

电子科技大学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

硕士学位论文

MASTER THESIS



论文题目 面向长尾分布数据的自监督学习故障
诊断方法研究

学科专业 电子信息

学 号 *

作者姓名 *

指导老师 * 教授

学 院 电子科技大学(深圳)高等研究院

分类号 _____ 密级 _____

UDC 注 1 _____

学 位 论 文

面向长尾分布数据的自监督学习故障诊断方法研究

(题名和副题名)

*

(作者姓名)

指导老师

* 教授

电子科技大学 成都

(姓名、职称、单位名称)

申请学位级别 硕士 学科专业 电子信息

提交论文日期 _____ 论文答辩日期 _____

学位授予单位和日期 电子科技大学 年 月

答辩委员会主席 _____

评阅人 _____

注 1：注明《国际十进分类法 UDC》的类号。

A Master Thesis Submitted to
University of Electronic Science and Technology of China

Discipline: **Electronic Information**

Author: *

Student ID: *

Supervisor: *

School: **Shenzhen Institute for Advanced Study,
UESTC**

摘要

随着工业系统的不断复杂化，系统中可能出现的故障类型也越来越多，导致故障诊断数据集常呈现“长尾分布”，即大多数样本集中在少数几类(头部类)，而其他多数类别(尾部类)样本稀少。这种不平衡会使模型倾向于识别头部类，而忽视尾部类。然而，尾部类常包含危害更大的罕见故障，因此在故障诊断中更具价值。为应对这一问题，自监督学习成为主流方法，其训练分为两个阶段：首先通过无标签数据进行特征提取预训练，再在有标签数据上进行微调分类。

本文旨在研究面向长尾分布数据的自监督学习故障诊断方法，主要内容包括：(1)针对简单暹罗孪生网络对比学习直接应用于故障诊断领域无法达到预期性能效果的问题，从理论上分析了简单暹罗孪生网络关键的数据增强模块所需具备的特质，以此提出了基于协方差矩阵适应进化策略的数据增强策略搜索算法。该算法通过在小规模数据上的迭代搜索得到一个最适用当前数据对象的增强策略参数，减轻了人工设计数据增强模块的时间成本。该算法训练的编码器层与 SOTA 算法 SimCLR 和 BYOL 相比有更加卓越的特征提取能力，并通过消融实验探讨了 SimCLR 原文中的“较大的 Batch Size 对对比学习模型性能有益”观点，发现与之不一致的趋势；(2)针对简单暹罗孪生网络在微调分类器层时更容易受到长尾效应的影响，从而分类准确率低于 SimCLR 的问题引入了 MARC 决策面校准算法。但 MARC 决策面校准算法并没有合理应用大规模无标签数据集中的信息，因此结合半监督学习的思想，将原模型预测的无监督数据的类别作为该数据的伪标签，并与原有标签数据集构成新的有标签数据集微调并校准网络。实验证明，本研究的方法在构建的长尾分布轴承故障诊断数据集上，成功超越了现有 SOTA 方法 SimCLR 和 BYOL，取得了更优的性能。

关键词：故障诊断，长尾学习，自监督学习，孪生网络

ABSTRACT

As modern industrial systems become increasingly complex, the number of fault types in a single system continues to grow, resulting in datasets that often exhibit a long-tailed distribution. Training models with long-tailed data introduces a significant class imbalance problem, and learning under long-tailed distributions is referred to as long-tail learning. Head classes typically refer to categories with a large number of samples, while tail classes refer to those with fewer samples. The goal of long-tail learning is to enable the model to perform well not only on head classes but also to effectively recognize data from tail classes. In this context, traditional intelligent fault diagnosis models often suffer from the issue of misclassifying tail-class samples as head-class samples. However, rare fault types in tail classes usually pose a greater threat to the system, making them more valuable in fault diagnosis.

Self-supervised learning has become a mainstream approach to addressing the long-tailed distribution problem. Its training process consists of two stages: self-supervised pre-training and fine-tuning. In the self-supervised pre-training stage, the model’s feature layers are trained using an unlabeled dataset and a proxy task. In the fine-tuning stage, a linear classification layer is attached to the feature layers and trained on a labeled dataset to build the final classification model. This thesis aims to study self-supervised learning-based fault diagnosis methods for long-tailed distribution data. The main contents include: (1) Data Augmentation Strategy Search Algorithm: Since directly applying a simple Siamese network-based contrastive learning method to the fault diagnosis domain does not achieve the expected performance, we theoretically analyze the essential characteristics required for the key data augmentation module in a simple Siamese network. Based on this, we propose a data augmentation strategy search algorithm using a covariance matrix adaptation evolution strategy (CMA-ES). This algorithm iteratively searches for the optimal augmentation strategy parameters on a small-scale dataset, reducing the time cost associated with manually designing augmentation modules. The encoder trained with this algorithm demonstrates superior feature extraction capabilities compared to state-of-the-art (SOTA) methods such as SimCLR and BYOL. Additionally, through ablation studies, we challenge the claim in SimCLR’s original paper that “a larger batch size benefits contrastive learning models.” (2) Semi-Supervised Decision Boundary Cal-

ABSTRACT

ibration: The simple Siamese network is more susceptible to the long-tail effect when fine-tuning the classifier layer, leading to lower classification accuracy than SimCLR. To mitigate this issue, we introduce the MARC decision boundary calibration algorithm. However, the MARC algorithm does not fully leverage the information from large-scale unlabeled datasets. To address this, we incorporate semi-supervised learning, where the model’s predictions on unlabeled data are used as pseudo-labels. These pseudo-labeled samples are then combined with the original labeled dataset to fine-tune and calibrate the network. Experimental results demonstrate that our proposed method outperforms existing SOTA methods, including SimCLR and BYOL, on a constructed long-tailed bearing fault diagnosis dataset, achieving superior performance.

Keywords: Fault Diagnosis, Long-Tail Learning, Self-Supervised Learning, Siamese Network

目 录

第一章 绪 论	1
1.1 研究工作的背景与意义	1
1.2 故障诊断基本流程	3
1.3 故障诊断的研究现状	4
1.4 长尾学习的发展历程和研究现状	5
1.5 自监督学习的发展历程和研究现状	6
1.6 对比自监督学习和孪生网络的发展历程和研究现状	7
1.7 本文的主要贡献与创新	8
1.8 本论文的结构安排	9
第二章 长尾学习及相关方法的理论基础与轴承故障数据集	10
2.1 自监督学习相关理论	10
2.2 孪生网络与对比学习相关理论	11
2.3 t-SNE 降维原理	11
2.4 K-Means 聚类原理	12
2.5 协方差矩阵适应进化策略原理	13
2.6 数据集介绍	15
2.6.1 凯斯西储大学 CWRU 数据集	15
2.6.2 帕德伯恩大学 PU 数据集	16
2.6.3 长尾数据集构造	18
第三章 基于暹罗孪生网络对比学习的自监督学习故障诊断方法	21
3.1 模型整体架构及其模块设计	21
3.1.1 简单暹罗孪生故障诊断网络整体结构	21
3.1.2 简单暹罗孪生网络对比学习	21
3.1.3 简单暹罗孪生网络的数据增强模块	26
3.1.4 实验设置	30
3.2 实验与分析	33
3.2.1 数据增强策略搜索的结果	34
3.2.2 在 CWRU 数据集和 PU 数据集上的实验结果	34
3.2.3 消融实验	36
3.2.4 准确率、损失函数和特征标准差之间的相关性分析	44

3.3 本章小结	48
第四章 半监督学习与边距调整改进的暹罗孪生故障诊断网络	51
4.1 模型架构	52
4.1.1 半监督学习与边距调整改进的暹罗孪生故障诊断网络整体架构	52
4.1.2 半监督学习模型架构	53
4.1.3 经半监督优化的决策面调整算法架构	54
4.1.4 实验设置	56
4.2 实验与分析	57
4.2.1 在 CWRU 数据集和 PU 数据集上的实验结果	57
4.2.2 消融实验	59
4.3 本章小结	62
第五章 全文总结与展望	64
5.1 全文总结	64
5.2 后续工作展望	65
致 谢	66
参考文献	67
攻读硕士学位期间取得的成果	72

第一章 绪论

1.1 研究工作的背景与意义

智能故障诊断是预测性健康管理 (Prognostics and Health Management, PHM) 系统中的核心组成部分，旨在实现设备故障的高效检测与精准识别，从而保障系统的安全与稳定运行。PHM 技术在各类关键工业装备中得到广泛应用，如复杂机械系统和高精度运转设备，具有重要的工程实践价值。智能故障诊断的有效性直接关系到生产过程的安全性、连续性与可靠性，具有重要的现实意义，例如旋转机械是现代工业中的核心设备，长期在高温、疲劳、重载等复杂工况下运行，若发生故障，不仅可能引发严重事故，还会造成巨大的经济损失，甚至威胁人员安全。传统的智能故障诊断方法主要依赖于信号处理技术进行特征提取，并结合机器学习算法实现故障类型的判别与分类。相关研究在特征工程与分类性能方面已取得了诸多进展。然而，面对异构且海量的数据时，基于人工设计的特征提取方法通常高度依赖领域知识，不仅开发周期较长，还对专家经验提出较高要求，限制了其在实际应用中的灵活性与可扩展性，而且人工提取的特征在某一特定工况下可能表现良好，但在其他工况或不同设备上可能失效，缺乏泛化能力。在高维特征空间中，许多人工构造的特征可能存在冗余性或与目标任务无关，进而导致分类器性能波动，甚至引发过拟合问题，影响模型的泛化能力。特征提取的难题不仅增加了诊断系统的复杂性和实施成本，还影响了故障诊断模型的稳定性、可靠性和适用性。因此，如何自动化、智能化地提取具有鲁棒性和泛化能力的故障特征，成为智能故障诊断领域亟待解决的重要问题。

现实中的故障检测数据通常呈现长尾分布^[1]，这是一种偏态分布，其中头部类包含大量正常数据，而尾部类则包含较少的故障数据。如何基于长尾分布的故障检测数据训练有效的故障检测模型，是一个具有现实意义的难点问题^[2]。长尾分布数据在现实世界中广泛存在，大规模数据集通常呈现出长尾分布的特性。长尾分布的学习是故障诊断领域中的一项常见挑战。例如，正常工作状态的数据通常远多于故障工作状态，或者常见故障状态的数据量远多于罕见故障状态。现实中的故障检测数据往往服从长尾分布，这种分布是一种典型的偏态分布，如图1-1所示：大规模训练数据中，头部类别的样本数量占据主导地位，而尾部类别的样本数量相对稀少，并呈现逐渐递减的趋势。长尾分布的数据常伴随显著的样本不平衡问题，导致训练过程中模型倾向于关注头部类别，从而削弱了对尾部类别的识别性能。如图1-2(a)所示，当样本分布均衡时，各类别在特征空间中具有清晰的区

分边界，并占据较宽广的特征空间。然而，当数据分布失衡呈现长尾分布时(如图1-2(b)所示)，尾部类别的特征分布变得狭窄，并依附于头部类别附近，从而扭曲了特征空间结构，降低了类别间的多样性与区分性。在使用长尾数据训练神经网络时，模型性能容易受到头部类别影响，导致尾部类别的表现较差。然而，对尾部类别样本的错分往往带来更大的实际损失，因此研究尾部类别样本具有重要的现实意义。如何基于长尾分布的故障检测数据有效训练故障检测模型，已成为一个具有重要价值的研究难题。

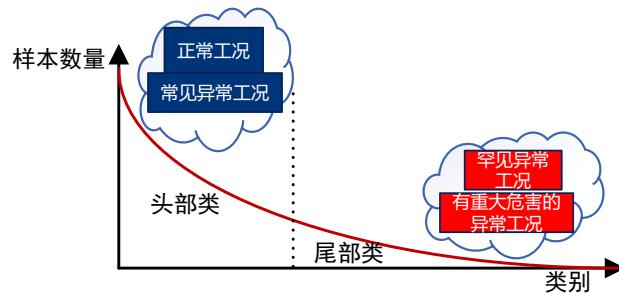


图 1-1 长尾分布示意图

长尾分布数据对现代深度学习框架提出了巨大的挑战。即使使用数据重采样、类平衡损失等专门技术，在极端类不平衡的情况下，模型性能仍然显著下降。因此，为了更有效地应对这一挑战，深入理解长尾学习对特征分布的影响至关重要。然而，与平衡数据的学习情景不同，长尾学习中的标签扮演着一个复杂而矛盾的角色，形成了标签价值的困境。一方面，有标签的监督学习通常比无监督学习能够训练出更准确的分类器，这突显了标签的积极作用；另一方面，长尾的标签分布会自然引入“标签偏差”，其中头部类别在很大程度上驱动决策边界的决策，导致对尾部类别的压制。这表明，标签既是推动学习的动力，也可能成为性能下降的原因，可以说是一把双刃剑。自监督学习的提出旨在减少对人工标注数据的依赖，使得即便在缺乏标注数据的情况下，网络仍能高效地进行训练。自监督学习在预训练阶段通过人工构造输入数据的代理任务使特征层网络学习到有效的特征表示，完全摆脱了对标签的依赖。只在对最终的分类器层进行微调过程中运用了标签，从而最大限度地减小“标签偏差”带来的影响。文献[3]指出，额外的自监督预训练能够显著提升常规长尾学习算法的性能，并已在CIFAR-LT、ImageNet-LT等视觉识别长尾数据集上得到验证。尽管在故障诊断的长尾学习领域，半监督学习、集成学习、样本加权等方法已经取得一定进展，但应用自监督学习方法的研

究仍相对较少。因此，应用自监督学习解决长尾分布数据故障诊断问题仍是一个具有重要意义的研究方向。

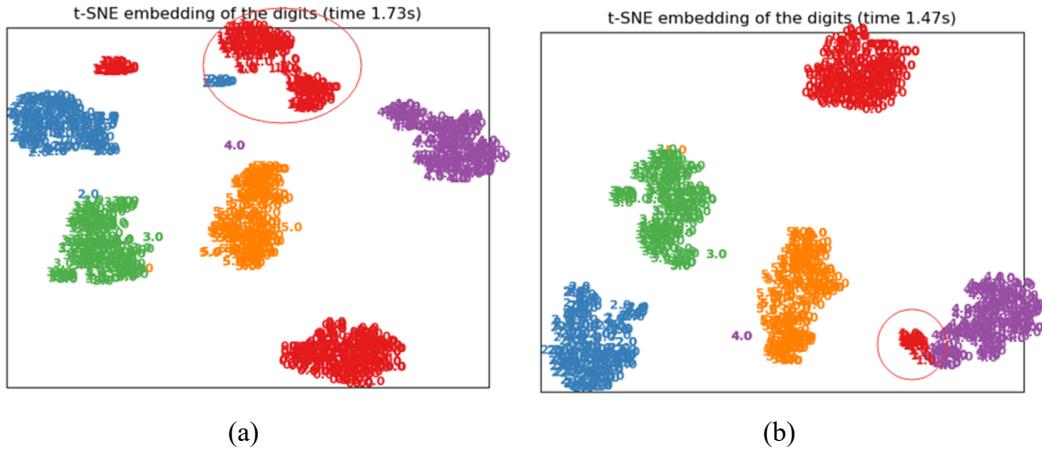


图 1-2 均匀/长尾分布数据的 tsne 图 (a) 样本均匀分布时的 T-SNE 分布图;
(b) 尾部类“1”类样本个数为 30, 头部类样本个数为 180 个时的 T-SNE 分布图

1.2 故障诊断基本流程

故障诊断的基本流程通常包括数据采集、信号预处理、特征选择、模型训练与优化、以及故障识别与诊断结果分析等步骤。在现代工业系统中，设备状态的监测依赖于多种传感器，如加速度传感器、声发射传感器、电流传感器和温度传感器等，以采集振动、声音、电流和温度等信号，从而提供反映设备健康状况的数据基础。由于采集到的原始信号往往受到噪声干扰，信号预处理成为数据分析的关键步骤，常见的方法包括滤波、归一化、去趋势以及信号分解，以提高数据质量并增强关键特征的可识别性。

在信号预处理的基础上，特征提取是故障诊断中的核心环节，其目的是从原始信号中提取能够反映设备运行状态的关键信息。根据信号分析方式的不同，特征可以分为时域、频域和时频域特征。时域特征如均值、均方根、偏度、峰值因子和峭度等，主要用于描述信号的统计特性；频域特征通过傅里叶变换提取主频分量、功率谱密度和边频带特征，以揭示设备的动态特性；时频域特征则通过小波变换、短时傅里叶变换或其他时频分析方法提取，分析信号的非平稳性和时变特征。由于部分手工提取的特征可能存在冗余或无关信息，特征选择成为优化故障诊断性能的重要手段，通常采用主成分分析、互信息分析或递归特征消除等方法筛选最具诊断价值的特征，本文构造了深度神经网络并采用自监督学习方法训练

特征层网络提取深层特征。

在获得高质量的特征后，模型训练是智能故障诊断系统构建的核心环节。传统的机器学习方法，常见的分类方法包括支持向量机、随机森林、K 近邻和决策树等，依赖于人工构建的特征，能够在特定工况下实现较高的分类精度。然而，随着深度学习的发展，卷积神经网络、循环神经网络、Transformer 和图神经网络等深度学习模型展现出更强的特征自动提取和故障识别能力，特别是在面对复杂工况和海量数据时，其优势尤为明显。由于工业环境中的数据分布通常存在不均衡问题，模型训练过程中需要考虑数据重采样、类别平衡损失等策略，以降低类别不平衡对故障识别能力的影响。此外，超参数优化在提升模型泛化能力和稳定性方面也起着至关重要的作用。

在训练得到故障诊断模型后，测试数据的输入可实现自动故障识别与分类。常见的故障类型包括轴承故障（如滚动体故障、内圈故障和外圈故障）、齿轮故障（如齿面点蚀、断齿和偏磨）以及电机故障（如定子故障、转子故障）等。故障分类的准确性通常采用准确率、召回率、F1-score 等指标进行评估，以确保模型的可靠性和实用性。最终，故障诊断的核心目标是为设备维护和运行决策提供有效支持。通过诊断结果分析，可以实现设备健康状态评估、故障预警、剩余寿命预测以及故障溯源分析等功能，从而优化维护策略，提高设备运行的安全性和可靠性。随着深度学习、迁移学习和自监督学习等技术的不断发展，智能故障诊断方法正在逐步突破传统方法的局限，未来研究将进一步探索如何在复杂工况下提升诊断模型的鲁棒性和适用性，以推动智能运维的发展。

1.3 故障诊断的研究现状

近年来，故障诊断的研究如火如荼。薛阳等人^[4]基于 CWRU 滚动轴承故障数据提出了将多模态融合与卷积神经网络相结合，提取了时域、频域两个模态的特征，实现时频域双模态融合对故障类型的联合诊断，相较基于单时域和单频域的卷积神经网络故障诊断模型都有提升。郭文军等人^[5]构建了经验模态分解与自回归相融合的深度特诊提取方法，提升了隐层筛选故障特征的能力。在构造故障诊断模型时特征提取显得颇为重要，因此故障诊断技术的特征提取得到了国内外学者的广泛关注。传统的时域信号特征提取方法，如 FFT，小波包分解，小波变换，经验模态分解都得到了较好的应用。近年来，随着神经网络的发展，基于机器学习的特征提取方法的研究也如火如荼，如基于粒子群优化的自适应多稳态欠阻尼随机共振方法^[6]，堆叠自动编码器^[7]，多尺度核卷积神经网络^[8]，递归图编码将振动信号转为增强信号特征的二维图像^[9]。

1.4 长尾学习的发展历程和研究现状

国内外在处理长尾问题时通常从三个方面着手，如图1-3所示：一是样本层面，采取欠采样（如随机欠采样、NearMiss^[10]、ENN^[11]）或过采样（如随机过采样、SMOTE^[12]、变分自编码器^[13]、生成对抗网络^[14]）的方法。然而，由于新生成的样本中可能掺杂噪声，仍可能导致模型出现误判；二是损失函数层面，使用类平衡的损失函数，如基于数据频率逆加权^[15]，OHEM(Online Hard Example Mining)、Focal Loss^[16]，以及最新提出的CB(Class-Balanced)Loss^[17]；三是模型层面，如通过集成学习、半监督和自监督学习^[18]、MARC 决策面调整算法^[19]，采用少数类样本与等量多数类样本的组合进行模型训练，还有迁移学习^[20,21]、度量学习^[22,23]和元学习^[24,25]，也得到了探索。最近的研究还发现，解耦特征层与分类器的结构可以带来更好的长尾学习结果^[26,27]。在长尾分布问题的研究上，国内外已有大量的探索。例如，Liu J 等人^[28]提出了一种基于特征云的数据增强方法，通过从方差较大的头部类中学习类内多样性，并将其转移到尾部类，从而改善尾部类的分类效果，该方法已成功应用于人脸识别领域。Yang Y 等人^[18]探讨了不平衡样本标签在模型训练中的价值，提出通过半监督学习方法利用原模型识别无标签数据，并通过伪标签扩充原数据集，同时验证了自监督预训练在长尾学习中的有效性，且证明样本维度越高，性能提升越显著。吴磊等人^[29]提出了一种针对长尾图像的个性化专家识别算法，基于残差网络构建多专家学习模型，并引入个性化学习、信息融合与增强模块。通过两阶段学习方法，不仅显著提升了整体识别精度，还特别增强了中尾部类别的识别效果。吴亮等人^[30]提出了多级分组分类器，以提升尾部类性能，同时避免头部类性能损失，并设计了基于多头注意力机制的分组特征重融合模块，为多级分类器输入更精细的特征。此外，基于多级分类器，他们还提出了Logit 联合调整方法，以缓解组间不平衡问题。Wang Y 等人^[19]介绍了适用于长尾视觉识别的MARC 决策面调整算法，通过对模型输出的预测分数进行额外训练，仅需三行代码便能实现。Cui Y 等人^[17]提出了CB(Class-Balanced) 损失函数，通过乘以与“独特”样本数量相关的因子，缓解了类别不平衡问题。CB 因子与超参数 β 有关，通过平滑 1 到最常见类别的不平衡因子 $1/n$ 来实现。

综上所述，文献[18]为自监督预训练在故障诊断长尾学习中的应用提供了良好的启示，但半监督学习方法在现实长尾学习模型中的适用性尚待验证。文献[29,30]介绍的集成学习方法有助于减弱样本不平衡效应，虽然是常规方法，但如果在训练过程中融合自监督预训练，性能可能会得到进一步提升。文献[19]提出的决策面调整算法适用于一般长尾学习模型，但其在故障诊断中的可行性仍需进一步探讨，且该算法对模型性能提升的理论性解释也尚不完善。文献[17]提出

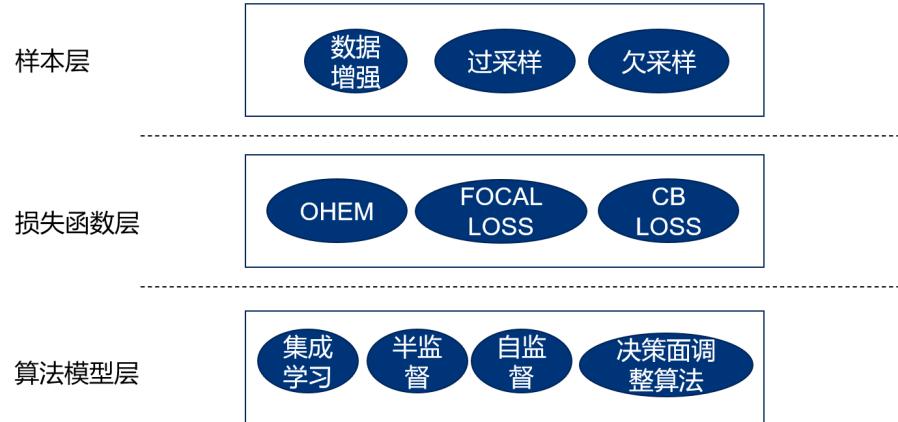


图 1-3 长尾学习常见方法

的损失函数方法具有创新性，但超参数 β 的人工选择在实际应用中可能面临一定的挑战。以上研究多集中于视觉长尾学习领域，并且在 CIFAR-LT、ImageNet-LT 等视觉识别数据集上取得了良好效果，而在故障诊断长尾学习领域，前沿的长尾学习算法应用较为匮乏。因此，将前沿的长尾学习算法应用于故障诊断领域，具有重要的研究价值，本文将着重研究模型层面的自监督长尾学习方法，分别研究自监督预训练和微调阶段过程的优化。

1.5 自监督学习的发展历程和研究现状

2019 年，MoCo^[31] 的提出掀起了视觉自监督学习的热潮，随后 SimCLR^[32]、BYOL^[33]、SwAV^[34] 等主流自监督学习算法相继问世，使得该领域呈现出百花齐放、百家争鸣的繁荣局面。2021 年末，MAE^[35] 的提出进一步促进了自监督学习的发展，并将其提升到了一个全新的高度。然而，这一成就的背后，自监督学习经历了长期的迭代与发展。

目前，国内外学者对自监督预训练方法展开了大量研究。例如，在计算机视觉领域，Yang 等人^[18] 研究了不平衡样本标签信息在模型训练中的价值，提出了一种半监督学习方法，该方法通过利用原模型识别无标签数据并生成伪标签来扩充原数据集，并验证了自监督预训练的有效性。此外，该研究还发现，样本维度越高，性能提升越显著，这为自监督预训练在故障诊断长尾学习领域的应用提供了有力支持。Doersch 等人^[36] 提出了基于局部图像块位置预测的自监督任务，随机抽取图像中的两个块，训练模型预测一个块相对于另一个块的位置。Zhang 等人^[37] 构建了基于图像上色的自监督任务，将图像转换到 CIE Lab 颜色空间后，提取其中的 L 通道作为模型输入，引导模型预测对应的 a 和 b 通道颜色信息。Gidaris 等人^[38] 通过人为旋转图像不同角度，并将旋转角度作为监督信号训练模型，从而实现图

像语义特征的自监督学习。文献 [39] 提出了一种新的表示学习方式，通过基于计数视觉原语的人工监督信号来进行训练，而无需任何人工标注。核心思想是利用图像变换与表示变换之间的等变关系。文献 [40] 提出了利用 KMeans 算法生成样本的伪标签，将其作为数据的真实标签训练网络，用网络新提取的特征重新生成 KMeans 伪标签，以此重复。

近年来，自监督预训练方法也逐步应用于故障诊断领域，并受到广泛关注。例如，Zhang 等人^[41]提出了基于先验知识的自监督任务，利用卷积自编码器进行训练，在小样本学习任务中取得了良好表现。Senanayaka 等人^[42]采用 One-Class SVM 输出的标签作为代理标签，使用自监督预训练的卷积神经网络进行故障特征提取。W. Zhang 等人^[3]提出了一种基于信号块交换的自监督任务，即通过打乱时域信号块顺序构造“伪数据”，并训练模型区分原始数据和伪数据。实验结果表明，该方法在特征提取任务中表现优异。

综上所述，尽管文献 [36–38] 等研究在计算机视觉领域取得了重要进展，但其方法在故障诊断任务中的直接应用仍面临一定困难。然而，这些研究的思路为故障诊断领域的自监督预训练任务设计提供了有益的启示。例如，文献 [42] 提出的 One-Class SVM 生成的标签可以结合聚类算法，使其适用于多分类任务；文献 [3] 提出的信号块交换方法与文献 [36] 提出的图像块位置预测任务在思路上具有一定相似性，值得探索其结合的可能性。此外，文献 [3, 41, 42] 提出的自监督预训练方法在任务设计上仍具有较大的创新空间，且在长尾学习任务中的应用尚未得到充分探索。因此，故障诊断领域的自监督预训练仍处于发展初期，借鉴计算机视觉领域的前沿自监督任务，提出更具创新性和更高性能的自监督预训练任务，仍具有重要的研究价值和挑战。

1.6 对比自监督学习和孪生网络的发展历程和研究现状

自监督学习的核心目标是在无人工标注的情况下自动提取有效特征^[43]。近年来，对比自监督学习 (contrastive self-supervised learning) 作为无监督学习的重要分支之一，通过构造正负样本对，使模型能够学习更具判别性的特征表示。它逐渐成为研究热点，并在图像识别任务中取得了最先进的 (SOTA) 性能。许多相关算法相继被提出，例如动量对比学习 (Momentum Contrast, MoCo)^[31]、对比预测编码 (Contrastive Predictive Coding, CPC)^[44] 以及用于视觉表征学习的简单对比学习框架 (SimCLR)^[32]。

这些方法的核心思想是通过拉近相同样本的不同视图 (正样本对) 并推远不同样本 (负样本对) 来学习有意义的特征表示。对比学习方法通常依赖存储在记忆库

(memory bank) 中的大量负样本，以提高对比学习效果^[45]。此外，一些其他对比学习方法不依赖负样本，而是计算两个正样本对之间的相似性，例如 SwAV(Swapping Assignments between multiple Views)^[34]、BYOL(Bootstrap Your Own Latent)^[33] 以及 SDCT(Self-supervised Deep Correlation Tracking)^[46]。

在对比自监督学习中，孪生网络 (Siamese Network) 因其适用于比较不同样本的网络结构，具有独特优势。Laine 和 Aila^[47] 提出了基于孪生网络的 π -model，用于训练深度神经网络，该方法在半监督学习场景下依赖不同的正则化规则与数据增强策略。Zheng 和 Yang^[48] 提出了一种“记忆正则化”机制，其核心思想是让主模型自身 (而非外部模块) 学习域内知识，以实现无监督场景自适应。Zhao 等人^[49] 提出了基于相互学习 (mutual learning) 和知识蒸馏 (knowledge distillation) 的训练方法，以提高视觉目标跟踪任务的性能。

上述研究均在各自领域中引入孪生网络的思想，以解决特定问题。本文受到 SimSiam^[50] 的启发，SimSiam 采用简化的孪生网络结构，并在计算机视觉 (CV) 任务中展现出卓越性能。基于此，本文提出了一种新颖的轴承故障诊断框架，该框架可适用于由不同神经网络层构建的多种模型。与其他主流方法相比，SimSiam 具有以下三大优势：

- (1) 无需构造负样本对；
- (2) 无需采用大批量训练；
- (3) 无需使用动量编码器 (momentum encoder)。

然而原有 SimSiam 模型过于依赖数据增强模块的增强质量，应用在故障诊断领域的效果较差。本文从理论上分析了 SimSiam 数据增强模块对训练的意义以及所需要的数据增强方法需要具备的特质，同时通过一种自动搜索最优增强参数的方法解决 SimSiam 需要经过人工挑选数据增强方法和设定数据增强模块的参数的问题。

1.7 本文的主要贡献与创新

本论文聚焦于面向长尾分布数据的自监督学习故障诊断，主要的创新点和贡献如下：

- (1) 提出了基于协方差矩阵适应进化策略的数据增强策略搜索算法优化的简单暹罗孪生网络的对比学习自监督预训练故障诊断框架。
- (2) 实验结果显示与 SimCLR 中提出的“较大的 Batch Size 对对比学习模型性能有益”的观点不一致的趋势。
- (3) 通过半监督学习优化了决策面调整算法从而减小微调过程受长尾分布的影响。

1.8 本论文的结构安排

本文的章节结构安排如下：

第一章介绍了研究的背景与意义，分析了传统故障诊断方法在处理长尾分布数据时面临的挑战，并重点回顾了国内外在长尾学习算法领域的研究进展与现有局限性。

第二章阐述了长尾学习与自监督学习范式的基本概念，介绍了为提出简单孪生网络故障诊断框架的理论基础，并介绍了实验使用的数据集和研究对象。

第三章提出了基于简单孪生网络为骨干网络的自监督预训练故障诊断框架，详细描述了框架各模块的设计与实现，并从实验与理论两个层面验证了框架的有效性。通过准确率、t-SNE 分布图以及对 Batch Size 等参数的敏感性分析，证明了该方法在性能和稳定性上优于现有 SOTA 的自监督学习方法。

第四章提出了基于半监督学习与决策面调整相结合的微调框架，旨在减轻简单孪生网络在微调阶段受长尾效应影响的问题。实验结果表明，该微调框架不仅提升了简单孪生网络的整体性能，而且在其他模型中同样具有较好的效果，展示了其良好的普适性。

第五章总结了全文的研究工作，并对未来可能的研究方向进行了展望。

第二章 长尾学习及相关方法的理论基础与轴承故障数据集

2.1 自监督学习相关理论

自监督学习是一种范式，利用“代理任务”(pretext)自监督学习旨在挖掘大规模无标签数据中所蕴含的潜在结构，通过构造辅助任务生成伪监督信号，进而引导网络学习具备迁移至下游任务能力的通用表征。如图2-1所示，该方法通常包含两个阶段：首先进行自监督预训练以学习初始表征，然后通过微调阶段在特定下游任务上进一步优化模型性能。在自监督预训练阶段，完全忽略数据的标签信息，专注于数据本身的特征学习。通过设计合理的自监督任务，引导模型从不平衡数据中提取与标签无关的潜在特征表示，以此减少类别偏差对模型初始化的影响，同时增强模型的泛化能力。在微调阶段，使用自监督预训练得到的网络权重作为初始化，并结合任务相关的监督信息进行优化，以提升模型在特定任务上的表现，同时保留预训练阶段学到的通用特征，从而提高模型的稳健性和泛化能力。在此基础上，可结合任意标准的不平衡学习技术进行后续训练，以学习最终的分类模型。由于预训练阶段和正常训练阶段是独立的，自监督学习可以与现有的不平衡学习方法无缝结合，增强模型性能，且自监督预训练阶段不依赖标签，使得网络能学习到更通用、更鲁棒的特征表示，从而避免了类别不平衡对特征学习的负面影响。

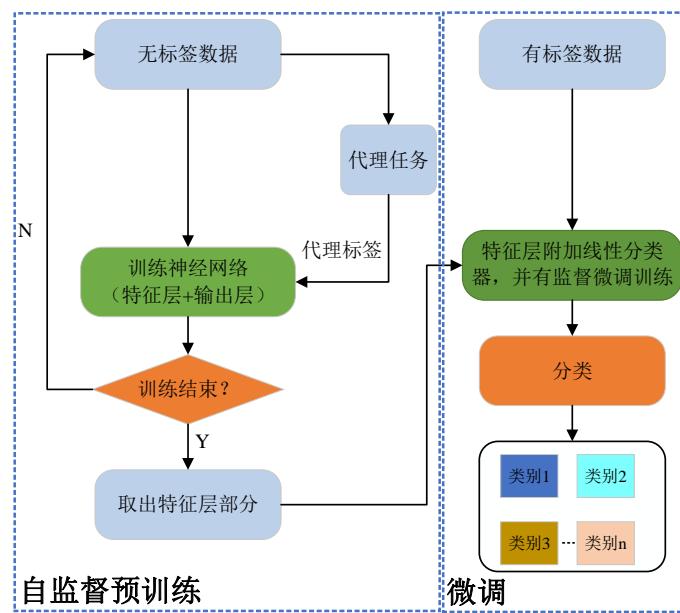


图 2-1 自监督训练范式流程

2.2 孪生网络与对比学习相关理论

对比学习是一种通过比较正负样本对来提取有意义特征的学习方法。其基本假设是在学习到的特征嵌入空间中，相似样本应聚集在一起，而不相似样本应远离彼此。通过将学习任务视为辨别任务，对比学习能够帮助模型识别数据中的潜在特征和相似性。对比学习可分为监督对比学习和自监督对比学习两类。

监督对比学习是对比学习的一个子领域，依赖于标注数据来明确区分相似与不相似的样本。在监督对比学习中，模型通过训练数据对及其对应标签来学习哪些数据点是相似的，哪些是不相似的。其目标是学习一个表示空间，其中具备相似特征的实例聚集在一起，而特征不相似的实例则被推开。一种常用的优化目标是信息噪声对比估计 (InfoNCE) 损失函数，它通过最大化正样本对的相似性并最小化负样本对的相似性来优化模型。通过优化该目标，模型能够有效区分正负样本，进而提升下游任务的性能。

自监督对比学习则不同于监督学习，它从未标注的数据中学习深层且有效的特征表示。自监督对比学习通过设计代理任务，从未标注数据中生成正负样本对。这些代理任务旨在促使模型捕获数据中有意义的特征和相似性。自监督对比学习中常见的策略之一是利用数据增强技术构造同一样本的多个不同视图，并将这些视图作为正样本对，而来自不同样本生成的实例则作为负样本对。通过训练模型识别并区分正负样本对，模型能够学习到更丰富的语义信息，并在下游任务中取得较好的推广效果。

2.3 t-SNE 降维原理

t-SNE(t-分布随机邻域嵌入) 是一种广泛应用的非线性降维方法，常用于将高维数据映射到二维或三维空间，便于通过散点图直观地展示数据分布和聚类情况。它通常用于处理如图像、文本或高维特征数据的可视化任务。t-SNE 通过以下三个步骤来进行降维：

- (1) **计算高维相似度：** t-SNE 首先计算每对数据点之间的相似度。在高维空间中，对于任意两点 x_i 和 x_j ，其相似度通过条件概率 p_{ij} 表示，计算公式为：

$$p_{ij} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{j \neq i} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)} \quad (2-1)$$

其中， $\|x_i - x_j\|$ 表示两点之间的欧氏距离， σ_i 是与点 x_i 相关的局部尺度参数，反映了该点的邻域大小。通过这种方式，t-SNE 保证了高维空间中相邻的数

据点具有较高的相似度。

- (2) **映射到低维空间:** 接着, t-SNE 将数据映射到二维或三维的低维空间中。在低维空间中, t-SNE 使用 t 分布 (Student's t-distribution) 来计算数据点之间的相似度, 公式为:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{j \neq i} (1 + \|y_i - y_j\|^2)^{-1}} \quad (2-2)$$

其中, y_i 和 y_j 是低维空间中的数据点的向量, $\|y_i - y_j\|$ 是它们之间的欧氏距离。t 分布的重尾特性使得它能够更加有效地处理聚类之间的距离, 同时避免了在低维空间中过度拥挤的现象。

- (3) **最小化 KL 散度:** t-SNE 的目标是最小化高维空间和低维空间中数据点相似度分布的差异。为此, t-SNE 通过最小化 Kullback-Leibler(KL) 散度来优化模型的映射过程, KL 散度的计算公式为:

$$\text{KL}(P||Q) = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (2-3)$$

通过迭代优化, t-SNE 会调整低维空间中的数据点位置, 直到高维和低维空间的相似度分布尽可能接近。

t-SNE 降维伪代码见表2-1。

2.4 K-Means 聚类原理

K-Means 是一种经典的无监督学习算法, 其核心思想是通过最小化样本点与各自簇中心之间的距离, 将数据划分为 K 个簇。该算法在图像分割、客户细分、文档聚类以及特征提取等多个领域被广泛应用。算法步骤如下:

- (1) **初始化:** 从数据集中随机选取 K 个样本作为初始簇中心。
- (2) **簇分配:** 依据最小欧氏距离准则, 将每个样本分配至其最近的簇中心所属的簇, 欧氏距离计算为:

$$\text{Distance}(x_i, c_j) = \|x_i - c_j\|$$

其中, x_i 为数据点, c_j 为簇的质心。

- (3) **更新质心:** 重新计算每个簇的质心, 即簇内所有数据点的均值:

$$c_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i$$

其中, S_j 为第 j 个簇的数据点集合, $|S_j|$ 为该簇的数据点数。

表 2-1 t-SNE 算法伪代码

t-SNE 算法伪代码

```

# 输入：高维数据 X = {x1, x2, ..., xn}
# 输出：低维数据 Y = {y1, y2, ..., yn}
# Step 1: 计算高维空间中数据点之间的相似度 p_ij
for i = 1 to n: # 对每个数据点 xi
    for j = 1 to n:
        # 计算 xi 和 xj 之间的欧氏距离，并转化为条件概率 p_ij
        dist_ij = norm(xi - xj)
        p_ij = exp(-dist_ij^2 / (2 * sigma_i^2)) / \
            sum(exp(-dist_ij^2 / (2 * sigma_i^2)) for all j)
        P[i][j] = p_ij # 存储 p_ij
# Step 2: 初始化低维空间中的数据点 Y
Y = random_initialization(n, d) # 随机初始化低维空间中的数据点
# Step 3: 迭代优化：最小化KL散度
for t = 1 to T: # 最大迭代次数 T
    # 计算低维空间中的相似度 q_ij
    for i = 1 to n:
        for j = 1 to n:
            dist_ij = norm(yi - yj)
            q_ij = (1 + dist_ij^2)^(-1) / \
                sum((1 + dist_ij^2)^(-1) for all j)
            Q[i][j] = q_ij # 存储 q_ij

    # 计算KL散度并计算梯度
    KL = sum(P[i][j] * log(P[i][j] / Q[i][j]) for all i and j)
    gradients = compute_gradients(P, Q, Y) # 计算梯度
    # Step 4: 更新低维空间中的数据点 Y
    Y = Y - learning_rate * gradients # 梯度下降更新低维数据点
    # Step 5: 终止条件判断
    if converged(KL): # 如果KL散度收敛
        break
return Y # 返回降维后的低维数据 Y

```

- (4) **重复迭代：**重复簇分配与簇质心更新步骤，直至簇质心收敛或达到预设的最大迭代次数。
- (5) **使用匈牙利算法找到最佳匹配：**在完成迭代后，使用匈牙利算法来优化簇标签的匹配，以确保最合适的标签分配。匈牙利算法可以通过最小化簇间的匹配成本来实现最佳匹配。

KMeans 算法的伪代码见表2-2。

2.5 协方差矩阵适应进化策略原理

协方差矩阵适应进化策略 (CMA-ES) 是一种进化算法，专用于解决连续优化问题。它是一种基于种群的优化方法，通过不断调整种群的协方差矩阵来引导搜

表 2-2 KMeans 算法伪代码

KMeans 算法伪代码

```

# 输入：数据集 X = {x1, x2, ..., xn}, 簇数 K, 最大迭代次数 T
# 输出：簇标签 {y1, y2, ..., yn}

# 初始化：随机选择 K 个数据点作为初始质心 C = {c1, c2, ..., cK}
for t = 1 to T: # 最大迭代次数 T
    for i = 1 to n: # 对每个数据点 xi
        # 计算每个质心的距离，分配数据点给最近的质心
        distances = [] # 初始化一个空列表
        for cj in C: # 对每个质心 cj
            dist = norm(xi - cj) # 计算 xi 到 cj 的距离
            distances.append(dist) # 添加到距离列表
        yi = argmin(distances) # 分配数据点 xi 到最近的质心
        # 更新数据点的簇标签
        labels[i] = yi

    for j = 1 to K: # 对每个簇
        # 更新质心为簇中所有数据点的均值
        cluster_points =
            [xi for xi, label in zip(X, labels) if label == j]
        # 计算簇内数据点的均值作为新质心
        C[j] = mean(cluster_points)

    # 使用匈牙利算法找到最佳匹配
    optimal_labels = hungarian_algorithm(labels, C)

    # 如果质心不再变化
    if no_change_in_centroids(C): # 如果质心不再变化
        break # 退出

return optimal_labels # 返回优化后的簇标签 {y1, y2, ..., yn}

```

索过程。CMA-ES 在解决非线性、非凸、高维优化问题时表现出色，广泛应用于机器学习、工程优化和科学研究中心。CMA-ES 的更新规则如下：

(1) 采样新解：

$$\mathbf{x}_k^{(g+1)} = \mathbf{m}^{(g)} + \sigma^{(g)} \cdot \mathcal{N}(0, \mathbf{C}^{(g)}) \quad (2-4)$$

其中： $\mathbf{x}_k^{(g+1)}$ 是第 $g+1$ 代中的第 k 个候选解， $\mathbf{m}^{(g)}$ 是第 g 代的均值向量， $\sigma^{(g)}$ 是第 g 代的步长， $\mathbf{C}^{(g)}$ 是第 g 代的协方差矩阵， $\mathcal{N}(0, \mathbf{C}^{(g)})$ 是从多元正态分布中采样的随机向量。

(2) 更新均值:

$$\mathbf{m}^{(g+1)} = \sum_{i=1}^{\mu} w_i \mathbf{x}_{i:\lambda}^{(g+1)} \quad (2-5)$$

其中: μ 是选择的父代数量, w_i 是权重系数, $\mathbf{x}_{i:\lambda}^{(g+1)}$ 是第 $g+1$ 代中适应度排名前 μ 的候选解。

(3) 更新协方差矩阵:

$$\mathbf{C}^{(g+1)} = (1 - c_1 - c_\mu) \mathbf{C}^{(g)} + c_1 \mathbf{p}_c^{(g+1)} (\mathbf{p}_c^{(g+1)})^\top + c_\mu \sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda}^{(g+1)} (\mathbf{y}_{i:\lambda}^{(g+1)})^\top \quad (2-6)$$

其中: c_1 和 c_μ 是学习率, $\mathbf{p}_c^{(g+1)}$ 是进化路径, $\mathbf{y}_{i:\lambda}^{(g+1)} = (\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}) / \sigma^{(g)}$ 。

(4) 更新步长:

$$\sigma^{(g+1)} = \sigma^{(g)} \exp \left(\frac{c_\sigma}{d_\sigma} \left(\frac{\|\mathbf{p}_\sigma^{(g+1)}\|}{\mathbb{E}[\|\mathcal{N}(0, \mathbf{I})\|]} - 1 \right) \right) \quad (2-7)$$

其中: c_σ 是步长学习率, d_σ 是阻尼系数, $\mathbf{p}_\sigma^{(g+1)}$ 是步长进化路径, $\mathbb{E}[\|\mathcal{N}(0, \mathbf{I})\|]$ 是标准正态分布向量的期望范数。

2.6 数据集介绍

2.6.1 凯斯西储大学 CWRU 数据集

所使用的数据集由凯斯西储大学 (Case Western Reserve University) 提供。图 2-2 展示了该实验平台的结构配置, 其主要包括一台 2 马力的电动机、扭矩编码器、测功机及若干控制电路。实验中使用的轴承为 6205-2RS JEM 型深沟球轴承, 分别安装在电机的驱动端和风扇端位置。

轴承故障是通过单点电火花放电加工 (Electro-discharge Machining) 制造的。根据故障位置, 故障类型可分为以下四种: 外圈故障 (Outer Raceway Fault, OF); 内圈故障 (Inner Raceway Fault, IF); 滚动体故障 (Roller Faults, RFs); 正常状态 (Normal Condition, NC)。轴承故障的严重程度可由故障直径 (Fault Diameter) 描述, 故障直径指轴承或其他机械部件上出现的故障或损伤的直径尺寸。故障直径通常用来描述故障的大小和程度, 对于故障诊断和预测维护非常重要。故障分类见表2-3。

表 2-3 CWRU 轴承数据集故障类型表，其中 0 为正常类

序号	损伤部位	损伤直径 (mm)
0	-	-
1	外圈 OF	0.007
2	外圈 OF	0.014
3	外圈 OF	0.021
4	内圈 IF	0.007
5	内圈 IF	0.014
6	内圈 IF	0.021
7	滚珠 BF	0.007
8	滚珠 BF	0.014
9	滚珠 BF	0.021

本研究采集了驱动端轴承的数据，采样频率为 12 kHz，故障类型包括内圈故障、外圈故障和滚动体故障，故障直径分别为 0.007 英寸、0.014 英寸和 0.021 英寸。不同故障类型的波形如图 2-3 所示。有关该数据集的更多详细信息，可访问 CWRU 轴承数据中心网站^[51]。

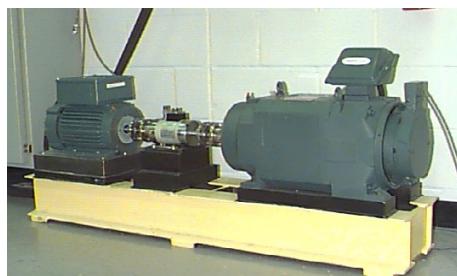


图 2-2 CWRU 轴承测试设备

2.6.2 帕德伯恩大学 PU 数据集

PU 轴承试验台由电机、测矩轴、滚动轴承试验模块、飞轮和负载电机组成，所有测试轴承均为 6203 型号滚动轴承。故障类别以不同损伤部位、损伤程度、损伤方法区分，加上正常类，共 13 类，具体故障类型如表2-4所示。

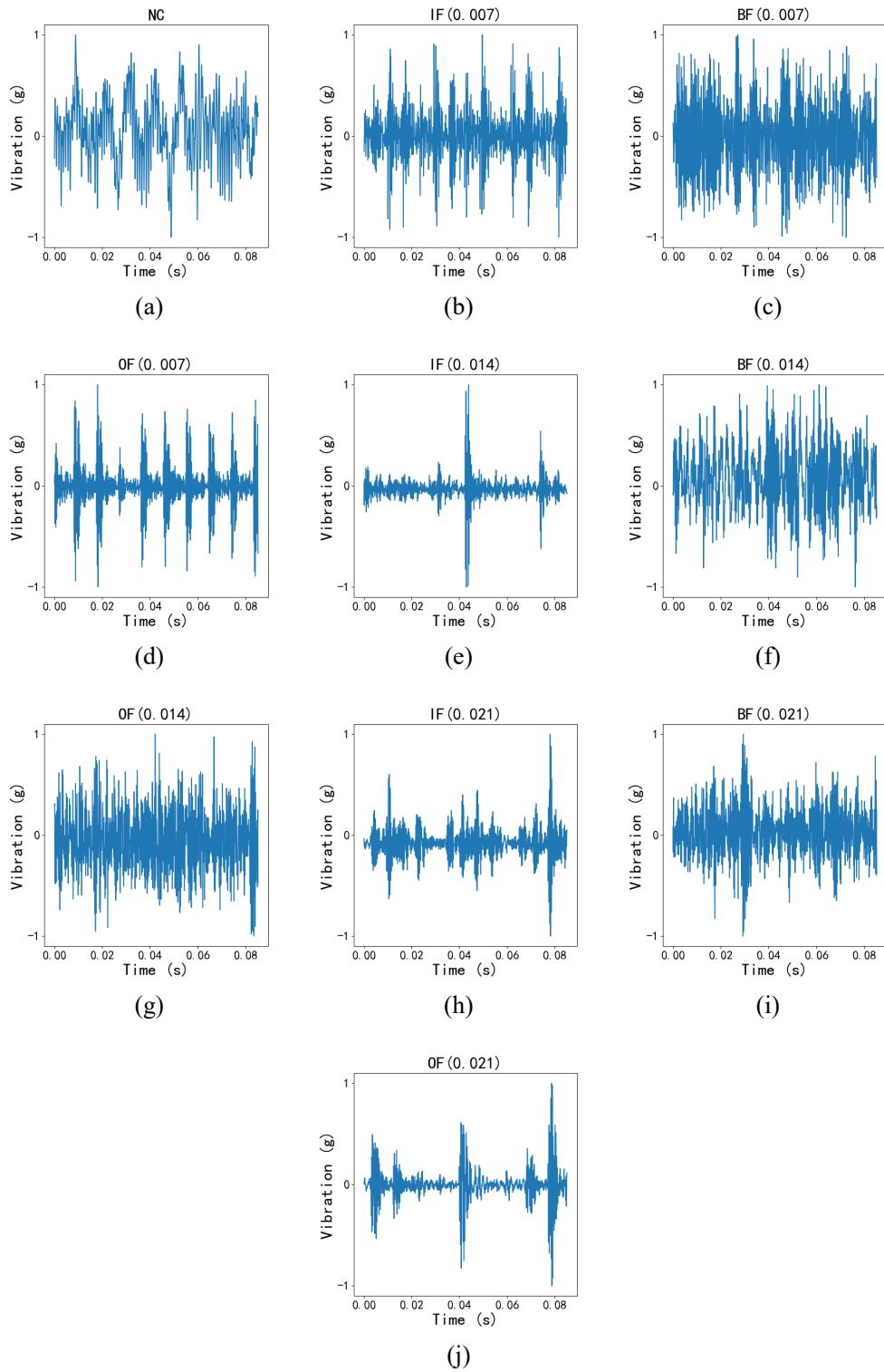


图 2-3 每个类别的样本信号示例: (a) NC; (b) IF(0.007); (c) BF(0.007)
 (d) OF(0.007); (e) IF(0.014); (f) BF(0.014); (g) OF(0.014); (h) IF(0.021); (i)
 BF(0.021); (j) OF(0.021)

