

中国科学技术大学

硕士学位论文



PCB 缺陷元件复检及其可修复性判定 方法研究

作者姓名： 陈龙鑫

学科专业： 控制科学与工程

导 师： 赵云波 教授

完成时间： 二〇二五年五月二十四日

University of Science and Technology of China

A dissertation for master's degree



Research on Reinspection and Repairability Determination Methods for Defective PCB Components

Author: Longxin Chen

Speciality: Control Science and Engineering

Supervisor: Prof. Yun-Bo Zhao

Completion date: May 24, 2025

中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：陈龙鑫

签字日期：2025.5.24

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

控阅的学位论文在解密后也遵守此规定。

☒ 公开 ☐ 控阅（____ 年）

作者签名：陈龙鑫

导师签名：袁波

签字日期：2025.5.24

签字日期：2025.05.24

摘要

表面贴装技术（Surface Mount Technology, SMT）是一种将电子元件焊接到印刷电路板（Printed Circuit Board, PCB）表面指定位置的技术。为把控焊接质量，标准 SMT 工艺采用自动光学检测（Automated Optical Inspection, AOI）机器对完成焊接的 PCB 元件进行初步缺陷检测。然而，据调研，超过 95% 的 AOI 初检缺陷元件实际并无缺陷。因此，需由专人对 AOI 初检缺陷元件进行缺陷复检，并对复检为缺陷的元件进行可修复性判定。这两个环节高度依赖人工，会耗费较大人力成本，影响生产效率。因此，实现自动的缺陷元件复检及其可修复性判定，可以节省人力成本，并提高产线的自动化水平及生产效率。

但是，在实现自动的缺陷元件复检及其可修复性判定的过程中，面临以下难点：在缺陷元件自动复检任务中，复检无缺陷元件的数量远超复检缺陷元件的数量，存在数据类别高度不平衡的难点；在缺陷元件可修复性自动判定任务中，复检缺陷元件的数量少，存在数据样本量小的难点。目前相关研究较少，尚处于初步探索阶段，未充分关注并深入解决上述难点，因此取得的效果有限。

因此，本文基于 SMT 生产过程数据，针对缺陷元件自动复检任务和缺陷元件可修复性自动判定任务展开研究。本文的具体研究工作包括以下两个方面：

（1）考虑到数据不平衡的难点，提出了基于深度自编码合成少数类过采样技术（Synthetic Minority Oversampling Technique, SMOTE）的缺陷元件复检方法，实现了比现有方法更优的自动复检性能。该方法采用引脚行重组后的数据构建深度自编码器，并利用 SMOTE 方法在深度自编码空间内对少数类元件数据进行过采样，以平衡数据集。同时，设计了层次特征融合分类模型，利用神经网络分别在引脚层和元件层进行深度特征融合，以进一步缓解不平衡数据带来的影响。在 PHME2022 数据集上进行了实验，结果表明，相比于现有方法的最佳性能，所提方法的 F1 分数提高了 0.09，精确率提高了 6.9%，召回率提高了 8.3%。

（2）考虑到数据样本量小的难点，提出了基于任务演进 TabNet 训练策略的缺陷元件可修复性判定方法，实现了比现有方法更优的自动判定性能。设计了元件分层特征提取方法，从不同层级全面提取元件的缺陷特征。同时，通过双阶段任务演进 TabNet 训练策略，缓解小样本数据带来的影响。在第一阶段，利用大量 AOI 初检缺陷元件数据预训练 TabNet，提升其缺陷感知能力；在第二阶段，利用少量复检缺陷元件数据对 TabNet 进行再训练，提升其对元件可修复性的判定能力。在 PHME2022 数据集上进行了实验，结果表明，相比于现有方法的最佳性能，所提方法的 F1 分数提高了 0.08，精确率提高了 7.9%，召回率提高了 6.5%。

关键词：表面贴装技术；缺陷复检；可修复性判定；不平衡数据；小样本数据

ABSTRACT

Surface Mount Technology (SMT) is used to solder electronic components onto specified locations on a Printed Circuit Board (PCB). In order to control the quality of soldering, the standard SMT process uses Automated Optical Inspection (AOI) machines to carry out initial defect detection on the soldered PCB components. However, research has found that more than 95% of the defective components marked by AOI machines are actually not defective. Therefore, dedicated personnel must reinspect the components flagged as defective by AOI machines and determine the repairability of those confirmed to be truly defective. These two processes heavily rely on manual labour, leading to high labour costs and low production efficiency. Therefore, automatic reinspection of defective components and determination of their repairability can reduce labour costs and improve production line automation and efficiency.

However, the automation of reinspection and repairability determination of defective components faces the following challenges. Firstly, in the reinspection task, the number of components reinspected as non-defective far exceeds the number of components reinspected as defective, resulting in a high degree of data imbalance. Secondly, in the repairability determination task, the number of components reinspected as defective is small, resulting in a small sample size. Currently, few relevant studies exist. And these studies remain at a preliminary stage, without sufficient attention and in-depth solution to the challenges mentioned above, thus achieving limited results.

Therefore, this thesis addresses the task of automatic reinspection and repairability determination of defective components, based on the SMT production process data. The specific research work of this thesis includes the following two parts:

(1) Considering the challenge of data imbalance, a reinspection method for defective PCB components based on deep autoencoding Synthetic Minority Oversampling Technique (SMOTE) is proposed, achieving superior performance compared to existing methods. This method uses the data after pin row reorganization to construct a deep autoencoder, and applies SMOTE to oversample the minority component data in the deep autoencoding space to balance the data. In order to further mitigate the impact of imbalanced data, a hierarchical feature fusion classification model is designed. The model performs deep feature fusion through neural networks at both the pin and component levels. Experiments were conducted on the PHME2022 dataset. The results showed that the proposed method outperformed existing methods, achieving a 0.09 in-

crease in F1 score, along with 6.9% and 8.3% improvements in precision and recall, respectively.

(2) Considering the challenge of small sample size, a repairability determination method for defective PCB components based on the task-evolving TabNet training strategy is proposed, achieving superior performance compared to existing methods. A component hierarchical feature extraction method is designed to comprehensively extract component defect features from different levels. In addition, a two-stage task-evolving TabNet training strategy is adopted to mitigate the impact of small sample data. In the first stage, a large amount of data on the defective components marked by AOI machines is used to pretrain TabNet to improve defect detection capability. In the second stage, TabNet is retrained using a small amount of data on components reinspected as defective to improve its capability of determining the repairability of components. Experiments were conducted on the PHME2022 dataset. The results showed that the proposed method outperformed existing methods, achieving a 0.08 increase in F1 score, along with 7.9% and 6.5% improvements in precision and recall, respectively.

KEY WORDS: Surface Mount Technology; Defect Reinspection; Repairability Determination; Imbalanced Data; Small Sample Data

目录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	4
1.2.1 PCB 缺陷元件复检研究现状	4
1.2.2 PCB 缺陷元件可修复性判定研究现状	5
1.2.3 不平衡数据分类方法研究现状	5
1.2.4 小样本数据分类方法研究现状	9
1.3 研究内容与结构安排	11
1.3.1 研究内容	11
1.3.2 结构安排	12
第 2 章 相关基础知识	14
2.1 SMT 工艺介绍	14
2.2 自编码器介绍	16
2.3 TabNet 模型介绍	17
2.3.1 激活函数	18
2.3.2 特征选择	20
2.3.3 特征处理	21
2.3.4 可解释性	22
2.4 本章小结	23
第 3 章 基于深度自编码 SMOTE 的 PCB 缺陷元件复检方法	24
3.1 引言	24
3.2 问题描述	25
3.3 SPI 与 AOI 数据特性分析	26
3.3.1 SPI 数据特性	27
3.3.2 AOI 数据特性	30
3.4 PCB 缺陷元件复检方法设计	32
3.4.1 数据预处理	33
3.4.2 深度自编码 SMOTE 数据增强方法设计	37
3.4.3 层次特征融合分类模型设计	42
3.5 实验设计与分析	45
3.5.1 实验数据	45

3.5.2 实验环境	46
3.5.3 评价指标	46
3.5.4 实验结果分析	48
3.6 本章小结	52
第 4 章 基于任务演进 TabNet 训练策略的 PCB 缺陷元件可修复性判定方法 ..	54
4.1 引言	54
4.2 问题描述	55
4.3 PCB 缺陷元件可修复性判定方法设计	56
4.3.1 元件分层特征提取方法设计	57
4.3.2 双阶段任务演进 TabNet 训练策略设计	62
4.4 实验设计与分析	64
4.4.1 实验数据	64
4.4.2 实验环境	65
4.4.3 评价指标	65
4.4.4 实验结果分析	65
4.5 本章小结	70
第 5 章 总结与展望	71
5.1 总结	71
5.2 展望	72
参考文献	73
致谢	80
在读期间取得的科研成果	81

插图清单

图 1.1	PCB 实例及其广泛应用示意图	1
图 1.2	标准 SMT 工艺的主要生产环节	1
图 1.3	锡膏印刷、元件贴装与回流焊接示意图	2
图 1.4	SMT 产线缺陷元件复检及其可修复性判定流程	4
图 1.5	论文的结构安排及创新点（结合 SMT 产线现状说明）	13
图 2.1	不同的 PCB 元件组装方式	14
图 2.2	单层自编码器的训练流程	16
图 2.3	深度自编码器的结构组成与数据处理流程	17
图 2.4	TabNet 网络结构图	17
图 2.5	ReLU 函数示意图	18
图 2.6	Sigmoid 函数示意图	19
图 2.7	Attentive Transformer 网络结构图	20
图 2.8	Feature Transformer 网络结构图	21
图 3.1	PCB 面板示意图	26
图 3.2	SPI 特征相关性可视化热图	28
图 3.3	Volume 与 Height 特征关系散点图	29
图 3.4	Volume 与 Area 特征关系散点图	30
图 3.5	AOI 缺陷数量与元件复检类别的关联性分析图	31
图 3.6	PCB 缺陷元件复检方法整体框架	32
图 3.7	引脚级多源数据融合示意图	33
图 3.8	数据组织结构	36
图 3.9	深度自编码 SMOTE 数据增强框架	38
图 3.10	基于引脚行重组的数据扩充示意图	40
图 3.11	层次特征融合分类模型框架	43
图 3.12	AOI 初检缺陷元件在 PCB 面板上的位置分布图	46
图 3.13	二分类混淆矩阵示意图	47
图 3.14	缺陷元件复检方法对比实验结果	49
图 3.15	深度自编码 SMOTE 方法消融实验结果	50
图 3.16	层次特征融合模型消融实验结果	52
图 4.1	缺陷元件可修复性判定方法整体框架	56

图 4.2	任务演进 TabNet 训练策略示意图（结合 SMT 产线现状说明）	62
图 4.3	复检缺陷元件在 PCB 面板上的位置分布图	64
图 4.4	缺陷元件可修复性判定方法对比实验结果	66
图 4.5	分层特征消融实验结果	67
图 4.6	双阶段任务演进 TabNet 训练策略消融实验结果	68
图 4.7	特征相对重要性示意图	69

附表清单

表 1.1	SPI 数据示例	2
表 1.2	AOI 数据示例	3
表 3.1	SPI 数据特性	27
表 3.2	AOI 数据特性	31
表 3.3	实验环境配置参数	47
表 4.1	引脚级特征列表	57
表 4.2	元件级特征列表	60

第 1 章 绪论

1.1 研究背景与意义

随着电子信息产业的快速发展，电子设备的需求日益增长^[1-2]。如图1.1所示，印刷电路板（Printed Circuit Board, PCB）作为智能手机、计算机、电视、汽车系统、医疗设备和工业机械设备等各种电子设备的核心组成部分，是这些设备运行的基础^[3]。表面贴装技术（Surface Mount Technology, SMT）是一种 PCB 元件贴装技术，用于将表面贴装元件焊接到 PCB 裸板表面预定的位置^[4-7]。

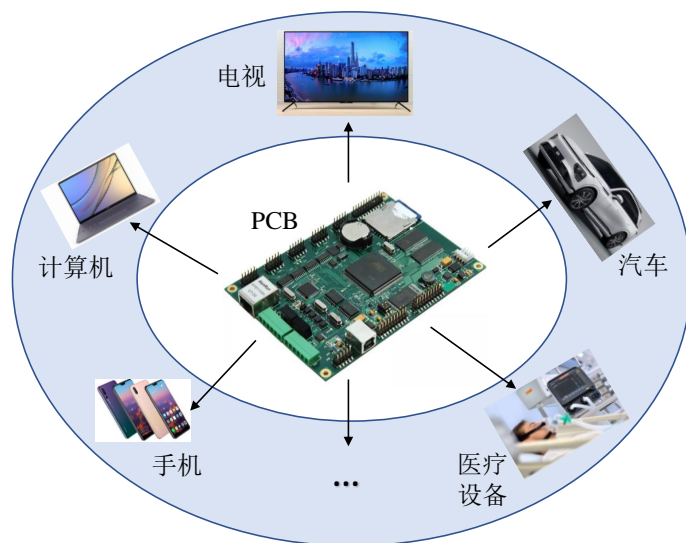


图 1.1 PCB 实例及其广泛应用示意图

标准 SMT 工艺通常包括五个主要生产环节：锡膏印刷、锡膏检测、元件贴装、回流焊接和自动光学检测，如图1.2所示^[8-9]。



图 1.2 标准 SMT 工艺的主要生产环节

首先，锡膏印刷环节的任务是将锡膏精准印刷至 PCB 的焊盘位置。焊盘是 PCB 上专门设计的金属化区域，在电子元件引脚与 PCB 之间起到机械固定和电气连接的关键作用。该环节采用专用的锡膏印刷机进行操作，其核心部件包括钢网和刮刀。钢网依据 PCB 的几何布局设计，在焊盘位置设置特定的开孔图案。而刮刀在压力作用下以特定角度推动锡膏流动，通过钢网上的开孔将锡膏精准、

定量地漏印到 PCB 的焊盘上，如图1.3(a)所示。锡膏印刷环节对最终 PCB 元件的焊接质量有至关重要的影响。统计数据显示，50%-70% 的元件焊接缺陷与该环节有关^[10-12]。为把控印刷质量，通常在锡膏印刷机后配置锡膏检测机（Solder Paste Inspection Machine, SPI Machine），对锡膏的体积、面积、高度及偏移等特性进行检测，并生成相应的 SPI 数据，如表1.1所示。

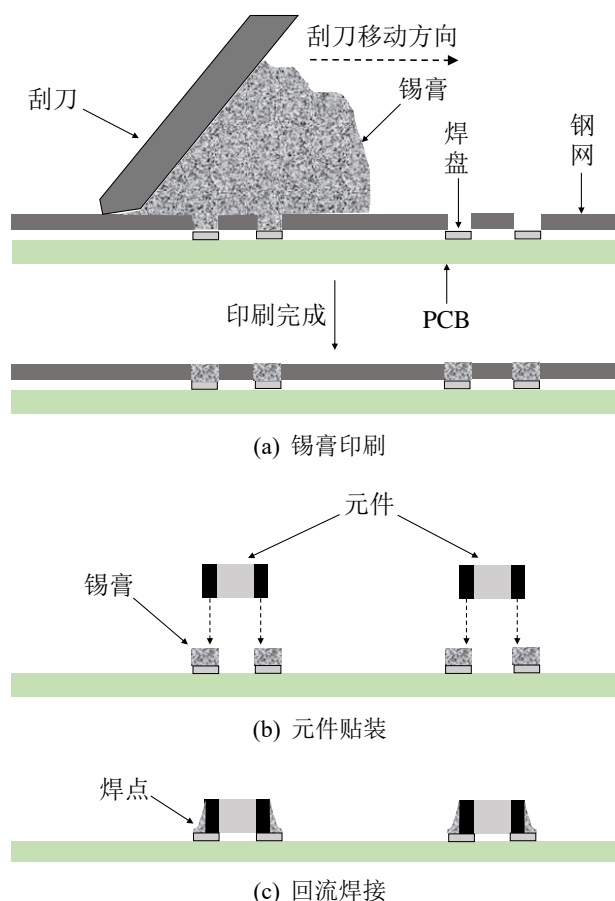


图 1.3 锡膏印刷、元件贴装与回流焊接示意图

表 1.1 SPI 数据示例

面板标识符	PCB 标识符	元件标识符	引脚号	体积 (μm^3)	面积 (μm^2)
2844	2	R11	1	3506371000	27013470
2844	2	R11	2	3369075000	26957750
2844	2	R12	1	3691364000	27189650
2844	2	R12	2	3645320000	27252060

随后，进入元件贴装环节。在元件贴装环节，元件贴装机将各类元件精确地拾取并贴装至 PCB 表面的指定焊盘位置，如图1.3(b)所示。接着，在回流焊接环节，PCB 被送入回流炉中^[13]。锡膏在回流炉中被加热熔化并形成可靠的焊点，从而将元件引脚牢固焊接到 PCB 焊盘上，如图1.3(c)所示。为把控元件焊接质量，标准 SMT 工艺通常在回流炉之后放置自动光学检测机（Automated Optical

Inspection Machine, AOI Machine), 对已完成元件焊接的 PCB 进行检测, 以确认其是否存在元件焊接缺陷^[14]。AOI 机器能够检测 PCB 表面的元件焊接缺陷, 如焊点桥接和焊点偏移等, 并生成相应的 AOI 数据, 如表1.2所示。

表 1.2 AOI 数据示例

面板标识符	PCB 标识符	元件标识符	引脚号	AOI 缺陷标签
2844	2	D1	1	焊点桥接
2844	2	D1	1	焊点偏移

但在实际生产过程中, AOI 机器存在大量的误报情况, 即将大量实际无焊接缺陷的 PCB 元件误报为存在焊接缺陷^[15-16]。在对 X 公司^①的 SMT 产线开展深入调研后发现, 在 AOI 机器初次检测出的缺陷元件中, 实际无缺陷的元件占比超过 95%, 这使得完成元件焊接的 PCB 在 AOI 机器上的首次通过率低于 20%。换言之, 超过 80% 的 PCB 会被 AOI 机器误报为不合格产品并被拦截。由于 AOI 机器实际存在大量的误报情况, 对于其初次检测为缺陷的 PCB 元件, 需要 AOI 缺陷复检操作员对其进行实时复检分类, 以确认这些 AOI 初检缺陷元件是否存在真实焊接缺陷。在这些缺陷元件的复检分类过程中, 需要耗费大量的人力成本, 尤其是在大规模生产的场景下。以 X 公司为例, 其每条 SMT 产线在任意时刻配备 2 名 AOI 缺陷复检操作员。X 公司共有 30 条 SMT 产线, 这意味着 X 公司在任意时刻需安排 60 名 AOI 缺陷复检操作员。考虑到白夜班轮班制度, 每天实际所需的 AOI 缺陷复检操作员数量达到 120 人次。同时, AOI 缺陷复检操作员的工作效率会制约产线的生产效率。

之后, 对于 AOI 缺陷复检操作员复检为缺陷的元件, 需要安排修复操作员进一步检查并判定其可修复性, 并对其中可修复的缺陷元件进行修复。而修复操作员判定复检缺陷元件可修复性的过程是耗时的, 这会影响其实际修复工作的效率。此外, 在大规模生产场景下, 为维持修复流程的正常运转, 需配备较多修复操作员, 会明显增加人力成本。以 X 公司为例, 其每 2 条 SMT 产线配备 1 名修复操作员。X 公司共有 30 条 SMT 产线, 这意味着 X 公司在任意时刻需安排 15 名修复操作员。考虑到白夜班轮班制度, X 公司每天实际所需的修复操作员数量为 30 人次。

在实际产线中, 缺陷元件复检及其可修复性判定的具体流程如图1.4所示。前述分析表明, 现有产线在缺陷元件的复检与可修复性判定环节高度依赖人工, 这不仅增加了人力成本, 还在一定程度上制约了生产效率的提升。

基于上述考虑, 同时鉴于 SPI 数据可以提供锡膏印刷质量信息, AOI 数据可

^①X 公司是一家专注于笔记本电脑生产的先进制造企业, 拥有 30 条 SMT 产线。本文研究课题依托于与 X 公司合作开展的实际项目, 因此对其 SMT 产线的实际运行状况进行了调查。但受商业机密限制, 本文不公开公司名称。

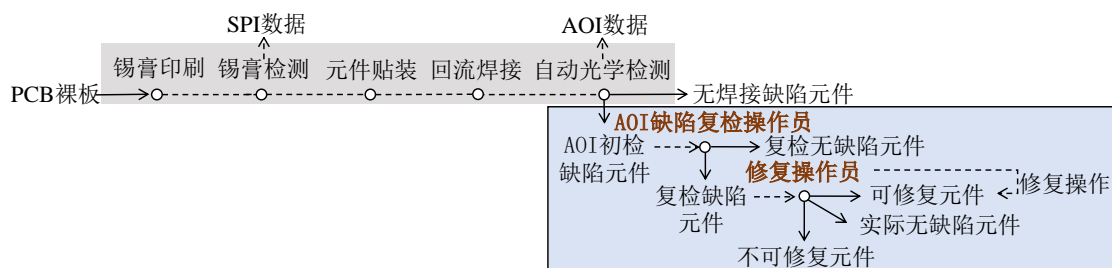


图 1.4 SMT 产线缺陷元件复检及其可修复性判定流程

以在一定程度上反映表面缺陷信息，本文基于工厂现有标准 SMT 产线中可获得的生产过程数据，即 SPI 和 AOI 数据，在 AOI 机器初检结果的基础上，利用数据驱动算法，对 AOI 初检缺陷元件进行自动复检，并实现复检缺陷元件可修复性的自动判定。通过对 AOI 初检缺陷元件进行自动复检，可以取代实际产线中的 AOI 缺陷复检操作员，以省去相应的人力成本；通过对复检缺陷元件的可修复性进行自动判定，可以使修复操作员不再耗费时间于复检缺陷元件的可修复性判定，使其专注于可修复缺陷元件的修复工作，提高实际修复效率，进而减少其数量的配置，降低相应的人力成本。同时，通过对数据驱动算法的研究，可以进一步提升现有 SMT 产线的自动化水平和生产效率。

1.2 国内外研究现状

1.2.1 PCB 缺陷元件复检研究现状

PCB 缺陷元件复检，指对 AOI 机器初检为缺陷的元件进行复检，并根据复检结果将其分类为复检缺陷元件和复检无缺陷元件。现有产线通过 AOI 缺陷复检操作员进行人工复检。同时，已有研究人员基于 SPI 和 AOI 数据，开展缺陷元件自动复检方面的研究工作。但目前相关研究较少，尚处在初步探索阶段。

Gore 等^[17]利用元件的 AOI 数据统计特征，基于 LightGBM 算法对缺陷元件进行复检分类，但该研究未能利用 SPI 数据。Tang 等^[18]利用了 SPI 数据中的单一特征和 AOI 数据特征，基于随机森林算法对缺陷元件进行复检分类，但该研究未能充分利用 SPI 数据中丰富的锡膏特性数据。Gaffet 等^[19]基于 XGBoost 算法，在单一引脚层次使用 SPI 和 AOI 数据训练复检模型，但未能考虑不同引脚间锡膏焊接质量的相互影响。Schmidt 等^[20]在原始 SPI 数据和 AOI 数据上提取元件级统计特征，并利用随机森林算法对缺陷元件进行复检分类。但该研究在提取特征的过程中损失了很多与焊接质量相关的细节信息，未能充分利用现有数据。

同时，AOI 机器存在大量的误报情况，会将大量实际没有焊接缺陷的 PCB 元件误报为存在焊接缺陷，这导致 AOI 缺陷复检操作员确认为无缺陷的元件数量远超于其确认为缺陷的元件数量。因此，在缺陷元件自动复检任务中，获取自

实际生产过程的数据集存在数据类别高度不平衡的问题，即复检无缺陷元件的数量要远超于复检缺陷元件的数量。但 Gore 等^[17]和 Gaffet 等^[19]未能关注该任务中的数据不平衡问题。Tang 等^[18]试图通过设置类别阈值来简单缓解数据不平衡问题，但未能较好地应对数据不平衡问题。Schmidt 等^[20]通过设置损失函数中的类别权重来简单缓解复检任务中的数据不平衡问题，并未取得较好效果。

总体而言，现有研究尚处于初步探索阶段，主要采用 LightGBM、XGBoost 和随机森林等传统的树模型算法，并结合简单的特征提取方法进行缺陷元件的自动复检。但是，现有研究未能充分关注并深入研究缺陷元件自动复检任务中数据类别高度不平衡的难点。因此，现有方法的效果有限，仍有进一步改进和深入研究的空间。

1.2.2 PCB 缺陷元件可修复性判定研究现状

PCB 缺陷元件可修复性判定，指对复检缺陷元件进行检测，判定其可修复性。现有实际产线通过修复操作员进行复检缺陷元件的可修复性判定。同时，已有研究人员基于 SPI 和 AOI 数据，开展了缺陷元件可修复性自动判定方面的研究工作。但目前相关研究较少，尚处在初步探索阶段。

Schmidt 等^[20]提出了一种基于单一 AOI 统计特征的决策树模型，用于复检缺陷元件的可修复性判定，这相当于一种基于规则的方法。Gore 等^[17]和 Gaffet 等^[19]分别对 SPI 和 AOI 数据采取了不同的元件级特征提取方法，并利用 XGBoost 算法对元件的可修复性进行判定。Tang 等^[18]采用元件级统计特征提取方法，并利用随机森林算法对元件的可修复性进行判定。但是这些研究均在元件层级上提取统计特征，会丢失一些与焊接质量相关的细节信息。

同时，SMT 产线中真正出现缺陷的元件数量较少，使得复检过程中被确认为缺陷的元件数量较少，这导致缺陷元件可修复性自动判定任务中存在小样本数据问题。然而，上述这些研究均未对该任务下的小样本数据问题加以关注，未提出任何解决方案。

总体而言，现有研究尚处于初步探索阶段，主要采用决策树、XGBoost 和随机森林等传统的树模型算法，并结合简单的元件级统计特征提取方法进行缺陷元件可修复性的自动判定。然而，现有研究未能充分关注并深入研究缺陷元件可修复性自动判定任务中数据样本量小的难点。因此，现有方法的效果有限，仍有进一步改进和深入研究的空间。

