类选型，并构建了统一的异构数据持久化层。

2.3.1 数据特征分类与存储选型

系统从数据结构特性、读写吞吐需求和一致性要求三个维度，对上报数据进行分类分析，具体的多源数据特征与混合存储选型对比见表 2.3。

表 2.3 多源数据特征与混合存储选型对比

数据分类	数据特征	存储需求难点	选型方案
时序监测数据	高频写入	传统关系型数据库写入扩展性不足，难以适应灵活的数据结构	文档数据库 (MongoDB)
	追加式存储		
	时间序列结构		
结构化业务数据	低更新需求	事务一致性与严格关系约束	关系型数据库 (MySQL)
	模型固定，关系明确		
	读多写少，一致性要求高		
瞬态热点快照数据	时效性短	极低延迟与高并发响应	内存数据库 (Redis)
	Key-Value 结构		
	高并发访问		

(1) 强一致性结构化元数据

此类数据涵盖仪器资产台账、终端硬件标识以及两者间的逻辑绑定关系。其特征是写操作频率低、数据模型相对固化且关联查询频繁。为了保障多表关联的事务完整性，系统选用具备严格 ACID 特性的关系型数据库 MySQL 进行存储。该选型为平台静态资源管理提供了稳定的数据基础，确保了资产管理逻辑的准确性。

(2) 高吞吐半结构化时序与算法数据

该类数据包括终端上报的分钟级电流序列、清洗后的状态特征以及算法训练产生的中间标签。其核心挑战在于写操作密集且数据量随监测时长线性累积。基于此，系统摒弃了传统建表的模式，选用以 BSON 为原生格式的文档型数据库 MongoDB，利用其无固定模式（Schema-free）特性应对灵活多变的数据结构需求。

(3) 高并发瞬态热点数据

为支撑前端大屏的实时监控与预警推送，系统需频繁获取设备的当前上下线状态、瞬时电流等指标。此类数据具有时效性强、生命周期短的特点，通常以简单的键值对（Key-Value）形式存在，但极易成为系统的并发访问瓶颈。针对高频读写的热点访问场景，系统引入了基于内存计算的 Redis 数据库，以极低的访问延迟构建数据缓冲层^[51]。

2.3.2 混合持久层实现

为保障平台在海量设备接入场景下的高可用性与容灾能力，混合持久层采用了分布式集群架构。

(1) 关系型存储层

针对平台中模型固化、生命周期长且对事务强一致性要求较高的结构化元数据，系统选用 MySQL 作为核心关系型存储引擎，并搭建了图 2.12 所示一主多从的 MHA（Master High Availability）架构，以降低单点故障带来的风险。MHA 通过后台守护脚本实时监控节点健康状况。当主库发生宕机时，MHA 能够在 30 秒内自动完成主从角色的故障转移，并尽可能保障二进制日志的一致性。

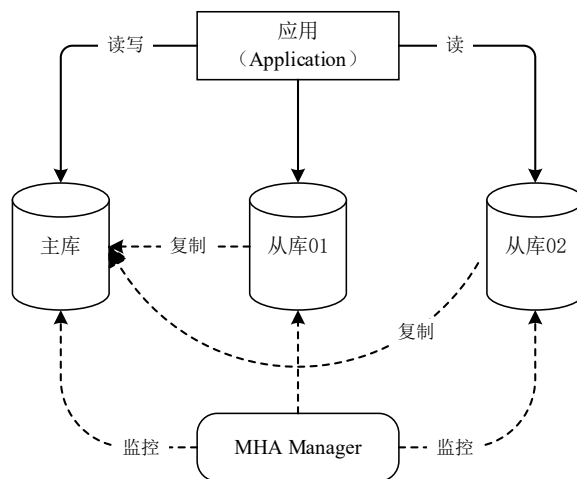


图 2.12 MySQL MHA 分布式集群架构图

在底层表结构设计上，严格遵循数据库范式，主要构建了三大核心实体表：

1) 科学仪器资产表。该表记录科研仪器的物理位置、归属部门、负责人、价值及注册时间等信息，为平台的资产管理提供基础数据支撑。

表 2.4 科学仪器资产数据表

字段名	数据类型	字段注释	字段名	数据类型	字段注释
id	Int(11)	唯一标识	address	Varchar(40)	所在地址
asset_id	Varchar(100)	仪器编码	worth	Varchar(40)	仪器价值
name	Varchar(100)	仪器名称	user_name	Varchar(100)	负责人姓名
asset_type	Varchar(100)	仪器类型	user_phone	Varchar(100)	负责人联系方式
department	Varchar(100)	所属部门	binding_id	Varchar(40)	关系编码
school	Varchar(100)	所属高校	install_time	Timestamp	安装时间

2) 终端设备表。该表用于记录非侵入式检测终端的出厂溯源信息（IMEI）、硬件通信参数以及 GPS 空间坐标，为设备的入网鉴权提供依据。

表 2.5 终端设备数据表

字段名	数据类型	字段注释	字段名	数据类型	字段注释
id	Int(11)	唯一标识	device_type	Tinyint(2)	终端类型
imei	Varchar(40)	终端编码	lon_range	Varchar(255)	经度位置
product_id	Varchar(40)	生产编码	lat_range	Varchar(255)	纬度位置
master_key	Varchar(100)		binding_id	Varchar(40)	关系编码
app_key	Varchar(40)	入网相关编号	outstock_time	Timestamp	出库时间
nb_id	Varchar(40)		-	-	-

3) 科学仪器与终端关联数据表。该表动态维护仪器与终端间的物理绑定状态。针对设备更迭频繁的实验室场景，系统引入雪花算法（Snowflake）生成全局唯一的绑定主键。即使发生终端硬件替换，也只需在映射表中更新关联关系，即可确保历史运行数据在逻辑上的连续性与数据的完整性，从而避免因硬件变更导致的数据中断。

表 2.6 科学仪器与终端关联数据表

字段名	数据类型	字段注释	字段名	数据类型	字段注释
id	Int(11)	唯一标识	imei	Varchar(40)	终端编码
binding_id	Varchar(40)	关系编码	is_binding	Tinyint(2)	是否解绑
unbinding_id	Varchar(40)	解绑编码	binding_time	Timestamp	最近绑定时间
asset_id	Varchar(100)	仪器编码	unbinding_time	Timestamp	最近解绑时间

(2) 文档型存储层

针对高吞吐的时序电流特征、结构多样的运行状态快照以及随算法迭代不断变化的特征工程数据，系统选用支持原生 BSON 结构的 MongoDB 构建文档型存储层。为避免单节点的 I/O 吞吐瓶颈，系统部署了 MongoDB Sharded Cluster（分片集群）架构，其逻辑拓扑如图 2.13 所示。该架构利用配置服务器（Config Server）管理集群元数据，并通过路由节点（Mongos）根据预设片键（Shard Key）将海量时序数据自动分发至多个物理独立的副本集（Replica Set）中。这种分片机制实现了存储容量与并发写入能力的线性扩展，并为第三章中海量时序样本的特征检索提供了高性能的数据访问支持^[52]。

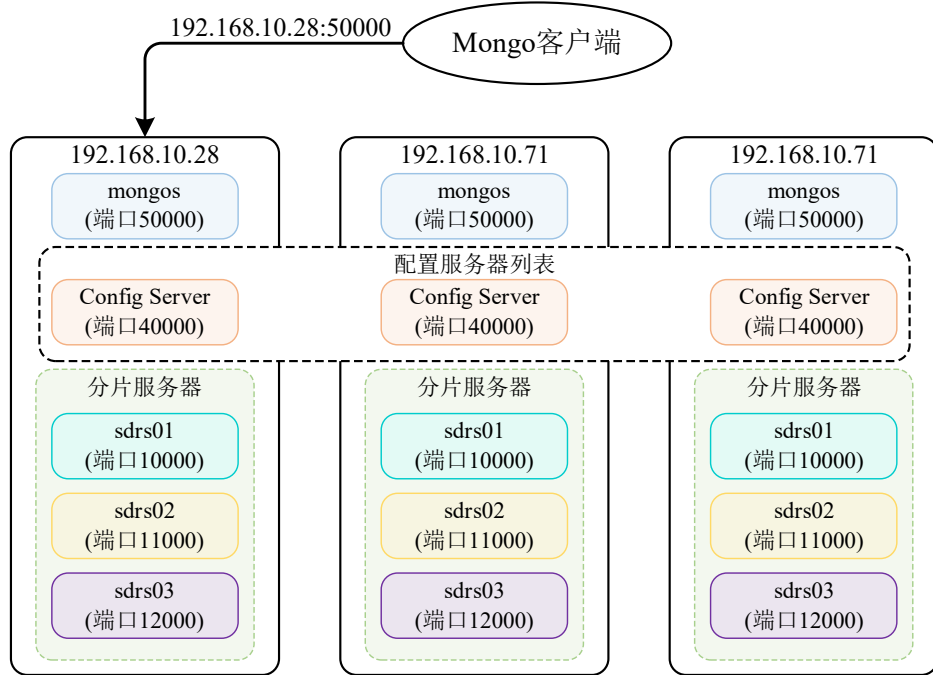


图 2.13 MongoDB 分片集群架构图

MongoDB 中主要包含以下数据集合：

1) 终端电流数据集合。终端采集的数据按日进行集合划分，并将高频上报的离散数据组织为包含 startTime、duration 和 value 字段的数据块，以降低存储碎片率。

表 2.7 终端电流数据集合结构表

字段名	数据类型	字段注释			
_id	Object	唯一标识			
imei	String	终端编码			
time	Long	该条记录时间戳			
dataList	startTime	Long	起始时间		
	value	Object	Double	数据集合	值
	duration		Int		持续时间

2) 仪器状态数据集合。该集合记录终端上传电流所对应的仪器运行状态，可通过终端编码进行查询。

表 2.8 状态数据集结构表

字段名	数据类型	字段注释
_id	Object	唯一标识
imei	String	终端编码
time	Long	该条记录时间戳
dataList	startTime	Long 起始时间
	label	Object Double 数据集合 状态标签
	duration	Int 持续时间

3) 算法模型数据集。该集合用于存储算法推理结果与专家先验知识的关联信息。按时间窗口粒度,集合持久化存储 QA-TF-CL 模型输出的融合特征指纹、空间偏离度及重构误差,还同步记录用于前端的 t-SNE 降维坐标。此外,集合预留了 expert_status 与 semantic_label 字段,用于记录运维专家的交互反馈结果,从而支持无监督特征向动态故障知识表示的持续转化。

表 2.9 算法模型数据集结构表

字段名	数据类型	字段注释
_id	Object	唯一标识
imei	String	终端编码
window_start	Long	窗口起始时间戳
window_end	Long	窗口结束时间戳
quality_mask	Double	数据质量掩码评分
feature_fingerprint	Array[Double]	128 维融合特征向量
tsne_coords	Object	t-SNE 降维二维坐标
cluster_id	Int	聚类编号
cluster_distance	Double	到聚类中心的距离
anomaly_score	Double	综合异常评分
anomaly_level	Int	异常等级
expert_status	Int	专家确权状态 (0-待办, 1-确认故障, 2-标记误报/负样本, 3-忽略)
semantic_label	String	专家语义标签
model_version	Varchar(32)	算法版本号
create_time	Timestamp	创建时间

(3) 内存缓存层

为应对突发的高频并发请求,系统构建了去中心化的 Redis Cluster 架构。如图 2.14 所示,该架构通过 Gossip 协议实现节点间的状态探测与同步,并利用 16384 个哈希槽 (Hash Slot) 实现数据的自动分片寻址,从而避免了中心化代理模式下的性能瓶颈与单点故障风险。

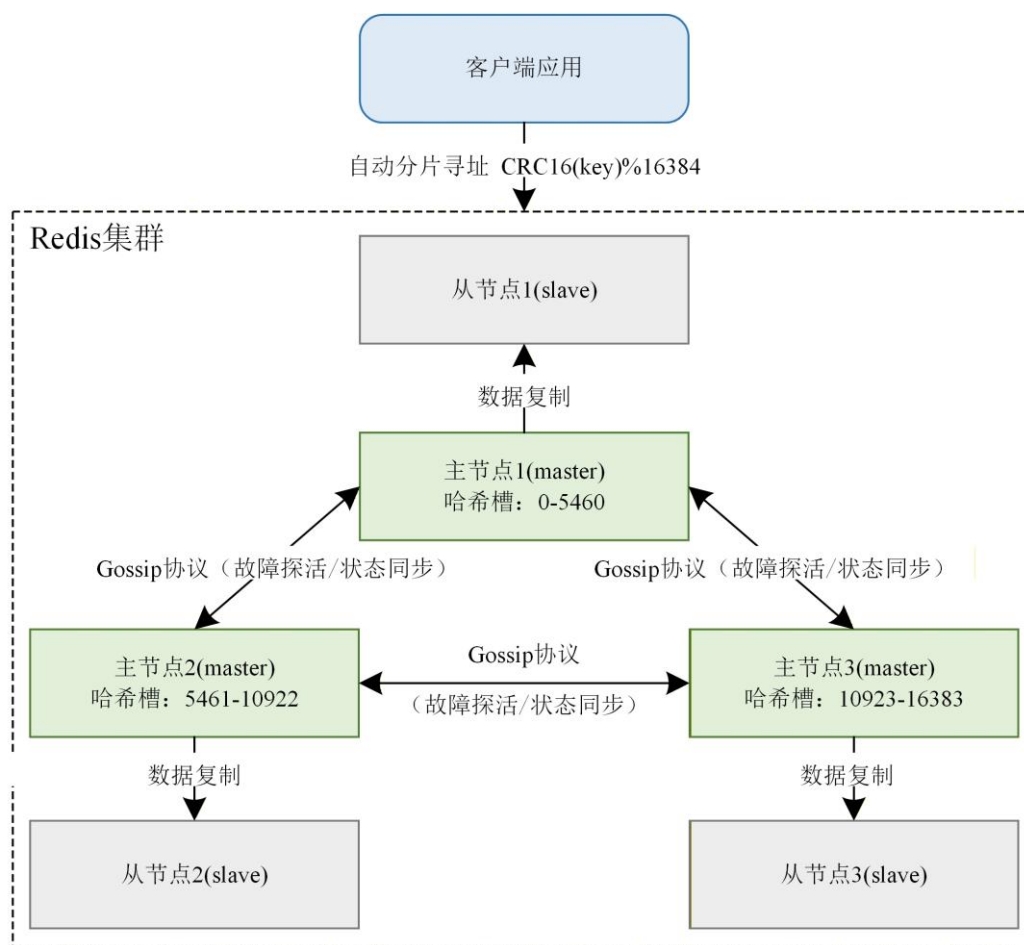


图 2.14 Redis 集群架构图

在实际业务处理中,系统以仪器的全局唯一编码作为 Hash Key,将终端的在线/离线标记、最新电流有效值及工况信息(如待机、运行)等热点快照存储于内存中。该设计显著降低了下层 MySQL 与 MongoDB 的查询承载压力,提升了平台在万级设备并发监测场景下的低延迟与高吞吐响应能力。

表 2.10 科学仪器在线状态表

字段名	数据类型	字段注释
key	String	终端编码
value	Int	0-离线 1-在线

2.4 业务逻辑与算法引擎的集成

为同时支撑海量设备并发接入带来的 I/O 密集型任务和深度学习推理带来的计算密集型任务,系统采用了异构计算架构,如图 2.15 所示。Java 端用于构建高吞吐的业务中台,Python 端用于承担算法推理任务,并通过共享状态机制实现两者之间的异步解耦与协同运行。

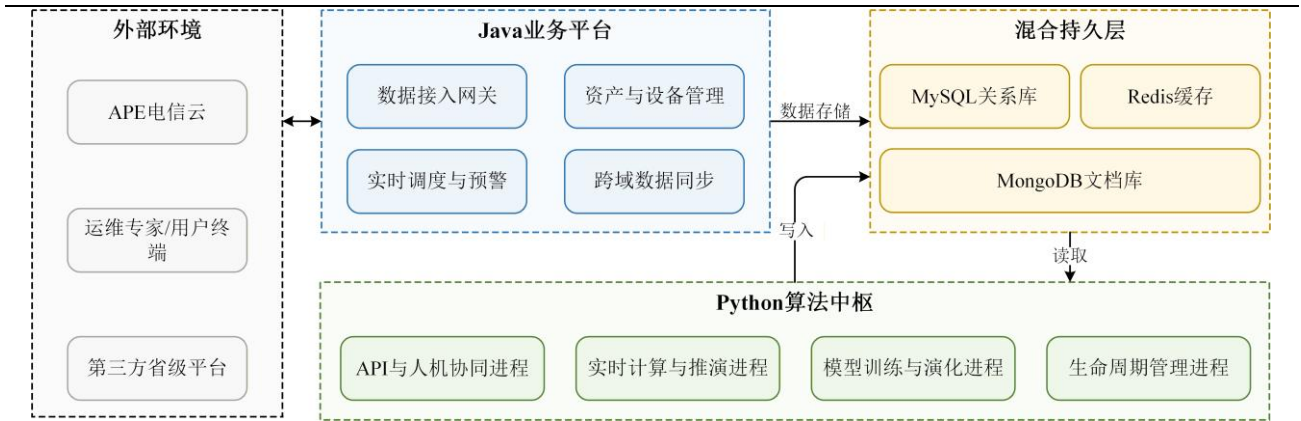


图 2.15 异构计算架构图

2.4.1 后端业务系统构建

Java 后端作为整个平台的数据入口与业务调度中心，基于 Spring Boot 与 MyBatis-Plus 框架构建服务化体系。为支撑全省范围内的设备接入与复杂运维需求，其主要职责可归纳为以下四个功能模块。

(1) 数据接入模块

该模块主要负责对接电信 AEP 等第三方物联网平台，接收底层传感器上报的脉冲数据。如图 2.16 所示，为应对海量设备并发上报引起的流量高峰，网关层引入 RabbitMQ 消息队列作为缓冲机制，以实现削峰填谷。数据报文在消费端经过鉴权、LwM2M 协议解析与数据校验后，被统一格式化并存储至 MongoDB 的待归档集合中，为后续算法分析提供干净的原始数据源。

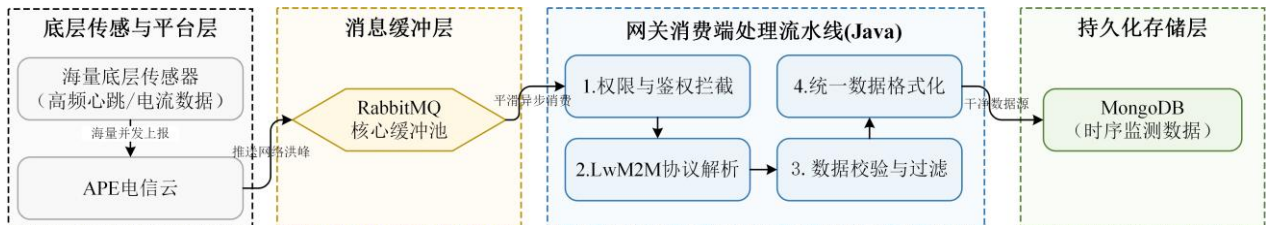


图 2.16 数据接入流程图

(2) 资产与设备管理模块

该模块负责与底层 MySQL 关系型数据库交互。该模块不仅承担仪器基础台账录入与终端设备出厂信息管理，还负责维护科研仪器与监测终端之间的动态物理映射关系。系统通过集成雪花算法生成全局唯一的绑定关系 ID，确保在设备损坏更换或跨实验室迁移时，历史运行状态数据仍能够保持逻辑连续性并实现准确溯源。

(3) 实时调度与预警模块

该模块负责接收并处理算法引擎输出的分析结果，进而驱动业务流转。一方面，模块通过 WebSocket 全双工通道将设备的实时状态变化主动推送至前端监控大屏，以实现快速页面响应；另一方面，模块内置规则引擎，针对算法端判断为高置信度告警的异常事件，自动触发短信通知，并向设备第一责任人发送分级故障告警信息。

(4) 跨域数据同步模块

业务端内置 Quartz 定时任务调度引擎，可定时执行机时统计数据汇总等任务，例如每 15 分钟触发一次。该模块负责将系统生成的仪器在线率、有效机时、异常评分等核心指标，通过标准化的 HTTP/WebService 接口定时推送至浙江省科技厅大仪网和市场监督管理局等第三方省级监管平台，以实现跨平台数据同步和监管协同。

2.4.2 算法引擎系统构建

针对第三章提出的 QA-TF-CL 时频双视图对比学习算法及第四章的知识检索与演化机制，考虑到系统对 PyTorch、NumPy 等矩阵计算库的依赖，算法层采用 Python 作为主要开发语言。为充分利用计算节点的 CPU/GPU 资源，并减小 Python 全局解释器锁带来的性能限制，算法服务采用了多进程并发架构。在服务启动时，系统主进程会同时启动四个独立子进程协同运行，以实现特征提取、异常判别与人机协同等任务的流水线式处理。

(1) API 与人机协同进程

该进程主要运行基于 ASGI 标准的 FastAPI 应用，除提供手动重算、模型配置下发等同步接口外，其核心功能是接收第四章中的人机协同反馈。当运维专家在前端 Web 页面对灰区疑似样本进行确权并赋予语义标签后，该进程负责接收相关反馈，并将其映射存入底层动态故障知识库，为后续案例检索与模型微调提供高质量标注数据

(2) 实时计算与推演进程

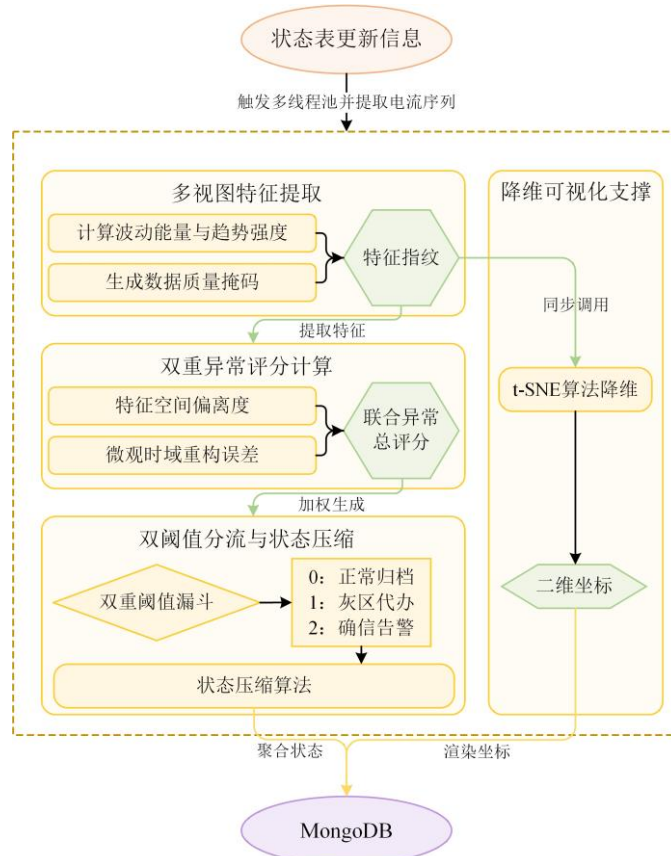


图 2.17 实时计算与推演进程核心逻辑

作为系统的后台守护进程，该进程持续监听 MySQL 中的状态更新信息，并利用多线程池并发执行推理任务。其核心逻辑如图 2.17 所示，首先计算数据的波动能量与趋势强度，并生成数据质量掩码以抑制插值噪声；随后通过双流网络提取融合特征向量。基于提取的特征，实时计算当前状态与基准簇的特征空间偏离度及时域重构误差，加权生成联合异常总评分。随后引入第四章的双重阈值漏斗机制对评分进行分级判定。同时，系统调用状态压缩算法，对相邻且类别相同的冗余时序散点进行聚合，将其转化为带有持续时间属性的紧凑状态块，从而大幅降低存储压力。此外，为支持前端大屏对特征分布的可视化展示，后台进程还会调用 t-SNE 算法对高维特征进行降维，生成二维坐标并与结果一同存入持久层。

(3) 模型训练与知识演化进程

该进程按预设时间在后台触发数据聚合与模型演化任务。该进程不仅负责聚类基准簇的更新，还承担 QA-TF-CL 模型的增量微调任务。该进程会抽取近期的专家确权样本，构建正负样本库，并利用对比损失引导模型重新优化特征空间边界，从而持续提升系统性能。

(4) 生命周期管理进程

该进程负责冷热数据的生命周期管理，定期将超过保存期限的原始高频数据从 MongoDB 热库迁移至冷存储或执行物理删除，同时清理失效的中间特征向量，以保障内存资源与数据库引擎的长期运行效率。

2.4.3 跨语言通信与异步解耦机制

在 Java 业务层与 Python 算法层的数据交互上，系统未采用强耦合的同步调用模式，而是设计了基于 MySQL 核心监控表 (tb_update_record) 的状态轮询机制，其核心闭环流转流程如图 2.18 所示。

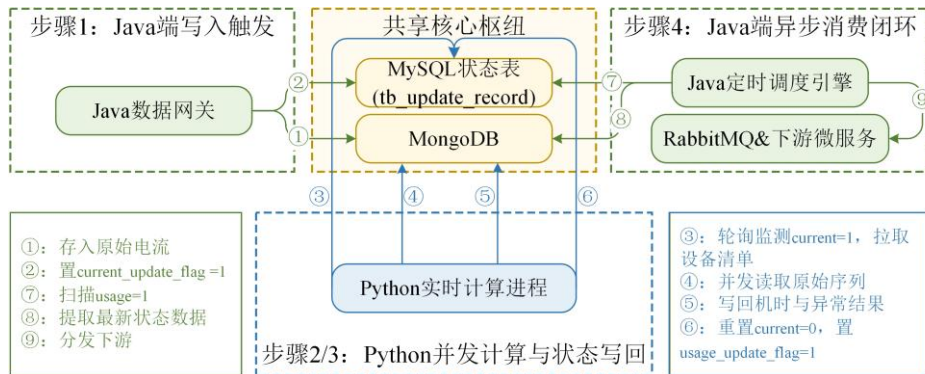


图 2.18 跨语言通信闭环流转图

当 Java 数据网关接收到终端的新增电流数据并成功存入 MongoDB 后，会同步更新 MySQL 数据库，将 tb_update_record 表中对应设备的 current_update_flag 标志位置为 1。该操作可视为向算法层发出异步计算请求。