表 2-4 PU 数据集故障类型分类表，其中 0 为正常类

序号	轴承编码	制造商	损伤程度	损伤部位	损伤方法
0	K001	IBU	-	-	-
1	KA01	MTK	1	外圈	电火花加工
2	KA03	LBU	2	外圈	电雕刻
3	KA05	LBU	1	外圈	电雕刻
4	KA06	LBU	2	外圈	电雕刻
5	KA07	LBU	1	外圈	转孔
6	KA08	LBU	2	外圈	转孔
7	KA09	LBU	2	外圈	转孔
8	KI01	MTK	1	内圈	电火花加工
9	KI03	LBU	1	内圈	电雕刻
10	KI05	LBU	1	内圈	电雕刻
11	KI07	LBU	2	内圈	电雕刻
12	KI08	LBU	2	内圈	电雕刻

2.6.3 长尾数据集构造

模拟构建长尾分布数据集流程如图2-4所示。设定不平衡因子 $\beta = x_{\max}/x_{\min}$ 为数据集中样本数量最多的类与样本数量最少的类的数量之比。帕累托分布的概率密度函数为 $p(x) = \frac{\alpha x_{\min}^{\alpha}}{x^{\alpha+1}}$ ，令 $x_{\min} = 1$ ，则 $p(x) = \frac{\alpha}{x^{\alpha+1}}$ 。以不平衡因子构建服从帕累托分布的长尾分布数据集，不同不平衡因子的帕累托分布概率密度函数如图2-5所示，可以看到当最大类与最小类样本数的比值 β 增大时，帕累托分布的形状参数 α 也随之增大。这表明分布的头部类别占比提高，而尾部类别占比显著降低，长尾效应越发显著。下面将介绍形状参数 α 的求解方法，以及各类别样本占比的计算步骤。

已知帕累托分布的累积分布函数为：

$$F(x) = 1 - x^{-\alpha}, \quad x > 1 \quad (2-8)$$

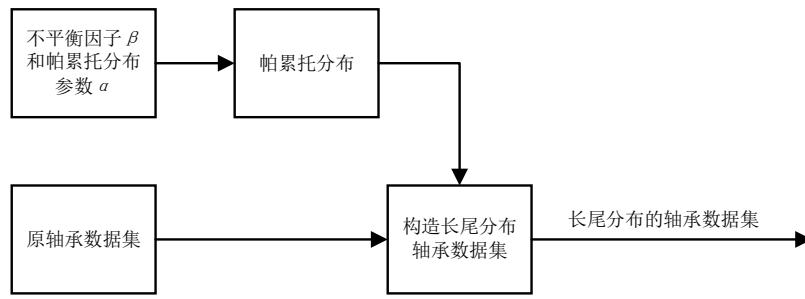
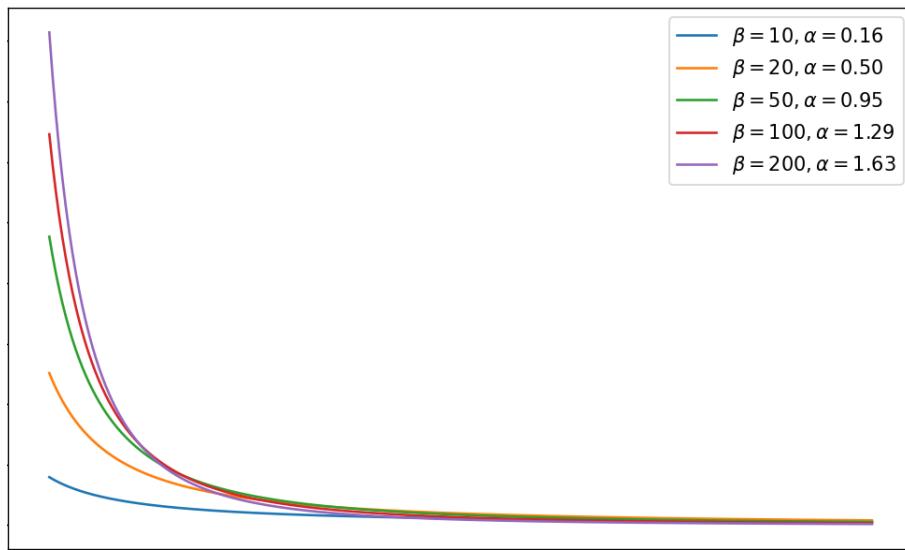


图 2-4 基于帕累托分布构造长尾分布的轴承数据集流程图


 图 2-5 不同不平衡因子 β 的帕累托分布概率密度图

每类的概率定义为：

$$P(n \leq x < n+1) = F(n+1) - F(n) = n^{-\alpha} - (n+1)^{-\alpha} \quad (2-9)$$

设最大类与最小类的样本数比值为 β ，则有以下关系式：

$$\beta = \frac{P(1 \leq x < 2)}{P(n \leq x < n+1)} = \frac{1 - 2^{-\alpha}}{n^{-\alpha} - (n+1)^{-\alpha}} \quad (2-10)$$

目标是根据给定的 β 和类别总数 n ，求解形状参数 α 。其次计算每类的样本占比 $P(n \leq x < n+1)$ 。最后将所有类别的概率归一化，使其和为 1。具体步骤如下：

首先求解 α ，根据 β 的定义，解以下非线性方程以确定 α ：

$$\beta = \frac{1 - 2^{-\alpha}}{n^{-\alpha} - (n+1)^{-\alpha}} \quad (2-11)$$

该方程一般无解析解，可以通过数值方法(如 Newton-Raphson 或其他优化算法)求解。

其次计算每类的概率，每类的概率可以通过以下公式计算：

$$P(n \leq x < n + 1) = n^{-\alpha} - (n + 1)^{-\alpha}, \quad n = 1, 2, \dots, N \quad (2-12)$$

最后概率归一化，将所有类别的概率归一化，计算归一化后的概率：

$$P_{\text{norm}}(n \leq x < n + 1) = \frac{P(n \leq x < n + 1)}{\sum_{k=1}^N P(k \leq x < k + 1)} \quad (2-13)$$

其中 N 为类别总数，归一化后各类别的样本占比之和为 1：

$$\sum_{n=1}^N P_{\text{norm}}(n \leq x < n + 1) = 1 \quad (2-14)$$

第三章 基于暹罗孪生网络对比学习的自监督学习故障诊断方法

本章提出了一种基于暹罗孪生网络对比学习的自监督学习故障诊断框架，设计了一种新颖的数据增强策略搜索算法，并对数据增强模块在该框架中的必要性和设计方向进行了详细的理论分析和实验。

设计孪生网络的一个关键挑战是防止网络输出出现“坍塌”现象，即输出最终趋向于一个常数的平凡解。为了解决这一问题，本章首先介绍了简单暹罗孪生网络框架，并结合实验与理论分析，探讨了各个模块在防止网络“坍塌”方面的作用。通过一系列实验证明了框架的有效性和鲁棒性，特别是在实际应用中如何通过合理设计避免模型的退化问题。

3.1 模型整体架构及其模块设计

3.1.1 简单暹罗孪生故障诊断网络整体结构

提出了一种用于故障诊断的简单暹罗孪生网络 (Simple Siamese, SimSiam)，其结构如图 3-1 所示。模型的训练过程包括三个阶段。首先是对比学习阶段，在该阶段，所有未标记的振动样本经过归一化预处理，并采用协方差矩阵适应进化策略 (CMA-ES) 搜索最优的数据增强策略，对输入样本进行增强，生成两个不同视图作为正样本对。然后，将正样本对输入模型进行 SimSiam 网络对比学习，以优化编码器的特征表示能力。接下来的微调阶段，提取第一阶段训练好的编码器，并在其最后一层附加一个线性分类层，随后使用标记样本对模型进行有监督训练。此外，线性分类层还可以被其他分类器替换，如 KNN 和 SVM。最后，在故障诊断阶段，将编码器和线性分类器层组合成一个完整的故障分类器，应用于实际的故障诊断任务，实现高效的故障识别。以下将对各个模块的具体实现细节进行详细介绍。

3.1.2 简单暹罗孪生网络对比学习

简单暹罗孪生网络 (SimSiam——Simple Siamese)^[50] 架构 (图3-2) 将信号 x 中的两个随机增强视图 x_1 和 x_2 作为输入。这两个视图由编码器 (Encoder) 网络 f 处理，该网络由骨干网络 (例如 ResNet^[52]) 和投影器 MLP(Projector) 组成。编码器 f 在两个视图之间共享权重。预测器 MLP(Predictor) h 转换一个视图的输出并将其与另一个视图进行匹配。将两个输出向量表示为 $p_1 \triangleq h(f(x_1))$ 和 $z_2 \triangleq f(x_2)$ ，最小化

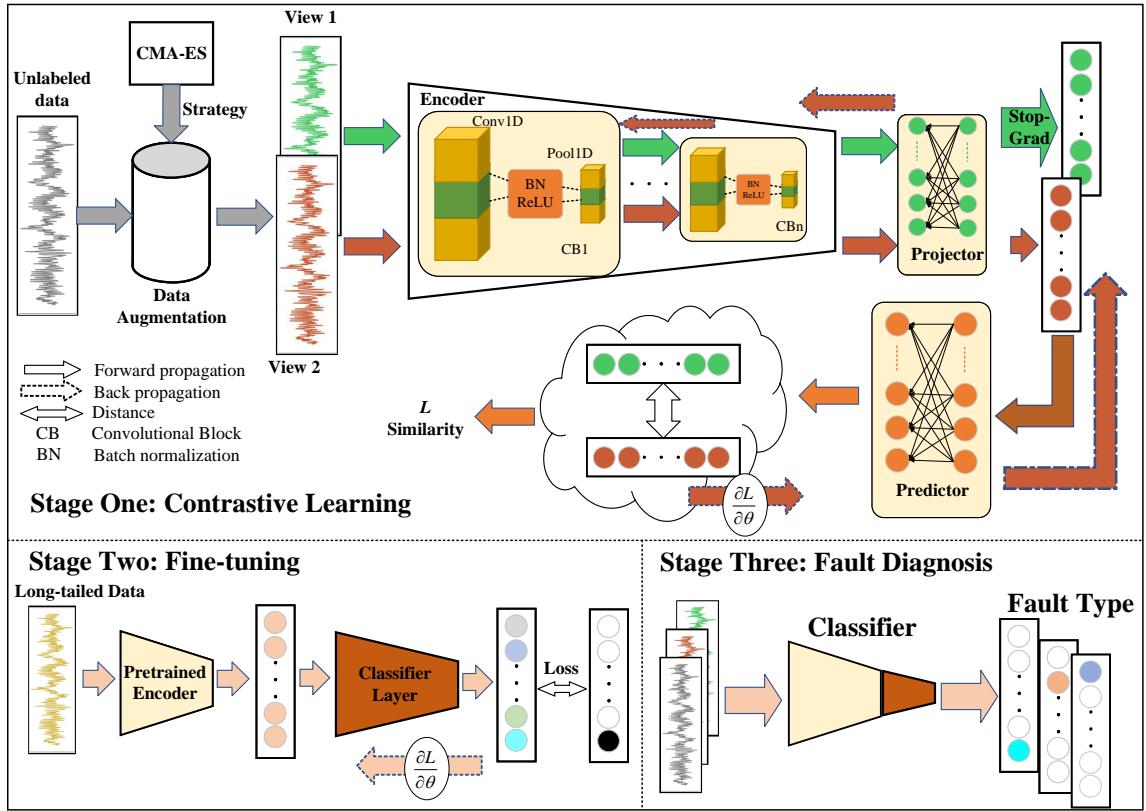


图 3-1 简单暹罗孪生故障诊断网络

它们的负余弦相似度：

$$\mathcal{D}(p_1, z_2) = -\frac{p_1}{\|p_1\|_2} \cdot \frac{z_2}{\|z_2\|_2} \quad (3-1)$$

其中， $\|\cdot\|_2$ 是 ℓ_2 范数。定义对称损失函数

$$\mathcal{L} = \frac{1}{2}\mathcal{D}(p_1, z_2) + \frac{1}{2}\mathcal{D}(p_2, z_1) \quad (3-2)$$

该损失函数作用于单段信号，总损失值在计算所有信号的损失后取平均值。损失函数的最小值为-1。

方法中一个重要的组件是梯度停止 (stop-grad) 操作 (图3-2)。通过修改公式 (3-1) 来实现它：

$$D(p_1, \text{stopgrad}(z_2)) \quad (3-3)$$

这意味着在这一项中， z_2 被视为常数。类似地，公式 (3-2) 的实现形式为：

$$\mathcal{L} = \frac{1}{2}D(p_1, \text{stopgrad}(z_2)) + \frac{1}{2}D(p_2, \text{stopgrad}(z_1)) \quad (3-4)$$

第一项中 x_2 的编码器不会从 z_2 接收梯度，但在第二项中会从 p_2 接收梯度（反之亦然，对于 x_1 也是如此）。即在这一项中， z_2 被视为常数。

简单暹罗孪生网络的伪代码如表3-1所示。

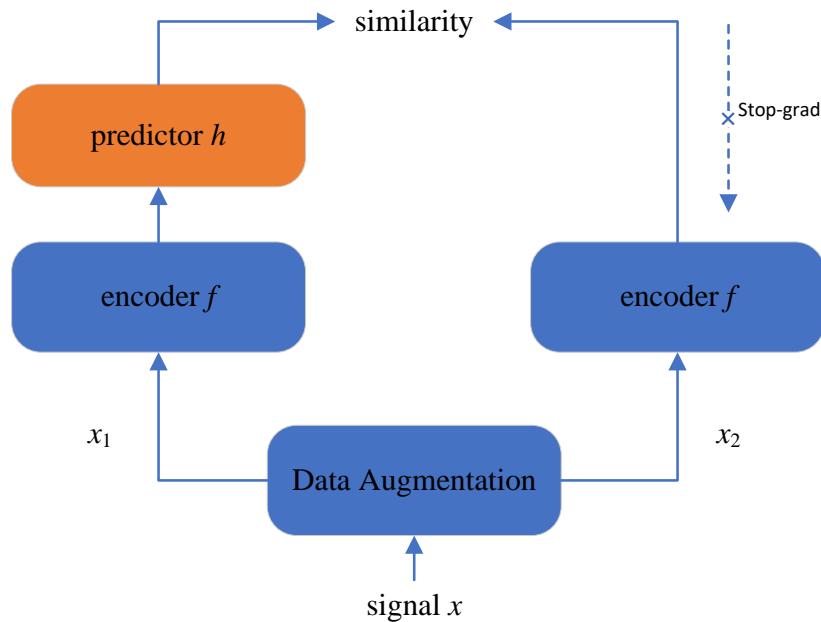


图 3-2 简单暹罗孪生网络

引入停止梯度的设计暗示了另一个潜在的优化问题正在被隐式解决。SimSiam是一种类似于期望最大化(EM)算法的方法，隐式涉及两组变量并解决两个潜在的问题。停止梯度操作的引入是为了引入额外的变量集。

考虑以下形式的损失函数：

$$\mathcal{L}(\theta, \eta) = \mathbb{E}_{x, \mathcal{T}} \left[\|\mathcal{F}_\theta(\mathcal{T}(x)) - \eta_x\|_2^2 \right]. \quad (3-5)$$

其中， \mathcal{F} 是由 θ 参数化的网络， \mathcal{T} 是数据增强， x 是输入样本。期望 $\mathbb{E}[\cdot]$ 是对输入样本和数据增强的分布进行的。为了便于分析，这里使用均方误差 $\|\cdot\|_2^2$ ，如果向量是 ℓ_2 -归一化的，则等效于余弦相似度。暂时不考虑预测器，稍后再讨论。

在式(3-5)中，引入了另一组变量，记为 η 。 η 的大小与输入样本数量成正比。直观上， η_x 是输入样本 x 的表示，下标 x 表示使用输入样本索引访问 η 的子向量。 η 不一定是网络的输出；它是一个优化问题的参数。

通过这种形式化，考虑解决以下问题：

$$\min_{\theta, \eta} \mathcal{L}(\theta, \eta). \quad (3-6)$$

表 3-1 简单暹罗孪生网络的伪代码，用 Pytorch 描述

简单暹罗孪生网络 Pytorch 伪代码

```

#f: 骨干网络 + 投影器 MLP
#h: 预测器 MLP
for x in loader: # 加载一个包含 n 个样本的小批量数据 x
    x1, x2 = aug(x), aug(x) # 随机数据增强
    z1, z2 = f(x1), f(x2) # 投影, 形状为 n-by-d
    p1, p2 = h(z1), h(z2) # 预测, 形状为 n-by-d
    L = D(p1, z2)/2 + D(p2, z1)/2 # 损失函数

    L.backward() # 反向传播
    update(f, h) # SGD 更新

def D(p, z): # 负余弦相似度
    z = z.detach() # 停止梯度
    p = normalize(p, dim=1) # 对 p 进行 L2 归一化
    z = normalize(z, dim=1) # 对 z 进行 L2 归一化
    return -(p * z).sum(dim=1).mean() # 计算负余弦相似度

```

这里的问题是针对 θ 和 η 的。这种形式化类似于 KMeans 聚类。变量 θ 类似于聚类中心：它是编码器的可学习参数。变量 η_x 类似于样本 x 的分配向量（在 KMeans 中是一个 one-hot 向量），它是 x 的表示。

同样类似于 KMeans，式 (3-6) 中的问题可以通过交替算法解决，固定一组变量并解决另一组变量。形式上，可以在以下两个子问题之间交替：

$$\theta^t \leftarrow \arg \min_{\theta} \mathcal{L}(\theta, \eta^{t-1}) \quad (3-7)$$

$$\eta^t \leftarrow \arg \min_{\eta} \mathcal{L}(\theta^t, \eta) \quad (3-8)$$

其中， t 是交替的索引，“ \leftarrow ” 表示赋值。

可以使用随机梯度下降 (SGD) 来解决子问题 (3-7)。停止梯度操作是一个自然的结果，因为梯度不会反向传播到 η^{t-1} ，而 η^{t-1} 在这个子问题中是一个常数。

子问题 (3-8) 可以独立地为每个 η_x 求解。现在的问题是最小化 $\mathbb{E}_{\mathcal{T}} [\|\mathcal{F}_{\theta^t}(\mathcal{T}(x)) - \eta_x\|_2^2]$ ，其中期望是对数据增强 \mathcal{T} 的分布进行的。由于使用均方误差，可以通过以下方式轻松求解：

$$\eta_x^t \leftarrow \mathbb{E}_{\mathcal{T}} [\mathcal{F}_{\theta^t}(\mathcal{T}(x))]. \quad (3-9)$$

这表明 η_x 被赋值为 x 在数据增强分布上的平均表示。

SimSiam 可以通过在式 (3-7) 和式 (3-8) 之间进行一次交替来近似。首先，通过仅对数据增强采样一次 (记为 \mathcal{T}') 并忽略 $\mathbb{E}_{\mathcal{T}}[\cdot]$ 来近似式 (3-9):

$$\eta_x^t \leftarrow \mathcal{F}_{\theta'}(\mathcal{T}'(x)). \quad (3-10)$$

将其代入子问题 (3-7)，得到：

$$\theta^{t+1} \leftarrow \arg \min_{\theta} \mathbb{E}_{x, \mathcal{T}} \left[\left\| \mathcal{F}_{\theta}(\mathcal{T}(x)) - \mathcal{F}_{\theta'}(\mathcal{T}'(x)) \right\|_2^2 \right]. \quad (3-11)$$

现在 θ' 在这个子问题中是一个常数，而 \mathcal{T}' 由于其随机性暗示了另一个视图。这种形式化展示了孪生网络架构。其次，如果通过一步 SGD 减少损失来实现式 (3-11)，那么可以接近 SimSiam 算法：一个自然带有停止梯度操作的孪生网络。

根据定义，预测器 h 期望最小化：

$$\mathbb{E}_x \left[\left\| h(z_1) - z_2 \right\|_2^2 \right] \quad (3-12)$$

其中， $z_i = \mathcal{F}(x_i)$ ， x_i 表示输入 x 的第 i 个视图。 h 的最优解应满足：对于任何输入 x ， $h(z_1) = \mathbb{E}_z[z_2] = \mathbb{E}_{\mathcal{T}}[\mathcal{F}(\mathcal{T}(x))]$ 。这个项类似于式 (3-9) 中的项。在式 (3-10) 的近似中，期望 $\mathbb{E}_{\mathcal{T}}[\cdot]$ 被忽略了。 h 的使用可能填补了这一空白。在实践中，实际计算期望 $\mathbb{E}_{\mathcal{T}}$ 是不现实的。但神经网络 (例如预测器 h) 可能能够学习预测期望，而 \mathcal{T} 的采样隐式分布在多个 epoch 中。

根据上述讨论， h 的最优解应满足 $h(z_1) = E_z[z_2]$ 。假设 h 为线性映射，则有 $h(E[z_1]) = E[z_2]$ 。同时，根据对称损失函数式 (3-2)， h 的最优解应满足 $h(z_2) = E_z[z_1]$ 。将此关系代入，得到 $h(E[z_1]) = E[z_2]$ ，进一步推导得 $h(h(z_2)) = E[z_2]$ 。因此，可表示为：

$$h(h(x)) = E[x] \quad (3-13)$$

这表明，在满足最优解的条件下， h 很可能是输入变量 x 到其期望值的平滑映射。假设 $h = E$ ，显然这是式 (3-13) 的一个特解。基于此，子问题 (3-7) 可以表述为：

$$\arg \min_{\theta} \left[\left\| \mathcal{F}_{\theta}(x_1) - E(z_2) \right\|_2^2 + \left\| \mathcal{F}_{\theta}(x_2) - E(z_1) \right\|_2^2 \right] \quad (3-14)$$

因此，SimSiam 的目标是最小化两个视图 x_1 和 x_2 输出特征期望值之间的距离，即使得 $E(\mathcal{F}_{\theta}(x_1)) = E[z_2]$ 且 $E(\mathcal{F}_{\theta}(x_2)) = E[z_1]$ ，从而促使模型学习到在数据增强下

保持不变的深层次语义。图 3-3 直观展示了模型训练前后特征分布的变化。在训练前，多个相似样本的两个视图的特征在特征空间中可能位于距离较远且形态各异的区域；而在训练后，两个视图的特征空间趋于重叠，且期望值之间的距离显著缩小。这表明，编码器成功学习到了对数据增强不敏感的深层次语义特征。此时，当 h 计算出 $E[z_1]$ 时，即可得到式 (3-12) 的最优解，对于 z_2 同理。

然而，如果数据增强模块没有有效地带来多样化的特征空间分布（例如，训练前两个视图的特征空间已经趋于重叠），那么当 h 学习到预测期望的能力时，编码器可能不再需要进一步学习那些与数据增强无关的深层语义特征。

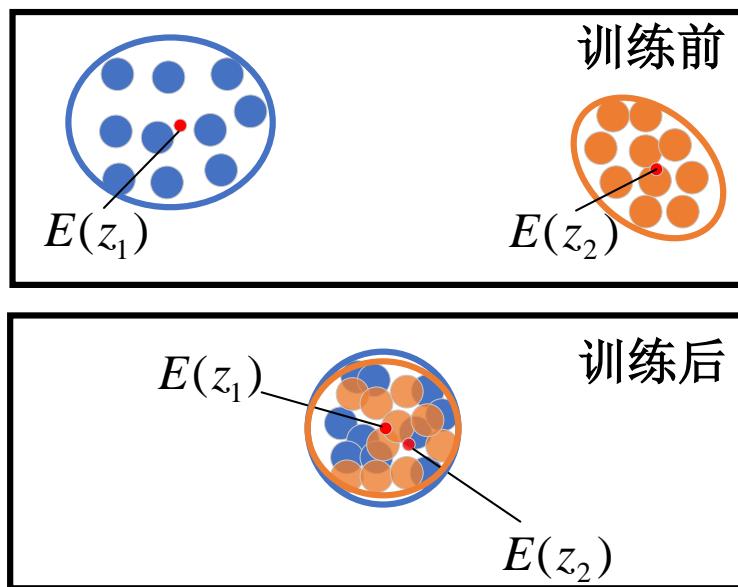


图 3-3 不同视图的模型特征输出分布的变化示意图

3.1.3 简单暹罗孪生网络的数据增强模块

在对比表示学习算法中，数据增强 (DA) 起着至关重要的作用。网络的编码器 (Encoder) 能否成功地从振动信号中提取可区分的故障特征，依赖于为同一样本生成不同视图的质量。注意到，面向图像的最新 (SOTA) 对比表示学习算法广泛使用多种数据增强方法，而针对序列数据的增强方法则相对较少。此外，用于对比学习生成正样本对的增强方法必须满足一个条件，即加入样本的噪声不能改变其语义含义。同时，数据增强模块需要有效地带来多样化的特征空间分布来迫使编码器学习深层语义特征（根据上一节对的理论分析）。因此选择了以下九种数据增强

方法：

- **掩码 (Random Masked):** 用于遮盖输入数据的一部分，即随机挑选信号的某些部分用 0 替代，通常是在序列数据或图像数据中随机选择部分区域进行掩码处理。这有助于模型在训练时学会忽略一些无关信息，从而提高其鲁棒性。
- **添加高斯噪声 (Adding Gaussian Noise):** 向数据中添加高斯噪声是一种常用的方法，旨在增强模型的鲁棒性。通过这种方式，模型能够在训练中适应噪声，从而提高其在噪声环境中的表现。
- **相位扰动 (Phase Perturbation):** 修改信号的频率域中的相位信息，而保持幅度不变来生成新的数据样本。这种方法保留了信号的整体结构，但引入了细微的扰动，用于提高模型的泛化能力。
- **块打乱 (RandomChunkShuffle):** 将时间序列数据分割成多个块，然后随机打乱这些块的顺序。这样可以创建不同的序列变体，增加模型对数据变异的适应能力，同时保持整体的语义不变。
- **随机缩放 (Random Scaled):** 对数据进行随机幅度缩放来增强数据的方法。通过改变数据的尺度，模型能够学习到不同幅度下的数据模式，从而提升其泛化能力。
- **随机绝对值 (Random Abs):** 对数据应用绝对值操作，将负值转换为正值或去掉负号。这个方法帮助模型处理包含负值的情形，增强其鲁棒性。
- **竖直翻转 (Random Vertical Flip):** 随机地将数据进行竖直翻转来创建新的样本。这对于一些对竖直方向变化不敏感的任务非常有效。
- **水平翻转 (Random Horizontal Flip):** 通过随机水平翻转图像或数据生成新样本，这对于具有较强对称性的数据特别有效，有助于提高模型的鲁棒性。
- **时移 (Time Shift):** 将数据在时间轴上平移一定的时间步长来生成新的样本。这种方法可以帮助模型学会适应时间序列数据中事件的变化位置，提升其对时间依赖的理解能力。

同一输入信号分别经过上述数据增强方法后的视图如图3-4所示。

简单暹罗孪生网络 (SimSiam) 的一个显著问题是可能出现所有输出“坍塌”为常数的平凡解。SimSiam 通过数据增强生成两个视图，并直接最大化同一输入的两个视图之间的相似性，而不依赖于负样本对。因此，其性能在很大程度上依赖于数据增强策略的质量(不改变原语义的同时丰富特征空间)。然而，人工设计数据增强策略不仅耗时耗力，还难以获得最优方案。如图3-5所示，基于人工设定数据增强策略训练的 SimSiam 网络在特征提取效果上存在明显不足，该策略未能有效提取样本间的区分性特征，导致相对大一部分类别的样本特征“糅杂”在一起。

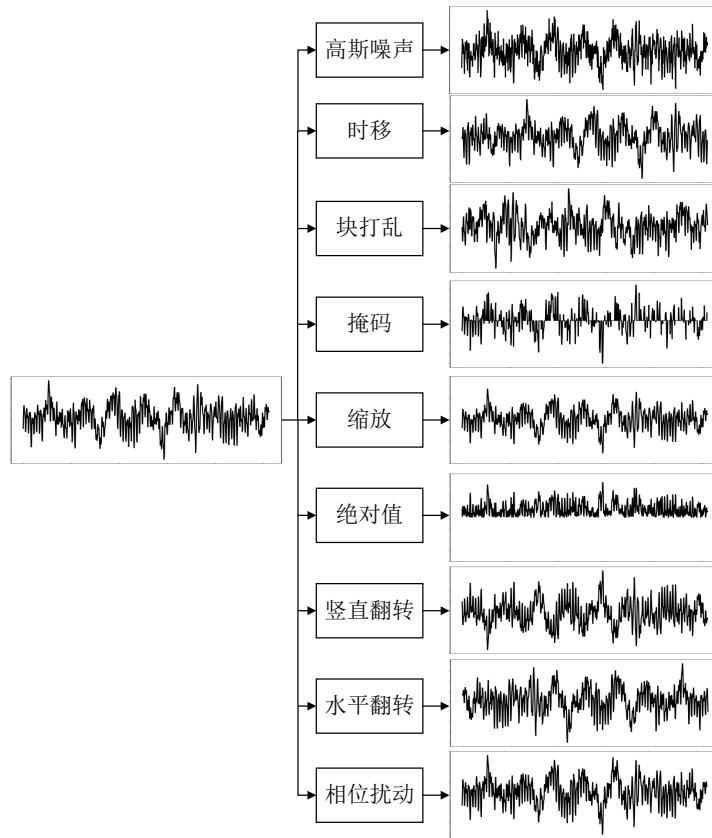


图 3-4 数据增强方法效果示意图

该节提出了基于协方差矩阵适应进化策略 (Covariance Matrix Adaptation Evolution Strategy, 下称 CMA-ES) 的最优数据增强策略搜索算法, 将寻找最佳数据增强策略的问题形式化为一个搜索问题(见图 3-6)。该方法由两个组件组成: 搜索算法和搜索空间。在高层次上, 搜索算法(实现为 CMA-ES)采样一个数据增强策略 S , 该策略包含有关使用哪种图像处理操作、每批次中使用该操作的概率以及操作幅度的信息。此方法的关键在于, 策略 S 将用于训练具有固定架构的神经网络, 其验证准确率 R 将返回以评估适应度。搜索空间细节: 在搜索空间中, 一个策略由 8 个信号数据增强子策略组成, 如图3-4所示。此外, 每个操作还关联两个超参数:

- (1) 应用操作的概率 $\in [0, 1]$,
- (2) 操作的幅度映射后 $\in [0, 10]$ 。

映射为线性变换, 如变量 $x \in [a, b]$ 到 $y \in [0, 10]$ 的变换为

$$y = \frac{10(x - a)}{b - a} \quad (3-15)$$

以下介绍各个数据增强子策略幅度的表示:

- **掩码 (Random Masked):** 随机选取 100 个不重叠的大小为 s 的子区间, 将其

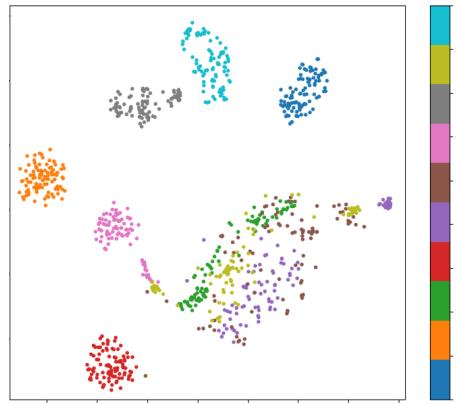


图 3-5 基于人工设定数据增强策略的 SimSiam 特征提取 t-SNE 图

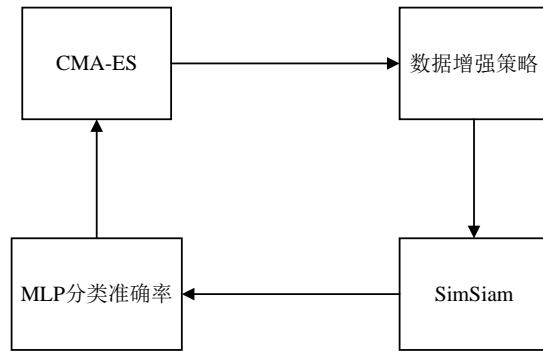


图 3-6 使用 CMA-ES 搜索最优的数据增强策略框架

置 0。 s 作为增强幅度由 $[1, 5]$ 映射为 $[0, 10]$ 。

- **添加高斯噪声 (Adding Gaussian Noise):** 将高斯噪声的信噪比 SNR 作为增强幅度，由 $[2, 6]$ 映射为 $[0, 10]$ 。
- **相位扰动 (Phase Perturbation):** 相位扰动服从均匀分布 $U(-perturb_{max}, perturb_{max})$ ， $perturb_{max}$ 作为增强幅度由 $[0.1, 0.5]$ 映射为 $[0, 10]$ 。
- **块打乱 (RandomChunkShuffle):** 将时间序列数据分割成大小均匀的 s 个块，然后随机打乱这些块的顺序。 s 作为增强幅度由 $[10, 100]$ 映射为 $[0, 10]$ 。
- **随机缩放 (Random Scaled):** 缩放的幅度服从均匀分布 $U(1.0 - scale, 1.0 + scale)$ 。 $scale$ 作为增强幅度由 $[0.05, 0.6]$ 映射为 $[0, 10]$ 。
- **随机绝对值 (Random Abs):** 无幅度值。
- **竖直翻转 (Random Vertical Flip):** 无幅度值。
- **水平翻转 (Random Horizontal Flip):** 无幅度值。

则 CMA-ES 优化器求解的问题可以描述为

$$\arg \max_{\mathbf{p}, \mathbf{s}} \text{Accuracy}(\mathbf{p}, \mathbf{s}, \text{SimSiam-Net}) \quad (3-16)$$

其中， \mathbf{p} 和 \mathbf{s} 为求解的参数，分别满足 $p_i \in [0, 1]$ 和 $s_i \in [0, 10]$ ，Accuracy 为简单暹罗网络验证的准确率。

3.1.4 实验设置

编码器 (Encoder), 预测器 (Predictor) 和分类器 (Classifier): 编码器是该算法中故障诊断的关键部分。本研究主要使用其潜在编码来完成分类任务。由于简单暹罗网络是一个框架，可以根据需要构建编码器模型和预测器，其中预测器是一个相对简单的非线性函数。在第二阶段，需要在预训练的编码器后添加一个分类器层用于故障诊断，如图3-1所示。为了方便起见，本研究使用多个常规的卷积神经网络 (CNN) 块来构建编码器，并使用全连接层来构建预测器和分类器。每个卷积块 (CB) 包括四个层：一个 1-D 卷积层 f_{BN} 、一个激活函数层 f_{ReLU} ，以及一个池化层 f_{Pool} 。方程 (3-17) 展示了 1-D 卷积层的操作，其中输入的形状为 (N, C_{in}, L) ，输出的形状为 $(N, C_{\text{out}}, L_{\text{out}})$ ，其中 N 是样本的数量， C 表示通道数， L 表示数据的长度。

$$\text{out}(N_i, C_{\text{out}_i}) = \text{bias}(C_{\text{out}_i}) + \sum_{k=0}^{C_{\text{in}}-1} \text{weight}(C_{\text{out}_i}, k) \otimes \text{input}(N_i, k). \quad (3-17)$$

BN 层 f_{BN} 用于加速训练过程并减少由于层输入分布不同而导致的内部协变量偏移 (ICS) 的影响 [53]，当模型较深时，这种影响尤为严重。公式 (3-18) 展示了批归一化的过程。

$$f_{\text{BN}}(x_i, \mathbf{x}) = \gamma \frac{x_i - \mathbb{E}(\mathbf{x})}{\sqrt{\sigma(\mathbf{x})^2 + \varepsilon}} + \beta \quad (3-18)$$

$$\mathbb{E}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N x_i \quad (3-19)$$

$$\sigma(\mathbf{x}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mathbb{E}(\mathbf{x}))^2} \quad (3-20)$$

其中 x 是整个小批量 (mini batch) 的向量，批量大小为 N ， x_i 表示小批量中的第 i 个样本。 $\mathbb{E}(x)$ 和 $\sigma(x)$ 分别表示向量批量的均值和方差，如公式 (3-19) 和 (3-20) 所示。 ε 是一个超参数，通常设置为 10^{-5} ，而 γ 和 β 分别是缩放因子和偏移因子，它们是可学习的参数，初始值分别设置为 1 和 0。激活函数层 f_{ReLU} 可以表示为公式

(3-21):

$$f_{\text{ReLU}} = \max(0, x) \quad (3-21)$$

$x_i^{(n)}$ 表示通过第 n 个卷积块 (CB) 的第 i 个输出向量，可以用公式 (3-22) 表示：

$$x_i^{(n)} = f_{\text{Pool}} \left\{ f_{\text{ReLU}} \left[f_{BN} \left(f_{\text{conv}} \left(\mathbf{x}^{(n-1)} \right), f_{\text{conv}} \left(x_i^{(n-1)} \right) \right) \right] \right\} \quad (3-22)$$

在该实验中，堆叠了十个 CNN 块作为编码器，其中卷积层的核大小设置为 3，填充设置为 1，步幅设置为 1，隐藏层采用 256 个核。使用两个线性全连接层构建预测器，分别具有 256×512 和 512×256 个神经元。在它们之间放置了一个批归一化层和一个修正线性单元 (ReLU) 层。分类器层部署了两个线性全连接层，分别具有 256×256 和 $256 \times \text{num_class}$ 个神经元，其中 num_class 表示故障类别的数量。在两个线性层之间也放置了一个批归一化层和一个 ReLU 层。

优化器 (Optimizer): 模型的训练不需要使用大批量优化器，例如分层自适应速率缩放 (LARS)^[54]，因为所提出的方法可以在典型批量大小下工作，而不依赖于大批量训练。本研究使用 SGD 优化器训练模型。网络参数通过公式 (3-23) 更新。

$$\theta_{l+1} = \theta_l - \eta_l \cdot \nabla_{\theta} \mathcal{L}(\mathbf{x}; \theta_l) \quad (3-23)$$

其中 θ_t 表示时间 t 时的可学习参数， η_t 表示时间 t 时的学习率， $L(\cdot)$ 表示损失函数。

学习率初始设置为 $0.05 \times \frac{\text{BatchSize}}{256}$ ，学习率采用余弦衰减调度。余弦衰减调度的公式为：

$$\eta_t = \eta_{\min} + \frac{1}{2} (\eta_{\max} - \eta_{\min}) \left(1 + \cos \left(\frac{t\pi}{T} \right) \right) \quad (3-24)$$

其中： η_t 是第 t 步的学习率， η_{\max} 是初始学习率 (最大学习率)， η_{\min} 是最小学习率， t 是当前训练步数， T 是总训练步数 (衰减周期)， \cos 是余弦函数。权重衰减为 0.0001，SGD 动量为 0.9。

损失函数： 对比学习预训练阶段的损失函数为余弦相似度见式 (3-1)，微调阶段的损失函数为交叉熵损失函数 (Cross-Entropy Loss Function)。对于多分类问题，交叉熵损失函数表示为

$$\mathcal{L}_{\text{CE}} = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (3-25)$$

其中： C 是类别数量， y_i 是真实标签的 one-hot 编码 (第 i 类的真实概率)， \hat{y}_i 是模型预测的第 i 类的概率。

数据细节： 根据不同不平衡因子构造帕累托分布的有标签数据集，测试集为

均匀分布。每个样本包含 1024 个数据点。训练过程分为两个阶段：

- 在第一阶段，随机选择每个故障类别的 100 个无标签样本进行模型训练，训练总计 500 个周期。
- 在第二阶段，根据式 (2-13) 构造服从帕累托分布的长尾分布有标签样本，以此微调模型，微调总计 150 个周期。

在整个训练过程中，每个输入模型的 mini-batch 大小设定为 **64**。**微调 (Fine-tuning)**：在对比学习阶段完成后，从模型中取出训练好的编码器 f ，并将其与一个多层次感知器 (MLP) 模块连接，该 MLP 部署了两个线性全连接层，分别具有 256×256 和 $256 \times \text{num_class}$ 个神经元，其中 num_class 表示故障类别的数量。在两个线性层之间也放置了一个批归一化层和一个 ReLU 层。用带标签的数据微调构建一个完整的分类器用于故障诊断任务，从而使模型获得分类能力。以准确率为性能指标，如公式 (3-26)：

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN} \quad (3-26)$$

其中， TN 表示真阴性，即正确预测为阴性的样本数； TP 表示真阳性，即正确预测为阳性的样本数； FP 表示假阳性，即错误分类为阳性的阴性样本数； FN 表示假阴性，即错误分类为阴性的阳性样本数。

当验证集的样本均匀采样后(本研究采用此方法)，准确率等于宏平均召回率 (Macro-Averaged Recall)。宏平均召回率是对每个类别单独计算召回率，然后取平均值：

$$\text{Macro-Recall} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FN_i} \quad (3-27)$$

其中， C 是类别的总数， TP_i 和 FN_i 分别是第 i 类的真正例和假负例。

性能评估：测试集由每个故障类别均匀地随机选取 **100** 个样本组成。需要注意的是，微调过程中使用的子集、对比学习过程中用到的无监督数据集与测试集完全独立。用到的三种性能评估指标如下：

- 验证集准确率/宏平均召回率：在验证集上评估模型的分类准确率，以衡量模型的泛化能力。该指标表示了模型对未知数据的分类能力。计算公式如式 (3-26)。宏平均召回率用于评估模型对各个类别的分类准确率，由于验证集的样本均匀分布，其数值上与准确率相等，计算公式如式 (3-27)。
- t-SNE 特征提取可视化：使用 t-SNE 方法将高维特征降维至二维空间进行可视化分析，直观展示不同类别之间的聚类情况。该方法有助于分析模型对不同故障类别的区分能力。

- 特征标准差：如果输出 z 具有零均值各向同性高斯分布，可以证明 $\frac{z}{\|z\|_2}$ 的标准差为 $\frac{1}{\sqrt{d}}$ 。下证：假设 $z \in \mathbb{R}^d$ 是一个零均值各向同性高斯分布，即 $z \sim \mathcal{N}(0, I_d)$ ，其中 I_d 是 d 维单位矩阵，表示协方差矩阵为单位矩阵，意味着每个分量 z_i 独立且服从标准正态分布。需要证明的是， $\frac{z}{\|z\|_2}$ 的标准差（即每个分量的标准差）为 $\frac{1}{\sqrt{d}}$ ，其中 $\|z\|_2 = \sqrt{\sum_{i=1}^d z_i^2}$ 是 z 的 L_2 范数。以下证明：

设 $z = (z_1, z_2, \dots, z_d)$ ，则 z 的每个分量 z_i 都是独立的标准正态随机变量 $z_i \sim \mathcal{N}(0, 1)$ ，并且 $\|z\|_2 = \sqrt{\sum_{i=1}^d z_i^2}$ 是 z 的欧几里得范数。关注向量 $\frac{z}{\|z\|_2}$ 的每个分量。令 $\hat{z} = \frac{z}{\|z\|_2}$ ，即 \hat{z} 是单位向量，表示将 z 投影到单位球面上的结果。因为 z 是各向同性高斯分布，其方向是均匀分布在单位球面上的。对于标准化后的向量 \hat{z} ，每个分量 $\hat{z}_i = \frac{z_i}{\|z\|_2}$ 的分布不再是标准正态分布，而是受到范数约束的结果。由于 z 是各向同性的， \hat{z}_i 的分布是对称的，所有分量 \hat{z}_i 的期望值为 0。因此，接下来计算 \hat{z}_i 的方差。对于 \hat{z}_i 的方差，有：

$$\text{Var}(\hat{z}_i) = \mathbb{E}[\hat{z}_i^2] - (\mathbb{E}[\hat{z}_i])^2 \quad (3-28)$$

由于 $\mathbb{E}[\hat{z}_i] = 0$ ，只需要计算 $\mathbb{E}[\hat{z}_i^2]$ 。

$$\mathbb{E}[\hat{z}_i^2] = \mathbb{E}\left[\frac{z_i^2}{\|z\|_2^2}\right] = \frac{\mathbb{E}[z_i^2]}{\mathbb{E}[\|z\|_2^2]} = \frac{1}{d} \quad (3-29)$$

其中 $\mathbb{E}[z_i^2] = 1$ （因为 z_i 是标准正态分布），而 $\mathbb{E}[\|z\|_2^2] = \mathbb{E}[\sum_{i=1}^d z_i^2] = d$ ，因为每个 z_i^2 的期望为 1。由于方差 $\text{Var}(\hat{z}_i) = \frac{1}{d}$ ，因此标准差是：

$$\text{std}(\hat{z}_i) = \frac{1}{\sqrt{d}} \quad (3-30)$$

- 验证集 KMeans 分类准确率：在验证集上使用 KMeans 算法对 Encoder 提取的特征进行无监督分类，并计算分类准确率。通过与真实标签对比，评估模型在无监督情境下的表现，进一步验证其分类能力。

故障诊断 (Fault Diagnosis): 在完成微调后，即可将模型应用于真实输入信号的故障诊断，以预测故障类别。

3.2 实验与分析

为了验证所提出方法的有效性，本研究选择了 CWRU 数据集和 PU 数据集进行实验。为了验证所提方法的优越性，选择了几种流行的对比方法，分为传统机器学习算法和基于深度学习的智能方法两类。

(1) **SVM**^[55]：多类 SVM 是一种强大且多功能的机器学习模型，能够处理线性或

非线性的分类任务，通过构建多个二分类器来实现对多个类别的区分。首先，将计算原始数据的 16 个时域指标(均值、平方根、偏度等)，并将它们作为 SVM 分类器的输入。

- (2) **CNN**^[56]: CNN 是一种监督学习方法。在这里，CNN 的网络架构与所提出方法的编码器 f 相同。
- (3) **BYOL**^[33]: BYOL(Bootstrap Your Own Latent) 是一种基于自监督学习的表示学习方法。与传统的对比学习方法不同，BYOL 不依赖于负样本对进行训练，而是通过最大化正样本对之间的相似度，从而学习到更具判别性的图像表示。该方法采用两个神经网络模块，其中一个作为在线网络，另一个作为目标网络，通过不断更新目标网络来提高模型的稳定性和性能。两个网络共享相同的编码器 f ，但参数更新的方式不同：在线网络的参数通过反向传播进行更新，而目标网络的参数则通过指数滑动平均进行更新。编码器 f 和数据增强方法与所提出方法相同。
- (4) **SimCLR**^[32]: SimCLR 是一种基于对比学习的自监督表示学习方法，通过构造正负样本对并最大化正样本对的相似度，从而学习出具有高质量表示的特征。该方法使用一个编码器 f ，通常是一个卷积神经网络(CNN)，并通过数据增强技术生成不同的图像视图。SimCLR 的训练目标是最大化正样本对的相似性，同时最小化负样本对的相似性，来优化表示学习的质量。该方法的关键创新是引入了基于温度缩放的对比损失函数，进而提升了模型的表达能力。编码器 f 和数据增强方法与所提出方法相同。

3.2.1 数据增强策略搜索的结果

采用 CWRU 数据集作为对比学习预训练的无标签数据，样本总数为 **500** 个。用于微调的有标签数据共 **100** 个，涵盖 **10** 个类别，每类包含 **10** 个样本。各类别信号的时域波形如图 2-3 所示。

在优化过程中，基于协方差矩阵适应进化策略 (CMA-ES) 进行最优数据增强策略的搜索。CMA-ES 的超参数设定如下：变量维度为 **14**(即 $\text{len}(p) + \text{len}(s)$)，种群规模设定为 **11**，最大迭代次数设定为 **10** 轮。经过优化后，所获得的最优数据增强策略如表 3-2 所示。

3.2.2 在 CWRU 数据集和 PU 数据集上的实验结果

在本节中，采用表3-2 所示的数据增强参数，并使用 CWRU 数据集和 PU 数据集作为数据集。对比学习预训练的无标签数据集的样本总数为 $100 \times$ 类别数个，用于微调的有标签数据共计 $10 \times$ 类别数个，并通过帕累托分布生成不同不平衡因

表 3-2 CMA-ES 搜索最优的数据增强策略参数

增强编号	增强名称	概率 p	幅度 s
DA0	高斯噪声	0.9761	9.6284
DA1	相位扰动	0.9899	3.4435
DA2	块打乱	0.9687	1.3436
DA3	掩码	0.7873	8.1961
DA4	缩放	0.5776	7.8206
DA5	绝对值	0.4020	—
DA6	竖直翻转	0.4392	—
DA7	水平翻转	0.4420	—