1.2.3 不平衡数据分类方法研究现状

在不平衡数据中，各类别的样本数量往往相差较大，其中部分类别的样本数量明显少于其他类别的样本数量^[21]。通常，样本量少的类别被称为少数类，而

样本量多的类别被称为多数类^[22]。

在本文的缺陷元件自动复检分类任务中，存在数据类别高度不平衡的难点。数据不平衡是分类任务中常见的挑战之一，会使分类模型受到类别分布的偏斜影响，导致分类效果不理想。在缺陷与故障检测领域，由于缺陷或故障情况的发生频率较低，数据不平衡现象较为普遍。

当前主流的不平衡数据分类方法可以从数据与算法两个层面加以划分，分别从数据类别分布调整和模型结构优化的角度出发，提高模型的分类性能。

1. 数据层面的方法

数据层面的方法在模型训练之前调整不平衡数据，以缓解数据类别不平衡问题。主流的处理方法可以分为重采样方法和数据增强方法。

(1) 重采样方法

数据重采样通过调整各类别样本的数量分布来平衡数据集。重采样方法可分为以下三类：欠采样、过采样和混合采样^[23]。

欠采样，其核心思想是通过减少多数类的样本数量来平衡数据集的类别分布^[24]。随机欠采样是最基本的欠采样方法，其通过随机的方式在多数类数据中移除样本^[25]。常用的欠采样方法还有 Tomek Links 方法和基于聚类的方法^[26]。在 Tomek links 方法中，对属于不同类别的样本 M 和 N，如果 M 和 N 互为最近邻，那么 M 和 N 组成 Tomek link。此时，如果其中一个样本属于多数类，那么便可以将该样本删除。基于聚类的欠采样方法则将多数类样本聚成特定数量的簇，随后从每个簇中删除部分样本，为后续分类任务做准备。在基于聚类的欠采样方法中，最为常见的是 K-means 聚类方法^[26]。

然而，最近关于欠采样的大多数研究得出的结论是：由于样本选择策略不佳，欠采样方法可能会丢弃重要样本，导致信息损失，进而影响最终模型性能^[24,27]。Azlim 等^[28]在研究药物-靶点相互作用预测分类时，发现在类别高度不平衡的情况下，随机欠采样方法表现不佳。因此，尽管欠采样方法操作简单直观，但在实际中需要谨慎应用，防止关键信息的丢失。

由于欠采样方法的优化空间有限，以及合成少数类过采样技术（Synthetic Minority Oversampling Technique, SMOTE）的成功，过采样方法受到了更多关注^[29]。过采样方法通过增加少数类样本的数量来平衡数据集中不同类别样本的分布。常见的过采样方法包括随机过采样^[30]、SMOTE 及其变体方法^[31-32]等。随机过采样通过随机的方式复制少数类样本。然而，随机过采样可能会增加数据集中噪声样本的比例，同时可能会导致模型过拟合，从而影响模型的泛化能力。

SMOTE 方法由 Chawla 等人^[33]提出，其通过在少数类样本之间进行插值，生成新样本以平衡数据集。这一方法的核心思想是随机从当前少数类样本的邻居中选择一个少数类样本，并根据其与当前样本的差异生成新样本，从而有效地增

加少数类样本的数量。然而, SMOTE 方法也存在局限性。比如, SMOTE 方法只适用于连续特征, 不支持分类特征。因此, 研究人员纷纷引入了 SMOTE 的变体改进方法。用于类别型和连续型数据的合成少数类过采样技术 (Synthetic Minority Oversampling Technique for Nominal and Continuous, SMOTE-NC) 是 SMOTE 的变体, 专门为具有分类和连续特征的数据集而设计, 使得生成的样本在语义上更加合理^[34]。但是, SMOTE-NC 在处理分类特征时, 未考虑类别之间的关联性。针对 SMOTE-NC 的局限性, Mukherjee 等^[34]提出了增强型 SMOTE-NC。当分类特征与目标类别之间存在某种关联时, 使用增强型 SMOTE-NC 的模型比使用 SMOTE-NC 的模型取得了更好的分类效果。

但是, 基于 SMOTE 的过采样方法通过线性插值的方式生成新样本, 容易引入噪声, 生成不符合实际分布的新样本, 并且线性插值的方式难以有效捕捉原始数据的复杂非线性结构。同时, 由于高维空间中样本之间的欧式距离失去代表性, 使得插值策略难以生成有效的少数类样本。因此, 基于 SMOTE 的过采样方法难以有效应对高维数据。

混合采样则同时结合上述的欠采样和过采样方法。Shamsudin 等^[25]采用随机欠采样和过采样相结合的方式处理不平衡欺诈数据集, 提升了模型检测信用卡欺诈的性能。Zhou 等^[35]混合使用 SMOTE 和 Tomek Links 方法来解决不平衡数据对质量预测的影响。

(2) 数据增强方法

在传统的重采样方法中, 过采样方法无法有效捕捉到数据的复杂分布, 其合成样本的多样性和代表性不足, 无法有效地增强数据集^[36-37], 这引发了研究人员对基于生成的数据增强方法的兴趣。简单的数据增强方法会对样本进行变换, 通过预设的规则来生成和扩充少数类数据表示, 以平衡数据集。例如在计算机视觉领域, 常用旋转、灰度化、裁剪、加入噪声和颜色变化等方式生成新样本^[38]。但是这种简单的数据增强方法并不能实质性地改变样本模式, 未必能有效地增加样本特征多样性^[38]。

近年来, 随着人工智能技术的快速发展, 深度生成模型凭借其卓越的性能成为主流的数据增强技术, 被研究人员广泛关注^[39]。深度生成模型可以利用神经网络学习到数据的复杂分布, 生成高质量的新样本, 并增加样本特征多样性, 可用于少数类别样本数据的生成, 成为解决数据类别不平衡的主流方法^[40]。目前主流的深度生成网络为生成对抗网络 (Generative Adversarial Network, GAN)^[41]和自编码器 (Autoencoder, AE)^[42], 已成功用于包括故障诊断和缺陷检测在内的诸多领域, 并展示出卓越的数据生成能力^[24]。

GAN 由生成器和判别器组成, 通过生成器和判别器之间的博弈生成高质量且具有多样性的新样本。其中生成器从随机噪声中生成伪造数据, 目标是“骗

过”判别器；判别器区分真实数据与生成数据，目标是提高判别能力。在对抗训练过程中，二者交替优化，不断提高各自的能力。GAN 具有强大的生成能力，可以有效捕捉数据的复杂分布。Luo 等^[43]为了提高数据不平衡情形下的故障诊断性能，提出了一种基于 GAN 的生成模型，并研究了基于两级 GAN 的故障诊断方法，为数据不平衡情形下的故障诊断问题提供了一种有效的解决方案。Chen 等^[44]提出了一种名为紧凑卷积 Transformers-GAN 的方法，以缓解工业机器人复合故障诊断建模中的数据不平衡问题，该方法可以实现高质量的假数据生成和复合故障诊断建模。然而，GAN 的训练过程相对复杂且不稳定，可能会出现模式崩塌的现象。

AE 是一种无监督学习模型，其主要作用是学习数据的潜在表示，并可以应用于数据压缩、特征提取及数据生成领域^[42]。AE 由编码器和解码器两部分组成，其中编码器将输入数据映射为低维表示特征，而解码器基于低维表示特征重构输入数据^[45]。AE 训练过程中使用重构误差作为损失函数。在数据生成时，随机生成一个潜在表示特征向量，然后输入给编码器，便可以得到一个新的生成样本。AE 架构训练容易，但由于其设置简单，用于数据生成的效果并不好^[42]。

考虑到标准自编码器的缺点，一些自编码器的变体相继被提出。变分自编码器（Variational Autoencoder, VAE）通过最大化数据对数似然的变分下限来运行^[46-47]。VAE 中的损失函数通常是通过将重构损失与 KL 散度相结合来实现的。Fan 等^[48]使用 VAE 作为数据增强策略，生成缺陷样本，以解决晶片分类中的数据不平衡问题，极大地提升了分类模型的性能。Abdulganiyu 等^[49]提出了一种基于类别聚焦损失的 VAE 方法，该方法利用 VAE 来有效解决入侵检测系统中的网络流量数据不平衡问题。

Wasserstein 自编码器（Wasserstein Autoencoders, WAE）也是一种常见的自编码器变体，可用于数据生成^[50]。WAE 的损失函数是通过将重构损失与惩罚项相结合来实现的。在 WAE 中，惩罚项被解释为鉴别器网络的输出。Pei 等^[51]基于 WAE 解决滚动轴承诊断中的数据不平衡问题，取得了较好效果。Harikrishnan 等^[52]使用基于 Transformer 的 WAE 生成受控文本，帮助分类器克服了原始数据不平衡的挑战。

但是，基于自编码器的数据增强方法的主要优化目标是还原原始输入，并不引导模型学习类别判别特征。因此，生成的数据虽然在结构上与原始数据相似，但不一定包含足够的类别区分性，缺乏针对性地生成少数类样本的能力。

2. 算法层面的方法

针对数据类别不平衡问题，算法层面的方法通过优化模型结构及训练过程，增强模型对少数类数据的学习能力。

加权损失函数是常用的处理不平衡数据的算法级方法。该方法通过对损失

函数赋予类别权重，增加少数类误分类的惩罚，使模型更加关注少数类。其中，加权交叉熵损失函数和焦点损失函数^[53]是两种常用的加权损失函数。加权交叉熵损失函数根据类别不平衡程度动态调整权重，而焦点损失函数进一步减少易分类样本的影响，专注于难分类样本的优化。Roy 等^[54]提出了一种边际感知自适应加权损失函数，该方法通过动态调整每个小批次的权重，有效地缓解了类别不平衡带来的偏差问题。李长洪等^[55]将焦点损失函数引入到梯度提升决策树算法中，使模型在处理不平衡数据的分类问题时具有较好的效果。然而，加权损失函数对超参数的设置非常敏感，需要进行大量实验才能确定最优配置。这不仅增加了试错成本，还无法从根本上解决信息不足的问题。

集成学习方法通过结合多个弱分类器提升整体分类性能，也可用于解决数据不平衡问题^[24]。为了处理不平衡的数据，通常会使用 Bagging 和 Boosting 等技术。Bagging 创建原始数据集的子集，并在这些子集上训练多个模型，而子集中少数类别和多数类别的样本数量比例均衡。Boosting 侧重于在先前模型表现不佳的情况下训练后续模型，从而赋予少数群体更多的权重，逐步优化模型对少数类样本的分类能力。当进行分类时，这些个体模型分别会对最有可能的类别进行“投票”，从而产生从不同角度考虑信息的分类结果。这能够有效缓解单一模型对多数类的过拟合问题，进行更平衡和准确地分类。但该方法计算复杂度较高，同时会增加过拟合风险^[56]。

此外，良好的特征提取性能是实现卓越性能的前提。通过对特定数据进行算法设计，提高主干网络的特征提取和表示性能可以有效缓解数据不平衡问题^[24]。

总体而言，在数据层面的不平衡数据处理方法中，传统的重采样技术存在较大局限性，而基于深度生成模型的数据增强方法呈现出更大的应用潜力。在算法层面，可以通过设计加权损失函数、集成学习策略或增强网络的特征提取能力等方式，有效提升模型的分类性能。在实际应用中，数据层面和算法层面的方法可结合使用，以更有效地解决不平衡数据分类问题。

1.2.4 小样本数据分类方法研究现状

在小样本数据场景下，可用的训练样本总量非常有限，难以有效支持模型的训练。小样本数据会使模型无法充分捕捉到数据的内在特征，导致模型性能不稳定或效果不佳。

在本文的缺陷元件可修复性判定分类任务中，存在小样本数据的挑战。小样本数据是机器学习任务中常见的难点之一^[57]。在缺陷与故障检测领域，由于实际场景中获取样本（特别是缺陷样本或故障样本）的成本较高，小样本数据现象较为普遍。针对小样本数据问题，主流的解决方法可从数据和算法两个层面加以划分。

1. 数据层面的方法

小样本数据问题的本质是缺乏训练样本，因此可以通过数据增强方法进行数据扩充。简单的数据增强方法包括旋转、平移和翻转等操作，这些方法简单实用，可以在一定程度上丰富数据，但其生成的样本与原始样本过于相似，因此效果有限^[58]。

随着人工智能技术的快速发展，深度生成模型凭借其卓越的性能成为主流的数据增强方法，可用于解决小样本数据分类问题。常用的深度生成模型有 GAN^[41]和 AE^[42]。这几种方法已在 1.2.3 节中进行了介绍，故此处不再赘述。

针对不平衡数据，深度生成模型能够充分利用多数类样本，学习到多数类样本的特征分布，并基于此生成与少数类样本相似的数据。而在小样本数据场景下，尽管 GAN 和 AE 等深度生成模型能够进行数据生成，但这些模型的训练通常需要依赖于充足的数据支持。在小样本数据场景下，由于数据量不足，深度生成模型难以充分学习到数据的复杂分布，从而导致生成数据的质量和多样性受限。因此，这类方法在解决小样本数据分类问题时存在局限性^[58]。

2. 算法层面的方法

数据增强方法在小样本数据处理上有一定的局限性，难以生成高质量和多样性的数据。因此，研究人员尝试在算法层面更好地处理小样本数据。当前，主流的方法分为三类：基于元学习的方法、基于迁移学习的方法和基于度量学习的方法^[59]。此外，图神经网络也逐渐被用于小样本数据分类问题。

元学习方法尝试解决模型如何学习的问题，其核心目标是使模型具备自动获取元知识的能力。元知识是指模型在训练过程之外所能习得的信息，如超参数设置等^[60]。元学习方法可从大量前置任务中学习元知识，并利用这些元知识来指导模型在小样本任务中更好地学习^[61]。Santoro 等^[62]提出了基于记忆增强的神经网络，这是元学习方法在单样本学习问题上较早且较成功的一次尝试。此后，元学习方法开始进入研究人员视野，成为解决小样本数据分类问题的新兴方法。之后，Finn 等^[63]提出了未知模型的元学习方法，通过少步数的训练得到较好的小样本数据分类效果。Jamal 等^[64]提出了一种算法未知任务元学习方法用于小样本数据学习，并取得了较好效果。但是在某些小样本数据分类问题中，元知识难以被准确表征，使模型的设计较为困难，且在学习迭代过程中易出现不收敛问题^[59]。

迁移学习通过借助已有的知识来学习新知识，其主要应用场景之一便是小样本数据分类问题^[65-66]。只要源领域与目标领域之间有一定关联，迁移学习便可将模型在源领域的数据中学到的知识应用到目标领域，从而实现不同领域间知识的迁移。基于模型预训练和微调的方法^[61,67]是最常用的迁移学习方法，该方法通常在相对更大规模的源数据集上预训练模型，之后在小样本目标数据集

上对模型的结构或参数进行微调。微调后的模型最终用于小样本数据集对应的目标任务。Nakamura 等^[68]分析了微调方法的性能,并表明微调方法可以实现比常见的小样本学习方法更高的分类精度,甚至可以达到最先进算法的精度。基于模型预训练和微调的方法是解决小样本数据分类问题简单有效的方法,但是微调的效果高度依赖于预训练任务与目标任务间的匹配程度^[61]。

度量学习通过给定的距离函数度量两个样本间的相似度^[69-70]。在小样本数据分类问题中,度量学习通过计算待分类样本和已知分类样本间的相似度来确定待分类样本的类别^[71]。Koch 等^[72]提出了一种相似性度量模型,用于单样本图像识别。Vinyals 等^[73]提出了匹配网络,可将带标签的数据和不带标签的数据均映射到对应的标签上,实现小样本数据分类。Li 等^[74]提出了一种协方差度量网络用于小样本数据分类,其利用了协方差矩阵来衡量样本间的分布一致性。但是度量学习方法的效果受距离度量方式和数据分布的影响较大^[75]。

图神经网络专门用于处理图结构数据,因其优秀的性能和较强的可解释性而受到广泛关注^[76],开始被研究人员尝试用于处理小样本数据。图神经网络有多种变体,最常见的变体为图卷积神经网络。Gidaris 等^[77]将降噪自编码器融入图神经网络,以调整小样本类别的权重。李凡等^[78]针对小样本数据下的图像分类问题,设计了一种残差图卷积网络,提高了分类模型的分类准确率。但是图神经网络仅适用于可以被抽象成图模型的分类问题^[79]。

总体而言,数据层面的小样本数据处理方法难以生成高质量和多样性的数据,其实际应用效果有限。相比之下,算法层面的方法是解决小样本数据分类问题的主流手段,并展现出较大的应用潜力。在算法层面,不同类型的方法各具优势与局限性,在实际任务中需结合问题特性,合理选择最适合的方法,以提升模型的泛化能力和分类性能。

1.3 研究内容与结构安排

1.3.1 研究内容

本文基于标准 SMT 产线中的 SPI 和 AOI 数据,针对缺陷元件自动复检任务和缺陷元件可修复性自动判定任务展开研究。本文具体研究内容如下:

(1) 基于深度自编码 SMOTE 的 PCB 缺陷元件复检方法

针对缺陷元件自动复检任务,考虑到数据高度不平衡的难点,提出了一种基于深度自编码 SMOTE 的缺陷元件自动复检方法,实现了比现有方法更优的自动复检性能。首先,结合深度自编码器与 SMOTE 方法,提出了一种深度自编码 SMOTE 数据增强方法,用于少数类元件数据生成,以平衡数据类别。该方法通过引脚行重组的方式扩充数据集,并以此构建基于自编码器架构的深度重构模

型,从而提高重构模型的泛化能力。为平衡数据类别,将原始的复检缺陷元件数据输入编码器提取低维特征,并通过 SMOTE 方法生成新特征向量,再由解码器将新特征向量重构为新的复检缺陷元件数据。该方法可以有效避免 SMOTE 方法在原始数据上直接进行过采样所面临的噪声引入、无法捕捉原始数据中的复杂非线性结构及难以应对高维数据等问题,同时可以弥补单一自编码器方法缺乏针对性地生成少数类样本的能力。其次,为充分挖掘 SPI 和 AOI 数据中的有效信息,并进一步缓解不平衡数据带来的影响,设计了一种层次特征融合分类模型。该模型考虑元件数据的层次关系和特征间关联性,借助全连接和卷积神经网络分别在引脚级与元件级进行深度特征融合操作,可以提高缺陷元件的复检分类性能。

最后,在 PHME2022 数据集上进行相关实验,验证了本文提出的基于深度自编码 SMOTE 的缺陷元件复检方法的有效性^①。

(2) 基于任务演进 TabNet 训练策略的 PCB 缺陷元件可修复性判定方法

针对缺陷元件可修复性自动判定任务,考虑到数据样本量小的难点,提出了一种基于任务演进 TabNet 训练策略的缺陷元件可修复性自动判定方法,实现了比现有方法更优的自动判定性能。首先,设计了一种元件分层特征提取方法,可以同时引脚层级和元件层级构建和提取特征,充分利用了原始数据信息,并从不同层级全面表征元件的缺陷程度,从而便于模型判定复检缺陷元件的可修复性。同时,该方法所具备的高效特征提取能力,在一定程度上能够减轻小样本数据的影响。其次,基于分层特征,提出任务演进 TabNet 训练策略,以充分缓解小样本数据带来的影响。该策略分为两个阶段:在缺陷感知预训练阶段,利用大量的 AOI 初检缺陷元件数据对 TabNet 进行预训练,从而提升 TabNet 对元件缺陷的鉴别能力;在可修复性认知迁移阶段,调整 TabNet 的输出层结构,并结合小样本复检缺陷元件数据对 TabNet 进行再训练,进一步提升 TabNet 在可修复性判定任务中的性能。通过任务演进 TabNet 训练策略,能够在较大程度上缓和小样本数据所造成的不利影响。

最后,在 PHME2022 数据集上进行相关实验,验证了本文提出的基于任务演进 TabNet 训练策略的缺陷元件可修复性判定方法的有效性。

1.3.2 结构安排

论文共五章,结构安排及创新点如图1.5所示,具体描述如下:

第1章:绪论。介绍了数据驱动的 PCB 缺陷元件自动复检及其可修复性自动判定方法的研究背景与研究意义,以及 PCB 缺陷元件复检问题、PCB 缺陷元

^①PHME2022 数据集是一个公开数据集,收集自 Bitron Spa 公司(机电设备制造领域的领先公司)的实际 SMT 产线,包含 SPI 数据和 AOI 数据。获取链接:<https://github.com/PHME-Datachallenge/Data-Challenge-2022>

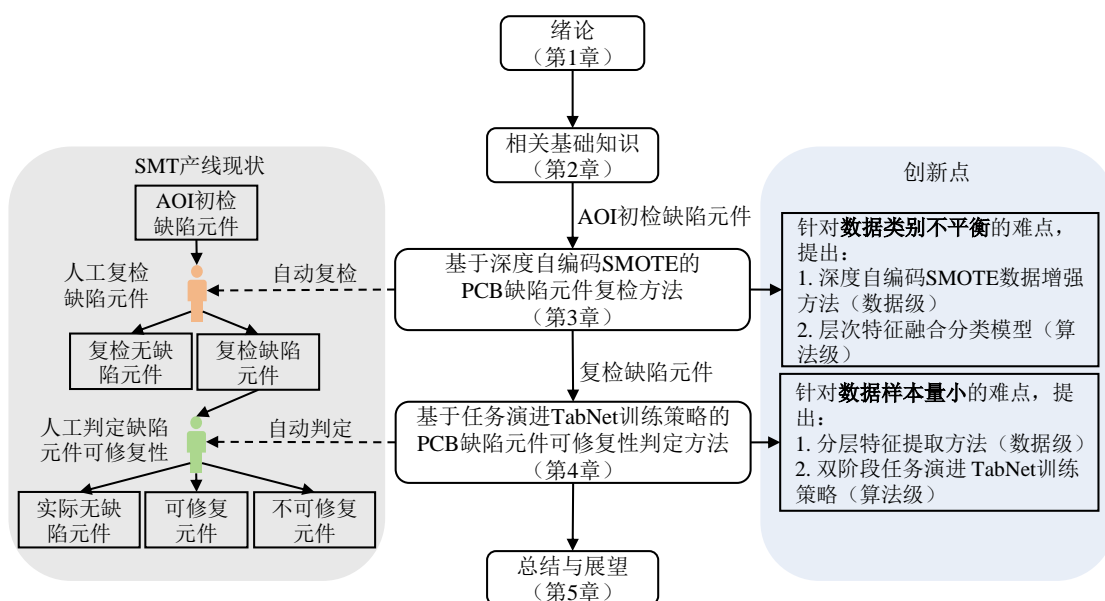


图 1.5 论文的结构安排及创新点（结合 SMT 产线现状说明）

件可修复性判定问题、不平衡数据分类方法和小样本数据分类方法的研究现状，最后给出了全文的研究内容与结构安排。

第 2 章：相关基础知识。首先，介绍了 SMT 工艺；其次，介绍了自编码器的架构和原理；最后，详细介绍了 TabNet 的网络结构和原理。

第 3 章：基于深度自编码 SMOTE 的 PCB 缺陷元件复检方法。首先，分析了标准 SMT 工序中的 SPI 及 AOI 数据特性；其次，设计了一种专用于复检缺陷元件数据生成的深度自编码 SMOTE 方法；接着，提出了层次特征融合分类模型；最后，通过实验验证所提方法的有效性。

第 4 章：基于任务演进 TabNet 训练策略的 PCB 缺陷元件可修复性判定方法。首先，设计了一种分层特征提取方法；其次，提出了双阶段任务演进 TabNet 训练策略；最后，通过实验验证所提方法的有效性。

第 5 章：总结与展望。对全文工作进行总结，分析当前工作的不足，并展望了本文工作在未来进一步的改进方向。

第2章 相关基础知识

本章介绍了本文研究所涉及到的相关基础知识。首先，介绍了 SMT 工艺；其次，介绍了自编码器的网络架构和原理；最后，介绍了 TabNet 模型的网络结构和原理。

2.1 SMT 工艺介绍

SMT 与通孔插装技术（Through-Hole Technology, THT）是 PCB 元件组装的两种主要技术。SMT 工艺将表面贴装元件直接放置在 PCB 表面的焊盘上，并利用回流焊工艺将提前印刷在焊盘上的锡膏熔化，使元件的引脚与 PCB 上的焊盘连接在一起，实现机械固定和电气连接。THT 工艺将电子元件的引脚插入 PCB 上预先钻好的通孔中，然后在 PCB 的另一侧通过焊接工艺，实现元件引脚与 PCB 上铜箔线路的机械固定和电气连接。SMT 工艺与 THT 工艺之间最显著的区别在于元件的组装方式：SMT 工艺采用贴装方式，而 THT 工艺则采用插装方式。图2.1直观展示了这两种 PCB 元件组装方式的不同。

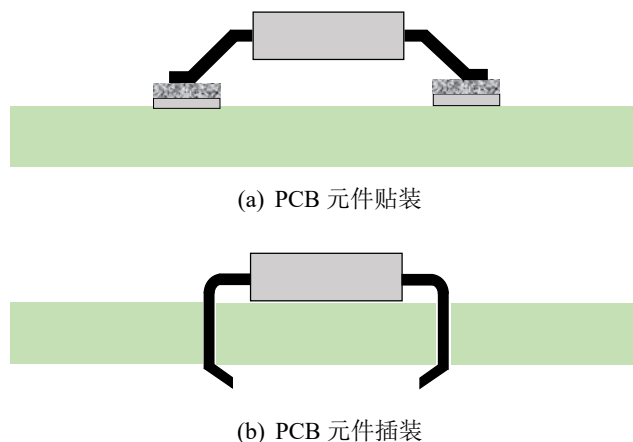


图 2.1 不同的 PCB 元件组装方式

THT 工艺在电子制造行业的早期发展阶段曾发挥关键作用。但随着电子制造行业的进步，电子产品向小型化和高密度方向发展，THT 工艺在生产效率、成本控制和集成度等方面的局限性逐渐显现。相比之下，SMT 工艺由于无需钻孔、可实现双面组装，并具备更高的自动化程度，逐步取代 THT 工艺，成为当前电子制造行业中应用最为广泛的 PCB 元件组装技术。