Python 算法服务的实时计算进程不断轮询状态监控表，并通过“SELECT imei FROM tb_update_record WHERE current_update_flag = 1”语句批量获取需要进行计算的设备清单。随后，算法服务开启多线程并发读取 MongoDB 中的原始电流序列，进行状态特征分类与机时

聚合计算。

计算完成后, Python 进程将机时结果与异常标签写回 MongoDB 的模型数据集合中。同时, 执行 SQL 更新操作以重置相关状态标志位: 将 `current_update_flag` 置为 0, 表示该批次电流数据已处理完成; 将 `usage_update_flag` 置为 1, 表示机时结果已生成, 等待上层业务模块进一步处理。

同时, Java 端的定时调度引擎 (CalculateUsageJob) 定期扫描 `usage_update_flag=1` 的记录。一旦发现标志位变更, 立即从 MongoDB 中提取最新的设备状态数据, 并通过 RabbitMQ 将统计结果分发至各个订阅了告警和统计队列的下游服务, 用于大屏渲染及第三方系统推送。

通过上述基于状态监控表的异步交互机制, Java 与 Python 服务各自依靠 flag 标志位独立运行、互不干扰。该架构不仅提高了跨语言通信的可靠性, 也增强了系统的故障容错能力。当任一服务发生重启或异常退出时, 未处理任务的状态标志位仍保存在数据库中, 待服务恢复后可继续执行后续处理流程, 从而保障监测平台的数据完整性与最终一致性。

2.5 交互层设计与实现

针对实验室精细化管理与移动化运维场景, 本系统采用跨平台架构设计, 构建了网页 Web 端与微信小程序端两类交互应用。

2.5.1 网页端实现

网页 Web 端主要用于全局监控、数据分析与专家知识录入。前端采用响应式框架 Vue 3.0 与 Element Plus UI 组件库进行开发, 并结合 ECharts 数据可视化引擎完成时序波形与散点图的渲染。

(1) 多维数据可视化大屏

可视化大屏是平台进行全局状态展示与监控的核心界面。前端通过异步接收后端推送的统计结果, 将海量物联网数据与算法推理结果以图形化方式展示。如图 2.19 所示, 可视化大屏主要划分为以下四个模块。

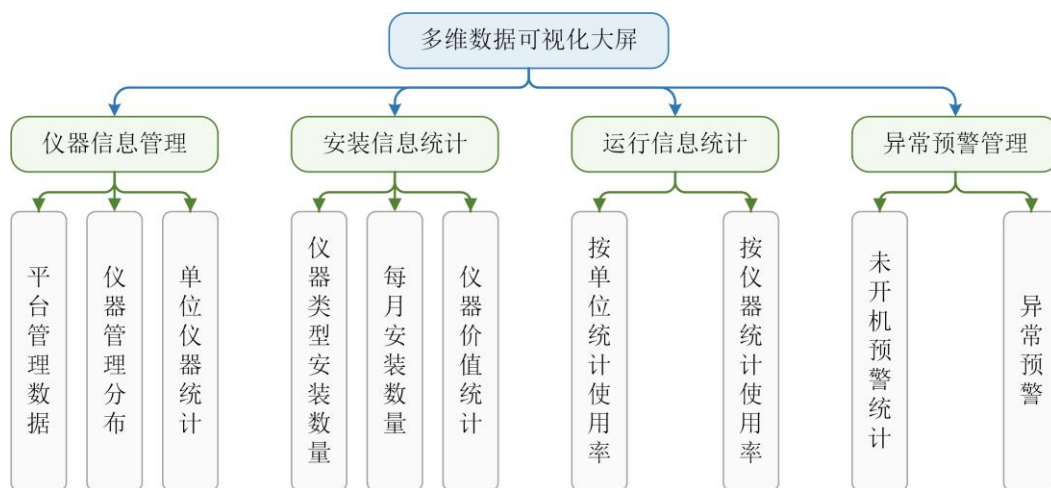


图 2.19 可视化大屏功能模块图

1) 仪器信息管理视图模块包含平台管理数据、仪器管理分布与单位仪器统计三个子功能项。如图 2.20 所示, 该界面通过地理拓扑图与柱状占比图, 直观展示接入仪器的总量、各类科研设备在不同实验室的空间分布, 以及各归属单位名下的仪器规模情况, 为管理者提供清晰的资产分布信息。



图 2.20 仪器信息管理视图

2) 安装信息统计视图侧重于呈现物联网监测终端的现场部署进度。内部分为仪器类型安装数量、每月安装数量与仪器价值统计三个子项。利用折线图展现各月份终端装机的历史走势, 并直观展示不同类型、不同价值区间的监控覆盖情况, 辅助相关部门进行工程验收与采购预算规划。

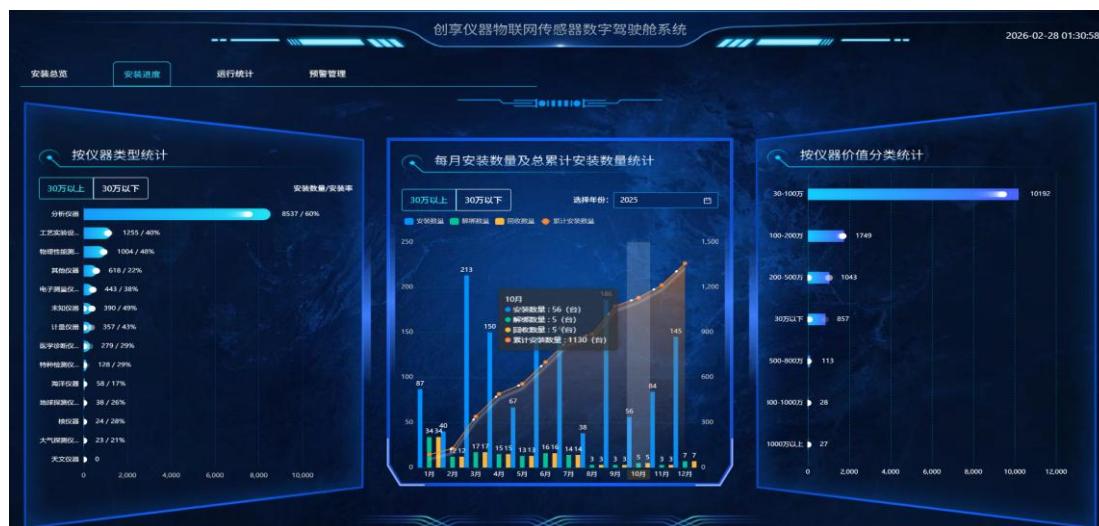


图 2.21 安装信息统计视图

3) 运行信息统计视图主要分为按单位和按仪器分类的使用情况列表。如图 2.22 (b) 所示, 界面上方利用饼图、柱状图和折线图, 宏观展示了特定单位内仪器的分类占比、资产价值分布以及整体的实时开机率趋势; 下方的数据列表则微观记录了单台重点仪器的实时运行状态与年有效使用机时。如图 2.22 (c) 所示, 该视图转换分析视角, 以特定仪器类别为切入点进行展示。界面上方通过柱状图详细刻画了该类仪器在全天 24 小时各时间段内的机时分布

规律，以及跨月份的使用率演变趋势，直观反映了同类设备的时间利用情况；下方列表则展示该类别下的具体设备明细，用于横向对比各台仪器的原值、年总机时与年平均使用率。



(a) 运行信息统计主视图



(b) 某单位仪器运行信息统计视图



(c) 某仪器分类运行信息统计视图

图 2.22 运行信息统计视图

4) 异常预警管理视图模块主要用于展示 QA-TF-CL 算法输出的联合异常判断结果。该模块包含未开机预警统计与异常预警两大子项。前者针对长时间未产生有效工作电流的闲置仪器进行预警汇总；后者则以预警列表的形式，实时展示触发高置信度阈值的设备故障。



图 2.23 异常预警管理视图

(2) 安装运维管理工作台

安装运维管理工作台是平台实现数字资产流转与设备全生命周期维护的管理后台。该模块主要面向省级监管人员、校级设备负责人以及一线运维专员等不同层级的用户群体，提供资产与智能终端的精细化管理功能。为确保数据的安全性与管理规范性，工作台底层引入了基于角色的访问控制机制，通过分级权限管理实现数据的逻辑隔离。根据实际的物联网运维业务流程，该工作台划分为如图 2.24 所示的四大核心功能。

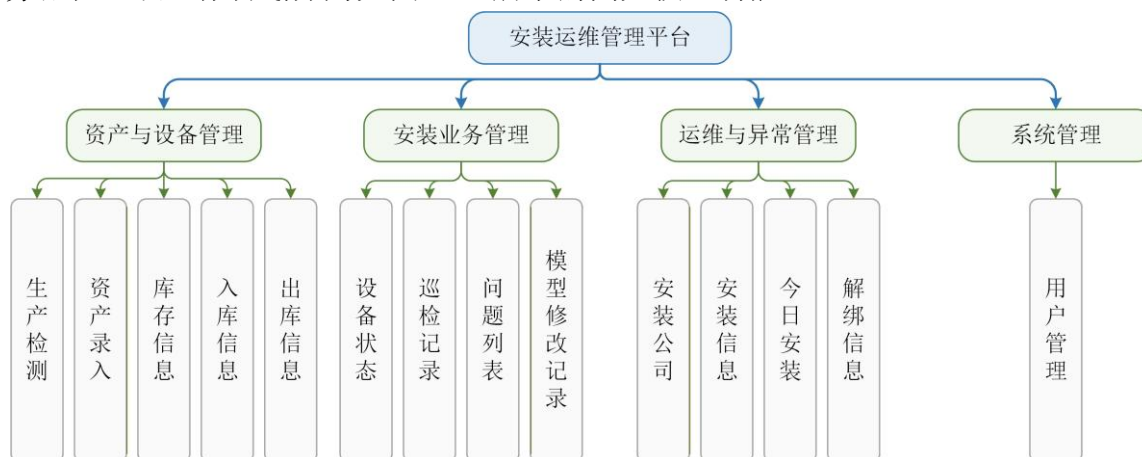


图 2.24 安装运维管理平台架构图

1) 资产与设备管理

该模块对应底层 MySQL 的资产与设备实体表，覆盖监测终端从出厂到仓储的完整链条。该功能涵盖终端生产环节的质量与性能检测登记、新资产基础档案的录入、统计与管理，以及对库存余量、入库明细与出库动态的实时追踪，从而确保硬件资产信息的一致性与流转过

程的可追溯性。

序号	设备IMEI码	设备类型	标准电流值	电流误差范围	当前电流值	标准信号量	信号误差范围	信号质量	是否合格	检测时间	所属平台	操作
7	869275070780729	单桩	0.87	5	0.87	10	1	22	合格	2025-04-24 19:48:49	浙江大仪平台	迁移
8	869275070772593	单桩	0.87	5	0.89	10	1	25	合格	2025-04-24 19:47:40	浙江大仪平台	迁移
9	869275070776123	单桩	0.87	5	0.86	10	1	27	合格	2025-04-24 19:02:15	浙江大仪平台	迁移
10	869275070780026	单桩	0.87	5	0.89	10	1	31	合格	2025-04-24 18:16:14	浙江大仪平台	迁移
11	869275070780061	单桩	0.87	5	0.89	10	1	24	合格	2025-04-24 18:05:06	浙江大仪平台	迁移
12	869275070772171	单桩	0.87	5	0.89	10	1	26	合格	2025-04-24 18:01:02	浙江大仪平台	迁移
13	869275070778251	单桩	0.87	5	0.90	10	1	24	合格	2025-04-24 17:58:13	浙江大仪平台	迁移
14	869275070774318	单桩	0.87	5	0.90	10	1	20	合格	2025-04-24 17:56:52	浙江大仪平台	迁移
15	869275070775646	单桩	0.87	5	0.89	10	1	31	合格	2025-04-24 17:55:36	浙江大仪平台	迁移

图 2.25 资产与设备管理视图

2) 安装业务管理

该部分支持外部合作安装公司的资质与档案管理，并提供每日任务分配与待执行任务展示功能。同时，该模块详细记录设备安装现场的工单信息（人员、地点、设备状态等）。针对设备报废或跨实验室迁移场景，系统提供解绑信息台账，详细记录拆除操作与解绑原因，以确保仪器与终端物理关系的逻辑连续性。

创享仪器物联网传感器安装运维管理平台

退出 物拓

生产检测

资产录入

安装公司

库存信息

入库信息

出库信息

设备信息

设备状态

今日安装

维修信息

维修记录

问题列表

模型修改记录

用户管理

安装列表

安装统计

查询相关：

开始时间：

选择开始时间

结束时间：

选择结束时间

选择学校：

全部

选择场地：

请选择场地

资产编码：

请输入资产编码

盒子编码：

请输入盒子编码

imei码：

请输入imei码

入库人员：

请选择入库人员

安装公司：

请选择安装公司

领取人员：

请选择领取人员

安装人员：

请选择安装人员

设备类型：

请选择设备类型

是否有故障：

是否有故障

电流计在线状态：

电流计在线状态

资产开机状态：

资产开机状态

资产属性状态：

资产属性状态

资产上报：

资产上报

安装类型：

安装类型

费用类型：

费用类型

资产价值：

资产价值

复制

删除安装

导出数据

序号	设备imei码	资产编号	资产名称	gsiCode	学校	使用管理部门	存放地址	设备类型	领取人	安装公司	安装类型	安装人	安装时间	电流计在线状态	资产开机状态	资产属性	资产云上报	上线次数	操作	
1	869658032094774	2025125916	多功能酶标仪	4700518570000000760510	杭州医学院	科研实验中心	杭州医学院黄友校区(天目山路182号)1号楼13楼1301室	单桩	孙勃	浙江创享	新装	钟乃飞	2026-02-04 10:56:58	在线	关机	未知	未上报	0	详情	
2	866497066692345	2025074116	蓝装液相色谱仪		杭州医学院	药学院	杭州医学院黄友校区(天目山路182号)1号楼1013室	单桩		浙江创享	浙江创享	新装	冷广洋	2026-02-04 10:49:36	离线	关机	未知	未上报	0	详情
3	866028076587819	2025125616	蛋白质谱分析系统	4700518570000000760503	杭州医学院	科研实验中心	杭州医学院黄友校区(天目山路182号)1号楼	单桩		浙江创享	浙江创享	新装	钟乃飞	2026-02-04 10:43:50	在线	关机	未知	未上报	0	详情

< 1 2 3 4 5 6 ... 934 >

前往 1 页 15条/页 共 14008 条

图 2.26 安装业务视图

3) 运维与异常管理

该部分除提供设备状态（在线、离线、故障）实时监控列表外，还支持一线巡检人员日常巡检记录的电子化填报与问题工单的流转追踪。同时，该模块设置了模型修改记录子功能，用于追溯底层异常诊断模型或阈值配置的变更历史，为算法的持续演进提供追溯依据。

创享仪器物联网传感器安装运维管理平台

在线设备 离线设备 故障设备

查询相关: 开始时间: 选择开始时间 结束时间: 选择结束时间 选择学校: 全部 安装公司: 请选择安装公司 安装人员: 请选择安装人员 设备类型: 请选择设备类型

新增 重置查询

序号	设备IMEI码	资产编号	资产名称	使用管理部门	存放地址	设备类型	安装人	安装时间
1	86958632094774	2025129816	多功能检测仪	科研实验中心	杭州医学院莫龙校区(天目山路182号)1号楼13楼1301室	单桩	钟乃飞	2026-02-04 10:56:58
2	866028670587829	2025125616	蛋白质组分析系统	科研实验中心	杭州医学院莫龙校区(天目山路182号)1号楼5楼503室	单桩	钟乃飞	2026-02-04 10:43:50
3	866028670580697	2025322916	蛋白质组分析系统	科研实验中心	杭州医学院莫龙校区(天目山路182号)1号楼5楼502室	单桩	钟乃飞	2026-02-04 10:39:44
4	860275070780653	2024271116	多角度粒度与流式联用 Zeta电位分析仪	科研实验中心	杭州医学院莫龙校区19号楼506室	单桩	钟乃飞	2026-02-04 10:09:51
5	860275070772106	2022399800	流式细胞仪	0221004重点实验室	天目山路182号(707495)809	单桩	潘广泽	2026-02-04 10:09:42
6	860275070777915	2020524100	超高压液相色谱三重四级杆质谱	杭州师范大学实验室与设备管理中心	杭州师范大学仓前校区德园15号楼402室	单桩	潘广泽	2026-01-29 13:32:25
7	866028670580848	20250762	激光打标机	中心区: 芯片化量子传感技术创新中心	象牙楼303	单桩	钟乃飞	2026-01-29 10:43:06
8	860275070780471	20251719	3D打印机	中心区: 芯片化量子传感技术创新中心	白马湖5-801	单桩	钟乃飞	2026-01-29 10:10:38
9	866028670581747	20251685	3D打印机	中心区: 芯片化量子传感技术创新中心	白马湖5-304	单桩	钟乃飞	2026-01-29 09:59:47
10	866028670575137	20252983	恒电声腔体大器	中心区: 芯片化量子传感技术创新中心	白马湖5-202	单桩	钟乃飞	2026-01-29 09:43:21

1 2 3 4 5 6 ... 624 前往 1 页 15条/页 共 9348 条

(a) 设备状态监控列表

创享仪器物联网传感器安装运维管理平台

添加问题

选择学校: 全部 资产编号: 选择资产编号 问题类型: 选择问题类型

备注:

设备插入照片:

设备故障照片:

取消 确定

(b) 添加问题

创享仪器物联网传感器安装运维管理平台

查询相关: 选择月份: 选择月份 选择学校: 全部 选择场地: 选择场地 资产编号: 选择资产编号 imei码: 选择imeimei码

安装公司: 请选择安装公司 巡检人员: 请选择巡检人员 设备类型: 请选择设备类型 故障类型: 请选择故障类型 巡检状态: 请选择巡检状态

新增 重置查询 导出数据

序号	设备IMEI码	资产编号	资产名称	学校	使用管理部门	存放地址	设备类型	总价	故障类型	联系人	联系方式	巡检状态	巡检人员	巡检时间	备注
1	866028670578966	2024059712	元素分析仪	浙江大学	海洋科学与技术学院	(860104)海科楼104-1	单桩	43.90	仪器用电, 流量计断电	马伟伟	17864272997	巡检通过	凌广泽	2025-12-18 13:48:37	
2	864823041517079	201801040					三相		其它			巡检通过	凌广泽	2025-08-26 9:42:30	有使用记录, 但没有照片
3	864823041531518	TY2018000201	高光谱成像仪	中国农业科学院茶叶研究所(农业农村部茶叶所茶业质量监督测试中心)	加工中心	加工中心201	单桩	32.75	仪器用电, 流量计断电	李佳	0571-864050637	巡检通过	凌广泽	2025-03-27 4:59:05	
4	866497056668482	20227868	时间分辨光子成像测量系统	北京航空航天大学	先进飞行研究中心	先进飞行研究中心	单桩	144.80	其它	董庆飞	13426315077	巡检通过	凌广泽	2025-03-13 9:44:09	实际未使用
5	863430069132258	52314711	高通量自动化核酸提取平台	之江实验室	科研条件管理部	南湖1418区-11号楼-C102-一层(17)基础实验室	单桩	156.19	仪器用电, 流量计断电	张鹏	0571-98005097	巡检通过	凌广泽	2025-02-19 1:17:28	
6	863430069135230	52302353	紫外扫描系统	之江实验室	科研条件管理部	4号楼B座205	单桩	46.50	仪器用电, 流量计断电	孙艺璇	18069860089	巡检通过	凌广泽	2025-02-19 0:14:34	
7	869858033749301	2014048500	半制备高光谱仪	浙江大学	食品科学与工程学院	(11330A)机械楼300-1	单桩	66.27		魏高深	13958119342	待巡检			
8	869858033749763	2018789501	纤维轴10光学	浙江大学	分析测试中心	浙江杭州西湖区388号浙工大(之江校区)	单桩	30.00		李纪云	13588407011	待巡检			

1 2 3 4 5 6 ... 32671 前往 1 页 15条/页 共 190055 条

(c) 巡检记录

序号	ID代码	资产编码	资产名称	资产类型	学校	维护管理部门	维护原因	维护时间	是否确认	操作
221	866028076586200	ZDWYJC00729	微量紫外光谱仪	光谱仪器	浙江大学温州研究院	基础部理化与材料研究中心	老样反馈调整仪器模型	2025-02-14 19:01:08	未确认	确认处理
222	86649706662600	ZDWYJC00263	飞秒激光系统	光谱仪器	浙江大学温州研究院	新材料创新创业中心	老样反馈调整仪器模型	2025-02-14 18:48:05	未确认	确认处理
223	866497062500898	ZDWYJC001560	多功能X射线衍射仪	X射线仪器	浙江大学温州研究院	新材料创新创业中心	老样反馈调整仪器模型, 修正其机头温度	2025-02-14 18:36:23	未确认	确认处理
224	866497062510251	ZDWYJC001487	紫外可见红外分光光度计	光谱仪器	浙江大学温州研究院	新材料创新创业中心	老样反馈调整仪器模型	2025-02-14 18:27:03	未确认	确认处理
225	863450069132852	ZDWYJC001459	气相色谱三通道气相色谱仪	色谱仪器	浙江大学温州研究院	新材料创新创业中心	老样反馈调整仪器模型	2025-02-14 18:07:46	未确认	确认处理
226	866497062510889	ZDWYJC00196	高速离心分离瓶离心机	光谱仪器	浙江大学温州研究院	新材料创新创业中心	老样反馈调整仪器模型, 该设备1.5A电流为待机	2025-02-14 17:43:23	未确认	确认处理
227	866497062502987	ZDWYJC00161	场发射扫描电子显微镜	电子光学仪器	浙江大学温州研究院	新材料创新创业中心	老样反馈调整仪器模型, 该设备关机状态下电压偏低	2025-02-14 17:31:55	未确认	确认处理
228	866497062502987	ZDWYJC00161	场发射扫描电子显微镜	电子光学仪器	浙江大学温州研究院	新材料创新创业中心	老样反馈调整仪器模型, 该设备关机状态下电压偏低	2025-02-14 17:25:59	未确认	确认处理
229	866028076578917	A02102024080055	微光光谱仪	光谱仪及图像分析仪器	浙江中医药大学金宇研究院	细胞子平台	老样反馈小电流为工作状态	2025-02-14 17:06:51	未确认	确认处理
230	866028076578917	A02102024080055	微光光谱仪	光谱仪及图像分析仪器	浙江中医药大学金宇研究院	细胞子平台	老样反馈0.05A为工作状态	2025-02-14 17:06:51	未确认	确认处理

(d) 模型修改记录列表

图 2.27 运维与异常管理视图

4) 系统权限管理

该部分为平台的基础支撑模块，负责构建“省-校-院系-实验室”的多级树状组织架构。超级管理员可在此功能中统一管理系统用户的账号、角色及权限菜单，从而实现省级用户全局监管、校级用户管辖本校、运维专员定点操作的差异化赋权机制。

序号	用户账号	名称	单位	操作
1				修改 分配单位 删除
2				修改 分配单位 删除
3				修改 分配单位 删除
4				修改 分配单位 删除
5				修改 分配单位 删除
6				修改 分配单位 删除
7				修改 分配单位 删除
8				修改 分配单位 删除
9				修改 分配单位 删除

图 2.28 权限管理视图

(3) 时序波形与流形拓扑展示视窗

该视图是实现深度学习算法结果可视化的重要模块。当用户查看单台设备时，前端不仅通过时间轴加载 MongoDB 中的电流波形，还会调用 ECharts 的散点图组件，读取后台 Python 算法生成的二维坐标。将高维抽象的运行特征降维展平为状态地图，使正常工作簇、待机簇与异常游离点在空间中一目了然，降低用户对黑盒模型的理解门槛。

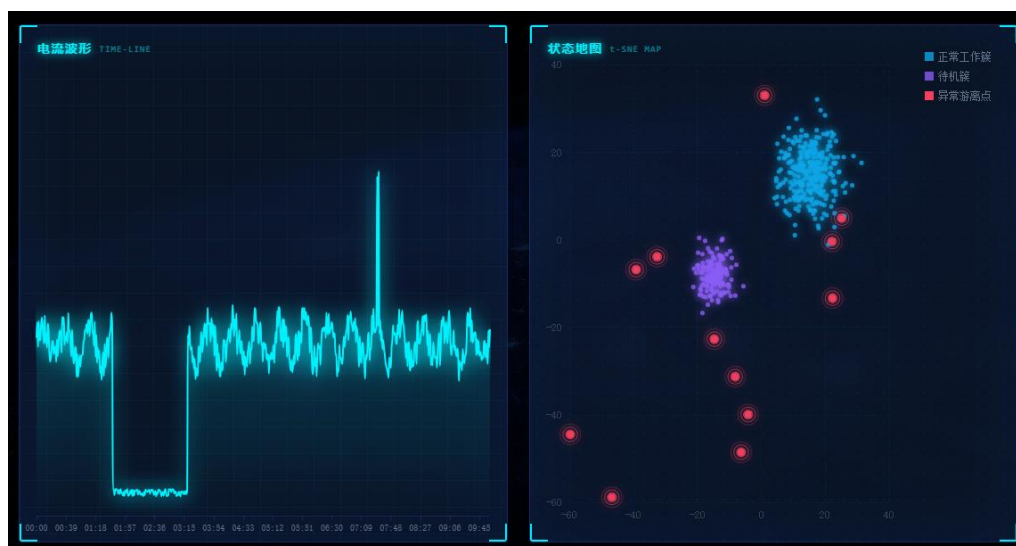


图 2.29 时序波形与流形拓扑渲染视窗

(4) 人机协同确权与知识映射模块

该模块直接对接 Python FastAPI 提供的微服务。针对算法判定为灰区的疑似异常，界面会自动生成待办卡片。领域专家可在此视窗内结合波形与状态地图进行人工判别。一旦完成确认，专家可通过下拉表单为该数据段标上具体的物理语义标签并提交。这一交互过程将专家反馈同步写入底层动态故障知识库，从而实现由机器预警到专家修正的闭环更新。



图 2.30 人机协同确权与知识映射模块视图

2.5.2 微信小程序端实现

为满足不同业务角色在移动办公场景下的使用需求，本系统设计了运行小程序与运维小程序两类微信小程序。两端基于统一的后端数据与 RESTful API 构建，但在交互界面与业务流程上实现了严格的物理与逻辑隔离。