子的长尾分布标记样本。最终，在样本分布均匀的测试集上计算分类准确率，并进行 5 次重复试验取平均值。表3-3 和表3-4 展示了在不同不平衡因子 β 下，五种方法 (SimSiam、SimCLR、BYOL、CNN 及 SVM) 在 CWRU 数据集与 PU 数据集上的分类准确率与宏平均召回率。宏观上看，对比学习方法 (尤其是 SimCLR 和 SimSiam) 在不同程度的不平衡条件下均表现出较强的鲁棒性与稳定性，显著优于传统的浅层方法 SVM 及基于监督学习的 CNN。

在 CWRU 数据集上，SimCLR 在所有 β 值下均取得了最优或接近最优的结果，平均准确率/召回率达到 0.9265，表现最为稳健。SimSiam 紧随其后，平均值为 0.9131，亦具备良好的不平衡适应能力。相较之下，BYOL 平均性能为 0.8610，略低于前两者，而 CNN 和 SVM 的平均值分别为 0.7420 与 0.4305，显示出其对不平衡分布较为敏感，尤其在 $\beta = 100$ 时性能大幅下降。

类似地，在 PU 数据集上，SimCLR 与 SimSiam 分别以 0.6494 和 0.6412 的平均值占据前两位，显著高于 BYOL(0.6001)、CNN(0.5211) 与 SVM(0.2989)。尤其值得注意的是，随着不平衡程度的加剧 (β 从 1 增至 100)，SVM 表现急剧下滑，说明其对类别分布极为敏感，而对比学习方法仍能保持相对稳定的性能。

图3-7 展示了 SimSiam、SimCLR、BYOL、CNN 和 SVM 五种方法 (CNN 和 SVM 基于不平衡因子 β 为 1 训练) 在均匀分布与长尾分布数据集上的 t-SNE 特征投影。通过这些图示，可以直观地观察不同类别样本在低维空间中的分布情况，从而评估各方法的特征提取效果。SimSiam 的 t-SNE 图 (图3-7(a)) 展示了相对最佳的特征分布，其中各类别之间的边界清晰，且不同类别的样本未出现明显的混合，类别间的聚类效果较好。相比之下，其他模型的 t-SNE 分布图普遍存在不同类别样本的交叉重叠现象，这对线性分类器进行分类时构成了较大挑战。但这与表3-3 的结果出入，在下一章中我们将着重解决该问题。

表 3-3 CWRU 数据集 SimSiam、SimCLR、BYOL、CNN 与 SVM 在不同 β 值下的准确率/宏平均召回率

不平衡因子 β	SimSiam	SimCLR	BYOL	CNN	SVM
1	0.9208	0.9398	0.8904	0.8979	0.7198
10	0.9423	0.9458	0.8942	0.8538	0.2990
50	0.9133	0.9198	0.8481	0.6804	0.3917
100	0.8760	0.9004	0.8112	0.5358	0.3115
平均值	0.9131	0.9265	0.8610	0.7420	0.4305

表 3-4 PU 数据集 SimSiam、SimCLR、BYOL、CNN 与 SVM 在不同 β 值下的准确率/宏平均召回率

不平衡因子 β	SimSiam	SimCLR	BYOL	CNN	SVM
1	0.6443	0.6543	0.6246	0.6273	0.5036
10	0.6627	0.6597	0.6229	0.6018	0.2046
50	0.6418	0.6511	0.5870	0.4770	0.2710
100	0.6158	0.6324	0.5660	0.3784	0.2163
平均值	0.6412	0.6494	0.6001	0.5211	0.2989

3.2.3 消融实验

本小节基于 CWRU 数据集对 SimSiam 网络的停止梯度模块、预测器模块和数据增强模块的作用进行了实验分析，并对无标签数据集规模、Batch Size 等训练关键参数进行了实验。

停止梯度和预测期模块

图3-8展示了“使用停止梯度(Stop-Grad)”与“不使用停止梯度”在 CWRU 数据集上的比较。架构和所有超参数保持不变，唯一的区别是是否使用停止梯度或预测器。图3-8(a)显示了训练损失。如果不使用停止梯度，优化器可能会迅速找到一个退化解，并使损失函数达到最小值 -1 ，但这通常并不是一个有效的解，因为模型可能会过拟合或陷入局部最优解。为了证明这种退化是由坍缩引起的，以下研究了 ℓ_2 归一化输出 $z/\|z\|_2$ 的标准差。如果输出坍缩到一个常数向量，则所有样本的标准差在每个通道上应该为零，这可以从图3-10(b)中的蓝色或绿色曲线中观察到。作为对比，如果输出 z 具有零均值各向同性高斯分布， $\frac{z}{\|z\|_2}$ 的标准差为 $\frac{1}{\sqrt{d}}$ ，从图3-10(b)的红色曲线可以观察到。

图3-8(b)的红色曲线显示，使用停止梯度时，标准差接近 $\frac{1}{\sqrt{d}}$ 。这表明输出没有坍缩，而是分散在单位超球面上。图3-8(c)绘制了 KMeans 聚类的验证准确率。

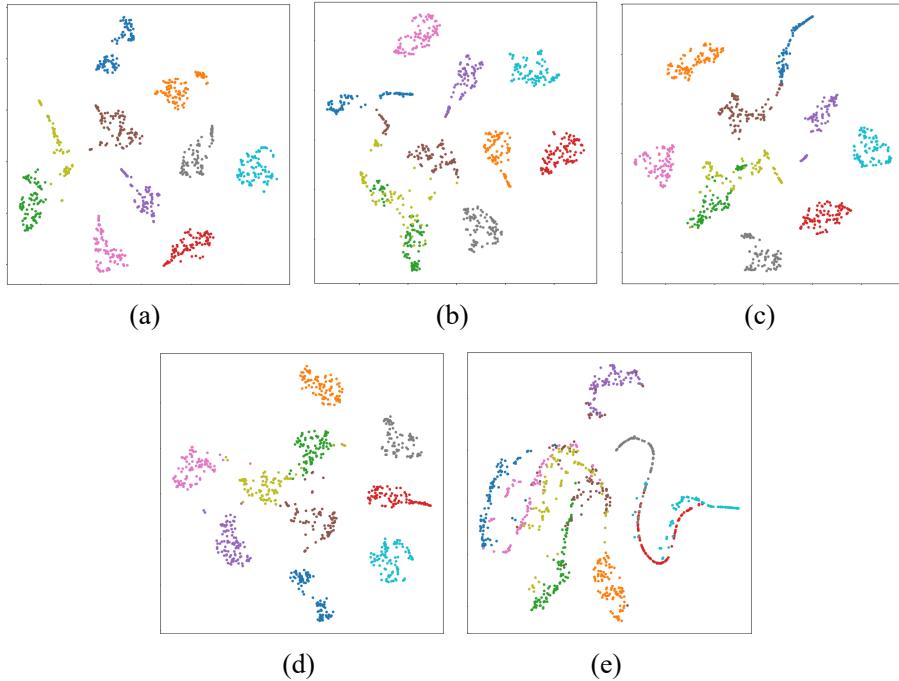


图 3-7 验证集数据经特征提取的 T-SNE 可视化: (a) SimSiam 的 T-SNE 分布图; (b) SimCLR 的 T-SNE 分布图; (c) BYOL 的 T-SNE 分布图; (d) CNN 的 T-SNE 分布图; (e) SVM 的 T-SNE 分布图

KMeans 聚类可以作为进展的监控工具。使用停止梯度时, 监控显示出准确率稳步提高。图3-8的绿色曲线显示模型在移除预测器 (Predictor) 后的表现, 可以看出移除预测器后发生了与移除 Stop-Grad 类似的坍塌。事实上, 如果使用对称损失 (式 (3-2)), 这一现象是可以预期的。现在的损失函数是

$$\frac{1}{2}D(z_1, \text{stopgrad}(z_2)) + \frac{1}{2}D(z_2, \text{stopgrad}(z_1)). \quad (3-31)$$

其梯度的方向与 $D(z_1, z_2)$ 的梯度相同, 只不过幅度被缩放了 $\frac{1}{2}$ 。在这种情况下, 移除 Stop-Grad 相当于将损失函数缩放为 $\frac{1}{2}$ 。这一调整导致了模型的崩溃 (见图3-8)。图3-9进一步可视化了该崩溃对特征提取的影响: 在移除 Stop-Grad 或预测器后, 特征空间中的大量样本交叉重叠, 导致其变得难以区分。表3-5也显示了移除 Stop-Grad 或预测器后的巨大性能损失。实验表明, Stop-Grad 模块对防止模型塌陷起到至关重要的作用, 当移除该模块后, 通过观察最小可能的损失值以及模型输出的常数性 (标准差), 可以发现这种塌陷现象。塌陷解的存在表明, 仅通过架构设计 (如预测器、BN、 ℓ_2 -范数) 来防止塌陷是不充分的。

数据增强模块

3.2.1小节在限定范围内求得了各个数据增强子策略的最优参数解, 并在之后

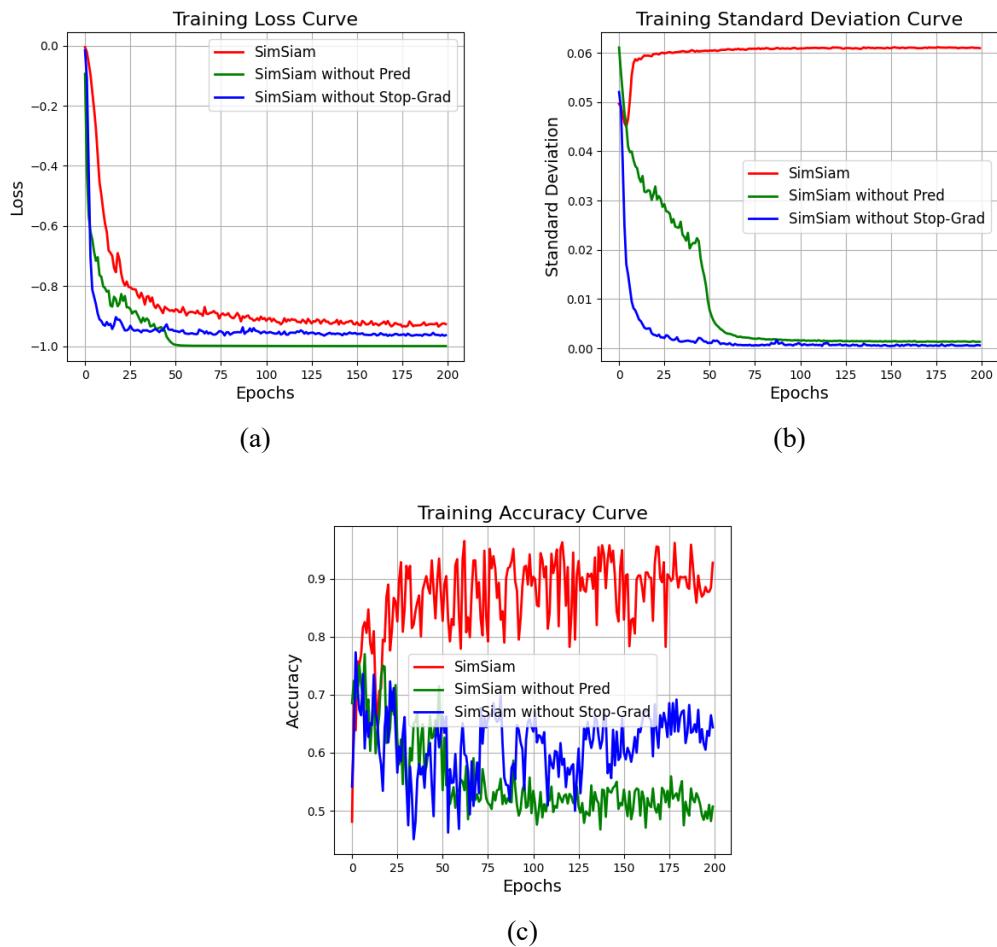


图 3-8 SimSiam, SimSiam(无 Stop-Grad 模块)与 SimSiam(无预测器模块)的训练过程对比 (a) 损失函数; (b) 提取特征的标准差; (c) K-Means 聚类准确率

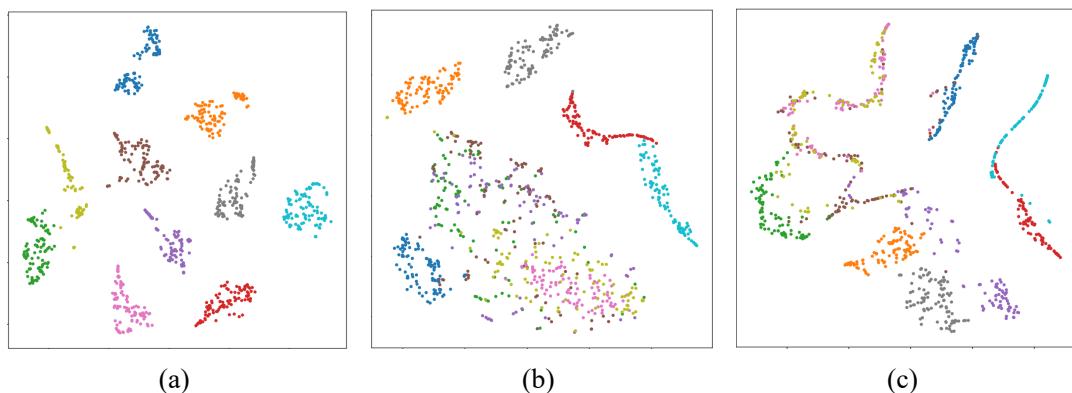


图 3-9 SimSiam, SimSiam(无 Stop-Grad 模块)与 SimSiam(无预测器模块)的 T-SNE 图 (a) SimSiam; (b) SimSiam(无 Stop-Grad 模块); (c) SimSiam(无预测器模块)

表 3-5 SimSiam、SimSiamw/o SG 与 SimSiamw/o Pred 在不同 β 值下的准确率/宏平均召回率

不平衡因子 β	SimSiam	SimSiamw/o SG	SimSiamw/o Pred
1	0.9515	0.5338	0.5217
10	0.9402	0.4919	0.4604
50	0.9131	0.3579	0.3892
100	0.8450	0.3412	0.3704
平均值	0.9125	0.4312	0.4354

的实验中取得了良好的结果。以下通过实验分析表3-2中的8个数据增强子策略对模型性能的影响。图3-10显示了移除单个子策略后模型训练过程中的损失函数及特征标准差。表3-6移除不同数据增强子策略的准确率性能变化。以下分析了不同数据增强方法对模型性能的影响。具体而言，讨论了以下几种增强技术：缩放、块打乱和掩码、高斯噪声和相位扰动、绝对值以及竖直翻转和水平翻转。

首先，缩放操作对模型性能的影响最为显著。从表3-6的结果可以看出，移除缩放数据增强后，模型的性能显著下降。振动信号的幅值在时间轴上并非稳定，而是存在波动，这表明信号的特征在不同的尺度和幅度上有变化。引入缩放操作后，模型能够在较小的时间范围内提取出更大幅值范围的特征，从而增强了模型的泛化能力。缺少缩放操作后，模型在不同幅度范围内的鲁棒性降低，导致性能下降。通过增加数据的多样性，缩放操作帮助模型适应了不同的目标尺度和视野范围，进而提升了模型的表现。

其次，块打乱和掩码操作通过局部区域的扰动提高了模型的鲁棒性和对局部特征的提取能力。具体而言，块打乱操作通过打乱图像或信号的局部区域，迫使模型学习局部特征而非依赖全局信息，从而避免模型对局部区域的过度拟合。而掩码操作通过对部分区域进行遮挡，模拟数据中的缺失或遮挡情况，促使模型从有限的信息中恢复或推断出其余部分。这两种数据增强方法在特定场景下，尤其是在存在局部信息缺失或遮挡的情况下，能够有效提升模型的鲁棒性和泛化能力。

高斯噪声和相位扰动主要通过引入轻微的随机噪声或频域变化来增加数据的扰动。尽管这类增强方法能够增加数据的多样性，但生成的样本与原始数据的差异较小，可能无法充分捕捉到数据的多种变异情况，从而限制了模型的泛化能力，未能有效扩展模型的学习能力，因此对模型性能的影响较为有限。

绝对值操作通过将信号中的负值转化为正值，从而扩展了模型所接触到的样本空间。尤其是在处理具有周期性或对称特征的信号时，绝对值能够引入更多样化的训练样本。这种方法帮助模型适应不同的变化模式，并有效防止模型过早陷

表 3-6 移除不同数据增强子策略在不同 β 值下的准确率/宏平均召回率

不平衡因子 β	DA0	DA1	DA2	DA3	DA4	DA5	DA6	DA7
1	0.9279	0.9127	0.9254	0.9204	0.9106	0.9298	0.9394	0.9338
10	0.9258	0.9348	0.8944	0.9298	0.8792	0.9358	0.9333	0.9442
50	0.8563	0.8379	0.8221	0.8467	0.7490	0.8165	0.8573	0.8387
100	0.7935	0.8169	0.7356	0.7194	0.7119	0.7912	0.8333	0.7869
平均值	0.8759	0.8756	0.8444	0.8541	0.8127	0.8683	0.8908	0.8759
相比 Baseline	-0.0366	-0.0369	-0.0681	-0.0584	-0.0998	-0.0442	-0.0217	-0.0366

入“塌陷”现象。此外，绝对值操作增强了模型对信号反转的鲁棒性，有助于提升模型的泛化能力，特别是在面对多变输入数据时。

最后，竖直翻转和水平翻转操作在处理周期性信号时效果较为有限。周期性信号通常具有特定的模式或周期性变化，这些翻转操作改变了信号的空间结构，但对信号的本质特征影响较小。因此，这两种操作对模型性能的提升相对较弱，尤其是在周期性信号的情况下。

综上所述，不同的数据增强方法对模型性能的提升作用不同。缩放操作显著增加了数据的多样性，提升了模型的泛化能力，因此对模型性能的提升最为显著。其次，块打乱和掩码操作在局部信息缺失或遮挡的情况下有助于提高模型的鲁棒性。尽管高斯噪声和相位扰动对模型性能的提升作用较小，但它们仍然能够增加数据的多样性。最后，绝对值操作有效防止了模型的“塌陷”现象，增强了模型对信号反转的鲁棒性，而竖直翻转和水平翻转的作用较为有限，特别是在处理周期性信号时。

无标签数据集规模

根据前文的讨论，SimSiam 网络的性能可能在很大程度上依赖于编码器在无标签数据上进行自监督预训练的质量，现在讨论无标签样本规模对 SimSiam 网络性能的影响。在以下实验中，固定了训练集的样本规模为 100，并分别设置了无标签数据集的规模为 100、200、500、1000 和 2000。通过对比分析，评估了不同无标签数据集规模下模型的性能表现。从表3-7和表3-8中可以看出，随着无标签数据集规模的增大，模型的性能普遍有所提升，尤其是在较大的数据集规模下，准确率的提升更为显著。

在无标签数据集规模为 100 时，SimSiam 网络的准确率相对较低，尤其是在较大 β 值配置下（例如， $\beta = 100$ 时，准确率仅为 0.7200）。这一现象表明，在较小的数据集规模下，模型的自监督预训练效果不佳，未能有效地学习到足够的特征，从而导致性能较差。

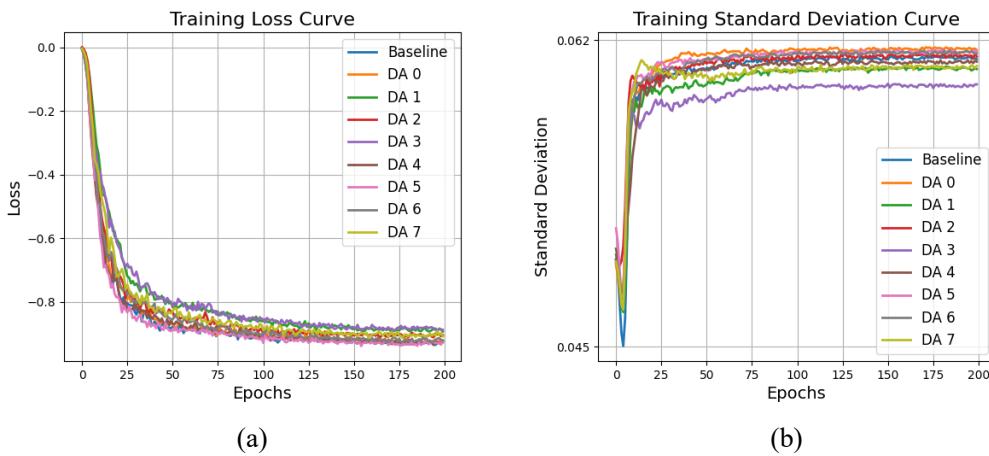


图 3-10 移除不同数据增强子策略的训练过程对比 (a) 损失函数; (b) 提取特征的标准差

随着无标签数据集规模的增大，模型的准确率逐渐提升(见图3-11)，尤其在数据集规模达到2000时，网络的表现达到最佳，准确率为0.9238。这表明，增加无标签数据集的规模对于模型性能的提升有显著作用。表格中的平均值进一步验证了这一点，随着数据集规模的增加，SimSiam网络的整体表现也逐步提高。在数据集规模为2000时，所有不平衡因子下的平均准确率均较高，凸显了大规模无标签数据集在自监督学习中的重要性。

然而，值得注意的是，当数据规模达到 500 及以上时，模型性能的提升趋于平稳。通过 t-SNE 分布图(图3-12)，可以观察到特征提取质量的改善相对有限，特别是在数据规模从 200 增加到 500 时，性能提升最为显著，而从 500 到 2000 的提升则较为缓慢。这进一步表明，虽然大规模无标签数据集对模型性能有积极作用，但在一定规模后，性能提升的边际效益逐渐递减。

表 3-7 CWRU 数据集不同无标签数据集规模配置下的 SimSiam 准确率/宏平均召回率

无标签数据集规模	$\beta = 1$	$\beta = 10$	$\beta = 50$	$\beta = 100$	平均值
100	0.8988	0.8942	0.8088	0.7200	0.8305
200	0.8710	0.8642	0.7479	0.7454	0.8071
500	0.9463	0.9577	0.8713	0.7929	0.8921
1000	0.9513	0.9508	0.9025	0.8746	0.9198
2000	0.9548	0.9538	0.8946	0.8919	0.9238

Batch Size

表 3-8 PU 数据集不同无标签数据集规模配置下的 SimSiam 准确率/宏平均召回率

无标签数据集规模	$\beta = 1$	$\beta = 10$	$\beta = 50$	$\beta = 100$	平均值
100	0.6187	0.6221	0.5740	0.5213	0.5840
200	0.6297	0.6351	0.5815	0.5348	0.5953
500	0.6624	0.6704	0.6099	0.5549	0.6244
1000	0.6659	0.6656	0.6318	0.6103	0.6434
2000	0.6684	0.6677	0.6253	0.6243	0.6464

文献 [32] 提到， 较大的批大小 (下称 Batch Size) 能够显著提高对比学习模型的性能。在本节中，通过实验探讨了不同 Batch Size 配置下 SimSiam 网络的表现。表3-9展示了不同 Batch Size(16, 32, 64, 128, 256, 512) 下，SimSiam 网络在不同不平衡因子下的准确率。

对于 $\beta = 1$ 和 $\beta = 10$ 的情况，SimSiam 网络在不同 Batch Size 下的表现相对稳定，唯一的显著性能下降发生在 Batch Size 为 512 且 $\beta = 1$ 时。这表明，较大的 Batch Size 可能对模型的训练稳定性产生负面影响，尤其是在低不平衡因子的配置下，可能导致训练过程中出现不稳定现象。

另一方面，针对 $\beta = 50$ 和 $\beta = 100$ 的配置，当 Batch Size 较小 (16 和 32) 时，模型的准确率表现出较大的波动。这种波动可能与小 Batch Size 导致的训练不充分或梯度估计不稳定有关。然而，当 Batch Size 增大至 64 及以上时，模型表现趋于稳定，说明较大的 Batch Size 有助于减小训练过程中波动，特别是在高不平衡因子的情况下。

从 t-SNE 分布图 (图3-14) 中也可以观察到，当 Batch Size ≥ 64 时，特征提取的质量保持稳定，各个类别的边界更加清晰；而在 Batch Size 为 16 或 32 时，部分点的分布呈现出交叠现象，表明个别点特征分离度较差。综合来看，SimSiam 网络的性能对 Batch Size 的依赖相对较小，整体表现较为稳定，这进一步证明了在一定范围内，Batch Size 的增大能够优化训练过程并稳定模型的性能。

反观 SimCLR 在不同 Batch Size 下的表现 (见图 3-13)，随着 Batch Size 的增大，SimCLR 模型的准确率呈现出明显的下降趋势，而 SimSiam 和 BYOL 的表现稳定，BYOL 的性能在 Batch Size 为 128 时达到峰值，且是所有不平衡因子下都为最高性能 (见表3-11)。从表3-10也可以看到，当 Batch Size 增大，SimCLR 在不同不平衡因子的准确率全线下降。表明较大的 Batch Size 可能导致 SimCLR 模型学习到的特征质量下降。

此外，通过观察 t-SNE 分布图 (见图 3-15)，进一步验证了这一现象。在较小的

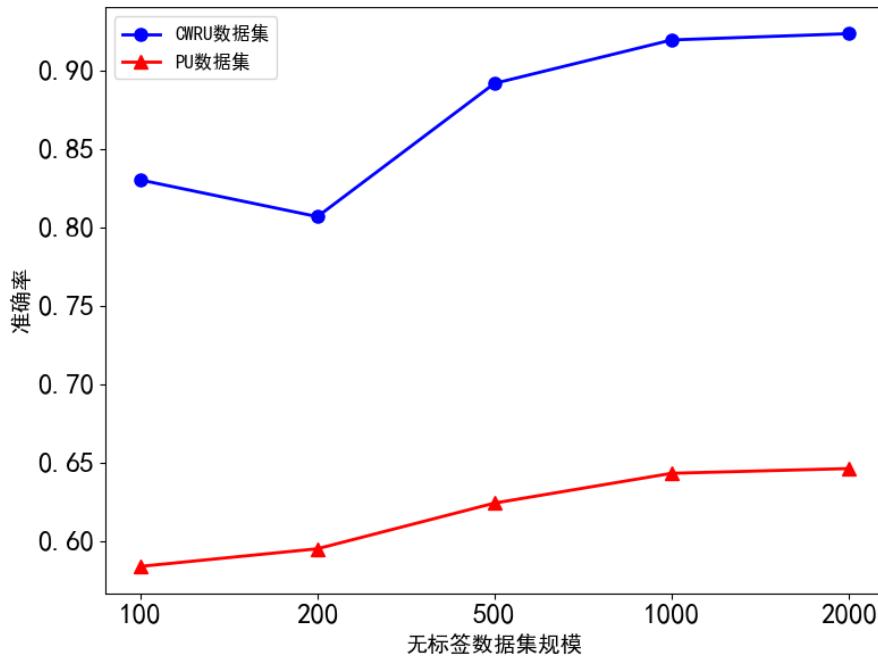


图 3-11 不同无标签数据集规模配置下的 SimSiam 准确率/宏平均召回率变化图

Batch Size 下，特征的分布较为紧凑，且样本间的区分度较高。然而，随着 Batch Size 的增大，特征分布逐渐变得难以区分，特征提取的质量显著下降。据推测，这一变化可能与较大 Batch Size 与样本类别总数之比有关。在 SimCLR 中，同一批次中的样本通常被视为负样本，而当 Batch Size 增大时，同一批次中同类别样本的比例更高，从而导致模型倾向于增大同类样本之间的距离，这与对比学习的目标相悖——即相似样本应该聚集。此现象与文献 [32] 中提出的“较大的 Batch Size 对对比学习模型性能有益”的观点存在不一致之处，表明过大的 Batch Size 可能会损害对比学习的训练效果。

表 3-9 不同 Batch Size 配置下的 SimSiam 准确率/宏平均召回率

Batch Size	$\beta = 1$	$\beta = 10$	$\beta = 50$	$\beta = 100$	平均值
16	0.9554	0.9567	0.9188	0.8798	0.9277
32	0.9492	0.9519	0.8600	0.7967	0.8895
64	0.9513	0.9508	0.9025	0.8746	0.9198
128	0.9433	0.9498	0.8804	0.8521	0.9064
256	0.9508	0.9692	0.8900	0.8131	0.9058
512	0.9258	0.9421	0.8748	0.8435	0.8966

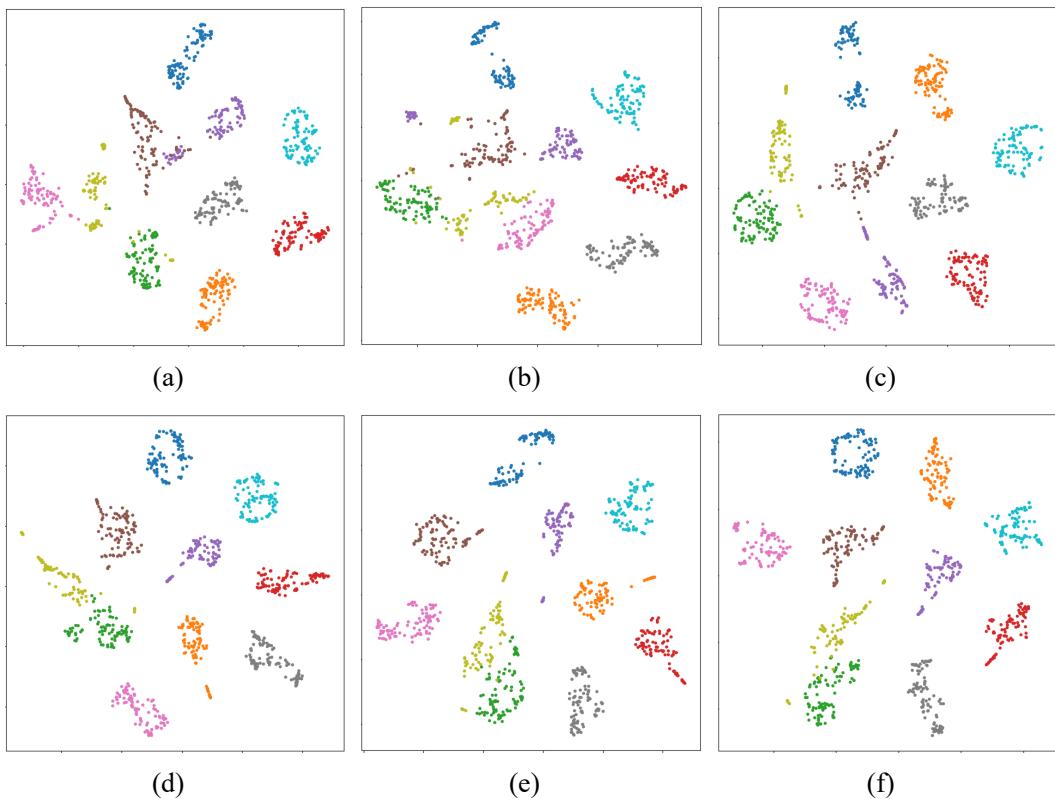


图 3-12 不同无标签数据规模的 SimSiam 网络输出特征的 T-SNE 分布图: (a) 数据规模为 100; (b) 数据规模为 200; (c) 数据规模为 500; (d) 数据规模为 1000; (e) 数据规模为 1500; (f) 数据规模为 2000

3.2.4 准确率、损失函数和特征标准差之间的相关性分析

准确率、损失函数以及特征标准差是评估机器学习和深度学习的模型性能和学习过程的重要指标。准确率作为衡量模型预测正确性的标准，通常用于反映模型在分类任务中的表现；而损失函数则用来度量模型输出与真实标签之间的差距，指导优化器调整模型参数；特征标准差则能够从一定程度上反映模型“塌陷”的程度，同时反映特征提取的质量。

尽管这些指标分别在不同方面评价模型的表现，但它们之间是否存在内在的相关性，以及这种相关性是否能够提供更深刻的洞察，仍然是值得探讨的问题。准确率与损失函数之间的关系通常较为直观：随着损失函数的减少，模型的预测精度有望提高。然而，特征标准差在此过程中所起的作用可能更加复杂。特征标准差的变化可能在一定程度上影响模型对数据分布的适应能力，进而间接影响最终的准确率。

因此，本节旨在深入探讨准确率、损失函数和特征标准差之间的相关性，通

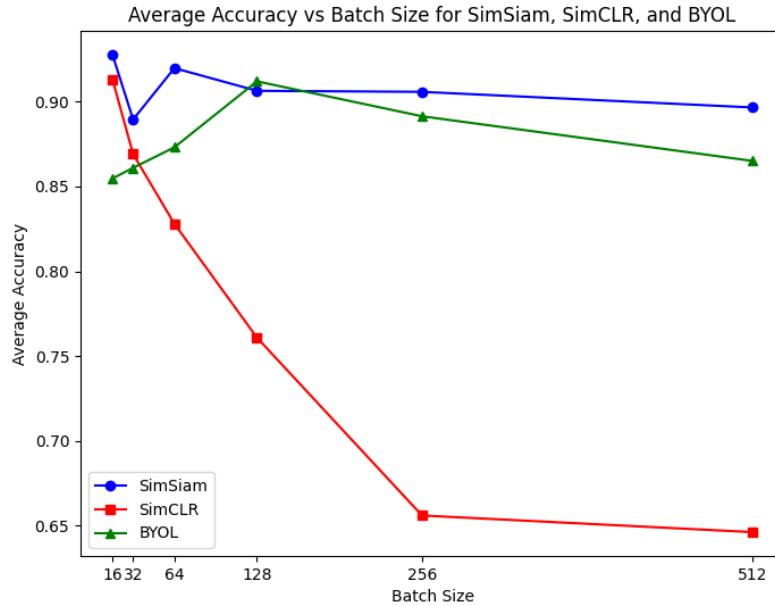


图 3-13 不同 Batch Size 下 SimSiam, SimCLR, BYOL 的准确率/宏平均召回率变化图

表 3-10 不同 Batch Size 配置下的 SimCLR 准确率/宏平均召回率

Batch Size	$\beta = 1$	$\beta = 10$	$\beta = 50$	$\beta = 100$	平均值
16	0.9208	0.9423	0.9133	0.8760	0.9131
32	0.9108	0.9090	0.8556	0.8010	0.8691
64	0.8402	0.8535	0.8323	0.7860	0.8280
128	0.7935	0.8038	0.7388	0.7073	0.7609
256	0.7100	0.7292	0.6229	0.5617	0.6560
512	0.7185	0.7275	0.5940	0.5448	0.6462

过量化这些指标的关系，揭示它们在模型训练中的互动机制，以便为模型优化和性能提升提供理论依据。通过这一分析，可以更全面地理解训练过程中各项指标之间的相互作用，进而指导实际应用中的模型调整与改进。

Loss 与 Std: 图3-17显示了图3-10中的 Loss 曲线与标准差曲线的皮尔森相关系数 (Pearson Correlation) 与斯皮尔曼相关系数 (Spearman Correlation)。以下是对结果的分析：

在 Baseline 中，模型包含了所有的数据增强策略，Pearson 相关系数为-0.9448，Spearman 相关系数为-0.9232。这表明，在包含所有增强策略的情况下，Loss 与标准差之间有较强的负相关性。这个相关性表明，随着训练过程中的损失降低，特征的标准差也相应增大，表明模型的提取特征的能力越来越稳定。

表 3-11 不同 Batch Size 配置下的 BYOL 微调准确率/宏平均召回率

Batch Size	$\beta = 1$	$\beta = 10$	$\beta = 50$	$\beta = 100$	平均值
16	0.8977	0.8977	0.8354	0.7881	0.8547
32	0.8904	0.8942	0.8481	0.8112	0.8610
64	0.9200	0.9050	0.8646	0.8029	0.8731
128	0.9392	0.9529	0.8846	0.8712	0.9120
256	0.9248	0.9246	0.8517	0.8644	0.8914
500	0.9267	0.9213	0.8179	0.7946	0.8651

Pearson 相关系数和 Spearman 相关系数在许多情况下呈现出相似的趋势，这表明数据增强策略对 Loss 和标准差的关系影响较为一致。从结果可以看出，不同的增强策略对 Loss 与标准差的相关性产生了不同的影响。总的来说，删除高斯噪声、相位扰动和缩放等增强策略在影响 Loss 与标准差的关系时，表现出了较强的负相关性，表明这些操作增加了数据的多样性或扰动，使得模型更容易适应各种尺度和噪声变化。而像竖直翻转和水平翻转等操作对 Loss 与标准差的影响较小，表明这些操作对于周期性信号的处理可能并没有显著改变数据的基本特征。因此，在设计数据增强策略时，应该根据具体任务的特性选择合适的增强方法，以最大程度提升模型的鲁棒性和泛化能力。

同时，图3-17显示，SimSiam 在去除 Stop-Grad 模块或 Predictor 模块后，表现出较高的正相关性，尤其是斯皮尔曼相关系数几乎接近 1，这表明准确率与损失/标准差之间呈现较强的等级相关性。这意味着去除这些模块后，损失和标准差的变化对模型性能的影响呈现出更加明显的等级趋势。具体来说，去除 Stop-Grad 模块后，模型的性能与损失、标准差之间的关系变得更加显著，尤其在 SimSiam w/o SG 中，斯皮尔曼相关系数高达 0.9935，接近 1，说明这种关联非常强。

对于去除 Predictor 模块的情况，尽管皮尔森相关系数仍然较高 (0.9735)，但斯皮尔曼相关系数较低 (0.7907)，这表明去除 Predictor 模块后，损失和标准差的变化对准确率的影响在等级上的一致性有所减弱。这个变化可能揭示了模型在去除 Predictor 模块后表现出的一定程度的性能下降，尤其是在捕捉复杂模式时的能力受限。

这个相关性也揭示了模型在去除 Stop-Grad 模块或 Predictor 模块后的潜在模型崩塌现象。去除这些模块后，模型的性能变化更加显著，且损失与准确率的关系变得更加直接，暗示着模型在这些模块缺失的情况下，可能缺乏必要的学习能力和稳定性，导致性能严重下降或过拟合。这种分析进一步表明，Stop-Grad 模块和 Predictor 模块在 SimSiam 模型中的重要性，去除这些模块后可能影响模型的稳定性和泛化能力，进而导致性能的大幅波动和模型崩塌现象。这与前文的实验结

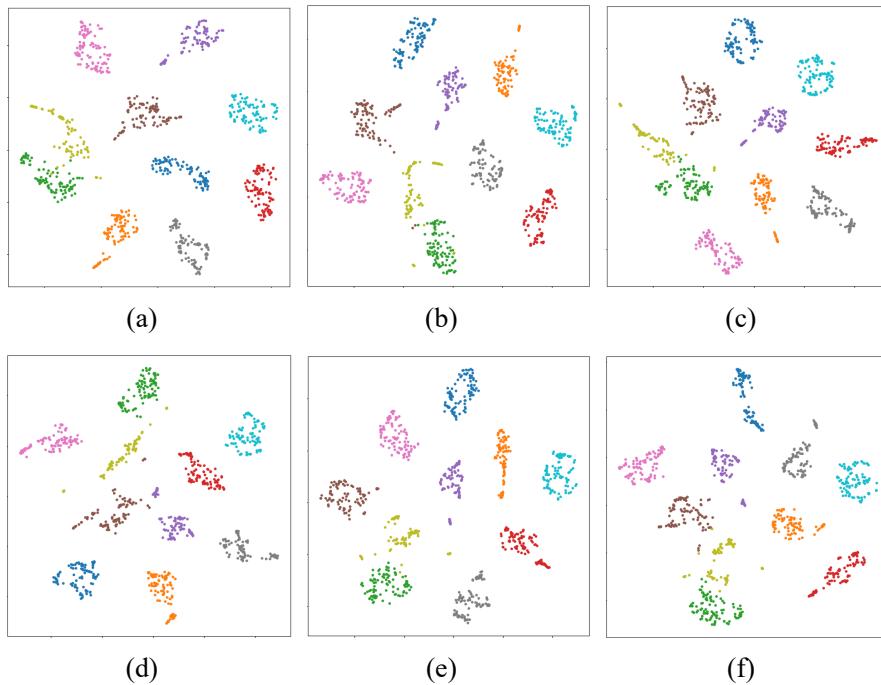


图 3-14 不同 Batch Size 配置下的 SimSiam 网络输出特征的 T-SNE 分布图:

(a) Batch Size 为 16; (b) Batch Size 为 32; (c) Batch Size 为 64; (d) Batch Size 为 128; (e) Batch Size 为 256; (f) Batch Size 为 512

论相一致。

准确率与 Loss/Std: 表3-12展示了准确率与损失函数以及标准差之间的相关性分析。具体分析如下:

首先，准确率与损失函数的相关性分析表明，无论是使用皮尔森相关系数 (-0.0048) 还是斯皮尔曼相关系数 (-0.1557)，两者的相关系数都非常接近零，且对应的 p-value 分别为 0.9910 和 0.7128。这些结果表明，准确率与损失函数之间几乎没有线性或等级上的相关性。因此，尽管损失值反映了模型在训练过程中的误差大小，但它与最终的准确率之间并没有明显的联系。

其次，准确率与标准差之间的相关性也表现出较弱的关系。皮尔森相关系数为 0.1972，斯皮尔曼相关系数为 0.2994，二者均较小，说明准确率与标准差之间的线性或等级相关性较弱。同时，p-value 分别为 0.6398 和 0.4713，这意味着这些相关性并不显著。因此，标准差可能未能有效反映模型预测的波动性与准确率之间的关系。

综合来看，表格中的结果表明，准确率与损失以及标准差之间的相关性较低，无论是线性还是非线性关系均不显著。这可能意味着，尽管损失和标准差是训练过程中非常重要的指标，但它们并不能直接有效地预测最终模型的准确率。在实

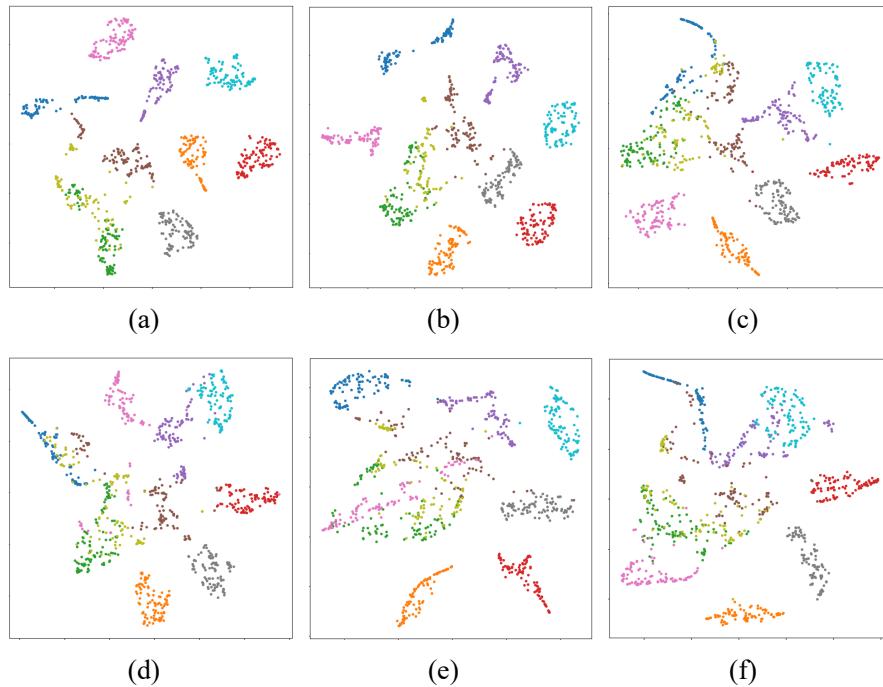


图 3-15 不同 Batch Size 配置下的 SimCLR 网络输出特征的 T-SNE 分布图:

(a) Batch Size 为 16; (b) Batch Size 为 32; (c) Batch Size 为 64; (d) Batch Size 为 128; (e) Batch Size 为 256; (f) Batch Size 为 512

际应用中，可能需要依赖其他更直接的性能指标来评估模型的表现。

表 3-12 SimSiam 准确率与不同指标的相关性分析

指标	相关系数类型	相关系数	p-value
准确率与 Loss	皮尔森相关系数	-0.0048	0.9910
准确率与 Loss	斯皮尔曼相关系数	-0.1557	0.7128
准确率与标准差	皮尔森相关系数	0.1972	0.6398
准确率与标准差	斯皮尔曼相关系数	0.2994	0.4713

3.3 本章小结

本章节提出了一种结合简单暹罗孪生网络 (SimSiam) 与基于 CMA-ES 优化器搜索最优数据增强策略的对比学习自监督预训练模型架构，并在构建的符合帕累托长尾分布的 CWRU 数据集和 PU 数据集上取得了显著的性能提升。通过实验，系统地分析了模型架构中各模块的作用，并从理论上揭示了 Predictor 模块与 Stop-Grad 模块之间的内在联系。与当前 SOTA 的方法对比，进一步验证了所提出模型的有效性。此外，通过消融实验，系统探讨了无标签数据集的引入、Batch Size

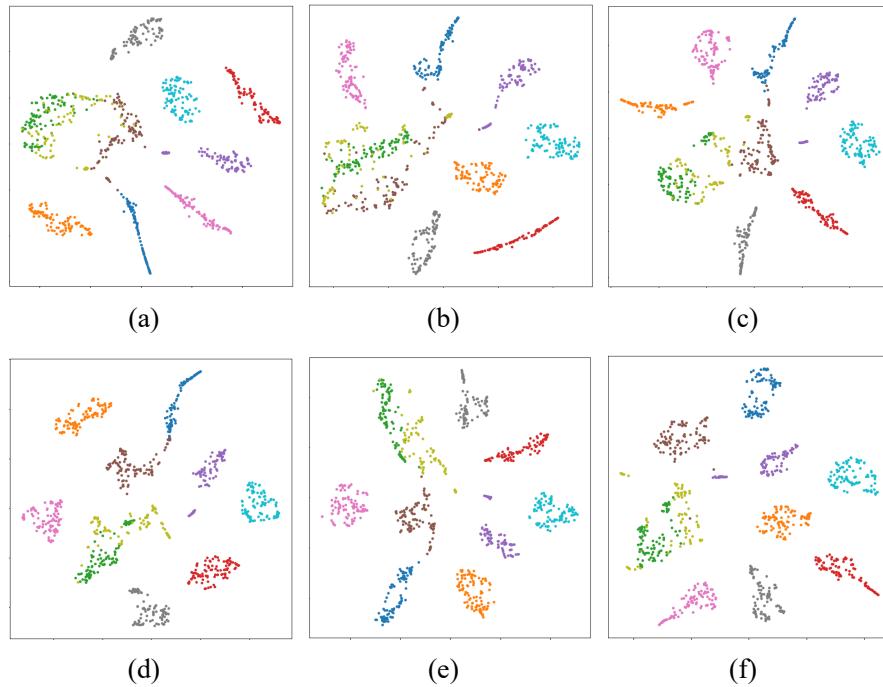


图 3-16 不同 Batch Size 配置下的 BYOL 网络输出特征的 T-SNE 分布图: (a) Batch Size 为 16; (b) Batch Size 为 32; (c) Batch Size 为 64; (d) Batch Size 为 128; (e) Batch Size 为 256; (f) Batch Size 为 512

等关键参数对模型性能的影响, 观察结果与文献 [32] 中关于“较大 Batch Size 有助于对比学习模型性能”的结论存在一定差异, 表明在某些情况下, 过大的 Batch Size 可能对模型训练效果产生负面影响。然而, 研究过程中发现, 模型在微调阶段仍然受到由样本分布不平衡引起的偏差影响。因此, 下一章将重点针对这一问题提出改进方案, 以进一步提升模型的分类性能。

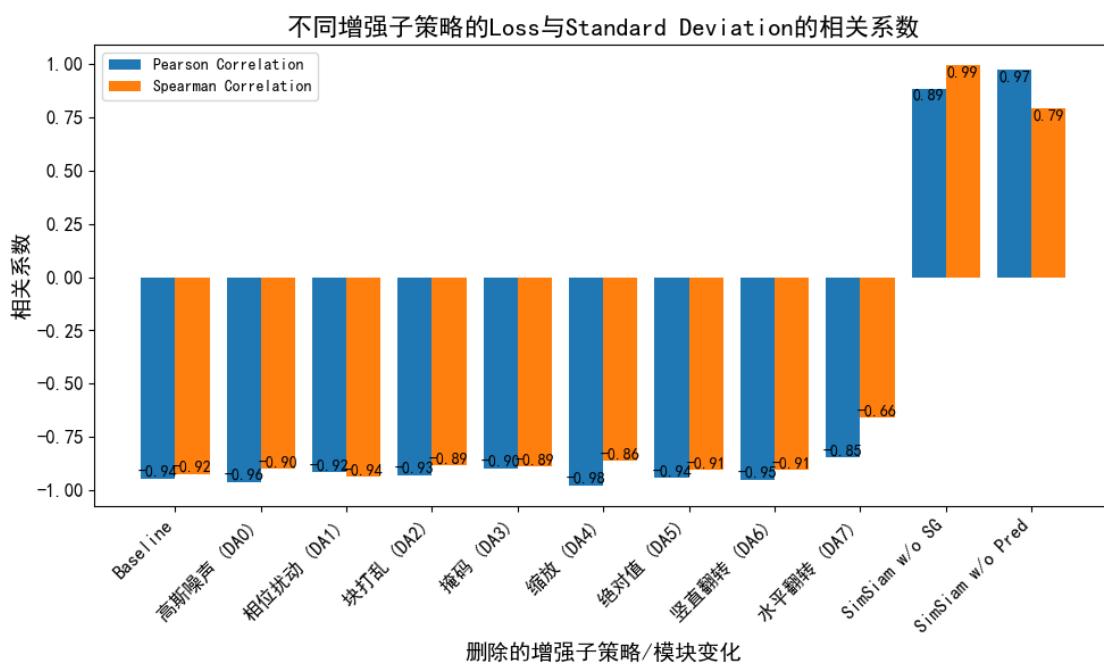


图 3-17 不同增强子策略的 Loss 与 Standard Deviation 的相关系数

第四章 半监督学习与边距调整改进的暹罗孪生故障诊断网络

上一章展示了自监督学习范式下的二阶段式训练方法(如图2-1所示)，其在故障诊断任务的长尾学习场景中显著提升了模型性能。然而，在微调阶段，样本分布的不平衡问题仍对模型的性能产生不利影响，导致分类器的决策边界发生偏移。如图4-1所示，决策边界向尾部类特征空间发生偏移，使头部类在特征空间中占据更大的体积，从而降低了尾部类的分类准确率。本章节将重点探讨一种结合半监督学习与 MARC(Margin Calibration) 决策边界调整的微调算法，以缓解长尾分布对微调阶段决策边界偏移的影响，并进一步提升尾部类的识别能力。在处理长尾分布和样本不平衡问题时，决策面调整方法 (MARC) 提供了一种有效的方式来改善分类器的性能，特别是在少数类别(尾部类别)上。文献 [19] 介绍的决策面调整算法通过引入边距校准方法，调整分类器的决策面，使其更加均衡地处理各个类别的数据。研究发现长尾学习中的边距 (Margin) 和 logits 存在偏差，通过训练 ω 和 β 两个参数校准这种偏差。而该校准算法并没有很好地运用无标签数据集中的信息发挥完整的性能，因此下文结合半监督学习的思想对此进行改进。