在第 1 章中，仅对 SMT 工艺的 5 个主要生产环节进行了简要介绍。下面给出 SMT 工艺整个生产流程的介绍：

- 自动上板环节：通过自动化设备，实现 PCB 的自动装载与传输。

- 激光打印环节：运用激光技术，将产品序列号精准地打印在 PCB 表面，以实现产品的追溯与管理。
- PCB 清洁环节：借助离子风设备，对 PCB 进行深度清洁，有效去除表面附着的灰尘颗粒，同时消除静电隐患。
- 锡膏印刷环节：以钢网为媒介，将锡膏均匀且准确地印刷至 PCB 的焊盘之上，确保锡膏在焊盘上的覆盖与分布符合工艺要求，为后续的元件焊接做好准备。
- SPI 环节：采用光学扫描技术，对锡膏印刷质量展开全面检测。检测范围涵盖锡膏的体积、面积、高度以及印刷位置偏移等关键特性，以此保证锡膏印刷质量的可靠性。这一环节依赖人工经验设置检测阈值。统计数据显示，对于通过 SPI 环节的 PCB，仍有 50%-70% 的元件焊接缺陷与锡膏印刷质量有关^[12]。
- 元件贴装环节：依据预设程序，自动化设备将各类元件精准贴装到已完成锡膏印刷的 PCB 上，该环节严格按照设计布局进行元件放置，保障元件贴装的准确性与高效性。
- 回流焊接环节：将 PCB 送入回流炉中，使锡膏受热融化完成焊接。回流炉由加热系统、传输系统和温度控制系统组成，其核心工艺是加热空气或氮气至高温后吹向 PCB，使锡膏熔化并流动，以完成回流焊接。
- 下板缓冲环节：作为生产流水线中的缓冲部分，自动化设备完成 PCB 的传输与临时缓存工作，有效协调生产节奏，确保整个生产线的流畅运行。
- AOI 环节：AOI 机器对完成焊接的 PCB 元件进行光学检测，并报告表面焊接缺陷，该检测过程通常基于模板匹配算法。之后，被 AOI 机器判定存在缺陷元件的 PCB 会被拦截并进入缺陷复检环节，其余 PCB 进入后续功能测试环节。
- 缺陷复检环节：AOI 缺陷复检操作员对 AOI 机器报告的缺陷元件进行目视检查，确认其是否真正存在缺陷。之后，复检缺陷元件所在的 PCB 进入修复环节，其余 PCB 进入后续功能测试环节。
- 修复环节：修复操作员借助放大镜等专业设备，检查并判定复检缺陷元件的可修复性，同时对可修复的复检缺陷元件进行修复。之后，存在不可修复元件的 PCB 会被送入技术中心，其余 PCB 进入后续功能测试环节。技术中心配备专业技术人员及先进设备，会对修复操作员无法修复的元件作进一步处理。
- 功能测试环节：通过施加信号和检查响应来测试 PCB 的功能，可以帮助识别断路或短路等 PCB 功能缺陷。通过实施功能测试，可以验证 PCB 的整体功能，并检测出非表面缺陷。

2.2 自编码器介绍

自编码器是一种无监督神经网络，由编码器和解码器构成^[80]。其中编码器负责将原始样本数据映射至低维编码空间，而解码器则利用这些编码特征重构样本数据^[81-82]。自编码器可学习数据的深层特征并生成高度相似的重构样本。这一特性使其在图像处理和异常检测等多个领域得到广泛应用。同时，基于这一特性，自编码器可有效重建受噪声干扰的原始样本^[83]。

单层自编码器是结构最为简单且最早被提出的一类自编码器，其网络仅包含一个隐藏层。其中，输入层至隐藏层的映射构成编码器，而隐藏层至输出层的映射则构成解码器^[84]。单层自编码器的训练流程如图2.2所示。

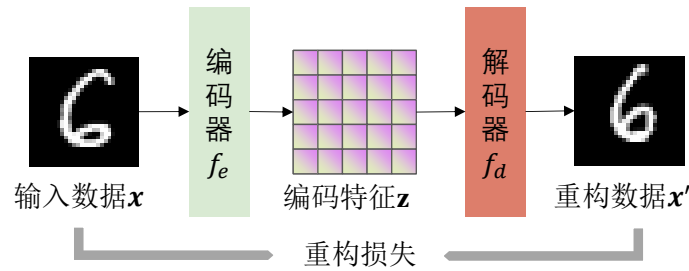


图 2.2 单层自编码器的训练流程

设输入数据为 $\mathbf{x} \in \mathbb{R}^d$ ，编码器 f_e 将其映射为低维编码特征 $\mathbf{z} \in \mathbb{R}^{d'}$ ：

$$\mathbf{z} = f_e(\mathbf{x}) = \sigma(\mathbf{W}_e \mathbf{x} + \mathbf{b}_e) \quad (2.1)$$

其中 $\mathbf{W}_e \in \mathbb{R}^{d' \times d}$ 为编码器权重矩阵， $\mathbf{b}_e \in \mathbb{R}^{d'}$ 为编码器偏置向量， $\sigma(\cdot)$ 为非线性激活函数。

解码器 f_d 基于编码特征 \mathbf{z} 进行数据重构，得到重构数据 $\mathbf{x}' \in \mathbb{R}^d$ ：

$$\mathbf{x}' = f_d(\mathbf{z}) = \sigma(\mathbf{W}_d \mathbf{z} + \mathbf{b}_d) \quad (2.2)$$

其中 $\mathbf{W}_d \in \mathbb{R}^{d \times d'}$ 为解码器权重矩阵， $\mathbf{b}_d \in \mathbb{R}^d$ 为解码器偏置向量。

自编码器的目标是最小化输入数据 \mathbf{x} 与重构数据 \mathbf{x}' 之间的误差，即图2.2中的重构损失^[80]。常用的重构损失函数为均方误差函数。

单层自编码器的网络层数少，难以捕捉到数据中的深层结构信息。为此，可以增加自编码器的层数，增强其特征提取能力。深度自编码器通常指网络层数超过七层的自编码器^[80]。图2.3展示了一个深度自编码器，其中编码器和解码器各自包含四层网络结构。

近年来，研究人员对自编码器展开了深入研究，并衍生出多种变体，如卷积自编码器和变分自编码器等。这些不同类型的自编码器在各类应用场景中展现出独特优势。例如，在卷积自编码器中，为应对高维且分布复杂的样本数据，研究人员引入了卷积层和池化层^[85]。通过卷积和池化操作，编码器可有效降低数

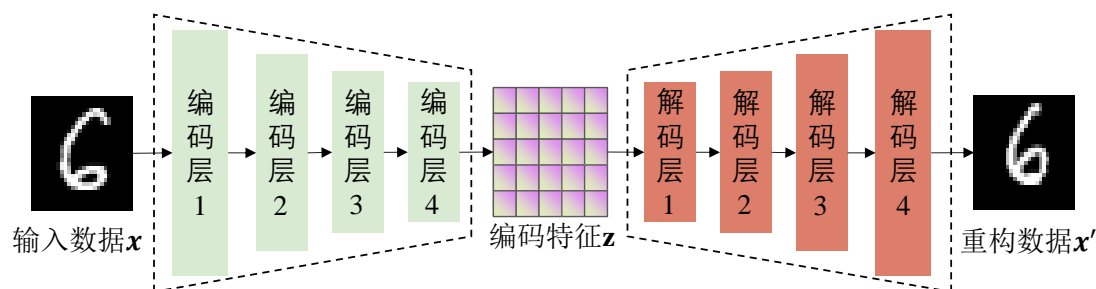


图 2.3 深度自编码器的结构组成与数据处理流程

据的维度，同时保留数据中的关键信息；相应地，解码器通过反池化和反卷积操作对编码特征进行重构，得到重构数据。

2.3 TabNet 模型介绍

TabNet 是一种专为表格数据设计的深度学习模型，由 Google Cloud AI 提出^[86-87]。TabNet 是一种具有稀疏特征选择和可解释性特点的深度学习模型，其网络结构如图2.4所示。

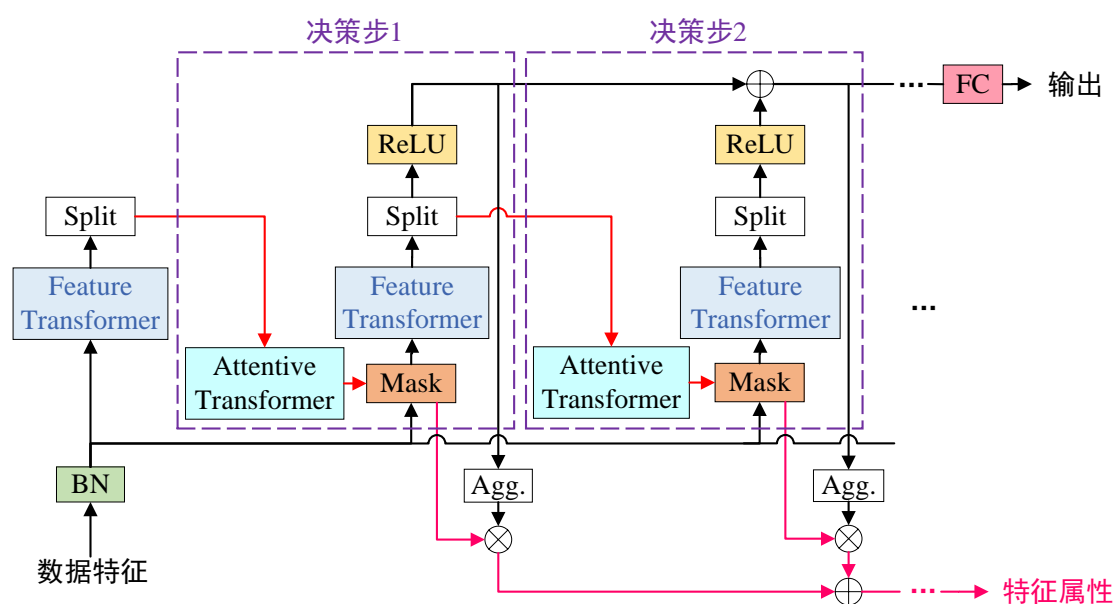


图 2.4 TabNet 网络结构图

TabNet 结合了决策树模型的高效特征选择能力与神经网络的特征表达能力，其核心思想是通过注意力机制对输入特征进行动态选择，从而实现对特征的解釋性建模与高效学习。TabNet 模仿树的决策行为，可以将其视为一个多步神经网络，在每一步采用两个关键操作：通过 Attentive Transformer 模块挑选重要特征；通过 Feature Transformer 模块处理特征。

TabNet 还包含用于表格数据自监督学习的网络模块，但本文仅基于 TabNet 进行监督学习，在此不对自监督学习模块作进一步探讨。

TabNet 凭借其独特的特征选择机制和决策路径解释性，在表格数据处理任务中表现出与现有主流模型相当甚至更优的效果。

接下来对 TabNet 进行详细介绍。

2.3.1 激活函数

激活函数是神经网络的重要组成部分，为网络引入了非线性特性，使模型能够拟合复杂的映射函数。在阐述 TabNet 的详细原理前，有必要对 TabNet 所涉及的几个激活函数展开介绍，为后续理解 TabNet 的网络结构和运行原理奠定基础。

1. ReLU 函数

修正线性单元（Rectified Linear Unit, ReLU）的主要作用是将神经元输出值限制在一个非负的范围。其公式如下所示：

$$\text{ReLU}(x) = \max(0, x) \quad (2.3)$$

根据公式画出 ReLU 函数示意图如图2.5所示。可以观察出 ReLU 函数的特点为：当 $x > 0$ ，输出为 x ；当 $x \leq 0$ ，输出为 0。

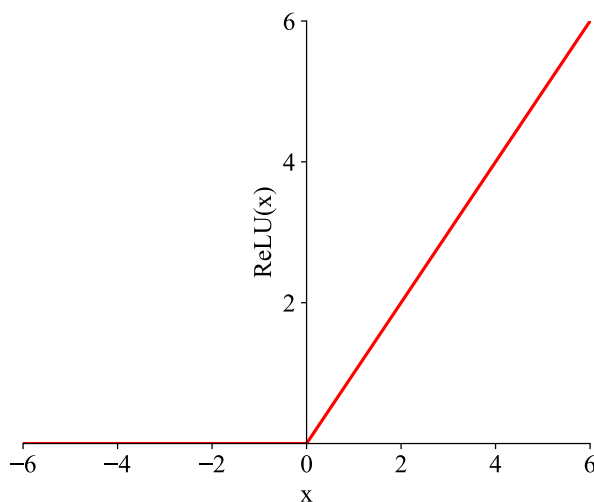


图 2.5 ReLU 函数示意图

2. Sigmoid 函数

Sigmoid 是一种常用的激活函数，其将神经元输出映射到 $(0, 1)$ 的范围，常用于二分类任务的输出层或其他需要归一化的场景。其数学表达式为：

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.4)$$

根据公式画出 Sigmoid 函数示意图如图2.6所示。可以总结得到 Sigmoid 激活函数具有以下特点：输出值范围在 $(0, 1)$ ，可解释为概率值；函数平滑且连续，适用于梯度计算。

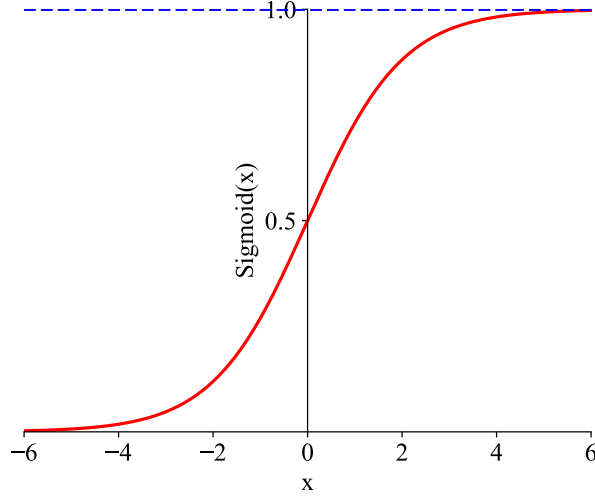


图 2.6 Sigmoid 函数示意图

3. Softmax 函数

Softmax 激活函数主要用于多分类任务中的输出层，用于将神经元输出值转换为概率分布。其公式为：

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (2.5)$$

其中 x_i 表示第 i 个神经元的输出值， n 表示类别的总数。

Softmax 函数将每个神经元的输出值转化为分布于 $(0, 1)$ 区间内的数值，且它们的和为 1。因此，这些数值可看作对应类别的概率。

4. Sparsemax 函数

Sparsemax 是 Softmax 的一种变体，适用于需要稀疏输出的场景。给定输入向量 $\mathbf{z} \in \mathbb{R}^d$ ，其定义为：

$$\text{Sparsemax}(\mathbf{z}) = \arg \min_{\mathbf{p} \in \Delta^{d-1}} \|\mathbf{p} - \mathbf{z}\|^2 \quad (2.6)$$

其中 Δ^{d-1} 是一个 d 维概率单纯形，满足 $\Delta^{d-1} = \{\mathbf{p} \in \mathbb{R}^d \mid p_i \geq 0, \sum_{i=1}^d p_i = 1\}$ 。

Sparsemax 的一种闭式解法如下：

$$\text{Sparsemax}_i(\mathbf{z}) = [z_i - \tau(\mathbf{z})]_+ \quad (2.7)$$

其中 τ 是满足 $\sum_j [z_j - \tau(\mathbf{z})]_+ = 1$ 的函数 ($\tau : \mathbb{R}^d \rightarrow \mathbb{R}$)。

对 \mathbf{z} 进行排序，使得： $z_{(1)} \geq z_{(2)} \geq \dots \geq z_{(K)}$ 。定义 $k(\mathbf{z}) = \max \left\{ j \in \{1, \dots, d\} \mid 1 + jz_{(j)} > \sum_{i=1}^j z_{(i)} \right\}$ 。那么， $\tau(\mathbf{z})$ 可以被表示为：

$$\tau(\mathbf{z}) = \frac{\sum_{i=1}^{k(\mathbf{z})} z_{(i)} - 1}{k(\mathbf{z})} \quad (2.8)$$

2.3.2 特征选择

TabNet 设计了一个可学习的掩码向量 $M[i] \in \mathbb{R}^{B \times D}$ (B 代表批量大小, D 为特征维度), 该掩码向量对应图2.4中的 Mask 模块。掩码向量在每个决策步骤中为特征分配不同的选择系数, 从而使模型能够自动淘汰不相关的特征。这样, TabNet 在每一轮决策都能聚焦于最具信息量的特征, 从而减少冗余信息并加速模型学习。

注意力处理模块即 Attentive Transformer, 是特征选择的核心模块, 其被用于计算获取掩码向量 $M[i]$ 。该模块的网络结构如图2.7所示。

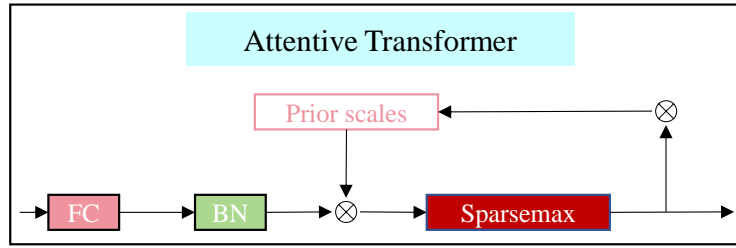


图 2.7 Attentive Transformer 网络结构图

Attentive Transformer 计算掩码 $M[i]$ 的公式如下：

$$M[i] = \text{Sparsemax}(P[i-1] \cdot h_i(a[i-1])) \quad (2.9)$$

其中 i 表示当前决策步, $a[i-1]$ 表示前一个决策步划分出的特征信息, $h_i(\cdot)$ 表示由全连接层和批归一化层组成的可训练函数, $P[i-1]$ 表示先验信息。掩码的规范化条件为: $\sum_{j=1}^D M[i]_{bj} = 1$, 这意味着每个决策步中的掩码向量都是概率分布, 掩码向量中各元素的和为 1, 从而使得特征选择的过程是一个归一化的决策过程。

上式中, 先验信息描述了每个特征在前一决策步中被使用的频率, 其计算方式如下：

$$P[i] = \prod_{j=1}^i (Y - M[j]) \quad (2.10)$$

其中 Y 是松弛参数。直观来说, 如果某特征已在前一决策步中被使用, 则模型应更聚焦于剩余特征, 以避免过拟合^[88]。例如, 当 $Y = 1$ 时, 具有较大激活的特征将具有较小的先验尺度。较小的先验比例确保不会在当前步骤中选择该特征。初始化时, $P[0]$ 是一个全为 1 的向量, 表示没有使用任何特征的先验信息。

同时在 Attentive Transformer 中, 采用 Sparsemax 函数代替了传统的 Softmax 函数, 用于对特征选择进行稀疏化处理。Attentive Transformer 通过 Sparsemax 函数将输入向量中的每个元素限制在 0 到 1 之间, 并确保所有非零元素的和为 1。Sparsemax 的稀疏性有助于模型选择少数几个重要的特征而忽略大部分无关特征, 使得每一轮决策的特征选择更加高效和精确。

通过 Attentive Transformer, TabNet 能够自适应地选择特征进行处理, 使得模型在每一步决策中能聚焦于当前最相关的信息。

2.3.3 特征处理

TabNet 通过构建特征处理模块 Feature Transformer 进行特征处理。Feature Transformer 的网络结构如图2.8所示。

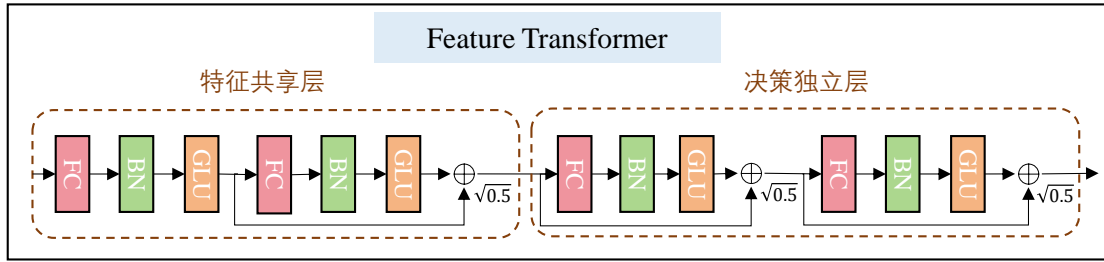


图 2.8 Feature Transformer 网络结构图

由图2.8可见, Feature Transformer 具体由全连接层、批归一化层及门控线性单元 (Gated Linear Unit, GLU) 组成。全连接层和批归一化层是深度学习中的基础网络结构, 已在众多研究中被广泛应用, 因此本章不再赘述其基本原理与实现细节。GLU 是一种改进的激活机制, 是由特定的门控机制实现的, 能够动态调整输入信号的传递, 进而提升模型的性能。其核心思想是通过门控机制选择性地过滤输入特征, 从而学习到更有用的特征表示。

GLU 的计算公式为:

$$\text{GLU}(X) = (X \cdot W + b) \odot \sigma(X \cdot V + c) \quad (2.11)$$

其中 X 表示输入张量, W 和 V 表示权重矩阵, b 和 c 表示偏置向量, \odot 表示逐元素乘法 (也称为哈达玛积), $\sigma(\cdot)$ 特指 Sigmoid 激活函数。

从公式中可以看出, GLU 将输入分为两部分: 一部分为 $(X \cdot W + b)$, 用于传递信号; 另一部分为 $\sigma(X \cdot V + c)$, 用于生成门控值。

总体而言, Feature Transformer 由特征共享层和决策独立层两部分组成^[88]。特征共享层在所有决策步上共同训练, 其参数是共享的。这种设计可以提取出特征的共性, 并使参数更新量更少, 减轻模型的复杂度。决策独立层在每一个决策步上分开训练, 参数独立使得其在每个决策步中可能具有不同的特征处理能力, 特征处理更加有效。对于每一个决策步而言, 其输入特征相同 (掩码层仅屏蔽部分特征, 未改变其余特征), 因此可以用特征共享层处理特征计算的共性部分, 之后再通过决策独立层处理每个决策步的特性部分^[88]。

同时, Feature Transformer 中使用了残差连接, 并在残差连接过程中选择了 $\sqrt{5}$ 的系数, 以稳定训练过程, 防止训练过程中模型参数剧烈变化。

Feature Transformer 处理经过掩码向量过滤后的特征, 其输出会通过 Split 层。

Split 层将 Feature Transformer 输出的向量切成两部分 $a[i]$ 和 $d[i]$ 。其中, $a[i]$ 将被输入到 Attentive Transformer, 用来计算下一个决策步的掩码; 而 $d[i]$ 将用于计算模型的最终输出。最后, 不同决策步的 $d[i]$ 通过 ReLU 激活函数后被相加, 作为最终输出, 如下式所示:

$$d_{\text{out}} = \sum_{i=1}^{N_{\text{steps}}} \text{ReLU}(d[i]) \quad (2.12)$$

2.3.4 可解释性

通过上述 TabNet 中的特征处理模块 Feature Transformer 和特征选择模块 Attentive Transformer, 可以明显看到 TabNet 和多层感知机 (Multilayer Perceptron, MLP) 等普通神经网络的差异。在 MLP 中, 模型对于输入特征是无差别地对待; 而 TabNet 通过学习得到用于特征选择的掩码矩阵, 该掩码矩阵可以体现各个特征在不同决策步中的重要性。TabNet 的特征选择掩码矩阵可以在每个步骤中显示所选特征, 因此可以用于评估决策过程中不同特征的重要性。 $M[i]_{bj}$ 一定程度上可以代表第 b 个样本的第 j 个特征对于当前决策步的贡献度。比如当 $M[i]_{bj} = 0$ 时, 表示第 b 个样本的第 j 个特征对于当前决策步的决策没有任何贡献度。

不过如果想量化总体特征的重要性, 对于不同决策步的掩码, 分别需要一个相应的系数来衡量决策中每个决策步的相对重要性。由式 (2.12) 可知, 不同决策步的输出 $d[i]$ 最终通过线性方式加权在一起作为输出, 其中 $d[i] \in \mathbb{R}^{N_d}$ 。因此, 给出第 b 个样本在第 i 个决策步内的贡献如下, 并作为最终贡献度计算的系数:

$$\eta_b[i] = \sum_{j=1}^{N_d} \text{ReLU}(d_{bj}[i]) \quad (2.13)$$

直观地说, 当 $\eta_b[i]$ 越大, 该决策步对于模型最终决策的贡献也越明显。因此, $\eta_b[i]$ 可以作为第 i 个决策步的权重, 用于对第 i 个决策步的掩码 $M[i]$ 进行加权。而掩码 $M[i]$ 内的权重参数对应了特征的重要性, 因此可以利用如下公式来定义样本 b 中特征 j 的重要性值:

$$M_{\text{agg-bj}} = \sum_{i=1}^{N_{\text{steps}}} \eta_b[i] \cdot M_{bj}[i] \quad (2.14)$$

上式的归一化形式为:

$$M_{\text{agg-bj}} = \frac{\sum_{i=1}^{N_{\text{steps}}} \eta_b[i] \cdot M_{bj}[i]}{\sum_{j=1}^D \sum_{i=1}^{N_{\text{steps}}} \eta_b[i] \cdot M_{bj}[i]} \quad (2.15)$$

鉴于 TabNet 独特的网络结构设计，其具备一般深度学习模型所没有的可解释性优势。通过上述方法，可以计算最终决策过程中不同特征的重要性，为深入理解模型的决策机制以及进一步优化模型提供了关键依据。

2.4 本章小结

本章系统地梳理了本文研究所涉及的相关基础知识。首先，对 SMT 工艺及其生产流程进行了概述；其次，介绍了自编码器的网络架构与工作原理；最后，从激活函数、特征选择、特征处理以及可解释性四个方面，详细解析了 TabNet 模型的网络结构与原理。