(1) 运行小程序

运行小程序主要面向科研人员与实验室管理者，提供轻量化的设备状态掌上看板。用户

可通过该端便捷查看单台仪器的历史电流波形与对应运行状态，并按照自定义时间段进行机时统计分析。此外，考虑到大型科学仪器价值高昂，该端还集成了资产保障相关流程。在设备发生故障或损毁时，用户可实现从问题反馈、一键报案对接保险到理赔进度查询的完整业务流程。



图 2.31 运行小程序视图

(2) 运维小程序

运维小程序主要面向一线实施工程师与设备巡检专员，用于物联网终端的现场部署与全生命周期维护。实施人员无需携带电脑，可通过调用微信原生扫码功能读取终端编号，并一键完成科研仪器与监测终端的物理关系绑定与解绑操作。同时，该端以电子化巡检任务替代传统纸质巡检单，支持现场工况核验与拍照上传，从而提高线下巡检作业与线上数据流转的一致性，并有助于提升前端感知硬件的在线稳定性和数据采集质量。



图 2.32 运维小程序视图

2.6 本章小结

本章从工程实现的角度出发，详细阐述了科学仪器运行在线监测平台的整体架构与系统

设计方案。针对科研仪器机理不透明和物理接入受限的问题，本文设计了基于 NB-IoT 的非侵入式电流监测终端，实现了底层设备运行状态的稳定采集。面对物联网感知层产生的海量数据，系统设计了由 MySQL、MongoDB 和 Redis 联合构成的混合持久层，有效解决了高频时序波形、半结构化模型特征以及强一致性资产数据的高效协同存储问题。为协调业务流转与算法推理之间的性能需求，平台进一步构建了 Java 业务中台与 Python 算法服务协同运行的异构计算架构。通过基于数据库状态轮询的异步机制，实现了跨语言模块之间的解耦，提升了平台在海量设备并发监测场景下的吞吐能力与可用性。此外，系统构建了 Web 端和小程序的交互展示层，不仅满足了多层级用户的全局监控与移动运维需求，也为后续人机协同机制提供了界面基础。综上所述，本章所构建的系统架构不仅实现了科学仪器在线监测的完整业务流程，也为第三章和第四章中复杂深度学习算法的运行、验证与演化提供了可靠的工程基础与数据支撑。

3 基于质量感知对比学习的仪器状态辨识与异常检测方法

本章针对科研仪器分钟级采样电流数据在状态辨识与异常检测中存在的采样率低、数据质量不均以及频域信息缺失等问题，提出一种基于质量感知的时频双视图对比学习（Quality-Aware Time-Frequency Contrastive Learning, QA-TF-CL）监测方法。该方法在不依赖高采样率假设的前提下，通过引入等效频域统计特征和质量约束机制，实现了对仪器运行状态的鲁棒表征学习，是本文的核心创新点之一。

3.1 问题描述与总体架构

3.1.1 问题描述

在实际科研仪器运行监测中，受限于通信带宽和长期数据存储成本，仪器电流通常按分钟级时间进行采样，其采样频率可表示为

$$F_s \approx \frac{1}{60} \text{ Hz} \quad (3.1)$$

在这种采样条件下，电流数据失去了高频波形细节，更多反映的是仪器运行过程中的低频趋势变化和状态切换行为。此外，受实验室复杂电磁环境干扰及无线传输丢包影响，原始时序中常含有随机离群点与数据缺失。

基于上述现实约束，本章的研究重点并非对电流信号进行精细波形建模，而是在无监督条件下，从包含噪声和不完整信息的低频趋势数据中学习能够稳定刻画科研仪器工作状态的鲁棒特征表示，并在此基础上实现仪器运行状态的自动聚类与异常监测。

具体而言，设一段长度为 T 的科研仪器电流时间序列为：

$$X = \{x_1, x_2, \dots, x_T\} \quad (3.2)$$

针对采样过程中的非理想因素，引入数据质量掩码 M ，并构建如下映射函数：

$$f_\theta(X, M) \rightarrow Z \quad (3.3)$$

利用数据质量掩码 M ，模型从含噪数据中提取具有物理可解释性的高维状态表征 Z ，并使得到的特征表示 Z 在特征空间中呈现出良好的结构性：一方面，同一工作状态下的特征应尽可能聚集；另一方面，当仪器出现异常或运行模式发生明显变化时，其特征应与正常状态保持足够的区分度。在此鲁棒表征基础上，后续的运行状态聚类与异常检测任务无需显式依赖标签信息，可通过无监督或弱监督方式实现，从而满足科研仪器长期自动化健康监测的实际需求。

3.1.2 总体架构

本节介绍 QA-TF-CL 方法的整体技术路线，如图 3.1 所示。

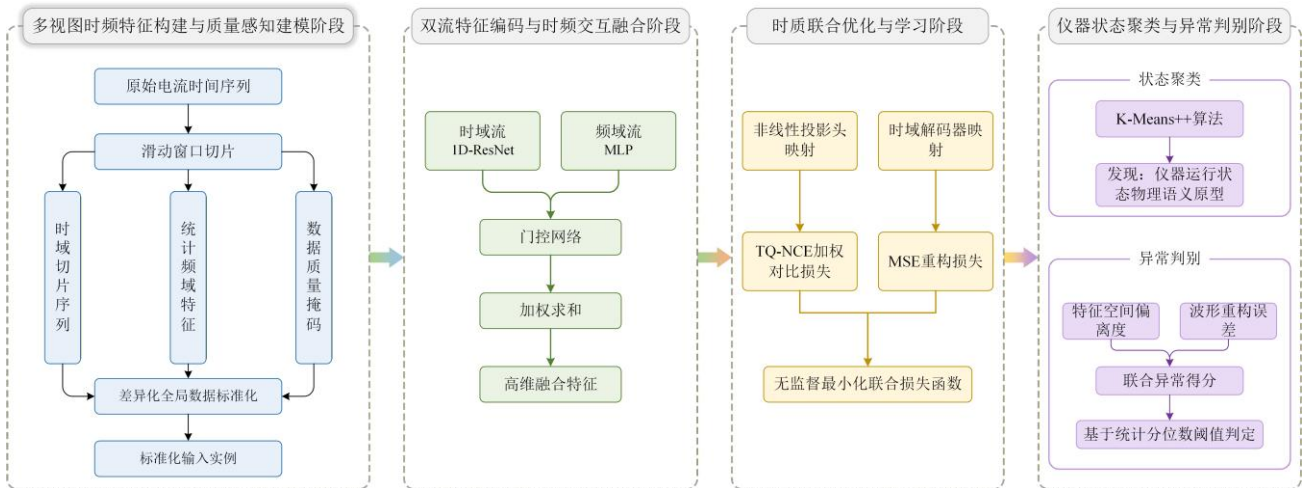


图 3.1 技术路线图

该方法主要由以下四个阶段组成。

(1) 在输入端对原始电流时间序列进行多视图特征构建。一方面，通过滑动窗口方式生成时域切片，用于描述电流随时间变化的动态过程；另一方面，在相同时间窗口内计算一组统计特征，构成统计频域视图，用以刻画电流波动幅度和变化节律。此外，为了评估低频数据的可靠性，系统会在同一时间窗口内同步生成数据质量掩码。由此，每个时间窗口不仅对应一对时频互补的特征表示，还具备了一个量化数据质量的指标。

(2) 引入双流编码器结构。针对时域序列和统计频域特征的不同特点，分别设计时域编码器和频域编码器，对两类特征进行独立建模，得到时域和频域的隐空间表示。在此阶段，数据质量掩码并不直接参与前向传播与特征计算，以免在特征提取层引入额外的干扰。它将被作为一种软约束，被引入到后续的优化阶段。双流特征输出后，进入门控网络进行交互融合，并通过自适应的加权求和，最终输出统一的高维融合特征。

(3) 时质联合优化与学习阶段。特征学习采用双分支的多任务联合优化策略。在对比学习分支中，融合特征经过非线性投影头后，结合数据质量掩码计算 TQ-NCE 加权对比损失；在重构分支中，融合特征输入时域解码器，计算均方误差重构损失。系统最终以无监督方式最小化联合损失函数，从而实现鲁棒特征的提取。

(4) 状态辨识与异常判别阶段。基于学习到的鲁棒表征，系统利用无监督聚类实现工况的自动化辨识，并进一步计算特征空间偏离度与波形重构误差，将二者融合为联合异常得分，最后基于统计分位数确定的阈值进行最终的异常判定与预警。

3.2 时频特征构建与质量感知建模

在分钟级采样条件下，科研仪器电流数据所能提供的信息，与传统电能分析场景存在明显差异。此时，信号中已经不再包含可解析的工频谐波结构，若仍然沿用傅里叶变换等经典频域分析手段，不仅难以得到有意义的频谱信息，反而可能引入额外噪声干扰。

基于这一现实约束，本节从工程实践角度出发，重新构建低采样率条件下的频域信息表

达方式，并进一步引入数据质量感知机制，为后续特征学习提供可靠支撑。

3.2.1 统计频域特征构建

由于采样率较低，传统基于傅里叶变换（Fast Fourier Transform, FFT）的精确频谱分析不仅会因奈奎斯特采样定理受限而产生严重混叠，而且在短窗口条件下也难以保持明确的物理意义。本研究通过引入统计频域这一概念，用以描述电流信号在时间窗口内的波动模式与变化节律。

根据帕塞瓦尔定理（Parseval's Theorem）与随机信号处理理论，时域的特定统计矩与频域的功率谱密度分布存在着内在的能量等效关系。因此，本文对于任意时间窗口 $X = \{x_1, x_2, \dots, x_N\}$ ，构建统计特征向量 $X_{freq} \in \mathbb{R}^D$ （ $D=3$ ），主要包括以下三类指标：

（1）波动能量。该指标通过计算窗口内电流的标准差来量化负载的动态稳定性：

$$f_{std} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (3.4)$$

其中， x_i 为窗口内的第 i 个电流采样点， μ 为该窗口内的电流均值， N 为窗口长度。在物理层面，波动能量反映了设备运行过程中的能量离散程度。在稳态运行或待机工况下，电流波动较小，该值趋近于 0；而在加热、变速离心等非平稳工况下，该值会明显增大，可作为识别复杂工况的重要依据。

（2）趋势强度。为捕捉电流演化的动态速率，本算法引入一阶差分的平均绝对值来表征趋势强度：

$$f_{diff} = \frac{1}{N-1} \sum_{i=1}^{N-1} |x_{i+1} - x_i| \quad (3.5)$$

该指标对状态切换瞬态较为敏感。它能够有效区分稳态高功耗工况与频繁状态调整过程，为模型提供了关于状态迁移边界的重要信息。

（3）极值分布特征。为了进一步辨识信号中的脉冲冲击特征，本文引入峰值因子来刻画电流极值的分布特性：

$$C_{factor} = \frac{\max(|x_i|)}{X_{rms}} \quad (3.6)$$

其中， X_{rms} 为窗口内的均方根值。峰值因子能够量化波峰相对于有效值的极端程度，从而在特征空间中实现瞬态冲击负荷与持续性平稳负荷的解耦。

综上所述，虽然这些指标并未直接映射至特定的频率分量，但在低采样环境下，它们共同构成了电流信号的统计频域特征。这种多维统计归约方式弥补了时域原始序列在整体变化模式表征上的不足，并为后续双流编码器的隐空间建模提供了互补信息。

3.2.2 数据质量掩码生成

在实际工程环境中，科研仪器监测系统常处于复杂的电磁条件与网络环境中，电流数据在采集与传输过程中不可避免地受到网络波动、网关拥塞和传感器漂移等因素的影响。这导致原始数据中常混杂着数据丢包、插值伪影以及非平稳噪声。若直接将此类低质量数据输入模型，容易导致模型在训练阶段拟合噪声分布，从而降低状态辨识的鲁棒性。

基于此，本文提出一种显式数据质量建模策略，通过引入数据质量掩码来量化每个时间窗口的数据可靠性。本文定义第 t 个时间窗口的质量置信度评分 m_t 如下：

$$m_t = \lambda_1 \cdot (1 - R_{miss}) + \lambda_2 \cdot \mathcal{F}_{norm}(SNR_{est}) \quad (3.7)$$

其中，数值 $m_t \in [0,1]$ 越接近 1 表示该窗口数据质量越高，越能真实反映仪器运行状态。 λ_1, λ_2 为平衡系数，满足 $\lambda_1 + \lambda_2 = 1$ （本文实验中均取 0.5）。具体指标构建如下：

（1）丢包率惩罚项(R)

该项用于评估数据的完整性。对于长度为 N 的时间窗口，若其中实际采集到的有效点数为 N_{valid} ，其余 $N - N_{valid}$ 个点均为通过线性插值或零阶保持生成的填充点，则丢包率定义为：

$$R_{miss} = \frac{N - N_{valid}}{N} \quad (3.8)$$

插值点占比越高，数据包含的真实信息越少，因此该项对质量评分 m_t 的贡献越低。

（2）归一化信噪比估计项(SNR)

该项用于评估数据的纯净度。由于低频采样数据无法直接通过频谱分析获取噪声基底，本文采用趋势-残差分解的方法估计信噪比。

首先，利用 Savitzky-Golay 滤波器^[53]对窗口内序列 x 进行平滑处理，得到趋势项 x_{trend} ，则残差项为 $x_{noise} = x - x_{trend}$ 。信噪比估计值 S 的计算如下：

$$S = 10 \cdot \log_{10} \left(\frac{\sum x_{trend}^2}{\sum x_{noise}^2 + \epsilon} \right) \quad (3.9)$$

其中 ϵ 为防止分母为零的微小量。为了将 SNR 映射到 $[0,1]$ 区间并参与加权计算，系统采用 Sigmoid 函数进行归一化处理：

$$SNR = \frac{1}{1 + e^{-(S - \mu_{snr})/\sigma_{snr}}} \quad (3.10)$$

其中， μ_{snr} 和 σ_{snr} 分别为历史数据集中信噪比的均值与标准差。

生成的数据质量掩码 m_t 并非用于简单的阈值剔除，而是作为一种软约束参与到后续的对比学习网络中。它将以损失函数权重的形式参与梯度计算：在计算对比损失时，质量低的样本对所产生的梯度会被 m_t 衰减，从而使得模型训练过程更加关注高质量的、具有明确物理特征的样本。

3.2.3 多视图数据的标准化与输入实例构建

经过上述特征提取与质量评估步骤后，对于每一个时间窗口，均可得到原始时域序列 X_{time} 、

统计频域特征 X_{freq} 以及数据质量掩码 m_t 。由于时域电流数据与各类统计特征之间存在明显的量纲差异,若直接将其输入神经网络,不同数量级的特征可能导致模型收敛困难或梯度震荡。为此,本文针对不同类型的数据设计了差异化的标准化策略。为了防止数据泄露,所有标准化统计量(均值 μ 、标准差 σ)均仅基于训练集计算,并固定用于验证集与测试集的预处理。

(1) 时域序列

对于时域电流序列,传统的滑动窗口级标准化会削弱数据的绝对幅值信息,导致 0.5A 待机状态与 10A 满载状态在归一化后波形趋同,从而丢失了判别工况强度的关键信息。为了解决这一问题,本文采用全局标量标准化策略。计算训练集整体数据的全局均值 μ_{train} 与全局标准差 σ_{train} , 对任意时间窗口 t 的时域序列进行统一缩放:

$$x_{time}^{(t)} = \frac{x_{time}^{(t)} - \mu_{train}}{\sigma_{train} + \epsilon} \quad (3.11)$$

其中 ϵ 为防止除零而引入的微小量。该方法在将数据映射到统一尺度空间的同时,保留了不同运行状态下的绝对电流强度差异,使模型能够感知负载大小。

(2) 统计频域

统计频域特征向量 $X_{freq}^{(t)}$ 由多类统计指标构成,各维度的数值分布差异较大,因此不能共享同一组统计量。为此,本文采用逐维度独立标准化策略。针对特征向量的第 d 个维度 ($d \in \{1,2,3\}$),分别统计其在训练集上的均值 $\mu_{freq}^{(d)}$ 与标准差 $\sigma_{freq}^{(d)}$,进行独立归一化:

$$x_{freq}^{(t,d)} = \frac{x_{freq}^{(t,d)} - \mu_{freq}^{(d)}}{\sigma_{freq}^{(d)} + \epsilon} \quad (3.12)$$

该操作确保了后续编码器能够均衡地提取波动能量、趋势强度与极值分布信息,避免低数值特征在融合过程中被弱化。

(3) 数据质量掩码

数据质量掩码本质上表示样本的置信度,其原始定义域为 $[0,1]$ 。由于 m_t 将直接作为后续损失函数的加权系数,因此必须保证其非负性。若对其进行 Z-Score 标准化,部分低质量样本的权重可能会变为负数,进而破坏损失函数的优化方向。因此,本文不对 m_t 进行标准化处理,仅在工程实现中进行数值截断,以消除计算机浮点运算可能产生的微小精度误差,从而保证数据质量掩码始终保持非负,并维持损失函数的正确优化方向:

$$m_t = \text{Clip}(m_t, 0, 1) \quad (3.13)$$

(4) 标准输入实例构建

经过上述步骤,第 t 个时间窗口被封装为标准输入实例 S_t :

$$S_t = \{\tilde{X}_{time}^{(t)}, \tilde{X}_{freq}^{(t)}, m_t\} \quad (3.14)$$

其中, $\tilde{X}_{time}^{(t)}$ 表示标准化后的时域分量, $\tilde{X}_{freq}^{(t)}$ 表示标准化后的统计频域分量, m_t 表示该窗口对应的质量权重分量。

该标准化实例 S_t 实现了时域信息、频域统计规律与数据质量先验的统一表达，为下一节中双流编码器的特征提取与加权对比学习提供了规范的输入基础。

3.3 质量感知的双视图对比学习模型

在上节中，本文完成了低采样率电流的时域切片、统计频域特征构建以及数据质量建模，并将它们整合为标准化输入实例 $S_t = \{\tilde{X}_{time}^{(t)}, \tilde{X}_{freq}^{(t)}, m_t\}$ 。接下来，本节介绍如何利用这些多视图数据，在无监督条件下通过对比学习获得鲁棒的状态表征。

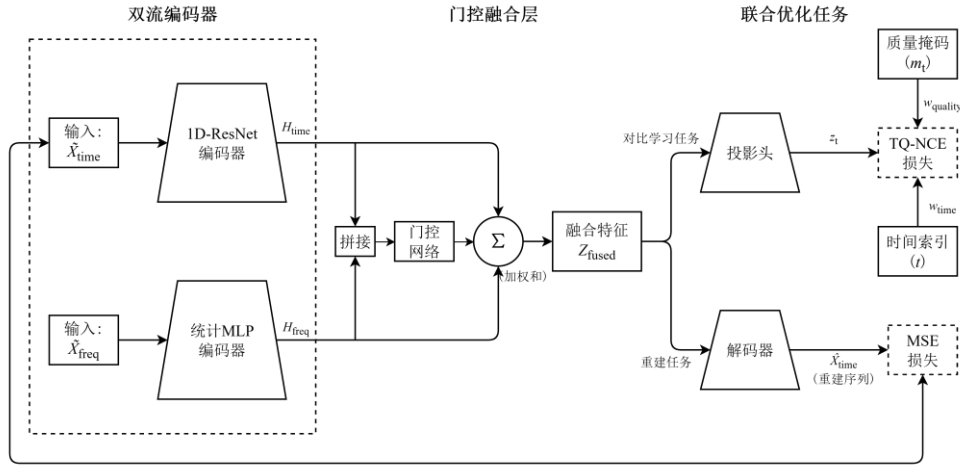


图 3.2 双视图对比学习总体架构

3.3.1 双流编码器设计

对于标准化后的时域序列 $\tilde{X}_{time}$ 与统计频域特征 $\tilde{X}_{freq}$ ，首先需明确二者在信息结构与语义属性上的差异：时域信号 $\tilde{X}_{time}^{(t)}$ 是高维、连续且具有较强的时序依赖性的过程量，其主要作用在于刻画电流随时间演化的动态行为；而频域特征 $\tilde{X}_{freq}^{(t)}$ 则是经过人工设计并高度压缩的状态量，虽维度低且不具备时序性，但蕴含了对运行模式具有判别力的全局统计规律。

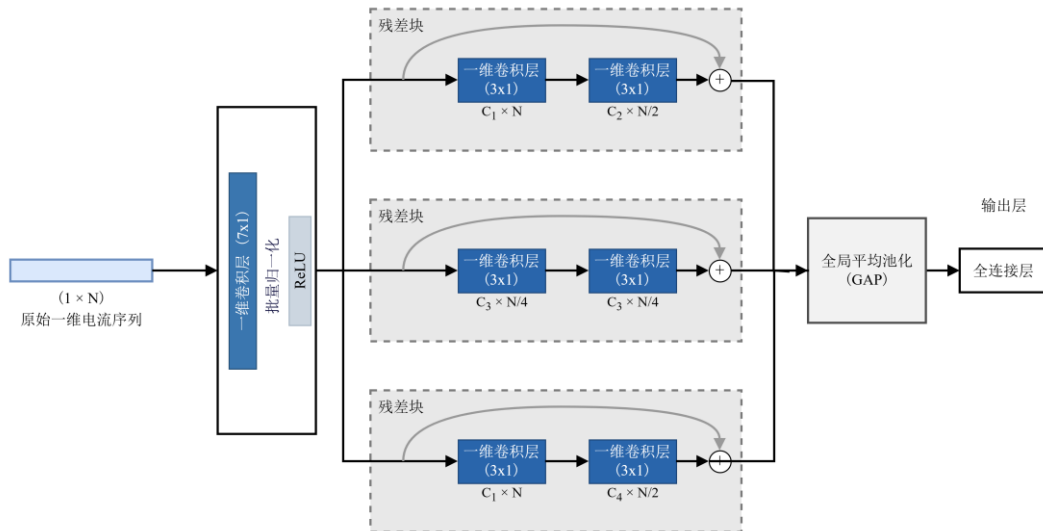


图 3.3 1D-ResNet 编码器结构图

鉴于两类输入在数据形态与信息密度上的不同，若将其统一输入同一特征提取器，不仅难以兼顾各自的建模需求，还可能因归纳偏置错配而损害表征质量。为此，本文提出一种双流编码策略，为时域与频域信息分别构建专用的特征提取路径。

在时域分支中，核心目标是捕获电流序列中的长期趋势与局部瞬态模式。给定长度为 N 的标准化时域序列 $\tilde{X}_{\text{time}}^{(t)}$ ，系统采用一维残差网络（1D-ResNet）作为骨干编码器^[54]。如图 3.3 所示，该网络通过堆叠卷积残差块，有效建模多尺度时间依赖关系，并缓解深层网络中的梯度消失问题。该向量旨在紧凑地编码仪器在时刻 t 前后所表现出的动态运行节奏。

最终，时域流输出隐向量 $H_{\text{time}}^{(t)}$ ：

$$H_{\text{time}}^{(t)} = \text{ResNet}(\tilde{X}_{\text{time}}^{(t)}) \quad (3.15)$$

相比之下，频域分支的处理逻辑更为简洁。由于 $\tilde{X}_{\text{freq}}^{(t)}$ 已是经过去时序化处理的统计摘要，不具备序列结构，使用递归或卷积结构反而容易引入不必要的参数冗余与过拟合风险。因此，仅采用一个轻量级多层感知机（Multi-Layer Perceptron, MLP）作为频域编码器^[55]：

$$H_{\text{freq}}^{(t)} = \text{MLP}(\tilde{X}_{\text{freq}}^{(t)}) \quad (3.16)$$

该 MLP 通过非线性变换将离散统计指标映射至与 $H_{\text{time}}^{(t)}$ 同维的语义空间，从而为后续融合奠定基础。

3.3.2 时频交互门控融合

经过双流编码器的处理，系统分别获得了时域特征 $H_{\text{time}}^{(t)}$ 与频域特征 $H_{\text{freq}}^{(t)}$ 的高维表示。在分钟级监测场景下，两类模态对状态判别的贡献会随运行状态动态变化：在仪器启停或状态切换的非平稳瞬态，时域序列的演化趋势具有更高的决策权重；而在长时间稳态运行阶段，统计频域指标则能更稳定地反映设备健康度。

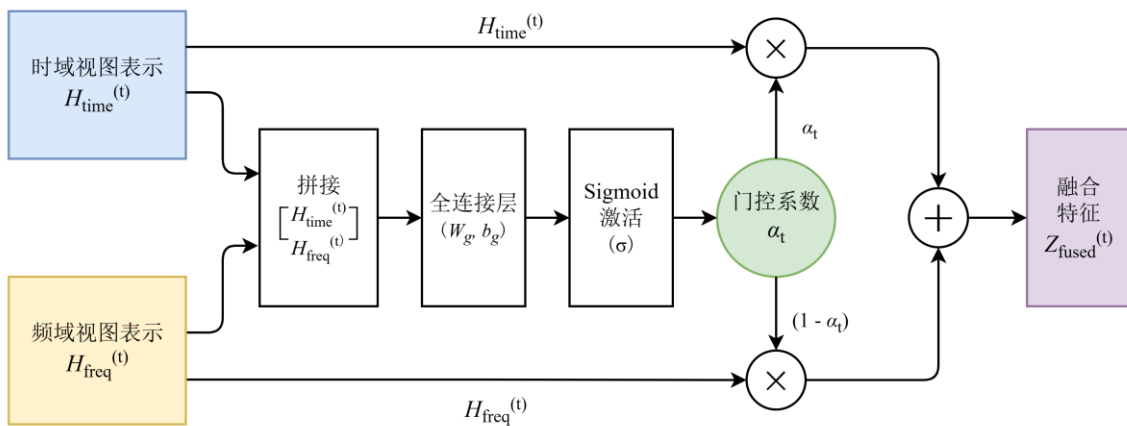


图 3.4 时频自适应门控融合模块结构图

为了应对这一动态特性，系统引入可学习的门控融合机制，如图 3.4 所示，使网络能够根据不同运行状态自适应调整时域与频域特征的重要性。

首先，将两个流的特征进行拼接：

$$\left[H_{time}^{(t)} \parallel H_{freq}^{(t)} \right] \quad (3.17)$$

其中, $[\bullet \parallel \bullet]$ 表示向量拼接操作。随后, 通过一个全连接层与逐元素 Sigmoid 激活函数生成动态门控系数:

$$\alpha_t = \sigma \left(W_g \left[H_{time}^{(t)} \parallel H_{freq}^{(t)} \right] + b_g \right) \in [0, 1] \quad (3.18)$$