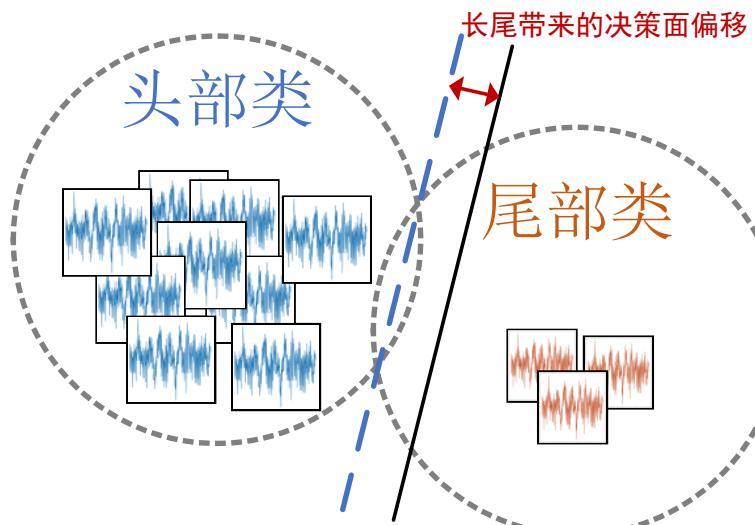


图 4-1 由长尾分布带来的决策面偏移示意图

4.1 模型架构

4.1.1 半监督学习与边距调整改进的暹罗孪生故障诊断网络整体架构

针对原有 MARC 算法不能充分运用无监督样本的信息，本研究在 MARC 算法中引入半监督学习的思想，优化了上一章的 SimSiam 架构的微调过程，提出了一种结合两者的 SimSiam 模型架构，称为基于半监督 MARC 微调的 SimSiam 模型 (Semi-MARC SimSiam)。该模型通过三阶段的逐步训练和优化，旨在解决长尾分布和样本不平衡问题，并显著提升分类器在尾部类别上的性能。其具体模型训练架构流程如图 4-2 所示。该模型架构包括以下四个阶段：

- (1) **阶段一 (自监督学习):** 在第一阶段，基于 SimSiam 架构对模型进行预训练。具体而言，利用大量无标签样本和少量有标签的长尾样本，训练一个性能较优的特征提取网络和初步分类器。SimSiam 的自监督特性能够在数据量受限的情况下充分挖掘数据的潜在分布信息，为后续的半监督学习提供高质量的特征表示。
- (2) **阶段二 (半监督学习):** 在第二阶段，基于阶段一训练得到的 SimSiam 模型，使用其生成无标签样本的伪标签 (Pseudo Labels)，将伪标签视为无标签数据的“真实标签”。随后，将补充有伪标签的无标签数据与原始有标签数据集相结合，通过最小化标记数据和伪标签数据的损失函数，重新优化分类器的全连接层参数，得到半监督微调后的 Semi-SimSiam 模型。这一阶段的目标是榨取无标签数据集的信息，扩充原有数据集。
- (3) **阶段三 (半监督优化的 MARC 决策边界调整):** 在第三阶段，利用经过半监督学习优化后的分类器及其输出的 logits 和经伪标签扩充的数据集，对决策边界进行微调。具体而言，通过训练两个向量参数 ω 和 β ，分别控制分类器输出的尺度因子和偏置项，调整分类器的决策面，以提升模型在尾部类别上的分类准确率。MARC 的引入有效平衡了模型在头部类别和尾部类别上的性能差异。
- (4) **阶段四 (故障诊断):** 最终，将阶段二训练得到的 Semi-SimSiam 模型与阶段三中学习到的尺度因子 ω 和偏置项 β 相结合，构建完整的故障诊断模型 Semi-Marc SimSiam。该模型综合利用自监督特征学习、伪标签生成与决策边界优化三者的优势，对输入样本进行精确分类，从而实现更为稳健且性能均衡的故障诊断。

整体流程通过结合 SimSiam 的特征提取能力、半监督学习对伪标签的利用，以及 MARC 算法对决策边界的微调，形成了一个多阶段优化框架。该方法在处理长尾分布和样本不平衡问题时，展现出显著的优势，尤其是在提高尾部类别分类

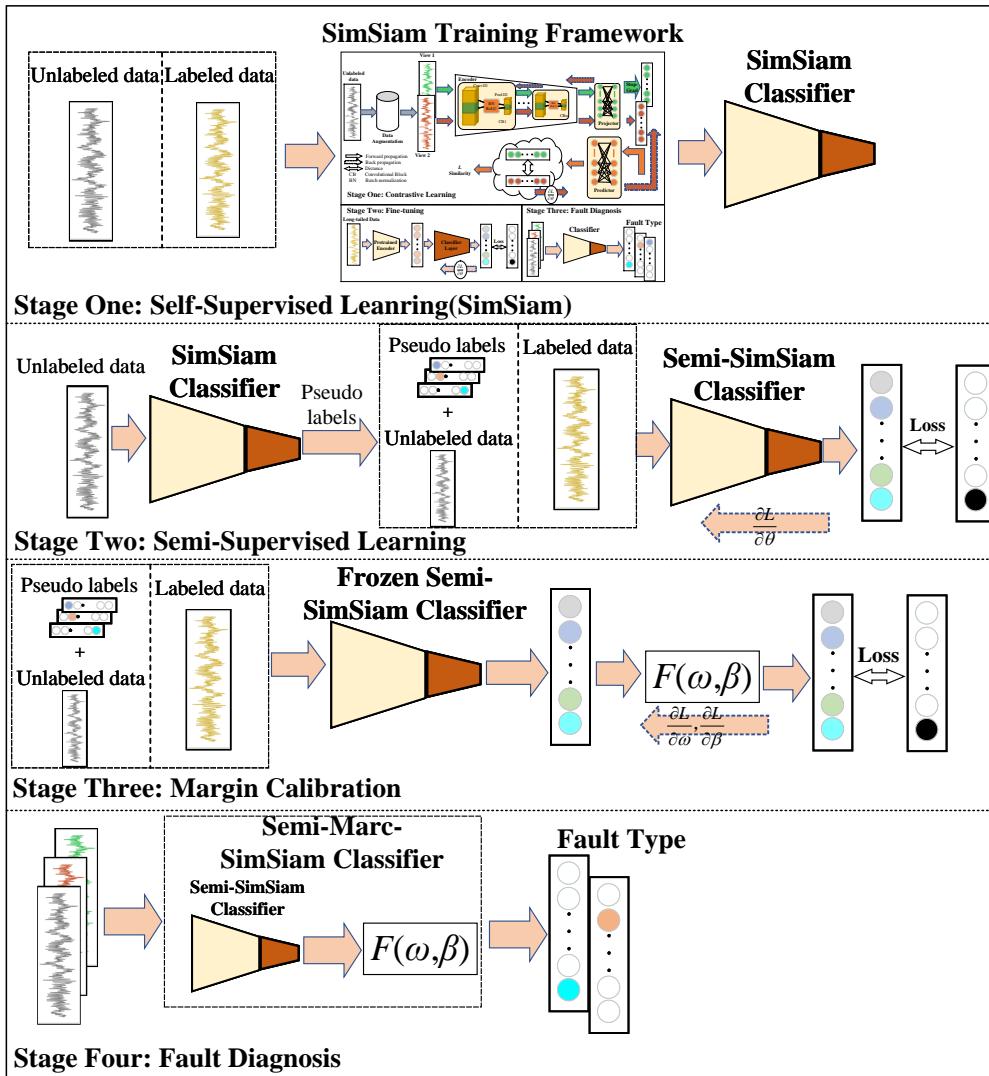


图 4-2 基于半监督学习和 MARC 微调的 SimSiam 模型流程框图

准确率方面效果显著。

4.1.2 半监督学习模型架构

本研究提出了一种结合半监督学习与决策边界调整算法的框架，以有效解决长尾分布和样本不平衡问题。受到文献 [18, 19] 的启发，研究借鉴了半监督学习 (Semi-Supervised Learning, 下称 Semi) 的经典流程 (如图4-3所示)。该流程分为两个阶段：

在第一阶段，使用原始的不平衡标记数据集 \mathcal{D}_L 训练一个初步分类器 f_θ 。随后，在第二阶段，利用该分类器对未标记数据集 \mathcal{D}_U 进行推断并生成伪标签 \hat{y} 。伪标签

数据与标记数据结合，通过最小化以下损失函数优化模型：

$$\mathcal{L}(\mathcal{D}_L, \theta) + \omega \mathcal{L}(\mathcal{D}_U, \theta), \quad (4-1)$$

其中， ω 为未标记数据的损失权重， $\mathcal{L}(\mathcal{D}_L, \theta)$ 和 $\mathcal{L}(\mathcal{D}_U, \theta)$ 分别代表基于标记数据和伪标签数据的损失项。通过整合这两部分损失，最终得到优化后的分类器 $f_{\hat{\theta}}$ ，以更好地建模数据分布，尤其是尾部类别（少数类别）。该过程通过充分利用未标记数据 \mathcal{D}_U ，提升模型对少数类别的分类能力，从而显著优化其决策边界。

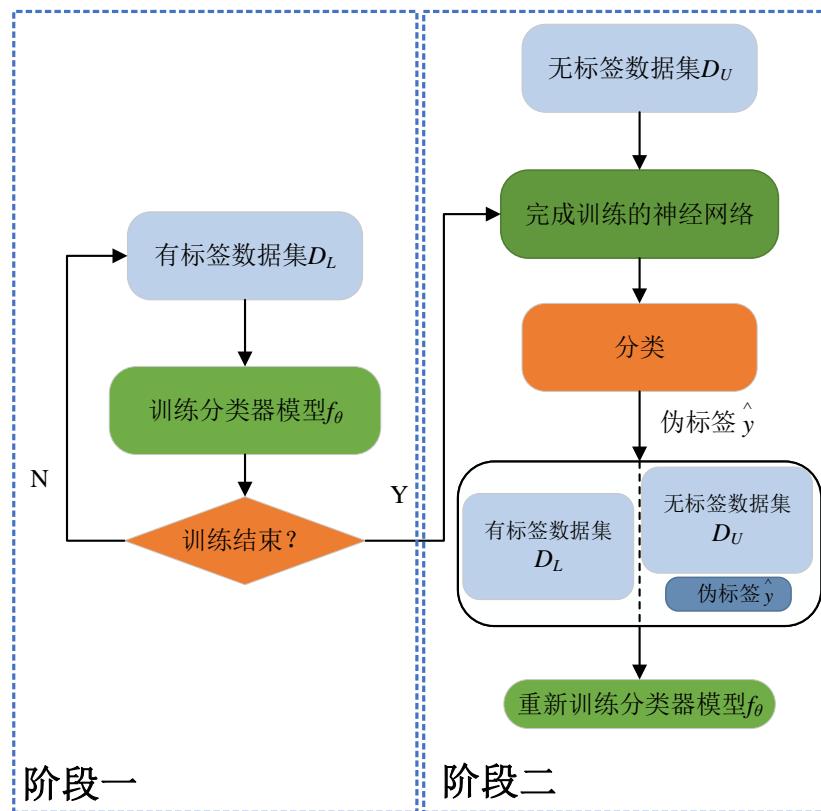


图 4-3 半监督学习流程图

4.1.3 经半监督优化的决策面调整算法架构

与此同时，在处理长尾分布和样本不平衡问题时，决策面调整方法 (MARC) 提供了一种有效的方式来改善分类器的性能，特别是在少数类别（尾部类别）上。文献 [19] 介绍的决策面调整算法通过引入边距校准方法，调整分类器的决策面，使其更加均衡地处理各个类别的数据。

研究发现长尾学习中的边距 (Margin) 和 logits 存在偏差。边距在图4-4中进行了说明。定义类别 j 的仿射超平面 $H_j \in \mathbb{R}^{p-1}$ 为 $W_j z + b_j = 0$ ，即任何落在超平面

H_j 正侧的表示点可以归类为类别 j 。假设 z_0 是一个满足 $W_j z_0 + b_j = 0$ 的点，即 z_0 在超平面 H_j 上。假设 z_1 是特征空间中的任意点。构造从 z_0 指向 z_1 的向量 $z_1 - z_0$ ，并将其投影到法向量 W_j 上。投影向量 $\text{proj}_{W_j}(z_1 - z_0)$ 的长度即是从 z_1 到 H_j 的边距。边距 d_j 的计算如下：

$$\begin{aligned}
 d_j &= \|\text{proj}_{W_j}(z_1 - z_0)\| \\
 &= \left\| \frac{W_j \cdot (z_1 - z_0)}{\|W_j\|} \right\| \\
 &= \frac{W_j \cdot z_1 - W_j \cdot z_0}{\|W_j\|} \\
 &= \frac{W_j z_1 + b_j}{\|W_j\|} \quad (\text{由于 } W_j z_0 + b_j = 0)
 \end{aligned} \tag{4-2}$$

其中 $\|\cdot\|$ 表示 L2 范数。因此， $\text{logit} = W_j \cdot z_1 + b_j$ 也可以表示为 $\|W_j\|d_j$ 。

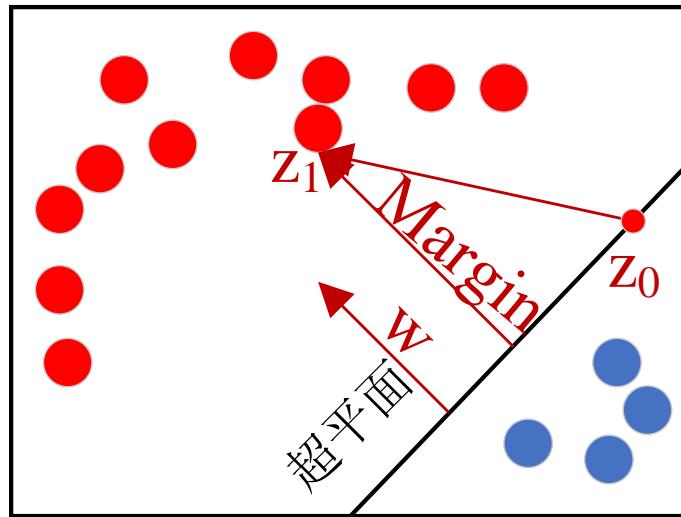


图 4-4 决策面与决策边界示意图，红色点代表头部类，蓝色点代表尾部类

MARC 算法包括两个关键参数： ω 和 β ，其中 ω 是用于调整输出的尺度因子，而 β 则是用于调整输出的偏置项。调整后的边距 \hat{d}_j 如下：

$$\hat{d}_j = \omega_j \cdot d_j + \beta_j \tag{4-3}$$

这两个参数通过最小化类平衡的损失函数（如 CB LOSS）来进行优化，从而调整获

得适合当前任务的最优决策面。校准后的 logits 通过下式计算

$$\begin{aligned}
 \text{logits}_{\text{marc}} &= \|\mathbf{W}_j\| \hat{d}_j = \|\mathbf{W}_j\| (\omega_j \cdot d_j + \beta_j) \\
 &= \omega_j \cdot \|\mathbf{W}_j\| d_j + \beta_j \cdot \|\mathbf{W}_j\| \\
 &= \omega_j \cdot \eta_j + \beta_j \cdot \|\mathbf{W}_j\|,
 \end{aligned} \tag{4-4}$$

其中 η_j 是原始的 logits。算法的具体实现步骤参考表4-1。在 MARC 伪代码中， ω 和 β 是通过 `torch.nn.Parameter` 初始化为可训练的参数，它们分别表示边距的缩放因子和偏置项。使用 `torch.no_grad()` 关键字来避免计算梯度，确保在前向传播时不会影响到梯度计算。接下来，代码通过计算网络的权重范数以及模型在给定输入 x 上的输出结果，来调整输出的决策面，生成校准后的输出 `logit_after`。在实际应用中，MARC 算法有助于减小尾部类别的分类误差，尤其是在样本不平衡的情况下，通过有效地调整决策边距来改善模型的泛化能力和分类精度。

MARC 伪代码 (用 Pytorch 描述)

```

1: 初始化边界校准方法：
   omega = torch.nn.Parameter(torch.ones(1, num_classes))
   beta = torch.nn.Parameter(torch.zeros(1, num_classes))

2: 输入：训练数据 x(由标签数据集D_L和带伪标签的无标签数据集D_U组成)，
   标准的预训练神经网络模型。

3: with torch.no_grad():
4:     w_norm = torch.norm(model.fc.weight, dim=1)
5:     logit_before = model(x)
6:     logit_after = omega * logit_before + beta * w_norm
7: 计算损失并更新参数 omega 和 beta。

```

表 4-1 MARC 算法的实现伪代码

4.1.4 实验设置

模型配置：阶段一 SimSiam 的网络设置与上一章相同。在决策面调整过程中，将 ω 初始化为大小为 $1 \times \text{num_classes}$ 的全 1 向量，而 β 则初始化为相同大小的全 0 向量。

优化器：本研究使用 SGD 优化器训练模型。网络参数通过公式 (3-23) 更新。学习率初始设置为 $0.05 \times \frac{\text{BatchSize}}{256}$ ，学习率采用余弦衰减调度。余弦衰减调度的公式如式 (3-24) 所示。

损失函数：阶段一 SimSiam 对比学习使用的损失函数如上一章所介绍。阶段二、阶段三的损失函数为式 (4-1)，其中 \mathcal{L} 为交叉熵损失函数式 (3-25)， ω 设为 0.9。

数据细节：无标签数据集的规模为类别数 *100，有标签数据集为不同不平衡因子构造帕累托分布的数据，测试集为均匀分布。每个样本包含 1024 个数据点。每个阶段分别训练 150 个周期。在整个训练过程中，每个输入模型的 mini-batch 大小设定为 64。

4.2 实验与分析

4.2.1 在 CWRU 数据集和 PU 数据集上的实验结果

本研究基于 CWRU 数据集和 PU 数据集进行实验，数据集的构造方式与3.2.2节描述一致。每个训练阶段均运行 150 个周期。不同不平衡因子条件下的分类准确率结果如表4-2所示。图4-5展示了在不同 β 值下，采用 Semi-Marc 微调方法后，各类别准确率在三个训练阶段 (SimSiam、Semi、Marc) 中的变化趋势。值得注意的是，Class 0 为头部类别，Class 9 为尾部类别，各类别样本数量呈递减趋势。

表 4-2 与表 4-3 分别展示了不同模型在 CWRU 与 PU 数据集上，针对不同不平衡因子 β 所获得的准确率与宏平均召回率。从整体结果来看，本文提出的 Semi-Marc SimSiam 模型在两个数据集上均取得了最优的平均表现，分别达到 0.9454 和 0.6614，显著优于基础模型 SimSiam 及其两种单独增强版本。这表明，结合半监督机制与 MARC 特征增强策略能够有效缓解样本不平衡问题所带来的性能退化，提升模型的鲁棒性与泛化能力。

在 CWRU 数据集中，随着不平衡因子逐渐增大，所有模型的性能均呈现一定程度下降，其中基础的 SimSiam 模型在 $\beta = 100$ 时准确率降至 0.8450，而 Semi-Marc SimSiam 在相同条件下依然维持在 0.9142，表现出对极端不平衡场景更强的适应能力。与之相对，PU 数据集整体难度更大，所有模型的准确率普遍低于 CWRU，反映出该数据集的复杂性和挑战性。然而，即使在此背景下，Semi-Marc SimSiam 依然在所有 β 设置中保持领先，且波动幅度较小，说明其具有更好的稳定性与迁移能力。

进一步对比可以发现，单独引入半监督策略或 MARC 技术虽能在一定程度上改善模型性能，但其提升幅度相较联合方法较为有限，尤其是在高度不平衡情形下难以维持较高的准确率。这说明两者在机制上具有互补性：半监督策略能够引导模型更充分地利用未标记样本以增强泛化能力，而 MARC 则能通过特征层的增强提升模型对少数类的区分能力，从而在联合使用时实现协同增益，显著提升分类性能。

从图4-2可以清晰地观察到，尾部类别 Class 8 在经过 Semi-Marc 微调后，其分类准确率显著提升。此外，除了 Class 8 和 Class 5 外，其余类别的分类准确率不仅

未下降，反而有所提升。具体变化趋势见图4-5(b)、图4-5(c)和图4-5(d)，均验证了 Semi-Marc 微调方法在提升尾部类分类性能上的有效性，同时对头部类和中部类的分类性能也具有积极影响。

表 4-2 CWRU 数据集 SimSiam 与 Semi-Marc SimSiam 在不同 β 值下的准确率/宏平均召回率

不平衡因子 β	SimSiam	Semi SimSiam	Marc SimSiam	Semi-Marc SimSiam
1	0.9515	0.9685	0.94896	0.9781
10	0.9402	0.9606	0.9575	0.9623
50	0.9131	0.9248	0.91687	0.9271
100	0.8450	0.9090	0.87896	0.9142
平均值	0.9125	0.9407	0.9256	0.9454

表 4-3 PU 数据集 SimSiam 与 Semi-Marc SimSiam 在不同 β 值下的准确率/宏平均召回率

不平衡因子 β	SimSiam	Semi SimSiam	Marc SimSiam	Semi-Marc SimSiam
1	0.6661	0.6801	0.6635	0.6821
10	0.6581	0.6764	0.6696	0.6747
50	0.6392	0.6474	0.6408	0.6499
100	0.5915	0.6363	0.6152	0.6390
平均值	0.6387	0.6601	0.6473	0.6614

Semi-MARC 是一种架构，本身就可以作用于其他任何有无标签样本的网络。因此，以下研究了 Semi-MARC 对 SimCLR, BYOL, CNN 网络的作用。其中，SimCLR 和 BYOL 取性能最高的 Batch Size 训练，即 SimCLR 的 Batch Size 为 16，BYOL 的 Batch Size 为 128。

从图4-6结果可以观察到，四种模型在引入 Semi-Marc 方法之后均表现出不同程度的性能提升，尤其是 SimSiam 和 BYOL 的准确率显著提高，分别从 0.9131 提升至 0.9454，以及从 0.8610 提升至 0.9297。值得注意的是，在未引入 Semi-Marc 时，SimCLR 以 0.9265 的准确率略高于 SimSiam，但在引入 Semi-Marc 后，SimSiam 凭借更显著的性能增益反超 SimCLR，最终以 0.9454 的准确率成为表现最优的模型。这一现象表明 Semi-Marc 对于基于对比学习的结构具有良好的兼容性与提升效果，尤其在 SimSiam 这类不依赖负样本对比的轻量化结构中，其辅助作用更加明显。同时，传统的 CNN 模型也获得了一定的性能改善，从 0.7420 提升至 0.7783，说明 Semi-Marc 并非仅对自监督模型有效，也具备一定的通用性。整体而言，Semi-Marc

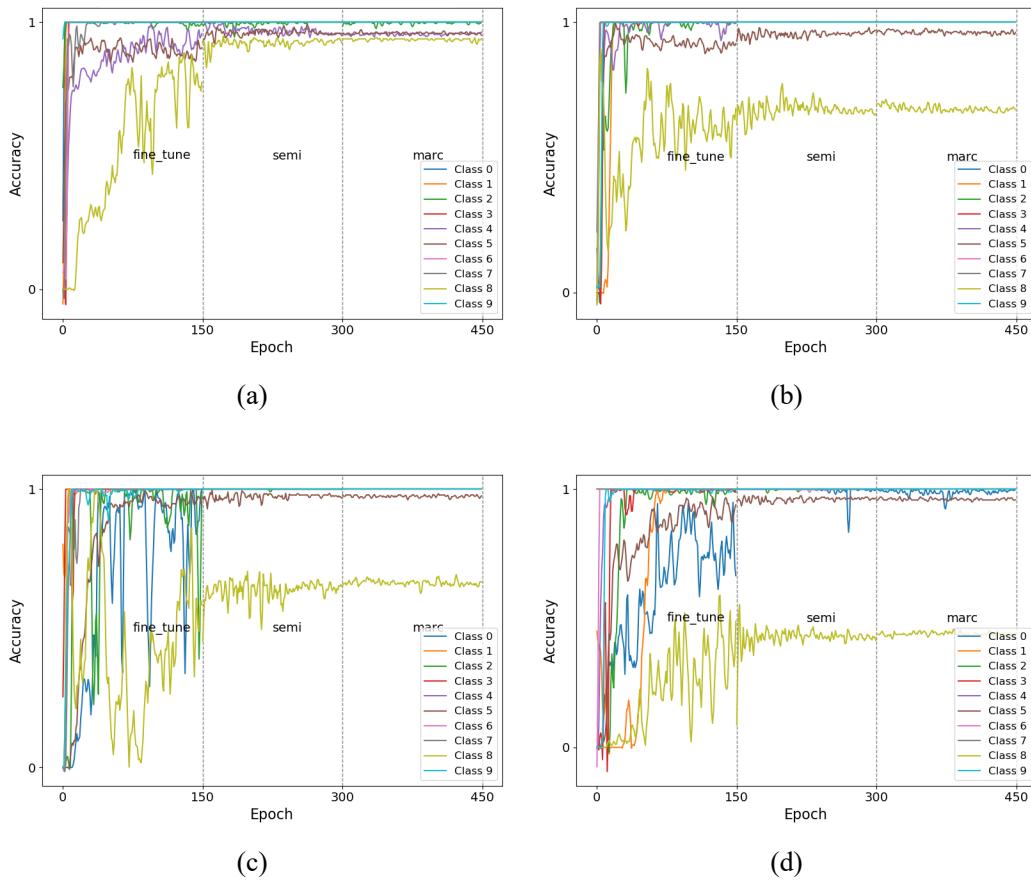


图 4-5 不同 β 值下基于 Semi-Marc 微调的各类别准确率/宏平均召回率的变化 (a) $\beta = 1$; (b) $\beta = 10$; (c) $\beta = 50$; (d) $\beta = 100$

能够有效缓解数据不平衡对模型训练的影响，并提升模型在长尾分布数据集条件下的特征表达能力，其在多个结构中的广泛适配性进一步验证了其作为一类通用增强策略的潜力与价值。

总的来说，Semi-Marc 作为一种有效的模型微调架构，不仅能够提升网络对抗长尾效应的鲁棒性，而且能够对不同网络架构产生明显的性能优化效果，这为自监督学习模型提供了一种有力的优化手段。

4.2.2 消融实验

在本小节，基于 CWRU 数据集对训练过程中的参数进行了消融实验，主要包含各个阶段的“需要梯度”(Require Grad)参数、无标签数据集规模、Batch Size、阶段一的网络模型。

“需要梯度”

编码器参数“需要梯度”(Require Grad)是一个布尔类型的参数，它代表了在

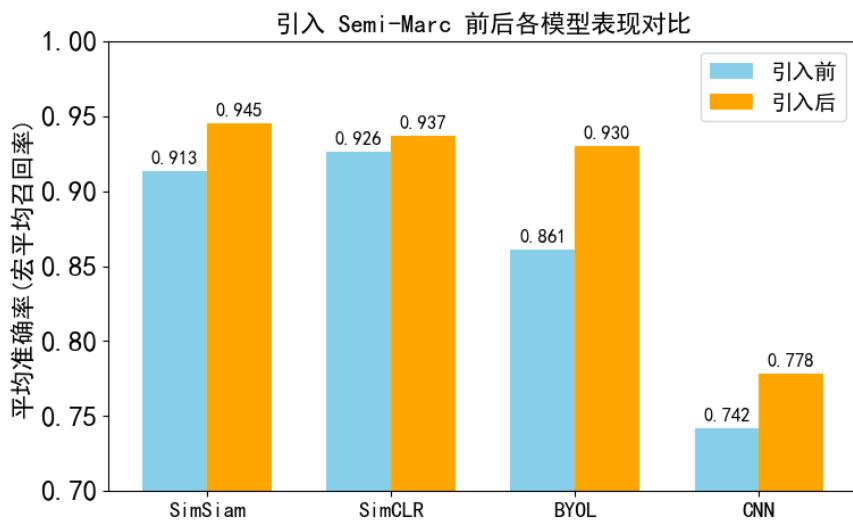


图 4-6 SimSiam, SimCLR, BYOL, CNN 在 Semi-Marc 微调前后的平均准确率/宏平均召回率柱状图

训练中是否对该模块的网络权重进行反向传播更新。在 SimSiam 对比学习阶段和半监督学习阶段都需要对该参数进行设置，表4-4对该参数进行了实验。表中，`fine_tune_requires_grad` 表示在 SimSiam 阶段是否对编码器 (Encoder) 的参数进行微调，`semi_requires_grad` 则表示在半监督学习阶段是否对编码器的参数进行微调。通过对比不同配置下的准确率，可以得出以下结论：

- SimSiam 阶段梯度更新对性能的影响：**当 `fine_tune_requires_grad` 为 True 时，模型在 SimSiam 阶段对编码器参数进行微调。从表中可以看出，无论 `semi_requires_grad` 的配置如何，`fine_tune_requires_grad=True` 时的平均准确率 (0.9396 和 0.9398) 略低于 `fine_tune_requires_grad=False` 时的平均准确率 (0.9409 和 0.9454)。这表明在 SimSiam 阶段冻结编码器参数可能有助于提升模型的整体性能。
- 半监督学习阶段梯度更新对性能的影响：**当 `semi_requires_grad` 为 False 时，模型在半监督学习阶段冻结编码器参数。从表中可以看出，`semi_requires_grad` 为 False 时的平均准确率 (0.9398 和 0.9454) 略高于 `semi_requires_grad` 为 True 时的平均准确率 (0.9396 和 0.9409)。这表明在半监督学习阶段冻结编码器参数可能有助于进一步提升模型的性能。
- 综合配置的影响：**当 `fine_tune_requires_grad=False` 且 `semi_requires_grad=False` 时，模型在两个阶段均冻结编码器参数，此时的平均准确率最高 (0.9454)。这表明在 SimSiam 和半监督学习阶段均冻结编码器参数可能是最优的配置。

方案。

综上所述，通过对不同梯度配置的实验分析，可以得出在 SimSiam 和半监督学习阶段冻结编码器参数有助于提升模型的整体性能。这一发现为后续研究提供了重要的参考依据。

表 4-4 不同配置下 Semi-Marc 方法的准确率/宏平均召回率，
fine_tune_requires_grad 表示是否在 SimSiam 微调时微调编码器，
semi_requires_grad 表示是否在 Semi 时微调编码器

fine_tune_requires_grad	semi_requires_grad	$\beta = 1$	$\beta = 10$	$\beta = 50$	$\beta = 100$	平均值
True	True	0.9727	0.9650	0.9185	0.9021	0.9396
True	False	0.9681	0.9669	0.9229	0.9013	0.9398
False	True	0.9746	0.9744	0.9175	0.8973	0.9409
False	False	0.9781	0.9623	0.9271	0.9142	0.9454

无标签数据集规模

由于无标签数据集的规模直接决定了半监督学习所扩增的带有伪标签的数据集大小，因此无标签数据集规模对 Semi-Marc SimSiam 模型性能的影响至关重要。在以下实验中，固定了训练集的样本规模为 100，并分别设置了无标签数据集的规模为 100、200、500、1000 和 2000。根据图4-7的实验结果，随着无标签数据集规模的增加，Semi-Marc SimSiam 的性能逐步提高。这表明，较大的无标签数据集有助于提升模型在自监督预训练阶段学习到的特征质量。

然而，值得注意的是，当数据规模达到 500 及以上时，模型性能的提升趋于平稳，甚至在数据规模达到 2000 时，性能出现了微弱的下降。这一现象可能与以下几个因素有关：

- (1) **数据冗余性：**随着数据集规模的增大，训练数据的冗余性也随之增加，尤其是当无标签数据集的质量相对较低时，更多的无标签数据并没有提供足够的信息来进一步提升模型的性能。此时，模型可能受到冗余信息的干扰，导致训练效率降低，从而影响了性能提升。
- (2) **数据质量：**在某些情况下，较大的无标签数据集可能包含更多噪声数据，这些噪声数据可能对模型学习到的深层特征产生负面影响。因此，随着数据规模的增加，模型的性能可能不再持续提高，甚至出现性能下降的情况。
- (3) **过拟合风险：**当数据集规模过大时，模型可能过度依赖于训练数据中的某些特征，导致过拟合的风险增加，从而影响了模型在测试集上的泛化能力。

综合来看，虽然增加无标签数据集的规模对模型性能有积极的促进作用，但当数据规模超过一定阈值后，性能提升趋于平稳，甚至可能出现下降。因此，在

实际应用中，选择合适的无标签数据集规模，确保数据的多样性和质量，可能比单纯地增加数据量更为重要。

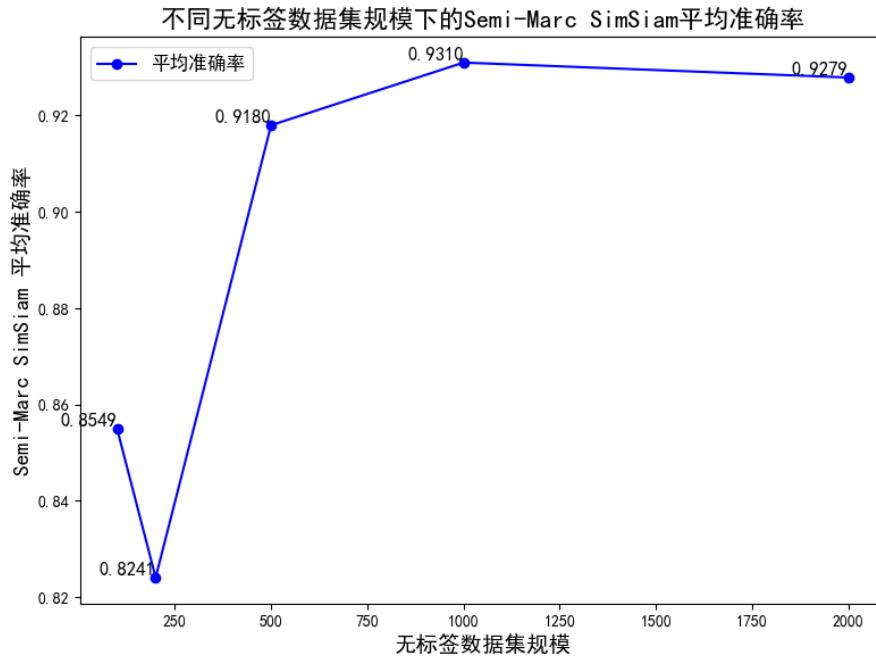


图 4-7 不同无标签数据集规模和不平衡因子 β 配置下的 Semi-Marc SimSiam 平均准确率/宏平均召回率折线图

Batch Size

以下探讨在自监督学习中，Batch Size 对 Semi-Marc 微调网络过程的影响。从图4-8可以看出，Semi-Marc SimSiam 对 Batch Size 的变化表现出较强的稳定性，而 Semi-Marc SimCLR 的性能则呈现出逐渐下降的趋势。具体而言，Semi-Marc SimCLR 在 Batch Size 为 16 时达到了最佳性能，而 BYOL 则在 Batch Size 为 128 时表现最优。因此，本文提出的 Semi-Marc SimSiam 网络在不同 Batch Size 设置下展现了更好的稳定性。相比之下，SimCLR 和 BYOL 在训练过程中对 Batch Size 的选择较为敏感，这可能会增加网络训练人员的工作量。

4.3 本章小结

本章通过结合半监督学习和 MARC 决策面调整的三阶段微调架构 (Semi-Marc)，有效地减轻了长尾分布在有标签微调阶段对 SimSiam 分类层的负面影响，显著提高了 SimSiam 在 CWRU 数据集和 PU 数据集上的性能，超越了 SimCLR。此外，通过消融实验分析了编码器的“需要梯度”(Require Grad) 参数、无标签数据集规模、Batch Size 对模型性能的影响。最后，作为一种通用架构，Semi-Marc

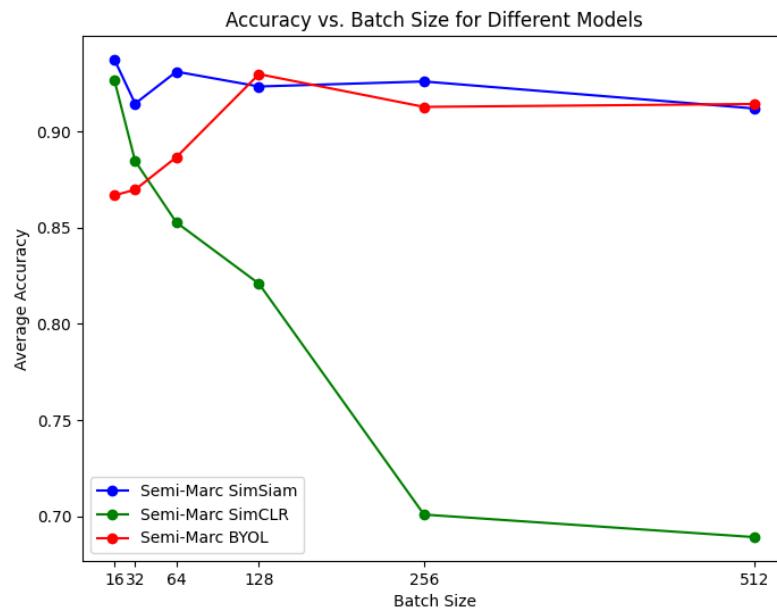


图 4-8 不同 Batch Size 下 SimSiam, SimCLR, BYOL 经过 Semi-Marc 微调后的准确率/宏平均召回率变化图

在其他模型上的应用也证明了其有效性，表明该架构具有良好的普适性，能够在不同模型中提升性能。

第五章 全文总结与展望

5.1 全文总结

随着技术的不断发展，现代工业中的机械设备逐渐变得更加复杂，相应地，机械设备系统的故障种类也在不断增加。这导致在现实中，故障检测数据通常呈现长尾分布。然而，传统的智能故障诊断算法在处理此类长尾分布数据时，容易出现类别偏差，即模型倾向于将样本归类为头部类，而那些样本较少、但更具诊断价值和潜在危害的类别往往被忽视。此外，标注数据的获取通常需要付出高昂的时间和经济成本。

针对这一问题，本文以机械部位轴承为研究对象，首先调研了故障诊断领域及其他领域(如计算机视觉)在解决长尾学习问题时采用的方法和思路，提出了一种基于自监督学习的长尾学习故障诊断框架。接着，本文提出了基于孪生网络对比学习的自监督预训练框架，并结合半监督学习与决策面调整算法的微调框架，针对 CWRU 数据集和 PU 数据集进行了实验与分析。

本文的主要研究内容及结论如下：

- (1) 提出了基于 SimSiam 简单孪生网络为骨干网络的自监督学习故障诊断框架，该框架通过对比学习进行训练。并提出了基于 CMA-ES 搜索最优数据增强策略的框架，为 SimSiam 提供了丰富有效的样本空间，解决了人工设计数据增强方法准确性和多样性不足的问题，从而使 SimSiam 能够学习到更深层次的语义特征。实验表明，相较于 SOTA 方法 SimCLR 和 BYOL，SimSiam 在 CWRU 数据集和 PU 数据集上的准确率较高仅次于 SimCLR，成功减弱了长尾效应。SimSiam 的优势在于无标签数据集的规模，但也存在边际效应。此外，SimSiam 在 Batch Size 变化下仍能保持稳定性能，而 SimCLR 在 Batch Size 增大时性能明显下降，这一现象与文献 [32] 中关于“大 Batch Size 有益于对比学习”的结论存在一定差异。
- (2) 结合优化了半监督学习与 MARC 决策面调整相结合的微调方法 (Semi-Marc)，该方法与 SimSiam 自监督预训练组成了完整的三阶段自监督学习框架，显著减轻了长尾分布对 SimSiam 分类层的负面影响。实验证明，Semi-Marc 方法是一种通用方法，能有效提升其他模型的性能，具有较好的普适性。经过该方法优化后 SimSiam 的性能成功超越了 SimCLR。

5.2 后续工作展望

虽然本文通过实验证明了该模型在轴承故障诊断中的有效性，但仍可进一步探索其适用性与稳定性。未来的研究方向如下：

- (1) 本文提出的模型仅在轴承故障诊断中进行了验证，实际工业系统中涉及更多复杂的对象，如无人机系统等。因此，模型在其他领域中的有效性仍需进一步验证。
- (2) 本文虽然在理论上提出了 SimSiam 数据增强模块的设计方向，但关于 SimSiam 简单孪生网络所需的数据增强方法尚未得出明确结论。并且，CMA-ES 算法在搜索最优数据增强解时可能存在计算效率问题。未来研究可能会尝试将 CMA-ES 算法替换为强化学习算法，以提高效率。
- (3) 本文在微调阶段未对原始数据集进行数据增强，因此，微调阶段的数据增强可能成为下一步的研究重点，例如通过生成对抗网络 (GAN) 进行数据增强。

致 谢

在攻读硕士学位期间，衷心感谢老师、同学们的关心、支持和帮助！

参考文献

- [1] 南京大学,南京智谷人工智能研究院有限公司.一种适用于长尾带噪分布数据的故障检测方法 [P].中国,发明专利,CN202211435489.6,2023-2-3.
- [2] 中国矿业大学.一种面向长尾分布数据的工业过程故障诊断方法 [P].中国,发明专利,CN202310182227.1,2023-6-2.
- [3] Zhang W, Li X, Ma H, et al. Federated learning for machinery fault diagnosis with dynamic validation and self-supervision[J]. Knowledge-Based Systems, 2021, 213: 106679.
- [4] 薛阳,雷文平,岳帅旭.多模态学习方法在滚动轴承故障诊断中的应用 [J].机械科学与技术, 2022, 41(8): 1149-1153.
- [5] 郭文军,张自来,陈丽君. Emd-ar 和 grnn 算法下的航空液压泵多模态故障诊断分析 [J]. 液压与气动, 2022, 46(006): 151.
- [6] 杨波,夏虹,尹文哲. 自适应随机共振在信号特征提取中的应用 [J]. 哈尔滨工程大学学报, 2022, 43(12): 1750-1758.
- [7] 蒋爱国,符培伦,谷明. 基于多模态堆叠自动编码器的感应电机故障诊断 [J]. 电子测量与仪器学报, 2018, 32(8): 17-23.
- [8] 李永刚,侯岳佳. 一种基于 msk-cnn 和多源机电信息融合的同步电机故障诊断方法 [J]. 电机与控制学报, 2023, 27(1): 11.
- [9] 张龙,胡燕青,赵丽娟. 采用递归图编码技术与残差网络的滚动轴承故障诊断 [J]. 西安交通大学学报, 2023, 57(2): 110-120.
- [10] Shen X, Marks E. Near-miss information visualization tool in bim for construction safety[J]. Journal of construction engineering and management, 2016, 142(4): 04015100.
- [11] Wilson D L. Asymptotic properties of nearest neighbor rules using edited data[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1972, SMC-2(3): 408-421.
- [12] Chawla N V, Bowyer K W, Hall L O, et al. Smote: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357.
- [13] Kingma D P. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.
- [14] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

- [15] Huang C, Li Y, Loy C C, et al. Learning deep representation for imbalanced classification[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 5375-5384.
- [16] Lin T. Focal loss for dense object detection[J]. arXiv preprint arXiv:1708.02002, 2017.
- [17] Cui Y, Jia M, Lin T-Y, et al. Class-balanced loss based on effective number of samples[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 9268-9277.
- [18] Yang Y, Xu Z. Rethinking the value of labels for improving class-imbalanced learning[J]. Advances in neural information processing systems, 2020, 33: 19290-19301.
- [19] Wang Y, Zhang B, Hou W, et al. Margin calibration for long-tailed visual recognition[C]. Asian Conference on Machine Learning, 2023: 1101-1116.
- [20] Liu Z, Miao Z, Zhan X, et al. Large-scale long-tailed recognition in an open world[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 2537-2546.
- [21] Yin X, Yu X, Sohn K, et al. Feature transfer learning for face recognition with under-represented data[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 5704-5713.
- [22] You C, Li C, Robinson D P, et al. Scalable exemplar-based subspace clustering on class-imbalanced data[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 67-83.
- [23] Zhang X, Fang Z, Wen Y, et al. Range loss for deep face recognition with long-tailed training data[C]. Proceedings of the IEEE international conference on computer vision, 2017: 5409-5418.
- [24] Jamal M A, Brown M, Yang M-H, et al. Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 7610-7619.
- [25] Shu J, Xie Q, Yi L, et al. Meta-weight-net: Learning an explicit mapping for sample weighting[C]. Advances in Neural Information Processing Systems, 2019: .
- [26] Zhou B, Cui Q, Wei X-S, et al. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 9719-9728.
- [27] Kang B, Xie S, Rohrbach M, et al. Decoupling representation and classifier for long-tailed recognition[J]. arXiv preprint arXiv:1910.09217, 2019.

- [28] Liu J, Sun Y, Han C, et al. Deep representation learning on long-tailed data: A learnable embedding augmentation perspective[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 2970-2979.
- [29] 吴磊, 韩华, 黄丽. 面向长尾图像的个性化专家识别算法 [J]. 电光与控制, 2023, 30(11): 62-66.
- [30] 吴亮, 梁振, 张燚鑫. 基于多级学习的长尾分布下交通多目标检测 [J]. 计算机与数字工程, 2023, 51(3): 599-605.
- [31] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: .
- [32] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]. International conference on machine learning, 2020: 1597-1607.
- [33] Grill J-B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning[J]. Advances in neural information processing systems, 2020, 33: 21271-21284.
- [34] Caron M, Misra I, Mairal J, et al. Unsupervised learning of visual features by contrasting cluster assignments[J]. Advances in neural information processing systems, 2020, 33: 9912-9924.
- [35] He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022: 16000-16009.
- [36] Doersch C, Gupta A, Efros A A. Unsupervised visual representation learning by context prediction[C]. Proceedings of the IEEE international conference on computer vision, 2015: 1422-1430.
- [37] Zhang R, Isola P, Efros A A. Colorful image colorization[C]. Computer Vision – ECCV 2016, Cham, 2016: 649-666.
- [38] Gidaris S, Singh P, Komodakis N. Unsupervised representation learning by predicting image rotations[J]. arXiv preprint arXiv:1803.07728, 2018.
- [39] Noroozi M, Pirsiavash H, Favaro P. Representation learning by learning to count[C]. Proceedings of the IEEE international conference on computer vision, 2017: 5898-5906.
- [40] Caron M, Bojanowski P, Joulin A, et al. Deep clustering for unsupervised learning of visual features[C]. Proceedings of the European conference on computer vision (ECCV), 2018: 132-149.
- [41] Zhang T, Chen J, He S, et al. Prior knowledge-augmented self-supervised feature learning for few-shot intelligent fault diagnosis of machines[J]. IEEE Transactions on Industrial Electronics, 2022, 69(10): 10573-10584.

- [42] Senanayaka J S L, Khang H V, Robbersmyr K G. Toward self-supervised feature learning for online diagnosis of multiple faults in electric powertrains[J]. IEEE Transactions on Industrial Informatics, 2020, 17(6): 3772-3781.
- [43] Min S, Dai Q, Xie H, et al. Cross-modal attention consistency for video-audio unsupervised learning[J]. arXiv preprint arXiv:2106.06939, 2021.
- [44] Oord Avan den, Li Y, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint arXiv:1807.03748, 2018.
- [45] 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, 2018: 3733-3742.
- [46] Yuan D, Chang X, Huang P-Y, et al. Self-supervised deep correlation tracking[J]. IEEE Transactions on Image Processing, 2020, 30: 976-985.
- [47] Laine S, Aila T. Temporal ensembling for semi-supervised learning[J]. arXiv preprint arXiv:1610.02242, 2016.
- [48] Zheng Z, Yang Y. Unsupervised scene adaptation with memory regularization in vivo[J]. arXiv preprint arXiv:1912.11164, 2019.
- [49] Zhao H, Yang G, Wang D, et al. Deep mutual learning for visual object tracking[J]. Pattern Recognition, 2021, 112: 107796.
- [50] Chen X, He K. Exploring simple siamese representation learning[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021: 15750-15758.
- [51] Loparo K. Case western reserve university bearing data center[J]. Bearings Vibration Data Sets, Case Western Reserve University, 2012, 22-28.
- [52] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.
- [53] Ioffe S. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.
- [54] You Y, Gitman I, Ginsburg B. Large batch training of convolutional networks[J]. arXiv preprint arXiv:1708.03888, 2017.
- [55] Ziani R, Felkaoui A, Zegadi R. Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized fisher' s criterion[J]. Journal of Intelligent Manufacturing, 2017, 28: 405-417.

- [56] Eren L, Ince T, Kiranyaz S. A generic intelligent bearing fault diagnosis system using compact adaptive 1d cnn classifier[J]. Journal of signal processing systems, 2019, 91(2): 179-189.