第3章 基于深度自编码 SMOTE 的 PCB 缺陷元件复检方法

针对缺陷元件复检问题，考虑到数据类别高度不平衡的难点，本章提出了一种基于深度自编码 SMOTE 的缺陷元件复检方法，实现了比现有方法更优的自动复检性能。首先，深入分析了 PCB 元件的 SPI 及 AOI 数据特性；其次，针对缺陷元件复检问题中的数据类别高度不平衡难点，结合深度自编码器与 SMOTE 方法，提出了一种深度自编码 SMOTE 数据增强方法，旨在扩充少数类复检缺陷元件数据，有效缓解数据不平衡的影响；接着，为了充分挖掘 SPI 和 AOI 数据中的有效信息，并进一步缓解数据不平衡带来的影响，提出了一种层次特征融合分类模型，该模型通过神经网络在引脚层级和元件层级分别进行深度特征融合，并进行复检分类；最后，在 PHME2022 数据集上进行对比实验和消融实验，验证了所提方法的有效性，并分析说明了所提方法在实际产线中的应用可行性。

3.1 引言

在实际 SMT 产线中，AOI 机器存在大量的误报情况，即将大量实际没有焊接缺陷的 PCB 元件误报为存在焊接缺陷。因此，经 AOI 机器初次检测为存在焊接缺陷的元件，需要由 AOI 缺陷复检操作员对其进行实时复检，以确认其是否真实存在焊接缺陷，并根据复检结果将其划分为复检缺陷元件和复检无缺陷元件。在缺陷元件复检流程中，需要大量的人力投入，导致人力成本显著增加，特别是在大规模生产的情况下。同时，AOI 缺陷复检操作员的工作效率会制约产线的生产效率。因此，本章基于标准 SMT 产线的生产过程数据，即 SPI 和 AOI 数据，对 AOI 初检缺陷元件进行自动复检，判断相应元件是否真正存在缺陷，旨在取代实际产线上的 AOI 缺陷复检操作员，以降低相应的人力成本，并提高产线的自动化水平和生产效率。

根据第1章所述，AOI 机器的误检占比可高达 95%，即在 AOI 机器初检出的缺陷元件中，实际无缺陷的元件占比超过 95%。因此，AOI 缺陷复检操作员确认的缺陷元件数量远少于其确认的无缺陷元件数量，表明该任务存在数据类别高度不平衡的难点。针对该任务中的数据不平衡难点，本章结合深度自编码器与 SMOTE 方法，提出了一种深度自编码 SMOTE 数据增强方法，旨在通过扩充少数类复检缺陷元件数据来平衡训练集数据。首先，该方法通过引脚行重组模块扩充原始不平衡训练集。其次，利用扩充的不平衡训练集训练一个由编码器和解码器构成的数据重构模型。其中，编码器负责从原始数据中提取低维特征表示，有

效去除数据中的冗余信息和噪声；而解码器则负责将这些特征重构为原始元件数据。接着，将原始不平衡训练集中的复检缺陷元件数据输入编码器，获得编码空间内的低维特征，并利用 SMOTE 方法在该编码特征空间进行过采样。最后，解码器将 SMOTE 生成的新特征解码为新的复检缺陷元件样本，从而得到类别平衡的训练集。

在此基础上，为了充分挖掘 SPI 和 AOI 数据中的潜在有效信息，并进一步缓解不平衡数据带来的影响，本章提出了一种层次特征融合分类模型。该模型通过神经网络，在引脚层级和元件层级分别进行深度特征融合，以实现更精确的复检分类。具体而言，所提出的特征融合分类模型包括三个主要模块，分别是：共享引脚特征重构器、元件缺陷特征提取器和元件缺陷评估器。其中，共享引脚特征重构器的功能是对每个引脚的 SPI 特征和 AOI 特征进行融合和重构，旨在从原始数据中提取有效信息，并去除噪声与信息冗余；元件缺陷特征提取器在元件层级对所有引脚的重构特征进行融合，提取出描述整个元件缺陷特性的全局特征；元件缺陷评估器接收元件缺陷特征提取器输出的全局特征，综合评估元件的整体焊接质量并输出最终的复检分类结果。

本章结构安排如下：第 3.2 节给出了本章缺陷元件复检问题的描述；第 3.3 节深入分析了 SPI 和 AOI 数据的特性；第 3.4 节给出了包括数据预处理、深度自编码 SMOTE 数据增强方法与层次特征融合分类模型在内的方法设计；第 3.5 节给出了相关的实验设计与结果分析；第 3.6 节给出了本章研究内容的总结。

3.2 问题描述

本章针对 AOI 初检缺陷元件，开展缺陷元件自动复检工作。通过给定 AOI 初检缺陷元件的 SPI 数据和 AOI 数据，利用算法对元件进行复检分类，判断其为复检无缺陷元件或复检缺陷元件。该任务的核心目标是对 AOI 初检缺陷元件进行焊接质量的自动复检评估，从而尝试替代现有产线上的 AOI 缺陷复检操作员，提高产线的自动化水平和生产效率。

具体而言，本章基于给定的 AOI 初检缺陷元件的 SPI 和 AOI 数据，构建一个深度学习分类模型，用于缺陷元件的自动复检。给定某元件的数据表示为 $\mathbf{x} = [\mathbf{x}_{\text{SPI}}, \mathbf{x}_{\text{AOI}}]$ ，其中 \mathbf{x}_{SPI} 表示该元件的 SPI 数据特征， \mathbf{x}_{AOI} 表示该元件的 AOI 数据特征。分类模型的目标是构建一个函数 $f(\cdot)$ ，该函数通过学习输入数据与输出类别之间的映射关系，实现对元件类别的复检分类：

$$\mathbf{y} = f(\mathbf{x}) \quad (3.1)$$

其中 $\mathbf{y} \in \{0, 1\}$ 表示模型的输出（即分类结果），类别 0 表示复检无缺陷元件，类别 1 表示复检缺陷元件。

在实际应用中，复检缺陷元件和复检无缺陷元件的分类并不依赖于简单的基于规则的阈值判断，而是涉及对 SPI 和 AOI 数据中高维复杂特征的深度挖掘和融合。特别是，SPI 数据提供了锡膏印刷质量信息，而 AOI 数据反映了焊点表面的光学特征。二者在特征空间中共同决定了元件焊接质量的分类边界。

3.3 SPI 与 AOI 数据特性分析

本节系统性地介绍并分析 SPI 数据与 AOI 数据的特性，为后续研究奠定基础。在介绍和分析 SPI 和 AOI 数据的特性之前，有必要先了解 PCB 拼板技术的基础概念。PCB 拼板技术是一种广泛应用于 SMT 生产线的关键工艺，其核心思想是将多个独立 PCB 组合成一个 PCB 面板，以提升生产效率并优化制造工艺。通常，每个 PCB 面板由多个相同结构的独立 PCB 组成，面板上独立 PCB 的具体数量根据产品设计与生产需求而定。例如，图3.1 展示了一个典型的 PCB 面板，其由 4 个独立 PCB 拼接而成。

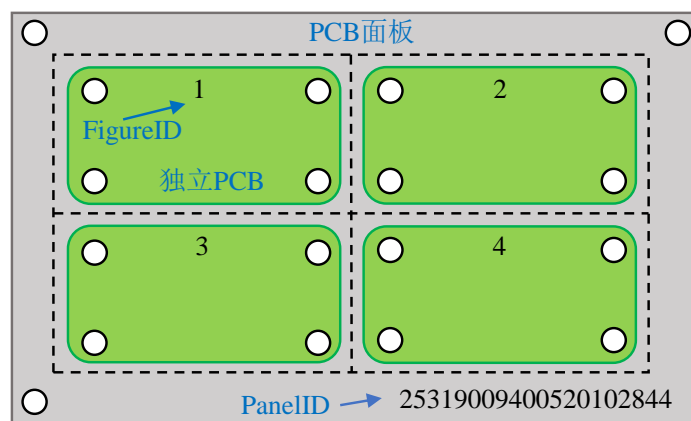


图 3.1 PCB 面板示意图

为了方便管理和追踪，PCB 拼板技术采用了层次化的标识系统。每个面板被分配一个唯一的标识符，称为 PanelID，用于不同面板的区分。面板上的每个 PCB 被分配到一个标识符，称为 FigureID。每个 PCB 则通过 PanelID 和 FigureID 的组合进行唯一标识。进一步地，每个 PCB 上的元件通过标识符 ComponentID 进行单独标识，而元件上的每个引脚则通过标识符 PinNumber 加以区分。因此，利用三元组标识符（PanelID, FigureID, ComponentID），可以唯一定位每个元件；而每个元件的引脚则可通过四元组标识符（PanelID, FigureID, ComponentID, PinNumber）进行精确定位。

这种层次化标识系统不仅有助于在制造过程中进行精准定位和追踪，还为数据的层次化管理奠定了基础。通过该层次化标识系统，可以将生产和检测数据精确地映射到特定的面板、PCB、元件和引脚层级，从而为后续的质量分析和故障诊断提供了可靠的数据支撑。

3.3.1 SPI 数据特性

表3.1列出了 SPI 数据中包含的关键特性及其描述。

表 3.1 SPI 数据特性

名称	单位	描述
PanelID	-	面板标识符
FigureID	-	PCB 标识符
ComponentID	-	元件标识符
PinNumber	-	元件引脚编号
PadID	-	焊盘编号
Date	-	执行检测的时间，具体到天
Time	-	执行检测的时间，具体到秒
PosX	mm	从左下角开始的焊盘的 X 方向坐标
PosY	mm	从左下角开始的焊盘的 Y 方向坐标
Height	um	锡膏的高度
Volume	um ³	锡膏的体积
Area	um ²	锡膏的面积
Shape	um	锡膏形状规则的变化
SizeX	mm	锡膏在 X 方向的长度
SizeY	mm	锡膏在 Y 方向的长度
VolumeP	%	锡膏体积占目标体积的百分比
AreaP	%	锡膏面积占目标面积的百分比
OffsetXP	%	锡膏在 X 方向与预定义位置的差异
OffsetYP	%	锡膏在 Y 方向与预定义位置的差异

SPI 机器对 PCB 焊盘上所印刷锡膏的质量进行精确检测，进而生成详尽的引脚级 SPI 数据。在数据生成过程中，每条 SPI 记录通过一套层次化标识符体系进行关联和区分，主要包括 PanelID、FigureID、ComponentID、PinNumber 和 PadID。通过这些标识符，能够将每条 SPI 记录精准对应到特定的元件引脚或焊盘位置。因此，SPI 数据中包含 PanelID、FigureID、ComponentID、PinNumber 和 PadID 等关键信息。

与此同时，SPI 数据记录了一系列与锡膏质量相关的关键特性。具体而言，这些特性包括锡膏的体积、面积、高度和偏移量等。这些特性相互关联，从多个维度完整地描绘出锡膏印刷的状态，能够全面精准地反映锡膏的印刷质量，为后续的元素焊接质量评估提供了数据支撑。

为探索 SPI 数据集中描述锡膏质量特性的连续特征间的关系，在此以 PHME2022 数据集为例，使用皮尔逊相关系数来衡量其中任意两特征之间的相关程度。记特征 x 与特征 y 之间的皮尔逊相关系数为 r ，其计算公式如下：

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.2)$$

其中 n 表示数据点数量, x_i 和 y_i 分别表示两特征在第 i 个数据点的值, \bar{x} 和 \bar{y} 分别表示两特征的平均值。

皮尔逊相关系数 r 的取值范围为 $[-1, 1]$ 。当 $r < 0$ 时, 表示两个特征间关系呈负相关; 当 $r > 0$ 时, 表示两个特征间关系呈正相关。相关系数 r 的绝对值越大, 表示两个特征间的相关性就越强。

本文根据皮尔逊相关系数计算结果, 绘制了如图3.2所示的特征相关性热图, 来直观检查不同 SPI 特征间的相关性。

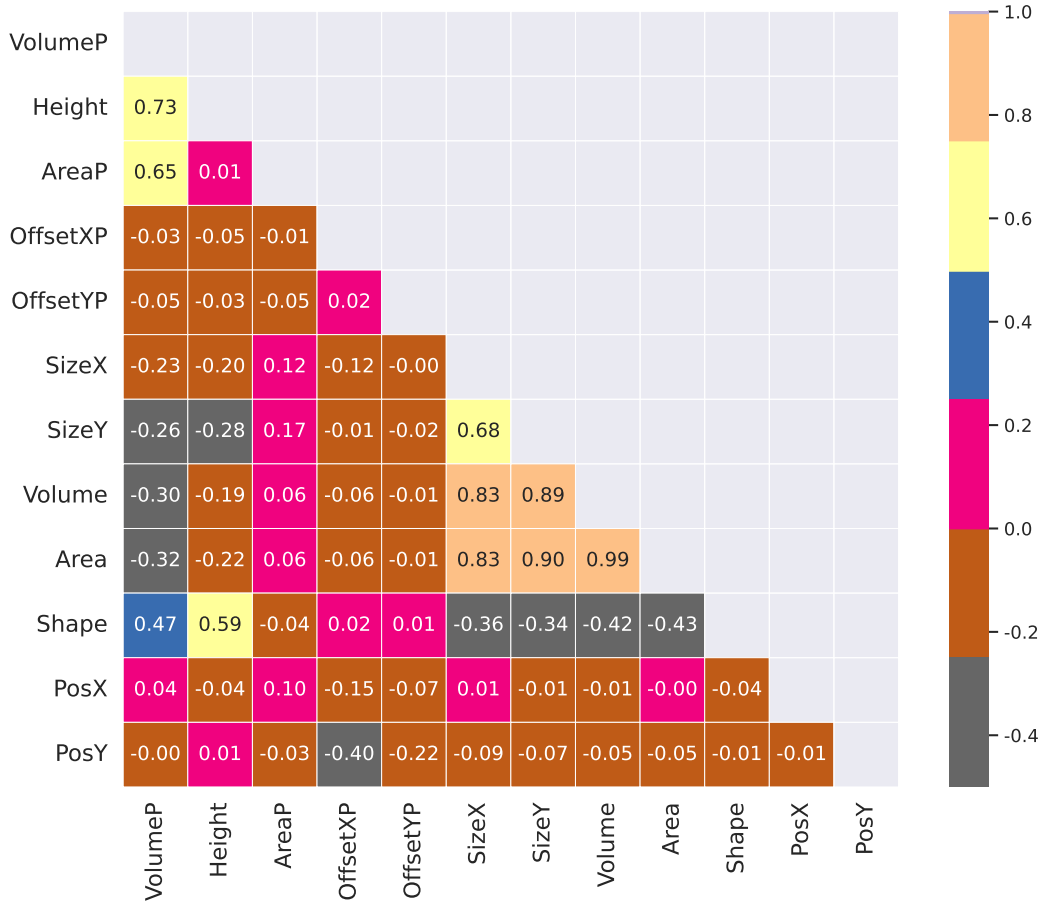


图 3.2 SPI 特征相关性可视化热图

值得注意的是, 在特征相关性可视化热图中, 特征 SizeX、SizeY、Volume 和 Area 之间呈现出较高的相关系数, 表明这些特征之间存在显著且稳健的相关性。通过考虑锡膏的固有几何特性, 可以更深入地理解这种现象。从本质上讲, SizeX 表示锡膏在 X 维度的长度, SizeY 表示锡膏在 Y 维度的长度, Area 表示锡膏的面积。理论上, 对于印刷在矩形焊盘上的锡膏, Area 应等于 SizeX 与 SizeY 之积。因此, Area 与 SizeX 或 SizeY 之间均表现出显著的正相关性。鉴于体积本质上是面积向第三维的延伸, 描述锡膏体积特性的特征 Volume 与 Area 也必然高度相关。若把锡膏视为理想的矩形柱体, 锡膏的体积 Volume 理论上等于 SizeX、SizeY 与锡膏高度之积。因此, 如果 SizeX 或 SizeY 发生变化, Volume 也会对应

变化，从而表现出显著的正相关性。

同时，通过上述分析，本文认为锡膏的体积与高度之间应存在较强的正相关性。然而，进一步观察图3.2，发现描述锡膏体积的特征 **Volume** 与描述锡膏高度的特征 **Height** 之间的皮尔逊系数为-0.19，并未表现出较强的正相关性。接着，继续分析特征 **VolumeP** 和 **Height** 之间的相关性，发现二者之间的皮尔逊系数为0.73，表现出较强的正相关性。**VolumeP** 表示的是锡膏体积占目标体积的百分比。因此，推测 **Volume** 与 **Height** 之间未呈现出正相关性的原因可能是由于不同元件间锡膏目标体积的差异所导致。而锡膏目标体积的差异取决于元件类型的差异。因此，本文推测特征相关性可视化热图中 **Volume** 与 **Height** 之间未呈现出显著相关性的原因，可能源于元件类型的差异。对于同一类型的元件，**Volume** 与 **Height** 之间可能会呈现出较强的正相关性。

为验证上述推测，本文进一步分析了 SPI 数据中特征 **Volume** 与 **Height** 之间的关系。具体而言，本文利用 SPI 数据中的 **Volume** 和 **Height** 特征列，绘制了如图3.3所示的特征关系散点图。

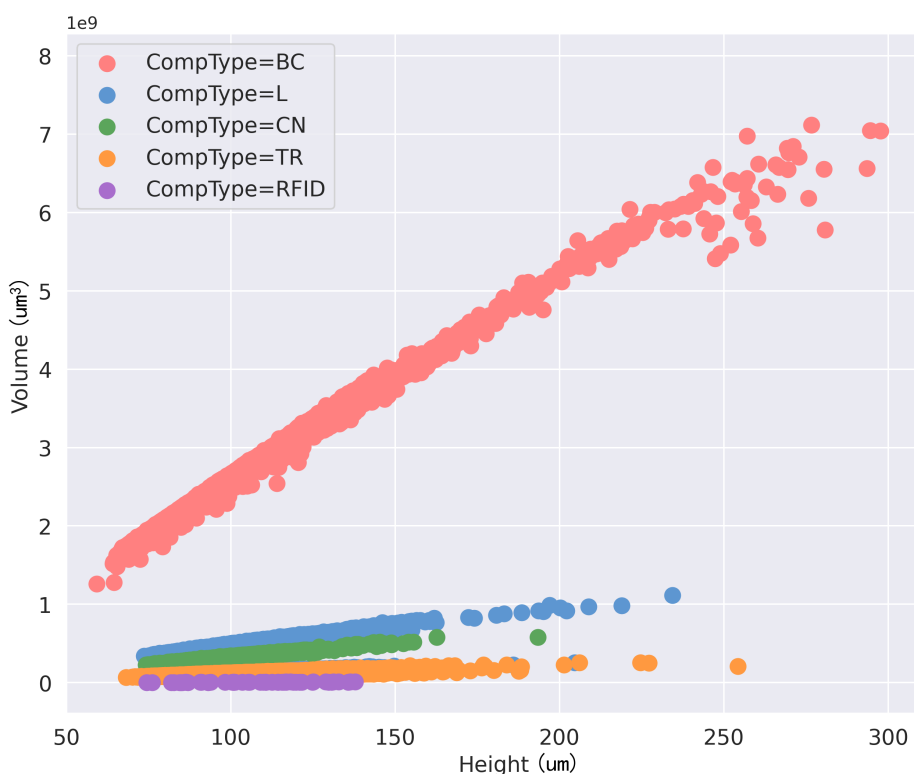


图 3.3 **Volume** 与 **Height** 特征关系散点图

图3.3以 **Height** 为横坐标，以 **Volume** 为纵坐标。在绘制该特征关系散点图时，对数据中的元件标识符 **ComponentID** 提取字母部分作为其元件类型。例如，对于 **ComponentID** 为 BC1、BC2、BC3 和 BC4 的元件，统一提取元件类型为 BC。为了便于观察，图中仅展示了 5 种元件类型，分别为 BC、L、CN、TR 和 RFID。关于元件类型的具体描述可查看 PHME2022 原始数据集，在此不进行赘述。同时，

为便于直观比较不同元件类型之间的锡膏特性差异，本文使用了不同颜色标识不同元件类型的散点数据。通过观察图3.3可以发现，对于同一类型元件，Volume 与 Height 之间呈现出显著的线性正相关性，这一结果验证了前述推测。

此外，通过图3.3可以观察到，不同类型元件的 Volume 与 Height 之间呈现出不同的线性趋势。基于锡膏体积与高度及面积之间的关系，本文推测不同类型元件引脚对应焊盘上的锡膏面积存在较大差异。为验证这一推测，本文利用 SPI 数据中的 Volume 与 Area 特征列，绘制了如图3.4所示的散点图。

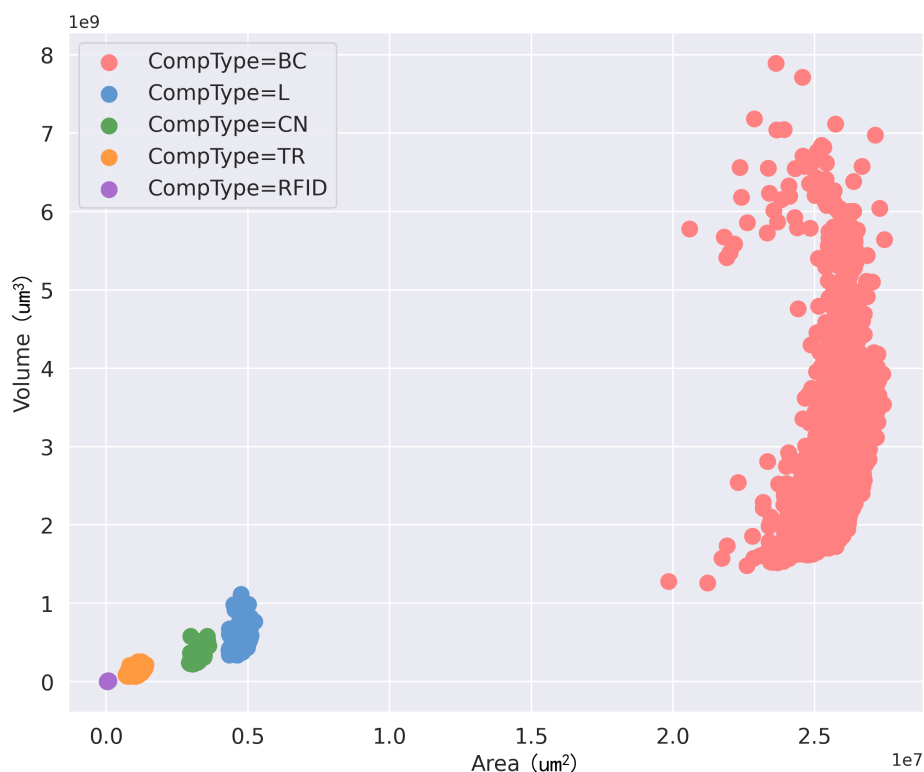


图 3.4 Volume 与 Area 特征关系散点图

图3.4中同样展示了 BC、L、CN、TR 和 RFID 这 5 种不同类型元件的特征关系散点图。通过观察图3.4可以发现，不同类型元件对应的锡膏面积 Area 差异显著，而同一类型元件对应的锡膏面积则维持在一定的范围内。这一现象再次验证了本文推测。

3.3.2 AOI 数据特性

在回流焊接环节后，AOI 机器对焊接后的 PCB 元件进行焊接缺陷检测，并针对检测出的缺陷生成详细的 AOI 数据。AOI 机器利用视觉技术可以识别多种表面焊接缺陷，如焊点桥接和焊点偏移等。与 SPI 数据相同，不同的 AOI 数据记录均通过层次化标识符体系进行关联和区分。该标识符体系包括 PanelID、FigureID、ComponentID 和 PinNumber，可以确保每条记录能准确对应到特定的元件及其引脚。除了这些标识符外，AOI 数据中还包含具体的缺陷标签 AOILabel，用于进

一步描述检测到的缺陷类型。

AOI 数据中具体包含的关键特性及其描述可见表3.2。

表 3.2 AOI 数据特性

名称	描述
PanelID	面板标识符
FigureID	PCB 标识符
ComponentID	元件标识符
PinNumber	元件上的引脚编号
MachineID	执行自动光学检测的 AOI 机器标识符
AOILabel	AOI 机器根据检测结果分配的缺陷标签 一般常见分配的标签为： <ul style="list-style-type: none"> • 焊点桥接 • 焊点偏移 • 锡膏不足 • 锡膏过量 • 共面问题 • 元件缺失 • 极性相反 • 立碑效应

在元件表面焊接缺陷检测过程中，AOI 机器针对每个检测到的元件缺陷都会生成一条 AOI 数据。直观而言，尽管 AOI 机器存在较高的误报率，但当同一元件被报告的 AOI 缺陷数量增加时，该元件实际存在缺陷的可能性也会相应提高，其在复检过程中被判定为缺陷元件的概率应当更大。为验证这一推测，本文基于 PHME2022 数据集，绘制了 AOI 缺陷数量与元件复检类别的关联性分析图（如图3.5所示）。

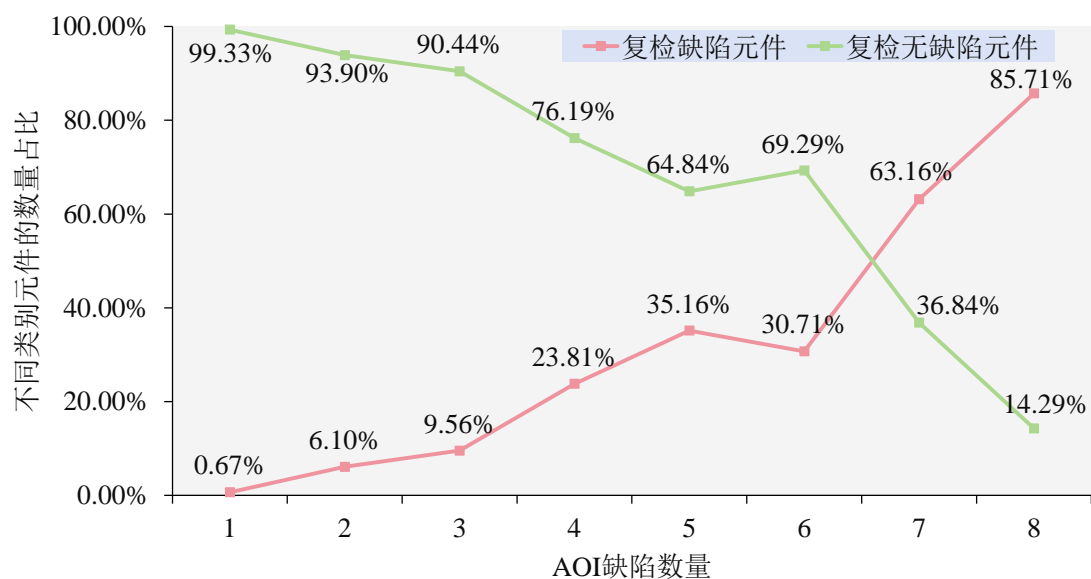


图 3.5 AOI 缺陷数量与元件复检类别的关联性分析图

在图3.5中，横坐标表示同一元件上 AOI 机器报告的缺陷数量，纵坐标则表示在具有相同 AOI 缺陷数量的元件中，复检缺陷元件和复检无缺陷元件分别占该类元件总数的比例。