其中, W_g 和 b_g 为可学习参数; $\sigma(\cdot)$ 表示逐元素 Sigmoid 函数。最终的融合特征通过加权求和得到:

$$Z_{fused}^{(t)} = \alpha_t \odot H_{time}^{(t)} + (1 - \alpha_t) \odot H_{freq}^{(t)} \quad (3.19)$$

其中, $\odot$ 表示按元素乘。当 α_t 某些维度接近 1 时, 模型认为这些维度的判别依据主要来源于时域特征; 反之, 当 α_t 接近 0 时, 模型则更依赖频域特征。通过这种自适应加权, 网络能够在不同状态下自动聚焦于更具判别力的特征视图, 从而获得更有效的状态表征。

3.3.3 时质联合加权对比损失函数

通过双流编码器与门控融合, 已获得了时域与频域的高维融合特征, 但如何在无标签条件下进一步学习到稳定且具有判别力的状态表征, 仍是该方法的关键问题。本研究采用对比学习范式。该方法通过构建正负样本对, 使同一实例在不同增强视图下的表示在特征空间中保持接近, 同时增大不同实例之间的表示距离, 从而学习具有判别力的特征表示^[56]。从信息论角度看, 对比学习可以理解为通过最大化不同视图之间的互信息来增强表示的一致性^[57]。近年来, 该思想已被广泛应用于时间序列自监督表示学习, 通过对时序片段或多视图信号施加对比约束, 模型能够学习到稳定的时序结构特征^[58]。

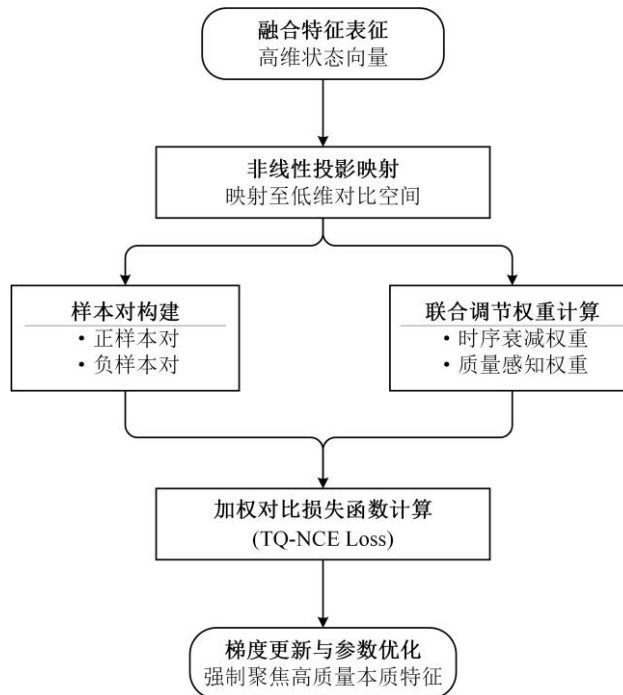


图 3.5 时质联合加权对比损失函数流程

然而，传统 InfoNCE 损失函数建立在独立同分布的假设上，认为所有非锚点样本均为等价的负样本^{[59][60]}。在科研仪器监测场景中，这一假设往往难以成立：在时序连续性方面，仪器运行状态具有惯性，时间上邻近的样本（如 t 与 $t \pm 1$ ）往往属于同一工况。若将邻近样本视为完全独立的负样本进行强行推远，会破坏特征空间中的时序拓扑结构。在数据质量方面，由于传输丢包和插值补全的存在，部分窗口中会包含较多伪影噪声。若仍要求模型对这类低质量样本进行拟合或区分，可能导致特征提取器学习到错误的噪声模式，而非真实的物理状态特征。

为此，本文提出一种时质联合加权对比损失函数（Time-Quality Joint Weighted TQ-NCE Loss）。该函数在特征空间中引入时间衰减与质量感知双重调节机制，引导模型关注高质量且时序独立的本质特征。其流程如上图 3.5 所示。

（1）投影空间映射与正负样本构建

为了确保梯度能够有效回传至门控融合模块并更新其参数 W_g 与 b_g ，系统将对对比约束直接作用于最终的融合特征 $Z_{fused}^{(t)}$ 上。为此，引入非线性投影头，将融合特征映射至超球面对比空间，以提升特征表示能力：

$$z_t = g\left(Z_{fused}^{(t)}\right) \quad (3.20)$$

其中，投影头 $g(\cdot)$ 由两层全连接层与 ReLU 激活函数构成。

对于当前训练批次中的任意一个原始输入窗口 S_t ，应用 3.3.5 节所述的两种随机数据增强变换，生成两个增强视图 $S_t^{(1)}$ 与 $S_t^{(2)}$ 。将这两个视图分别输入“双流编码器-门控融合-投影头”网络后，得到的输出向量 $z_t^{(1)}$ 与 $z_t^{(2)}$ 构成一对正样本，而由其他时间索引 k ($k \neq t$) 的输入窗口所生成的特征向量 z_k 则为负样本。

（2）时质联合权重定义

为缓解标准对比损失在处理时序数据时的局限性^[61]，本文定义了两类权重系数，分别从时序邻域和样本置信度两个维度对损失函数进行加权调节。

1) 时间衰减权重

由于时间序列样本间的相关性随时间距离的增加而衰减，为了避免将潜在的假负样本（即时间上极度接近、状态可能相同的样本）强行推远，引入时间衰减权重 w_{time} 。设 t 为当前锚点样本在原始序列中的绝对时间索引， k 为批次中作为对比对象的负样本的绝对时间索引。定义二者的时间相关性权重为：

$$w_{time}(t, k) = \exp(-\gamma \cdot |t - k|) \quad (3.21)$$

其中， $|t - k|$ 表示样本 k 与锚点 t 的时间步差值（单位分钟）； $\gamma > 0$ 为时间衰减系数，用于控制权重的衰减速率。当 $|t - k| \rightarrow 0$ 时， $w_{time} \rightarrow 1$ ，表明样本 k 极有可能与锚点属于同一工况，因而在损失函数中降低对其推远的惩罚力度。当 $|t - k| \rightarrow \infty$ 时， $w_{time} \rightarrow 0$ ，表明样本 k 是统计独立的负样本，应在特征空间中与锚点保持分离。

2) 质量感知权重

结合 3.2.2 节生成的数据质量掩码 m_t ，系统进一步从数据置信度的角度对损失进行约束。定义锚点 t 与对比样本 k 之间的成对质量权重 $w_{quality}$ 为二者掩码的几何平均：

$$w_{quality}(t, k) = \sqrt{m_t \cdot m_k} \quad (3.22)$$

该权重反映了样本对 (t, k) 的联合置信度。仅当两个样本均具备高质量 ($m \approx 1$) 时，该权重才接近 1，此时模型会更充分地优化该样本对之间的距离关系；若任一样本质量较低 ($m \rightarrow 0$)，则 $w_{quality} \rightarrow 0$ ，从而自动降低该样本对在梯度更新中的贡献。

(3) 联合加权 TQ-NCE 损失函数

结合上述两类权重，本文构建了 TQ-NCE 损失函数。对于批次 $\mathcal{B}$ 中的每个锚点样本 t ，其损失计算公式如下：

$$\mathcal{L}_{contra}^{(t)} = -\log \frac{\exp\left(\text{sim}\left(z_t^{(1)}, z_t^{(2)}\right) / \tau\right)}{\sum_{k \in \mathcal{B}, k \neq t} \left(\underbrace{(1 - w_{time}(t, k)) \cdot w_{quality}(t, k)}_{\text{联合调节因子 } \lambda_{tk}} \cdot \exp\left(\text{sim}\left(z_t^{(1)}, z_k^{(1)}\right) / \tau\right) \right)} \quad (3.23)$$

其中， $\text{sim}(u, v) = u^\top v / (\|u\| \|v\|)$ 表示余弦相似度， $\tau > 0$ 为温度参数，用于控制相似度分布的平滑程度。 λ_{tk} 为联合调节因子，用于动态调整负样本 k 在损失分母中的权重。

当样本 k 与锚点 t 时间间隔较远且数据质量较高时（即 $w_{time} \rightarrow 0, w_{quality} \rightarrow 1$ ），此时 $\lambda_{tk} \rightarrow 1$ ，损失函数退化为标准的 InfoNCE 形式，模型正常执行负样本推远操作。当样本 k 为邻近样本（ $w_{time} \rightarrow 1$ ）或低质量噪声（ $w_{quality} \rightarrow 0$ ）时，联合因子 $\lambda_{tk} \rightarrow 0$ 。该样本对在损失函数分母中的贡献趋近于零，从而显著减弱其在反向传播中的梯度影响，避免模型对同类状态施加过强分离约束，或对噪声模式进行过度拟合。

(4) 物理意义与作用机制

该损失函数在物理意义上引入了时间与质量两类联合调节机制，并通过对梯度贡献的动态控制实现鲁棒特征学习。

质量调节项由 $w_{quality}$ 控制。对于高质量的真实观测数据，该权重较高，从而保留其在反向传播中的梯度贡献；对于插值补全较多或噪声干扰严重的数据，该权重趋于减小，从而抑制其对特征空间学习的干扰。由此，模型能够更加关注可信样本中包含的物理规律。

时序调节项由 w_{time} 控制。对于时间上高度邻近的样本，该权重会自动减弱分离约束的强度，从而使特征在时间维度上保持相对平滑的过渡，并有助于维持仪器状态演化的局部拓扑一致性。

综上所述，借助本节提出的时质联合加权对比损失函数，QA-TF-CL 网络在低采样率、数据缺失与噪声混杂的科研仪器监测场景中，能够自适应地抑制干扰并保留时序结构，最终学习到兼具有判别力与物理一致性的鲁棒状态表征。

3.3.4 联合优化目标

单纯依赖对比学习虽然能提升特征的判别性，但可能导致模型过度关注某些区分度高的局部特征，而忽略了能够描述信号完整形态的细节信息。为了兼顾特征的判别能力与物理信息完整性，本文采用多任务联合优化策略^[62]。

为了引入物理约束，系统在门控融合模块之后引入了一个轻量级的时域解码器^[63]。该解码器由三层全连接层构成，旨在将高维融合特征映射回时间序列空间：

$$\hat{X}_{time}^{(t)} = \text{Decoder}\left(Z_{fused}^{(t)}\right) \in R^N \quad (3.24)$$

其中， $\hat{X}_{time}^{(t)}$ 表示模型对当前输入窗口的重构序列。

总损失函数由重构损失与对比损失加权组成：

$$\mathcal{L}_{total} = \lambda_{recon} \cdot \mathcal{L}_{recon} + \lambda_{contra} \cdot \mathcal{L}_{contra} \quad (3.25)$$

其中，重构损失 $\mathcal{L}_{recon}$ 采用均方误差（Mean Squared Error, MSE），用于约束重构序列 $\hat{X}_{time}^{(t)}$ 与原始标准化输入 $\tilde{X}_{time}^{(t)}$ 之间的一致性：

$$\mathcal{L}_{recon} = \frac{1}{|\mathcal{B}|} \sum_{t \in \mathcal{B}} \|\tilde{X}_{time}^{(t)} - \hat{X}_{time}^{(t)}\|_2^2 \quad (3.26)$$

该损失项作为正则化约束，促使融合特征 $Z_{fused}^{(t)}$ 保留足够的原始波形信息（如幅值、趋势），从而避免编码器在对比训练中丢失关键的物理信息。对比损失 $\mathcal{L}_{contra}$ 即 TQ-NCE 损失函数，其作用在于优化特征空间分布，拉近同类状态，推远异类状态，并自动抑制噪声样本的干扰。 λ_{recon} 与 λ_{contra} 为平衡超参数，用于调节两个任务的相对贡献。在实验中，设定 $\lambda_{recon} = 1.0$ 以保证基础表征能力，并设定 $\lambda_{contra} = 0.3$ 以引入判别性约束。若对比损失权重过高，如接近 1.0，模型可能过度捕捉噪声引起的波动而忽略基础物理分布，导致表征失真；若权重过低，如小于 0.1，则特征空间的判别力不足。

通过联合优化，模型在特征空间中实现了双重优化目标，从而兼顾了多工况状态分类与异常检测两类核心任务的需求。对比损失促使不同运行模式（如工作与待机）的样本在特征空间中相互推远，形成边界清晰的独立簇结构，显著提升了无监督聚类的准确性。重构损失则约束特征保留原始信号中的物理信息。当发生故障或异常波动时，模型难以利用所学习的正常模式对输入进行有效重构，从而表现出较大的重构误差或特征偏移，这有助于增强模型对异常状态的感知能力。

3.3.5 模型训练与实现细节

为了确保所提出的 QA-TF-CL 网络能够稳定收敛并具备良好的泛化能力，本文在训练与推理阶段设计了相应的策略与实现方案。

（1）信号特异性的数据增强策略

在对比学习范式中，正样本对的构建质量直接影响模型能否学习到稳定有效的状态特征。鉴于科研仪器电流数据具有独特的物理结构特性，图像或音频领域常用的增强方法（如旋转、

裁剪、随机裁切)并不适用于时序序列信号,对此需要设计针对时序数据的增强策略。

如图 3.6 所示,在本文对标准化后的时域序列引入轻微的随机缩放以模拟仪器负载轻微变化并叠加微小高斯噪声^[64],模拟传感器热噪声及微观波动,这类时间序列增强方法已被证明能够有效提升模型的泛化能力。同时,对统计频域特征向量施加随机掩码,避免模型过度依赖单一统计量做判别^[65]。通过上述增强方式,使模型能够学习仪器状态的稳定不变特征,而非依赖采集过程中的随机噪声模式,从而提升状态表征的鲁棒性。

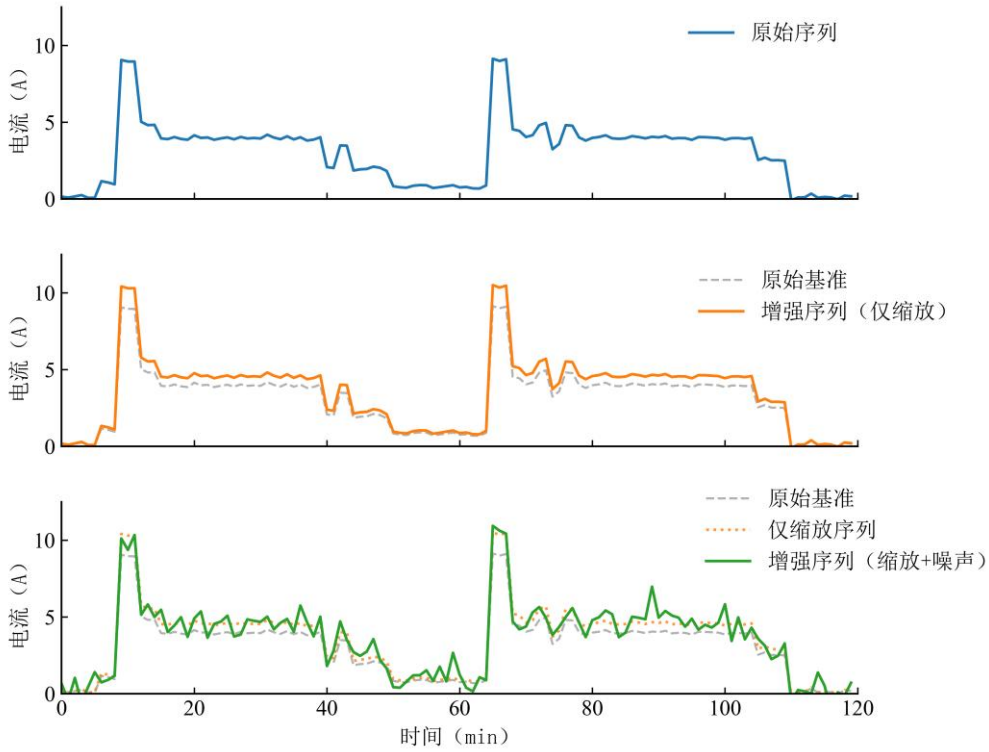


图 3.6 时域波形增强示意图

(2) 优化器与超参数配置

在优化器选择上,本文采用 AdamW 优化器^[66],相较于传统 Adam,AdamW 通过解耦权重衰减,能够更有效地抑制过拟合^[67],同时保持梯度更新的稳定性。在学习率调度上,初始学习率设置为 1×10^{-3} ,并结合余弦退火策略^[68],使得学习率在训练后期平滑下降,帮助模型更稳定地逼近最优解。批次大小设为 128。对于对比学习任务而言,较大批次能够提供更多负样本,从而提升 TQ-NCE 损失的互信息下界估计精度。此外,在训练过程中,本文在全连接层引入 Dropout 机制(比率为 0.2)以减少过拟合,并结合 $\lambda = 1 \times 10^{-4}$ 的权重衰减进一步增强模型的泛化能力。

(3) 早停机制的实施

整个训练过程通常迭代 100 至 200 个 Epoch。为了防止模型在训练集上过拟合,本文划分了 20% 的验证集,并实时监控总损失、对比损失以及正负样本对的相似度分布,以确保模型在降低重构误差的同时,逐渐提升正样本的聚合度并拉开负样本距离。当验证集损失连续 10 个 Epoch 不再下降时,触发早停机制,并保存验证集表现最优的模型权重作为最终结果。

(4) 推理阶段的特征获取

网络末端的投影头仅用于在训练阶段，将特征映射到对比空间以计算损失。在训练完成后的推理与应用阶段，会移除这一非线性映射层，直接截取融合层输出的 $Z_{fused}^{(t)}$ 作为最终的状态表征。该操作使这组特征最大程度地保留了原始数据的语义信息，并直接输入到后续的聚类算法与异常检测模块中。

3.4 仪器状态聚类与异常判别

经过前述 QA-TF-CL 网络的特征提取、门控融合与质量加权对比训练，原本受噪声干扰且时序特征模糊的分钟级电流数据，已被编码为具有较强区分能力的融合特征向量 $Z_{fused}^{(t)}$ 。模型在特征空间内能够较好的隔离不同的运行工况。然而，由于实际应用现场通常无法提供充足的人工标签，仅依靠特征可分性无法直接输出诊断结果。因此，本节采用无监督聚类方法来解决这一问题。系统将通过挖掘特征向量在空间分布上的聚簇结构，来识别仪器的常规工作模式，并据此设计一套结合空间距离与特征重构的异常评分指标。

3.4.1 基于特征空间的无监督聚类

在推理阶段，实时采集的电流数据需经历与训练阶段严格一致的预处理流程，以保证特征空间分布的统一。构建完成的标准输入实例 S_t （此时不包含 m_t ）送入网络后，在双流编码器与门控融合模块参数冻结的状态下，网络会输出融合特征 $Z = \{Z_{fused}^{(1)}, Z_{fused}^{(2)}, \dots, Z_{fused}^{(T)}\}$ 。该特征可作为表征仪器当前运行工况的状态指纹。

表 3.1 K-Means++算法流程

Algorithm 1 K-Means++(k) initialization
1. $C \leftarrow$ sample a point uniformly at random from X
2. while $ C < k$ do
3. Sample $x \in X$ with probability $\frac{d^2(x, C)}{\phi_x(C)}$
4. $C \leftarrow C \cup \{x\}$
5. end while

对比学习通过拉近正样本、推远负样本的机制，会使同类样本在超球面空间内聚集成簇。鉴于目标仪器的标准运行模式为几种已知的离散固定工况，本文选用基于欧氏距离度量的 K-Means++算法进行状态划分，算法如表 3.1 所示，具体的算法流程为：从数据集中随机选择一点作为首个中心点；对于数据集中的每一个样本点 x 分别计算它与当前已有质心之间的最短距离，记为 $D(x)$ ；选择一个新的点作为新的中心点：根据概率公式， $D(x)$ 越大的点作为中心点的概率越大；重复步骤 2 和 3 直到选择出 K 个中心点^[69]。

算法收敛后，将得到 K 个聚类中心：

$$C = \{C_1, C_2, \dots, C_K\} \quad (3.27)$$

每一个聚类中心 C_k 不仅是几何意义上的质心，也对应着仪器在特定工况下的物理语义原型。通过分析簇内样本的原始电流分布特征，可以进一步解释这些聚类簇的实际工程含义，并完成状态标注。

相比直接对原始数据进行聚类，基于 QA-TF-CL 融合特征的聚类方式具有更好的表现。原始分钟级电流序列易受插值噪声和随机波动干扰，直接处理往往会产生无物理意义的噪声簇。而质量加权机制在训练阶段已有效过滤了这些干扰信息，使得最终的特征聚类能够更精准地还原仪器的真实工作状态。

3.4.2 基于距离重构的异常评分机制

在复杂科研仪器的监测过程中，单一度量指标往往难以覆盖形态各异的异常表现。基于特征空间距离的检测方法对未见工况或整体运行模式发生偏移的异常较为敏感，但在面对已知工况内部的细微波形畸变或局部非物理扰动时，往往存在检测能力不足的问题。相反，基于重构误差的生成式方法能够有效捕捉波形层面的异常扰动，但对宏观运行状态的整体偏移敏感性不足^[70]。

为充分发挥 QA-TF-CL 模型在对比学习与时域重构方面的双重优势，本文采用一种基于距离-重构的联合异常评分机制，从特征空间几何偏离与原始物理波形一致性两个互补视角，对仪器运行状态进行综合评估^[71]。

在推理阶段，对于在线监测得到的新样本 S_{new} ，首先通过冻结的双流编码器与门控融合模块，获得其融合特征表示 Z_{new} 。基于上节中构建的正常状态原型集合 $\{C_k\}_{k=1}^K$ ，定义特征空间偏离度为当前样本到最近状态簇中心的欧氏距离：

$$S_{dist} = \min_{k \in \{1, \dots, K\}} \|Z_{new} - C_k\|_2 \quad (3.28)$$

该指标用于衡量当前样本在隐空间中相对于已知正常工况流形的偏离程度。当仪器进入训练阶段未覆盖的运行模式，或由于设备老化、参数漂移导致整体状态发生显著变化时，样本特征往往会落入特征空间的低密度区域，从而引起 S_{dist} 的明显增大。

不过仅依赖特征空间距离难以识别工况不变但波形局部失真的异常情况。为此，本文进一步引入基于时域解码器的重构误差作为补充指标。利用 3.3.4 节中在正常样本上训练得到的解码器，将融合特征 Z_{new} 映射回原始时间序列空间，得到重构序列 $\hat{X}_{time}$ ，并将其与输入序列 $\tilde{X}_{time}$ 之间的均方误差定义为重构得分：

$$S_{recon} = \frac{1}{N} \|\tilde{X}_{time} - \hat{X}_{time}\|_2^2 \quad (3.29)$$

由于解码器在训练过程中主要学习了符合正常物理运行规律的波形流形，当输入序列中存在异常振荡、非物理跳变或破坏时间连续性的噪声扰动时，模型将难以实现有效重构，导致重构误差显著增大^[72]。因此，细微波形畸变及局部异常扰动可以通过重构误差得到有效表征。

为消除不同指标之间的量纲差异并增强可比性，本文用验证集中正常样本的统计信息，对 S_{dist} 与 S_{recon} 进行 Z-Score 标准化。随后，通过加权融合得到最终异常得分：

$$S_{total} = \alpha \cdot \mathcal{N}(S_{dist}) + (1 - \alpha) \cdot \mathcal{N}(S_{recon}) \quad (3.30)$$

其中， $\mathcal{N}(\cdot)$ 表示 Z-Score 标准化操作， α 为平衡系数，用于调节两类指标在最终得分中的相对权重。为避免模型对任一异常指标产生偏置，本文采用等权融合策略，并在实验中固定取 $\alpha = 0.5$ 。

由于科研仪器在不同工况下的异常得分基准存在明显差异，采用预设的固定阈值难以兼顾检测的灵敏度与普适性。为此，本研究利用验证集中正常样本的统计特征，实现决策边界的自适应估计。考虑到异常得分 S_{total} 在实际场景下往往呈现非高斯的长尾分布特征，传统的参数化方法易受到极值干扰。本文采用非参数化的百分位法来确定基础报警阈值^[73]。具体而言，选取验证集 $\mathcal{D}_{val}$ 中正常样本得分的 $1 - \delta$ 分位数作为判别上界：

$$\tau_{th} = P_{1-\delta} \left(\left\{ S_{total}^{(i)} \mid i \in \mathcal{D}_{val} \right\} \right) \quad (3.31)$$

在基准实验中，设置 $\delta = 0.01$ ，即在 99% 的置信水平下确定正常波动的判别边界。对于实时监测样本，其判定逻辑如下：当 $S_{total}(t) \leq \tau_{th}$ 时，判定当前状态为正常，并将其归类为距离最近的状态簇 C_k 所对应的运行工况；当 $S_{total}(t) > \tau_{th}$ 时，判定当前状态为异常，触发异常报警并记录异常发生时刻。同时，为降低瞬时噪声引起的误报，引入时间窗口平滑策略：仅当连续 M 个时间窗口的异常得分均超过阈值 τ_{th} 时，才将该异常判定为有效异常事件^[74]。

然而，在实际的科研仪器在线监测工程中，这种一刀切的单阈值二元判定机制面临着严峻挑战：若阈值设定过高，极易漏报微小的早期故障；若阈值设定过低，又会因仪器的正常动态扰动引发大量误报。因此，如何将 S_{total} 评分更加柔性、科学地应用于实际运维系统，仍是需要进一步解决的问题。针对这一问题，本文将在第四章中打破单阈值的局限，提出基于 T_{low} 与 T_{high} 的自适应双重阈值分流机制，并结合人机协同完成最终的异常确权。

3.5 实验验证与分析

为系统评估所提出的 QA-TF-CL 算法在科研仪器运行状态识别与异常检测任务中的有效性与鲁棒性，本节基于真实采集的分钟级电流数据开展了多组对比实验。实验设计重点围绕以下三个核心问题展开：

- (1) 低采样率条件下，引入时频双视图表征是否能够有效提升仪器运行状态的可分性；
- (2) 相较于典型的无监督时序建模方法，QA-TF-CL 是否在状态辨识与异常检测性能上表现更优；
- (3) 所提出的统计频域建模与质量加权机制，是否能够在数据丢包与噪声干扰加剧的情况下提升模型的鲁棒性。