攻读硕士学位期间取得的成果

电子科技大学
UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE



论文题目 基于领域泛化的轴承
故障诊断方法研究

专业学位类别 电子信息
学 号 _____
作者姓名 _____
指导教师 _____
学 院 电子科技大学（深圳）高等研究院

分类号 V279 密级 公开
UDC^{注1} 681.5

学位论文

基于领域泛化的轴承故障诊断方法研究

指导教师

电子科技大学 成都

电子科技大学 成都

申请学位级别 硕士 专业学位类别 电子信息

提交论文日期 论文答辩日期

学位授予单位和日期 电子科技大学 2025年6月

答辩委员会主席

评阅人

注1：注明《国际十进分类法 UDC》的类号。

Research on Bearing Fault Diagnosis Method Based on Domain Generalization

A Master Thesis Submitted to
University of Electronic Science and Technology of China

Discipline Electronic Information

Student ID _____

Author _____

Supervisor _____

School Shenzhen Institute for Advanced Study, UESTC

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

作者签名：_____ 日期： 年 月 日

论文使用授权

本学位论文作者完全了解电子科技大学有关保留、使用学位论文的规定，同意学校有权保留并向国家有关部门或机构送交论文的复印件和数字文档，允许论文被查阅。本人授权电子科技大学可以将学位论文的全部或部分内容编入有关数据库进行检索及下载，可以采用影印、扫描等复制手段保存、汇编学位论文。

(涉密的学位论文须按照国家及学校相关规定管理，在解密后适用于本授权。)

作者签名：_____ 导师签名：_____

日期： 年 月 日

摘要

轴承作为工业设备的核心传动部件，其故障诊断对保障生产安全具有重要工程价值。传统方法在跨域泛化场景下面临单源域数据稀缺、跨工况分布差异较大及多源域过拟合特定源域的三重问题。本研究从领域泛化理论出发解决上述问题，主要创新如下：

(1) 针对变工况轴承故障诊断中单源域数据稀疏性与工况负载不平衡导致的领域泛化难题，构建了融合多模态特征与对抗生成网络的跨域数据增强框架 MDWGANGP(Multi-Domain WGAN-GP)，旨在通过结合一维时序信号和二维图像特征，并且该方法通过结合梯度惩罚的 WGAN 网络生成质量较高的训练数据。实验结果显示，MDWGANGP 平均准确率达到 85.78%，比最优基线方法提升 3.6 个百分点，在准确性和可靠性上优于单一信息域特征提取，有效解决源域数据不足问题，并为后续诊断提供可靠的训练数据集。

(2) 针对不同工况下分布差异较大导致域偏移和泛化鲁棒性不足的问题，提出了多损失协同优化的跨域轴承故障诊断方法 HAF-MDWGANGP (Hierarchical Alignment Framework for Multi-Domain WGAN-GP)。该方法在交叉熵损失的基础上结合最大均值差异损失、三元组损失和裕度差异损失。通过交叉熵损失保持分类判别性，最大均值差异损失对齐全局分布，三元组损失约束类内紧凑性，裕度差异损失增强决策边界可分性。实验证明，该方法在复杂工况下的平均准确率达 89.24%，较 MDWGANGP 模型提升 3.46 个百分点，证明多损失协同优化策略对域不变特征提取的有效性。

(3) 针对现有方法易过拟合多源域中的特定源域导致对不可见目标域预测准确率差的问题，提出了基于自适应多源集成的跨域轴承故障诊断方法 AMSEDG(Adaptive Multi-Source Ensemble Domain Generalization)。区别于传统的单一策略方法，AMSEDG 通过多层次不变特征学习和基于动态相似性的加权集成策略，解决多源域异构条件下旋转机械故障诊断的领域泛化问题。在多源域情况下，AMSEDG 以 95.1% 的平均准确率超越 8 种对比方法，AMSEDG 相较于表现最佳的 HAF-MDWGANGP 高出 2.2 个百分点，验证了其在多源协同泛化方面的优势。

关键词：轴承故障诊断，领域泛化，最大均值差异损失，三元组损失，裕度差异损失

ABSTRACT

Bearing as the core transmission component of industrial equipment, its fault diagnosis has important engineering value to ensure production safety. Traditional methods face the triple problems of scarcity of data in single source domain, large differences in distribution across industrial conditions and overfitting of specific source domains in multiple source domains in cross-domain generalization scenarios. This study addresses the above problems from the domain generalization theory, and the main innovations are as follows:

(1) Aiming at the difficult problem of domain generalization caused by the sparsity of single-source domain data and workload imbalance in fault diagnosis of bearings with variable operating conditions, a cross-domain data enhancement framework, MDWGANGP (Multi-Domain WGAN-GP), which fuses multimodal features and adversarial generative networks, was constructed, which aims at combining the one-dimensional time-series signals and the two-dimensional image features, and which generates higher-quality training data by combining the gradient-penalty of the WGAN network to generate higher quality training data. The experimental results show that the average accuracy of MDWGANGP reaches 85.78%, which is 3.6 percentage points higher than the optimal baseline method, and it outperforms the single-information-domain feature extraction in terms of accuracy and reliability, effectively solves the problem of insufficient source-domain data, and provides a reliable training dataset for subsequent diagnosis.

(2) Aiming at the problem of insufficient robustness of domain bias and generalization due to large differences in distributions under different operating conditions, a multi-loss co-optimization cross-domain bearing fault diagnosis method HAF-MDWGANGP (Hierarchical Alignment Framework for Multi-Domain WGAN-GP) is proposed. The method combines maximum mean difference loss, triad loss and margin difference loss on the basis of cross-entropy loss. The cross-entropy loss maintains the categorical discriminative property, the maximum mean difference loss aligns the global distribution, the triad loss constrains the intra-class compactness, and the margin difference loss enhances the decision boundary divisibility. Experimental validation shows that the average accuracy of this method under complex working conditions

ABSTRACT

reaches 89.24%, which is 3.46 percentage points higher than that of the MDWGANGP model, proving the effectiveness of the multi-loss collaborative optimization strategy for domain-invariant feature extraction.

(3) Aiming at the problem that existing methods are prone to overfitting specific source domains in multi-source domains, which leads to poor prediction accuracy for invisible target domains, AMSEDG (Adaptive Multi-Source Ensemble Domain Generalization) is proposed as a cross-domain bearing fault diagnosis method based on adaptive multi-source integration. Distinguished from the traditional single-strategy approach, AMSEDG solves the domain generalization problem of rotating machinery fault diagnosis under the heterogeneous conditions of multi-source domains through multi-level invariant feature learning and a weighted integration strategy based on dynamic similarity. In the multi-source domain case, AMSEDG outperforms the eight compared methods with an average accuracy of 95.1%, and AMSEDG is 2.2 percentage points higher than the best-performing HAF-MDWGANGP, which verifies its advantages in multi-source collaborative generalization.

Keywords: Bearing Fault Diagnosis, Domain Generalization, Maximum Mean Difference Loss, Triad Loss, Margin Difference Loss

目 录

第一章 绪论	1
1.1 研究背景与意义	1
1.2 国内外相关研究工作	3
1.2.1 基于数据操作的故障诊断算法现状研究	3
1.2.2 基于表征学习的故障诊断算法现状研究	4
1.2.3 基于学习策略的故障诊断算法现状研究	6
1.3 本文的章节安排	6
第二章 跨工况轴承故障诊断知识背景	9
2.1 领域泛化相关的概念	9
2.2 故障增强相关理论	10
2.2.1 格拉姆角场(GAF)	10
2.2.2 马尔可夫迁移场(MTF)	11
2.2.3 递归图(RP)	13
2.2.4 短时傅立叶变换(STFT)	14
2.3 集成学习相关理论	16
2.3.1 集成学习概述	16
2.3.2 Bagging 原理	16
2.3.3 Boosting 原理	17
2.3.4 Stacking 原理	19
2.4 轴承故障诊断概述	20
2.4.1 轴承的基本构成	20
2.4.2 轴承典型故障形式	20
2.4.3 轴承故障诊断基本流程	21
2.5 实验平台介绍	23
2.6 本章小结	25
第三章 基于多模态数据增强的跨域轴承故障诊断方法	26
3.1 引言	26
3.2 MDWGANGP 模型框架	26
3.2.1 数据预处理模块	28
3.2.2 特征提取模块	29

3.2.3 生成对抗模块	32
3.2.4 故障分类模块	33
3.3 轴承故障诊断泛化实验验证	35
3.3.1 数据集样本说明	35
3.3.2 对比实验	36
3.3.3 消融实验	39
3.4 本章小结	40
第四章 基于多损失协同优化的跨域轴承故障诊断方法	42
4.1 引言	42
4.2 HAF-MDWGANGP 模型框架	42
4.2.1 MMD 损失	43
4.2.2 三元组损失	44
4.2.3 裕度差异损失	46
4.2.4 综合损失	47
4.2.5 训练流程	47
4.3 实验结果与分析	48
4.3.1 对比实验	48
4.3.2 消融实验	49
4.3.3 可视化实验	50
4.3.4 超参数实验	52
4.4 本章小结	53
第五章 基于自适应多源集成的跨域轴承故障诊断方法	54
5.1 引言	54
5.2 AMSEDG 模型框架	54
5.2.1 多层次不变特征学习	56
5.2.2 基于动态相似性的加权集成	57
5.3 实验结果与分析	58
5.3.1 对比实验	58
5.3.2 消融实验	62
5.4 本章小结	63
第六章 总结与展望	65
6.1 全文工作总结	65
6.2 下一步工作展望	65

目录

致 谢	67
参考文献	68
攻读硕士学位期间取得的成果	73

图目录

图 1-1 轴承在各个领域的应用	1
图 1-2 论文结构	7
图 2-1 领域自适应与领域泛化的对比图	9
图 2-2 四种故障形式的 GASF 和 GADF 效果图	11
图 2-3 四种故障形式的 MTF 效果图	13
图 2-4 四种故障形式的 RP 效果图	14
图 2-5 四种故障形式的 STFT 效果图	15
图 2-6 Bagging 流程图	16
图 2-7 Boosting 流程图	18
图 2-8 滚动轴承典型结构	20
图 2-9 轴承故障诊断基本流程图	21
图 2-10 动力传动故障诊断与仿真实验平台	23
图 2-11 传感器位置	24
图 3-1 MDWGANGP 模型框架	27
图 3-2 BiLSTM 结构	30
图 3-3 CNN 结构	30
图 3-4 生成对抗模块的生成器	33
图 3-5 生成对抗模块的鉴别器	33
图 3-6 MLP 分类器网络结构	34
图 3-7 预测值与真实值的混淆矩阵	36
图 3-8 各领域泛化的方法在实验中的平均准确率	38
图 3-9 不同方法在 A→D 任务中的迭代曲线	39
图 4-1 HAF-MDWGANGP 模型框架	43
图 4-2 原始数据的 t-SNE 可视化结果	51
图 4-3 C2,C3,C4,C5 处理后的可视化结果	51
图 5-1 训练流程	55
图 5-2 测试流程	55
图 5-3 特征提取器的网络结构	55
图 5-4 分类器的网络结构	56
图 5-5 各方法在不同迁移任务实验中的准确率	60
图 5-6 AMSEDG 在四个迁移任务中的混淆矩阵	61

图 5-7 不同方法在 T1 任务中的迭代曲线 62

表目录

表 2-1 动力传动故障诊断实验平台参数	23
表 3-1 数据集工况概述	35
表 3-2 不同跨域迁移任务下目标域诊断准确率 (%)	38
表 3-3 消融实验结果.....	40
表 4-1 对比实验结果.....	49
表 4-2 消融实验结果.....	50
表 4-3 各权重组合的结果	52
表 5-1 迁移任务设计.....	58
表 5-2 各方法在不同迁移任务实验中诊断准确率 (%)	59
表 5-3 AMSEDG 消融实验结果	63

第一章 绪论

1.1 研究背景与意义

在智能制造技术和工业 4.0 的深化发展背景下, 轴承作为工业旋转机械的核心传动部件, 其运行可靠性直接影响生产系统的稳定性。轴承是减少摩擦、使元件平稳运转的关键摩擦部件, 如图 1-1 所示, 轴承广泛应用于各个领域的装备机械上, 如水力发电的大型水电机组、工业生产的制造机器人、风力发电中的风发电机等。旋转机械的正常运行机械设备高度依赖于轴承的健康状态。与其他机械元件相比, 轴承由于长期处于高温、高转速、高载荷等恶劣的工作条件下, 故障会很容易出现。据统计, 导致机械设备停机的主要原因之一是轴承故障, 约占所有机械故障的 30%-40%^[1]。



图 1-1 轴承在各个领域的应用

轴承的稳定运行对机械设备的安全性具有重要意义。然而, 在工业生产中, 轴承通常在关键设备中长期承受高转速、重载及腐蚀性环境等苛刻条件, 这些因素可能对轴承的运行状态造成显著影响, 从而增加故障发生的风险。一旦滚动轴承发生故障, 不仅会造成生产停滞、效率下降, 还可能带来经济损失。因此, 及时识别并解决轴承故障对机械设备, 尤其是大型精密设备的正常运行至关重要。通过开展轴承故障诊断相关研究, 可以有效降低机械设备事故的发生概率^[2]。轴承故障诊断中常用的输入信号包括振动、声音、温度、转速和电流等多种类型, 这些信号通常通过传感器和监测系统进行采集, 并利用数据分析工具进行处理, 以实现故障类型的预测和及时的维护干预。随着人工智能和传感器技术的高速发展, 轴承故障诊断领域发生了深刻变革。传感器性能的提升使得高精度的振动、温度和声音等多维数据能够被实时获取, 为故障诊断提供了充足的信息支持。根据已有研究, 当前智能诊断技术主要可分为三大类别: 经典机器学习方法、深度学习方法以及迁移学习方法。

传统经典机器学习方法普遍依赖于学习样本与验证样本具备一致的统计特性, 然而在动态工业环境下的实际操作中, 这一前提往往难以完全满足, 尤其是在跨场

景诊断任务中传统的基于机器学习的机械设备故障诊断包括信号获取、特征提取以及故障识别寿命预测三个步骤^[3]。在信号采集阶段，通过利用振动、电流、温度及声发射等多种传感器，获取能够表征轴承运行状态的原始数据。在特征提取阶段，传统技术通常基于专家经验设计的预设公式，对数据特性进行分析并人工筛选特征。然而，这种手动提取方式容易混入与故障诊断无关或多余的信息，进而影响分类模型的预测效果，甚至带来不利影响。

为了解决经典机器学习方法的缺点，基于深度学习的智能诊断方法应运而生，旨在直接构建故障信号与故障类别之间的映射关系。相较于传统技术，该方法无需依赖预定义的知识，能够自主从输入数据及其运行状态中提取关键故障特征。近年来，该方法在复杂工况下的应用逐渐显现出显著效果，其主要优势在于能够高效挖掘数据中的故障规律，并展现出处理大规模数据的能力，因而成为旋转机械故障诊断领域的重点研究方向。然而，基于深度学习的诊断技术仍面临两大关键技术挑战：数据分布差异和数据可获取性。首先，该技术通常期望训练数据和测试数据的分布特性保持相似，但在实际工业场景中，由于负载、转速等运行条件的多样性和复杂性，测试数据的分布往往与训练数据存在较大偏差，从而可能导致诊断精度的显著下降。另外，模型的效果很大程度上取决于充足且优质的标注数据集，而在故障工况下运行设备以获取数据，不仅可能加速设备的磨损，还可能带来安全隐患和经济成本。因此，若不能有效解决不同运行场景下的数据采集与标注问题，该技术的整体效果将难以得到保证。

近些年，研究人员开始积极探索迁移学习，这项技术通过使用源域的标注数据来训练模型，并将其迁移到目标域的预测工作中，特别适合解决变工况和数据标注不足的场景。在轴承故障诊断中应用迁移学习，可以有效降低维护费用，减少设备停机时间，同时提高设备使用效率和经济收益。然而，常规的迁移学习方法虽然能够缓解数据分布不一致的问题，但其性能受限于领域间的相似性^[4]。

为了解决上述缺点，领域泛化(Domain Generalization, DG)研究开始受到关注，其旨在通过从多个源领域学习，提升模型在未见过的目标领域上的表现^[5]。相比迁移学习中的领域自适应(Domain Adapatation, DA)^[6]技术，领域泛化无需利用目标域的数据进行训练，而是通过挖掘跨域共享特征或构建适应性更强的模型架构，实现不同领域间的性能稳定表现。将这一技术应用于轴承故障诊断中，有助于缓解因运行条件差异引发的模型性能波动，从而增强诊断系统的适应稳定性和应用灵活性。同时，该技术能够大幅降低对海量标记数据的依赖，减少数据采集和标注过程中的资源投入，为工业设备的智能化运维提供了全新路径。在风力发电、轨道交通以及航空航天等领域的实际场景中，轴承运行环境通常复杂多变，传统诊断方法往往难

以满足需求。通过引入领域泛化技术，可以大幅增强这些场景下的故障识别性能，确保设备运行的安全性并延长其使用寿命^[7]。

总的来说，基于领域泛化技术的轴承故障诊断研究不仅在理论上具有重要的学术意义，还在实际应用中表现出显著的工程价值。该技术为故障诊断领域的发展提供了新动力，同时为工业设备的智能维护开辟了创新路径，有助于实现工业 4.0 的愿景。

1.2 国内外相关研究工作

领域泛化（Domain Generalization, DG）的关键难点在于其训练阶段只能依赖已有源域中的数据集，无法获取目标域的任何数据，无论这些数据是否带有标签。而相比之下，领域自适应（Domain Adaptation, DA）则具备利用源域和目标域数据的优势。领域泛化研究主要可归纳为三大类别：数据操作、表征学习和学习策略^[8-11]。

1.2.1 基于数据操作的故障诊断算法现状研究

数据操作指的是在源域训练样本有限的情况下，通过数据增强和变换等方法扩充并平衡新数据，从而提升模型的领域泛化能力。数据操作主要通过数据增强（Data Augmentation）和数据生成（Data Generation）等方式来扩充和平衡数据集，主要面对图像样本，以减少数据分布的偏差。

数据增强指的是主要基于输入数据的增强、随机化和转换，方法包括领域随机化（Domain Randomization）和对抗数据增强（Adversarial Data Augmentation）。领域随机化通过调整对象的纹理、位置、数量及形状，或向数据引入各类随机噪声，来实现对现有样本的扩展。Tobin 等人^[12]率先采用该技术，从仿真环境中生成更多训练样本，以提升在真实场景中的泛化能力。Peng 等人^[13]提出应同时增强特征和标签，而非仅限于特征本身。对抗数据增强则旨在提升数据的多样性，同时确保其可靠性。Shankar 等人^[14]通过贝叶斯网络构建标签、领域及输入实例间的依赖模型，并开发了 CrossGrad 这一独特策略，通过沿最大域变化方向扰动输入，同时尽量减少对类别标签的干扰。Zhou 等人^[15]则引入了弱增强与强增强相结合的正则化方法，替代直接通过梯度上升更新输入，而是对抗性地训练一个变换网络来实现数据扩充。

数据生成指的是通过生成不同的样本以帮助泛化，依靠梯度或者形变网络来增广样本，可以使用一些生成模型来实现，例如变分自编码器（Variational Autoencoder, VAE）^[16]和生成对抗网络（Generative Adversarial Networks, GAN）^[17]。此外，它也可以使用 Mixup^[18]策略来实现，Mixup 是一种通过对特征和标签对进行

凸组合来创建新样本的简便方法。具体而言，它利用从 Beta 分布中抽取的权重，在任意两个实例及其对应标签间进行线性插值，以生成新的数据，这一过程无需训练生成模型。此外，SMOTE(Synthetic Minority Over-sampling Technique)^[19]算法通过对少数类样本插值生成合成数据，以实现样本均衡。这种方法能够显著提升模型的训练效果、泛化性能以及分类准确性。

在故障诊断中 Zhang 等人^[20]利用条件扩散模型生成了真实而多样的故障信号，这些信号扩展到源域的数据分布之外扩散模型显着增强了故障诊断系统的泛化能力。Zhao 等人^[21]提出了一种用于不平衡域泛化故障诊断的语义判别增强驱动网络。设计了一种基于语义正则化的混合策略来合成足够可靠的样本来补偿少数类。Liu 等人^[22]提出了一种新的数据增强结构 VAEGAN-DRA，在 GAN 结构中引入编码器以学习原始数据的深层特征，并将编码的潜在变量馈传输到生成器以增强生成样本的逼真度，即利用 VAE 对原始数据的潜在特征的学习能力，同时克服了随机噪声采样导致生成过程不稳定、生成数据质量低的问题。Guo 等人^[23]提出辅助分类器生成对抗网络 AC-GAN，除了噪声外，每个生成的样本还拥有相应的类标签，通过嵌入随机噪声标签，可以提取有效的故障数据。

虽然现有的基于数据操作的领域泛化方法取得一定成就，但传统的数据增强方法多依赖于单一信息域（如时域或频域），难以充分挖掘数据中的多层次特征，导致增强效果有限。为此，本文在第三章提出了一种多模态的数据增强跨域轴承故障诊断方法 MDWGANGP，通过融合一维时序信号与二维图像特征生成优质训练样本，以提高模型在变工况下轴承故障诊断的泛化能力。

1.2.2 基于表征学习的故障诊断算法现状研究

表征学习指的是通过模型学习领域不变的特征，以提高特征的可迁移性，从而在领域泛化上展现较高的泛化性。表征学习主要包括域不变表示学习(Domain Invariant Representation Learning)和特征解纠缠(Feature Disentanglement)。

在域不变表示学习方面，Ben-David 等人^[24]通过理论分析表明，若特征在不同域中保持一致，则该表征具有普适性，并可应用于多种领域。同样，在领域泛化研究中，其目标是通过减少多个源域在特定特征空间中的表示差异，实现域不变性，从而使训练得到的模型具备对未知域的适应能力。基于这一思路，主要有四类方法：(1) 方法一是基于核的方法(Kernel-based Method)，主要利用核函数减小数据在特征空间中的分布差异，比如 Hu 等人^[25]提出了多域判别分析对领域泛化进行分类核学习；(2) 方法二是领域对抗学习(Domain Adversarial Learning)，最著名的是 Ganin 和 Lempitsky^[26]以及 Ganin 等人^[27]提出了邻域对抗神经网络(Domain Adversarial

Neural Network, DANN), 该网络通过对抗性方式训练生成器和鉴别器。鉴别器通过训练来辨别不同域之间的差异, 而生成器则通过训练试图欺骗鉴别器, 从而提取域不变的特征表示。Gong 等人^[28]通过逐步缩小流形空间中的域间差异, 运用对抗性训练来实现这一目标; (3) 方法三是显式特征对齐(Explicit Feature Alignment), 旨在扩大类间差异, 同时缩小不同域之间的类内差异。例如, Zhou 等人^[29]通过最小化 Wasserstein 距离, 利用最优传输技术对齐多个源域的边缘分布, 从而构建域不变的特征空间。方法四是不变风险最小化Invariant Risk Minimization), 它对域不变性提供了新的视角。方法四为基于不变性风险优化的技术Invariant Risk Minimization), 其通过针对多源故障诊断任务的域不变性分析, 带来了创新性的视角解读, 特别是在针对跨域性能优化的探索中展现出独特价值。相较于传统观点, 即在同一特征空间中对齐不同域的数据分布以获得域不变特征, 不变风险最小化提出了不同的解释。根据 IRM 理论, 域不变性要求分类器在所有领域中均能达到最优表现, 从而将问题转化为一个带有约束条件的优化任务。

特征解纠缠, 指的是将学习到的特征进行解耦, 分解成领域相关和领域不相关的两种特征。(1) 方法一为多成分分析(Multi-component Analysis), 其通常使用域共享和域特定的网络参数提取域共享和域特定的特征, 例如, Zunino 等人^[30]通过人工对比不同域中特定区域的注意力热图, 以提取解纠缠的表征; (2) 方法二为生成建模方法(Generative Modeling), 从数据生成的视角出发, 我们可以利用生成模型来解决纠缠问题。该方法着眼于从领域、样本与标签三个层面构建样本的生成机制, 例如, Nam 等人^[31]提出了一种基于生成模型的方法, 用以解纠风格与其他信息。该策略不仅适用于领域自适应问题, 同时也能应用于领域泛化任务。此外, 生成模型的使用不仅能够提升模型性能, 还可扩展到生成任务中; (3) 方法三为因果关系启发方法(Causality-inspired Methods)。该模型预测不会被其他不相关的干扰所影响。对于领域泛化, He 等人^[32]重新加权输入样本, 使加权相关性反映因果效应。Zhang 等人^[33]将傅里叶特征视为图像的关键因素, 并确保这些特征间保持相互独立。

在故障诊断领域, Shi 等人^[34]提出了一种融合动态加权机制与批量谱惩罚正则化项的对抗训练方法。该方法能够从存在潜在差异的多个源域中提取兼具域不变性和判别性的特征, 从而在域泛化故障诊断中应对类别偏移难题, 并通过利用更多源域数据, 获取更全面的诊断信息。Zheng 等人^[35]引入用于故障诊断的深度域泛化网络(DDGFD), 从振动信号中获取具有跨域统一性的 DDGFD 输入, 进而使 DDGFD 充分展现其核心优势, 在源域中提取兼具区分度和域无关性的故障特征, 并将习得的知识应用于未见目标样本的识别, 在轴承数据集的跨域场景中展现更优异的表现。

虽然现有的基于表征学习的领域泛化方法已经取得了一些成功，但它们未充分考虑模型在领域泛化过程中对领域偏移的鲁棒性，从而导致泛化性能不理想。为此本文在第四章提出基于多损失协同优化的跨域轴承故障诊断方法 HAF-MDWGANGP，通过在 MDWGANGP 的交叉熵损失的基础上引入多损失函数（MMD 损失、三元组损失、裕度差异损失），构建“数据增强-特征对齐-决策优化”三级协同机制，显著提升模型对复杂领域偏移的鲁棒性。

1.2.3 基于学习策略的故障诊断算法现状研究

学习策略指的是利用集成学习和元学习等成熟的学习模式算法来提取多个源域的信息，以此提高模型的鲁棒性和泛化性。学习策略包括基于集成学习的领域泛化(Ensemble Learning-based DG)^{[36][37]}，基于元学习的领域泛化(Meta-learning-based DG)^[38]，基于梯度操作的领域泛化(Gradient Operation-based DG)^{[39][40]}，基于分布鲁棒优化的领域泛化(Distributionally Robust Optimization-based DG)^{[41][42]}，基于自监督学习的领域泛化(Self-supervised Learning-based DG)^[43]。在故障诊断中，Ren 等人^[44]提出了一种与模型无关的基于元学习的训练框架 Meta-GENE，可用于同构和异构域泛化，并将梯度对齐算法引入元学习框架中以进行鲁棒预测的域不变策略。

虽然现有的基于学习策略的领域泛化方法也已经取得了一些成果，但目前的基于学习策略的领域泛化方法都主要关注领域级特征学习，而忽略了更细微的样本级特征学习，这导致对域不变特征的挖掘不够，尤其是在存在多个源域的情况下。此外，这些方法在诊断目标数据时没有考虑每个目标样本与来自不同源域的数据之间的相似性，导致泛化性能不理想。为此本文在第五章提出基于自适应多源集成跨域轴承故障诊断方法 AMSEDG，并通过实验结果表明，所提出的方法在多源域情况下领域泛化性能上优于现有方法。

1.3 本文的章节安排

本文围绕领域泛化技术在轴承故障诊断中的应用展开，共分为六章，论文结构如图 1-2 所示：

第一章为综述，阐述轴承故障诊断的研究背景与意义，分析传统方法局限性，轴承故障诊断概述综述领域泛化技术的国内外研究现状，明确本文章节安排。

第二章为跨工况轴承故障诊断知识背景，介绍领域泛化的理论基础，故障增强相关理论基础，集成学习的相关理论基础，并且介绍轴承故障诊断内容，除此之外

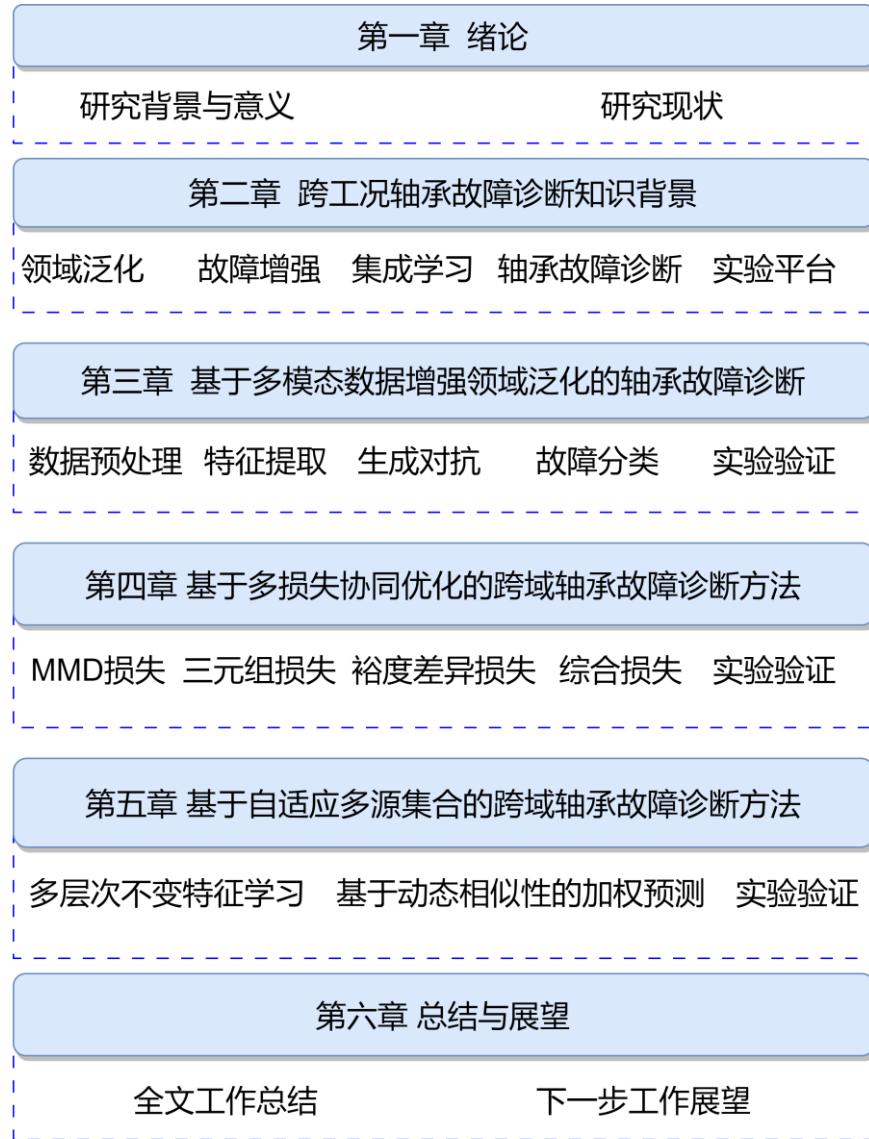


图 1-2 论文结构

还详细说明实验平台架构、数据采集流程等方法，为后续实验提供理论支撑与数据基础。

第三章为基于多模态数据增强领域泛化的轴承故障诊断，提出 MDWGANGP 框架，通过 SVD-NEU 特征重构、BiLSTM-CNN 多模态融合与 WGAN-GP 对抗生成，解决源域数据不足问题，设计 12 组跨域任务验证方法有效性。

第四章为基于多损失协同优化的跨域轴承故障诊断方法，构建 HAF-MDWGANGP 框架，引入 MMD 损失、三元组损失与裕度差异损失的多损失协同优化机制，通过消融实验与特征可视化分析，揭示多损失协同作用机理。

第五章为基于自适应多源集成的跨域轴承故障诊断方法，构建 AMSEDG 框架，通过多层次特征对齐与动态相似性的加权预测，展示在多源域任务中的泛化优势。

第六章为总结与展望，总结全文研究成果，分析现有方法的局限性，从实时性优化、跨设备泛化、开放集诊断等方向展望未来工作。

第二章 跨工况轴承故障诊断知识背景

2.1 领域泛化相关的概念

领域泛化(Domain Generalization, DG) 是其中一种迁移学习方法，旨在训练模型在未见过的目标领域上具有良好的泛化性能，要求模型具备较强的适应能力，能够处理新领域中的数据，而无需再次进行大规模训练，在实际工业场景比领域自适应(Domain Adaptation)更加实用。如图 2-1 所示，与领域自适应不同，领域泛化在训练过程中无法访问目标领域的数据，而是通过利用多个源领域的数据来学习通用的特征表示，从而提高模型在未知领域上的表现。

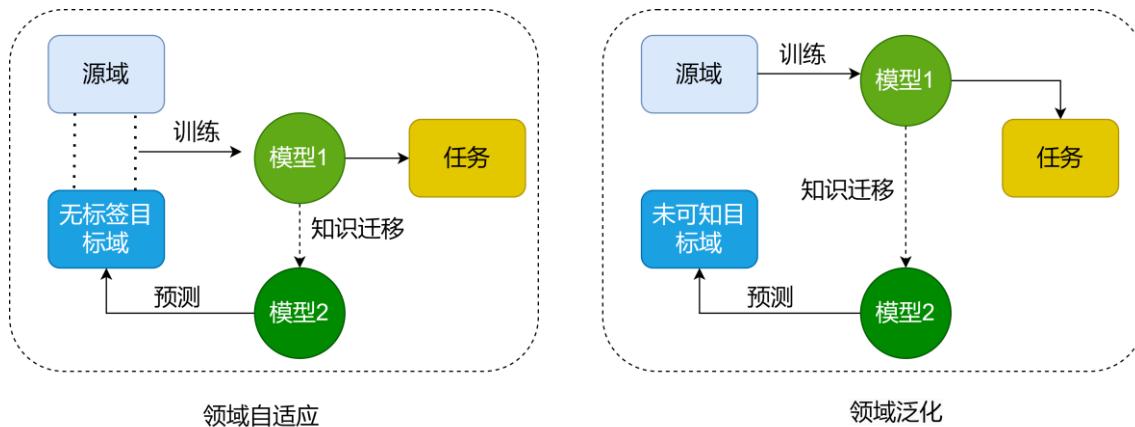


图 2-1 领域自适应与领域泛化的对比图

领域泛化的目标是学习一个模型 f ，使其在多个源领域 $\{D_1, D_2, \dots, D_N\}$ 上训练后，能够在未见过的目标领域 D_t 上表现良好。假设有 N 个源领域 $\{D_1, D_2, \dots, D_N\}$ ，每个领域 D_n 包含输入空间 X_n 和输出空间 Y_n 。领域泛化的目标是学习一个模型 $f: X \rightarrow Y$ ，使得在未见过的目标域 D_t 上的风险最小化。领域泛化的目标是最小化模型在未可见的目标领域上的期望风险：

$$\min E_f(x, y) \sim D_t[L(f(x), y)] \quad (2.1)$$

其中： L 是损失函数（如交叉熵损失或均方误差）， D_t 是目标域的分布。

2.2 故障增强相关理论

2.2.1 格拉姆角场(GAF)

格拉姆角场 (Gramian Angular Field, GAF) 是一种创新的方法，通过将一维时间序列转换为二维图像，清晰展示时间序列的特征，为后续分析与研究提供便利。这种方法的核心在于利用时间序列数据点之间的角度关系来构建图像。格拉姆角和场(GASF)、格拉姆角差场(GADF)方法都隶属于格拉姆角场(GAF)，其区别在于最终生成图像步骤时所用的正余弦函数以及相位参量的求和求差关系。上述方法是由 Wang 和 Tim 等^[45]人在 2015 年提出的。假设有一段时间序列 $X = \{x_1, x_2, x_3, x_n\}$ ，其中 x_i 是标量， n 为序列长度。具体步骤如下：

(1) 最小-最大归一化(Min-Max Normalization)

首先，将原始时间序列 $X = \{x_1, x_2, x_3, x_n\}$ 数据归一化到特定的区间(如[0, 1]或[-1, 1])，得到缩放后的时间序列 \tilde{X} ，以便于后续的角度计算。

$$\tilde{x}_i = \frac{(x_i - \min(X)) + (\max(X) - x_i)}{\max(X) - \min(X)} \quad (2.2)$$

其中，在公式(2.1)中， $\min(X)$ 和 $\max(X)$ 分别是序列 X 的最小值和最大值。

(2) 极坐标转换

将归一化后的时间序列数据转换为极坐标形式。每个数据点的值被映射为极径（半径），而其在时间序列中的位置则被映射为极角。

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in X \\ r_i = \frac{t_i}{n}, & t_i = 1, 2, \dots, n \end{cases} \quad (2.3)$$

其中，对时间序列 X 进行极坐标转换，每个 \tilde{x}_i 值对应一个以弧度为单位的角余弦 ϕ_i ，而每个时间步 t_i 就关联一个极径 r_i ，其中 n 作为一个常数因子，能够调节极坐标系的规模与形态。

(3) 计算 GASF 和 GADF

GASF 是通过计算每两个点之间的角度和的余弦值来构建的，GADF 是通过计算每两个点之间的角度差的余弦值来构建的，公式如下：

$$\begin{aligned} \text{GASF} &= [\cos(\phi_i + \phi_j)] = \tilde{X}^T \cdot \tilde{X} - \sqrt{I - \tilde{X}^2} \cdot \sqrt{I - \tilde{X}^2} \\ \text{GADF} &= [\sin(\phi_i - \phi_j)] = \sqrt{I - \tilde{X}^2} \cdot \tilde{X} - \tilde{X}^T \cdot \sqrt{I - \tilde{X}^2} \end{aligned} \quad (2.4)$$

其中， I 为单位行向量 $[1, 1, 1, \dots, 1]$ 。

如图 2-2 所示，将 GASF 和 GADF 的数值直接映射为伪彩色图像，即实现一维时间序列信号到图像的变换。

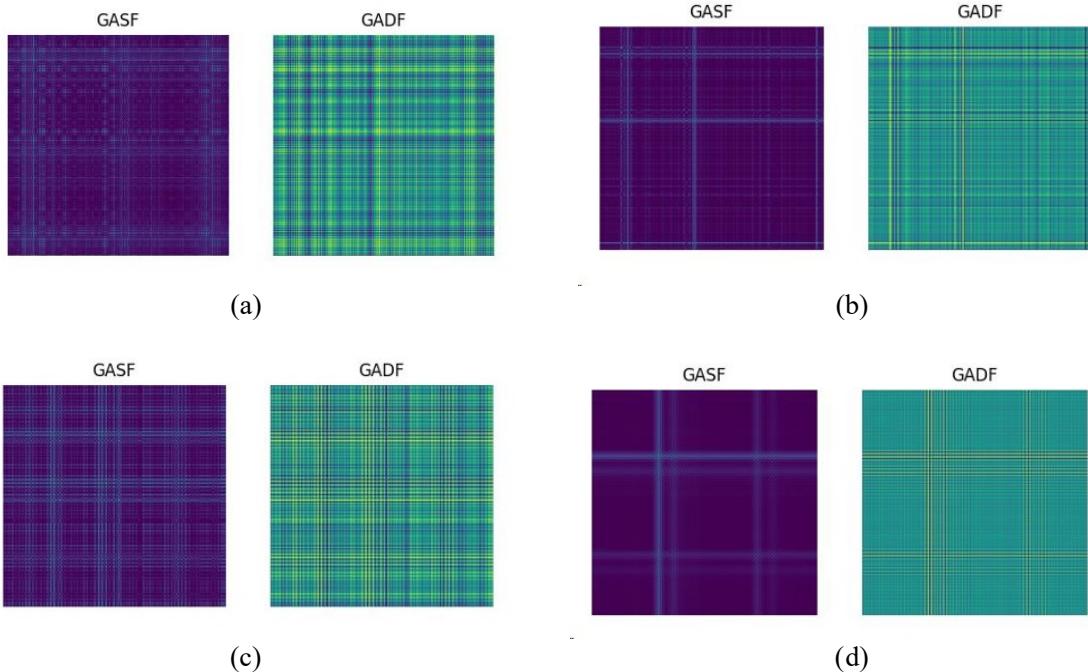


图 2-2 四种故障形式的 GASF 和 GADF 效果图。(a)正常; (b)内圈故障;
(c)滚动体故障; (d)外圈故障

2.2.2 马尔可夫迁移场(MTF)

马尔可夫转移场（Markov Transition Field, MTF）是一种将时间序列数据映射为图像表达的技术。该方法通过分析时间序列中状态转移的概率分布，将其动态特性投影到二维图像空间。MTF 的核心理念在于运用马尔可夫链模型来表征时间序列中状态间的转移模式，并将这些模式以图像像素值的形式呈现。通过构建马尔可夫转移矩阵，MTF 能够有效反映时间序列中的状态转移规律，从而捕捉其动态特征，同时也能保留时间依赖性。MTF 步骤如下：

(1) 时间序列离散化

首先，将原始时间序列 $X = \{x_1, x_2, x_3, \dots, x_n\}$ 离散化为有限的状态。离散化的方法可以是等宽分箱（将值域均匀划分为多个区间）或等频分箱（使每个区间包含相同数量的数据点）。离散化后的时间序列记为 $Q = \{q_1, q_2, \dots, q_n\}$ ，其中每个 q_i 表示 x_i 所属的状态。

(2) 构建马尔可夫转移矩阵

马尔可夫转移矩阵 W 描述了时间序列中状态之间的转移概率。转移概率可以通过统计时间序列中状态转移的频率来估计。假设离散化后的状态共有 m 个，

则 W 是一个 $m \times m$ 的矩阵，其中每个元素 W_{ij} 表示从状态 i 转移到状态 j 的概率：

$$W_{ij} = P(q_t = j | q_{t-1} = i) \quad (2.5)$$

(3) 构建马尔可夫转移场(MTF)

MTF 是一个 $n \times n$ 的矩阵，其中每个元素 MTF_{ij} 表示在时间点 i 和时间点 j 之间的状态转移概率。对于时间点 i 和 j ，找到它们对应的离散状态 q_i 和 q_j ；然后，从马尔可夫转移矩阵 W 中提取对应的转移概率 W_{q_i, q_j} ；之后将 W_{q_i, q_j} 作为 MTF 矩阵的元素 MTF_{ij} 。

MTF 矩阵的定义为：

$$MTF_{ij} = W_{q_i, q_j} \quad (2.6)$$

其中， q_i 和 q_j 分别是时间点 i 和 j 对应的离散状态。

(4) 图像表示

如图 2-3 所示，将 MTF 矩阵可视化为图像，其中每个像素的值对应于矩阵中的转移概率。通过这种方式，时间序列的动态特征（如状态转移模式和周期性）被转换为图像的纹理和模式。

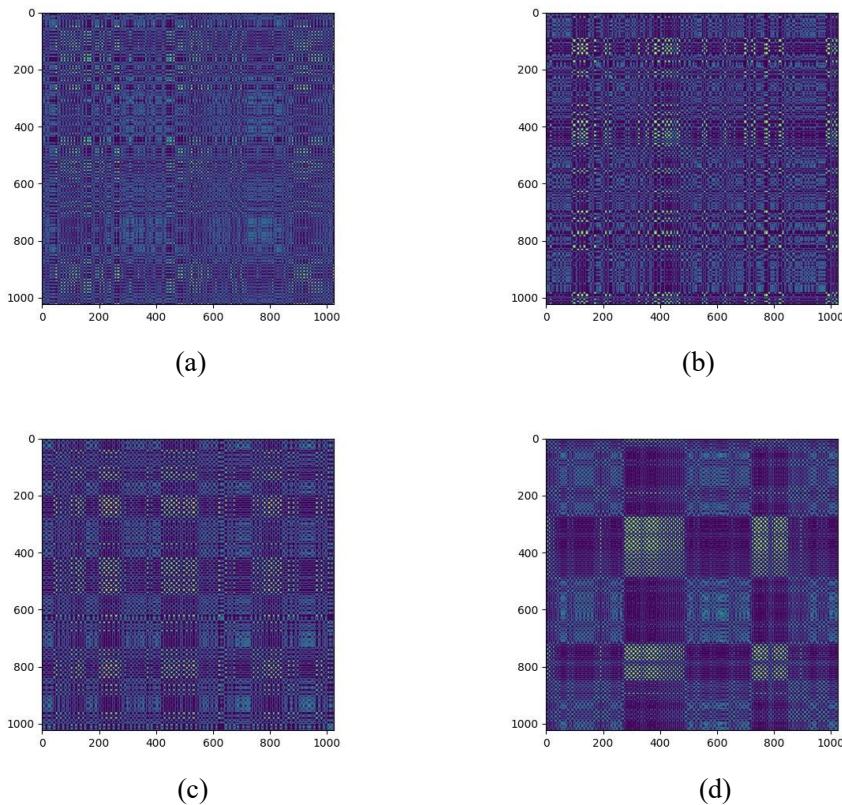


图 2-3 四种故障形式的 MTF 效果图。 (a)正常; (b)内圈故障; (c)滚动体故障; (d)外圈故障

2.2.3 递归图(RP)

递归图(Recurrence Plot, RP) 是一种基于 Poincare 定理的用于可视化时间序列中递归行为的图形化方法。RP 的核心是一种时域信号处理技术，通过将时间序列的周期性、趋势及非线性动态特征转化为二维图像表达。借助递归图中展现的独特规律（如周期性、混沌特性或非规则周期性），可实现时间序列数据的分类、异常检测或模式识别。RP 构建步骤如下：

(1) 时间序列嵌入

对于原始时间序列 $X = \{x_1, x_2, \dots, x_n\}$ ，首先通过时间延迟嵌入将其转换为相空间中的轨迹。相空间中的每个点 y_i 定义为：

$$y_i = (x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}) \quad (2.7)$$

其中： m 是嵌入维度（通常通过经验或算法选择）， τ 是时间延迟，通常通过自相关函数或互信息确定。

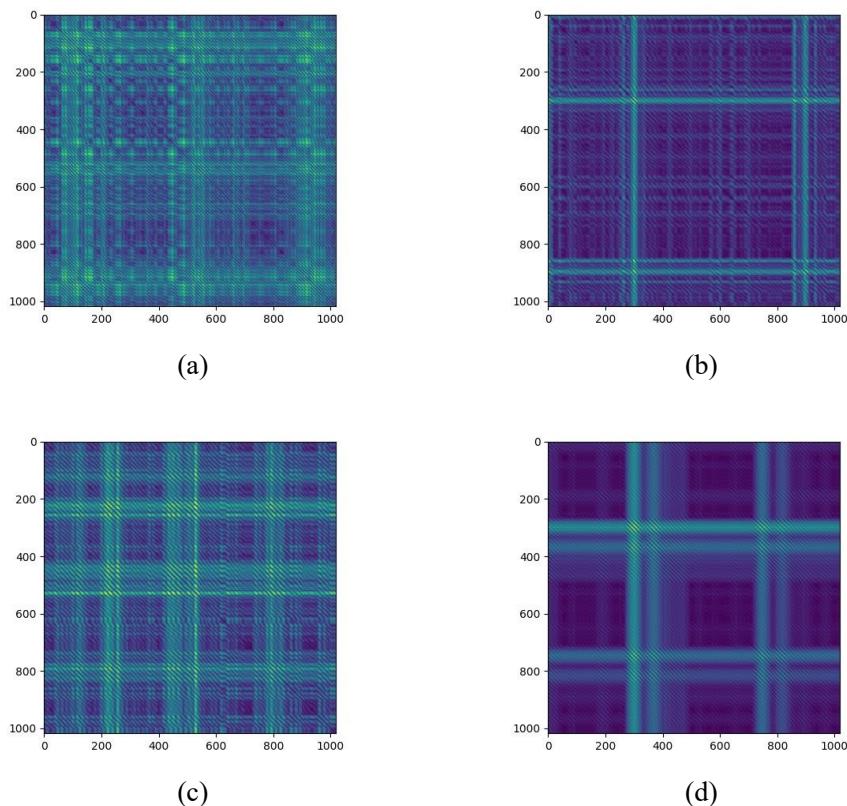


图 2-4 四种故障形式的 RP 效果图。(a)正常; (b)内圈故障; (c)滚动体故障; (d)外圈故障

(2) 计算递归矩阵

递归矩阵 R 是一个 $n \times n$ 的矩阵, 其中每个元素 R_{ij} 表示时间点 i 和时间点 j 的状态是否相似。相似性通常通过欧氏距离或阈值来判断:

$$R_{ij} = \Theta(\epsilon - \|y_i - y_j\|) \quad (2.8)$$

其中: $\Theta(\cdot)$ 是 Heaviside 阶跃函数(当 $x \geq 0$ 时, $\Theta(x)=1$; 否则 $\Theta(x)=0$), ϵ 是一个预设的阈值, 用于判断两个状态是否相似, $\|y_i - y_j\|$ 是相空间中两个点之间的欧氏距离。

(3) 可视化递归图

如图 2-4 所示, 将递归矩阵 R 可视化为图像, 其中每个像素的值对应于矩阵中的元素值 1 或 0。通常, 递归图中的点表示时间序列中的递归行为, 而空白区域表示非递归行为。

2.2.4 短时傅立叶变换(STFT)

短时段傅里叶变换 (Short-Time Fourier Transform, STFT) 能够将信号转化为二维的时频分布图, 具体而言, 通过将信号分割为多个短时片段并对每一片段进行频域变换分析, 从而实现对非平稳信号的时域与频域特性的同步捕捉。与传统的傅里叶变换(FT)不同, STFT 能够处理非平稳信号 (即频率随时间变化的信号), 并提供信号在时间和频率上的局部特征。STFT 步骤如下:

(1) 信号分段

将原始信号 $x(t)$ 分成多个短时段。通常使用滑动窗口函数 $w(t)$ 对信号进行加窗处理。

$$x_w(t, \tau) = x(t) \cdot w(t - \tau) \quad (2.9)$$

其中: $x(t)$ 是原始信号, $w(t - \tau)$ 是窗口函数 (如汉宁窗、汉明窗或矩形窗), τ 是窗口的滑动位置。

(2) 傅里叶变换

对每个加窗后的信号段进行傅里叶变换, 得到该时段的频谱。

$$X(f, \tau) = \int_{-\infty}^{\infty} x_w(t, \tau) \cdot e^{-j2\pi ft} dt \quad (2.10)$$

其中: $X(f, \tau)$ 是时频域表示, f 是频率变量, $e^{-j2\pi ft}$ 是傅里叶变换的核函数。

(3) 计算幅值谱

将 STFT 的复数结果转换为幅值谱，反映信号在不同频率和时间上的能量分布。幅值谱的计算公式为：

$$|X(f, \tau)| = \sqrt{\operatorname{Re}(X(f, \tau))^2 + \operatorname{Im}(X(f, \tau))^2} \quad (2.11)$$

其中： $\operatorname{Re}(X(f, \tau))$ 是实部， $\operatorname{Im}(X(f, \tau))$ 是虚部。

(4) 生成时频图

将幅值谱 $|X(f, \tau)|$ 可视化为时频图。其中时频图的横轴（时间轴）：表示时间 τ 。纵轴（频率轴）：表示频率 f 。颜色或灰度：表示幅值 $|X(f, \tau)|$ 。