通过图3.5可观察到，在 AOI 缺陷数量仅为 1 的元件中，超过 99% 的元件为复检无缺陷元件，而复检缺陷元件的比例极低。相比之下，当 AOI 缺陷数量增加至 8 时，超过 85% 的元件为复检缺陷元件，而复检无缺陷元件的比例显著降低。

进一步观察图中的整体趋势可以发现，随着 AOI 缺陷数量的增加，复检缺陷元件的比例整体呈上升趋势，这表明元件被判定为复检缺陷元件的概率随之增加。该结果验证了前述推测，即 AOI 机器报告的缺陷数量在一定程度上反映了元件实际存在缺陷的可能性。

总体而言，本节分别介绍了 SPI 数据和 AOI 数据的特性，并明确了其中所包含的有效特征，为本章及第 4 章的方法设计奠定了基础。通过对 SPI 数据的深入分析，发现部分数值特征之间存在较强的相关性。这种相关性并非偶然，而是反映了锡膏不同特性在物理层面上的内在联系。这一发现表明，在使用这些特征时，应保持谨慎，并充分考虑特征间的内在关联性，以避免冗余信息或潜在的误导性影响。此外，本节分析还表明，元件类型是影响特征分布的重要因素，不同类型元件的 SPI 特征呈现出不同的数据分布特性。同时，基于 AOI 数据提取的同一元件缺陷数量在一定程度上能够反映元件的缺陷程度。

综上所述，本节对数据特性的深入分析不仅为本章及第 4 章的特征选择与计算提供了重要指导，同时也为分类模型的设计提供了有力支持。

3.4 PCB 缺陷元件复检方法设计

针对缺陷元件复检任务，本章提出了一种基于深度自编码 SMOTE 的 PCB 缺陷元件复检方法，其整体框架如图3.6所示。

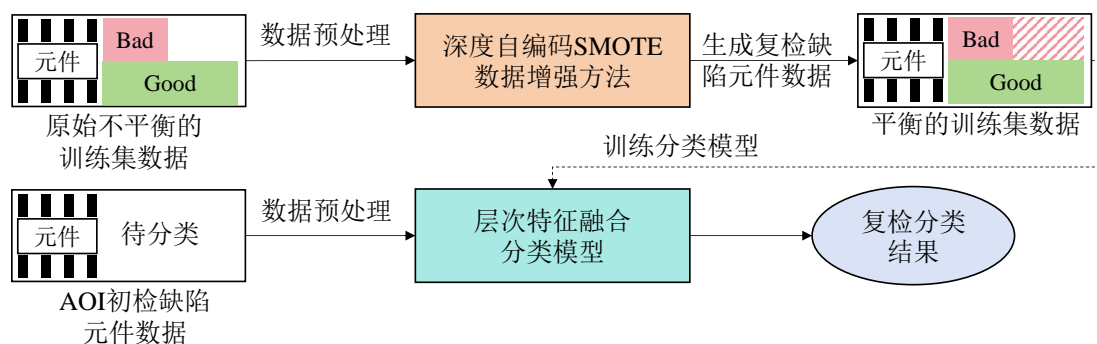


图 3.6 PCB 缺陷元件复检方法整体框架

所提方法首先对原始的不平衡数据集进行预处理，之后通过深度自编码 SMOTE 数据增强方法生成少数类复检缺陷元件数据，以得到平衡数据集。接

着，基于平衡数据集训练层次特征融合分类模型，使其能够对 AOI 初检缺陷元件进行复检分类。

本节将从数据预处理、深度自编码 SMOTE 数据增强方法以及层次特征融合分类模型三个方面，对所提方法进行详细介绍。

3.4.1 数据预处理

1. 引脚级多源数据融合

如前所述，SPI 和 AOI 数据分别获取自 SMT 产线的锡膏检测和自动光学检测两个不同的生产环节。因此，从 SMT 产线中直接获得的 SPI 和 AOI 数据来源不同，具有异质性，属于多源数据。同时，其在存储形式上是相互独立的，分别存储在不同的文件中。

尽管 SPI 和 AOI 数据来源于不同的生产环节，属于多源数据，但它们均在引脚级进行记录，并且采用第 3.3 节所述的层次化标识符体系。具体来说，SPI 和 AOI 数据集中的每一条记录均包含标识符 PanelID、FigureID、ComponentID 和 PinNumber。因此，可以将具有相同四元组标识符 (PanelID, FigureID, ComponentID, PinNumber) 的 SPI 数据行和 AOI 数据行进行合并，从而在引脚层级实现 SPI 和 AOI 数据的融合。本章复检任务针对的是被 AOI 机器初检为缺陷的元件，即 AOI 数据中记录的元件。因此，在数据处理过程中，需要忽略仅在 SPI 数据中出现的元件。

图3.7展示了引脚级多源数据融合的示意图。图中，不同引脚对应的数据行采用不同颜色进行标注。

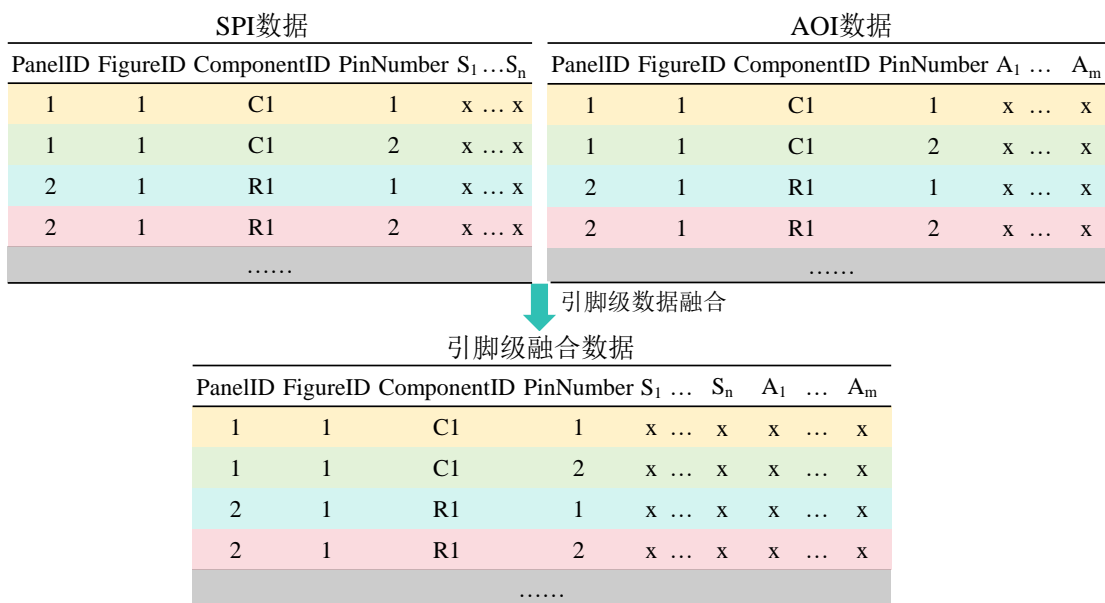


图 3.7 引脚级多源数据融合示意图

引脚级多源数据融合算法的流程如算法3.1所示。该算法通过数据集中的层

次化标识符，逐一遍历每个元件引脚的 SPI 和 AOI 数据条目，并在引脚层级对其进行融合。

算法 3.1 引脚级多源数据融合算法

输入: SPI 数据集 $DataS$

AOI 数据集 $DataA$

输出: SPI 与 AOI 数据的引脚级融合数据集 $DataSP$

```

1 从  $DataA$  提取 PanelID 列所有唯一值的集合  $P$ 
2 for  $p \leftarrow P$  do
3   从  $DataA$  提取 PanelID 值为  $p$  的数据条目  $DataA_p$ 
4   从  $DataS$  提取 PanelID 值为  $p$  的数据条目  $DataS_p$ 
5   从  $DataA_p$  提取 FigureID 列所有唯一值的集合  $F$ 
6   for  $f \leftarrow F$  do
7     从  $DataA_p$  提取 FigureID 值为  $f$  的数据条目  $DataA_f$ 
8     从  $DataS_p$  提取 FigureID 值为  $f$  的数据条目  $DataS_f$ 
9     从  $DataA_f$  提取 ComponentID 列所有唯一值的集合  $C$ 
10    for  $c \leftarrow C$  do
11      从  $DataA_f$  提取 ComponentID 值为  $c$  的数据条目  $DataA_c$ 
12      从  $DataS_f$  提取 ComponentID 值为  $c$  的数据条目  $DataS_c$ 
13      从  $DataS_c$  提取 PinNumber 列所有唯一值的集合  $N$ 
14      for  $n \leftarrow N$  do
15        从  $DataA_c$  提取 PinNumber 值为  $n$  的数据条目  $DataA_n$ 
16        if  $DataA_n$  为空 then
17          | 用 0 填充  $DataA_n$  中的相应 AOI 特征
18        end
19        从  $DataS_c$  提取 PinNumber 值为  $n$  的数据条目  $DataS_n$ 
20         $DataSP.Insert(MergeFeatures(DataA_n, DataS_n))$ 
21      end
22    end
23  end
24 end

```

2. 类别特征编码

在本研究中，所涉及的 SPI 和 AOI 数据包含的特征可分为两类：数值特征和类别特征。对于类别特征，由于它们本身以离散的类别形式存在，因此需要将其转换为数值特征，以便能够被后续的深度学习模型有效处理。

编码是这一转换过程中的关键步骤，其主要目的是将非数值型的类别特征转化为数值形式，从而使得模型能够在训练过程中识别和处理这些信息。数据编码是机器学习和深度学习领域中的常见操作，大多数模型（尤其是基于数学计算的模型）要求输入数据为数值型。特别是深度学习模型，它们通常通过权重和激活函数等数值计算方式进行学习和推理，因此输入数据必须符合数值化的要求。类别特征通过合适的编码方式转化为数值形式后，模型便可以对其进行有效的学习，从而捕捉其中的模式和规律。

本章方法采用的有效类别特征为 `CompType` 和 `AOILabel`。其中 `CompType` 特征表示元件类型，是本章从元件标识符 `ComponentID` 中提取的有效信息。例如，对于 `ComponentID` 为 BC1、BC2、BC3 和 BC4 的元件，统一提取元件类型 `CompType` 为 BC。`AOILabel` 是 AOI 数据中描述缺陷类型的特征。在本研究中，针对类别特征 `CompType` 和 `AOILabel`，分别采用了不同的编码方法，以确保模型能够有效地处理各类数据。

首先，对于 `CompType` 特征，采用了 `CatBoost` 编码方法。`CatBoost` 编码是一种类别特征编码方法，它能够有效捕捉类别特征与目标变量之间的潜在关系。与传统的编码方法不同，`CatBoost` 编码不仅考虑类别出现的频率，还结合目标变量与类别变量之间的关系，通过训练模型来对类别进行转换，从而生成更加紧凑且富有信息的特征表示。与下述独热编码方法相比，`CatBoost` 编码方法能够避免高维稀疏性问题，提升特征表达的有效性。

接着，对于 `AOILabel` 特征，采用了独热编码方法。由于某些引脚可能会同时被检测出多个焊接缺陷，`AOILabel` 特征中会同时出现多个类别。独热编码方法适用于类别之间独立且没有顺序关系的情况。独热编码将类别特征转换为一个独立的二进制向量，其中某类别存在时，对应位置的值设置为 1，否则为 0。由于 `AOILabel` 特征中会同时出现多个缺陷类别，独热编码能够有效地将每个引脚对应的多个缺陷类别进行表示，从而为后续的模型学习提供合适的输入格式。值得注意的是，`AOILabel` 特征不适合采用 `CatBoost` 编码方法。原因在于，`AOILabel` 特征中会同时出现多个缺陷类别，即每个元件引脚可能同时被检测出多个缺陷类别，而 `CatBoost` 编码方法通常假设每个类别的出现是互斥的。

总体而言，本章根据不同类别特征的特点，分别采用了 `CatBoost` 编码和独热编码两种方法，以确保特征能够被有效地输入到后续的深度学习模型中，从而提升模型的性能。

3. 数据组织结构

在实际中，同一 PCB 上通常包含不同引脚数量的元件，而每个引脚的 SPI 和 AOI 数据的总特征数量是固定的。因此，不同引脚数量的元件会具有不同维度的原始特征。而深度学习算法通常要求输入特征具有相同的维度。因此，为确

保持特征能够有效输入到深度学习模型，并充分利用元件的原始特征，本章根据元件的引脚数量对其进行分类处理，针对相同引脚数量的元件训练相同的独立模型。

对于元件上每个引脚的特征，除上述类别特征外，本章仅保留 SPI 数据中用于描述锡膏质量的连续数值特征。接着，针对某特定引脚数量的元件，本章按照引脚行顺序排列的方式组织该元件的数据。比如针对引脚数量为 n 的元件，假设其每个引脚的特征总数量为 m ，本章将数据组织为 $n \times m$ 的二维结构，如图3.8所示。数据的每一行表示该元件特定引脚的焊接质量特征，数据的每一列表示元件不同引脚的相同特征。图中相邻行对应的引脚在物理上也相邻。

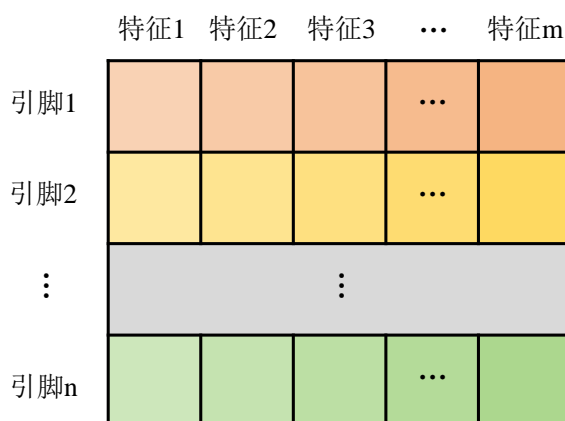


图 3.8 数据组织结构

4. 特征缩放

特征缩放是机器学习中至关重要的预处理步骤，用于消除数据集中不同特征间量纲差异的影响，这可以显著提高许多机器学习和深度学习算法的性能。鉴于本章所用数据的特征间有不同的量纲和相差较大的取值范围，因此这里需要采用特征缩放方法对数据进行处理。

特征缩放方法有很多种，但合适的特征缩放方法与数据的潜在分布、异常值的存在以及所用模型的具体要求有关。这里简要总结以下三种众所周知且广泛使用的特征缩放方法，并分析其中适用于本文数据的特征缩放方法。

(1) Z-score 缩放

Z-score 缩放，是一种用于缩放数据特征值的预处理技术，其缩放后的数据服从平均值为 0、标准差为 1 的分布。该方法广泛应用于对输入特征尺度敏感的机器学习算法，如支持向量机和 k 近邻算法。Z-score 缩放确保所有特征对算法中的距离计算都有同等的贡献，这可以在学习过程中带来更好的性能和稳定性。在数据近似正态分布时，该方法效果好。然而，这种方法很容易受到异常值的影响，因为它依赖于数据的均值和标准差，而异常值可能会对其产生重大影响。这会导致数据被缩放到一个小范围内，而异常值主导了缩放，从而影响了转换后的

数据。使用这种方法进行缩放的公式如下：

$$x_{scaled} = \frac{x - u}{\sigma} \quad (3.3)$$

其中 x 表示原始特征值， u 表示 x 所在特征列的平均值， σ 表示 x 所在特征列的标准偏差。

(2) Min-Max 缩放

这种方法通过线性变换来缩放数据，缩放后的数据通常在 $[0, 1]$ 范围内，有时在 $[-1, 1]$ 范围内。在对输入特征尺度敏感的算法中，Min-Max 缩放方法尤其有益。通过将所有特征调整到统一的尺度，该方法确保没有任何单个特征仅仅因为其数值的大小而主导模型的训练过程。因此，这有助于重新调整学习过程，通常会提高模型的性能。另一方面，该方法对异常值特别敏感，因为异常值会改变最小值和最大值，从而影响缩放。Min-Max 缩放的公式为：

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.4)$$

其中 x_{min} 和 x_{max} 分别表示 x 所在特征列的最小值和最大值。

(3) Robust 缩放

Robust 缩放方法利用中位数和四分位数间距对特征进行缩放，使其能够抵抗异常值。当数据中包含可能扭曲缩放过程的极值时，Robust 缩放的效果明显。采用这种方法计算的缩放值如下：

$$x_{scaled} = \frac{x - median}{IQR} \quad (3.5)$$

其中 $median$ 表示 x 所在特征列的中位数， IQR 表示四分位数间距。 IQR 是衡量数据离散程度的统计指标，基于第三四分位数 $Q3$ 与第一四分位数 $Q1$ 进行计算，计算方式如下式所示：

$$IQR = Q3 - Q1 \quad (3.6)$$

通过使用对异常值不太敏感的统计数据，Robust 缩放可在一定程度上消除异常值的影响。

在选择特征缩放方法时，应深入了解数据的特点和算法的要求。相比于 Min-Max 缩放和 Z-score 缩放，Robust 缩放对异常值不敏感，并且不依赖于数据的特定分布假设。对于具有显著异常值的数据集，如本文涉及的数据集，Robust 缩放通常是更合适的选择。因此，本文选择使用 Robust 缩放对数据集进行特征缩放。

3.4.2 深度自编码 SMOTE 数据增强方法设计

缺陷元件复检任务中存在数据类别高度不平衡的难点。其中，复检缺陷元件属于少数类，复检无缺陷元件属于多数类。数据类别不平衡可能引发训练过程中

梯度更新的不稳定性，导致模型难以有效学习少数类的特征表示，使模型在训练过程中更倾向于将元件判定为复检无缺陷元件类别，从而降低对复检缺陷元件的识别能力。因此，为了提升模型的泛化性能和复检缺陷元件的检出率，需要针对该任务中的数据类别不平衡问题采取有效的应对策略。

为此，本章提出了一种结合深度自编码器和 SMOTE 的深度自编码 SMOTE 数据增强方法，用于生成少数类复检缺陷元件数据，以平衡数据集。图3.9给出了深度自编码 SMOTE 数据增强方法的框架。

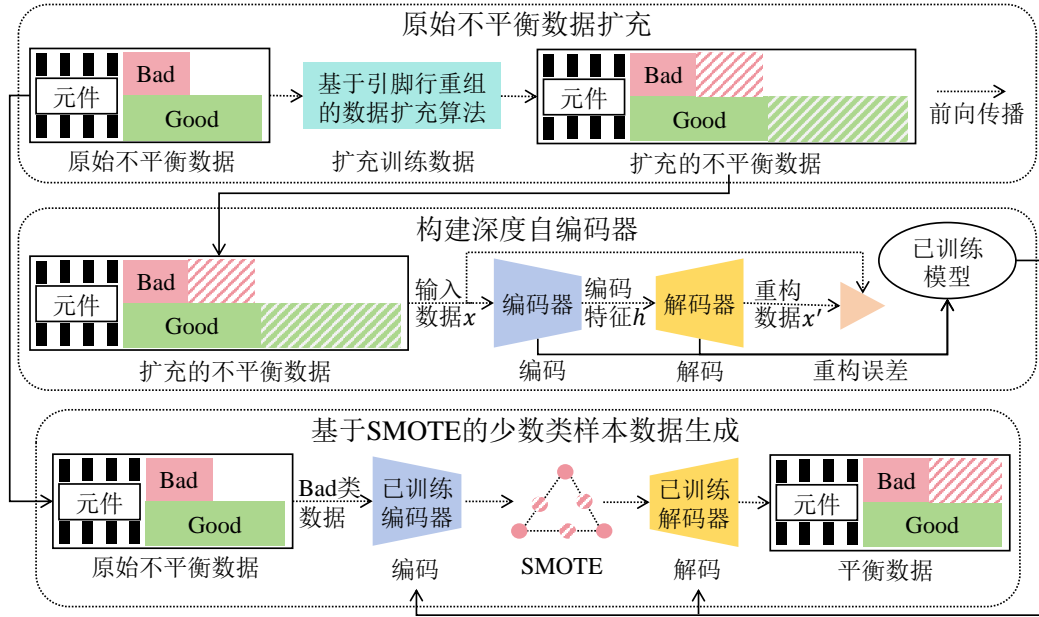


图 3.9 深度自编码 SMOTE 数据增强框架

在图3.9中，Bad 元件表示少数类复检缺陷元件，Good 元件表示多数类复检无缺陷元件。该方法旨在生成少数类复检缺陷元件数据，从而有效缓解任务中的数据不平衡问题。可以看到，深度自编码 SMOTE 数据增强方法可分为以下三部分：基于引脚行重组的数据扩充算法、深度自编码器模型以及基于 SMOTE 的少数类样本数据生成方法。该方法的核心思想是将自编码器与 SMOTE 方法相结合（该思想参考自文献 [46]），本文基于此提出了适用于 PCB 缺陷元件数据生成的深度自编码 SMOTE 数据增强方法。

首先，该方法通过基于引脚行重组的数据扩充算法来扩充原始不平衡训练集。其次，利用扩充的不平衡训练集训练一个由编码器和解码器构成的深度自编码器模型。其中，编码器负责从原始数据中提取低维特征表示，有效去除数据中的冗余信息和噪声；而解码器则负责将这些低维特征表示重构为原始元件数据。接着，将原始不平衡训练集中的复检缺陷元件数据输入到已训练编码器，获得编码空间内的低维特征表示。最后，利用 SMOTE 方法在低维编码空间生成新特征，并通过已训练解码器将 SMOTE 生成的新特征解码为新的复检缺陷元件数据，从而获得类别平衡的训练集。

深度自编码 SMOTE 数据增强方法的流程如算法3.2所示。

算法 3.2 深度自编码 SMOTE 数据增强算法

输入: 原始不平衡数据集 A

输出: 平衡数据集 F

参数: 模型参数: $\Theta = \{\Theta_0, \Theta_1, \dots, \Theta_j\}$; 学习率: α

符号: D - 扩充后的不平衡数据集; B - 扩充后不平衡数据集 D 的分批,
 $B = \{b_1, b_2, \dots, b_n\}$; R_L - 重构误差; C - A 中的类别集合; C_M - A
 中的少数类类别集合; G - 生成的编码数据; S - 生成的解码数据

// 扩充深度自编码器的训练集

```

1  $D \leftarrow$  基于引脚行重组的数据扩充算法 ( $A$ )
2 构建编码器 Encoder、解码器 Decoder, 随机初始化  $\Theta$ 
  // 训练深度自编码器模型
3 for  $e \leftarrow epochs$  do
4   for  $b \leftarrow B$  do
5      $E_b \leftarrow \text{Encoder}(b)$ 
6      $D_b \leftarrow \text{Decoder}(E_b)$ 
7      $R_L = \frac{1}{n} \sum_{i=1}^n (D_{bi} - b_i)^2$ 
8      $\Theta := \Theta - \alpha \frac{\partial R_L}{\partial \Theta}$ 
9   end
10 end
  // 少数类样本生成
11  $F = A$ 
12 for  $m \leftarrow C_M$  do
13    $A_m \leftarrow \text{Select}(C_m \text{ 类不平衡数据})$ 
14    $E_m \leftarrow \text{Encoder}(A_m)$ 
15    $G_m \leftarrow \text{SMOTE}(E_m)$ 
16    $S_m \leftarrow \text{Decoder}(G_m)$ 
17    $F = F + S_m$ 
18 end

```

接下来, 将分别详细介绍基于引脚行重组的数据扩充算法、深度自编码器模型以及基于 SMOTE 的少数类样本数据生成方法。

1. 基于引脚行重组的数据扩充算法

在对深度自编码器模型进行训练之前, 首先采用基于引脚行重组的数据扩充算法对原始不平衡数据进行扩充, 以生成更多的样本。这些扩充后的不平衡数

据仅用于深度自编码器的训练阶段。图3.10以三引脚元件为例，展示了通过重组引脚行数据来扩充样本的过程。

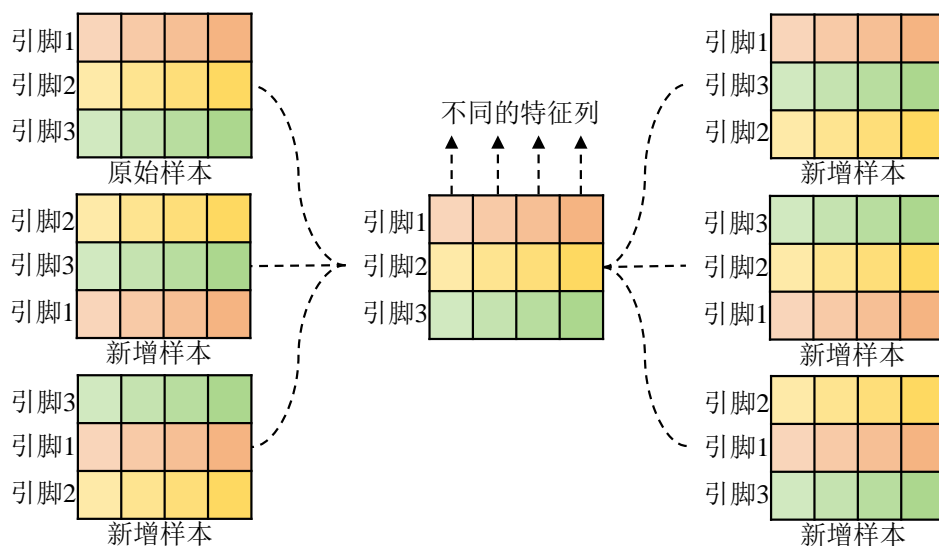


图 3.10 基于引脚行重组的数据扩充示意图

元件的引脚编号是从标记点附近的第一个引脚开始，逆时针方向递增编号。引脚编号在一定程度上反映了元件引脚间的相对物理位置。本章将原始的元素数据按照引脚编号递增的方式进行组织，其中数据的每一行表示该元件特定引脚的焊接质量特征，具体的组织结构已在第 3.4.1 小节进行了详细介绍。基于引脚行重组的数据扩充算法通过对引脚行顺序进行重组，交换不同引脚编号对应的行数据，从而得到新的元件数据。