3.5.1 实验设置

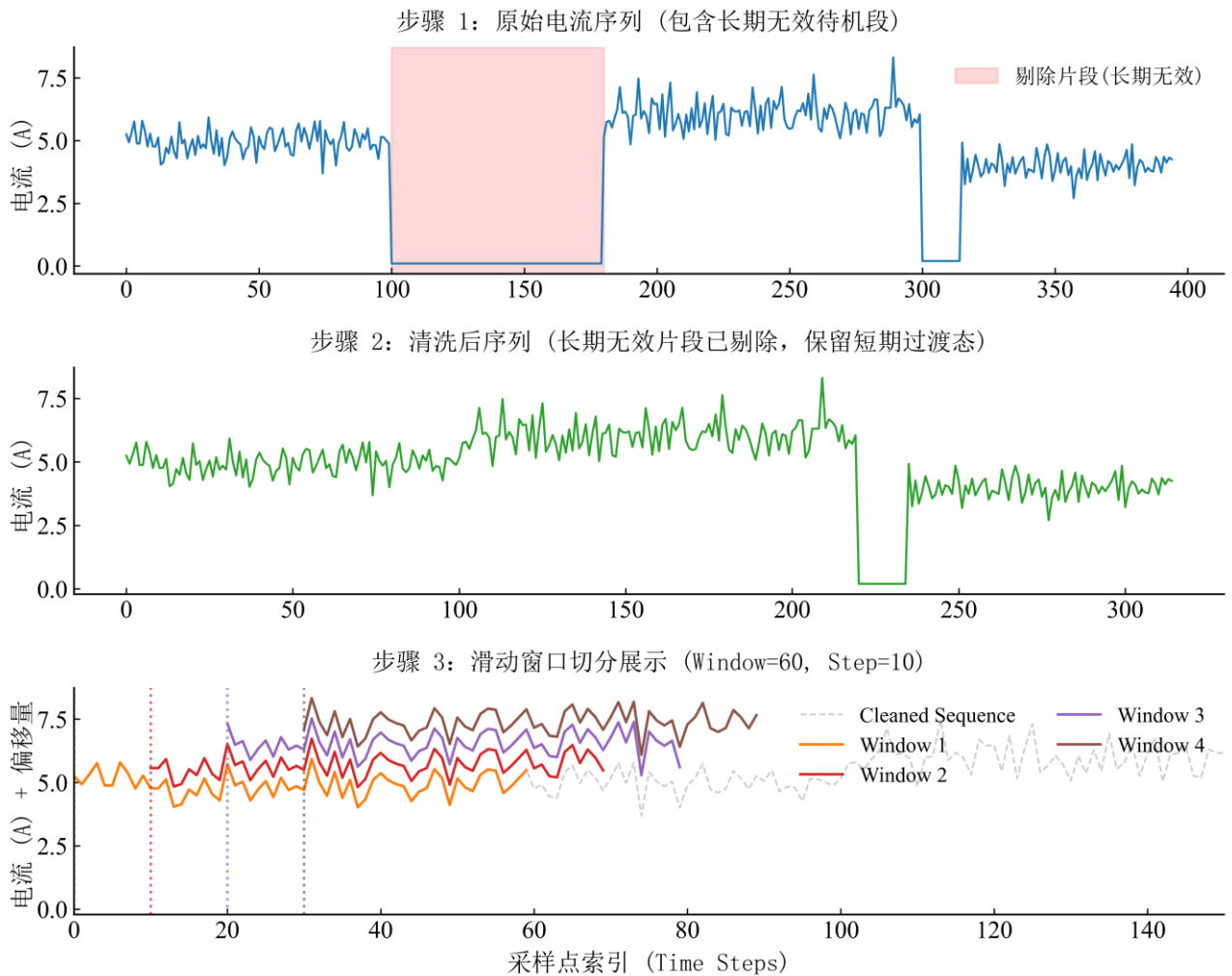
(1) 数据集与样本构建

实验数据来源于平台数据库，涵盖了 8 台具有代表性的科研仪器，累计采集时长超过 180 天。为确保实验覆盖了不同类型的运行工况，这些仪器根据其能量转换模式与电流波动特征被分为以下四类：

表 3.2 不同仪器类型的电流特征描述

仪器类型	代表仪器	电流特征
动力驱动型	离心机、真空泵	明显的启动升速、稳态运行与降速阶段
精密测量型	气质联用仪、质谱仪	电流背景复杂，包含多组件协同工作的微弱脉冲特征
热能转换型	高温管式炉、干燥箱	长周期的缓慢升温与 PID 调节下的脉冲波动
循环切换型	超低温冰箱、恒温培养箱	间歇性工作特征，电流在待机与全速运行之间频繁切换

如图 3.7 所示，为了捕获电流序列中的时序依赖性与波动纹理，数据预处理阶段首先剔除了一些长期无效片段，并采用滑动窗口方式对原始时间序列进行切分。窗口长度设为 60 个采样点，对应 1 小时物理运行时长，以保证窗口能覆盖完整的工况切换过程。滑动步长设为 10 个采样点，通过重叠采样提升模型对过渡态特征的捕捉能力。



在获得全量样本窗口后，本文按照约 8:2 的比例将其划分为训练集与测试集。训练集占比约 80%，总计包含 160000 个样本窗口，仅包含仪器正常运行状态下的电流序列，用于 QA-TF-CL 模型的无监督表征学习。测试集占比约 20%，总计包含 40000 个样本窗口，包含正常工况以及异常样本，用于评估模型的最终辨识与检测性能。模型训练过程完全基于无监督方式进行，不使用任何先验标签。

在测试阶段，为了定量评估性能，本研究结合运维日志及人工核验结果，对测试集样本进行了标注。测试集中正常工况（如待机、工作等已知模式）样本占比约为 88%，异常样本占比约为 12%。结合专家经验与常见故障机理，本文将异常样本划分为以下三类典型物理故障：第一类为突发物理掉电/断连，其表现为电流幅值瞬间跌落至零并持续，通常由设备供电中断、网关宕机或线路脱落引起。第二类为局部高频振荡/毛刺，其表现为在正常运行基线上附近出现了非物理特性的剧烈突变波动，多由传感器松动、外部电磁干扰或电路板老化导致；第三类为局部跳变，其表现为电流信号在短时间内出现明显的突升或突降，并在局部时间段内偏离正常变化趋势，通常与瞬时接触不良、采集链路扰动或局部器件异常有关。具体的测试集样本分布详情如表 3.3 所示。上述标签仅用于后验评价，不参与模型的梯度更新过程。

表 3.3 测试集样本类型与分布统计

样本类型	物理描述	样本数量	占测试集比例
正常工况	待机、预热、稳态运行等常规调度状态	35200	88%
异常：掉电/断连	突发电流跌落至零并维持	1200	3%
异常：振荡/毛刺	局部高频剧烈非正常波动	2000	5%
异常：局部跳变	短时间内出现明显的突升或突降	1600	4%

(2) 对比方法设置

为验证所提方法的有效性，本文选取以下五种具有代表性的基准方法进行对比，分别覆盖传统聚类方法、重构类时序模型以及典型自监督对比学习方法。具体方法如表 3.4 所示。

表 3.4 对比基准方法说明

方法	具体实现
K-Means++ ^[75]	直接对原始电流时间序列进行 K-Means++ 聚类，作为最基础的对照方法
LSTM-AE ^[22]	基于长短期记忆网络的自编码器模型，通过重构误差学习时序特征
Standard SimCLR ^[26]	标准对比学习框架，仅使用时域输入特征，且不引入数据质量加权机制
TF-C ^[29]	时序对比学习模型，通过最大化同一时间序列在时域和频域特征表示之间的一致性来学习深度表征
DCdetector ^[71]	时序异常对比学习方法，通过构造双重对比约束增强正常模式表征能力，从而提升异常识别效果

上述方法均在相同的数据划分和评价指标下进行测试，并尽量保持一致的实验设置，以

保证对比结果的公平性。

(3) 评价指标

为了全面评估 QA-TF-CL 模型在无监督工况划分与异常状态捕捉方面的表现。实验采用以下两类指标对模型性能进行综合评估：

1) 状态聚类性能指标

标准化互信息 (Normalized Mutual Information, NMI): 衡量聚类划分结果与真实状态标签之间的信息重合度。设 Y 为真实状态集合, C 为算法聚类结果, 其中 $I(Y;C)$ 为二者的互信息, $H(\bullet)$ 为信息熵。

$$NMI(Y, C) = \frac{2 \times I(Y; C)}{H(Y) + H(C)} \quad (3.32)$$

NMI 取值范围为 $[0, 1]$, 越接近 1 表示聚类划分与实际工况越吻合。

轮廓系数 (Silhouette Coefficient, SC): 从几何角度评估特征空间的质量。对于样本 i , 设 $a(i)$ 为其到同簇内其他样本的平均距离 (簇内紧凑性), $b(i)$ 为其到最近异簇样本的平均距离 (簇间分离度)。

$$SC(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3.33)$$

整体 SC 指标为全量样本的均值。取值越接近 1, 表示工况特征聚类越紧凑、边界越清晰。

2) 异常检测性能指标

准确率 (Precision, P): 在模型判定的所有异常样本中, 真实异常所占的比例。

$$P = \frac{TP}{TP + FP} \quad (3.34)$$

召回率 (Recall, R): 在所有实际存在的异常中, 被模型成功检出的比例, 反映了系统的查全能力。

$$R = \frac{TP}{TP + FN} \quad (3.35)$$

F1 分数 (F1-Score): 准确率与召回率的调和平均值, 是衡量异常检测模型整体效能的核心指标。

$$F1 = \frac{2 \times P \times R}{P + R} \quad (3.36)$$

误报率 (False Positive Rate, FPR): 衡量正常工况被错误触发告警的频率。

$$FPR = \frac{FP}{FP + TN} \quad (3.37)$$

漏报率 (False Negative Rate, FNR): 衡量实际故障未被模型识别出来的比例, 反映了系统的安全风险。

$$FNR = \frac{FN}{FN + TP} \quad (3.38)$$

其中, TP(True Positive)为真正例, FP(False Positive)为假正例, TN(True Negative)为真负例, FN(False Negative)为假负例。

3.5.2 预测结果对比分析

为综合评估 QA-TF-CL 在仪器运行状态划分与异常检测任务中的表现, 本文将状态聚类指标与异常检测指标统一汇总, 结果如表 3.5 所示。

表 3.5 状态辨识与异常检测性能对比

方法	NMI	SC	P	R	F1-Score	FPR	FNR
K-Means++ ^[75]	0.45	0.32	0.6336	0.5403	0.5832	0.1673	0.4597
LSTM-AE ^[22]	0.62	0.55	0.7102	0.6862	0.6980	0.1228	0.3138
Standard SimCLR ^[26]	0.71	0.64	0.7825	0.7503	0.7661	0.0937	0.2497
TF-C ^[29]	0.79	0.72	0.8427	0.8354	0.8390	0.0631	0.1646
DCdetector ^[71]	0.82	0.75	0.8715	0.8830	0.8772	0.0452	0.1170
QA-TF-CL	0.86	0.78	0.9102	0.9248	0.9174	0.0284	0.0752

从结果可以看出, K-Means++作为传统无监督聚类基线, F1-Score 仅有 0.5832, FNR 达到了 0.4597。由于科研仪器的电流数据通常伴随着复杂的环境噪声和非平稳特征, 若缺乏深层的特征提取, 直接基于原始序列去聚类难以获得理想结果。基于重构误差的 LSTM-AE 表现略有提升, F1-Score 为 0.6980, 说明用深度网络去捕捉时序依赖关系能够在一定程度上改善状态表征效果。不过, 生成式模型的一个通病是容易把异常噪声也一起学进去并重构出来, 所以在高噪环境下, 它的误报和漏报率依然偏高。

在对比学习框架下, 频域特征的引入表现出了明显优势。仅依赖时域特征的 Standard SimCLR 获得 0.7661 的 F1-Score; 而采用时频双视图的 TF-C 模型, 其 F1-Score 提升至 0.8390, NMI 与 SC 也分别达到 0.79 和 0.72。这表明, 面对仪器复杂的运行状态, 单一时域信息的表征能力仍然有限; 融合频域视图有助于捕获更全面的全局变化规律与局部异常特征, 从而增强模型对微弱异常的辨识能力。

此外, 引入近期的时序异常检测模型 DCdetector 进行对比可以发现, 该模型凭借双重注意力机制与对比表征, 在各项指标上整体优于 TF-C, 其 NMI 达到 0.82, F1-Score 达到了 0.8772, 表现出较强的异常检测能力。然而, 在本文所涉及的高丢包率、含有较多插值伪影的低质量物联网监测数据场景下, 其 FNR 仍然偏高。这说明, 尽管 DCdetector 在理想条件下具备较强性能, 但在复杂噪声与数据缺失条件下, 对异常样本的稳定辨识能力仍存在一定局限。

在 TF-C 与 DCdetector 已取得较好表现的基础上, QA-TF-CL 的 F1-Score 进一步提升至 0.9174。这一性能提升与质量感知机制的引入密切相关。由于实际监测数据中常夹杂丢包与插值伪影, TF-C 对不同时段样本采用统一处理方式, 容易受到低质量片段的干扰。相比之下, QA-TF-CL 通过动态评估数据置信度, 对低质量样本对的影响进行抑制, 从而降低了其对表

征空间学习的干扰，使模型在精确率（ $P=0.9102$ ）与召回率（ $R=0.9248$ ）之间达到了更好的平衡。

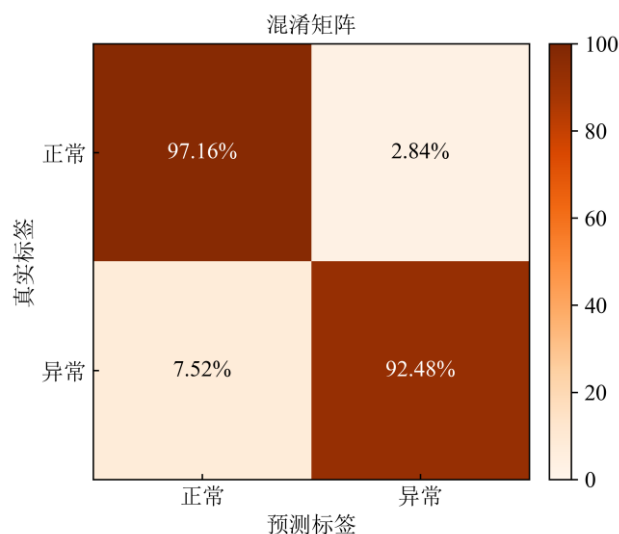


图 3.8 QA-TF-CL 模型检测结果混淆矩阵图

在可靠性方面，图 3.8 展示了 QA-TF-CL 模型在测试集上的归一化混淆矩阵，直观反映了模型对正常运行与仪器异常两类状态的分类效果。结果表明，模型在实际正常运行样本中的正确判定比例高达 97.16%，误报率被严格压制在 2.84%，说明其能够较好地抑制高频虚警，减轻告警疲劳；同时，模型对异常状态的检出率达到 92.48%，漏报率为 7.52%，较 DCdetector 的 11.70% 进一步降低。较低的漏报率有助于减少仪器在异常状态下持续运行所带来的潜在风险。剩余少量漏报样本主要表现为与正常工况信号高度相似的早期微弱漂移，后续可结合第四章提出的人机协同机制进行进一步复核与判别。

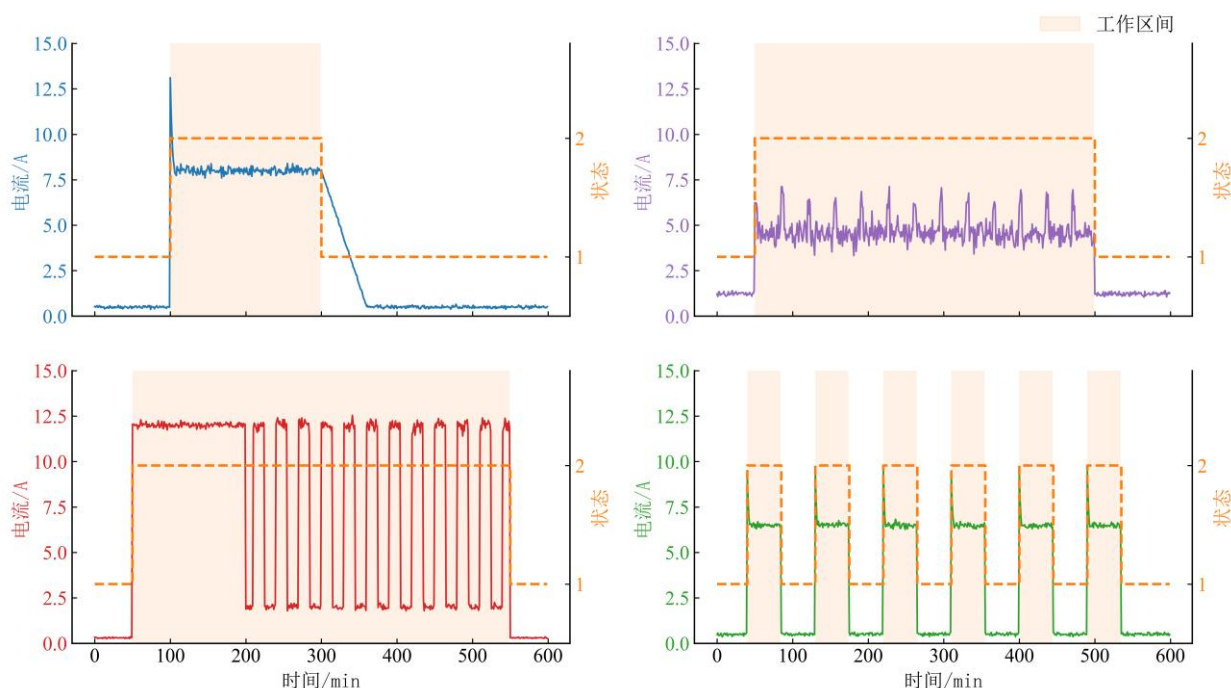


图 3.9 状态识别结果

图 3.9 展示了基于质量感知对比学习的仪器电流状态辨识方法对四类典型科研仪器的状态切分结果。从图中可以看出，得益于双视图对比学习构建的高内聚特征空间，模型不仅能够较好地刻画动力驱动型仪器在回落阶段的过渡态边界（左上），还能够在存在复杂毛刺噪声的工作区间内保持状态划分的相对平滑性（右上）。尤其在循环切换型仪器的 PID 调节场景下（右下），该方法能够将剧烈脉冲波动对应的运行过程识别为同一连续工况，较好地避免了传统固定阈值方法中容易出现碎片化切分现象，表明其在无监督条件下对复杂工况语义具有较好的适应能力。

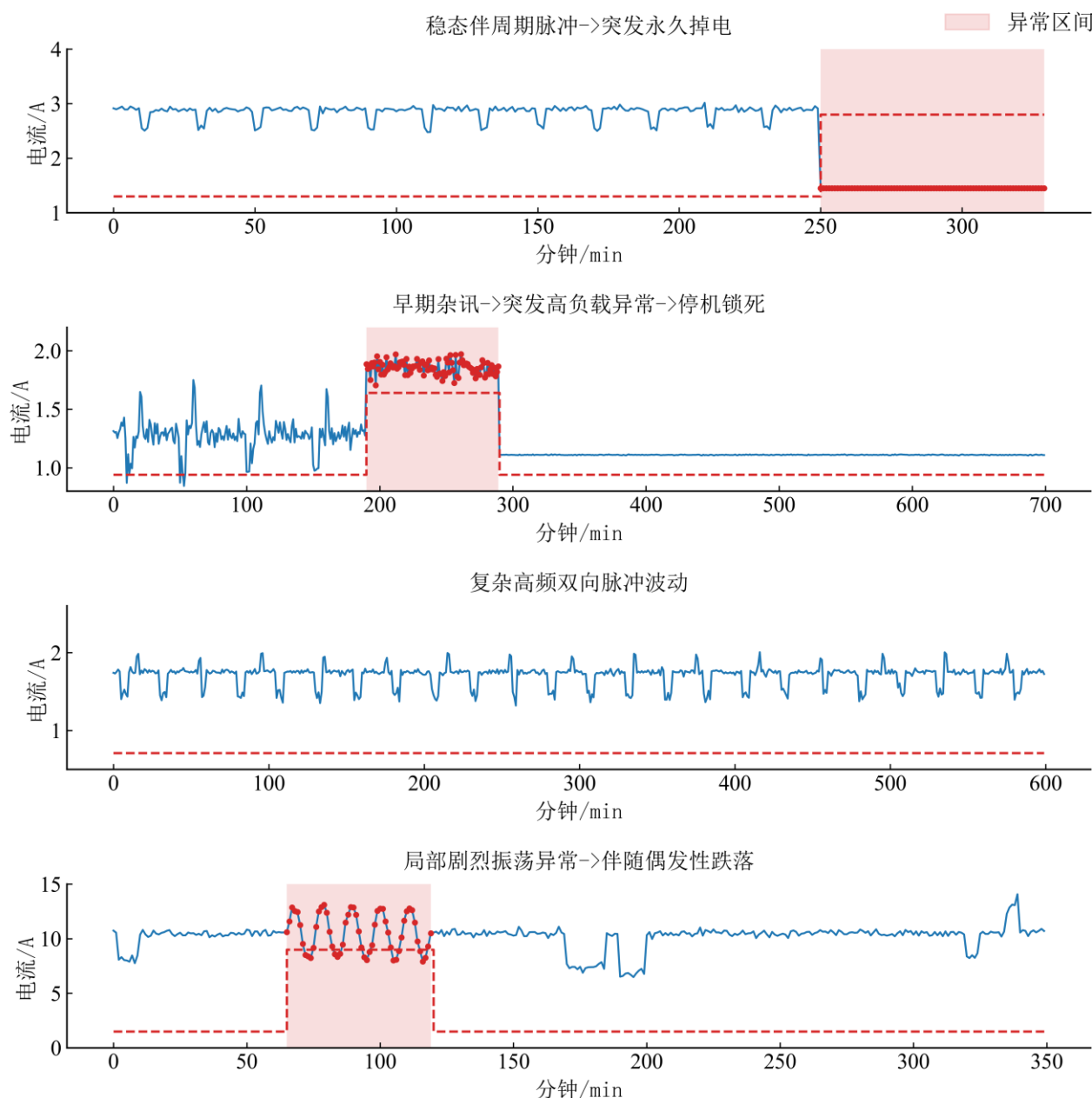


图 3.10 异常监测结果

图 3.10 展示了模型在典型场景下的异常监测结果。从结果可以看出，距离—重构联合得分不仅能够识别突发掉电等宏观状态异常，还能够借助时域解码器的重构误差约束检测到未

明显偏离幅值包络的局部波形畸变，如局部剧烈振荡。此外，由于 TQ-NCE 损失函数中的数据质量掩码对低质量样本进行了加权抑制，模型对早期剧烈插值噪声以及复杂高频双向正常脉冲表现出较好的抗干扰能力。整体来看，该方法在保持较低误报率的同时，兼顾了异常检出的敏感性。

3.5.3 核心模块的消融实验

为验证本章两个核心模块，即时频门控融合与质量加权损失，对模型性能的具体贡献，本文设计了逐步叠加模块的消融实验：Variant A (Time-Only) 仅使用时域编码器，不引入统计频域特征与质量加权；Variant B (Dual-View) 引入时频双流与门控融合，但采用标准 InfoNCE 损失；Variant C (Full Model) 是完整的 QA-TF-CL 模型。实验结果如表 3.6 所示。

表 3.6 不同模型变体的性能对比

模型变体	NMI	SC	P	R	F1-Score	FPR	FNR
Variant A	0.69	0.61	0.7765	0.7352	0.7553	0.1085	0.2648
Variant B	0.76	0.69	0.8214	0.7956	0.8083	0.0725	0.2044
Variant C	0.86	0.78	0.9102	0.9248	0.9174	0.0284	0.0752

从 Variant A 与 Variant B 的结果对比可以发现，引入频域特征后，模型的各项性能指标均得到了提升。作为仅依赖时域特征的基础对比学习网络，Variant A 的 F1-Score 为 0.7553，FNR 为 0.2648，表明在科研仪器非平稳运行场景下，单一时域序列的表征能力仍然有限。在此基础上，加入时频双流结构与门控融合机制后，Variant B 的 F1-Score 提高至 0.8083，FNR 则降至 0.2044。该结果说明，频域视图能够补充时域中较难表征的整体变化规律与局部异常特征，而门控机制则增强了两类特征之间的自适应互补能力，从而提升了模型对复杂工况的辨识效果。

虽然 Variant B 已体现了双视图架构的优势，但其误报率（0.0725）与漏报率（0.2044）在实际工程场景下仍存在优化空间。通过 Variant B 与 Variant C 的对比可见，质量感知加权损失（TQ-NCE）的引入进一步提升了模型性能。加入该机制后，Variant C 的 F1-Score 提升至 0.9174，同时 FNR 也下降至 0.0752。

在实际监测中，电流数据常伴有传感器丢包造成的插值伪影或电磁干扰。Variant B 采用标准的 InfoNCE 损失，对不同时间段样本使用统一的对比约束，容易将低质片段的噪声引入对比学习过程，从而干扰特征分布。而 Variant C 通过评估数据置信度，在损失计算中降低了这些劣质特征的权重，维持了表征空间的高内聚性。这使得模型在精确率（0.9102）和召回率（0.9248）上取得了更好的平衡，同时也提升了模型在强噪环境下的稳定性。

综合来看，消融实验结果表明，时频双流结构主要缓解了低采样率条件下信息表征不足的问题，而质量感知机制则进一步抑制了现实传输过程中的噪声干扰。两者协同作用，使最终模型在状态辨识与异常检测任务上取得了最佳表现。

3.5.4 数据质量敏感性与鲁棒性分析

在科研仪器真实的物联网监测场景中，受边缘侧通信链路波动和基站切换等因素影响，数据丢包与传输延迟是较为常见的干扰。为评估 QA-TF-CL 算法在极端恶劣数据环境下的性能边界，本节通过模拟实验对模型的鲁棒性进行了压力测试。

本实验在原始 1 分钟/点的电流序列中随机剔除部分采样点，并采用线性插值进行补全，以模拟边缘侧常用的数据对齐策略。在此基础上，构建了丢包率从 0% 逐步增加至 50% 的 6 组测试集，用于评估模型在不同数据质量条件下的性能变化。