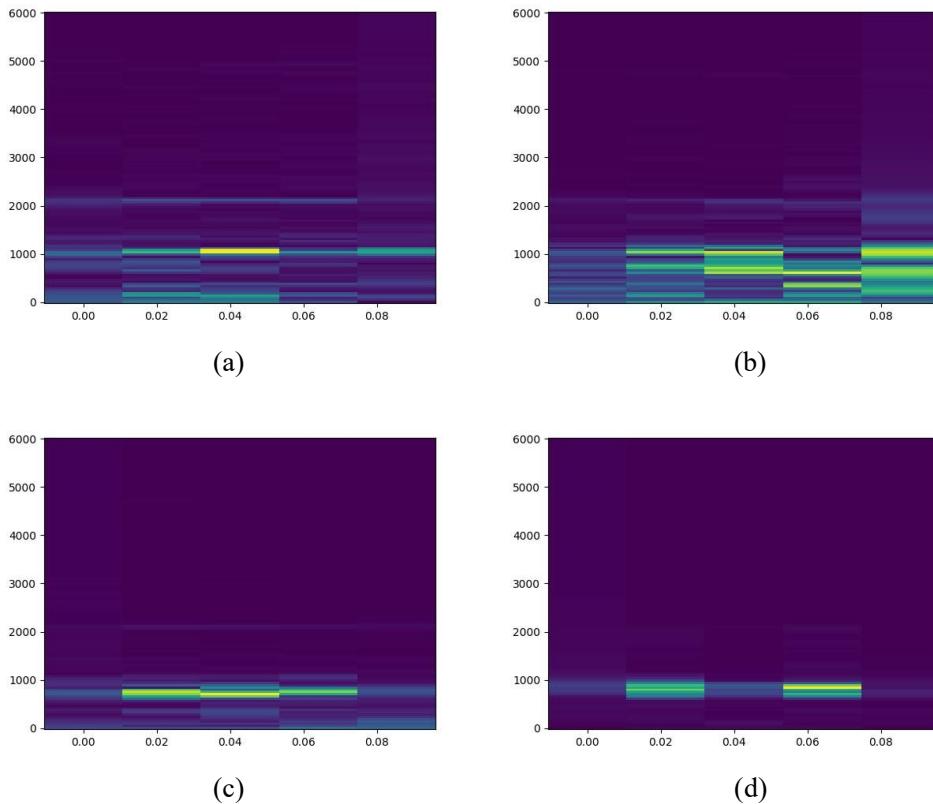


图 2-5 四种故障形式的 STFT 效果图。(a)正常；(b)内圈故障；(c)滚动体故障；(d)外圈故障

(5) 滑动窗口

将窗口函数 $w(t)$ 沿时间轴滑动，重复上述步骤，得到整个信号的时频域表示。最后得到 STFT 效果图如图 2-5 所示。

2.3 集成学习相关理论

2.3.1 集成学习概述

集成学习(Ensemble Learning)是一种通过协同组合多个弱学习器(Base Learner)构建强学习器的机器学习方法。其核心理念在于通过整合多个相似或不同的模型，弥补单一模型在数据特征、偏差或方差方面的局限，以获得更加稳健和精确的预测效果。集成学习主要分为两大类：

- 同质集成：多个弱学习器采用相同的模型结构，但通过不同的数据采样、特征子集等方式获得模型多样性，常见方法有 Bagging、Boosting 等方法。
- 异质集成：弱学习器可以采用不同的模型结构，不同模型在对数据理解上的侧重点不同，通过互补信息达到更优预测效果，常见方法有 Stacking 方法。

2.3.2 Bagging 原理

Bagging(Bootstrap Aggregation)算法最早由 Breiman 提出^[46]，属于一种并行式的同质集成学习技术。这种方法的核心思想是通过有放回抽样(Bootstrap Sampling)，从原始训练数据中生成多个子集，在每个子集上分别训练一个弱学习器，最后将各自的预测结果整合成一个强学习器，从而降低整体模型的方差。Bagging 的具体过程如图 2-6 所示：

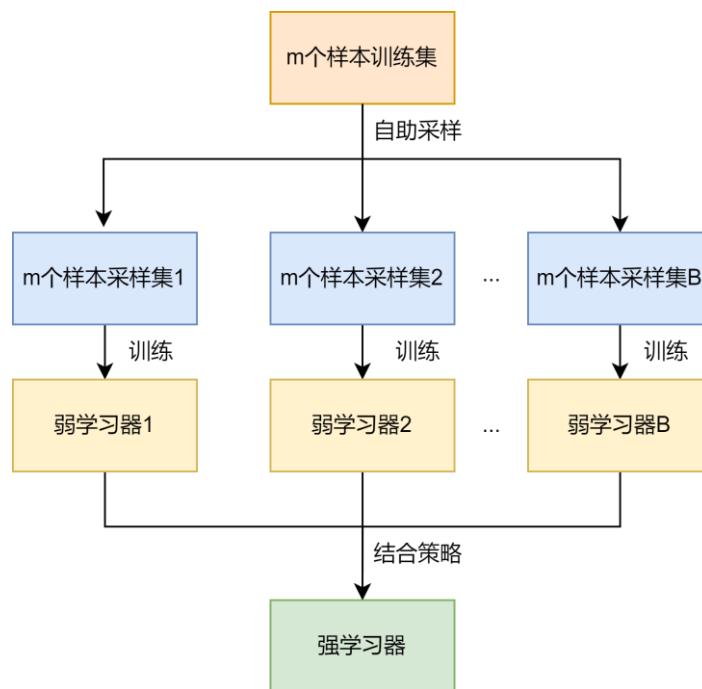


图 2-6 Bagging 流程图

Bagging 的核心流程可分为以下三个阶段：

- 自助采样(Bootstrap Sampling)

给定原始训练集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ ，采用有放回随机抽样生成 B 个大小为 m 的子训练集 D_1, D_2, \dots, D_B 。由于采样过程允许重复，每个子集的样本分布存在差异。依据概率论的结论，当原始训练集容量较大时，在每次采样过程中，大约有 36.8% 的样本不会被选中，这部分样本被称为袋外数据（Out-of-Bag, OOB），可用于对模型性能进行无偏估计。

- 基模型训练(Base Learner Training)

对每个子训练集 $D_b (b=1, 2, \dots, B)$ ，独立训练一个弱学习器 $f_b(x)$ 。弱学习器通常选择高方差、低偏差的模型（如未剪枝的决策树），以最大化集成的多样性。

- 预测聚合(Prediction Aggregation)

通过结合策略将弱学习器的输出统一为最终预测：若是回归任务，可以采用简单平均法通过平均操作平滑噪声，降低预测方差。

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (2.12)$$

若是分类任务，则可以采用多数投票法，能通过“少数服从多数”原则提升决策鲁棒性。

$$\hat{y} = \arg \max_{c \in C} \sum_{b=1}^B \mathbb{I}\{f_b(x) = c\} \quad (2.13)$$

其中， C 为类别集合， $\mathbb{I}\{\cdot\}$ 为指示函数，当且仅当弱学习器预测为类别 c 时取值为 1。

Bagging 通过样本扰动增强模型多样性，有效降低集成结果的方差，尤其适用于对噪声敏感的基模型（如深度决策树）。其泛化性能提升可直观理解为：多个独立弱学习器的集体决策抵消了单模型的过拟合风险。经典的随机森林(Random Forest)算法就是基于 Bagging 框架，并进一步通过特征随机扰动提升模型多样性。

2.3.3 Boosting 原理

Boosting 是一种串行同质集成学习方法，其核心思想是逐步纠正前一轮模型的预测误差，并动态调整训练样本权重，使后续模型更加关注难以分类的样本。简单来说，Boosting 算法根据前一弱学习器的表现重新分配样本权重，从而使后续弱学习器集中处理之前被误判的样本。AdaBoost 是 Boosting 的代表性算法之一，其基本流程如图 2-7 所示。

假设训练集包含 n 个样本，初始时对所有样本赋予均匀权重：

$$w_i^{(1)} = \frac{1}{n}, \quad i = 1, 2, \dots, n. \quad (2.14)$$

在第 t 轮迭代中 ($t=1, 2, \dots, T$)，执行以下步骤：先进行弱学习器训练，基于当前样本权重 $w_i^{(t)}$ ，训练弱学习器 $h_t(x)$ ，并计算其加权分类误差率：

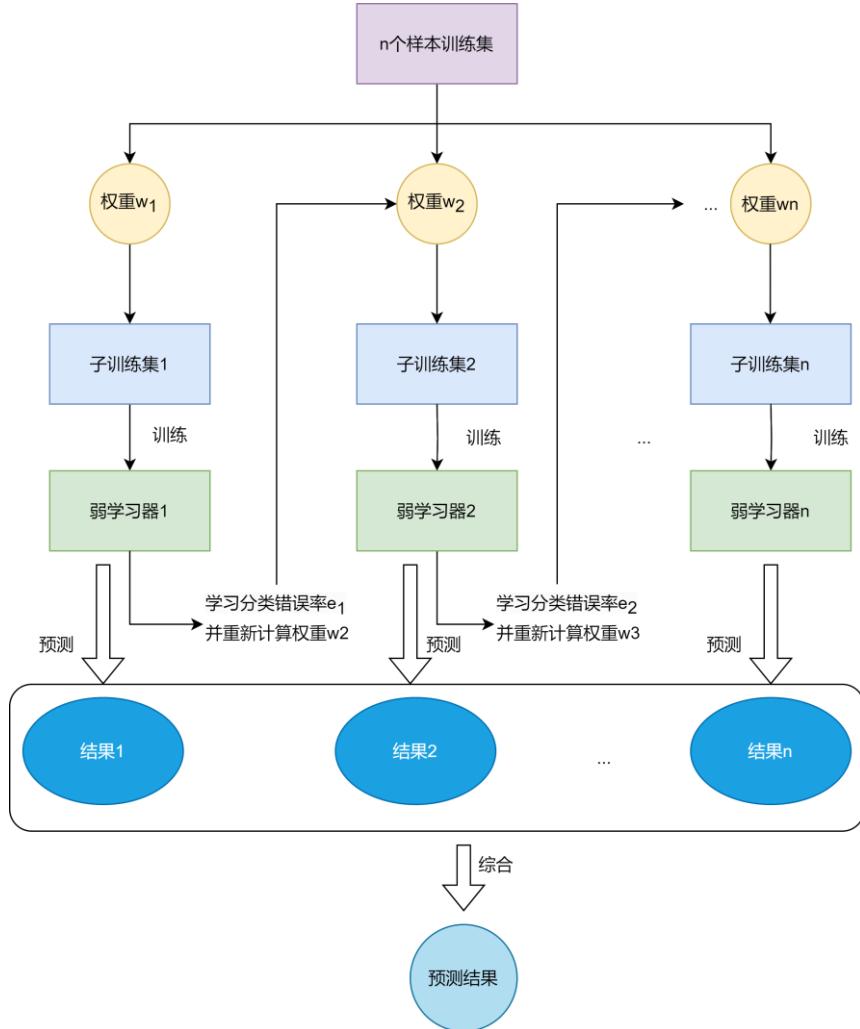


图 2-7 Boosting 流程图

$$\epsilon_t = \frac{\sum_{i=1}^n w_i^{(t)} \cdot \mathbb{I}\{h_t(x_i) \neq y_i\}}{\sum_{i=1}^n w_i^{(t)}} \quad (2.15)$$

其中， $\mathbb{I}\{\cdot\}$ 为指示函数，当条件满足时取 1，否则取 0。

之后，计算弱学习器权重，根据误差率 ϵ_t ，确定当前弱学习器的投票权重：

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (2.16)$$

其中，当 $\varepsilon_t \rightarrow 0$ 时， $\alpha_t \rightarrow +\infty$ ，即分类效果极佳，此时模型权重显著增加；因为弱学习器性能不优于随机猜测，当 $\varepsilon_t \geq 0.5$ 时，算法终止。

之后通过更新样本权重，调整错误分类样本的权重，使模型更加关注误分类的样本：

$$w_i^{(t+1)} = w_i^{(t)} \cdot \exp(-\alpha_t \cdot y_i \cdot h_t(x_i)) \quad (2.17)$$

由于 $y_i \in \{-1, +1\}$ 且 $h_t(x_i) \in \{-1, +1\}$ ，此时分类正确时候 $y_i h_t(x_i) = +1$ ，权重降低；分类错误时 $y_i h_t(x_i) = -1$ ，权重升高。为避免权重指数爆炸，对更新后权重进行归一化：

$$w_i^{(t+1)} \leftarrow w_i^{(t+1)} / \sum_{j=1}^n w_j^{(t+1)} \quad (2.18)$$

最后，通过最终强学习器为所有弱学习器的加权组合：

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (2.19)$$

其中 $\text{sign}(\cdot)$ 为符号函数，输出类别标签 (± 1)。通过加权投票机制，分类结果由高置信度弱学习器主导。

Boosting 通过逐步修正前序模型的错误，可以显著提升整体模型的预测性能，同时 Boosting 天然适合类别不平衡场景，可以通过增加少数类样本的权重，模型能够更关注稀有类别，提升召回率。但可能带来过拟合风险，需要在迭代轮次上进行适当控制，而且存储多个模型参数及中间状态需较高内存，尤其在超参搜索时加剧资源压力。

2.3.4 Stacking 原理

Stacking 是一种异质集成学习技术，通常把若干弱学习器（Base Learners）的预测结果当作新的特征，然后通过一个元学习器（Meta-learner）对这些结果进行综合，以生成最终的预测。其基本公式可表述为：

$$\hat{y} = g(f_1(x), f_2(x), \dots, f_M(x)) \quad (2.20)$$

其中， x 表示输入样本； $f_1(x), f_2(x), \dots, f_M(x)$ 分别是第 M 个弱学习器对样本 x 的预测输出； $g(\cdot)$ 是训练得到的元学习器，它利用弱学习器的输出作为输入特征，给出最终预测 \hat{y} 。通常分为两个阶段进行：

第一阶段，对原始数据集进行交叉验证或直接训练多个弱学习器，从而获得每个样本的预测结果。

第二阶段，使用第一阶段获得的预测结果构成的新特征矩阵，训练一个元模型 g 来学习如何最佳地整合各个弱模型的输出。

Stacking 能有效利用了各个模型的优势，并通过元学习器捕捉模型间的互补信息，从而通常能取得比单一模型更优的预测性能。

2.4 轴承故障诊断概述

作为旋转机械的重要组成部分，轴承的状态直接决定了设备运行的可靠性和安全性。本节从轴承的基本结构、典型故障形式及故障诊断流程三个方面展开论述，为后续基于领域泛化的诊断方法研究奠定理论基础。

2.4.1 轴承的基本构成

滚动轴承的典型结构如图 2-8 所示，主要由以下四部分组成：

- (1) 内圈(Inner Ring): 与旋转轴紧密配合，随轴转动，内圈滚道表面与滚动体接触传递载荷。
- (2) 外圈(Outer Ring): 固定于轴承座中，外圈滚道为滚动体提供运动轨道。
- (3) 滚动体(Rolling Elements): 包括球体、圆柱滚子或圆锥滚子等，通过滚动运动降低摩擦阻力，均匀分布载荷。
- (4) 保持架(Cage): 用于分隔滚动体，防止其相互碰撞并引导其沿滚道均匀分布。

轴承的几何参数（如滚道曲率、接触角）与材料性能（如硬度、抗疲劳性）共同决定了其承载能力与寿命。当轴承长期承受交变应力时，其内部组件易因疲劳、磨损或润滑失效而发生故障^[47]。

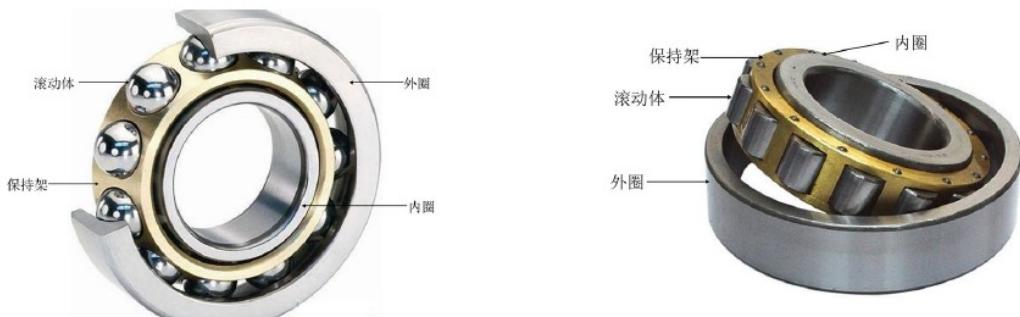


图 2-8 滚动轴承典型结构

2.4.2 轴承典型故障形式

轴承的常见的典型故障形式可分为局部损伤与退化性故障两类：

- 局部损伤型故障

- (1) 内圈故障：由过载或润滑不良引发，表现为滚道表面点蚀、剥落或裂纹，振动信号中会出现特征频率成分，比如内圈故障频率(BPFI)。
- (2) 外圈故障：多由安装不当或异物侵入导致，损伤集中在滚道区域，振动能量在外圈故障频率(BPFO)处显著增强。
- (3) 滚动体故障：滚动体表面损伤（如压痕、碎裂）会周期性冲击内外圈，产生以滚动体自转频率(BSF)为基频的调制信号^[48]。
- (4) 保持架故障：保持架断裂或变形将导致滚动体运动失稳，引发非周期性冲击振动。

- 退化性故障

综合故障：长期磨损导致轴承游隙增大、表面粗糙度增加，表现为振动幅值缓慢上升与谐波成分增多，最终发展为综合故障。

统计表明，约 70% 的轴承失效源于局部损伤，且故障初期特征微弱，易被噪声掩盖^[2]，这对轴承故障诊断方法提出了更高要求。

2.4.3 轴承故障诊断基本流程

轴承故障诊断是通过对运行状态信号的采集与分析，实现故障类型识别与健康状态评估的技术过程。其核心流程可分为信号采集、特征提取与状态识别三个阶段，如图 2-9 所示，各阶段的具体内容如下：

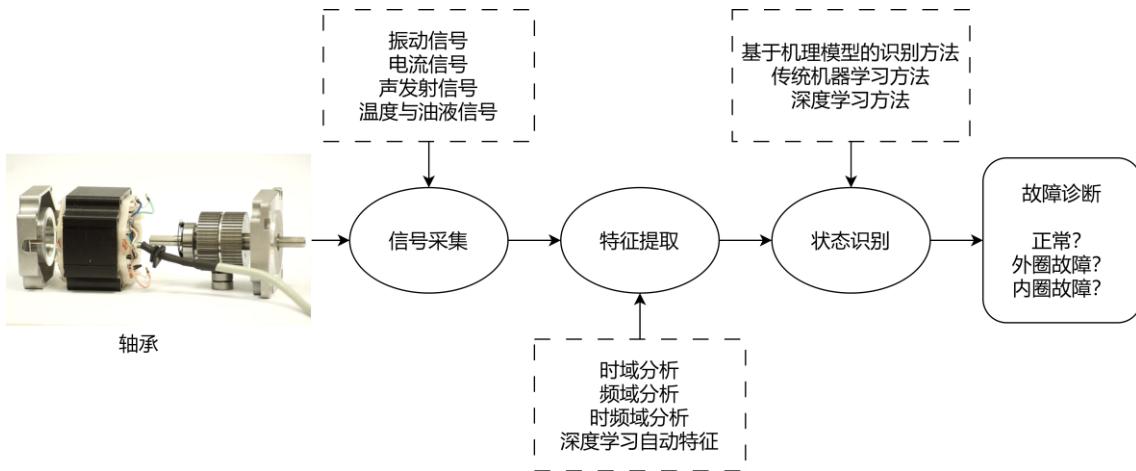


图 2-9 轴承故障诊断基本流程图

- 信号采集

信号采集是故障诊断的基础环节，其质量直接影响后续分析的可靠性。根据传感原理与物理量类型，主要采集以下信号：

(1) 振动信号：通过加速度传感器，比如压电式、MEMS 式加速度传感器采集轴承振动响应，能够敏感捕捉局部损伤引发的瞬态冲击，是故障诊断最常用的信号类型。

(2) 电流信号：利用电流互感器监测电机定子电流，通过调制效应间接反映轴承异常，适用于难以安装振动传感器的场景^[49]。

(3) 声发射信号：采用 20 kHz 以上的高频麦克风捕获材料微观变形或裂纹扩展释放的弹性波，会对早期故障敏感^[50]。

(4) 温度与油液信号：通过红外传感器或光谱仪监测轴承温升与润滑油磨粒成分，用于退化趋势分析^[51]。

● 特征提取

特征提取旨在从原始信号中获取与故障相关的敏感信息，方法可分为传统手工特征和深度学习自动特征两类：

(1) 传统手工特征：包括时域特征、频域特征和时频域特征。时域特征如均方根值(RMS)、峰峰值和峭度(Kurtosis)，用于表征信号的能量分布和冲击特性^[52]；频域特征通过快速傅里叶变换(FFT)提取故障特征频率的幅值，并结合包络分析提高信噪比；时频域特征则利用小波变换(WT)或希尔伯特-黄变换(HHT)分析非平稳信号的时变特性，以确定故障发生的时频范围^[53]。

(2) 深度学习自动特征：基于卷积神经网络(CNN)或自编码器(AE)，直接从原始信号中学习高维抽象特征，避免人工设计的局限性。例如，一维 CNN 能够自动识别振动信号中的局部共振模式^[54]。

● 状态识别

状态识别通过模式分类技术将提取的特征映射到故障类别，其发展历程可分为三个阶段：

(1) 基于机理模型的识别方法：根据轴承的几何参数和转速计算理论故障频率，通过频谱峰值匹配进行诊断。

(2) 传统机器学习方法：包括支持向量机(SVM)^[55]，随机森林(RF)^[49]，贝叶斯分类器^[50]。

(3) 深度学习方法：包括卷积神经网络(CNN)^[51]和长短时记忆网络(LSTM)^[56]。近些年，深度学习技术得益于计算硬件的提升和理论突破，取得了快速发展，为故障诊断领域带来了诸多创新方案。作为机器学习的关键分支，深度学习能够从海量数据中自主挖掘有效特征，显著提高故障诊断分类任务中的预测精度^[57]。

2.5 实验平台介绍

如图 2-10 所示，本次的采集数据集所用的轴承故障诊断实验平台是基于动力传动故障诊断与仿真实验台搭建的。该平台配备了双级定轴减速传动系统，专门用于安装故障轴承。其核心组件包括驱动电机、联轴器、行星齿轮箱、负载齿轮箱、负载电机以及加速度传感器等。此外，平台还集成了信号采集系统、控制柜和计算机等配套设备^[58]。各部件的详细技术参数可参见表 2-1。为实现轴承传动系统的驱动功能，实验平台选用了变频器与电机协同工作的方案作为动力来源。通过 PLC 对变频器的精确调控，实现了对电机转速的灵活调整，以满足恒定转速和可变转速等多种实验工况的要求。

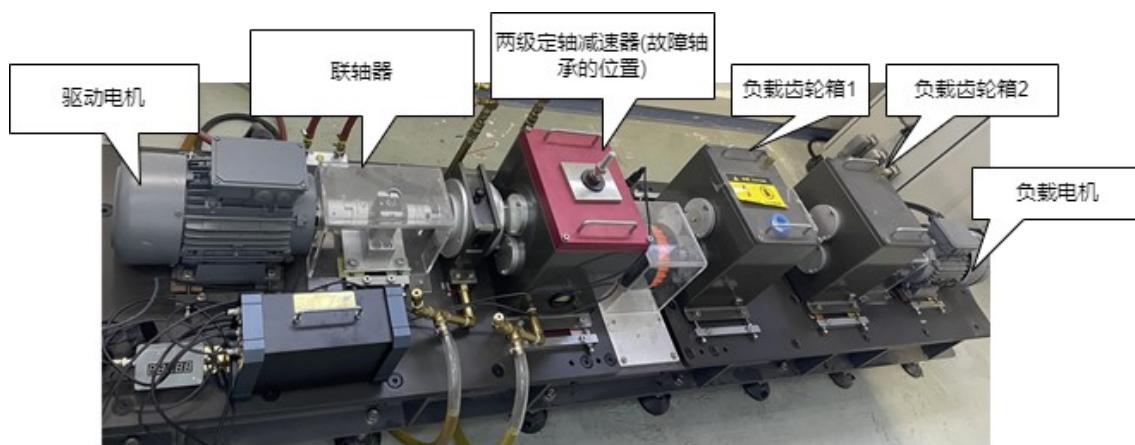


图 2-10 动力传动故障诊断与仿真实验平台

表 2-1 动力传动故障诊断实验平台参数

组成部分	说明
驱动电机、负载电机	三相，功率 10 马力，转速范围 0-3600 RPM
工作电压	50Hz, 220V 和 380V (三相)
变频器	3 马力变频交流驱动
行星齿轮箱	二级，传动比为 27:1，采用行星齿轮系
平行轴齿轮箱	二级，每级最大传动比为 2.5，配备直齿轮或斜齿轮
平行轴齿轮箱轴承	深沟球轴承
轴承负载	3000 磅
负载齿轮箱	两个两级负载齿轮箱，最大总传动比为 20:1
重量、尺寸	约 1000kg；长 2850mm，宽 560mm，高 560mm

振动传感器在该平台中负责采集滚动轴承的振动信号，确保信号输出的稳定性。在众多传感器类型中，ICP 型加速度传感器因其出色的抗干扰性能而成为市场主流。本课题在选择传感器时，重点考量了四个关键指标：高灵敏度、强抗干扰能

力、广泛的振动频率检测范围以及小巧的体积。综合评估后，选用了美国 PCB 公司生产的 604B31 三轴压电式加速度传感器。

这种传感器设计有磁性底座，安装简便，无需借助其他工具，可直接吸附在轴承座上，实现振动信号的无损检测。如图 2-11 所示，传感器的安装位置位于轴承座外侧，这种布置方式不仅便于操作，还能有效避免对旋转部件造成损伤，完美契合无损测量的设计理念。然而，在实际操作中，如果安装位置存在油污、灰尘或铁屑等杂质，会影响传感器与轴承座的紧密贴合，甚至可能在平台运行时导致传感器移位。因此，安装前必须彻底清洁安装位置，确保无任何杂质。

在滚动轴承故障诊断实验中，信号采集单元作为前端处理的重要环节，主要负责采集振动信号并存储数据。该模块先通过传感器获取信号，再将数据传输至计算机进行处理与存储。

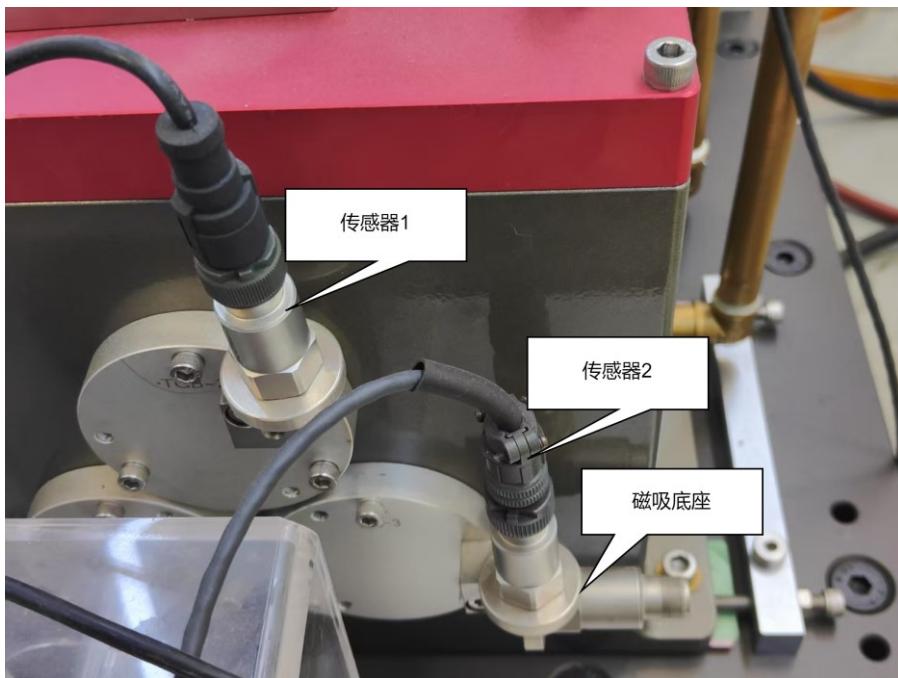


图 2-11 传感器位置

平台选用了晟阳科技生产的 USB-8-NFC8-0016 型动态信号数据采集系统。该系统具备 8 个通道，采样频率高达 256KHz，采用直流 24V 供电。

在滚动轴承故障诊断中，常见的实验方法包括预先设置故障实验和疲劳实验。预先设置故障实验通过对正常轴承进行人工加工，制造已知故障，然后在平台上测试并采集振动信号数据。该方法能快速获取故障频率数据，但人工故障过于典型，与实际生产中的故障有一定差异。疲劳实验更接近实际工作状态，但周期长且故障特征不够明显。综合考虑，本文采用了预先设置已知故障的实验方案。

实验的具体步骤如下：

- 1、轴承安装：将预设故障的轴承安装到指定轴承座上，确保牢固可靠。
- 2、清洁与传感器安装：彻底清洁轴承座，去除杂质后将传感器固定在预定位置。
- 3、系统连接：用联轴器连接电机轴与传动轴，确保传感器、采集器和计算机对接无误，检查接口牢固性后接通电源。
- 4、电机启动与参数设置：启动电机，设定转速和负载条件。
- 5、信号采集准备：在采集系统中预设参数，检查信号波形，待轴承运行平稳后开始采集。
- 6、数据处理与分析：对采集数据进行处理和分析，以诊断故障。

2.6 本章小结

本章探讨了领域泛化的相关知识，为后续分析领域泛化故障诊断技术奠定了理论基础，阐述了四种故障增强方法以及集成学习的相关技术，为后续内容建立了理论依据。同时，本章还涉及轴承故障诊断的相关内容，并介绍了本次实验所用的实验平台，为后面章节验证模型效果提供了数据支持和准备工作。总体来看，本章从理论、方法和实验三个方面搭建了完整的轴承故障诊断研究框架，为后续章节提供了理论支撑和数据基础。

第三章 基于多模态数据增强的跨域轴承故障诊断方法

3.1 引言

轴承作为旋转机械系统的核心传动部件，其运行状态的精确诊断直接决定了工业设备的安全性与可靠性。然而，实际工业场景中复杂的动态载荷与多源工况耦合效应，导致故障诊断模型面临单源域训练数据受限于设备监测成本与运行环境约束，呈现显著的数据稀疏性与类别不平衡特征的问题。在此背景下，如何通过有限源域数据构建具有跨工况适应能力的诊断模型，成为工业智能运维领域亟待突破的技术瓶颈。

针对数据稀缺问题，现有研究多采用单模态数据增强策略，比如时域信号重采样、频域噪声注入等，但此类方法存在本质局限，其一，单一信息域的特征生成难以全面表征轴承故障的多物理场耦合特性；其二，传统生成模型易受梯度消失问题影响，导致生成样本与真实数据分布偏离显著。为此，本章构建了融合多模态特征与对抗生成网络的跨域数据增强框架 MDWGANGP(Multi-Domain WGAN-GP)，其通过融合一维时序信号和二维图像特征，生成更高质量的训练数据，进而提高模型在单源域条件下的故障诊断跨域能力。

本章的研究不仅为变工况轴承故障诊断提供了一种新的数据增强思路，也为工业设备故障诊断领域的领域泛化问题提供了有益的参考。

3.2 MDWGANGP 模型框架

本次提出的 MDWGANGP 模型由数据预处理模块、特征提取模块、生成对抗模块以及故障分类模块这四大模块构成，其结构如图 3-1 所示，该框架的工作流程和关键模块：

1、数据预处理模块

MDWGANGP 框架首先对初始数据进行预处理和增强，以提升数据质量和多样性：第一步：数据增强，通过奇异值分解(SVD)对初始数据进行分解，提取数据的内在结构，减少噪声干扰，为后续处理奠定基础；第二步：网络嵌入更新(NEU)，对 SVD 分解后的矩阵进行优化，生成高阶重构数据。这一步骤不仅提高了数据的质量，还增加了数据的多样性，为特征提取提供了更丰富的输入。

2、特征提取模块

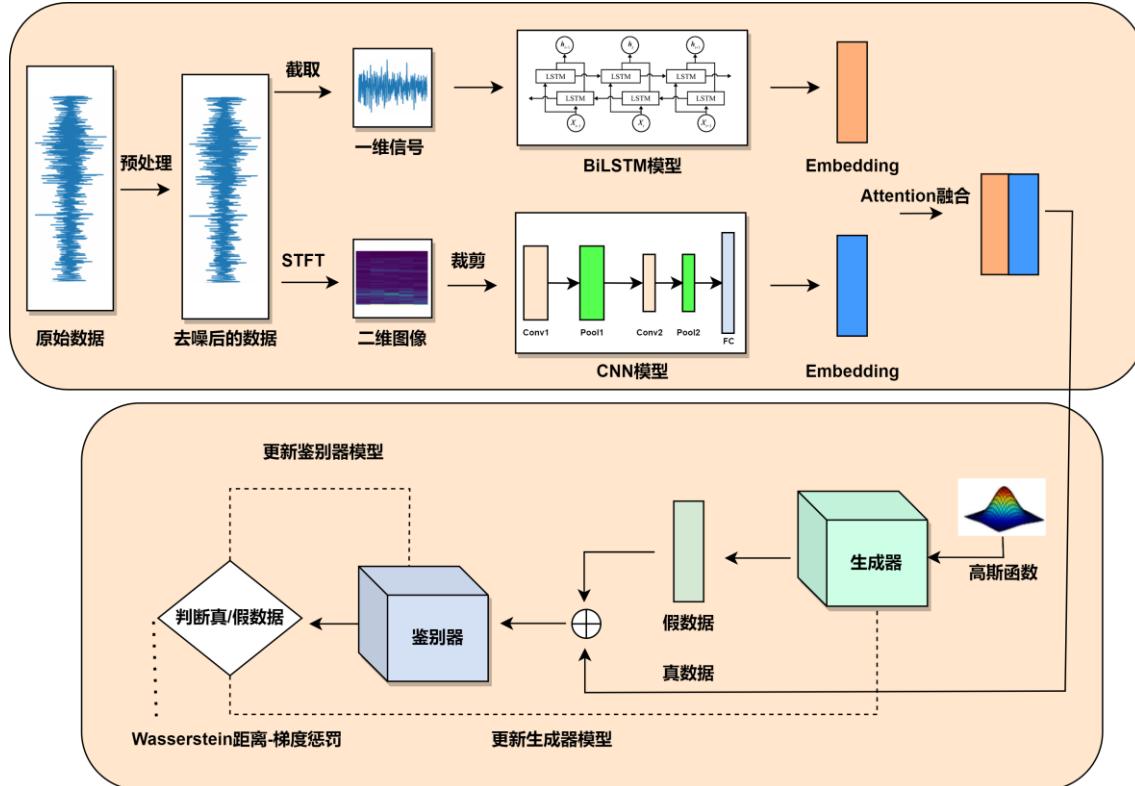


图 3-1 MDWGANGP 模型框架

为了全面捕捉数据的特性，MDWGANGP 采用双路径策略分别提取时域和频域特征：一方面，通过时序特征提取，采用双向长短期记忆网络（BiLSTM）对一维时序信号进行处理。BiLSTM 能够捕捉信号的时间依赖性，提取反映轴承运行状态的关键时序特征；另一方面，通过频域特征提取，采用短时傅里叶变换（STFT）将时序信号转换为二维时频图，并利用卷积神经网络（CNN）提取其空间特征。该过程可捕捉信号的频率变化及局部模式，为故障诊断提供频域信息。随后，框架引入注意力机制，实现时域与频域特征的融合。该机制自动学习并加权不同特征的重要性，充分挖掘数据的多层次信息，从而显著增强模型的泛化能力。

3、生成对抗模块

为应对工业环境中数据量有限及分布不均的挑战，MDWGANGP 框架引入了基于梯度约束的 Wasserstein 对抗网络（WGAN-GP）进行数据增强。相较于传统 GAN 模型，该技术通过采用 Wasserstein 度量作为损失函数，实现了更平滑的梯度更新，有效减少了训练过程中因梯度消失或爆炸导致的不稳定现象。此外，通过在损失函数中引入梯度约束策略，强制判别器满足 Lipschitz 连续性要求，避免了生成样本多样性不足的问题，确保了生成数据的真实性和高质量。该方法在轴承故障诊断领域的小样本场景中展现了显著的实用价值。

4、故障分类模块

利用 WGAN-GP 生成的高质量增强数据，故障分类模块对不可见的目标域进行故障分类。这一模块能够应对数据分布不一致的挑战，实现跨领域的轴承故障诊断。

MDWGANGP 框架具备以下优势：首先，通过 BiLSTM 和 CNN 的双路径特征提取策略，结合注意力机制，实现了时序和频域特性的全面捕捉；其次，通过数据增强和对抗学习，利用 WGAN-GP 技术扩展了训练数据，提高了模型在动态工况下的适应能力；最后，强大的跨域能力，特征融合与数据生成技术的协同作用，使框架在未知目标域中表现出色，显著提升了诊断的精确性和稳定性。

MDWGANGP 框架通过数据预处理、多层次特征提取及 WGAN-GP 数据生成策略，有效解决了工业设备故障诊断中数据量不足和工况复杂多变的问题。其特征融合与对抗学习机制显著增强了模型的跨域适应性，为复杂工业场景下的轴承健康监测和故障诊断提供了有力的技术保障。

3.2.1 数据预处理模块

本次的原始数据是连续的一维时序振动信号，是以连续 2048 个振动点作为一个数据样本。数据预处理是故障诊断流程中的关键环节，能够显著提升后续模型训练的效果。本章选用了 Min-Max 标准化作为预处理方法。该方法通过线性变换，将轴承数据的取值范围映射到 [0, 1] 区间内。这种标准化方式特别适用于数据分布未知或具有明显边界的情形，能够有效消除量纲差异，增强数据的可比性，为后续的特征提取和模型训练奠定基础。Min-Max 标准化公式是：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3.1)$$

其中， x 是原始数据， x_{\max} 和 x_{\min} 分别是数据在该特征上的最小值和最大值， x' 是标准化后的数据。

在完成数据标准化后，本研究对初始数据进一步应用了奇异值分解(SVD)转换。SVD 是一种高效的矩阵分解方法，可将任意实数矩阵分解为三个矩阵的乘积形式。这一过程不仅能挖掘数据中的潜在结构和特性，还能有效去除噪声，进而提升数据的整体质量。对大小为 $M \times N$ 的矩阵 M 进行 SVD 分解：

$$M_{m \times n} \approx U_{m \times k} H_{k \times k} (V_{k \times n})^T \quad (3.2)$$

其中, m 和 n 分别代表矩阵 M 的行向量和列向量; k 表示特征值个数; MM^T 的正交单位特征向量构成 U , 特征值构成 HH^T , MM^T 的正交单元特征向量构成 V , 特征值构成 HH^T 。

之后通过网络嵌入更新(NEU)优化分解矩阵, 得到高阶重构数据, NEU 是一种改进网络表示学习的有效方法, 通过隐式近似高阶邻近度, 能够更好地捕捉网络的全局结构和局部结构信息, NEU 具有通用性强、嵌入质量高的特点, 其目的是通过矩阵对向量进行变换, 更新方程如下:

$$R' = R + \lambda A \cdot R \quad (3.3)$$

$$C' = C + \lambda A^T \cdot C \quad (3.4)$$

其中, λ 是一个参数且 $\lambda \in (0, 0.5]$ 。 A 是奇异值分解矩阵中的对角矩阵。 R 和 C 分别是网络表示向量和上下文表示向量。

通过奇异值分解(SVD)与网络嵌入更新(NEU), 可以滤除噪声并重构高阶特征矩阵, 以此增强数据质量。

3.2.2 特征提取模块

特征提取模块作为 MDWGANGP 框架的核心组成部分, 通过时域和频域两个方向的多维度分析, 结合特征融合与优化技术, 实现了对轴承振动信号的全面表征。特征提取模块与框架其他部分的协同作用进一步增强了模型在复杂工况下的诊断能力, 为轴承故障诊断提供了高效、鲁棒的解决方案。

特征提取模块的一个分支利用双向长短期记忆网络 (BiLSTM) 来挖掘一维时间序列数据的特征。如图 3-2 所示, BiLSTM 通过结合正向 LSTM 和反向 LSTM 的处理能力, 分析一维时序数据。与单向 LSTM 不同, BiLSTM 能够同时利用数据的过去和未来信息, 从而更全面地捕捉时序信号中的上下文依赖性。

正向 LSTM 从序列的起点到终点处理数据, 生成隐藏状态序列:

$$\bar{h}_t = \text{LSTM}(x_t, \bar{h}_{t-1}) \quad (3.5)$$

反向 LSTM 从序列的终点到起点处理数据, 生成隐藏状态序列:

$$\tilde{h}_t = \text{LSTM}(x_t, \tilde{h}_{t-1}) \quad (3.6)$$

将正向和反向的隐藏状态组合, 得到最终的输出:

$$y_t = \sigma\left(W_y \cdot [\bar{h}_t, \tilde{h}_t] + b_y\right) \quad (3.7)$$

其中, W_y 表示权重矩阵, b_y 表示偏差项。

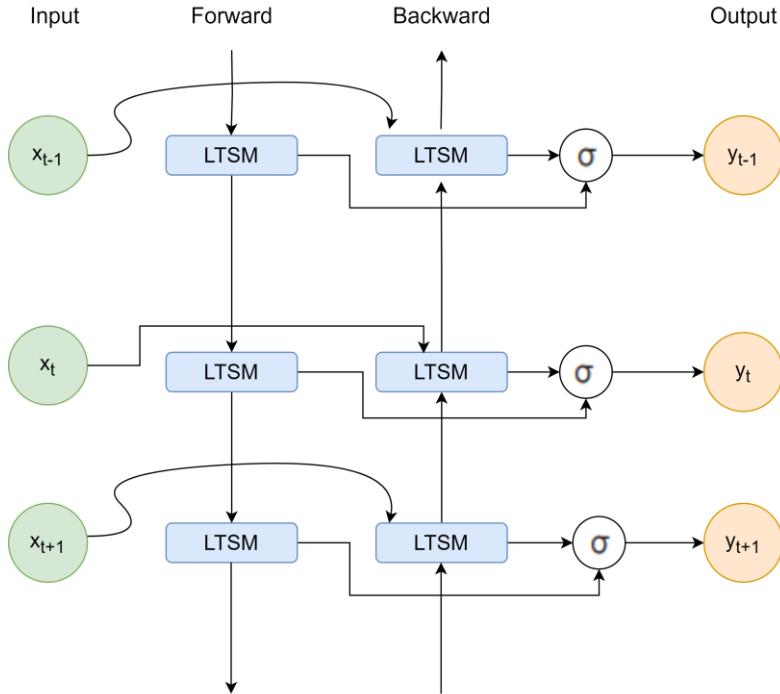


图 3-2 BiLSTM 结构

特征提取模块的另一分支首先采用短时傅里叶变换(STFT)将一维信号转化为二维图像，随后通过擅长图像特征识别的卷积神经网络(CNN)进行特征提取。具体而言，STFT生成时频图后，CNN 专注于挖掘频域中的局部特征，其结构如图 3-3 所示。

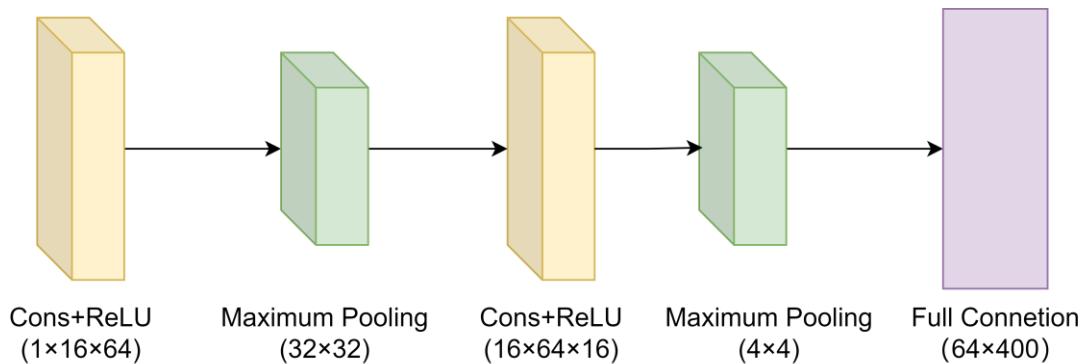


图 3-3 CNN 结构

卷积层通过卷积运算获取数据的空间或时间结构特征，每个卷积核均配备一个权重参数和一个偏置项，其计算方式如下：

$$s(i, j) = \sum_{k=1}^n (X_k * W_k)(i, j) + b \quad (3.8)$$

其中, n 表示输入矩阵的总数, X_k 表示第 k 个输入矩阵, W_k 表示卷积核中的第 k 个子核矩阵, b 代表偏置项。

池化层紧随卷积层之后, 利用不同尺寸的窗口进行下采样, 以筛选和精简特征, 从而降低数据规模。常见的池化方式包括平均池化和最大池化, 本研究选用的是平均池化, 其计算公式为:

$$Y(i, j) = \frac{1}{k \times k} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} X(i \cdot s + m, j \cdot s + n) \quad (3.9)$$

其中, k 表示池化窗口的尺寸, s 表示滑动步长, m 和 n 分别表示局部区域的索引位置。

全连接层通常负责将前层输出映射至最终的分类或回归结果, 能够融合特征并生成预测输出。输出结果一般通过激活函数 Softmax 进行非线性处理, Softmax 的计算公式如下:

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^m e^{x_j}} \quad (3.10)$$

其中, x_i 是第 i 个类别的原始输出(Logits), e^{x_i} 为指数函数, 将得分转换为正值范围, $\sum_{j=1}^m e^{x_j}$ 是所有类别的指数和, 用于归一化。

最后将 BiLSTM 提取的一维时序信号的嵌入(Embedding)与 CNN 提取的二维图像的嵌入(Embedding)进行注意力机制(Attention Mechanism)加权融合的方法, 能够动态地捕捉不同嵌入表示的重要性, 从而生成更有效的融合表示, 其方法是:

首先, 对于每个嵌入表示 e_i , 计算其注意力分数 s_i :

$$s_i = f(e_i, q) \quad (3.11)$$

其中, f 是注意力函数, q 是查询向量(Query Vector)。

随后, 采用 Softmax 函数将注意力分数转化为权重, 其计算方式如下:

$$\alpha_i = \frac{e^{s_i}}{\sum_{j=1}^n e^{s_j}} \quad (3.12)$$

最终, 根据注意力权重对嵌入表示进行加权求和, 得到结果:

$$e_{\text{fused}} = \sum_{i=1}^n \alpha_i e_i \quad (3.13)$$

3.2.3 生成对抗模块

生成对抗网络(GAN)最早由 Ian Goodfellow 等人^[17]于 2014 年提出，是一种基于无监督学习的深度学习架构。该框架包含生成器 G 和判别器 D 两个核心部分，其训练过程体现为一个极小化极大(Minimax)博弈问题，目标函数表达如下：

$$\min_G \max_D (V(D, G) = \mathbb{E}_{x \sim P_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim P_z(z)} [\log(1 - D(G(z)))]) \quad (3.14)$$

其中， $D(x)$ 表示判别器对真实样本 x 的判别结果， $D(G(z))$ 是判别器对生成样本 $G(z)$ 的输出， $P_{\text{data}}(x)$ 代表真实数据的分布， $P_z(z)$ 是噪声的分布。

传统 GAN 采用 JS(Jensen-Shannon)散度作为损失函数，用于衡量两个概率分布之间的相似程度。通过最小化生成样本与真实样本分布间的差异，JS 散度推动生成器 G 产出更高质量的样本。然而，JS 散度的应用存在局限性：当真实数据与生成数据的分布无重叠或重叠较少时，JS 散度趋于常数，容易引发梯度消失问题，导致训练过程不稳定。

为解决 JS 散度在模型训练中导致的不稳定性问题，Arjovsky 等人^[67]提出了一种创新方法，即采用 Wasserstein 度量替代 JS 散度，并在此基础上开发了 WGAN 模型。Wasserstein 度量能够更精确地衡量生成数据与真实数据之间的分布差异，同时为模型提供更具指导性的梯度信息，从而显著提升训练过程的稳定性和可靠性。其目标函数为：

$$L_{\text{WGAN}} = \min_G \max_{D \in \Delta} V(D, G) = \mathbb{E}_{x \sim P(x)} [D(x)] - \mathbb{E}_{z \sim P(z)} [D(G(z))] \quad (3.15)$$

其中， Δ 表示满足 1-Lipschitz 条件的函数集合。WGAN 通过对判别器施加权重剪裁，限制函数梯度的变化范围，避免局部变化过大，从而缓解梯度消失问题。此外，Wasserstein 距离提供了一个更平滑的优化目标，有助于提升模型的稳定性和生成样本的质量。然而，WGAN 中权重剪裁的实现方式无法精确约束判别器函数的梯度范围，可能引发梯度消失或爆炸的风险，进而影响生成样本的最终质量。

为此，本次生成对抗模型采用梯度惩罚(GP)代替权重裁剪，以克服 WGAN 的缺点，进一步提升了 WGAN 的性能，其目标函数为：

$$\begin{aligned} L_{\text{WGAN-GP}} = \min_G \max_D V(D, G) = & \mathbb{E}_{z \sim P(z)} [D(G(z))] + \\ & \mathbb{E}_{x \sim P(x)} [D(x)] + \lambda \mathbb{E}_{\hat{x} \sim P(\hat{x})} \left[(\|\nabla_x D(\hat{x})\|_2 - 1)^2 \right] \end{aligned} \quad (3.16)$$

其中， λ 为梯度惩罚系数； ∇ 为梯度算子； \hat{x} 为真实数据与生成数据间的随机插值，为

$$\hat{x} = \zeta x + (1 - \zeta)G(z) \quad (3.17)$$

其中 ζ 服从 $[0,1]$ 内的均匀分布。

生成对抗模块的生成器如图 3-4 所示，生成器采用三次批归一化层和三次 4×4 卷积层以及 LeakyReLU 激活函数，并且在最后卷积层前引入注意力机制，可以相对好地将经过高斯函数处理的随机噪声和辅助信息 y 的特征向量拼接在一起的新向量进行处理和转换，其中注意力机制可以很好增强网络对图像信息的处理能力，Tanh 激活函数可以进行归一化。

生成对抗模块的鉴别器如图 3-5 所示，鉴别器采用三次批归一化层和三次 4×4 卷积层，LeakyReLU 激活函数以及两次最大池化层，同样在最后卷积层前引入注意力机制，最后，将卷积运算的结果与辅助信息 y 进行融合，接着通过全连接层进行处理，生成一个标量值，以反映输入图像属于真实图像的概率。