基于引脚行重组的数据扩充算法流程如算法3.3所示。值得注意的是，在引脚行顺序重组的过程中，必须保持引脚行对应的各实际引脚之间的相对物理位置关系，而非随意交换。

这一扩充操作不仅增加了训练数据的规模，还在一定程度上提升了数据的多样性，从而为深度自编码器中编码器与解码器的训练提供了更加丰富的数据。这种数据扩充策略能够有效改善编码器对潜在特征的捕捉能力以及解码器的重构质量，从而显著提高整个自编码器模型的性能和泛化能力。

2. 深度自编码器模型

参照第 2.2 节内容可知，深度自编码器模型由编码器和解码器组成。在训练过程中，扩充的不平衡数据集被分批次输入编码器进行编码，随后编码特征被传递至解码器以重构数据。接着，基于原始数据和重构数据计算重构损失。训练过程中使用所有类别数据，以使编码器和解码器充分学习元件数据间的内在联系，并提取有效的特征表示，从而更准确地重构并生成符合真实数据分布的样本。

在编码器和解码器的训练过程中，并不存在数据类别不平衡的问题。这是因为对于多数类的复检缺陷元件和少数类的复检无缺陷元件，其原始数据具有相

算法 3.3 基于引脚行重组的数据扩充算法**输入:** 原始元件数据集 A **输出:** 扩充数据集 B

```

1 创建空的表格数据结构  $B$ 
2 for each  $a \leftarrow A$  do
3   由元件数据  $a$  获取该元件的引脚数  $N$ 
4   for  $i \leftarrow 1, 2, \dots, N$  do
5     // 以  $i$  引脚起始, 引脚号的逆时针顺序  $index_1$ 
6      $index_1 = \{i, i+1, \dots, N, 1, 2, \dots, i-1\}$ 
7     按引脚顺序  $index_1$  重新组织元件数据为  $a_1$ 
8      $B.Insert(a_1)$ 
9     // 以  $i$  引脚起始, 引脚号的顺时针顺序  $index_2$ 
10     $index_2 = \{i, i-1, \dots, 1, N, N-1, \dots, i+1\}$ 
11    按引脚顺序  $index_2$  重新组织元件数据为  $a_2$ 
12     $B.Insert(a_2)$ 
13  end
14 end

```

同的特征维度以及基本特征空间结构。这两类数据的主要区别在于特征空间中的分布态势,而非特征本身的固有属性。在深度自编码器框架中,编码器负责提取数据的潜在特征表示,有效去除原始数据中的冗余信息和噪声;而解码器则负责将编码特征空间中的特征表示重构为原始数据。编码和重构过程仅依赖于特征自身的表示及特征间的关联关系,并不受制于数据的类别标签。因此,多数类元件与少数类元件在编码和重构过程中没有本质上的区别。由此可见,在编码器和解码器的训练过程中,并不存在数据类别不平衡的问题。

编码器及解码器的骨干网络基于 Radford 等人建立的深度卷积 GAN 架构^[46,89]。Radford 等人在 GAN 中使用了鉴别器和生成器,其在功能上与编码器和解码器具有相似性。鉴别器在对输入数据进行特征提取和表示学习的过程中,实际上承担了编码器的作用(不考虑鉴别器最终的全连接层),而生成器基于潜在特征表示生成数据,与解码器的功能相对应。此外,如 3.4.1 小节所述,不同引脚数量的元件会具有不同维度的原始特征。为确保原始元件数据能够有效输入到深度学习模型,这里根据元件的引脚数量对其进行分类处理,针对相同引脚数量的元件训练同一编码器和解码器模型。

在编码器和解码器的训练过程中,采用重构误差函数作为损失函数。重构误差函数用于评估解码器生成的重构样本与原始样本之间的差异。通过最小化重

构误差,深度自编码器模型能够学习数据的低维潜在特征表示,从而有效重构原始样本或生成高质量的新样本。重构误差函数的公式如下所示:

$$L_{\text{recon}}(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \frac{1}{n} \sum_{i=1}^n \|x_i - \hat{x}_i\|^2 \quad (3.7)$$

其中 L_{recon} 表示重构误差, \mathbf{x} 表示原始样本特征向量, $\hat{\mathbf{x}}$ 表示解码器生成的重构样本特征向量, n 表示样本的特征维度, x_i 表示原始样本在第 i 个特征维度上的值, \hat{x}_i 表示重构样本在第 i 个特征维度上的值, $\|\cdot\|^2$ 表示向量或标量的平方范数。

重构样本 $\hat{\mathbf{x}}$ 的计算方式如下:

$$\hat{\mathbf{x}} = f_{\text{dec}}(f_{\text{enc}}(\mathbf{x})) \quad (3.8)$$

其中 $f_{\text{enc}}(\cdot)$ 表示编码器的映射函数, $f_{\text{dec}}(\cdot)$ 表示解码器的映射函数。

3. 基于 SMOTE 的少数类样本数据生成方法

如 1.2.3 小节所述, SMOTE 是一种合成少数类样本的过采样技术,其核心思想是在少数类样本的特征空间内合成新样本,而非简单复制已有实例。具体而言,该方法通过在少数类样本之间进行插值生成新样本,从而扩展少数类样本的数量。

一旦编码器和解码器经过训练,就可以利用其生成复检缺陷元件数据。这里将原始的少量复检缺陷元件数据输入已训练编码器,得到编码后的元件特征。接着,采用 SMOTE 方法在该编码空间内进行过采样,以生成新的特征向量。具体而言,在该编码空间内,对于任一复检缺陷元件样本 \mathbf{a}_i , SMOTE 方法首先从其 k 个属于复检缺陷元件类别的最近邻居中选择一个邻居为 \mathbf{a}_{zi} 。之后,使用以下公式生成新样本 \mathbf{s} :

$$\mathbf{s} = \mathbf{a}_i + \lambda \cdot (\mathbf{a}_{zi} - \mathbf{a}_i) \quad (3.9)$$

其中 λ 是取值范围介于 0 和 1 之间的随机权重。

最后,已训练解码器将 SMOTE 生成的新特征向量解码为新的复检缺陷元件数据,从而增加少数类样本数量。这有助于平衡原始数据集,以应对原始数据集中的类别不平衡问题。

3.4.3 层次特征融合分类模型设计

在进行数据增强之后,为了充分挖掘 SPI 和 AOI 数据中的潜在有效信息,并进一步缓解不平衡数据带来的影响,本章提出了一种层次特征融合分类模型,模型框架如图3.11所示。

层次特征融合分类模型是一个专门针对 PCB 缺陷元件复检评估任务而设计的深度学习模型。其设计充分考虑了元件的引脚级特征和元件级特征的层次化

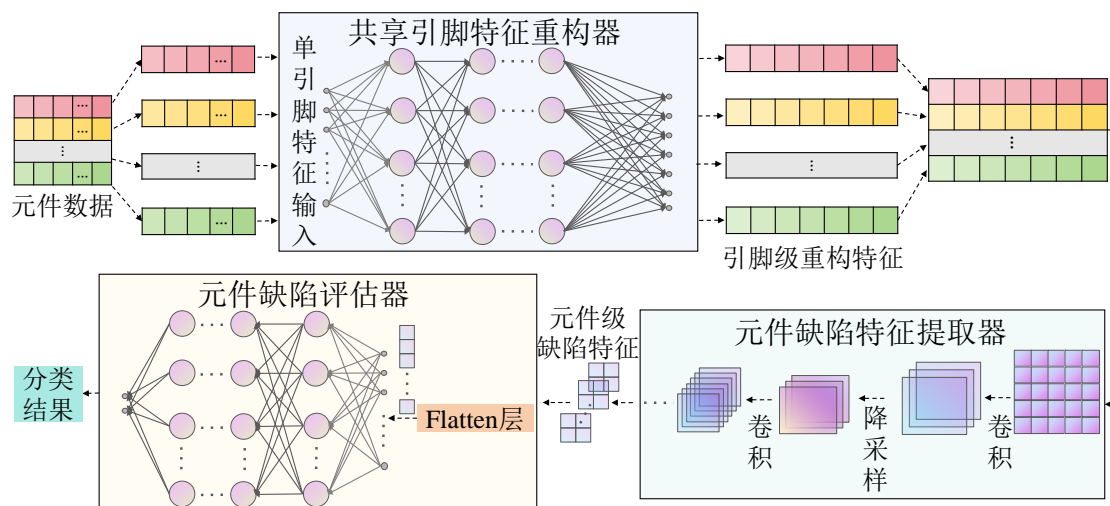


图 3.11 层次特征融合分类模型框架

组织方式，并通过分层次特征融合的策略，实现对元件焊接质量的复检。以下将详细介绍其原理及各组成模块的功能。

1. 输入数据结构

如 3.4.1 节所述，层次特征融合分类模型的输入数据采用二维表格的组织结构，其中每一行代表元件的一个特定引脚的特征，包含该引脚的 SPI 数据特征（如锡膏厚度等）和 AOI 数据特征（指示引脚是否存在缺陷）。对于每个元件，其引脚数量决定了数据表的行数，而每行的列数是固定的（即单一引脚的特征维度是固定的）。由于不同元件可能具有不同的引脚数量，例如两引脚元件和八引脚元件，导致输入数据的特征维度在整体上会随元件引脚数量的不同而变化。因此，本章根据元件的引脚数量对其进行分类处理，并对每类相同引脚数量的元件单独训练相应的模型。这一策略不仅有效解决了因引脚数量差异所引起的特征维度不一致问题，还使得本章方法能够更好地适应不同引脚数量元件的特性，从而提升复检任务中的分类性能。

2. 模型结构与模块功能

层次特征融合分类模型包含三个主要模块：共享引脚特征重构器、元件缺陷特征提取器和元件缺陷评估器。各模块按层次结构逐步处理数据，从引脚层级到元件层级依次进行特征融合和提取。这种层次化处理方式使得模型能够充分捕捉各引脚的局部特征，并有效整合元件的全局特征，进而提升对 AOI 初检缺陷元件的评估和复检能力。

（1）共享引脚特征重构器

共享引脚特征重构器的主要功能是对每个引脚的特征进行重构，旨在从原始数据中提取关键信息，并有效去除噪声和冗余信息。重构后的特征能够更精准地反映引脚的缺陷情况。

共享引脚特征重构器的设计理念主要包括以下三点。第一，每个引脚的特征分为 SPI 和 AOI 特征两部分，这两部分特征之间存在紧密的关联。同时，SPI 特征之间也具备内在的联系。因此，利用引脚特征重构器在引脚层级对不同特征进行深度融合，以实现信息的有效整合，并有效去除冗余信息。第二，引脚特征可能包含噪声，例如 SPI 数据中的测量误差或 AOI 结果的误判。通过特征重构，可以有效去除噪声。第三，不同元件的引脚特征在维度和特性上具有一致性，因此采用共享机制。共享引脚特征重构器的权重在所有引脚上通用，这不仅有效减少了模型的参数量，还提升了模型的泛化能力。该模块的主体网络结构为 MLP，通过非线性变换对每个引脚的原始特征进行重构。

(2) 元件缺陷特征提取器

元件缺陷特征提取器在元件层级对所有引脚的重构特征进行融合，从而提取出能够全面描述元件缺陷程度的全局特征。元件的整体缺陷程度不仅受单个引脚的焊接质量影响，还与各引脚间的交互特性密切相关。因此，设计一个能够捕捉局部与全局关联的模块显得尤为重要。该模块的主体网络结构为卷积神经网络。将重构后的引脚特征组织为二维特征图，卷积神经网络通过卷积和池化操作有效捕捉引脚特征的空间模式，进而生成用于评估元件缺陷程度的全局特征。

(3) 元件缺陷评估器

元件缺陷评估器接收缺陷特征提取器输出的全局特征，综合评估元件的整体焊接质量，并输出最终的分类结果。在评估阶段，模型需要以整体视角判断元件是否存在缺陷，而非仅聚焦于单个引脚的缺陷状况。因此，该模块利用最终的元件级特征进行分类。该模块的主体网络结构为 MLP，其最终输出表示缺陷元件的复检结果。

(4) 损失函数

在本章的二分类问题中，模型的输出层由两个神经元构成，其输出值分别对应复检无缺陷元件类别和复检缺陷元件类别的未归一化得分，记为 z_0 和 z_1 。通过 Softmax 函数将未归一化得分转换为概率，公式如下：

$$p(y = k|\mathbf{x}) = \frac{\exp(z_k)}{\exp(z_0) + \exp(z_1)}, k \in \{0, 1\} \quad (3.10)$$

其中 $p(y = 0|\mathbf{x})$ 表示元件为无缺陷元件的复检概率， $p(y = 1|\mathbf{x})$ 表示元件为缺陷元件的复检概率， \mathbf{x} 表示元件的输入特征。

为了优化网络参数，使用了交叉熵损失函数作为目标函数，其定义为：

$$\mathcal{L}_{class}(\theta) = -\frac{1}{N} \sum_{i=1}^N [y_i \log p(y = 1|\mathbf{x}_i; \theta) + (1 - y_i) \log p(y = 0|\mathbf{x}_i; \theta)] \quad (3.11)$$

其中 N 表示训练数据的样本总数， \mathbf{x}_i 表示第 i 个元件的输入特征， $y_i \in \{0, 1\}$ 表

示第 i 个元件的真实类别标签, $y_i = 1$ 表示复检缺陷元件, $y_i = 0$ 表示复检无缺陷元件, $p(y = k|\mathbf{x}_i; \theta)$ 表示模型判断样本属于类别 k 的概率, θ 表示模型参数。

交叉熵损失函数源自信息论中的交叉熵概念, 旨在衡量两个概率分布之间的差异。在二分类任务中, 真实标签 $y_i \in \{0, 1\}$ 表示样本 i 的类别, 其中 $y_i = 1$ 表示正类, $y_i = 0$ 表示负类。模型预测的概率分布 $p(y|\mathbf{x}_i; \theta)$ 则表示样本 i 属于正类或负类的预测概率。

对于每个样本, 交叉熵损失函数的作用是对元件真实类别的复检概率取对数并取反, 惩罚模型复检结果与真实类别的偏离。具体来说: 当真实类别为正类 ($y_i = 1$) 时, 交叉熵损失中保留 $-\log p(y = 1|\mathbf{x}_i; \theta)$, 即通过最小化损失, 最大化复检概率 $p(y = 1|\mathbf{x}_i; \theta)$; 当真实类别为负类 ($y_i = 0$) 时, 交叉熵损失中保留 $-\log p(y = 0|\mathbf{x}_i; \theta)$, 即通过最小化损失, 最大化复检概率 $p(y = 0|\mathbf{x}_i; \theta)$ 。

通过优化参数 θ , 交叉熵损失函数引导模型的复检概率 $p(y|\mathbf{x}_i; \theta)$ 逼近真实情况, 即通过最大化真实类别的复检概率, 确保模型复检的准确性。特别是在 $p(y = 1|\mathbf{x}_i; \theta)$ 或 $p(y = 0|\mathbf{x}_i; \theta)$ 接近 1 时, 损失函数值趋于最小, 模型达到了最优复检性能。

总体而言, 层次特征融合模型通过引脚级和元件级两个层次的特征融合, 使模型能够充分挖掘每个引脚的局部信息并刻画元件的整体缺陷程度, 从而提升复检性能。通过共享引脚特征重构器的共享机制, 模型能够对所有引脚进行统一建模, 有效减少参数量并避免过拟合, 同时保持对不同引脚特征输入的鲁棒性。同时, 共享引脚特征重构器对每个引脚的特征进行融合和重构, 可以从原始数据中提取关键信息, 并有效去除噪声和冗余信息。此外, 通过元件缺陷特征提取器在元件层级对所有引脚的特征进行深度融合, 进而提取出可以描述元件缺陷程度的全局特征。最终, 通过元件缺陷评估器综合评估元件的整体焊接质量, 给出元件复检结果。

3.5 实验设计与分析

3.5.1 实验数据

本章实验基于 PHME2022 数据集。该数据集中的 SPI 数据和 AOI 数据均来源于真实的 SMT 产线, 并分别以 CSV 文件格式提供。该数据集已经被划分好训练集和测试集。在训练集中, 包含 5985382 条有效 SPI 数据记录和 31617 条有效 AOI 数据记录。在测试集中, 包含 2586072 条有效 SPI 数据记录和 13684 条有效 AOI 数据记录。

本章任务面向的是被 AOI 机器初检为缺陷的元件。因此, 经过处理之后, 训练集中包含 27505 个 AOI 初检缺陷元件的数据可用于本章实验, 测试集中包含

12056 个 AOI 初检缺陷元件的数据可用于本章实验。在数据集中，AOI 初检缺陷元件的复检标签分为两类：复检缺陷元件和复检无缺陷元件。

图3.12展示了训练集中所有 AOI 初检缺陷元件在 PCB 面板上的分布情况。图中横坐标为 PoseX，纵坐标为 PoseY，二者的单位均为毫米（mm）。关于 PoseX 和 PoseY 的具体描述可参考表3.1。红色图形表示复检缺陷元件，绿色图形表示复检无缺陷元件。在训练集中，复检缺陷元件和复检无缺陷元件的数量比例约为 1:66，存在数据类别高度不平衡的问题，这在图3.12中可以明显看出。

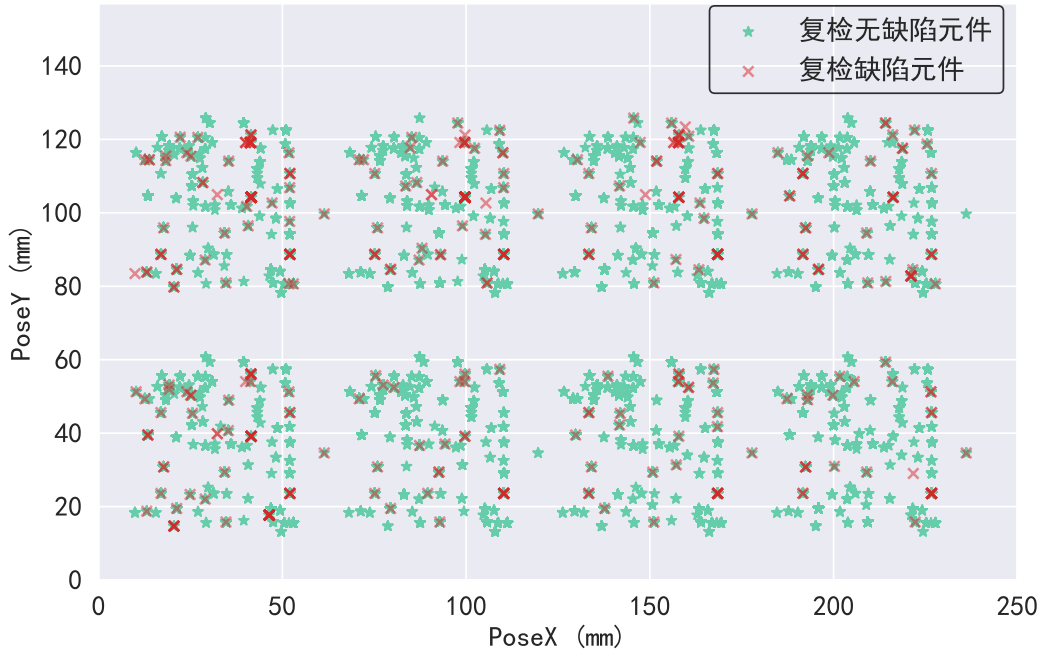


图 3.12 AOI 初检缺陷元件在 PCB 面板上的位置分布图

同时，可观察到图3.12中的元件主要分布于 8 个方形区域内，这源于该数据集所涉及的 PCB 采用了第 3.3 节所述的拼板技术，每块 PCB 面板由 8 块结构相同的独立 PCB 拼接而成。

3.5.2 实验环境

本章实验均在相同的软硬件环境下完成，以确保实验结果的公正性、可比性和一致性。表3.3详细列出了实验软硬件环境的关键配置参数，包括硬件设备的型号及其配置（如 CPU、GPU、内存等），以及软件的版本信息（如操作系统、编程语言等）。

3.5.3 评价指标

分类任务常用的评价指标为精确率（Precision）、召回率（Recall）和 F1 分数（F1-score）。同时，鉴于复检缺陷元件数目在整体元件中占比较少，以及实际应用中更关注复检缺陷元件，本章将复检缺陷元件的精确率、召回率和 F1 分数

表 3.3 实验环境配置参数

类别	规格/型号	详细配置
操作系统	Ubuntu	20.04
CPU	Intel(R) Core(TM) i9-10940X CPU @ 3.30GHz	6 核
GPU	NVIDIA Geforce RTX3090	24GB
内存	-	64GB
Python	-	3.8.0
CUDA	-	11.6
Pytorch	-	1.13.0
numpy	-	1.24.3
pandas	-	1.5.1

作为评价指标。

为了更好的理解上述指标，这里首先引入混淆矩阵的概念。图3.13展示了一个二分类混淆矩阵示意图。该混淆矩阵的不同行分别对应模型分类的不同类别，每一行的样本总数表示分类为相应类别的样本总数；不同列对应不同真实类别，每一列的样本总数表示相应类别样本的真实数目；每一行中的不同数值表示真实样本被分类为相应类别的数目。二分类混淆矩阵将模型的分类结果表示为以下 4 类：*TP* 类，表示真正类（True Positive），即实际为正类且被模型正确分类为正类的样本；*FP* 类，表示假正类（False Positive），即实际为负类而被模型错误分类为正类的样本；*FN* 类，表示假负类（False Negative），即实际为正类而被模型错误分类为负类的样本；*TN* 类，表示真负类（True Negative），即实际为负类且被模型正确分类为负类的样本。正类在本章任务中对应复检缺陷元件，负类在本章任务中对应复检无缺陷元件。

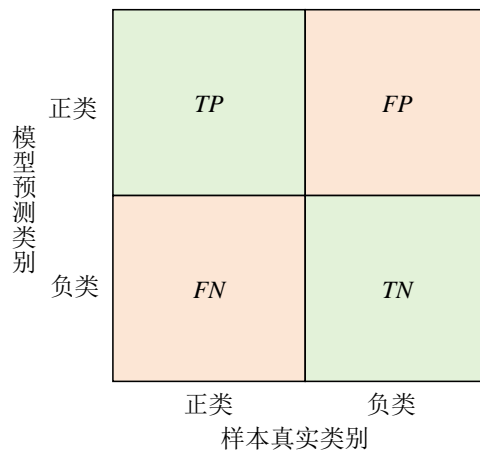


图 3.13 二分类混淆矩阵示意图

精确率表示在被模型分类为正类的样本中，实际为正类的样本比例。精确率反映了模型对正类检测结果的准确性，其值越高，表明模型在判定样本为正类时的结果更加可信。精确率用 P 表示，其计算方式如下：

$$P = \frac{TP}{TP + FP} \quad (3.12)$$

召回率表示在所有实际为正类的样本中，模型正确分类为正类的样本比例。召回率表示正类样本被检出的比例，其值越高，表示模型对正类样本的检测能力越强。召回率用 R 表示，其计算方式如下：

$$R = \frac{TP}{TP + FN} \quad (3.13)$$

F1 分数是精确率和召回率的调和平均数，用于平衡精确率和召回率。精确率和召回率各自衡量模型性能的不同方面。精确率关注模型的预测结果是否准确（正类预测中有多少为真），而召回率关注模型是否漏掉正类样本（实际正类中有多少被检出）。在实际中，精确率和召回率往往是矛盾的。当两者之间需要权衡时，单独考虑其中一个指标可能无法全面反映模型的性能，F1 分数可以综合这两个指标。F1 分数用 $F1$ 表示，其计算方式如下：

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (3.14)$$

这里通过设置精确率、召回率和 F1 分数为评价指标，能够全面地刻画不同方法的性能。

3.5.4 实验结果分析

为了验证所提方法的有效性，在此分别进行如下对比实验和消融实验，并给出相应实验结果和分析。

1. 对比实验

在对比实验部分，将本章所提方法与现有研究中的缺陷元件复检方法进行分类效果对比。具体对比的基线方法（现有最新研究中的方法）分别是：

- **Gaffet1-XGB**: Gaffet 等^[19]提出了一种基于 XGBoost 的引脚级缺陷元件复检方法，记该方法为 Gaffet1-XGB。该方法具体提取分类特征和数值特征，并复检单一引脚是否缺陷，如果复检某个引脚为缺陷，则判定该引脚所在的元件为缺陷。
- **Tang1-RF**: Tang 等^[18]提出了一种基于随机森林的缺陷元件复检方法，记该方法为 Tang1-RF。该方法构建和选取分类特征及数值特征，并在训练过程中通过调整阈值来缓解不平衡数据的影响。
- **Gore1-LGBM**: Gore 等^[17]提出了一种基于 LightGBM 的缺陷元件复检方法，记该方法为 Gore1-LGBM。该方法从 AOI 数据中提取了 15 个统计特征，从 SPI 数据中提取了 24 个统计特征。
- **Schmidt1-RF**: Schmidt 等^[20]提出了基于随机森林的缺陷元件复检方法，记该方法为 Schmidt1-RF。该方法计算元件中不同引脚对应锡膏的面积百分比、高度、X 和 Y 方向偏移等特征的最大值和最小值作为特征。同时该方

法计算元件的 AOI 条目数量，并从 AOI 缺陷条目中提取分类特征。该方法通过调整类别权重来缓解不平衡数据的影响。

本章提出的深度自编码 SMOTE 数据增强方法记为 DAESMOTE，层次特征融合模型记为 HFFNet。本章提出的完整的针对 AOI 初检缺陷元件的复检分类方法记为 DAESMOTE-HFFNet。通过对比试验，获得了如图3.14所示的实验结果。