在本小节的对比方法选择上，本文选取 Standard SimCLR、TF-C 和 DCdetector 作为对比对象。其原因在于：K-Means++ 与 LSTM-AE 在处理含有大量插值伪影的非平稳序列时，由于缺乏特征抽象能力和过度依赖时序连续性，其性能在低丢包率下发生严重退化，不具备极端环境下的参考价值。相比之下，Standard SimCLR 和 TF-C 与本文模型同属对比学习范式，且未引入质量感知机制；DCdetector 则代表了当前基于双重注意力机制的时序异常检测模型。通过引入上述方法进行比较，可以进一步分析在数据质量显著劣化时，模型结构复杂度的提升是否足以缓解噪声干扰，并更清晰地评估本文所提出质量感知权重在提升系统稳健性方面的作用。

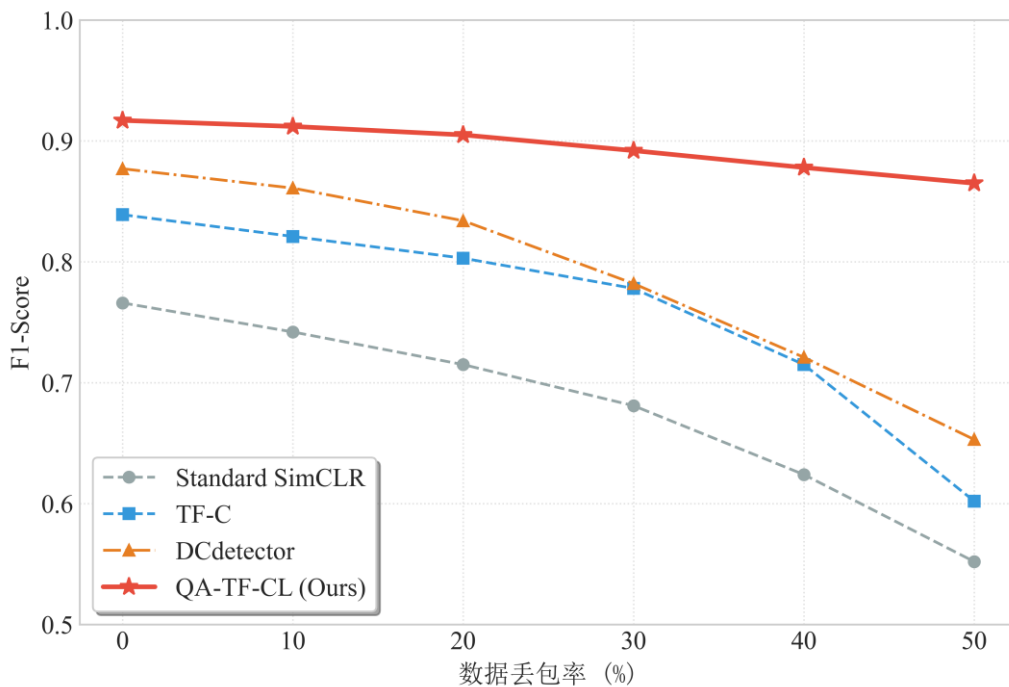


图 3.11 不同丢包率下 F1-Score 变化曲线

图 3.11 展示了 QA-TF-CL 与各对比模型在不同丢包率下的 F1-Score 变化趋势。整体来看，Standard SimCLR 对数据质量较为敏感，其性能随着丢包率增加而持续下降。当丢包率达到 30% 时，F1-Score 已降至 0.70 以下；在 50% 丢包率条件下，得分进一步下降至约 0.55。这说明，线性插值引入的平滑伪影会直接干扰时域特征提取，从而削弱单视图模型的代表能力。

TF-C 在低丢包率区间表现出一定鲁棒性，这主要得益于频域分支对全局统计特征的补充。

然而，由于该模型要求时域与频域特征保持较强一致性，当插值伪影导致时域信号失真时，这种一致性约束可能将噪声影响进一步传递至频域分支。因此，当丢包率超过 30% 后，TF-C 的性能下降趋势明显加快，难以在低质量数据条件下维持稳定表现。

DCdetector 在 0% 丢包的理想条件下表现较好，F1-Score 达到 0.88，在对比模型中处于较高水平。但随着丢包率提升，其性能也出现持续下降。当丢包率升至 30% 时，F1-Score 降至 0.78；在 50% 丢包率条件下进一步下降至 0.65。分析认为，DCdetector 所依赖的自注意力机制对局部细节较为敏感，而线性插值产生的平滑片段可能在注意力矩阵中形成具有误导性的规律特征，从而使模型在噪声环境下更容易出现判别偏移。

相比之下，本文提出的 QA-TF-CL 在整个丢包测试区间内表现出更好的稳定性。即使在 50% 丢包率条件下，其 F1-Score 仍保持在 0.85 以上。

上述结果表明，TQ-NCE 损失函数中的数据质量加权机制在高丢包率场景下发挥了重要作用。当检测到数据质量下降时，模型会自动减弱低质量样本在损失计算中的影响，并更多依赖高质量样本维持表征空间的稳定性，从而提高了整体模型对数据缺失与插值伪影的鲁棒性。

3.6 本章小结

本章聚焦于科研仪器在分钟级低频采样条件下，因数据缺失与插值噪声导致的状态辨识难题，提出了一种基于质量感知的时频双视图对比学习方法(QA-TF-CL)。该方法降低了传统状态监测对高频波形数据的依赖，在无监督条件下实现了对仪器运行状态的鲁棒表征与异常检测。本章的主要研究工作与结论概括如下：

(1) 构建了适应低频采样的多视图特征体系。针对分钟级数据频域信息缺失的问题，引入了统计频域特征，通过波动能量、趋势强度与峰值因子等统计指标，构建了信号的等效频域视图，有效弥补了单一时域信息表征能力不足的问题。

(2) 提出了质量约束下的鲁棒特征学习框架。设计了并行的双流编码器与自适应门控融合机制，实现了时频特征的动态互补。通过构建时间质量加权 InfoNCE 损失函数(TQ-NCE Loss)，将数据质量掩码显式引入到梯度更新过程。该机制能够在训练阶段自动抑制低质量插值样本的干扰，从而在噪声环境中学习到更稳定的状态特征。

(3) 实现了无监督的状态聚类与异常判别。基于 QA-TF-CL 提取的高内聚融合特征，利用 K-Means++ 算法自动发现仪器的典型物理工作模式，并建立了基于距离重构的异常评分机制，用于定量评估设备的健康状态。

(4) 验证了算法在恶劣工况下的有效性。真实数据集上的实验结果表明，QA-TF-CL 在状态辨识精度上显著优于多种典型时序建模方法。特别是在数据丢包率高达 50% 的压力测试中，本方法仍能保持较高的异常检测 F1-Score 值，证明了质量约束机制在低质工业数据环境下具有较好的鲁棒性。

综上所述，本章的 QA-TF-CL 算法解决了底层数据表征和基础异常检测问题。然而，要

将其真正落地到运维系统中，仅依赖纯算法的单阈值判定仍然不足。因此，下一章将重点探讨如何把这套算法与人工经验相结合，设计自适应双阈值与人机协同的异常确权机制，从而缓解实际应用中的误报压力。

4 基于人机协同的状态语义映射与知识演化机制

在第三章中，本文提出了基于质量感知的时频双视图对比学习方法（QA-TF-CL），通过在特征提取与对比学习过程中引入数据质量约束，较好缓解了科研仪器在低采样率与插值噪声环境下的鲁棒状态表征与无监督异常检测问题。实验结果表明，该方法能够学习到具有良好聚类结构的融合特征向量 Z_{fused} ，并基于距离度量较为准确地识别偏离稳态的异常波动。

然而，仅依靠数据驱动而无监督算法在实际工程应用中仍面临语义鸿沟与模型适应性不足两方面挑战。一方面，模型聚类生成的数字标签（如 Cluster 0、Cluster 1）缺乏明确的物理语义，运维人员难以直接理解其对应的仪器工况；另一方面，异常检测算法在实际应用中不可避免地会产生误报，若缺乏基于反馈的持续修正机制，模型将难以适应长期工况漂移带来的分布变化。为此，本章提出了一种专家反馈驱动的人机协同（Human-in-the-Loop, HITL）运维闭环框架，旨在将无监督特征转化为可解释的领域知识，并推动监测系统的持续优化与演化。

4.1 传统运维瓶颈与人机协同启示

为进一步回应上述问题，本节首先分析传统运维模式与静态模型的局限性，并结合人机协同技术在相关领域中的应用思路，引出本章的总体设计框架。

4.1.1 传统科研仪器运维模式的局限性

科研仪器的稳定运行直接关系到实验数据的有效性与可靠性。然而，当前主流的阈值报警结合人工巡检的运维模式在处理海量实时监测数据时，逐渐暴露出明显局限，难以满足现代科研对高精度、高效率运维的要求。该模式高度依赖静态阈值触发报警机制，缺乏对环境动态变化的自适应能力，容易因仪器瞬态波动或外部电磁干扰产生大量无效报警。运维人员长期处于高频误报的干扰环境中，容易形成告警疲劳，进而降低对真实故障信号的敏感性，并增加漏判风险。

与此同时，故障诊断过程在较大程度上依赖资深专家的隐性经验，而这些经验往往未被系统化沉淀与数字化传承。一旦关键人员流动或岗位变动，便容易出现知识断层，影响运维能力的可持续性。

此外，现有数据驱动模型在部署后参数固化，无法动态响应仪器运行环境、老化状态及工况条件的渐进性演变，致使模型检测精度随时间推移持续衰减，难以维持长期有效的故障预警能力。

4.1.2 跨领域人机协同的经验启示

近年来,人机协同、主动学习与持续学习等技术在医疗、工业等高风险领域的成功实践,为科研仪器运维的智能化转型提供了有益借鉴。

在医疗影像分析领域,深度神经网络虽能输出高精度预测结果,但其黑盒属性限制了临床决策的可解释性与可信度。为增强可解释性,研究者提出了基于梯度的可视化方法,用于标注模型关注的关键区域,从而辅助医生进行语义确认^[76]。这种算法预筛选和专家确权的协作机制,有效弥合了底层特征与高层语义之间的鸿沟,建立了可信的决策链条。其核心思想在于:算法输出必须经过领域专家的语义映射,方能转化为高风险场景下的可执行的决策依据。

在大规模 IT 运维与高端制造领域,面对海量日志与缺陷数据,全量人工标注成本极高。主动学习框架通过不确定性采样与边界样本优先选择,使少量高价值样本能够有效驱动模型优化。Human-in-the-Loop Machine Learning 相关研究阐述了这一范式,强调通过人机协作提升标注效率与决策边界刻画能力。该机制说明,在数据规模庞大且标注成本昂贵的场景中,精确筛选高信息增益样本,是实现模型持续优化的重要方式。

在复杂流程工业与智能制造的状态监测中,设备运行工况多变,且常伴随较强的工业背景噪声。纯数据驱动的异常检测模型往往缺乏对物理规律的先验认知,容易将设备合法的瞬态波动,如负载突变或加工模式切换,误判为异常状态,从而引发较多误报。为应对这一问题,近年来工业物联网运维框架逐步引入交互式可解释人工智能与专家确权机制^[77]。现场运维工程师依据实践经验对模型识别出的伪异常进行修正,并将确权后的真实故障模式作为先验知识反馈至知识库^[78]。这种融合专家物理经验与机器高维特征提取能力的双向交互范式,不仅有助于减少由动态工况引发的误报警,还推动了工业现场隐性经验向数字化模型的转化。

在预测性维护领域,由于设备生命周期较长、不可避免会面临老化和概念漂移问题,模型需具备适应渐进漂移的能力。对比学习框架为特征空间的微调提供了很好的数学基础,使模型无需重新整体训练,而可借助少量增量反馈样本完成局部表征空间的调整。这种基于反馈的持续演化机制,使监测系统能够随着设备状态变化不断更新。

上述跨领域实践表明,人机协同并非简单依赖人工干预,而本质上是一种结构化的反馈与迭代机制。通过将专家的语义认知与物理经验引入模型优化过程,可以有效弥补无监督学习在语义解释与持续适应方面的不足。这也为后续构建兼具自适应能力与可解释性的科研仪器监测体系提供了重要参考。

4.1.3 人机协同闭环监测框架设计

综合前述跨领域经验可以看出,在复杂多样的科研仪器运维场景中,不宜将无监督模型直接作为最终决策依据。更合理的方式是将其作为前端感知与筛选模块,用于挖掘高维数据结构并识别高风险样本。在此基础上,结合第三章提出的 QA-TF-CL 算法,本章构建了一套人机协同闭环监测框架。该框架涵盖异常分流、专家确权、知识沉淀与模型演化等关键环节。

整体业务流转逻辑大致分为四个环节,具体技术路线如图 4.1 所示:

(1) 异常发现与自适应分流：该环节主要依靠 QA-TF-CL 引擎计算得出的联合异常得分，并结合自适应双阈值机制进行分级过滤。系统对落入高风险区间的样本自动触发告警，而对处于灰区的疑似异常样本则转交人工审核，从源头上控制无效告警数量，减轻专家运维压力。

(2) 状态可视化与专家确权：被筛出的异常样本在底层通常表现为抽象的高维特征表示。系统通过流形降维技术将其映射为直观的拓扑散点分布，并提供融合多模态上下文信息的交互界面。专家依据可视化结果与波形信息，对样本对应的物理故障类别进行判断，从而完成异常确权。

(3) 知识沉淀与辅助决策：专家给出的每次诊断反馈都会被结构化处理并沉淀到故障知识库。该知识库不仅用于记录历史案例，还能够在未来遇到相似疑似异常时，基于特征相似度检索并自动向专家推送最接近的 Top-K 历史案例，从而形成机器辅助决策、专家反向指导模型的协同模式。

(4) 模型闭环增量演化：随着知识库中确权标签的不断积累，系统进一步利用难例挖掘策略筛选容易引起模型混淆的高价值样本，并基于这些样本通过有监督对比微调机制对底层特征提取器进行局部更新。由此，模型能够逐步修正原有表征盲区，并在仪器长期运行引发基线漂移时保持持续适应能力。

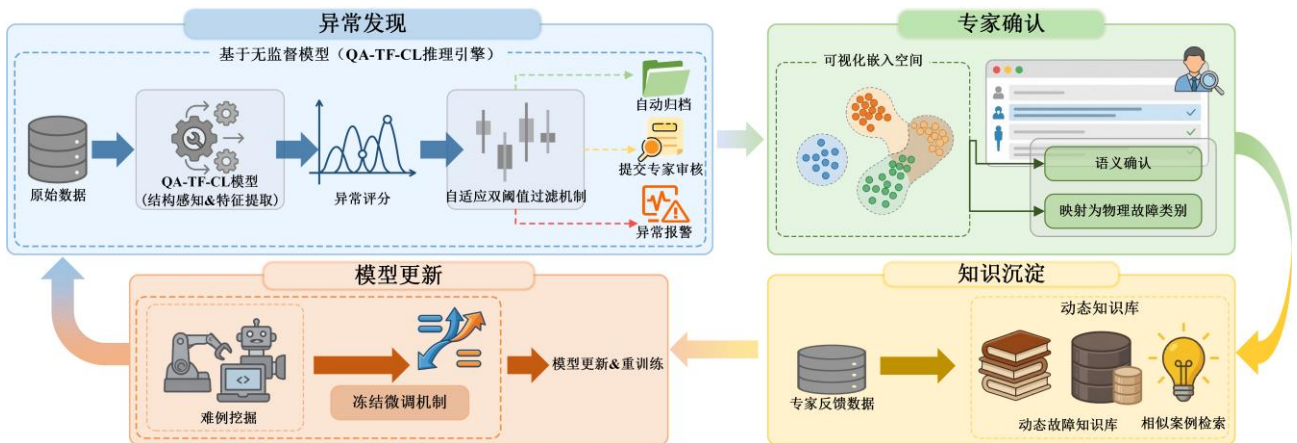


图 4.1 人机协同闭环监测技术路线图

4.2 无监督状态流形的语义重构与可视化

在监测系统上线初期，面对大量缺乏标注的历史数据，第三章提出的 QA-TF-CL 模型虽能够提取出具有良好聚类结构的 128 维融合特征，并将其划分为不同的数字簇（如 Cluster 0、Cluster 1），但这些底层特征表示与运维专家的工况认知之间仍存在明显的语义鸿沟。运维人员难以直接理解纯数学特征向量所对应的具体物理工况。

为缓解这一问题，本节构建了一套面向无监督状态流形的语义重构框架。首先，利用流形学习技术将高维特征映射为可视化的状态地图；随后，提出基于原型的批量语义传播策略，通过低成本的人机交互，将原本缺乏物理语义的数字簇转化为具有明确物理定义的标准工况

基线库。

4.2.1 状态流形降维与拓扑可视化

尽管 QA-TF-CL 输出的融合特征 $Z_{fused} \in R^{128}$ 在高维空间中已具备良好的类间分离度，但直接向运维人员展示高维特征向量既不直观，也难以辅助实际决策。因此，本系统引入流形学习技术，将高维特征空间映射为可视化的二维平面，同时尽可能保留样本间的局部邻域结构^[79]。

在降维算法的选型上，本文对比了 PCA^[80]与 t-SNE^[81]。传统的线性降维方法如 PCA 更侧重于保留数据的全局方差，但在处理非线性流形结构时，往往会导致局部邻域信息的丢失与扭曲。考虑到科研仪器的运行状态在高维空间中呈现复杂的非线性拓扑分布，例如同一故障模式在不同负载条件下形成的轨迹扭曲，线性投影难以有效分离不同性质的状态簇。相比之下，t-SNE 通过构建高维与低维空间中点对的概率分布，并最小化两个分布之间的 Kullback-Leibler 散度，能够较好地保持数据的局部邻域结构。这意味着，在原始特征空间中相似的仪器状态，在二维可视化平面上仍倾向于聚集，而差异较大的状态则会保持较明显的分离^[82]。因此，本文最终选用 t-SNE 作为状态映射的核心算法。

在此基础上，为了自动识别二维流形地图中的高密度区域，本文进一步引入 DPC 算法^[83]。

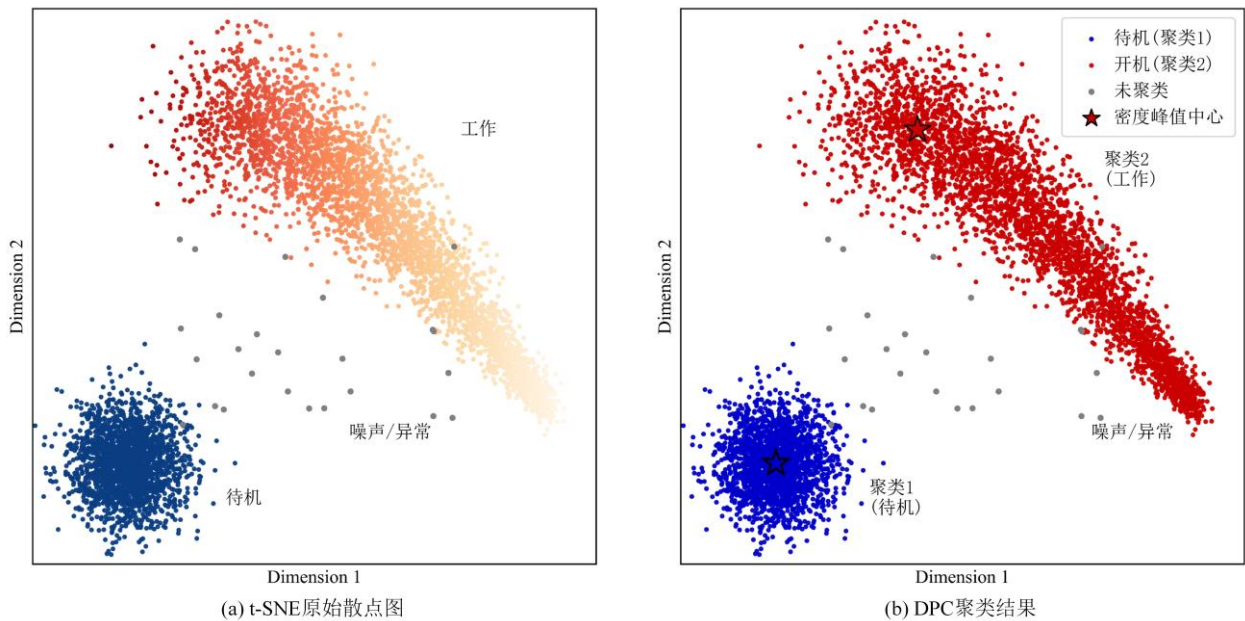


图 4.2 t-SNE 散点图与 DPC 聚类效果图

结合图 4.2 的可视化结果可看出，经过 t-SNE 映射后，历史监测数据在二维平面上形成了较为明显的岛屿状分布。其中，位于左下角的深色高密度簇呈紧密聚集状态，对应仪器最常见的待机状态，其分布方差较小，表明该状态下信号特征稳定；而右上角的点簇虽然聚类，但呈现出长条状的拖尾形态，反映了仪器在预热过程中电流随温度逐渐爬升的动态演变特性。此外，在两个主簇之间存在少量稀疏的散点，这些通常对应于仪器状态切换时的过渡数据或

偶发的噪声干扰。这种基于流形映射的可视化状态地图，直观地将抽象的数学特征转化为专家可感知的物理状态分布。

4.2.2 基于原型的批量语义传播策略

在获得无监督的状态簇后，核心挑战在于如何赋予这些数字簇具体的物理语义。针对传统全量人工标注成本过高、效率低下的问题，本文提出了一种基于原型的批量语义传播机制。该机制的核心思想是利用聚类结构的同质性，通过标注少量具有代表性的样本来推导整个簇的语义。

语义传播的具体流程始于典型样本的推荐。对于每一个由 DPC 识别出的聚类簇，系统首先自动计算其密度峰值中心，并依据欧氏距离筛选出距离中心最近的 k 个样本作为该状态的原型代表。关于 k 值的选择，若 k 过小（如 $k=1$ ），可能因单一样本的偶然偏差而导致语义判断不稳定；若 k 过大则增加了专家的阅读负担。综合考虑代表性与交互成本，本文设定 $k=3$ ，即选取最核心的三个样本作为原型集合。

随后进入专家语义赋予阶段。运维专家通过交互界面查看这几个典型样本对应的原始电流波形。由于这些样本位于特征空间的高密度核心区，较少受到边缘噪声的干扰，其波形往往具有较高的一致性。专家可以快速确认其物理含义，并将相应的语义标签赋予该簇。

最后执行标签自动扩散。基于聚类内部连通性与同质性假设，系统将专家赋予的语义标签自动传播至同一簇内的全部样本，从而完成历史数据的批量语义化标注，实现由少量原型样本到整个状态簇的知识扩展。

表 4.1 不同语义标注策略对比

标注策略	交互方式	交互成本	样本覆盖率	使用场景
全量人工标注	逐个样本确认	极高	100%	数据量较小且精度要求高
随机抽样标注	随机抽取并扩散	高	较低（易遗漏小簇）	数据分布均匀
原型语义传播	仅确认簇中心	低	近 100%	海量数据且存在明显聚类结构

为了量化评估该策略的优势，表 4.1 对不同语义标注策略的交互成本与覆盖能力进行了比较。结果表明，原型语义传播策略仅需少量专家交互，即可实现对海量历史数据的批量语义化标注，从而降低知识注入成本，并缓解冷启动阶段标注不足的问题。

4.3 专家反馈驱动异常确认机制

在上节中，系统通过离线冷启动构建了具备物理语义的标准工况基线库。然而，在进入在线监测阶段后，单纯依靠无监督模型进行异常预警仍存在局限。科研仪器的运行环境复杂多变，动态工况与瞬态扰动易引发模型的误报。为了过滤误报并赋予数据确定的物理属性，本节设计了专家反馈驱动的异常确认机制，通过自适应分级告警与人机交互接口，构建了从算法预测到专家确认的闭环流程。

4.3.1 基于统计分位数的自适应双重阈值

针对第三章提到的单阈值判定机制在在线监测中的局限性，为减弱人工设定阈值带来的主观性，本系统提出一种基于数据驱动的自适应双重阈值策略。该策略依据模型在验证集正常样本上的评分分布统计特性，动态划定正常、疑似与确信的决策边界。

延续 3.4.2 节中基于百分位统计刻画异常评分边界的思路，考虑到验证集正常样本的异常评分分布呈现明显长尾特征，本节在原有单一基准上界的基础上，重新界定两个关键的决策边界。

(1) 新增疑似异常阈值 (T_{low})。该阈值被锚定在基准分布的 95%分位数处，即 $T_{low} = \text{Percentile}(S_{val}, 95)$ 。该阈值在物理上定义了正常工况的边界，意味着在标准运行环境下，仅有少量由随机扰动引起的样本会超过该界限。凡是落在 $(T_{low}, T_{high}]$ 区间的样本，虽偏离了稳态中心，但尚未达到确信故障的程度，因此被定义为“灰色不确定区”，系统将此类事件触发黄色预警并推入人工审核队列，而非直接报警。

(2) 保留并重定义高置信度异常阈值 (T_{high})。该阈值等同于第三章中的 τ_{th} ，设定为分布的 99%分位数。该阈值对应统计意义上更严格的异常边界，超过该阈值的评分通常表明样本已明显偏离正常分布。此类样本不再需要人工确认，直接判定为“确信异常”并触发红色告警，以确保对严重事故的及时响应。

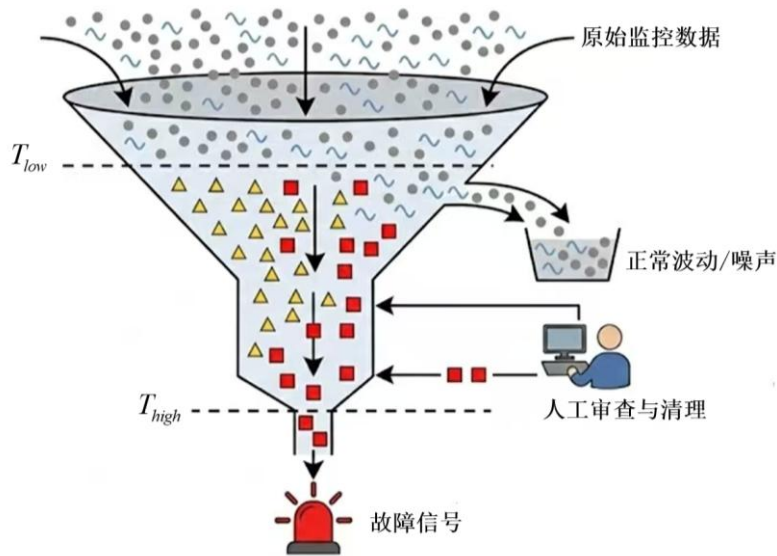


图 4.3 双重阈值机制

结合图 4.3 所示，通过引入这种双重阈值机制，系统有效地构建了一个漏斗形的过滤体系：绝大多数背景噪声被 T_{low} 屏蔽，模糊态数据被拦截在中间层进行人机交互清洗，而高置信度故障信号则畅通无阻地触发告警。该设计在保证较高异常检出率的同时，有效缓解了运维人员的告警疲劳问题。