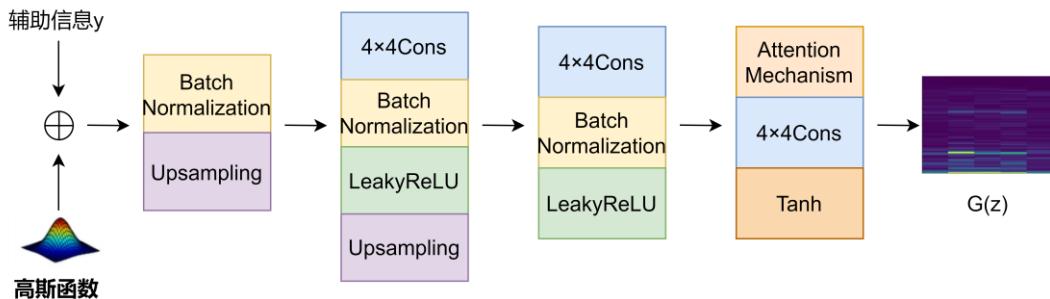


图 3-4 生成对抗模块的生成器

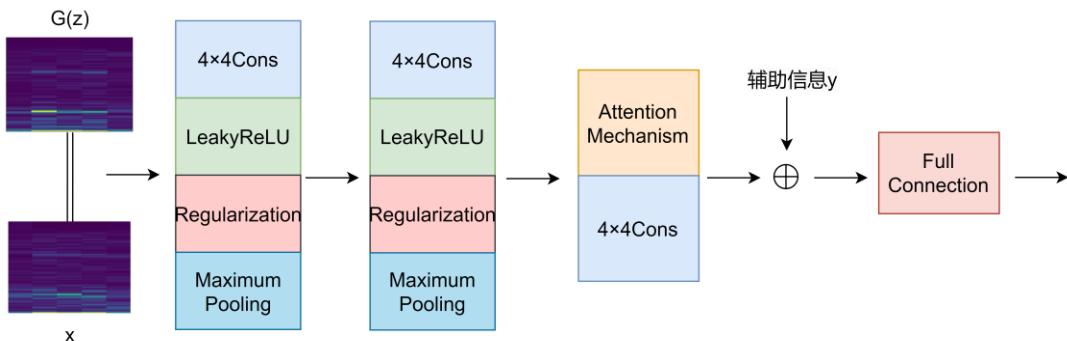


图 3-5 生成对抗模块的鉴别器

3.2.4 故障分类模块

在完成前两个模块的处理后，可利用扩展后的源域数据及其对应标签训练一个高精度的分类器。这一分类器能够在未见过的目标域中实现有效的预测，从而完成精确的故障诊断。

故障分类模块选用多层感知机(Multilayer Perceptron, MLP)作为分类器。MLP 是一种典型的前馈神经网络，由输入层、隐藏层和输出层组成，能够有效捕捉复杂的非线性关系，进而实现精准的分类判断。如图 3-6 所示，本研究构建的隐藏层包括三层全连接结构。设输入信号的长度为 I ，则前两层的节点数量分别定为 $I/4$ 和 $I/16$ ，最后一层的节点数与样本的类别数量相同。前两层全连接后选用 ReLU 函数作为激活函数，而最后一层全连接后则应用 Softmax 函数进行激活处理。

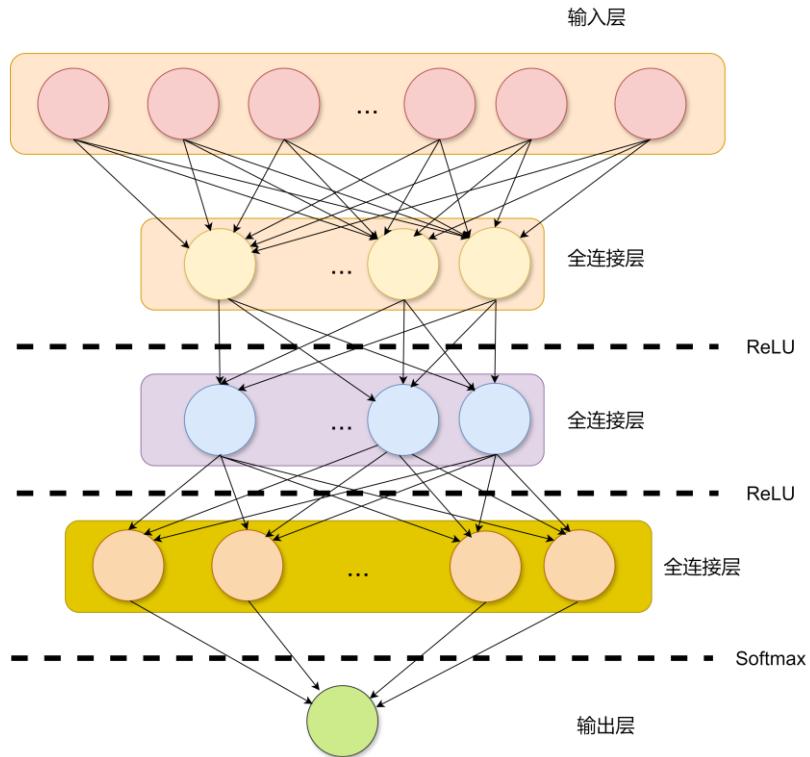


图 3-6 MLP 分类器网络结构

MLP 的前向传播过程可表达为：

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)} \quad (3.18)$$

$$a^{(l)} = \sigma(z^{(l)}) \quad (3.19)$$

其中： $a^{(l-1)}$ 是第 $l-1$ 层的输出， $a^{(l)}$ 是第 l 层的输出， $z^{(l)}$ 是第 l 层的线性变换结果。

Softmax 函数本质上是一个 Logistic 多分类器，其计算公式为：

$$a_i^{(L)} = \frac{e^{\tilde{z}_i^{(L)}}}{\sum_{j=1}^K e^{\tilde{z}_j^{(L)}}} \quad (3.20)$$

其中， K 表示类别数。

本研究采用交叉熵损失函数（Cross-Entropy Loss Function）作为损失函数，以衡量模型预测概率分布与真实标签分布之间的差异，该函数通常被认为是多分类任务中的一种常见评价指标。其数学公式表述为：

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (3.21)$$

其中， C 是类别的总数， $y_{i,c}$ 是一个指示器，如果样本 i 属于类 c 则为 1，否则为 0， $\hat{y}_{i,c}$ 是模型预测样本 i 属于类 c 的概率。当模型预测分布接近真实分布时，交叉熵损失值较小；反之，若分布差异较大，则损失值会显著增加。

3.3 轴承故障诊断泛化实验验证

3.3.1 数据集样本说明

为验证 MDWGANGP 框架在不同工况和转速下的故障诊断性能，本文选用第二章所述轴承故障诊断实验平台采集的数据集进行测试。如表 3-1 所示，该数据集包含五种故障情景（A 至 E），每种情景由特定的转速(rpm)和扭矩(N·m)参数界定。轴承状态列描述了具体的设备故障类型，而标签列则给出了故障的分类标记，分别为 N、IR、OR、B、C，依次表示正常、内圈故障、外圈故障、滚动体故障和综合故障。

表 3-1 数据集工况概述

数据集	转速(rpm)	扭矩(N · m)	轴承状态	标签
A	1800	1		
B	1800	3	正常/内圈故障/外	
C	1920	2	圈故障/滚动体故障	N/IR/OR/B/C
D	2400	2	/综合故障	

在数据集划分方面，本研究采用交叉验证(Cross-Validation)方法，将数据集按 70% 的训练集和 30% 的测试集比例分割，并取多次验证结果的平均值作为模型最终性能的评估依据。交叉验证通过反复训练和测试，能够充分利用数据资源，更准确地反映模型在未知数据上的表现能力。

本次评估采用的性能指标之一是准确率（Accuracy），该指标衡量模型正确预测的样本数与总样本数的比例。它能够直观地反映模型的整体预测能力，并有效评估其在未见目标域中的诊断表现。根据图 3-7 所示的划分，准确率的计算公式定义如下：

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.22)$$

	预测为正的类别	预测为负的类别
实际为正的类别	真正例(TP)	假负例(FN)
实际为负的类别	假正例(FP)	真负例(TN)

图 3-7 预测值与真实值的混淆矩阵

其中，TP、TN、FP、FN 含义如下：

TP(True Positive): 真正例，指模型正确识别为正类的样本数量；

TN(True Negative): 真反例，指模型正确识别为负类的样本数量；

FP(False Positive): 假正例，指模型错误预测为正类的样本数量；

FN(False Negative): 假反例，指模型错误预测为负类的样本数量。

模型设计统一基于 Python 3.11 和 PyTorch 2.5.1 框架，硬件配置包括：CPU 为 13th Gen Intel(R) Core(TM) i5-13400，内存容量为 16GB，GPU 型号为 NVIDIA GeForce RTX 3060。训练过程中，迭代次数设为 50 次，每批次样本量为 64，学习率定为 0.01。

3.3.2 对比实验

为全面评估 MDWGANGP 框架，设计了 12 组实验，基于 4 种工况的数据集，并与 5 种前沿领域泛化方法进行对比。

(1) Mixup

Mixup 算法是一种数据增强策略，通过对随机选取的两个训练样本及其标签进行线性插值，生成虚拟样本。这种方法简单有效，旨在通过平滑输入与标签间的关系，增强模型对领域偏移的鲁棒性^[18]。

(2) DIFEX

DIFEX 算法采用知识蒸馏框架，捕获高级傅里叶相作为内部不变特征，并通过跨域相关对齐学习互不变特征。同时，设计探索损失以增加特征多样性，从而提升泛化能力^[59]。

(4) VREx

VREx (Variance Risk Extrapolation) 算法针对域外泛化问题设计，通过对训练风险方差施加正则化约束，提升模型在未知测试域中的鲁棒性。该方法依托风险外推(Risk Extrapolation, REx)优化框架，相比其他泛化方法，在处理协变量偏移和概念偏移时表现出更强的稳定性^[41]。

(3) GroupDRO

Group DRO(Distributionally Robust Optimization)算法是由 Shiori Sagawa 等人提出，是一种分布式鲁棒优化方法，特别针对数据组间偏移问题。通过优化模型在不同数据组上的性能，Group DRO 在面对分布不均时表现出色^[42]。

(5) MLDG

MLDG(Meta-learning for Domain Generalization)算法是由 Li D 提出，是一种元学习方法，与传统专注于设计邻域转移鲁棒性模型的策略不同，MLDG 通过合成虚拟测试域模拟训练与测试域间的偏移。其元优化目标要求提升训练域性能的步骤同时改善测试域表现^[60]。

本研究采用“ $A \rightarrow C$ ”形式的跨域迁移任务表示法，其中源域(Source Domain)数据集 A 与目标域(Target Domain)数据集 C 组成迁移对。实验设计如下：训练阶段使用带标签的源域数据结合无标签的目标域训练集构成训练数据，而目标域测试集则用于评估模型性能。

在对比实验中，为确保实验可比性，采用以下方式：

- (1) 统一截取 2048 个采样点的时序信号作为模型输入；
- (2) 所有对比模型均在统一数据集上进行训练测试，通过控制变量法确保准确率差异仅反映模型迁移性能差异；
- (3) 为消除随机因素干扰，采用五折交叉验证并重复实验三次，最终取均值作为模型性能评估指标。

通过表 3-2 和图 3-8 的跨域诊断准确率对比可以发现，在 12 组迁移任务中 MDWGANGP 取得 8 项最优结果，展现出三类显著特征：

(1) 跨工况泛化优势显著：在转速和扭矩都差异较大的邻域泛化任务中 $C \rightarrow A$ 、 $A \rightarrow C$ ，MDWGANGP 表现尤为突出。以 $C \rightarrow A$ 任务为例，92.7% 的准确率较其它 5 种相对前沿方法中的最优方法 MLDG 提升 8.5 个百分点。

(2) 扭矩波动鲁棒性强：在转速相同，扭矩差距较大的邻域泛化迁移任务 $A \rightarrow B$ 和 $B \rightarrow A$ 中，各方法差距较小。例如 $B \rightarrow A$ 任务中 MDWGANGP (84.9%) 与 MLDG (85.2%) 基本持平，这说明当转速一致时，传统领域泛化方法已能满足基本需求，但本方法 MDWGANGP 仍保持了相当的竞争力。

表 3-2 不同跨域迁移任务下目标域诊断准确率 (%)

任务\模型	Mixup	DIFEX	GroupDRO	VREx	MLDG	MDWGANGP
A→B	79.6	84.1	82.3	85.4	86.7	88.1
A→C	77.2	81.5	78.9	83.6	84.9	91.8
A→D	73.8	79.2	75.4	80.1	82.3	86.4
B→A	78.4	83.9	81.6	84.7	85.2	84.9
B→C	75.9	80.6	77.1	82.3	83.8	85.0
B→D	72.1	78.3	74.8	79.5	81.6	81.4
C→A	76.8	82.4	79.5	83.1	84.2	92.7
C→B	74.5	79.8	76.3	81.7	82.9	82.2
C→D	70.3	77.1	73.9	78.4	80.1	82.5
D→A	69.8	85.4	72.6	76.9	78.3	81.6
D→B	67.5	73.8	70.2	75.2	76.8	82.0
D→C	65.3	71.5	68.9	73.6	74.9	81.9

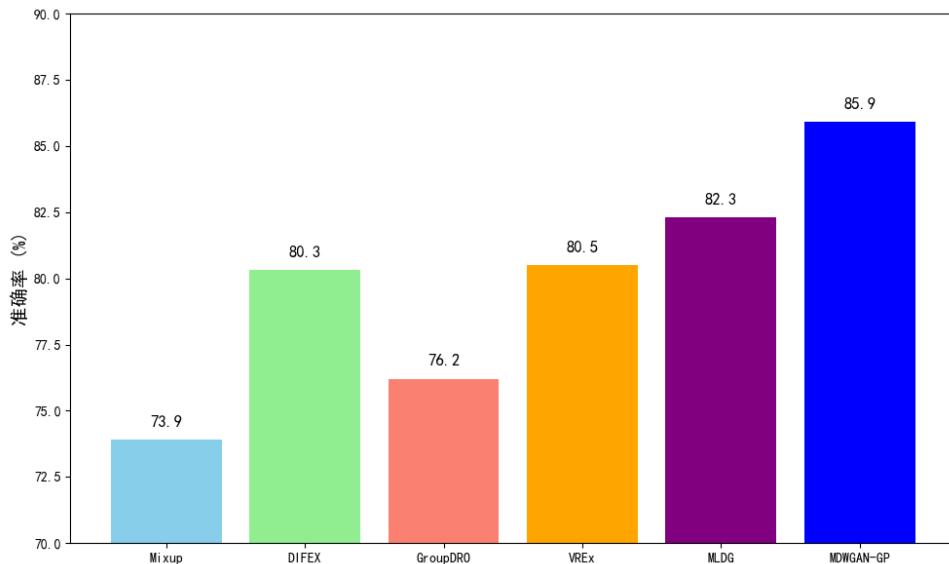


图 3-8 各领域泛化的方法在实验中的平均准确率

(3) 综合泛化能力最优：图 3-8 展示了不同方法在 12 个迁移任务中的平均准确率，可以看到，MDWGANGP 的平均准确率最高，达到 85.9%，高出其它 5 种相对前沿的方法至少 3.6 个百分点，这表明本章方法在此次对照实验中的性能显著优于其他方法，展现了 MDWGANGP 在综合泛化能力方面的优越性。

为了更好地对比分析不同方法在跨域故障诊断任务中的性能表现，本次选择 A→D 任务中的一次实验作为分析对象，统计在测试集的准确率随训练的变化情况，对比实验的迭代曲线如图所示，从图 3-9 中可以发现本章方法 MDWGANGP 收敛

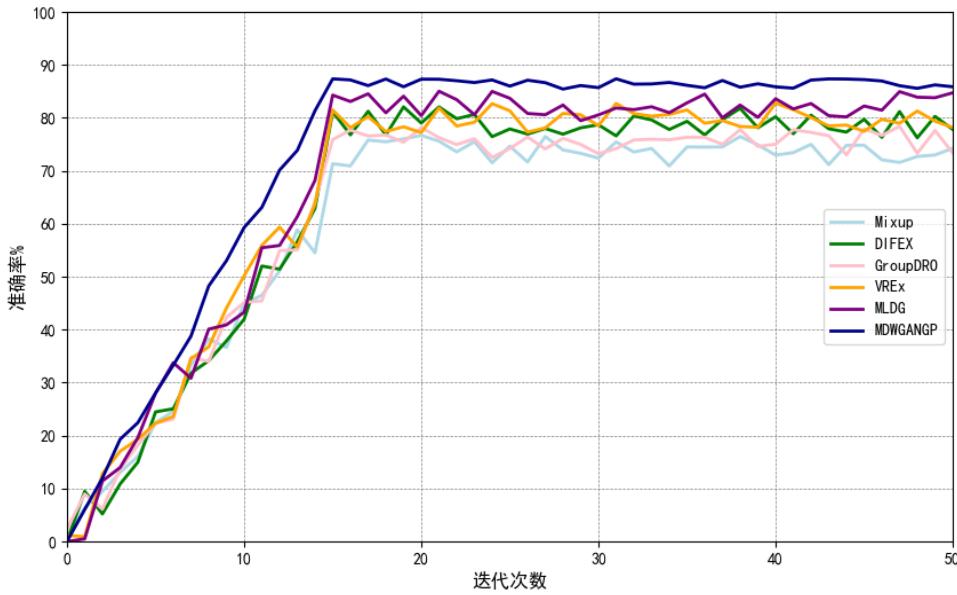


图 3-9 不同方法在 A→D 任务中的迭代曲线

较快，前 15 次迭代内快速提升至 85%，表明其多模态数据增强(WGAN-GP 生成)与特征融合机制(BiLSTM+CNN)能有效捕捉跨域不变特征。而在第 15 次迭代次数之后一直在 87% 左右稳定，准确率波动范围小于 $\pm 3\%$ ，得益于 WGAN-GP 的梯度惩罚机制，避免模式崩溃与训练震荡，显示 MDWGANGP 生成样本质量与特征对齐的鲁棒性。MDWGANGP 通过生成高质量伪目标域样本，与其它跨域任务平均准确率相比提升约 8%，验证了对抗生成在数据稀缺场景中的核心价值。

3.3.3 消融实验

为了证明本章方法 MDWGANGP 的各模块的有效性，进行了一系列消融实验，分为以下七种情况进行实验：

策略 1：本章提出的 MDWGANGP，命名为 D1；

策略 2：在 D1 的结构下，仅利用 BiLSTM 提取一维时序信号特征，而不通过 STFT 将数据转换为二维图像，命名为 D2；

策略 3：在 D1 的结构下，仅通过 STFT 将数据转换为二维图像，并利用 CNN 提取图像特征，而不利用 BiLSTM 提取一维时序信号特征，命名为 D3；

策略 4：在 D1 的结构下，不利用带有梯度惩罚的 Wasserstein 距离生成对抗网络(WGAN-GP)的数据生成对抗模块进行数据扩充，命名为 D4；

策略 5：将 D1 的结构下的 STFT 的方法改为 GAF；

策略 6：将 D1 的结构下的 STFT 的方法改为 MTF；

策略 7：将 D1 的结构下的 STFT 的方法改为 RP。

表 3-3 消融实验结果

模型配置	A→B	B→A	A→C	C→D	D→A	平均值	性能变化
D1	88.1	84.9	91.8	82.5	81.6	85.78	-
D2	78.3	77.4	81.3	75.4	70.7	76.62	▼9.16
D3	83.4	81.6	87.4	79.3	76.1	81.56	▼4.22
D4	74.5	72.1	80.3	73.8	72.3	74.60	▼11.18
D5	82.4	86.3	89.1	80.7	78.9	83.48	▼2.30
D6	80.2	83.7	88.5	78.9	77.4	81.74	▼4.04
D7	79.8	82.1	87.9	77.6	76.8	80.84	▼4.94

从上述表 3-3 的消融实验结果表明，本章提出的完整模型 D1 在跨域诊断任务中显著优于其他配置：

(1) 从 D1,D2,D3,D4 来看：D1 的平均准确率为 85.78%，较 D2（仅 BiLSTM）提高 9 个百分点，较 D3（仅 CNN）提高 4 个百分点，较 D4（无 WGAN-GP）提高 11 个百分点，可见通过 WGAN-GP 进行数据扩充是非常重要，可以很好地缓解数据量不足的问题，并增强模型对数据分布的鲁棒性，以此大幅度提升准确率。D2 与 D3 相比，可以看出通过 STFT 将数据转换为二维图像，并利用 CNN 提取图像特征比起仅利用 BiLSTM 提取一维时序信号特征可以更好的挖掘数据内部信号，从而提高准确率。

(2) 从 D1,D5,D6,D7 来看：D1 表现最佳，平均准确率为 85.78%，与 D5 (STFT→GAF) 的平均准确率差距最小，差距在 2 个百分点左右，与 D7(STFT→RP) 的平均准确率差距最大，差距在 5 个百分点左右，可能的原因是递归图 RP 对瞬态冲击波形的时延敏感性导致信息损失。总的来说，通过 STFT 转化为二维图像再进行 CNN 提取特征的方法比起其它三种方法更加能够将信号原本的信息挖掘出来，从而展现高鲁棒性和高准确性。

3.4 本章小结

本章提出了一种基于多模态数据增强的跨域轴承故障诊断方法 MDWGANGP，通过融合一维时序信号和二维图像特征并结合对抗生成技术进行数据扩充，这可以大幅提升了变工况场景下轴承故障诊断的跨域性能。针对传统方法因依赖单一信息域而导致特征提取不足及对分布偏移敏感的缺陷，本文创新性地引入了多模态特征融合机制（BiLSTM-CNN 动态加权）以及 WGAN-GP 对抗生成模块，有效缓解了源域数据不足和工况突变下的特征失配问题。实验结果表明，MDWGANGP 在 12 项跨域任务中的平均准确率达到 85.78%，比最优基线方法高出 3.6 个百分点，在动态工况（如 C→A 任务）中准确率更是突破 92.7%。消融实验进一步验证

了多模态融合和对抗生成技术的重要作用，同时显示 STFT 时频分析在瞬态特征提取上的明显优势（相较于 GAF/RP 提升 2.7% 至 5.0%）。MDWGANGP 为工业设备跨工况故障诊断提供了高鲁棒性的解决方案。相较于仅从单一信息域提取特征的诊断结果，该方法在轴承领域泛化故障诊断中展现出更高的精确性和可靠性，成功解决了单源域中泛化的数据稀缺问题，为后续诊断提供了可靠的训练数据支持。

第四章 基于多损失协同优化的跨域轴承故障诊断方法

4.1 引言

尽管上一章研究提出的基于多模态数据增强的跨域轴承故障诊断方法 MDWGANGP 在缓解源域数据稀缺性方面取得显著进展，但其仍面临两大核心瓶颈：其一，生成模型侧重于单源域数据的分布优化，而忽略跨工况场景下域间特征分布的非对齐性；其二，单一增强策略对复杂负载不平衡的鲁棒性不足。

为突破上一章 MDWGANGP 方法的局限性，本文在 MDWGANGP 原有的交叉熵损失的基础上，进一步提出了 HAF-MDWGANGP(Hierarchical Alignment Framework for Multi-Domain Wasserstein GAN-GP)，一种基于最大均值差异(MMD)损失、三元组损失和裕度差异损失联合优化的域间距离度量方法，以提升模型对领域偏移的适应性。通过协同优化这三种损失函数，HAF-MDWGANGP 能够提取更具普适性的域不变特征，从而大幅改善模型在未知目标域上的预测能力。

4.2 HAF-MDWGANGP 模型框架

如图 4-1 所示，为解决 MDWGANGP 方法在领域泛化过程中对领域偏移鲁棒性关注不足的缺陷，本章提出了 HAF-MDWGANGP。该框架在 MDWGANGP 的基础上新增了三种域间距离度量方式——最大均值差异(MMD)损失、三元组损失以及裕度差异损失，并将其与原有的交叉熵损失相结合，形成一个多层次的优化体系，以显著增强模型的跨域适应能力。

HAF-MDWGANGP 的设计理念是通过多维度的损失函数协同工作，全面提升模型在复杂工况下的泛化性能。具体而言：

- MMD 损失的作用在于减少源域与伪目标域之间的整体分布差异，从而实现全局特征对齐。通过量化并最小化两域特征空间的统计距离，该损失推动模型学习到跨域通用的特征表示。
- 三元组损失则聚焦于局部特征的对齐，通过拉近同类样本间的距离并推远异类样本间的距离，增强跨域同类型样本特征的相似性。这种机制有助于模型在不同领域中保持类内一致性。
- 裕度差异损失旨在优化分类边界，通过在不同域间保持合适的间隔，提升边界对分布扰动的适应性，从而减少模型因工况变化导致的性能波动。

最终，HAF-MDWGANGP 通过多任务损失的协同优化机制，将全局对齐、局部对齐和边界优化有效整合，有效解决了复杂分布偏移场景下模型跨域能力不足的问题。这一综合策略不仅弥补了 MDWGANGP 在适应性方面的不足，还为变工况轴承故障诊断提供了更加稳定和可靠的技术方案。

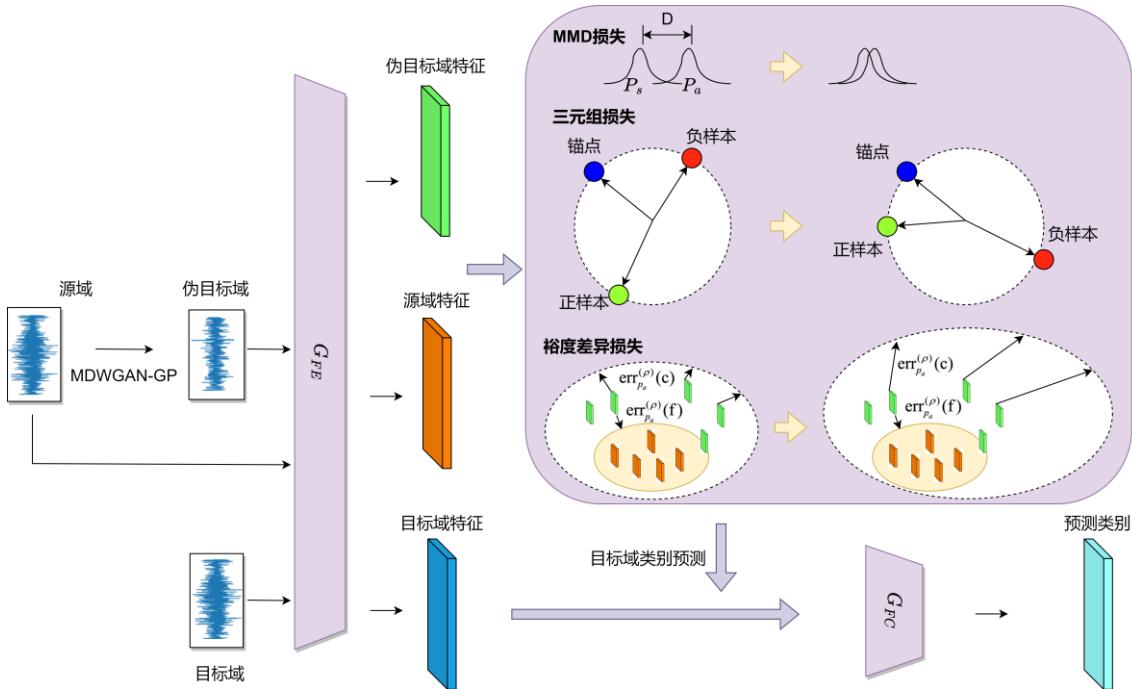


图 4-1 HAF-MDWGANGP 模型框架

4.2.1 MMD 损失

MMD 损失(Maximum Mean Discrepancy Loss) 是一种非参数化的评估手段，用于分析两个概率分布间的差异。其原理是通过核函数技术将数据投射至再生核希尔伯特空间 (RKHS)，并在该高维空间内计算两组分布的均值差距。MMD 本质上可视为一种分布间的距离度量，通过考察样本在特征空间中的统计特征，精确度量分布之间的不一致性。MMD 损失函数如下：

$$\mathcal{L}_{\text{MMD}} = \sum_{c=2}^C \left\| \frac{1}{n_c^s} \sum_{i=1}^{n_c^s} \phi(f_g(x_i^s)) - \frac{1}{n_c^m} \sum_{j=1}^{n_c^m} \phi(f_g(x_j^m)) \right\|^2 \quad (4.1)$$

其中， $\phi(\cdot)$ 为特征映射到希尔伯特空间的核函数； x_i^s 、 x_j^m 分别为源域样本和生成样本； C 为样本的类别。

在 MMD 损失体系中，核方法通过隐式变换技术，将数据从原始低维空间投影至高维特征空间，从而实现数据的线性可分，特别适用于解决复杂的非线性问题。

常用的核函数包括高斯核（RBF 核）、线性核和多项式核，这些技术各有特点，可适应不同的数据分布和分析场景。

高斯核通过指数函数将数据隐式映射到无限维的再生核希尔伯特空间(RKHS)，能捕捉复杂非线性模式，但对带宽 σ 敏感。高斯核适合处理低维复杂非线性数据。

高斯核公式如下：

$$k(x, y) = e^{\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)} \quad (4.2)$$

其中带宽 σ 控制局部敏感性，当 σ 过小时，MMD 值对噪声敏感、波动大。当 σ 过大时，所有样本相似度趋近 1，失去区分能力。

线性核对应的特征空间就是原始输入空间，没有进行非线性映射，只能捕捉数据中的线性关系，适用于高维稀疏数据、线性可分问题。线性核公式如下：

$$k(x, y) = x^T y = \sum_{i=1}^d x_i y_i \quad (4.3)$$

其中， x 和 y 是输入样本，维度为 d 。

多项式核通过引入多项式组合扩展特征空间，适合处理结构化数据、中等非线性问题，多项式核公式如下：

$$k(x, y) = (x^T y + c)^d \quad (4.4)$$

其中， d 为多项式次数（正整数），控制特征交叉的复杂度。 c 为常数项，通常 $c \geq 0$ ，调节低阶项的影响权重。

由于轴承故障信号往往具有非线性时频特征，不同工况（负载、转速）下分布差异显著，导致复杂性较高，为此，本次 MMD 损失函数采用的是多核高斯核函数，通过组合多个高斯核增强分布对齐的能力，能在轴承的跨领域泛化这一复杂场景中捕捉多样化的分布差异。公式如下：

$$k_{\text{multi}}(x, y) = \sum_{i=1}^K w_i e^{\left(-\frac{\|x-y\|^2}{2\sigma_i^2}\right)} \quad (4.5)$$

其中， $w_i \geq 0$ 为各核的权重，满足 $\sum w_i = 1$ 。所有核权重相等，为 $w_i = 1/K$ ，并且本次采用中位数启发式的方法，可以根据样本间距离的中位数动态调整各个高斯核的带宽 σ ，避免手动调参。

4.2.2 三元组损失

由于多源域中的同类别样本在映射分布上可能存在较大差异，为了提高所有域样本的聚类性能和类别可分性，本研究引入了三元组损失，三元组损失通过衡量

样本在欧式空间中的距离关系，缩小同类样本之间的距离并扩大异类样本之间的距离，从而间接提升类内紧凑性和类间分离度。该机制有效增强了模型在不同域内识别同类轴承故障特征的能力，并提升了跨域特征的相似性。

三元组损失的原理基于锚点样本、正样本和负样本的构造。锚点样本作为分类的基准，正样本与其属于同一类别，而负样本则与锚点分属不同类别。三元组损失的优化目标在于通过优化样本间的欧氏距离，尽可能缩短锚点与正样本的距离，同时拉大锚点与负样本的距离，并确保负样本对的距离始终超过正样本对一个预设的间隔阈值。

在欧式空间中，三元组损失的数学公式如下：

$$\mathcal{L}_{\text{tri}} = \sum_{i=1}^h \left[\|f_g(x_i^a) - f_g(x_i^p)\|^2 - \|f_g(x_i^a) - f_g(x_i^n)\|^2 + m \right] \quad (4.6)$$

其中，式中 $f_g(\cdot)$ 表示特征生成器，用于将样本映射到特征空间， m 为预设的边界裕度参数， h 表示训练集中选定的三元组数量， p 、 n 、 a 分别代表正样本、负样本和锚点样本。

为进一步提高模型的学习效率，本研究引入了批量困难采样（Batch Hard）策略。具体来说，在每批次数据中，为每个锚点样本选择距离最远的正样本和距离最近的负样本作为训练对象。该策略通过关注最难区分的样本对，增强模型对复杂样本的区分能力。批量困难采样的损失计算公式如下：

$$x^p = \arg \max_{x \in P} d(x^a, x) \quad (4.7)$$

$$x^n = \arg \min_{x \in N} d(x^a, x) \quad (4.8)$$

为平衡训练过程中的收敛速度与稳定性，本研究引入了自适应边距策略。该策略通过实时监测训练中难样本的比例，动态调整分类边界裕度参数。当难样本比例较高时，适当增加裕度以加速收敛；当难样本比例减少时，减小裕度以保障训练稳定性。这一动态调整机制使三元组损失能够根据训练进程自适应优化，避免因固定裕度引发的过快收敛或波动问题，从而进一步提升模型的学习性能。

在本次的跨域轴承故障诊断中，三元组损失在 HAF-MDWGANGP 框架中发挥了关键作用。它不仅促进了同类轴承故障特征在不同域内的聚类一致性，还通过类间分离度的提高了模型对分布偏移的适应能力。特别是在负载不均衡等复杂工况下，该损失策略能够有效降低特征分布的混乱性，为后续故障诊断提供更稳定的特征支持。

4.2.3 裕度差异损失

裕度是指样本点到分类边界的最短距离，其值直接反映了分类模型的鲁棒性和区分能力。裕度差异损失旨在保证不同域间分类间隔的统一性，尽管其主要目标是分布对齐，而非直接调整分类边界。具体来说，该损失通过增大伪样本与现有分类边界之间的距离，引导网络参数的后验分布向伪样本得分更高的方向优化。这种方式有助于增强模型在面对跨域扰动时的适应性和稳定性，从而更好地应对多个邻域之间的差异。

裕度差异损失的设计旨在通过约束伪样本与分类边界的关系，提升模型的跨域能力。它并不直接更改分类边界，而是通过优化模型参数，确保不同领域中的分类裕度保持一致。这种方法特别适用于变工况轴承故障诊断中因分布偏移引发的分布不一致问题。裕度差异损失的数学公式如下：

$$\mathcal{L}_{\text{mar}} = \frac{1}{J_c} \sum_{i=1}^{J_c} \left\| \varphi \left[\text{err}_{p_a}^{(\rho)}(c) \right] - \varphi \left[\text{err}_{p_a}^{(\rho)}(f) \right] \right\|_{\mathcal{H}}^2 \quad (4.9)$$

其中， $\text{err}_{p_a}^{(\rho)}(c)$ 、 $\text{err}_{p_a}^{(\rho)}(f)$ 分别表示与伪样本分布 p_a 相关的迁移分类模型裕度和预训练分类模型裕度； f 表示在源域上预训练的分类模型。

迁移分类模型裕度 $\text{err}_{p_a}^{(\rho)}(c)$ 是裕度差异损失的关键组成部分，其计算公式为：

$$\text{err}_{p_a}^{(\rho)}(c) = \frac{1}{n} \sum_{i=1}^n \Phi_{\rho} \left[\rho_c(x_{i,a}, y_{i,a}) \right] \quad (4.10)$$

其中， $\rho_c(x_a, y_a)$ 表示分类模型的决策边界，用于衡量样本与分类边界的距离关系。决策边界的具体计算方式为：

$$\rho_c(x_a, y_a) \triangleq \frac{1}{2} \left[c(x_a, y_a) - \max_{y_a' = y_a'} c(x_a, y_a') \right] \quad (4.11)$$

其中， y_a ， y_a' 分别表示分类模型对样本 x 输出的正确分类标签得分和错误分类标签得分。通过对比这两者的差值，决策边界能够反映样本被正确分类的难易程度。

为了进一步优化分类裕度，公式中引入了分段函数 $\Phi_{\rho}[\rho_c(x_a, y_a)]$ ，其作用是推动分类模型生成较大的正裕度值，确保正确类别的得分显著高于其他类别得分。其具体形式为：

$$\Phi_{\rho}[\rho_c(x_a, y_a)] \triangleq \begin{cases} 0, & \rho \leq \rho_c(x_a, y_a) \\ 1 - \frac{\rho_c(x_a, y_a)}{\rho}, & 0 \leq \rho_c(x_a, y_a) < \rho \\ 1, & \rho_c(x_a, y_a) < 0 \end{cases} \quad (4.12)$$

其中， ρ 表示裕度的下界参数，用于设定最小裕度要求。通过这一机制，分段函数限制了模型输出的裕度范围，避免过小的正裕度或负裕度，从而增强分类的可靠性。

本次采用的裕度差异损失，不仅实现了域间分类间隔的统一性，还间接增强了模型对伪样本的适应性。裕度差异损失能够有效应对复杂工况下的分布扰动。在本次的跨域轴承故障诊断中，裕度差异损失的计算需要平衡裕度下界 ρ 与模型训练的稳定性。若设置过大，可能会导致模型过于追求大裕度而影响收敛速度；若过小，则可能难以有效提升适应能力。

4.2.4 综合损失

在本次的跨域轴承故障诊断中，综合损失是将交叉熵损失作为主损失，其他损失作为正则化项，综合损失公式如下：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda_1 \mathcal{L}_{\text{MMD}} + \lambda_2 \mathcal{L}_{\text{tri}} + \lambda_3 \mathcal{L}_{\text{mar}} \quad (4.13)$$

其中， λ_1 表示 MMD 损失权重， λ_2 表示三元组损失权重， λ_3 表示裕度差异损失权重。MMD 损失权重 λ_1 能相对有效解决领域差异较大的任务中；三元组损失权重 λ_2 能相对有效解决类别重叠严重的任务中；裕度差异损失权重 λ_3 能相对有效解决领域扰动影响较大的任务。

4.2.5 训练流程

HAF-MDWGANGP 框架通过“数据增强-特征对齐-决策优化”三级协作机制，实现了跨域特征的高效整合与稳健分类。其运行流程划分为三个关键阶段，每一阶段均定义了明确的执行步骤和优化目标，以确保模型在多变工况下的泛化性能。以下是对该框架工作流程的详细解析：

- 阶段一：数据增强预训练

操作：在此阶段，仅激活 MDWGANGP 中的生成对抗模块，重点利用 WGAN-GP 的生成能力。

优化目标：通过最小化 WGAN-GP 损失（即公式 3.16），使生成样本的分布尽可能接近真实数据分布。这一过程旨在提升生成数据的质量和真实性，为后续训练提供多样化的样本支持。

输出：生成伪目标域样本 $\mathcal{D}'_t = \{G(z_i)\}_{i=1}^M$ ，并将其与原始样本集合 \mathcal{D}_s 混合形成增强后的混合数据集 $\mathcal{D}_{\text{train}}$

- 阶段二：多任务联合训练

操作：冻结生成器 G 参数，重点转向特征提取器和分类器的联合优化。此时，生成器不再更新，而是作为数据增强的固定模块支持后续训练。

优化目标：通过最小化综合损失函数 $\mathcal{L}_{\text{total}}$ （公式 4.13），同步更新特征提取器和分类器的参数。综合损失包括交叉熵损失、MMD 损失、三元组损失和裕度差异损失的加权组合，反向传播过程确保各模块参数协同优化。此外，每轮训练结束后，计算验证集上的 MMD 值，并采用早停策略以防止模型过拟合。

- 阶段三：对不可见目标域进行预测

操作：利用经过训练的网络对未见过的目标域样本进行预测。具体而言，输入目标域数据后，通过多层感知机(MLP)输出各类故障的概率分布，并据此计算最终的分类准确率。

输出：模型生成故障类别的概率分布，并基于最大概率原则确定预测类别，最终以准确率形式评估模型在目标域上的表现。

HAF-MDWGANGP 通过上述三阶段的逐步协作，完成了从数据生成到特征对齐再到决策优化的完整过程。数据增强阶段有效弥补了源域数据不足的局限，联合训练阶段通过多任务损失优化增强了特征的跨域一致性，预测阶段则验证了模型在未知目标域中的实用价值。这种三级协同策略不仅提升了模型的分类准确性，还显著增强了其在变工况下的适应能力，为工业设备故障诊断提供了高效的技术方案。

4.3 实验结果与分析

本次的实验设置仍沿用第三章的轴承数据集的 A/B/C/D 四工况，以及 $A \rightarrow B$ ， $A \rightarrow C$ ， $B \rightarrow A$ ， $C \rightarrow D$ ， $D \rightarrow A$ 这五组跨域任务，采用五折交叉验证，学习率设置为 0.001，批次大小设置为 64，Epoch 迭代为 50 轮。

4.3.1 对比实验

为全面评估 HAF-MDWGANGP 框架，本次引入 3 种在故障诊断中与本章方法相类似的前沿领域泛化方法进行对比，分别为 CCDG、RTDGN、DGNIS 方法。

CCDG 是一种用于提升故障诊断的条件对比域泛化方法，其设计了一种新的条件对比损失，旨在通过最大化同类别样本在不同域间的特征相似性，增强类内一致性，通过最小化不同类别样本的特征相关性，抑制类间干扰。CCDG 方法通过同时优化任务特定的交叉熵损失和条件对比损失来实现端到端的训练，提升模型在不同工作条件下的适应性和准确性^[61]。

RTDGN 是一种新颖的关系转移域泛化方法, 其包括任务无关领域适应 (TIDA) 和任务相关领域泛化 (TRDG) 两个阶段。在 TIDA 阶段, 构建多判别器对抗网络, 通过逆熵损失提升域混淆能力。在 TRDG 阶段, 设计残差融合分类器结构, 融合多层次特征, 并通过残差连接强化分类器的跨域判别能力^[62]。

DGNIS 是通过提取域不变特征并且保留域特定结构方法进行故障诊断, 其包括知识提取分支(KEB)和相似性指示分支(SIB)两个阶段。知识提取分支通过相关性对齐匹配多个源域之间的全局分布, 并引入三元组损失约束局部类内紧致性。相似性指示分支会估计每个目标测试样本与测试中所有可用源域的相似性, 最后合并具有相似性的多个预测分数而做出的^[63]。

表 4-1 对比实验结果

模型配置	A→B	A→C	B→A	C→D	D→A	平均准确率
CCDG	88.3	90.5	89.7	82.1	80.4	86.2
RTDGN	89.1	91.2	91.5	83.6	81.8	87.4
DGNIS	87.6	89.8	88.9	81.3	79.5	85.4
HAF-MDWGANGP	90.5	93.2	94.1	84.7	83.7	89.24

结果如表所示表 4-1, 可以看出 CCDG 虽然通过条件对比损失优化类间关系, 但缺乏对全局分布对齐的关注, 特别在 C→D 任务仅 82.1%, 可见在负载突变场景下泛化能力不足。RTDGN 尽管采用两阶段对抗训练, 但动态调整能力较弱, 无法充分应对多源域异构性。DGNIS 结合域不变特征与相似性指示分支, 但未显式优化决策边界, 特别在 D→A 任务仅为 79.5%, 是五组跨域任务中准确率最低。HAF-MDWGANGP 的平均准确率显著高于另外三种前沿领域方法, 特别在 B→A 任务中, HAF-MDWGANGP 准确率高达 94.1%, 验证了其多损失协同优化策略的有效性。HAF-MDWGANGP 通过多损失协同优化与层次化特征对齐策略, 显著提升了跨域轴承故障诊断的准确性和鲁棒性, 其性能全面超越现有 SOTA 方法, 验证了层次化特征对齐框架在跨域轴承故障诊断中的有效性实用性。

4.3.2 消融实验

为了分析各损失模块的贡献, 设计以下消融配置:

策略 1: 仅使用交叉熵损失的 MDWGANGP 作为基线, 命名为 C1;

策略 2: 在 C1 的结构下, 加上 MMD 损失, 命名为 C2;

策略 3: 在 C1 的结构下, 加上三元组损失, 命名为 C3;

策略 4：在 C1 的结构下，加上裕度差异损失，命名为 C4；

策略 5：将 C1 的结构下，加上 MMD 损失、三元组损失、裕度差异损失作为完整模型 HAF-MDWGANGP，命名为 C5。

表 4-2 消融实验结果

模型配置	A→B	A→C	B→A	C→D	D→A	平均准确率	性能变化
C1	88.1	91.8	84.9	81.9	81.6	85.78	-
C2	89.3	92.4	93.1	83.2	79.5	87.50	▲1.72
C3	88.7	92.1	92.9	82.5	76.6	86.55	▲0.77
C4	88.5	91.9	92.8	82.1	76.4	86.33	▲0.55
C5	90.5	93.2	94.1	84.7	83.7	89.24	▲3.46

从表 4-2 中可以得出有以下结论：

(1) MMD 损失贡献最大，单独引入 MMD 的 C2 能使平均准确率提升 1.72%，可以看出 MMD 主要优化域间分布对齐可以较好地提升平均准确率。

(2) 综合来看，单独引入三元组损失与单独引入裕度差异损失提升的程度较微弱，比基线模型 C1 提升不到 1 个百分点。

(3) 在 B→A 的任务下，C2,C3,C4,C5 比 C1 的准确率要高 7 个百分点到接近 10 个百分点，可见在转速相同时，扭矩不同的情况下，通过引入 MMD 损失、三元组损失、裕度差异损失可以大幅度提升领域泛化的预测准确性。

(4) 在 D→A 的任务下，C2,C3,C4 均比 C1 的准确率要低 2 个百分点到 5 个百分点，可见在差距较大、情况复杂的情况下，单纯引入 MMD 损失、三元组损失、裕度差异损失并不能提升领域泛化的预测准确性、反而会拉低准确率，而结合这三种损失反而比基线模型 C1 要高 2 个百分点，可能的原因是 MMD 损失约束全局分布对齐，三元组损失细化类内结构，裕度差异损失保障决策边界鲁棒性，三者形成“全局-局部-边界”三级优化体系，以此进行层次化对齐，联合优化使模型对领域差异大、噪声干扰强的情况容忍度得到提升。

(5) 通过完整模型 C5 性能显著优于其他单一模块，可以看出模块多损失协同优化的必要性。可以看出在领域差异大、噪声干扰强的工业场景中，需通过多目标协同优化而非单一损失增强，方能实现可靠的跨域泛化性能。

4.3.3 可视化实验

为更直观地展示 MDWGANGP 模型的优异性能，本研究采用了 t 分布随机邻域嵌入（t-SNE）技术，将提取的特征分布在二维空间中清晰呈现。t-SNE 是一种

广泛用于降维可视化的方法，由 Laurens van der Maaten 和 Geoffrey Hinton 于 2008 年提出的^[64]。该方法致力于将高维数据转换到二维或三维空间，尽量保持数据点之间的相对距离，使得相似点在降维后的空间内聚集，而不同点则被分开，以此清晰展示数据的内在模式和分布特征。

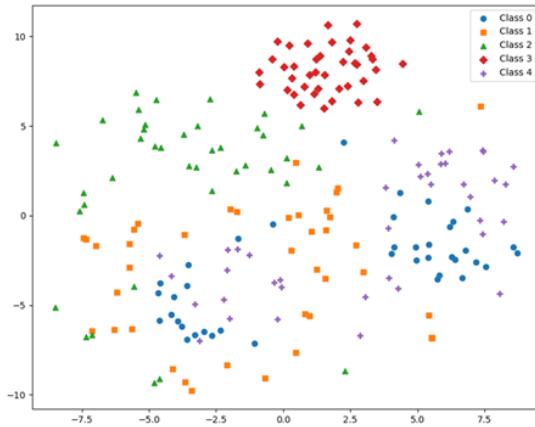


图 4-2 原始数据的 t-SNE 可视化结果

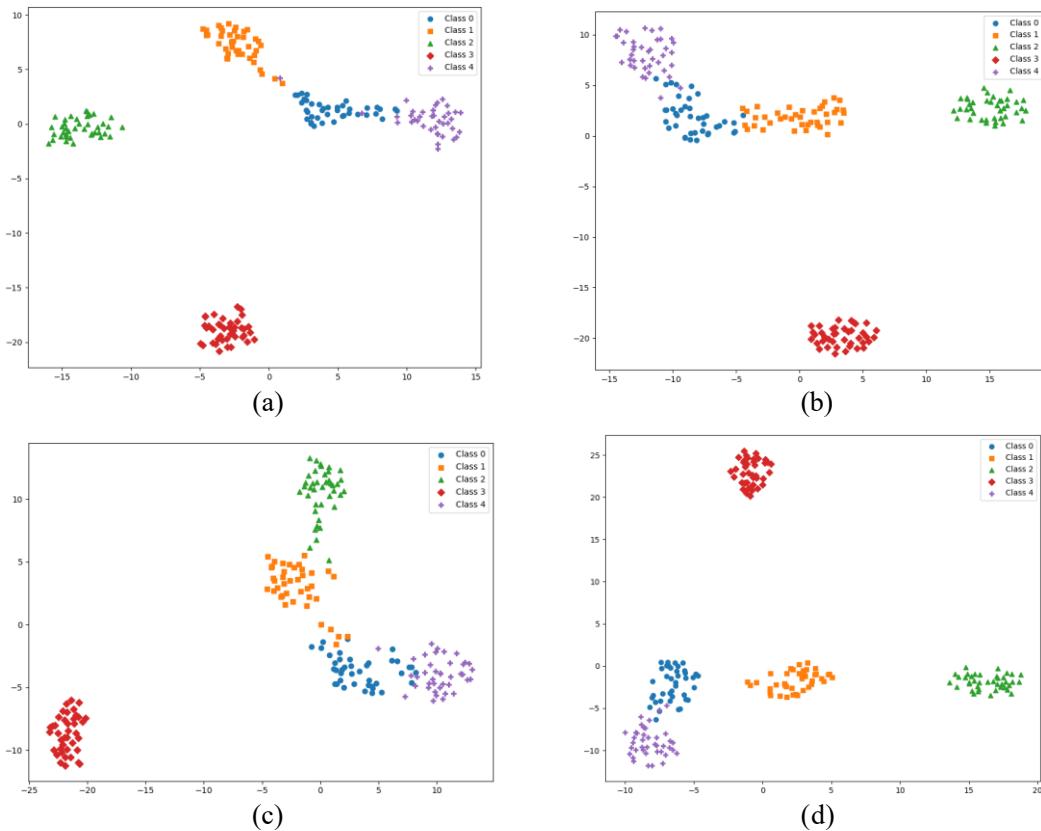


图 4-3 C2,C3,C4,C5 处理后的可视化结果。(a) C2; (b) C3; (c) C4; (d) C5

以 A→C 任务为例，将 C2,C3,C4,C5 输出的高维特征映射到二维上，可以在二维图像上展示数据分布图，图 4-2 为原始数据的 t-SNE 的目标域分布可视化结果，