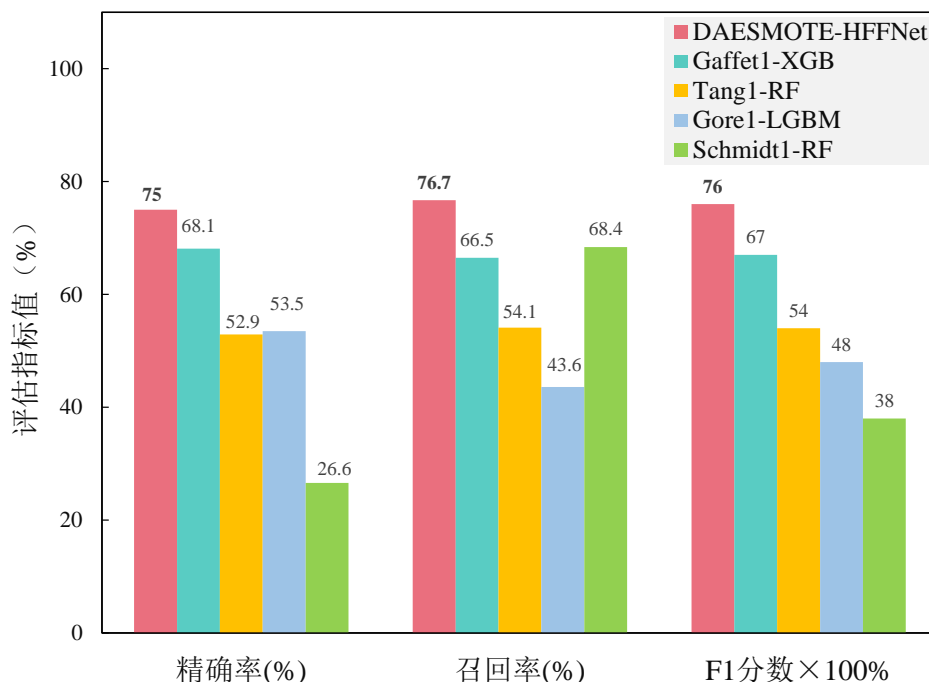


图 3.14 缺陷元件复检方法对比实验结果

图3.14中使用不同颜色的柱状图表示不同方法的性能。同时，在实验结果可视化时，为了使 F1 分数的数值范围与精确率、召回率保持一致，将 F1 分数乘以 100%，以百分比形式进行展示。

通过图3.14中的实验结果可以直观看出，所提方法优于现有研究中的相关方法。所提方法的精确率为 75.0%，召回率为 76.7%，F1 分数为 0.76。在基线方法中，Gaffet1-XGB 取得了最佳精确率为 68.1%，最佳 F1 分数为 0.67。而 Schmidt1-RF 取得了基线方法中的最佳召回率为 68.4%。因此，相比于基线方法的最佳性能，本章所提方法的精确率提升了 6.9%，召回率提升了 8.3%，F1 分数提升了 0.09。这证明了本章所提方法的优越性。

同时可以观察出，现有方法的性能相差较大，这是因为目前相关的研究尚处于初步探索阶段。

2. 消融实验

本章所提出的缺陷元件复检分类方法 DAESMOTE-HFFNet 由两部分组成：深度自编码 SMOTE 数据增强方法 DAESMOTE 及层次特征融合模型 HFFNet。在消融实验部分，为了分别证明 DAESMOTE 数据增强方法和 HFFNet 模型各自

的有效性,进行了以下消融实验。

(1) 深度自编码 SMOTE 数据增强方法消融实验

本消融实验旨在验证深度自编码 SMOTE 数据增强方法 DAESMOTE 的有效性。为此,以所提方法 DAESMOTE-HFFNet 作为基准参照,将数据增强方法作为控制变量,开展一系列消融实验。除所提方法外,其余消融实验方法如下:

- **SubDAESMOTE-HFFNet:** 在深度自编码 SMOTE 数据增强方法的基础上,将其中的引脚行重组数据扩充模块进行消除。分类模型仍采用 HFFNet。将该方法记为 SubDAESMOTE-HFFNet。
- **SMOTE-HFFNet:** 仅采用 SMOTE 进行数据增强,将引脚行重组数据扩充模块和编码器-解码器模块进行消除。分类模型仍采用 HFFNet。将该方法记为 SMOTE-HFFNet。
- **NoDAESMOTE-HFFNet:** 不使用任何数据增强方法,分类模型仍采用 HFFNet。将该方法记为 NoDAESMOTE-HFFNet。

该消融实验的结果呈现于图3.15。图中使用不同颜色的柱状图表示不同方法的性能。同时,在实验结果可视化时,为了使 F1 分数的数值范围与精确率、召回率保持一致,将 F1 分数乘以 100%,以百分比形式进行展示。

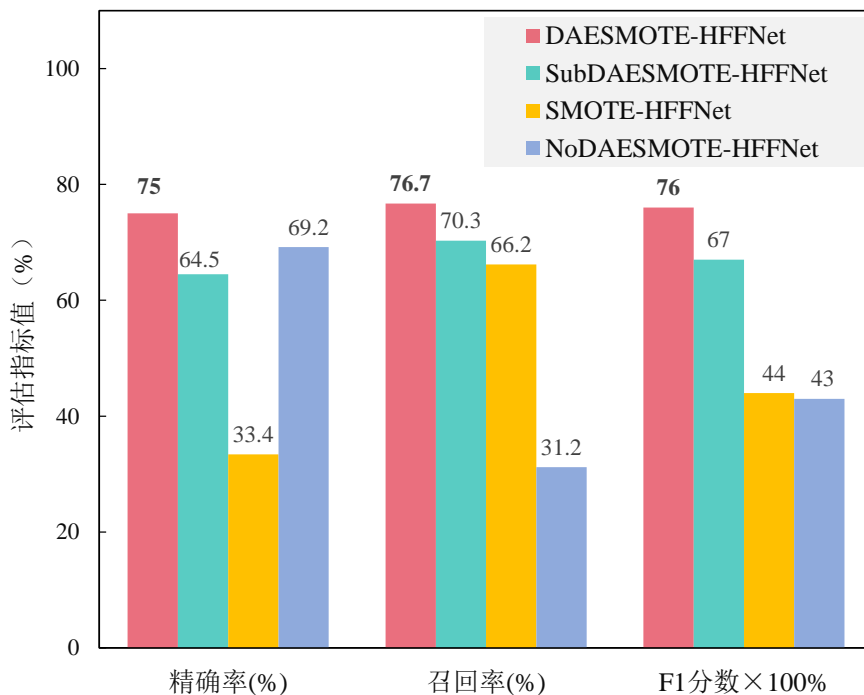


图 3.15 深度自编码 SMOTE 方法消融实验结果

通过观察图中的消融实验结果,可以发现 DAESMOTE-HFFNet 方法在精确率、召回率和 F1 分数上的性能均优于 NoDAESMOTE-HFFNet 方法。相比于 NoDAESMOTE-HFFNet, DAESMOTE-HFFNet 在采用深度自编码 SMOTE 数据增强方法后,模型的精确率提升了 5.8%,召回率提升了 45.5%,F1 分数提升了

0.33。这验证了本章所提出的深度自编码 SMOTE 数据增强方法的有效性,表明该方法能够有效应对数据不平衡问题。通过生成少数类样本,深度自编码 SMOTE 方法增加了少数类样本的数量和多样性,使模型在训练过程中能够更充分地学习不同类别样本的特征分布,从而提升了模型的整体性能。

同时,通过对比 DAESMOTE-HFFNet 方法和 SubDAESMOTE-HFFNet 方法的效果,发现 DAESMOTE-HFFNet 方法在精确率、召回率和 F1 分数上的性能均优于 SubDAESMOTE-HFFNet 方法。这验证了深度自编码 SMOTE 数据增强方法中数据行重组扩充模块的有效性。接着,对比 SubDAESMOTE-HFFNet 方法和 SMOTE-HFFNet 方法的效果,发现 SubDAESMOTE-HFFNet 方法在精确率、召回率和 F1 分数上的性能均优于 SMOTE-HFFNet 方法。这验证了深度自编码 SMOTE 数据增强方法中编码器-解码器模块的有效性。

综合以上分析,可以得出明确结论:在本章所提出的方法中,DAESMOTE 数据增强方法对于提升 HFFNet 模型的性能具有显著且积极的作用,是一种行之有效的数据增强策略。

(2) 层次特征融合模型消融实验

本消融实验旨在验证层次特征融合模型 HFFNet 的有效性。为此,以 DAESMOTE-HFFNet 方法作为基准参照,通过系统性地使用其他模型替换 DAESMOTE-HFFNet 方法中的 HFFNet 模型,开展了一系列实验并进行效果对比。除所提方法外,其余实验方法如下:

- DAESMOTE-XGB: 首先,采用 DAESMOTE 方法进行数据增强,生成少数类复检缺陷元件样本,以平衡数据集;随后,构建并训练 XGBoost 模型,以此代替 HFFNet 模型执行复检分类任务。为便于区分和引用,将该方法记为 DAESMOTE-XGB。
- DAESMOTE-LGBM: 首先,采用 DAESMOTE 方法进行数据增强,生成少数类复检缺陷元件样本,以平衡数据集;随后,构建并训练 LightGBM 模型,以此代替 HFFNet 模型执行复检分类任务。为便于区分和引用,将该方法记为 DAESMOTE-LGBM。
- DAESMOTE-RF: 首先,采用 DAESMOTE 方法进行数据增强,生成少数类复检缺陷元件样本,以平衡数据集;随后,构建并训练随机森林模型,以此代替 HFFNet 模型执行复检分类任务。为便于区分和引用,将该方法记为 DAESMOTE-RF。

该消融实验的结果如图3.16所示。

图3.16中使用不同颜色的柱状图表示不同方法的性能。同时,在实验结果可视化时,为了使 F1 分数的数值范围与精确率、召回率保持一致,将 F1 分数乘以 100%,以百分比形式进行展示。

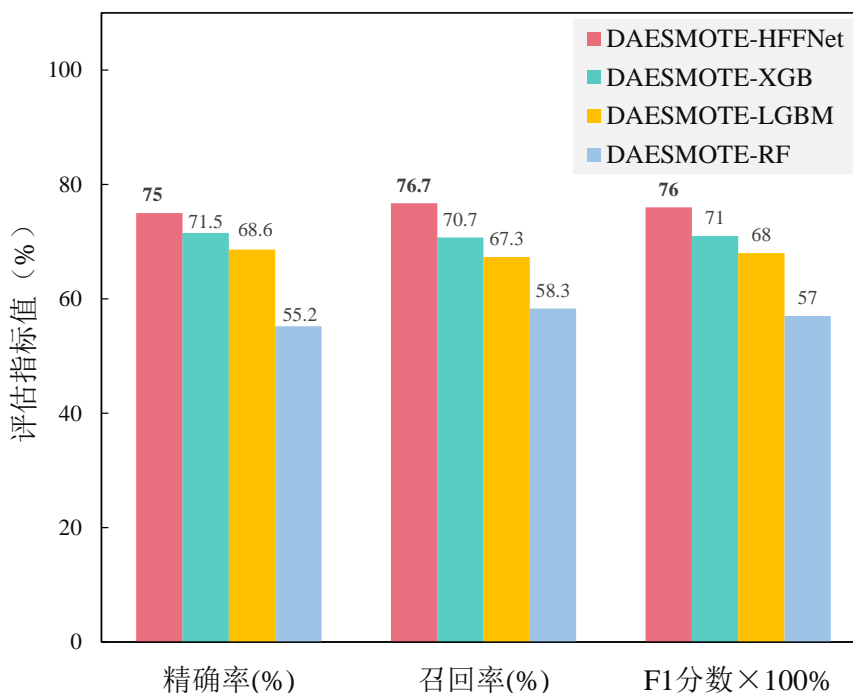


图 3.16 层次特征融合模型消融实验结果

在本实验中，保持数据增强方法的一致性，仅将分类模型作为变量。实验结果表明，DAESMOTE-HFFNet 方法在精确率、召回率和 F1 分数上均取得了最优性能。这充分验证了 HFFNet 分类模型在特征提取和复检分类方面的有效性，说明了在缺陷元件自动复检分类任务中 HFFNet 的性能要优于 XGBoost、LightGBM 和随机森林模型。同时也证实了 HFFNet 分类模型与 DAESMOTE 数据增强方法的良好兼容性。

3. 模型在实际产线中的应用可行性分析

为评估所提模型在实际产线中的部署与应用可行性，本文对模型的计算复杂度与实时性进行了测试与分析。考虑到工业现场通常不具备 GPU 或高性能计算资源，测试在一台搭载 Intel Core i5 处理器（16GB 内存，无 GPU 加速）的通用设备上进行。结果表明，所训练模型在推理阶段的内存占用约为 12.18MB，远低于当前常见处理器可用内存（通常为 GB 级别）；同时，模型对单个元件进行复检的平均推理耗时为 0.64ms，在无需硬件加速的条件下也能实现毫秒级响应。相比之下，人工复检过程通常耗时较长，时间单位多以秒甚至分钟计算。因此，本文方法具备良好的资源适配性与实时性，能够在不增加额外硬件成本的前提下部署于现有产线设备中，满足实际工业应用需求。

3.6 本章小结

针对缺陷元件复检问题，考虑到数据类别高度不平衡的难点，本章提出了一种基于深度自编码 SMOTE 的缺陷元件自动复检方法，实现了比现有方法更优

的自动复检性能。首先,介绍和分析了 PCB 元件的 SPI 及 AOI 数据特性。其次,提出了深度自编码 SMOTE 方法进行数据增强。该方法在深度编码空间内进行过采样,以扩充少数类复检缺陷元件数据,从而充分缓解不平衡数据带来的影响。接着,为充分分析利用 SPI 和 AOI 数据,并进一步缓解不平衡数据带来的影响,提出了层次特征融合分类模型,在引脚层级和元件层级深度融合特征。最后,在 PHME2022 数据集上进行了对比实验和消融实验。对比实验的结果表明,与现有方法的最佳性能相比,所提方法的 F1 分数提高了 0.09,精确率提高了 6.9%,召回率提高了 8.3%,这验证了所提方法的优越性。此外,消融实验进一步验证了所提出的深度自编码 SMOTE 数据增强方法和层次特征融合分类模型各自的有效性。

第4章 基于任务演进 TabNet 训练策略的 PCB 缺陷元件可修复性判定方法

针对缺陷元件可修复性判定问题,考虑到数据样本量小的难点,本章提出了一种基于任务演进 TabNet 训练策略的缺陷元件可修复性判定方法,实现了比现有方法更优的自动判定性能。首先,设计了分层特征提取方法,在引脚层级和元件层级分别构建并提取特征,从而充分利用原始数据信息,并从不同层次全面表征元件的缺陷程度,以便模型更好地判定缺陷元件的可修复性;其次,基于这些分层特征,提出任务演进 TabNet 训练策略,以充分缓解小样本数据带来的影响。该策略分为两个阶段:在缺陷感知预训练阶段,利用大量 AOI 初检缺陷元件数据对 TabNet 进行训练,使其识别其中的复检缺陷元件,从而增强模型对元件缺陷的识别能力;在可修复性认知迁移阶段,调整 TabNet 的输出层结构,并结合少量的复检缺陷元件数据对 TabNet 进行再训练,以进一步提升其在可修复性判定任务中的性能。最后,在 PHME2022 数据集上进行对比实验和消融实验,验证了所提方法的有效性,并分析说明了所提方法在实际产线中的应用可行性。

4.1 引言

在 SMT 实际产线中,被 AOI 机器初检为存在焊接缺陷的 PCB 元件,若在缺陷复检过程中被确认存在缺陷,则由修复操作员进一步检查并判定其可修复性,将其分类为可修复元件、不可修复元件和错误丢弃元件。其中,可修复元件和不可修复元件均为实际存在缺陷的元件;错误丢弃元件指实际没有焊接缺陷的元件,但其在缺陷复检过程中被错误判断为存在焊接缺陷。对于可修复元件,修复操作员会现场对其进行修复。

因此,修复操作员在 SMT 产线中的工作内容分为两部分:一是对复检缺陷元件进行可修复性的检查和判定;二是对可修复元件进行现场修复。而修复操作员对缺陷元件进行可修复性检查和判定的过程通常是繁琐且耗时的,该过程会影响其修复工作的整体效率。为维持修复流程的正常运转,需要配备较多的修复操作员,从而增加人力成本,尤其是在大规模生产的场景下。因此,本章基于标准 SMT 产线的生产过程数据,即 SPI 和 AOI 数据,利用算法对复检缺陷元件的可修复性进行自动判定。本章研究内容旨在减轻修复操作员的工作负担,使其专注于可修复元件的修复操作,从而提升其修复工作的整体效率,有效降低由缺陷元件的可修复性判定带来的额外人力成本,并进一步提高产线的自动化水平。

由于 SMT 产线中真正出现缺陷的元件数量较少,所以缺陷复检过程中被确

认存在缺陷的元件数量较少。因此，在复检缺陷元件的可修复性判定任务中，存在样本数据量小的难点。考虑到该难点，本章提出了一种基于任务演进 TabNet 训练策略的缺陷元件可修复性判定方法。

首先，设计了一种分层特征提取方法，该方法同时在引脚层级和元件层级构建和提取特征，充分利用原始数据提供的有效信息，从不同层级全面表征元件的缺陷程度，从而便于模型判定其可修复性。

随后，基于提取的分层特征，提出了双阶段任务演进 TabNet 训练策略。该策略分为缺陷感知预训练和可修复性认知迁移两个阶段。在缺陷感知预训练阶段，利用大量 AOI 初检缺陷元件数据对 TabNet 进行训练，实现对元件缺陷的自动复检分类。通过这一阶段的训练，TabNet 得以从大量的 AOI 初检缺陷元件数据中学习如何评估元件缺陷程度，从而提升模型对元件缺陷的感知能力。在可修复性认知迁移阶段，调整 TabNet 的输出层结构，并结合少量复检缺陷元件数据对 TabNet 模型进行再训练，以适应元件可修复性判定的目标任务。在这一阶段，TabNet 能够更准确地提取与元件可修复性相关的特征模式，从而在小样本数据条件下实现对缺陷元件可修复性判定任务的有效建模。

本章结构安排如下：第 4.2 节给出了本章任务的问题描述；第 4.3 节给出了包括分层特征提取方法和双阶段任务演进 TabNet 训练策略在内的缺陷元件可修复性判定方法设计；第 4.4 节给出了相关实验设计与结果分析；第 4.5 节给出了本章研究内容的总结。

4.2 问题描述

本章的任务是对 SMT 产线中的复检缺陷元件进行可修复性判定。通过分析给定复检缺陷元件的 SPI 和 AOI 数据，利用算法模型对其可修复性类别进行判定，将其分类为可修复元件、不可修复元件或错误丢弃元件。这一任务的核心目标是对复检缺陷元件的可修复性进行自动判定，从而使实际产线中的修复操作员专注于可修复元件的修复操作，进而提升修复工作的整体效率，有效降低由缺陷元件可修复性判定带来的额外人力成本，并进一步提高产线的自动化水平。

具体而言，给定某复检缺陷元件的数据为 $\mathbf{x} = [\mathbf{x}_{\text{SPI}}, \mathbf{x}_{\text{AOI}}]$ 。其中， \mathbf{x}_{SPI} 表示该元件的 SPI 数据， \mathbf{x}_{AOI} 表示该元件的 AOI 数据。模型的目标是构建一个函数 $f(\cdot)$ ，通过学习输入数据与输出类别之间的映射关系，对元件的可修复性类别进行判定：

$$y = f(\mathbf{x}) \quad (4.1)$$

其中 $y \in \{0, 1, 2\}$ 是模型的输出，表示分类结果。类别 0 表示错误丢弃元件，类别 1 表示可修复元件，类别 2 表示不可修复元件。

4.3 PCB 缺陷元件可修复性判定方法设计

针对缺陷元件可修复性判定任务，本章提出了一种基于任务演进 TabNet 训练策略的自动判定方法，其整体框架如图4.1所示。

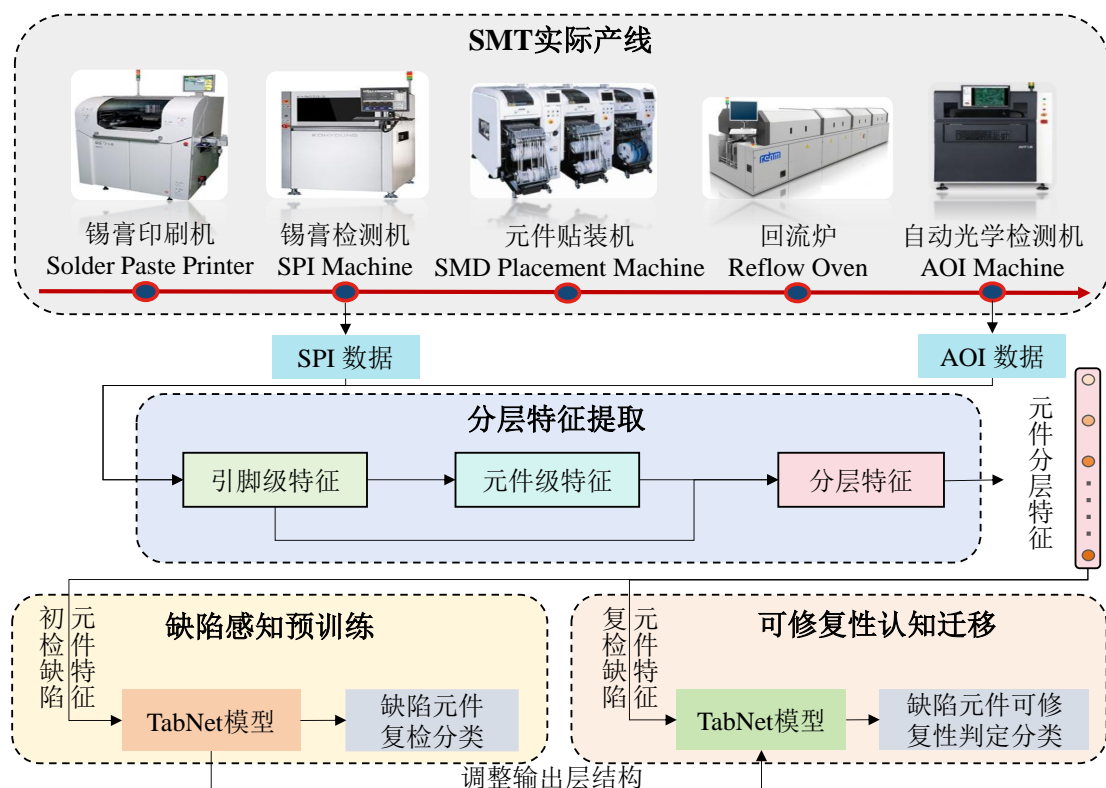


图 4.1 缺陷元件可修复性判定方法整体框架

首先，为了充分利用原始数据提供的信息，本章提出了一种分层特征提取方法。该方法在引脚层级和元件层级分别提取特征，充分考虑了 PCB 元件数据的特点及其物理特性。提取的分层特征能够从不同层次全面表征元件的缺陷程度，从而帮助模型判定复检缺陷元件的可修复性。同时，该方法通过设计统一维度的特征提取框架，确保能够对引脚数量不同的元件提取相同数量的特征。

接着，本章基于分层特征，提出了一种双阶段任务演进 TabNet 训练策略。该训练策略分为两个阶段，即缺陷感知预训练阶段和可修复性认知迁移阶段。在缺陷感知预训练阶段，利用大量 AOI 初检缺陷元件数据对 TabNet 进行预训练。通过这一预训练过程，TabNet 得以从大量的 AOI 初检缺陷元件数据中学习如何评估元件缺陷程度，从而提升模型对元件缺陷的感知能力。在可修复性认知迁移阶段，通过调整 TabNet 输出层结构，并结合少量复检缺陷元件数据对 TabNet 进行再训练，进一步提升 TabNet 在元件可修复性判定任务中的性能。在这一阶段，分类模型能够更准确地提取与元件可修复性相关的特征模式，从而在小样本数据条件下实现对缺陷元件可修复性判定任务的有效建模。

下面分别介绍元件分层特征提取方法和双阶段任务演进 TabNet 训练策略。

4.3.1 元件分层特征提取方法设计

为有效利用原始数据,本文提出了一种分层特征提取方法,分别在引脚层级和元件层级构建并提取特征,以便从多个层次全面表征元件的缺陷程度,从而便于模型判定其可修复性。同时,通过分层特征的有效信息提取能力,可以缓解小样本数据带来的影响。下文将直接阐述元件分层特征提取方法,关于 SPI 数据和 AOI 数据的详细介绍可参考第 3.3 节,此处不再赘述。

1. 引脚级特征提取方法

在实际生产中,大多数元件是两引脚元件^[18]。因此,在引脚级特征提取方法中,保留两引脚元件的原始引脚级测量值。对于具有两个以上引脚的元件,采用其引脚 1 和引脚 2 的特征值;若元件仅有一个引脚,则复制该引脚的特征作为引脚 2 的特征值,以形成统一的结构。

本章最终提取了 34 个引脚级特征,如表 4.1 所示。其中,编号为 01 的特征 Pin1_VolumeP 提取自引脚 1 的原始特征 VolumeP,编号为 02 的特征 Pin1_Height 提取自引脚 1 的原始特征 Height,编号为 03-12 的特征以此类推。编号为 13 的特征 Pin2_VolumeP 提取自引脚 2 的原始特征 VolumeP,编号为 14 的特征 Pin2_Height 提取自引脚 2 的原始特征 Height,编号为 15-24 的特征以此类推。编号为 25-29 的特征提取自引脚 1 的原始特征 AOI Label,并采用了独热编码的方式,这几个特征分别表示引脚 1 中相应 AOI 缺陷标签的存在情况。编号为 30-34 的特征提取自引脚 2 的原始特征 AOI Label,并采用了独热编码的方式,这几个特征分别表示引脚 2 中相应 AOI 缺陷标签的存在情况。

表 4.1 引脚级特征列表

特征编号	特征名称	描述
01-12	Pin1_{VolumeP, Height, AreaP, OffsetXP, OffsetYP, SizeX, SizeY, Volume, Area, Shape, PosX, PosY}	SPI 特征
13-24	Pin2_{VolumeP, Height, AreaP, OffsetXP, OffsetYP, SizeX, SizeY, Volume, Area, Shape, PosX, PosY}	
25-29	Pin1_{Translated, Coplanarity, Soldered, UnSoldered, Others}	AOI 特征
30-34	Pin2_{Translated, Coplanarity, Soldered, UnSoldered, Others}	