4.3.2 相似案例检索机制

对于被拦截在灰色不确定区的疑似异常样本，专家介入的效率与准确性往往成为影响系

统整体处理效率的重要因素。为了打破依赖个人经验的盲目排查模式，系统引入了相似案例检索机制，向专家智能推送高度相似的历史故障案例以辅助决策。

传统的工业时序检索通常采用基于原始波形的模板匹配方法（如动态时间规整算法，DTW）。然而，DTW 算法的计算复杂度为 $O(N^2)$ ，在面对海量分钟级长序列时不仅计算开销极大，且极易受到局部插值噪声和相位偏移的干扰，难以满足大规模设备集群下的实时响应需求。

为了实现高效检索，本系统摒弃了低效的原始波形匹配，转而利用前文提炼的深层表征，构建时频语义多模态混合索引架构。该架构将检索过程解耦为硬约束过滤与软匹配排序两个层级。

（1）基于元数据的物理隔离

科研仪器具有极强的物理异构性，不同类型设备即使在电流波形上偶然表现出相似的畸变，其底层的物理故障机理也截然不同。为了避免跨设备的语义混淆，系统必须建立物理隔离边界。本层抽取知识单元中的背景信息（包含设备类型、设备编号等），在关系型数据库中构建基于 B+树的标量倒排索引。在检索触发时，该层作为前置过滤器，能够快速将搜索空间从全局知识库缩减至当前设备的同构子集中，确保后续推荐结果的物理一致性。

（2）基于矩阵加速的特征匹配

在经过元数据过滤后的同构候选集中，系统不再比对原始波形，而是直接提取 QA-TF-CL 模型输出的融合特征向量作为核心检索指纹。采用精确最近邻搜索策略。对于查询样本的特征指纹 Z_{query} ，其与候选集中历史向量 Z_i 的欧氏距离平方，可通过代数展开转化为矩阵内积形式：

$$\|Z_{query} - Z_i\|_2^2 = \|Z_{query}\|_2^2 + \|Z_i\|_2^2 - 2Z_{query}^\top Z_i \quad (4.1)$$

设该候选集中包含 M 个历史记录，系统将候选集中的 M 个特征向量预先加载并拼接为特征矩阵 $Z_{mat} \in \mathbb{R}^{M \times 128}$ 。查询时，利用基础线性代数子程序库（BLAS），执行并行的矩阵-向量乘法运算：

$$D_{base}^2 = \text{broadcast}(\|Z_{query}\|_2^2) + \text{diag}(Z_{mat} Z_{mat}^\top) - 2Z_{mat} Z_{query} \quad (4.2)$$

通过这种矩阵形式的并行计算机制，系统在 $O(M \cdot d)$ 的线性时间复杂度下完成了高维特征的批量比对。

（3）基于时间衰减的加权推荐

考虑到科研仪器存在随服役周期老化的规律。即使是同一种物理故障，设备在不同时期的基线电流特征也可能存在微小的概念漂移。如果仅依赖纯特征距离进行排名，可能会导致早期的陈旧案例排名靠前，降低参考价值。为了优先推荐近期发生的、更符合当前设备状态的案例，本节在基础欧氏距离之上，引入基于时间衰减的惩罚机制，构建加权欧氏距离评估模型。对于候选集中的历史案例 i ，定义其最终推荐检索得分 $Score_i$ 为：

$$Score_i = D_{base} \cdot [1 + \gamma \cdot \ln(1 + \Delta t_i)] \quad (4.3)$$

其中, $\Delta t_i \geq 0$ 表示历史案例 i 的入库时间距今的时间差。 $\gamma > 0$ 为时间衰减调节系数, 用于控制时间权重的惩罚力度。

在完成评分计算后, 系统依据 $Score_i$ 对候选集进行升序排列, 截取得分最小的前 K 个案例。随后, 系统提取这 K 个案例对应的知识单元内容, 并将其设备编码及专家历史确权标签展示在交互界面中。

4.3.3 上下文增强的交互式反馈接口

在完成特征检索机制的构建后, 系统通过一个上下文增强的交互式反馈接口, 将底层算法逻辑与前端专家决策无缝拼接。

当专家处理灰色不确定区的告警时, 交互界面不仅展示当前疑似样本的原始电流波形与设备身份元数据, 还会同步渲染 3.3.4 节引擎实时推送的 Top-K 相似历史案例 (包含历史波形及其最终专家确权结论)。这为前端决策提供了一个动态的参考坐标系, 有助于辅助经验不足的运维人员做出更接近专家经验的判断。

当专家判定该样本为真实故障时, 可直接通过下拉菜单选择标准的故障类型标签。对于未曾记录过的异常模式, 专家还可手动输入新的故障描述及处置建议。生成的 (Z_i, y_{fault}) 数据对是构建故障知识库的核心来源。当专家判定该波动属于正常工况时, 下拉菜单选择排除。系统将对样本进行标记, 在后续模型更新中, 算法将学习抑制对此类特征模式的响应, 从而降低误报率。针对数据丢失严重或特征极其模糊无法判断的样本, 选择忽略。将该样本剔除出训练集, 防止低质量数据污染模型。

这种持续注入的领域知识, 使系统能够逐步积累具有明确物理语义的故障概念, 而不再仅停留于抽象异常评分层面。

4.4 动态故障知识库的构建与维护

在上节中, 系统通过人机协同交互获取了专家对疑似异常样本的语义裁决。然而, 离散的反馈信息若缺乏系统化组织与存储, 便难以沉淀为可复用的领域知识。因此, 本节进一步构建一个基于标注样本的动态故障知识库 (Fault Knowledge Base, FKB)。该知识库不仅用于将专家隐性经验转化为标准化的显性数据资产, 还通过动态清洗机制保障存储内容的准确性与时效性, 从而为后续的案例检索与模型演化提供可靠的数据基础。

4.4.1 故障知识单元

专家的每一次反馈, 本质上都是一次高价值的知识注入。系统将这些反馈信息实时结构化, 构建为故障知识单元, 作为后续模型演化的核心支撑。

对于每一个经过专家确认的样本, 系统生成如下记录:

$$\text{记录} = \{\text{原始波形, 特征指纹, 语义标签, 背景信息}\} \quad (4.4)$$

其中，原始波形完整保留采样周期内的时序电流数据。其作用不仅在于支持后续运维人员对故障现场进行可视化回溯，更在于当特征提取算法(QA-TF-CL)升级或需进行离线重训练时，能够提供最底层的原始数据支撑，确保了数据资产的向后兼容性。特征指纹指由 QA-TF-CL 编码器输出的融合特征向量 (Z_{fused})。该向量表征了信号的主要物理特性，不包含外部上下文信息，有利于保证数据资产的独立性与复用性。语义标签 (y_{label}) 表示由领域专家给出的最终诊断结论，可为具体故障类型或正常/误报等判定，是整条记录中承载核心语义的信息。背景信息则包含样本产生时的相关背景与运行上下文信息，如设备编号、时间戳及关键运行参数配置等，这些信息为后续的故障归因分析和跨场景对比提供了必要的辅助约束。通过这种结构化封装方式，专家经验不再局限于隐性知识形态，而是持续沉淀为可检索、可演化、可复用的显式知识单元，从而为后续模型的自适应更新与故障知识库构建奠定基础。

4.4.2 知识库自适应清洗

由于人的判断并非绝对无误，不同专家对同一波形的理解可能存在偏差，甚至同一专家在不同时期的判断标准也可能发生变化。为了减少错误经验进入知识库，系统引入了自适应清洗机制。

当新的知识单元入库时，系统会自动计算其与库中同类样本特征中心之间的距离。若某一样本被标记为正常，但其特征指纹却落入历史故障样本形成的高密度特征簇内部，系统将触发标签冲突预警，并提示专家进行二次复核，以尽可能从源头减少误判带来的错误标注数据。

对于历史久远的知识单元，系统引入时间衰减因子。设某历史知识单元距当前时刻的时间差为 Δt ，定义其在后续检索与微调训练中的有效权重 $w(\Delta t)$ 为：

$$w(\Delta t) = \exp(-\lambda \cdot \Delta t) \quad (4.5)$$

其中， λ 为依据设备生命周期设定的衰减率超参数。

通过该机制，当设备特征模式发生演变时，系统能够优先利用近期录入且更符合当前设备物理状态的知识单元，而早期陈旧知识的权重则随时间呈指数式平滑衰减。由此，知识库能够保持持续更新能力，从而保障在线演化模型的时效性与准确性。

4.5 基于难例挖掘的对比微调策略

现有的工业监测系统大多采用静态部署模式。一旦 QA-TF-CL 模型的参数被固化，它对物理工况的适应能力便随之受限。在实际长期运行中，科研仪器不可避免地会因自然磨损、传感器老化或外部环境参数（如季节性室温变化）的改变，产生新的正常基线波动。面对这些未知但合理的物理模式，静态的无监督模型会容易地将其推离正常流形并给出异常判定，从而引发持续性的误报。

在这种局面下，仅依靠前端运维人员在界面上的多次排除并不能从根本上解决问题。系统若无法从交互中完成持续学习，同样的误判必然反复上演。为了打破模型长期运行带来的静滞瓶颈，本文提出了一种基于难例挖掘的有监督对比微调机制。该机制旨在利用 4.4 节沉淀的高质量知识单元，在不破坏已建立的无监督特征空间拓扑的基础上，针对极易引起误判的难例进行局部表征优化。

4.5.1 难例挖掘与有监督对比损失构建

(1) 难例定义

在实际运行中，并非所有的反馈数据对模型演化都具有同等价值。那些位于决策边界附近、容易混淆的样本通常蕴含更高的信息增益。系统从动态故障知识库中定向挖掘以下两类高价值知识单元：

第一类为困难正样本 (Hard Positives)。该类样本在底层推理阶段产生了较高的异常评分，并被算法拦截，但经专家确权属于正常运行条件下的基线漂移。在几何特征空间中，这类样本被静态模型错误地判定为偏离正常流形。若不进行参数更新，系统可能对其持续产生误报。因此，微调的目标是将其产生的特征向量重新强行拉回正常簇的密度中心 C_{norm} 。

第二类为困难负样本 (Hard Negatives)。该类样本在模型中处于灰色不确定区，但专家确认其为真实的隐匿性故障或早期故障。若这类样本未能被模型有效区分，则可能增加漏报风险。因此，微调的目标是增强其与正常簇之间的分离程度，并使其特征表示逐步向对应故障簇中心 C_{fault} 靠近。

(2) 有监督对比损失函数构建

在第三章的无监督预训练阶段，由于缺乏真实的物理标签，QA-TF-CL 模型须依赖时间衰减权重 w_{time} 和数据质量掩码 $w_{quality}$ 来构建软置信度，以避免错误地推远同类样本或拟合噪声。然而，在人机协同的在线演化阶段，系统已获取了专家反馈的确定性标签 y_{label} 。专家给出的语义裁决具有更高的监督可信度，因此可以替代第三章中基于无监督数据驱动的时质双重软权重机制。

为此，本文在原有 TQ-NCE 投影空间的基础上，进一步构建了基于专家反馈的有监督对比微调损失 $\mathcal{L}_{fine-tune}$ 。

假设系统在近期的交互中，从动态故障知识库中提取了一个包含 M 个难例样本的微调批次：

$$B_{expert} = \left\{ \left(X_{time}^{(i)}, y_{label}^{(i)} \right) \right\}_{i=1}^M \quad (4.6)$$

为了执行前向传播，系统首先复用预处理模块，将知识单元中的原始时序波形 $X_{time}^{(i)}$ 重新计算并包装为包含时频双视图的标准化输入实例。随后将其送入 QA-TF-CL 网络，经过双流编码器与门控融合后，为保持与无监督预训练阶段几何空间的一致性，再通过非线性投影头映射为对比空间中的特征表示 $z_i = g\left(Z_{fused}^{(i)}\right)$ 。

对于微调批次 B_{expert} 中的任意锚点样本 i ，定义同属类别 $y_{\text{label}}^{(i)}$ 的样本索引集合为 $P(i)$ ，即正样本集包含批次中所有被专家赋予相同语义标签的样本，其余不同标签的样本为负样本集。有监督对比微调损失定义如下：

$$\mathcal{L}_{\text{fine-tune}} = \sum_{i=1}^M \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\text{sim}(z_i, z_p) / \tau)}{\sum_{k=1}^M \exp(\text{sim}(z_i, z_k) / \tau)} \quad (4.7)$$

其中， $\text{sim}(z_i, z_p)$ 为余弦相似度函数， τ 为温度超参数。

该损失相较于第三章无监督损失 $\mathcal{L}_{\text{contra}}$ ，其改进主要体现在以下两个方面。

首先，降低了假负样本的干扰。第三章中的无监督损失主要依赖时间衰减权重 w_{time} 在时间维度上对邻近样本之间的排斥关系进行软调节；而本公式通过确定的集合 $P(i)$ ，强制锚点 z_i 与所有同属类别 $y_{\text{label}}^{(i)}$ 的确权样本 z_p 拉近，即便它们在时间上相隔数月。该机制能够将被专家纠正为正常的误报难例（困难正样本）重新拉回正常流形附近。

其次，增强了对困难负样本的分离约束。本公式分母部分对所有语义标签不同的样本进行显式区分。特别是对于那些被模型误判为灰区、但经专家确认为真实故障的难例（困难负样本），该损失能够进一步增强其与正常样本之间的分离程度，从而促使其在特征空间中逐步远离正常流形，并形成更加清晰的故障簇结构。

4.5.2 冻结微调策略

在执行上述微调时，若利用有监督对比损失 $\mathcal{L}_{\text{fine-tune}}$ 对整个 QA-TF-CL 双流编码器进行全量梯度更新，容易引发灾难性遗忘现象。具体而言，模型可能为了拟合少量新增的难例样本，而显著改变底层特征提取参数，从而削弱其对大量历史正常波形的泛化表征能力。为兼顾在线演化能力与特征稳定性，本文在工程实现中采用冻结微调策略。

在演化阶段，系统冻结底层的 1D-ResNet 时域编码器与统计 MLP 特征编码器参数，仅对顶层的时频交互门控融合层与非线性投影头开放梯度，并采用极小的学习率（ 1×10^{-5} ）进行增量更新。

这种微调机制使模型能够在专家反馈的指导下持续修正其对边界样本的表征偏差。随着误报样本逐步被重新拉回正常簇附近，系统误警率有望逐渐下降，从而实现面向长期运行场景的闭环演化。

4.6 实验验证与分析

为了系统评估本章所提出的基于人机协同的闭环运维框架在实际工程场景中的有效性，本节在第三章 QA-TF-CL 算法性能验证的基础上，进一步围绕语义映射、异常确权及模型演化能力开展实验分析。

4.6.1 语义标注效率实验

本实验旨在评估 4.2 节提出的原型语义传播策略在处理海量无标签数据时的标注成本与

覆盖效果。实验选取了循环切换型仪器的 10000 个未标记的训练样本,通过 t-SNE 降维与 DPC 技术,系统自动识别出工作与待机两个主要状态簇。

(1) 评估指标定义

为了定量衡量语义映射的效率与质量,本实验引入以下三项核心指标:

标注压缩率 (Annotation Compression Ratio, ACR): 衡量人工介入成本的降低程度。定义为专家实际标注的原型样本数与待标注总样本数的比值,其公式为:

$$ACR = \frac{N_{anno}}{N_{total}} \quad (4.8)$$

其中, N_{anno} 为专家标注的原型样本数量, N_{total} 为待标注的原始样本总数。

语义覆盖率 (Semantic Coverage): 衡量通过标签扩散机制成功赋予物理语义的样本占比。其公式为:

$$Coverage = \frac{N_{covered}}{N_{total}} \quad (4.9)$$

其中, $N_{covered}$ 为通过原型扩散获得语义标签的样本数量。

标注准确率 (Label Accuracy, Acc): 衡量自动传播标签的正确性。以全量人工核验结果为基准,其公式为:

$$Acc = \frac{N_{correct}}{N_{covered}} \quad (4.10)$$

其中, $N_{correct}$ 为自动传播标签中与人工核验标签一致的样本数量。

(2) 实验结果分析

在实验中,设定每个簇仅提取 $K=3$ 个质心样本供专家审阅。将原型语义传播策略与全量人工标注、随机抽样标注进行了对比,结果如表 4.2 所示。

表 4.2 语义标注效率与准确性对比

标注策略	专家标注次数	ARC	Semantic Coverage	Acc
全量人工标注	10000	100%	100%	100%
随机抽样标注	100	1%	85.6%	91.2%
原型语义传播策略 ($K=3$)	6	0.06%	99%	98.5%

实验数据表明,在待机工作二元工况场景下,由于特征空间拓扑结构较为清晰,专家仅需分别对两个簇的 6 个质心样本进行语义赋予,即可实现对 10000 个历史样本约 99% 的语义覆盖。其标注压缩率仅为 0.06%,相较于全量人工标注,专家交互成本大幅降低。同时,由于质心样本位于簇内高密度区域,受边缘噪声干扰较小,因此该策略的标注准确率达到 98.5%,优于随机抽样标注方法。结果表明,原型语义传播策略能够实现由少量原型样本到整个状态簇的语义扩展,从而有效缓解科研仪器监测系统在冷启动阶段标签不足的问题。

4.6.2 双重阈值参数选择实验

本实验通过调节分位数参数 α ，观察系统性能指标的演变趋势，从而验证 4.3 节提出的自适应双阈值体系在全球视角下的科学性。

(1) 评估指标定义

为了刻画阈值连续变化对运维性能的影响，本实验定义了以下随分位数 α 滑动的函数指标：

滑动虚警率（FAR）：衡量随阈值 $T(\alpha)$ 提高，正常工况被误报为异常的比例：

$$FAR(\alpha) = \frac{FP(\alpha)}{FP(\alpha) + TN(\alpha)} \quad (4.11)$$

滑动漏报率（FNR）：衡量随阈值 $T(\alpha)$ 提高，真实物理故障未被检出的风险：

$$FNR(\alpha) = \frac{FN(\alpha)}{TP(\alpha) + FN(\alpha)} \quad (4.12)$$

其中， $TP(\alpha)$ 为实际异常且评分高于阈值的样本数， $FP(\alpha)$ 为实际正常但评分高于阈值的样本数， $TN(\alpha)$ 为实际为正常且评分低于阈值的样本数， $FN(\alpha)$ 实际为异常但评分低于阈值的样本数。

(2) 阈值灵敏度曲线分析

通过将分位数 α 在 $[0.80, 0.999]$ 范围内连续滑动，得到了图 4.4 所示的性能权衡曲线。

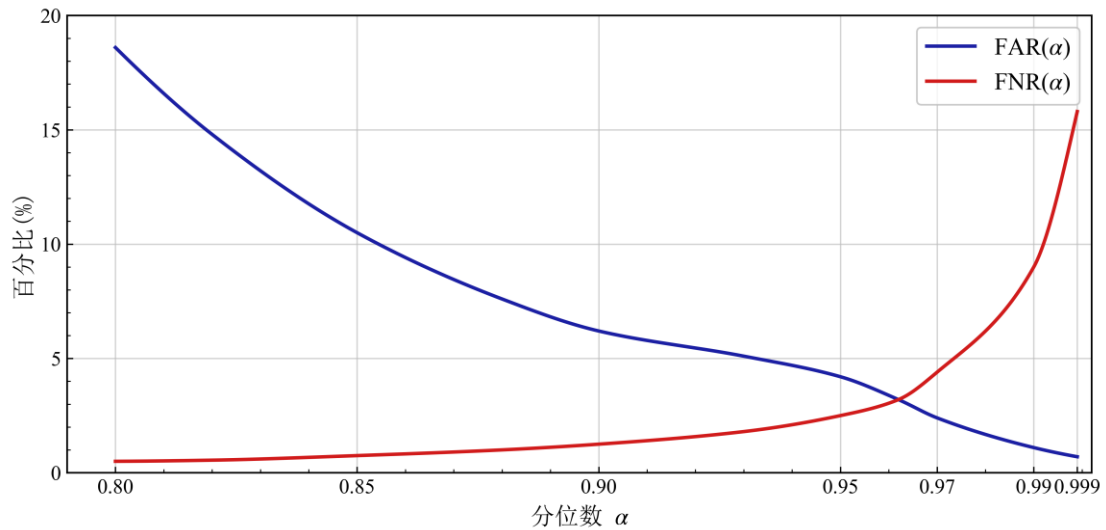


图 4.4 不同阈值性能变化曲线

分析结果表明，随着分位数参数的提高，误报率（FAR）持续下降，说明基于百分位统计的阈值设定能够较好地抑制背景噪声带来的误报。然而，当分位数超过 0.97 后，漏报率（FNR）的上升趋势明显加快，表明若单阈值设置过高，系统将面临更大的漏报风险。

基于灵敏度曲线的整体变化趋势，本文最终选取 $[0.95, 0.99]$ 作为双阈值的判定区间，旨在实现监测安全性与运维效率的最优平衡。其中，疑似异常阈值 T_{low} 设定为基准分布的 95%分位

数。这是因为该位置处于误报率下降趋势逐渐趋缓的拐点附近，能够在保持较低漏报风险的同时，有效过滤大部分背景随机扰动与传感器热噪声。高置信度异常阈值 T_{high} 设定为 99% 分位数，用于构建更严格的异常判定边界，使高风险故障样本能够直接触发红色告警，从而满足严重异常的快速响应需求。在此基础上，由双阈值区间形成的灰区有效拦截了一部分高不确定性样本。这类样本虽然已偏离稳态分布，但其物理语义尚不明确，因此更适合作为人机协同（HITL）审核的重点对象。一方面，这种分流机制有助于将运维人员的审核负荷控制在合理范围内；另一方面，这些经过专家确权的高价值样本也为后续的增量对比微调提供了可靠的数据来源。

4.6.3 模型在线演化与误报抑制实验

在科研仪器长期的物联网监测过程中，设备硬件老化或外部环境变化会导致基准状态发生漂移，从而使原本静态的无监督模型产生持续性误报。本实验旨在验证 4.5 节提出的基于难例挖掘的对比微调策略，考察专家反馈驱动的在线演化机制在抑制工况漂移引发的重复告警方面的作用，并进一步验证冻结微调策略在缓解灾难性遗忘方面的有效性。

（1）评估指标定义

误警下降率（False Alarm Reduction Rate, $FARR$ ）：用于衡量模型在接受 n 次专家确权后，对同类漂移工况的误报消除能力，其计算公式为：

$$FARR(n) = \frac{FP_{initial} - FP(n)}{FP_{initial}} \quad (4.13)$$

其中， $FP_{initial}$ 为基准偏移后初始产生的误报总数， $FP(n)$ 为经过 n 次专家确权并微调后的剩余误报数。

特征收敛距离（ D ）：衡量难例正样本在特征空间中向正常状态簇回归的程度。其公式定义为：

$$D_{conv} = \|z_{hard} - c_{normal}\|_2 \quad (4.14)$$

其中， z_{hard} 为被误报样本的融合特征向量， c_{normal} 为对应正常工况簇的质心。

基准准确率（Base Accuracy, Acc_{base} ）：衡量模型在接受在线微调演化后，对历史未发生漂移的原始正常与异常样本的整体识别保持能力，用于量化灾难性遗忘的严重程度。该指标在系统冷启动时划分的原始静态测试集上进行评估，其定义为：

$$Acc_{base} = \frac{N_{correct}}{N_{total}} \quad (4.15)$$

其中， N_{total} 为原始静态测试集的总样本数， $N_{correct}$ 为模型微调后能正确识别正常波形与原始异常的样本数。

遗忘衰退幅度（ Δ ）：衡量演化前后基准准确率的下降绝对值：

$$\Delta = Acc_{base}^{(n)} - Acc_{base}^{(Init)} \quad (4.16)$$

其中, $Acc_{base}^{(Init)}$ 为模型在初始状态下的基准准确率, $Acc_{base}^{(n)}$ 为经过 n 次增量微调后的基准准确率。 Δ 的绝对值越小, 说明模型防灾难性遗忘的能力越强。

(2) 实验设置

为了模拟由设备老化或环境波动引起的概念漂移现象, 本实验在原始 1 分钟/点的电流序列中人为引入了 10% 的基准偏移。在此场景下, 由于部署后的静态模型参数固化, 无法动态响应环境演变, 导致其将偏移后的正常待机波形误判为异常, 在初始阶段产生了 184 次持续误报。随后, 运维专家利用上下文增强的交互式反馈接口对系统推送的疑似异常进行语义确认; 当专家将特定的波动判定为正常/排除时, 该样本即作为困难正样本被纳入微调批次, 触发在线演化任务。系统严格执行冻结微调策略: 在冻结底层 1D-ResNet 时域编码器与统计特征 MLP 编码器参数、以避免灾难性遗忘的基础上, 仅对顶层的时频交互门控融合层与投影头开放梯度, 并通过增量对比微调损失引导模型持续修正特征空间中的判别边界。