图 4-3 分别表示 C2,C3,C4,C5 这四种领域泛化方法处理后的目标域分布可视化结果。目标域被聚类为五个簇，对应目标域样本的五种类别，正常状态用蓝色圆形表示，标记为 Class 0；内圈故障状态用橙色矩形表示，标记为 Class 1；外圈故障状态用红色菱形表示，标记为 Class 2；滚动体故障状态用绿色三角形表示，标记为 Class 3；综合故障状态用紫色十字架表示，标记为 Class 4。

从图 4-3 中的 t-SNE 可视化结果可以看出，本章的 HAF-MDWGANGP 方法不同类别的可视化特征比较明确，并且相互重叠比较少，仅仅正常状态的分布与综合故障状态分布有些许重叠，而从内圈故障状态和外圈故障状态以及滚动体故障状态可以看出分类的很好，综合来看类内距离小，类间距离大。此外，与 C2,C3,C4 相比，C5 能很好地对正常状态下与内圈故障状态的数据进行有效的分类。因此，C5 能够从源域学习到的足够的知识去对未知的目标域中的样本进行相对有效预测分类。总的来说，该 t-SNE 可视化结果进一步表明了 HAF-MDWGANGP 模型能够很好地提取源域中隐含的数据分布，以此更好地去预测未知的目标域的样本，达到强鲁棒性和高准确率的能力。

4.3.4 超参数实验

为了更好的展示此次的 MMD 权重 λ_1 、三元组权重 λ_2 、裕度差异权重 λ_3 对本次实验的影响，并以此确定权重 $(\lambda_1, \lambda_2, \lambda_3)$ 的最优组合。采用任务 A→C 任务，通过固定交叉熵权重为 1，对 λ_1 、 λ_2 、 λ_3 进行网格搜索，其中各自的范围如下 $\lambda_1 \in \{0.1, 0.5, 1.0\}$ ， $\lambda_2 \in \{0.05, 0.1, 0.2\}$ ， $\lambda_3 \in \{0.01, 0.05, 0.1\}$ ，实验结果如表 4-3 所示。

表 4-3 各权重组的结果

	λ_1	λ_2	λ_3	平均准确率
组合 1	0.5	0.1	0.05	93.2
组合 2	1.0	0.05	0.01	92.7
组合 3	0.5	0.2	0.1	92.4
组合 4	0.1	0.1	0.05	91.8

从表 4-3 可以看出本次最优配置为组合 1，即当 $\lambda_1=0.5$ ， $\lambda_2=0.1$ ， $\lambda_3=0.05$ ，平均准确率最高，可以达到 93.2%，此时模型在动态任务中平衡了领域对齐与类可分性，比起组合 4 可以提高 1.4%，其成功源于三者的动态平衡：1、MMD 权重主导全局对齐，适中的 $\lambda_1=0.5$ 既能有效约束域间分布差异，又避免过度抑制分类判别性特征；2、三元组权重优化类内结构，当 $\lambda_2=0.1$ 使同类样本在潜在空间的平均距

离缩减，可以显著增强特征紧凑性；3、裕度差异的边界调控，较低的 $\lambda_3=0.05$ ，在保证类间可分性的同时，可以很好缓解了决策边界过拟合问题。

4.4 本章小结

本章针对 MDWGANGP 在复杂领域偏移中鲁棒性不足的问题，提出了一种基于多损失协同优化的跨域轴承故障诊断方法 HAF-MDWGANGP，通过在 MDWGANGP 的交叉熵损失基础上引入其他损失函数，通过交叉熵损失保持分类判别性，最大均值差异损失对齐全局分布，三元组损失约束类内紧凑性，裕度差异损失增强决策边界可分性，这可以显著提升模型对复杂领域偏移的鲁棒性。实验表明，该方法在复杂工况下的平均准确率达 89.24%，较基线模型提升 3.46%，并且当权重设置为 $\lambda_1=0.5$, $\lambda_2=0.1$, $\lambda_3=0.05$ 时候，可达最优。从结果可以看出，本章提出的 HAF-MDWGANGP 框架通过多损失协同优化，实现了从数据增强到决策边界的全链路领域泛化能力提升，有效的解决了单源域泛化时模型对领域偏移、扰动差导致表现不佳的问题，为以后的工业设备智能运维提供了高鲁棒性解决方案。

第五章 基于自适应多源集成的跨域轴承故障诊断方法

5.1 引言

在第四章中，我们开发了 HAF-MDWGANGP 框架，通过整合最大均值差异 (MMD) 损失、三元组损失及裕度差异损失等分布间度量方法，显著提高了模型对分布偏移的适应能力。实验结果表明，HAF-MDWGANGP 结合多模态数据增强技术 (MDWGANGP) 及多损失协同优化策略，不仅有效解决了源域数据不足的挑战，还在未知目标域中实现了较高的故障诊断精度。该框架通过多层次对齐机制，成功增强了模型的跨域适应性，为变工况轴承故障诊断提供了稳健支持。

然而，随着工业环境中多源域数据的复杂性不断加深，现有技术仍面临以下挑战：(1)不同源域间数据分布差异显著，单一模型难以有效挖掘多源域间的互补性信息；(2)传统领域泛化方法通常依赖固定的特征对齐策略，难以灵活适应多源域的异质特性；(3)在高度动态的变工况场景下，模型对多源域特征融合的效率与稳定性仍需进一步改进。

针对这些问题，提出了一种自适应多源集成领域泛化方法 AMSEDG (Adaptive Multi-Source Ensemble Domain Generalization)，通过多模型协同及动态权重调整，旨在提升复杂工况下故障诊断的跨域能力。本章与第三章及第四章构成了逐步推进的技术闭环：第三章的 MDWGANGP 方法通过多模态数据增强克服了源域样本不足的限制；第四章加入基于多损失协同优化的分布间距离度量方法可以有效提升了模型对分布偏移的适应性；而本章的 AMSEDG 框架进一步整合多源域信息，通过多层次不变特征学习，包括域级与样本级这两个层次，同时采用基于动态相似性的加权集成策略，实现了从单一域增强向多源域协同泛化的跨越。

5.2 AMSEDG 模型框架

AMSEDG 方法的核心目标是通过多层次的不变特征学习和加权集成策略，解决多源域异构环境下旋转机械故障诊断中的领域泛化难题。如图 5-1 和图 5-2 所示，其工作流程分为训练阶段和测试阶段，并依托特定的网络架构和损失函数驱动。为充分发挥每个目标样本与不同源域间潜在相似性的差异，AMSEDG 为每种源域分别训练一个独立的分类器。在测试阶段，该方法根据目标样本与各源域的相似度，动态调整并整合所有分类器的预测结果，从而增强模型的跨域适应能力，实现高效的故障诊断。

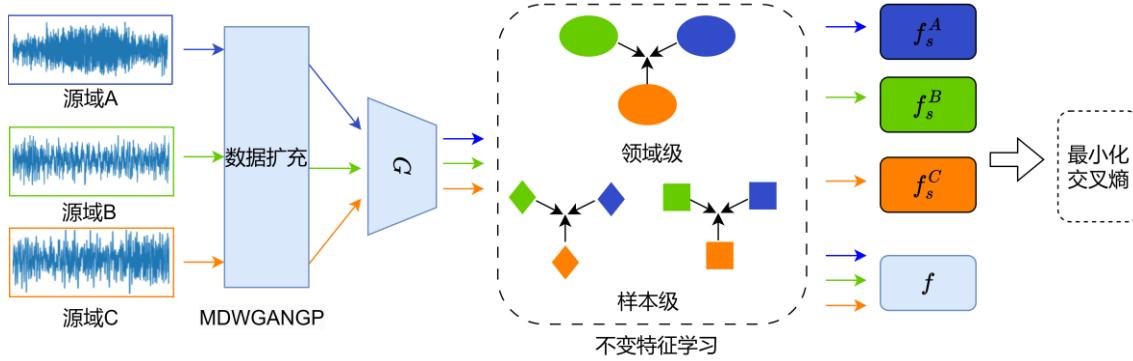


图 5-1 训练流程

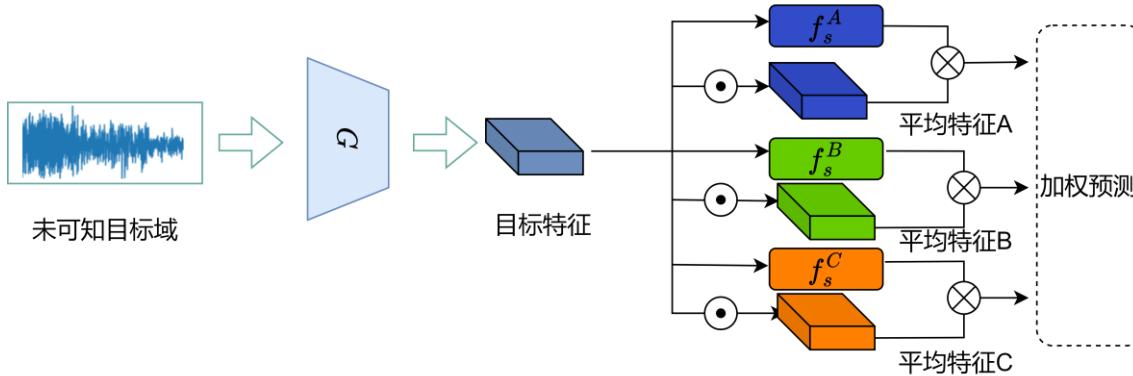


图 5-2 测试流程

AMSEDG 模型的网络结构和具体参数如图 5-3 和图 5-4 所示。该网络包括一个特征提取器 G 、一个域共享分类器 f 和特定于域的分类器 f_s^h , $h=1,\dots,H$ 。特征提取器由两个卷积单元组成，每个卷积单元包含一个卷积层、非线性激活单元 ReLU 和池化层。两种分类器的结构相同，由两个全连接层组成，并且在第一个全连接层之后连接非线性激活单元 ReLU 和 Dropout 层。

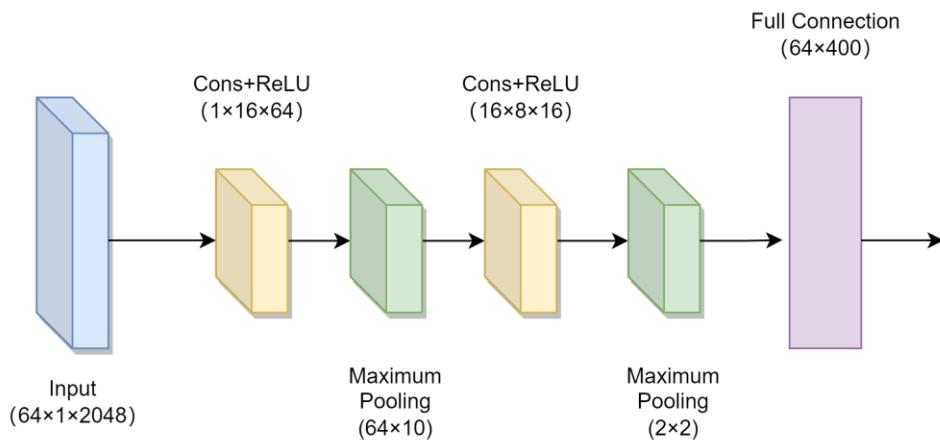


图 5-3 特征提取器的网络结构

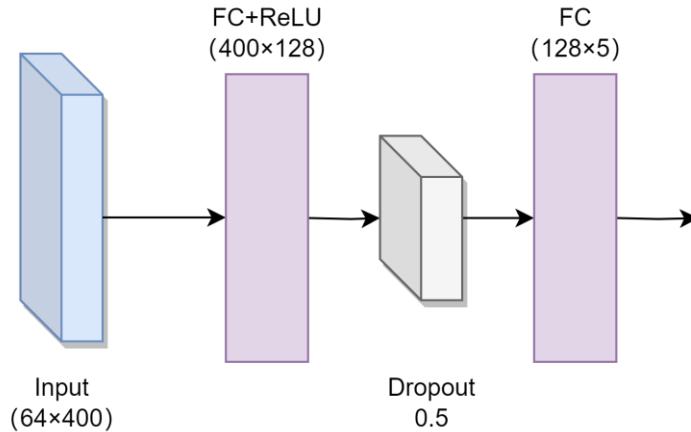


图 5-4 分类器的网络结构

5.2.1 多层次不变特征学习

本次的多层次不变特征学习方法分为领域级别域不变特征学习和样本级别域不变特征学习。引入领域级别域不变特征学习是为了适应不同源域的整体数据分布。引入了样本级别域不变特征学习是为了深入学习跨域的不同类别数据的域不变特征。

5.2.1.1 领域级别域不变特征学习

当将在一个领域上训练的模型应用于其他领域时，跨域数据分布的差异会导致性能下降。特征分为领域不变特征和特定领域特征类型。分布差异源于特定领域特征类型，影响模型表示学习并导致数据在特征空间上的映射不同，从而导致性能损失。因此。需要引入领域级别的域不变特征学习以此更好的学习具有类可区分性的领域不变特征，减少分布差异，并调整数据分布从而提高模型泛化能力。

本次的领域级别域不变特征学习采用相关性对齐方法(Correlation Alignment)，并结合 MMD 损失 (4.1)，以此来优化源域间分布对齐，其损失函数如下：

$$\mathcal{L}_{\text{domain}} = \frac{1}{4d^2 M(M-1)} \sum_{m_1=1}^M \sum_{m_2=m_1+1}^M \left\| \text{Corr}_S^{m_1} - \text{Corr}_S^{m_2} \right\|_F^2 + \lambda_{\text{MMD}} \cdot \mathcal{L}_{\text{MMD}} \quad (4.14)$$

其中， d 代表特征空间的维数， Corr_S 是源域的协方差矩阵，其表示为：

$$\text{Corr}_S^m = \frac{1}{n_S^m - 1} \left[D_S^{mT} D_S^m - \frac{1}{n_S^m} (\mathbf{1}^T D_S^m)^T (\mathbf{1}^T D_S^m) \right] \quad (4.15)$$

其中， $\mathbf{1}$ 是一个列向量，所有元素均为 1。为确保域不变特征具有类别可区分性，需要根据真实标签进行监督训练。损失函数可以表示为：

$$\mathcal{L}_{en} = - \left(\frac{1}{N_S} \sum_{i=1}^{N_S} y_S^i \log f(g(x_S^i)) + \sum_{m=1}^M \frac{1}{n_S^m} \sum_{m=1}^{n_S^m} y_{S_m}^i \log f(g(x_{S_m}^i)) \right) \quad (4.16)$$

其中， N_S 可以表示为：

$$N_S = \sum_{m=1}^M n_S^m \quad (4.17)$$

5.2.1.2 样本级别域不变特征学习

在领域级域不变特征学习之后，该模型将不同领域数据分布的中心对齐，从而减少差异。然而，不同类别样本的域不变特征提取各不相同，不同类别的样本可能会产生干扰。因此，需要引入了相应的样本级域不变特征学习，本次的样本级域不变特征学习则结合第四章的三元组损失和裕度差异损失，优化细粒度跨域特征，其损失函数可以表示为：

$$\mathcal{L}_{sample} = \lambda_{tri} \mathcal{L}_{tri} + \lambda_{mar} \mathcal{L}_{mar} \quad (4.18)$$

最后，多层次不变特征学习方法的损失函数可以表示为：

$$\mathcal{L}_m = \mathcal{L}_{domain} + \mathcal{L}_{en} + \mathcal{L}_{sample} \quad (4.19)$$

5.2.2 基于动态相似性的加权集成

考虑到目标数据与不同源域之间的相似性不同，AMSEDG 模型为每个源域分别训练额外的分类器，并根据目标数据与源域的相似性整合所有分类器的预测结果，以增强模型的最终泛化能力。在最终测试预测过程中，需要将更多的权重分配给与目标样本相似度更高的领域信息，因此为每个域训练单有的分类器，以学习特定域的信息。在测试过程中，目标样本的特征向量与每个源域样本的中心点特征向量之间的相似度决定了整合来自特定域分类器预测结果的权重。

目标样本的权重不仅基于初始特征相似性，还通过在线滑动窗口计算其与各源域样本的局部相似性。定义动态相似性为：

$$\lambda_m^d = \gamma \cdot \lambda_m + (1-\gamma) \cdot \frac{1}{L} \sum_{k=1}^L g(x_r^{(k)})^T g(x_{S_m}^i) \quad (4.20)$$

其中， $x_r^{(k)}$ 为目标样本的最近 L 个特征向量， $\gamma \in [0,1]$ 为平衡初始相似性与动态更新的超参数。

此外还在特征空间中引入多头注意力模块，捕捉目标样本与各源域样本间的细粒度交互。注意力权重计算为：

$$\alpha_{m,i} = \text{softmax}\left(\frac{W_q g(x_T) \cdot W_k g(x_{S_m}^i)}{\sqrt{d}}\right) \quad (4.21)$$

其中， W_q ， W_k 为可学习参数， d 为特征维度。

之后，构建 λ_m ，表示为第 m 种领域的相似度，公式如下：

$$\lambda_m = \lambda_m^d + \beta \cdot \sum_{i=1}^{n_S^m} \alpha_{m,i} \quad (4.22)$$

最终，基于具有预测值 \hat{y}_{Tm} 的领域特定分类器 f_S^m ，对目标样本 x_T 的预测可表示为：

$$\hat{y}_T = \sum_{m=1}^M \lambda_m \hat{y}_{Tm} \quad (4.23)$$

5.3 实验结果与分析

本次的实验设置仍然沿用第三和第四章的轴承数据集的A/B/C/D四工况，采用五折交叉验证，学习率设置为0.001，批次大小设置为64，Epoch迭代设置为50轮。

总共设置四个多源域迁移任务，任务名称如下表5-1所示分别为T1,T2,T3,T4，任务的结果为五次实验中最高的平均准确率。

表 5-1 迁移任务设计

任务名称	源域	目标域
T1	A、B、C	D
T2	A、B、D	C
T3	A、C、D	B
T4	C、B、D	A

5.3.1 对比实验

本次对比实验除了选择第三章对比实验中效果较好的前四种方法：DIFEX、VREx、MLDG、MDWGANGP和我们第四章的HAF-MDWGANGP，还选取了DANN以及AMINet、ANDmask。

DANN(Domain-Adversarial Neural Network)是一种基于对抗训练的领域泛化模型，旨在提取对域变化鲁棒的特征表示。该方法通过将特征提取器与域分类器巧妙

结合实现其目标。具体而言，DANN 设计了一个域分类器，其任务是识别特征向量所属的源域。为了使特征提取器生成域无关的特征，DANN 引入了梯度反转层（Gradient Reversal Layer），通过该层将域分类器的梯度信号反向传递至特征提取器。这种机制激励特征提取器生成能够混淆域分类器的特征，即削弱特征中与特定域相关的信息含量。通过这种对抗训练方式，DANN 成功学习到对领域偏移具有较强适应性的特征表示，从而显著提升模型在跨域任务中的泛化能力^[27]。

AMINet 是一种专为机械故障诊断开发的单域泛化网络，通过对抗性互信息引导技术增强模型的跨域适应性。该网络创新性地加入了域生成模块，生成与源域分布显著不同的伪目标域数据，进一步扩展了训练数据的多样性。在此基础上，AMINet 通过在域生成模块与任务诊断模块间建立迭代的最小-最大互信息博弈机制，有效提取鲁棒的通用特征，从而成功应对未知分布偏移的挑战。基于两台机械钻机的大规模实验验证，该方法的优异性能和实用价值得到了充分证明^[65]。

ANDmask 是一种基于梯度方向一致性验证的域泛化技术（Out-of-Distribution Generalization），其目标是通过筛选跨域梯度方向一致的参数更新，驱动模型学习具备跨域不变性的特征表示。该技术的核心在于运用逻辑“与”（AND）操作生成二进制掩码，仅在多个域的梯度方向高度一致时允许参数调整，从而降低模型对特定域特征的依赖，提升其在未知域中的泛化能力^[66]。

本次实验通过四个多源域迁移任务(T1、T2、T3、T4)对所提出的 AMSEDG 方法与八种对比方法(DIFEX、VREx、MLDG、DANN、AMINet、ANDmask、MDWGANGP 和 HAF-MDWGANGP)进行了性能评估。实验结果如表 5-2 所示和如图所示，所有方法的诊断准确率均基于五折交叉验证中最高的平均值，反映了各方法在轴承数据集 A、B、C、D 四种工况下的跨域泛化能力。以下从整体性能、任务间稳定性以及与其他方法的对比三个方面进行分析，以验证 AMSEDG 的优越性。

表 5-2 各方法在不同迁移任务实验中诊断准确率 (%)

方法	T1	T2	T3	T4	平均值
DIFEX	85.3	86.7	84.9	83.5	85.1
VREx	87.6	88.2	86.4	85.8	87.0
MLDG	89.1	90.5	88.7	87.3	89.2
DANN	88.5	89.0	87.1	86.4	87.8
AMINet	90.7	91.3	89.5	88.9	90.1
ANDmask	86.9	87.5	85.8	84.7	86.2
MDWGANGP	91.4	92.8	90.2	89.6	91.0
HAF-MDWGANGP	93.2	94.1	92.5	91.8	92.9
AMSEDG	95.6	96.2	94.8	93.9	95.1

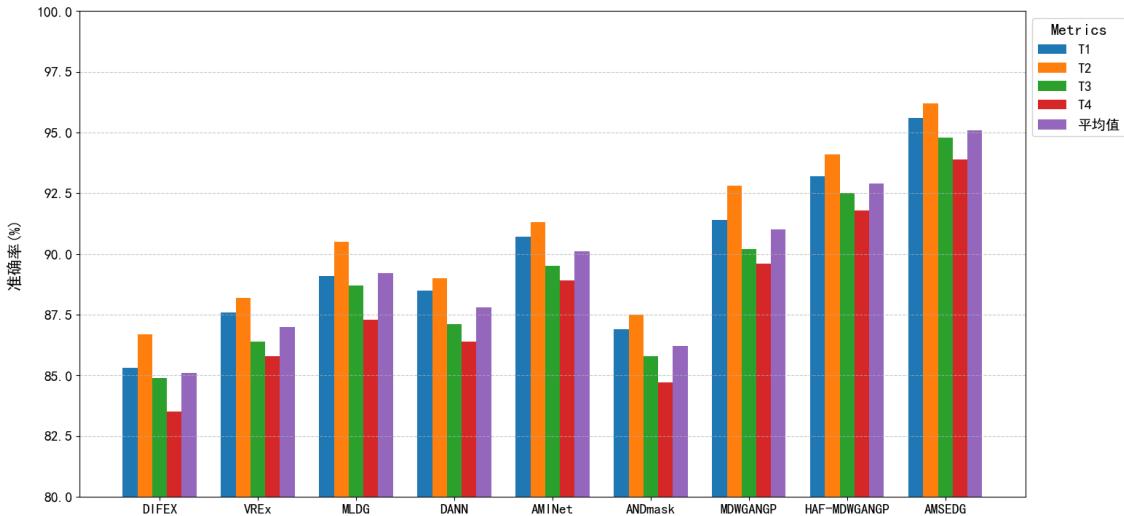


图 5-5 各方法在不同迁移任务实验中的准确率

从表 5-2 和图 5-5 来看，AMSEDG 以 95.1% 的平均诊断准确率明显优于所有对比方法，领先第二高的 HAF-MDWGAN (92.9%) 2.2 个百分点，相比表现最差的 DIFEX (85.1%) 提高了整整 10 个百分点。这一结果显示，AMSEDG 在多源域迁移任务中展现出卓越的跨域能力，有效应对了不同工况间分布偏移的挑战。相较之下，其他方法的平均准确率分布在 85.1% 至 92.9% 之间，凸显了 AMSEDG 在整体表现上的显著优势。

DIFEX、VREx 和 MLDG 是第三章中表现较优的传统域泛化方法，其平均准确率分别为 85.1%、87.0% 和 89.2%。然而，与 AMSEDG 相比，这三者在所有任务中的表现均较低，尤其是 DIFEX 在 T4 任务中仅达到 83.5%，远低于 AMSEDG 的 93.9%。这表明传统方法在处理多源域迁移任务时，可能无法充分捕捉跨域不变特征，而 AMSEDG 却可以较好的提取域不变特征可以显著提升了泛化性能。

DANN 和 AMINet 通过对抗性训练策略学习域不变特征，其平均准确率分别为 87.8% 和 90.1%。尽管 AMINet 在 T2 任务中达到 91.3%，表现较为优异，但其整体表现仍不及 AMSEDG，且在 T4 任务中降至 88.9%。DANN 的表现相对稳定但偏低，平均值仅为 87.8%。这表明，尽管对抗性训练在一定程度上提升了泛化能力，AMSEDG 凭借独特的多层次不变特征学习及基于动态相似性的预测加权，仍超越了这些方法。

MDWGANGP 和 HAF-MDWGANGP 是第三章和第四章提出的方法，在多源域的表现上，其平均准确率分别为 91.0% 和 92.9%，在对比方法中表现较为靠前。然而，AMSEDG 在所有任务中均超越这两者，这表明 AMSEDG 在继承前两章的

优点的基础上，进一步提出了独特的多层次不变特征学习和基于动态相似性的加权预测，从而提升预测准确性。

实验结果充分证明了 AMSEDG 在轴承故障诊断多源域领域泛化中的优势，为其在实际跨工况场景中的应用提供了强有力的支持。

为了深入验证 AMSEDG 模型在多源迁移任务中的泛化能力与鲁棒性，本研究通过四个迁移任务 T1-T4 的混淆矩阵对其分类性能进行全面分析。在其中一次测试中，AMSEDG 在四个多源迁移任务上的混淆矩阵如图 5-6 所示，其中正常状态轴承标记为 N、内圈故障轴承标记为 IR、外圈故障轴承标记为 OR、滚动体故障轴承标记为 B、综合故障轴承标记为 C。由图 5-6 可以看到，在四个迁移任务中，AMSEDG 在 T1 到 T4 任务中，对于五种轴承状态的预测准确率都比较高，几乎每个类型的轴承的正确预测数量 180 以上。四个任务中总共涉及到 4000 个测试样本的故障诊断，只有 246 个样本被错误分类，可见 AMSEDG 模型在不同的复杂多源域数据环境下的强鲁棒性和多源域泛化能力。

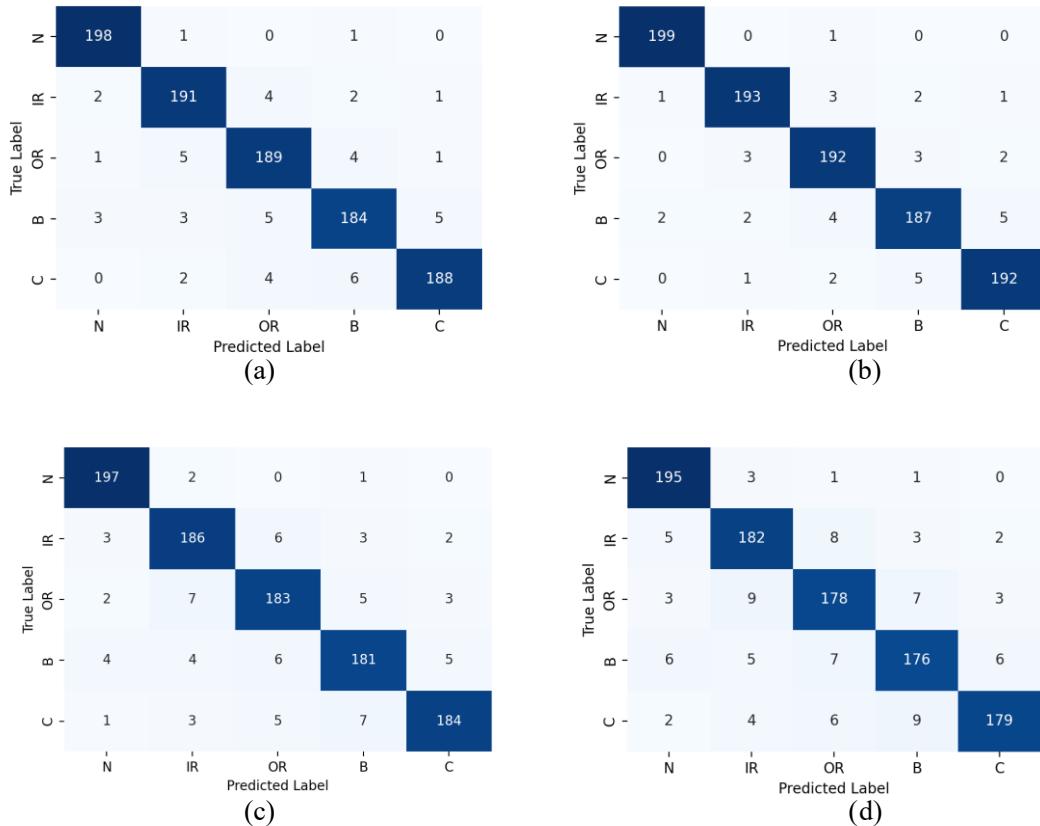


图 5-6 AMSEDG 在四个迁移任务中的混淆矩阵。(a)T1; (b)T2; (c)T3; (d)T4

为更清晰地比较不同方法在多源域跨域故障诊断任务中的性能，本研究选取 T1 任务中的一次实验作为研究对象，分析测试集准确率随训练过程的变化趋势，

并绘制了对比实验的迭代曲线，如图所示。从图 5-7 可见，本章提出的 AMSEDG 模型、第三章的 MDWGANGP 模型及第四章的 HAF-MDWGANGP 模型收敛速度较快，仅需 12 次迭代即可迅速达到 90% 的准确率。在第 12 次迭代次数之后 AMSEDG 模型一直在 95% 左右稳定，准确率波动范围小于 2%，波动较小，这显示了 AMSEDG 模型在复杂多源域数据环境下的稳定性和鲁棒性。AMSEDG 模型比 MDWGANGP 模型、HAF-MDWGANGP 模型平均准确率相比提升约 3%，与其它领域泛化的模型准确率相比提升约 8%。

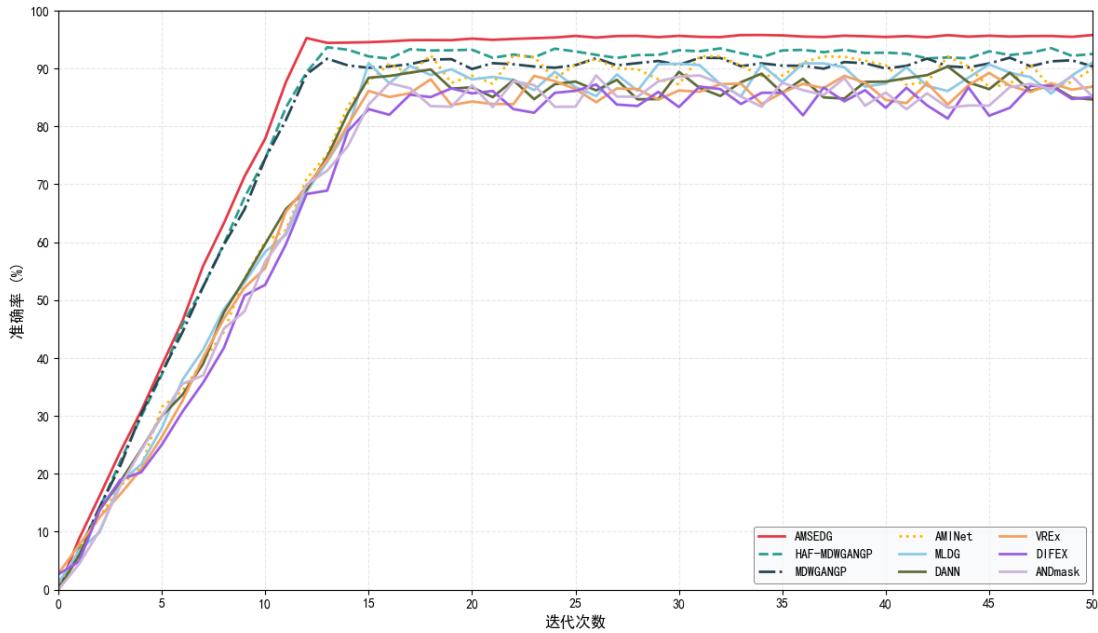


图 5-7 不同方法在 T1 任务中的迭代曲线

总的来说，AMSEDG 模型通过多层次不变特征学习及动态相似性加权融合策略，有效增强了模型对多源域数据的适应性和跨域能力。实验结果证实了 AMSEDG 模型的高效性和优越性，为复杂工况下的故障诊断提供了实用方案。

5.3.2 消融实验

为了证明本章方法 AMSEDG 框架中关键组件的有效性，本研究进行了一系列消融实验，分为以下五种情况进行实验：

策略 1：本章提出的 AMSEDG，命名为 B1；

策略 2：在 B1 的结构下，将动态相似性加权改为静态相似性加权，命名为 B2；

策略 3：在 B1 的结构下，去掉多头注意力模块，命名为 B3；

策略 4：在 B1 的结构下，不采用样本级域不变特征学习只采用领域级域不变特征学习，命名为 B4；

策略 5：将 B1 的结构下，不采用领域级域不变特征学习只采用样本级域不变特征学习，命名为 B5。

如表 5-3 所示，当将动态相似性权重替换为静态相似性权重时(B2)，平均准确率下降 2.4%。这表明在线滑动窗口的局部相似性计算与多头注意力模块能有效捕捉目标样本与源域的时变关联性，动态权重策略对复杂工况适应性更强。

表 5-3 AMSEDG 消融实验结果

模型变体	T1	T2	T3	T4	平均值	性能变化
B1	95.6	96.2	94.8	93.9	95.1	-
B2	93.1	94.0	92.3	91.5	92.7	▼2.4
B3	92.8	93.5	91.7	90.9	92.2	▼2.9
B4	91.5	92.4	90.1	89.3	90.8	▼4.3
B5	90.2	91.0	89.0	88.1	89.6	▼5.5

移除多头注意力模块(B3)导致平均准确率下降 2.9%，在 T4 任务中降幅达 3.0%。该模块通过多尺度交互建模增强了跨域特征的细粒度对齐能力，尤其在类间分布重叠严重的场景中作用显著。

仅使用领域级特征学习(B4)时，模型在细粒度特征对齐上的不足导致平均性能下降 4.3%，表明样本级三元组损失与裕度差异损失对类内紧致性和类间可分性具有关键作用。

仅保留样本级特征学习(B5)则使性能进一步恶化至 89.6%，较完整模型下降 5.5%。这验证了领域级相关性对齐是构建全局分布一致性的基础，两类特征学习存在互补增强关系。

消融实验显示：动态权重机制通过时域自适应调整权重分配，多头注意力模块强化跨域交互建模，多层次特征学习策略实现全局与局部特征的对齐协同优化，三者共同奠定了 AMSEDG 性能优势的基础。

5.4 本章小结

本章提出基于自适应多源集成的跨域轴承故障诊断方法 AMSEDG，通过多层次不变特征学习与基于动态相似性加权集成机制，有效解决了多源域异构场景下的旋转机械故障诊断泛化难题。引入双层次特征对齐架构，结构分别为（1）领域级特征学习通过协方差对齐与 MMD 损失实现跨域全局分布适配（2）样本级特征学习结合三元组损失与裕度差异损失，增强细粒度类间可分性；以及引入基于动态相似性的加权集成机制，引入滑动窗口局部相似性计算与多头注意力模块，构建时空双重权重优化。

实验验证可以证明在四组迁移任务中，AMSEDG 以 95.1% 的平均准确率超越 8 种对比方法，AMSEDG 相较于表现最佳的 HAF-MDWGANGP 提升 2.2%，验证了其在多源协同泛化方面的优势。消融实验揭示了各组件的贡献：动态权重机制（+2.4%）、多头注意力（+2.9%）以及双层次特征学习（+4.3%）共同形成了性能提升的技术闭环。

AMSEDG 为工业多源异构数据场景提供了高效且卓越的领域泛化框架，其动态权重机制对设备变工况和传感器等实际问题展现出较高的适应性。未来将进一步研究该框架在跨设备故障迁移诊断中的通用性，并优化实时推理效率，以满足在线监测的要求。

第六章 总结与展望

6.1 全文工作总结

随着工业智能化和数据采集技术的快速进步，旋转机械故障诊断领域积累了大量多源异构数据。然而，当前领域泛化仍然无法很好处理单源域数据稀缺、跨工况分布差异较大及多源域过拟合特定源域这三重问题。本文以轴承故障诊断为研究核心，聚焦领域泛化技术，提出了一套逐步优化的解决方案，旨在突破传统方法的局限，为复杂工业环境下的智能诊断开拓新思路。

本文的主要研究内容及结论如下：

(1) 基于多模态数据增强的跨域轴承故障诊断方法 MDWGANGP：针对单源域数据稀缺及负载不均衡问题，设计了奇异值分解(SVD)与网络嵌入更新(NEU)结合的数据重构方案，并整合 BiLSTM-CNN 多模态特征提取及 WGAN-GP 对抗生成技术，大幅提升了数据多样性。实验结果表明，该算法在 12 项跨域任务中的平均准确率达 85.78%，相较于其他前沿领域泛化方法 Mixup、GroupDRO 等可以提升 6-10 个百分点，实验结果验证了多模态协同增强的显著优势。

(2) 基于多损失协同优化的跨域轴承故障诊断方法 HAF-MDWGANGP：在 MDWGANGP 原有的交叉熵损失引入 MMD 损失、三元组损失与裕度差异损失，构建全局分布对齐-局部类可分性-决策边界鲁棒性三级优化机制，解决复杂领域偏移下的特征失配问题。在动态负载场景中，平均诊断准确率提升至 89.24%，消融实验表明多损失协同优化贡献了 3.46 个百分点的性能增益，尤其在 B→A 任务中准确率可达 94.1%。

(3) 基于自适应多源集成的跨域轴承故障诊断方法 AMSEDG：在结合前两章方法上提出多层次不变特征学习与动态相似性加权集成机制，以此更好地解决在多源域情况下易过拟合特定源域导致对不可见目标域预测准确率差的问题。在四组多源域任务中，AMSEDG 以 95.1% 的平均准确率显著领先，较其他最优方法 HAF-MDWGANGP 提升 2.2 个百分点，验证了动态权重策略与集成学习的协同优势。

本文研究从数据增强、特征对齐再到多源集成方向上，构建了一个逐步完善的领域泛化模型，实验结果可以得出本文研究可以为工业设备智能运维提供了理论基础和高鲁棒性解决方案。

6.2 下一步工作展望

尽管本文方法取得了显著进展，仍存在以下改进空间：

(1) 实时性优化: AMSEDG 的动态权重计算依赖在线滑动窗口与多头注意力, 导致推理延迟较高。未来可通过轻量化注意力模块或边缘计算部署提升实时性。

(2) 跨设备泛化: 当前实验限于同类型轴承数据, 需验证框架在跨设备(如齿轮箱-轴承联合诊断)中的普适性。引入拓扑结构先验知识或物理信息嵌入是潜在方向。

(3) 开放集泛化: 现有方法假设目标域类别与源域一致, 实际中可能面临未知故障类型。结合开集识别(Open-Set DG)与增量学习是未来重点。

(4) 理论解释性: 动态权重机制缺乏严格的泛化误差界分析。通过信息瓶颈理论或因果推断建模, 可进一步提升方法可解释性。

本文工作为领域泛化在故障诊断中的应用提供了新思路, 后续将围绕上述方向深化研究, 推动工业智能维护系统的实用化落地。

致 谢

无他，真挚地感谢*教授、*团队以及所有*班同学们、老师的关心、支持和帮助，同时无比珍惜在电子科技大学（深圳）高等研究院的三年研究生生涯时光。

正如习总书记所说，我们既要学习有字之书，也要学习无字之书，希望从象牙塔出来之后能够重新出发，再次奋斗，汲取来自社会的“知识”。

梦虽遥，追则能达；愿虽艰，持则可圆。希望未来大家能在坚守初衷的情况下，坚持追逐自己的梦想！

参考文献

- [1] Hu C, Wang Y. Multidimensional denoising of rotating machine based on tensor factorization[J]. Mechanical Systems and Signal Processing, 2019, 122: 273-289.
- [2] Smith W A, Randall R B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study[J]. Mechanical systems and signal processing, 2015, 64: 100-131.
- [3] 雷亚国, 贾峰, 孔德同, 等. 大数据下机械智能故障诊断的机遇与挑战[J]. 机械工程学报, 2018, 54(5): 94-104.
- [4] Xiao Q, Yang M, Yan J, et al. Feature decoupling integrated domain generalization network for bearing fault diagnosis under unknown operating conditions[J]. Scientific Reports, 2024, 14(1): 30848.
- [5] Liao Y, Huang R, Li J, et al. Deep semisupervised domain generalization network for rotary machinery fault diagnosis under variable speed[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(10): 8064-8075.
- [6] Zhang W, Peng G, Li C, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals[J]. Sensors, 2017, 17(2): 425.
- [7] Guo Y, Ju G, Zhang J. A domain generalization network for imbalanced machinery fault diagnosis[J]. Scientific Reports, 2024, 14(1): 25447.
- [8] Khirodkar R, Yoo D, Kitani K. Domain randomization for scene-specific car detection and pose estimation[C]. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 2019: 1932-1940.
- [9] Gan C, Yang T, Gong B. Learning attributes equals multi-source domain generalization[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 87-97.
- [10] Wang J, Lan C, Liu C, et al. Generalizing to unseen domains: A survey on domain generalization[J]. IEEE transactions on knowledge and data engineering, 2022, 35(8): 8052-8072.
- [11] Chen K, Zhuang D, Chang J M. Discriminative adversarial domain generalization with meta-learning based cross-domain validation[J]. Neurocomputing, 2022, 467: 418-426.
- [12] Tobin J, Fong R, Ray A, et al. Domain randomization for transferring deep neural networks from simulation to the real world[C]. 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), 2017: 23-30.

- [13] Peng X, Qiao F, Zhao L. Out-of-domain generalization from a single source: An uncertainty quantification approach[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 46(3): 1775-1787.
- [14] Shankar S, Piratla V, Chakrabarti S, et al. Generalizing across domains via cross-gradient training[J]. arXiv preprint arXiv:1804.10745, 2018.
- [15] Zhou K, Yang Y, Hospedales T, et al. Deep domain-adversarial image generation for domain generalisation[C]. Proceedings of the AAAI conference on artificial intelligence, 2020: 13025-13032.
- [16] Doersch C. Tutorial on variational autoencoders[J]. arXiv preprint arXiv:1606.05908, 2016.
- [17] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
- [18] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
- [19] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357.
- [20] Zhang J, Yang B. Enhancing Domain Generalization in Rotating Machinery Fault Diagnosis Through Diffusion Model-Based Data Augmentation[C]. 2024 IEEE 22nd International Conference on Industrial Informatics (INDIN), 2024: 1-7.
- [21] Zhao C, Shen W. Imbalanced domain generalization via semantic-discriminative augmentation for intelligent fault diagnosis[J]. Advanced Engineering Informatics, 2024, 59: 102262.
- [22] Liu S, Jiang H, Wu Z, et al. Rolling bearing fault diagnosis using variational autoencoding generative adversarial networks with deep regret analysis[J]. Measurement, 2021, 168: 108371.
- [23] Guo Q, Li Y, Song Y, et al. Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network[J]. IEEE Transactions on Industrial Informatics, 2019, 16(3): 2044-2053.
- [24] Ben-David S, Blitzer J, Crammer K, et al. Analysis of representations for domain adaptation[J]. Advances in neural information processing systems, 2006, 19.
- [25] Hu S, Zhang K, Chen Z, et al. Domain generalization via multidomain discriminant analysis[C]. Uncertainty in artificial intelligence, 2020: 292-302.
- [26] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation[C]. International conference on machine learning, 2015: 1180-1189.
- [27] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[J]. Journal of machine learning research, 2016, 17(59): 1-35.

- [28] Gong R, Li W, Chen Y, et al. Dflow: Domain flow for adaptation and generalization[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 2477-2486.
- [29] Zhou F, Jiang Z, Shui C, et al. Domain generalization via optimal transport with metric similarity learning[J]. Neurocomputing, 2021, 456: 469-480.
- [30] Zunino A, Bargal S A, Volpi R, et al. Explainable deep classification models for domain generalization[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 3233-3242.
- [31] Nam H, Lee H, Park J, et al. Reducing domain gap by reducing style bias[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021: 8690-8699.
- [32] He Y, Shen Z, Cui P. Towards non-iid image classification: A dataset and baselines[J]. Pattern Recognition, 2021, 110: 107383.
- [33] Zhang X, Cui P, Xu R, et al. Deep stable learning for out-of-distribution generalization[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021: 5372-5382.
- [34] Shi Y, Deng A, Deng M, et al. Domain transferability-based deep domain generalization method towards actual fault diagnosis scenarios[J]. IEEE Transactions on Industrial Informatics, 2022, 19(6): 7355-7366.
- [35] Zheng H, Yang Y, Yin J, et al. Deep domain generalization combining a priori diagnosis knowledge toward cross-domain fault diagnosis of rolling bearing[J]. IEEE Transactions on instrumentation and measurement, 2020, 70: 1-11.
- [36] Li J, Li Y, Tan J, et al. Bridging the gap with grad: Integrating active learning into semi-supervised domain generalization[J]. Neural Networks, 2024, 171: 186-199.
- [37] Wu G, Gong S. Collaborative optimization and aggregation for decentralized domain generalization and adaptation[C]. Proceedings of the IEEE/CVF international conference on computer vision, 2021: 6484-6493.
- [38] Khoei A G, Yu Y, Feldt R. Domain generalization through meta-learning: A survey[J]. Artificial Intelligence Review, 2024, 57(10): 285.
- [39] Shi Y, Seely J, Torr P H, et al. Gradient matching for domain generalization[J]. arXiv preprint arXiv:2104.09937, 2021.
- [40] Rame A, Dancette C, Cord M. Fishr: Invariant gradient variances for out-of-distribution generalization[C]. International Conference on Machine Learning, 2022: 18347-18377.

- [41] Krueger D, Caballero E, Jacobsen J-H, et al. Out-of-distribution generalization via risk extrapolation (rex)[C]. International conference on machine learning, 2021: 5815-5826.
- [42] Sagawa S, Koh P W, Hashimoto T B, et al. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization[J]. arXiv preprint arXiv:1911.08731, 2019.
- [43] Carlucci F M, D’Innocente A, Bucci S, et al. Domain generalization by solving jigsaw puzzles[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 2229-2238.
- [44] Ren L, Mo T, Cheng X. Meta-learning based domain generalization framework for fault diagnosis with gradient aligning and semantic matching[J]. IEEE Transactions on Industrial Informatics, 2023, 20(1): 754-764.
- [45] Wang Z, Oates T. Imaging time-series to improve classification and imputation[J]. arXiv preprint arXiv:1506.00327, 2015.
- [46] Breiman L. Bagging predictors[J]. Machine learning, 1996, 24: 123-140.
- [47] Harris T A, Kotzalas M N. Advanced concepts of bearing technology: rolling bearing analysis[M]. CRC press, 2006.
- [48] Randall R B, Antoni J. Rolling element bearing diagnostics—A tutorial[J]. Mechanical systems and signal processing, 2011, 25(2): 485-520.
- [49] CusidÓCusido J, Romeral L, Ortega J A, et al. Fault detection in induction machines using power spectral density in wavelet decomposition[J]. IEEE Transactions on Industrial Electronics, 2008, 55(2): 633-643.
- [50] Sikorska J Z, Hodkiewicz M, Ma L. Prognostic modelling options for remaining useful life estimation by industry[J]. Mechanical systems and signal processing, 2011, 25(5): 1803-1836.
- [51] Li X, Zhang W, Ding Q. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction[J]. Reliability engineering & system safety, 2019, 182: 208-218.
- [52] Antoni J. The spectral kurtosis: a useful tool for characterising non-stationary signals[J]. Mechanical systems and signal processing, 2006, 20(2): 282-307.
- [53] Peng Z, Peter W T, Chu F. A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing[J]. Mechanical systems and signal processing, 2005, 19(5): 974-988.
- [54] 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.

- [55] Jack L B, Nandi A K. Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms[J]. Mechanical systems and signal processing, 2002, 16(2-3): 373-390.
- [56] Zhao R, Yan R, Chen Z, et al. Deep learning and its applications to machine health monitoring[J]. Mechanical Systems and Signal Processing, 2019, 115: 213-237.
- [57] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [58] 王子栋.基于域自适应的轴承故障诊断方法研究及系统实现[D].电子科技大学,2024.DOI:10.27005/d.cnki.gdzku.2024.004818.
- [59] Lu W, Wang J, Li H, et al. Domain-invariant feature exploration for domain generalization[J]. arXiv preprint arXiv:2207.12020, 2022.
- [60] Li D, Yang Y, Song Y-Z, et al. Learning to generalize: Meta-learning for domain generalization[C]. Proceedings of the AAAI conference on artificial intelligence, 2018.
- [61] Ragab M, Chen Z, Zhang W, et al. Conditional contrastive domain generalization for fault diagnosis[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-12.
- [62] Qian Q, Zhou J, Qin Y. Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions[J]. IEEE transactions on industrial informatics, 2023, 19(9): 9898-9908.
- [63] Zhao C, Shen W. A domain generalization network combining invariance and specificity towards real-time intelligent fault diagnosis[J]. Mechanical Systems and Signal Processing, 2022, 173: 108990.
- [64] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of machine learning research, 2008, 9(11).
- [65] Zhao C, Shen W. Adversarial mutual information-guided single domain generalization network for intelligent fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2022, 19(3): 2909-2918.
- [66] Shahtalebi S, Gagnon-Audet J-C, Laleh T, et al. Sand-mask: An enhanced gradient masking strategy for the discovery of invariances in domain generalization[J]. arXiv preprint arXiv:2106.02266, 2021.
- [67] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]//International conference on machine learning. PMLR, 2017: 214-223.

攻读硕士学位期间取得的成果