(3) 演化效率与误报分析

实验记录了误报数量和特征收敛距离随专家交互次数 n 增加的动态变化曲线, 如图 4.5 所示。

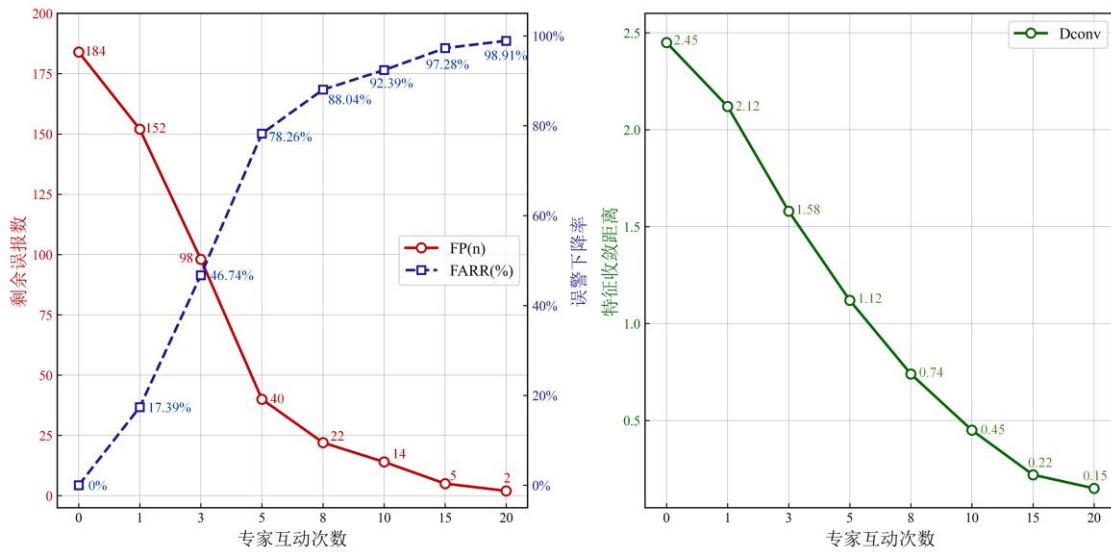


图 4.5 模型在线演化曲线图

实验数据表明, 系统展现出较强的在线学习能力: 在仅接受 5 次专家纠偏后, FARR 迅速跃升至 78.4%; 当交互次数达到 10 次时, FARR 超过 92%。这意味着对于同一类型的工况漂移, 专家只需进行极少量的点击确认, 模型即可自动修正对该类模式的认知偏见, 避免了同类误报的反复出现。与此同时, 特征收敛距离 D 从 2.45 持续下降至 0.15。该指标的减小从几何角度反映了漂移难例在隐空间中向正常稳态流形回归的过程。

(4) 灾难性遗忘抑制效果验证

神经网络在持续吸收新知识的过程中, 可能破坏原有参数分布, 从而削弱对历史常规工

况的识别能力，即产生灾难性遗忘。为验证本文所提出冻结微调策略在抑制灾难性遗忘方面的有效性，实验进一步补充了与全局全量微调策略的对比测试。模型在经历 10 次在线演化后，于原始静态测试集上的性能变化结果如表 4.3 所示。

表 4.3 冻结微调与全局微调策略的防遗忘性能对比

微调策略	$Acc_{base}^{(Init)}$	演化后 FARR	$Acc_{base}^{(n)}$	Δ
全局全量微调	96.16%	94.5%	81.34%	-14.82%
冻结微调	96.16%	92.39%	95.85%	-0.31%

从表中可以看出，虽然全局微调策略在消除新误报（FARR）上表现略高，但其对原有模型表征能力造成了较大影响。为了拟合少量漂移样本，底层特征提取参数发生明显变化，导致模型在历史静态测试集上的准确率下降近 15%，表现出较为明显的灾难性遗忘。相比之下，本文采用的冻结微调策略在冻结底层 1D-ResNet 与 MLP 编码器参数的基础上，仅对顶层可训练模块进行梯度更新，在保持较高误警抑制能力的同时，将原始测试集准确率的下降幅度控制在 0.31% 以内。

本组实验表明，通过引入人机协同驱动的闭环更新机制，系统能够在持续演化过程中有效缓解静态部署模型随时间退化的问题。模型不仅可以利用专家确权信息对局部工况漂移进行针对性修正，还能够在尽量保持原有泛化能力的前提下，实现对新变化模式的鲁棒适应。

4.7 本章小结

本章针对纯数据驱动算法在工程落地中面临的语义鸿沟与模型静滞问题，在第三章 QA-TF-CL 算法的基础上，构建了一套专家反馈驱动的人机协同（HITL）闭环演化框架。本章的主要研究工作概括如下：

（1）构建了状态流形的语义映射机制。利用 t-SNE 降维与 DPC 密度聚类技术，将高维无监督特征转化为直观可视的状态拓扑地图；并提出基于原型的批量标签传播策略，以极低的人工交互成本完成了对海量历史存量数据的语义重构，有效解决了系统冷启动阶段缺乏标准基线库的工程难题。

（2）设计了自适应异常确权与辅助检索。提出了基于统计分位数的双重阈值过滤体系，实现了漏斗式的高效分级拦截；同时，构建了检索架构，结合矩阵并行计算与时间衰减的加权评估模型，完成了 Top-K 相似案例的定位与推送。配合上下文增强的交互接口，有效降低了运维专家的认知负荷与报警疲劳。

（3）建立并维护了动态故障知识库。将前端专家的确权反馈结构化为包含多维物理与特征信息的故障知识单元，并引入了基于特征空间密度的冲突预警与基于时间衰减的权重平滑双重清洗机制，从源头防范了人为误差与设备概念漂移对知识库数据基础的污染。

（4）提出了基于难例挖掘的模型闭环增量演化策略。针对设备基线漂移导致的持续误报问题，定向提取知识库中的困难正负样本，以专家的确权标签为强监督信号，构建了增量对

比微调损失函数。通过冻结微调机制，系统在防范灾难性遗忘的前提下，实现了底层模型参数的自适应局部优化。

本章设计的协同机制赋予了静态监测模型动态演化的能力。系统构建了包含异常发现、专家确认、知识沉淀、模型更新的完整业务链条。这一链条打破了传统算法与物理语义之间的技术屏障，显著提升了科研仪器在复杂、动态物联网环境下的监测鲁棒性与工程可用性。本章所提出的框架不仅验证了人机协同在工业异常检测领域的应用价值，也为构建可持续进化的智能运维体系提供了相应的理论依据与实践参考。

5 总结与展望

5.1 总结

本文针对科研仪器在分钟级低频采样条件下，因数据缺失、插值噪声以及模型语义鸿沟和静态适应性不足所引发的运维监测难题，系统开展了基于质量感知对比学习与人机协同的异常检测研究。通过构建鲁棒的特征提取体系与闭环在线演化框架，实现了从原始电流信号到专家知识沉淀的业务逻辑闭环。主要研究成果如下：

(1) 设计并实现了一套科研仪器智能运维监测平台，构建了全链路的硬件感知与软件交互体系。该平台通过非侵入式物联网传感技术实现了电流数据的边缘侧采集与聚合，并针对多源数据设计了基于 MongoDB 与 MySQL 的混合存储治理方案。在业务逻辑层，利用 Spring Boot 和 Python 完成了后端业务系统构建与跨语言通信机制，最终通过网页端与微信小程序提供直观的交互界面，为后续算法验证与工程落地奠定了可靠的平台基础。

(2) 针对分钟级电流数据存在的采样率低、噪声干扰及丢包等问题，本文提出了基于质量感知的时频双视图对比学习算法 (QA-TF-CL)。该算法引入了统计频域特征以弥补单一时域信息的不足，并通过构建数据质量掩码建立了时间质量加权的对比损失函数 (TQ-NCE Loss)。实验结果显示，QA-TF-CL 在无监督工况划分与异常检测中表现优异，其 F1 分数达到 91.74%；特别是在数据丢包率高达 50% 的极端测试场景下，该算法仍能维持稳健的检测性能，有效解决了传统模型在低质工业数据环境下的性能退化问题。

(3) 为缓解纯数据驱动算法面临的语义鸿沟与模型无法持续适应环境漂移的问题，本文构建了一套专家反馈驱动的人机协同闭环演化框架。该框架利用 t-SNE 降维与密度峰值聚类实现无监督状态流形的语义重构，并通过原型标签传播策略将人工标注成本压缩至 0.06%。同时，提出基于统计分位数的自适应双重阈值过滤体系，并结合案例检索机制，有效缓解了运维人员的告警疲劳，提升了异常确权效率。在此基础上，进一步提出基于难例挖掘的冻结微调策略使系统仅需约 10 次专家确权即可抑制 92% 以上因工况漂移引发的重复误报，并在更新过程中防范了模型对历史正常模式的灾难性遗忘。该框架打通了从算法异常发现、专家语义确权到知识沉淀与模型反向更新的完整链路，提升了科研仪器运维监测系统的工程可用性与长期可靠性。

综上，本文围绕科研仪器智能运维场景，系统完成了平台设计、状态与异常检测算法构建以及人机协同闭环演化机制研究。研究结果表明，所提出的方法能够在低频、弱标签、强噪声的复杂物联网环境下实现较可靠的异常检测与持续优化，对提升科研仪器智能监测的可靠性、可解释性与长期运行能力具有一定的理论意义与工程应用价值。

5.2 未来展望

尽管本文在科研仪器状态监测与异常检测方面取得了一定的研究进展，并构建了具备在线演化能力的人机协同框架，但受当前的探索范围与现实工程条件的限制，面向更高水平的智能运维应用，仍有若干值得进一步研究的方向：

（1）多模态物理特征融合与多维关联分析。本文目前主要聚焦于非侵入式采集的电流侧信道数据进行工况辨识。未来的研究可尝试整合振动、噪声、环境温湿度以及设备自身的系统运行日志等多模态数据，利用多模态融合学习技术挖掘不同物理量间的深层关联，以构建更加全面、精细的设备健康评估画像。

（2）算法性能上限评估与有监督基线对比。受限于真实科研仪器异常标签的获取难度，本文目前的对比实验主要聚焦于同等弱标签约束条件下的基线算法。未来的研究将考虑构建一个包含高精度专家标注的小规模标准化测试集，引入主流的深度有监督异常检测模型进行量化性能压测，以进一步探索本文所提的无监督模型与理想全监督模型之间的性能边界差距。

（3）隐私保护下的联邦学习架构。科研仪器数据往往涉及实验室核心工艺参数或保密科研项目，数据共享受到严格限制。未来的研究可探索将联邦学习引入智能运维体系，实现在不移动原始数据的前提下，通过各实验室间的模型参数交换来共同训练更强大的全局异常监测模型，解决单一单位数据孤岛导致的样本稀缺问题，提升系统对罕见故障的预警能力。

（4）大语言模型辅助的交互式智能运维。在人机协同环节，目前仍依赖运维专家对波形和特征进行语义判定。未来可探索引入大语言模型（LLM），利用其强大的推理与理解能力，将复杂的异常评分与技术指标转化为通俗易懂的运维诊断报告，并结合本文构建的动态故障知识库，实现自动化的维修策略推荐与智能问答交互等功能。

（5）边缘侧轻量化部署与实时在线演化。目前本文所提出的算法主要运行于服务端，对边缘侧采集终端的计算能力利用尚不充分。后续研究可致力于 QA-TF-CL 模型的轻量化压缩技术，实现在边缘网关处的实时推理与初步训练，构建边缘感知与微调、云端存储与全局演化的端云协同架构，以进一步降低通信延迟并提升系统的响应速度。

致谢

三年一瞬，聚散有时。行文至此，敲下“致谢”二字时，意味着我三年研究生生涯即将画上句点。回首这段求学之路，有探索未知的迷茫，也有攻克难关的喜悦。在这篇论文即将付梓之际，我要向所有在此期间给予我指导、帮助与支持的人，致以最诚挚的感谢。

首先，我要向导师陈科明教授致以最深切的敬意。从最初课题的选定、研究框架的搭建，到最后论文的撰写与修改，导师始终给予我悉心的指导与无私的帮助。导师严谨的治学态度、渊博的学术视野以及精益求精的学者风范，将成为我终身受益的宝贵财富。

其次，我要感谢实验室的师兄师姐、师弟师妹以及我的同窗好友们。我们曾在无数个日夜里共同探讨学术难题，在求职的道路上互相打气、共享经验。那些一起攻克难关、相互陪伴的日子，是我研究生生涯中最温暖、最充满力量的青春记忆。

我还要特别感谢我的父母与家人。二十余载的求学之路，是你们在背后默默倾注了无尽的爱与包容。无论是学业上的挫折还是面对未来的焦虑，你们始终是最坚实的后盾和最温暖的港湾。这份深情厚恩，无以为报，唯有以更加踏实的脚步去走好未来的人生之路。

最后，感谢这段岁月里那个未曾轻言放弃的自己。凡是过往，皆为序章。带着这份对知识的敬畏与对未来的期许，我将继续砥砺前行，迎接新的人生篇章。

故事至此，承蒙关照。最后祝大家都好，世界和平，山河无恙。

参考文献

- [1] 高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗：在中国共产党第二十次全国代表大会上的报告[EB/OL]. (2022-10-25)[2022-10-25]. https://www.gov.cn/xinwen/2022-10/25/content_5721685.htm.
- [2] 教育部印发《关于加强高校有组织科研 推动高水平自立自强的若干意见》[EB/OL].(2022-08-29)[2022-10-25].
http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202208/t20220829_656091.html.
- [3] DataHorizzon Research. Scientific Research Instruments Market Size, Outlook, Growth Statistics & Forecast 2033 [EB/OL]. DataHorizzon Research, 2025-03-22.
<https://www.skyquestt.com/report/scientific-instruments-market>.
- [4] IMARC Group. Global Scientific Instrument Market to Reach USD 68.78 Billion by 2033, Fostered by Increasing Research Activities in the Pharmaceutical Industry [EB/OL]. IMARC Group, 2024-11-29. <https://www.imarcgroup.com/global-scientific-instrument-market>.
- [5] 中研普华. 2025-2030 年中国科研仪器行业市场全景研究与趋势展望报告 [EB/OL]. 中国行业研究网, 2025-04. <https://www.chinairn.com/report/20250423/171250824.html>.
- [6] 于若溪,黄涛,张胜.高校科研仪器共享平台赋能科技创新发展的现实困境和优化路径——基于扎根理论的案例研究[J].实验技术与管理, 2025, 42(3):246-252.
- [7] 王力清,毛荣,刘青,等.行业高校大型仪器开放共享现状及原因分析[J].实验室科学,2015,18(04):181-184.
- [8] 姜晓满,文武,翟宇.跨校区办学模式下科研仪器开放共享探索与实践[J].实验技术与管理,2024,41(10):255-259.
- [9] 蒋玲玲,马兰,严冬琳,等.省属高校大型科研仪器开放管理的实践探索[J].实验室研究与探索,2025,44(10):240-245.
- [10] Jardine A, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance[J]. Mechanical Systems and Signal Processing, 2006, 20:1483-1510.
- [11] Isermann, R. Process Fault Detection Based on Modeling and Estimation Methods[J]. IFAC Proceedings Volumes, 1982, 15(4):7-30.
- [12] Venkatasubramanian V, Rengaswamy R, Yin K, et al. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods[J]. Computers & Chemical Engineering, 2003, 27(3): 293-311.
- [13] Qin S J. Survey on data-driven industrial process monitoring and diagnosis[J]. Annual Reviews

in Control, 2012, 36(2):220-234.

- [14] 张成,戴絮年,李元.基于 DPCA 残差互异度的故障检测与诊断方法[J].自动化学报,2022,48(01):292-301.
- [15] Lei Y, Jia F, Lin J, et al. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data[J]. IEEE Transactions on Industrial Electronics, 2016, 63(5): 3137-3147.
- [16] Zhao R, Yan R, Chen Z, et al. Deep learning and its applications to machine health monitoring[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68(5): 1526-1546.
- [17] Janssens O, Slavkovikj V, Vervisch B, et al. Convolutional Neural Network Based Fault Detection for Rotating Machinery[J]. Journal of Sound and Vibration, 2016, 377:331-345.
- [18] 雷亚国,贾峰,周昕,等.基于深度学习理论的机械装备大数据健康监测方法[J].机械工程学报,2015,51(21):49-56.
- [19] 吴定海,任国全,王怀光,等.基于卷积神经网络的机械故障诊断方法综述[J].机械强度,2020,42(05):1024-1032.
- [20] Malhotra P, Vig L, Shroff G, et al. Long Short Term Memory Networks for Anomaly Detection in Time Series[C]//23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015.2015.
- [21] Zerveas G, Jayaraman S, Patel D, et al. A Transformer-based Framework for Multivariate Time Series Representation Learning[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021:2114-2124.
- [22] Malhotra P, Ramakrishnan A, Anand G, et al. LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection[EB/OL]. [2016-07-01]. arXiv:1607.00148.
- [23] Su Y, Zhao Y, Niu C, et al. Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 2828-2837.
- [24] Li D, Chen D, Jin B, et al. MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks[C]//Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series, Part 4: 28th International Conference on Artificial Neural Networks (ICANN 2019), September 17–19, 2019, Munich, Germany. :Springer Nature Switzerland AG, 2019:703-716.
- [25] Audibert J, Michiardi P, Guyard F, et al. USAD: UnSupervised Anomaly Detection on Multivariate Time Series[C]//KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, 2020.
- [26] Chen T, Kornblith S, Norouzi M, et al. A Simple Framework for Contrastive Learning of Visual

- Representations[C]//37th International Conference on Machine Learning: ICML 2020, Online, 13-18 July 2020, Part 3 of 15. :Curran Associates, Inc., 2021:1597-1607.
- [27] He K, Fan H, Wu Y, et al. Momentum Contrast for Unsupervised Visual Representation Learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020), 13-19 June 2020, Seattle, WA, USA. :Institute of Electrical and Electronics Engineers, 2020:9726-9735.
- [28] Eldele E, Ragab M, Chen Z, et al. Time-Series Representation Learning via Temporal and Contextual Contrasting[C]// Advances in Neural Information Processing Systems (NeurIPS). Red Hook, NY: Curran Associates, 2021: 1416-1429.
- [29] Zhang X, Zhao Z, Tsiligkaridis T, *et al.* Self-supervised contrastive pre-training for time series via time-frequency consistency[C]// Advances in Neural Information Processing Systems. New York: Curran Associates, 2022: 3988–4003.
- [30] Feng Wang, Huaping Liu. Understanding the Behaviour of Contrastive Loss[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition: CVPR 2021, Virtual Conference, 19-25 June 2021, [v.4]. 2021:2495-2504.
- [31] Monarch R M. Human-in-the-loop machine learning: Active learning and annotation for human-centered AI[M]. Shelter Island: Manning Publications, 2021.
- [32] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead[J]. Nature Machine Intelligence, 2019, 1(5):206-215.
- [33] Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop?[J]. Brain Informatics, 2016, 3(2):119-131.
- [34] Settles B. Active Learning Literature Survey[R]. Madison: University of Wisconsin-Madison, 2009.
- [35] Wu D, Lin C T, Huang J. Active Learning for Regression Using Greedy Sampling[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(5): 1336-1347.
- [36] Iscen A, Tolias G, Avrithis Y, et al. Label Propagation for Deep Semi-supervised Learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 5070-5079.
- [37] Lu J, Liu A, Dong F, et al. Learning under Concept Drift: A Review[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 31(12): 2346-2363.
- [38] Gama J, Liobait I, Bifet A, et al. A Survey on Concept Drift Adaptation[J]. ACM Computing Surveys, 2014, 46(4): 1-37.
- [39] Parisi G I, Kemker R, Part J L, et al. Continual Lifelong Learning with Neural Networks: A Review[J]. Neural Networks, 2019, 113: 54-71.

-
- [40] Sun W, Xiao Y, Jin H, et al. On the Feasibility of Reasoning about the Internal States of Blackbox IoT Devices Using Side-Channel Information[EB/OL]. arXiv:2311.13761, 2023.
 - [41] Jia G, Zhu Y, Li Y, et al. Analysis of the Effect of the Reliability of the NB-IoT Network on the Intelligent System[J]. IEEE Access, 2019, 7:1-1.
 - [42] Samatas G G, Moumgiakmas S S, Papakostas G A. Predictive Maintenance—Bridging Artificial Intelligence and IoT[C]// 2021 IEEE World AI IoT Congress (AIIoT). 2021: 413-419.
 - [43] Sang G M, Xu L, Vrieze P D, et al. Predictive Maintenance in Industry 4.0[C]//ICIST '20: 10th International Conference on Information Systems and Technologies.2020.
 - [44] Erickson R W, Maksimović D. Fundamentals of Power Electronics[M]. Norwell, MA: Kluwer Academic Publishers, 2001.
 - [45] Horowitz P, Hill W, Rollefson R J. The Art of Electronics[M]. Cambridge: Cambridge University Press, 1980.
 - [46] Schreier R, Temes G. Understanding Delta-Sigma Data Converters[M]. Piscataway, NJ: IEEE Press; Hoboken, N.J.: Wiley, 2005.
 - [47] Ratasuk R, Mangalvedhe N, Zhang Y, et al. Overview of Narrowband IoT in LTE Rel-13[C]// Proceedings of the IEEE Conference on Standards for Communications and Networking (CSCN). 2016: 1-7.
 - [48] Ratasuk R, Vejlgaard B, Mangalvedhe N, et al. NB-IoT system for M2M communication[C]//2016 IEEE Wireless Communications and Networking Conference (WCNC). 2016: 1-5.
 - [49] Pozar D M. Microwave engineering[M]. Beijing: Publishing House of Electronics Industry, 2004.
 - [50] Rycroft M J. Understanding GPS: Principles and Applications[J]. Journal of Atmospheric and Solar-Terrestrial Physics, 1996, 59(5):598-599.
 - [51] 朱亚兴,余爱民,王夷.基于 Redis+MySQL+MongoDB 存储架构应用[J].微型机与应用, 2014, 33(13):4.
 - [52] Thakare A O, Tembhurne O W, Thakare A R, et al. NoSQL Databases: Modern Data Systems for Big Data Analytics[J]. International Journal of Electrical and Computer Engineering Systems, 2023, 14(2):147-156.
 - [53] Savitzky A. Smoothing and Differentiation of Data by Simplified Least Squares Procedures[J]. Analytical Chemistry, 1964, 36(8):1627-1639.
 - [54] Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: A strong baseline[C]//2017 International Joint Conference on Neural Networks (IJCNN). 2017: 1578-1585.
 - [55] Ekambaram V, Jati A, Nguyen N, et al. TSMixer: Lightweight MLP-Mixer Model for

- Multivariate Time Series Forecasting[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023:459-469.
- [56] Wu Z, Xiong Y, Yu S X, et al. Unsupervised feature learning via non-parametric instance discrimination[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 3733-3742.
- [57] Hjelm R D, Fedorov A, Lavoie-Marchildon S, et al. Learning deep representations by mutual information estimation and maximization[C]//International Conference on Learning Representations (ICLR). 2019.
- [58] Eldele E, Ragab M, Chen Z, et al. Time-Series Representation Learning via Temporal and Contextual Contrasting[C]//Advances in Neural Information Processing Systems (NeurIPS). 2021.
- [59] Chuang C Y, Robinson J, Yen-Chen L, et al. Debaised Contrastive Learning[C]//Advances in Neural Information Processing Systems (NeurIPS). 2020.
- [60] Oord A V D, Li Y, Vinyals O. Representation Learning with Contrastive Predictive Coding[EB/OL]. arXiv:1807.03748, 2018.
- [61] Tonekaboni S, Eytan D, Goldenberg A. Unsupervised Representation Learning for Time Series with Temporal Neighborhood Coding[C]//International Conference on Learning Representations (ICLR). 2021.
- [62] Yuan H, He Y, Du P, et al. Multi-Task Learning Using Uncertainty to Weigh Losses for Heterogeneous Face Attribute Estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2018: 7567-7576.
- [63] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786):504-507.
- [64] Um T T, Pfister F M J, Pichler D, et al. Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks[C]//Proceedings of the 19th ACM International Conference on Multimodal Interaction. New York: ACM, 2017: 216-220.
- [65] Park D S, Chan W, Zhang Y, et al. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition[C]//Proceedings of Interspeech. 2019: 2613-2617.
- [66] Kingma D, Ba J. Adam: A Method for Stochastic Optimization[C]//International Conference on Learning Representations (ICLR). 2015.
- [67] Loshchilov I, Hutter F. Decoupled Weight Decay Regularization[C]//International Conference on Learning Representations (ICLR). 2019.
- [68] Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with Warm

-
- Restarts[C]//International Conference on Learning Representations (ICLR). 2017.
- [69] Harold. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms[C]. Philadelphia: SIAM, 2007.
- [70] Qi S. Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection[C]//International Conference on Learning Representations (ICLR). 2018.
- [71] Yang Y, Zhang C, Zhou T, et al. DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023:3033-3045.
- [72] Gong D, Liu L, Le V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 1705-1714.
- [73] Hundman K, Constantinou V, Laporte C, et al. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2018:387-395.
- [74] Schmidl S, Wenig P, Papenbrock T. Anomaly Detection in Time Series: A Comprehensive Evaluation[J]. Proc. VLDB Endow., 2022, 15(9): 1779-1797.
- [75] Arthur D, Vassilvitskii S. K-Means++: The Advantages of Careful Seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January 7-9, 2007.ACM, 2007.
- [76] Zhang W, Li X, Li J, et al. A time–frequency contrastive learning model for anomaly detection in multivariate time series[J]. Complex & Intelligent Systems, 2025, 11(12):475-475.
- [77] Amaliah R N, Tjahjono B, Palade V. Human-in-the-Loop XAI for Predictive Maintenance: A Systematic Review of Interactive Systems and Their Effectiveness in Maintenance Decision-Making[J]. Electronics, 2025, 14(17):3384-3384.
- [78] Kulkarni A, Terpenney J, Prabhu V. Leveraging Active Learning for Failure Mode Acquisition[J]. Sensors, 2023, 23(5):2818-2818.
- [79] Roweis S, Saul L. Nonlinear Dimensionality Reduction by Locally Linear Embedding[J]. Science, 2000, 290(5500):2323-2326.
- [80] Wold S, Esbensen K, Geladi P. Principal component analysis[J]. Chemometrics & Intelligent Laboratory Systems, 1987, 2(1-3):37-52.
- [81] Laurens V D M, Hinton G. Visualizing Data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(2605):2579-2605.
- [82] Wattenberg M, Viégas F, Johnson I. How to Use t-SNE Effectively[J]. Distill, 2016, 1(10):1-14.

- [83] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191):1492.

附 录

作者在读期间发表的学术论文、专利及参加的竞赛和科研项目

专利：

[1]王语瑛，陈科明，余佳楠，程飞. 基于信号质量感知与时频特征融合的低温制冷机氦气泄漏工况识别方法[P], 浙江省: CN121502512A, 2026-02-10.

科研项目：

[1]高校实验室管理系统开发.横向课题.2023.10-2024.04

[2]大型仪器运行状态监测系统开发.横向课题.2024.03-2025.07

[3]一源供暖智能体系统开发研究.横向课题.2025.03-2025.07