引脚级特征提取算法的流程见算法 4.1。

该引脚级特征提取方法充分考虑了实际 PCB 元件的特点,最大程度的保留了两引脚元件的原始数据信息,并通过统一的特征结构处理了其他引脚数量的元件。然而,对于具有多个引脚的元件,引脚 1 和引脚 2 的特征仅能从局部对其缺陷程度进行刻画。为了弥补这一不足,还需要通过接下来的元件级特征提取方法,进一步从整体上捕获元件特征的全貌和高层次信息,以更全面地表征元件的缺陷程度,提升模型的判别能力。

算法 4.1 引脚级特征提取算法**输入:** SPI 数据 $DataS$ 、AOI 数据 $DataA$ **输出:** 引脚级特征 $PinF$

```

1 从  $DataA$  提取复检缺陷元件的层次化标识符组合  $CompIds$ 
2 for  $ids \leftarrow CompIds$  do
    // 通过面板标识符、PCB标识符和元件标识符遍历复检缺陷元件
3    $Panel\_id = ids[0]$ 、 $Figure\_id = ids[1]$ 、 $Com\_id = ids[2]$ 
4   通过三元组标识符 ( $Panel\_id$ 、 $Figure\_id$ 、 $Com\_id$ ), 从  $DataS$  中提
      取该元件的 SPI 数据  $DataS_{Comp}$ 
5   通过三元组标识符 ( $Panel\_id$ 、 $Figure\_id$ 、 $Com\_id$ ), 从  $DataA$  中提
      取该元件的 AOI 数据  $DataA_{Comp}$ 
6    $Com\_pin\_num = \text{len}(DataS_{Comp})$  // 获取元件引脚数
7   if  $Com\_pin\_num \geq 2$  then
8       根据表4.1, 从  $DataS_{Comp}$  提取引脚 1 及引脚 2 的 SPI 特征分别
          为:  $S_1$ 、 $S_2$ 
9       根据表4.1, 从  $DataA_{Comp}$  提取引脚 1 及引脚 2 的 AOI 特征分别
          为:  $A_1$ 、 $A_2$ 
10       $PinF.Insert([S_1, A_1, S_2, A_2])$ 
11   end
12   else
13       根据表4.1, 从  $DataS_{Comp}$  提取引脚 1 的 SPI 特征  $S_1$ 
14       根据表4.1, 从  $DataA_{Comp}$  提取引脚 1 的 AOI 特征  $A_1$ 
15       $PinF.Insert([S_1, A_1, S_1, A_1])$ 
16   end
17 end

```

2. 元件级特征提取方法

本节在原始引脚级特征的基础上提取元件级特征。元件级特征的提取通过计算引脚级特征的统计值实现, 统计值包括均值、标准差、最小值和最大值等。

SPI 数据集提供的原始数据中, 包含的引脚级连续特征有: VolumeP、Height、AreaP、OffsetXP、OffsetYP、SizeX、SizeY、Volume、Area、Shape、PosX 和 PosY。这里首先利用原始引脚级特征进行计算, 生成新的引脚级特征: OffsetP、Size、AvgHeight、TgtVolume、PosXY、OffsetRatio、RelativeOffsetX、RelativeOffsetY、HeightP、SHR 和 OffsetR。之后, 计算引脚级原始特征和引脚级生成特征的统计值 (如均值、标准差、方差、最小值、最大值和中值等), 作为元件级特征。下

面首先介绍新生成引脚级特征的计算方式及其含义。

引脚级生成特征 OffsetP 用于反映锡膏偏移的综合程度。OffsetXP 和 OffsetYP 分别表示锡膏在 X 和 Y 方向的偏移百分比，无法全面描述偏移的严重程度。因此，计算 OffsetP 用于反映锡膏偏移的综合程度。这种计算方式类似于计算欧几里得距离，用于描述一个二维偏移向量的模长，具体 OffsetP 计算方式如下：

$$\text{OffsetP} = \sqrt{\text{OffsetXP}^2 + \text{OffsetYP}^2} \quad (4.2)$$

引脚级生成特征 Size 表示锡膏对角线的长度，其计算方式如下：

$$\text{Size} = \sqrt{\text{SizeX}^2 + \text{SizeY}^2} \quad (4.3)$$

引脚级生成特征 AvgHeight 反映锡膏的平均厚度，它不直接反映锡膏的最高厚度或最低厚度，而是一个综合指标，其计算方式如下：

$$\text{AvgHeight} = \frac{\text{Volume}}{\text{Area}} \quad (4.4)$$

引脚级生成特征 TgtVolume 反映锡膏的目标体积，实际锡膏的体积应尽可能接近该目标体积，其计算方式如下：

$$\text{TgtVolume} = \frac{\text{Volume}}{\text{VolumeP}} \quad (4.5)$$

引脚级生成特征 PosXY 反映焊盘距离参考点的直线距离，其计算方式如下：

$$\text{PosXY} = \sqrt{\text{PosX}^2 + \text{PosY}^2} \quad (4.6)$$

引脚级生成特征 OffsetRatio 揭示锡膏水平方向和垂直方向偏移的相对大小，用于捕捉锡膏在 X 和 Y 方向上的主导偏移趋势，其计算方式如下：

$$\text{OffsetRatio} = \frac{\text{OffsetXP}}{\text{OffsetYP}} \quad (4.7)$$

引脚级生成特征 RelativeOffsetX 描述锡膏水平方向偏移的相对大小，提供了水平方向上锡膏的偏移与其物理尺寸的比值，其计算方式如下：

$$\text{RelativeOffsetX} = \frac{\text{OffsetXP}}{\text{SizeX}} \quad (4.8)$$

引脚级生成特征 RelativeOffsetY 描述锡膏垂直方向偏移的相对大小，提供了垂直方向上锡膏的偏移与其物理尺寸的比值，其计算方式如下：

$$\text{RelativeOffsetY} = \frac{\text{OffsetYP}}{\text{SizeY}} \quad (4.9)$$

引脚级生成特征 HeightP 表示锡膏实际高度与目标高度的比值，其计算方式如下：

$$\text{HeightP} = \frac{\text{Height} \cdot \text{VolumeP} \cdot \text{Area}}{\text{Volume} \cdot \text{AreaP}} \quad (4.10)$$

引脚级生成特征 SHR 衡量锡膏形状的变化程度与厚度之间的关系，用于捕捉焊点不规则性，其计算方式如下：

$$\text{SHR} = \frac{\text{Shape}}{\text{Height}} \quad (4.11)$$

引脚级生成特征 OffsetR 用于综合反映锡膏偏移的相对大小，其计算方式如下：

$$\text{OffsetR} = \sqrt{\text{RelativeOffsetX}^2 + \text{RelativeOffsetY}^2} \quad (4.12)$$

最终提取了 158 个元件级特征，如表4.2所示。

表 4.2 元件级特征列表

特征编号	特征名称	描述
001-006	VolumeP_{mean, std, var, min, max, median}	SPI 特征
007-012	Height_{mean, std, var, min, max, median}	
013-018	AreaP_{mean, std, var, min, max, median}	
019-024	OffsetXP_{mean, std, var, min, max, median}	
025-030	OffsetYP_{mean, std, var, min, max, median}	
031-036	SizeX_{mean, std, var, min, max, median}	
037-042	SizeY_{mean, std, var, min, max, median}	
043-048	Volume_{mean, std, var, min, max, median}	
049-054	Area_{mean, std, var, min, max, median}	
055-060	Shape_{mean, std, var, min, max, median}	
061-066	PosX_{mean, std, var, min, max, median}	
067-072	PosY_{mean, std, var, min, max, median}	
073-078	OffsetP_{mean, std, var, min, max, median}	
079-084	Size_{mean, std, var, min, max, median}	
085-090	AvgHeight_{mean, std, var, min, max, median}	
091-096	TgtVolume_{mean, std, var, min, max, median}	
097-102	PosXY_{mean, std, var, min, max, median}	
103-108	OffsetRatio_{mean, std, var, min, max, median}	
109-114	RelativeOffsetX_{mean, std, var, min, max, median}	
115-120	RelativeOffsetY_{mean, std, var, min, max, median}	
121-126	HeightP_{mean, std, var, min, max, median}	AOI 特征
127-132	SHR_{mean, std, var, min, max, median}	
133-138	OffsetR_{mean, std, var, min, max, median}	
139-139	Pin_Count	
140-140	AOILabel_Count	
141-141	Pin_Count_Defective	
142-146	Count_{Translated, Coplanarity, Soldered, UnSoldered, Others}	
147-158	CompType_{C, L, R, D, U, TR, BC, RA, IC, CN, DZ, TRB}	

元件级特征提取算法的流程如算法4.2所示。

算法 4.2 元件级特征提取算法

输入: SPI 数据 $DataS$, AOI 数据 $DataA$

输出: 元件级特征 $CompF$

```

1 从  $DataA$  提取定位复检缺陷元件的标识符组合  $CompIds$ 
2 for  $ids \leftarrow CompIds$  do
3    $Panel\_id = ids[0]$  // 缺陷元件所在面板标识符
4    $Figure\_id = ids[1]$  // 缺陷元件所在 PCB 标识符
5    $Com\_id = ids[2]$  // 缺陷元件标识符
6   通过三元组标识符 ( $Panel\_id$ 、 $Figure\_id$ 、 $Com\_id$ ), 从  $DataS$  中提
    取该元件的 SPI 数据  $DataS_{Comp}$ 
7   通过三元组标识符 ( $Panel\_id$ 、 $Figure\_id$ 、 $Com\_id$ ), 从  $DataA$  中提
    取该元件的 AOI 数据  $DataA_{Comp}$ 
8   从  $DataA_{Comp}$  提取 SPI 数值特征列的数据  $Sf_1$ , 并根据式 (4.2) 至
    式 (4.12), 在引脚级计算生成新的 SPI 数据  $Sf_2$ 
9    $Sf = Sf_1 + Sf_2$ 
10  计算  $Sf$  中各特征的均值  $means$ 、标准差  $stds$ 、方差  $vars$ 、最小值
     $mins$ 、最大值  $maxs$  和中值  $medians$ 
11   $Pin\_Count = \text{len}(DataS_{Comp})$ 
12   $AOI\_Label\_Count = \text{len}(DataA_{Comp})$ 
13  从  $DataA_{Comp}$  中提取缺陷引脚的数量  $Pin\_Count\_Defective$ 
14  计算  $DataA_{Comp}$  中, 不同 AOI Label 的数量
15  从  $Com\_id$  提取元件类型, 并进行独热编码
16  将上述计算的该元件特征添加到  $CompF$ 
17 end

```

在编号为 001-006 的特征中, VolumeP_mean 表示该元件在引脚级特征 VolumeP 上的平均值, 即对该元件所有引脚的 VolumeP 取平均; VolumeP_std 表示该元件在引脚级特征 VolumeP 上的标准差, 即对该元件所有引脚的 VolumeP 取标准差; VolumeP_var 表示该元件在引脚级特征 VolumeP 上的方差, 即对该元件所有引脚的 VolumeP 取方差; VolumeP_min 表示该元件在引脚级特征 VolumeP 上的最小值, 即对该元件所有引脚的 VolumeP 取最小值; VolumeP_max 表示该元件在引脚级特征 VolumeP 上的最大值, 即对该元件所有引脚的 VolumeP 取最大值; VolumeP_median 表示该元件在引脚级特征 VolumeP 上的中值, 即对该元件所有引脚的 VolumeP 取中值。编号为 007-138 的特征以此类推。

此外, Pin_Count 表示元件包含的引脚数量; AOILabel_Count 表示元件的 AOI 缺陷条目数量; Pin_Count_Defective 表示 AOI 机器在该元件上检出的缺陷引脚数量; 编号为 142-146 的特征分别表示相应 AOI 缺陷条目的数量; 编号为 147-158 的特征表示元件类型, 采用了独热编码方式。

4.3.2 双阶段任务演进 TabNet 训练策略设计

在分层特征的基础上, 为充分缓解小样本数据带来的不利影响, 本章提出了一种双阶段任务演进 TabNet 训练策略。提取得到的分层特征具有典型的表格数据特征, 因此采用 TabNet 作为分类模型。所提策略通过在两个训练阶段对 TabNet 进行不同的目标任务设置, 使其能够更有效地完成缺陷元件可修复性自动判定任务。所提训练策略的示意图如图4.2所示。

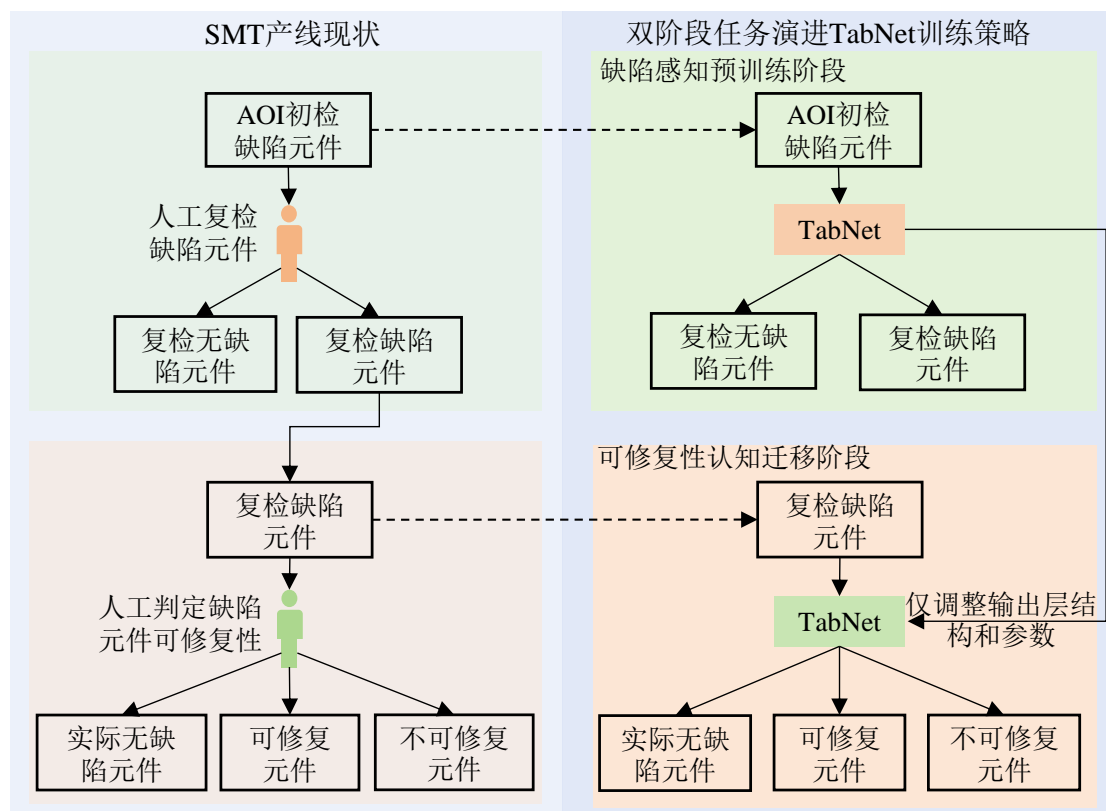


图 4.2 任务演进 TabNet 训练策略示意图 (结合 SMT 产线现状说明)

在缺陷感知预训练阶段, 采用大量 AOI 初检缺陷元件数据对 TabNet 进行预训练, 使其实现元件缺陷自动复检任务。通过缺陷感知预训练阶段的训练, TabNet 得以从大量的 AOI 初检缺陷元件数据中学习如何评估元件缺陷程度, 从而提升 TabNet 对元件缺陷的感知能力。

在可修复性认知迁移阶段, 基于少量复检缺陷元件数据, 对 TabNet 进行再训练, 以帮助其适配元件可修复性判定这一目标任务。在该阶段, 目标任务由元件缺陷复检演化为元件可修复性判定。通过可修复性认知迁移阶段的训练, TabNet

能够更准确地提取与元件可修复性相关的特征模式，从而在小样本数据条件下实现对缺陷元件可修复性判定任务的有效建模。

需要说明的是，在缺陷感知预训练阶段中训练的 TabNet 是一个二分类模型，旨在检测 AOI 初检缺陷元件是否属于复检缺陷元件或复检无缺陷元件。而在可修复性认知迁移阶段，TabNet 需要执行三分类任务，对复检缺陷元件的可修复性进行判定。因此，在可修复性认知迁移阶段，首先需要调整 TabNet 的模型结构。这里仅调节 TabNet 的最后一层输出层。具体而言，TabNet 的最后一层输出层是一个全连接层，本策略将该层的两个神经元调整为三个神经元，使 TabNet 适应三分类任务。同时，这三个神经元的参数会被随机初始化，而其余的网络结构及对应参数则保持不变。之后，利用少量复检缺陷元件数据，对 TabNet 进行再训练，从而使模型适应缺陷元件可修复性判定任务。

此外，在双阶段的模型训练过程中，均采用交叉熵损失函数作为训练损失。在缺陷感知预训练阶段，训练 TabNet 对 AOI 初检缺陷元件进行缺陷复检，目标任务为二分类任务，因此仅采用普通的二分类交叉熵损失函数，具体可参考公式 (3.11)。在可修复性认知迁移阶段，TabNet 需要对复检缺陷元件的可修复性进行判定，这是一个三分类任务。同时，在小样本复检缺陷元件数据集中，三个类别的样本数量通常不完全相同，这会导致轻微的数据类别不平衡问题。因此，在可修复性认知迁移阶段采用了加权的三分类交叉熵损失函数 \mathcal{L}_{class} ，计算方式如下：

$$\mathcal{L}_{class} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^3 w_j y_{ij} \log(p_{ij}) \quad (4.13)$$

其中 N 表示样本总数， $j \in \{0, 1, 2\}$ 为类别索引（对应元件类别分别为：错误丢弃元件、可修复元件、不可修复元件）， w_j 表示类别 j 的权重，用于平衡类别分布，权值通常与该类别样本数量的倒数相关， y_{ij} 表示样本 i 对应类别 j 的真实标签，以独热编码形式表示。当样本 i 属于类别 j 时， $y_{ij} = 1$ ，否则 $y_{ij} = 0$ 。 p_{ij} 是模型判定样本 i 属于类别 j 的概率，输出自 Softmax 函数。

同时，为了进一步控制 TabNet 所选特征的稀疏性，使用稀疏正则化项作为特征选择过程的约束。这种正则化通过最小化学习到的掩码 $M[t]$ 的熵，鼓励掩码向量更加稀疏，从而确保在每一轮决策中只选择最关键的特征。稀疏正则化项 \mathcal{L}_{sparse} 的公式如下：

$$\mathcal{L}_{sparse} = - \sum_{i=1}^{N_{steps}} \sum_{b=1}^B \sum_{j=1}^D \frac{M_{bj}[i] \log(M_{bj}[i] + \epsilon)}{N_{steps} \cdot B} \quad (4.14)$$

其中 T 表示决策步骤的总数， B 表示批大小， D 表示特征维度， $M_{bj}[i]$ 表示特征 j 在决策步 i 、批次 b 中的掩码值， ϵ 是一个用于数值稳定性的小数。

最终的损失函数 \mathcal{L}_{total} 计算如下：

$$\mathcal{L}_{total} = \mathcal{L}_{class} + \lambda \mathcal{L}_{sparse} \quad (4.15)$$

其中 λ 表示稀疏正则项权重，用来对特征选择阶段的特征稀疏性添加约束。 λ 越大，则特征选择越稀疏。

4.4 实验设计与分析

4.4.1 实验数据

本章实验基于 PHME2022 数据集。该数据集中的 SPI 数据和 AOI 数据均来源于真实的 SMT 产线。该数据集已经被划分好训练集和测试集。在训练集中，包含 5985382 条有效 SPI 数据记录和 31617 条有效 AOI 数据记录。在测试集中，包含 2586072 条有效 SPI 数据记录和 13684 条有效 AOI 数据记录。

本章任务面向的是被 AOI 缺陷复检操作员确认为存在缺陷的元件。因此，经过处理之后，训练集中包含 412 个复检缺陷元件的数据可用于本章实验，而测试集中包含 266 个复检缺陷元件的数据可用于本章实验。数据集中，复检缺陷元件的可修复标签分为三类：可修复元件、不可修复元件和错误丢弃元件。

图4.3展示了训练集中所有复检缺陷元件在 PCB 面板上的位置分布情况。图中横坐标为 PoseX，纵坐标为 PoseY，二者的单位均为毫米（mm）。关于 PoseX 和 PoseY 的具体描述可参考表3.1。图中采用不同的颜色和图形标注不同类别的元件。

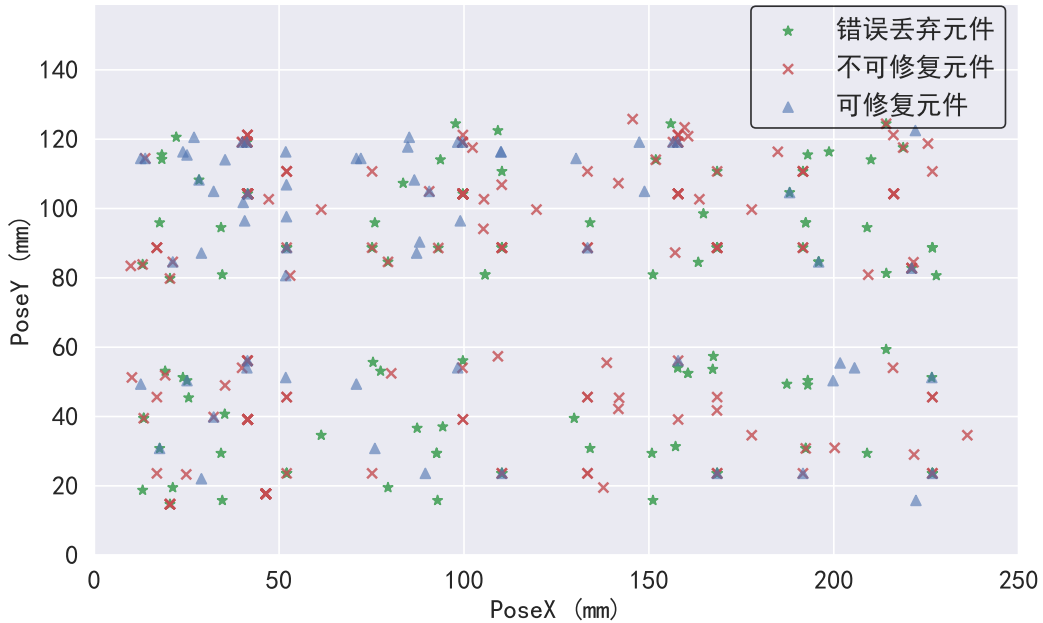


图 4.3 复检缺陷元件在 PCB 面板上的位置分布图

从图4.3中可以直观看出，训练集中的样本总数较少，存在数据样本量少的

问题。同时，可观察到图中的元件主要分布于 8 个区域，这源于该数据集所涉及的 PCB 采用了第 3.3 节所述的拼板技术，每块 PCB 面板由 8 块结构相同的独立 PCB 拼接而成。

4.4.2 实验环境

本章的所有实验均在相同的软硬件环境下进行，以确保实验结果的公正性、可比性和一致性。实验环境与第 3 章中的实验环境一致，具体配置参数可参见第 3 章的表 3.3。

4.4.3 评价指标

在本章的缺陷元件可修复性自动判定任务中，不可修复元件和错误丢弃元件是优化生产流程的核心目标类别，将这两种类别的元件分别精准地自动检测出来可以减少修复操作员的工作量，避免其在这两类元件上浪费时间。但如果这两种类别的元件检测精确率过低，则会使很多其他类别的元件被误判为不可修复类型，从而被错误处理，或使很多实际缺陷的元件流入后续正常工序，从而带来显著额外成本。至于可修复元件，则会由修复操作员进一步处理，其分类结果对任务的直接影响较小。

因此，本章主要关注不可修复元件和错误丢弃元件的分类性能。本章分别将不可修复元件和错误丢弃元件作为正类，计算其精确率、召回率和 F1 分数，并将其平均值作为评价指标。

将不可修复元件作为正类时的精确率、召回率和 F1 分数分别表示为： P_N 、 R_N 和 $F1_N$ 。将错误丢弃元件作为正类时的精确率、召回率和 F1 分数分别表示为： P_F 、 R_F 和 $F1_F$ 。

最终评价指标中的精确率 P_{avg} 计算公式如下：

$$P_{avg} = \frac{P_N + P_F}{2} \quad (4.16)$$

最终评价指标中的召回率 R_{avg} 计算公式如下：

$$R_{avg} = \frac{R_N + R_F}{2} \quad (4.17)$$

最终评价指标中的 F1 分数 $F1_{avg}$ 计算公式如下：

$$F1_{avg} = \frac{F1_N + F1_F}{2} \quad (4.18)$$

4.4.4 实验结果分析

为了验证所提方法的有效性，在此分别进行如下对比实验和消融实验，并给出相应实验结果和分析。

1. 对比实验

在对比实验部分，将本章所提方法与现有方法进行效果对比。具体对比的基线方法（现有最新研究中的方法）如下：

- **Gaffet2-XGB**: Gaffet 等^[19]提出了一种基于元件级统计特征和 XGBoost 模型的缺陷元件可修复性判定方法，记该方法为 Gaffet2-XGB。该方法在元件层级分别提取了相应的数值统计特征和分类特征。
- **Gore2-XGB**: Gore 等^[17]提出了一种基于元件级统计特征和 XGBoost 模型的缺陷元件可修复性判定方法，记该方法为 Gore2-XGB。该方法根据平均值、求和、标准差、最大值、最小值、中值和计数等统计指标计算统计特征，共提取了 153 个元件级统计特征。
- **Tang2-RF**: Tang 等^[18]采用引脚级特征提取方法，并利用随机森林算法对复检缺陷元件的可修复性进行自动判定，记该方法为 Tang2-RF。该方法针对不同元件，统一提取前两个引脚的特征，其中单个引脚上提取的特征数量为 15，总计提取的特征数量为 30。
- **Schmidt2-DT**: Schmidt 等^[20]提出了一种基于单一特征的决策树模型对复检缺陷元件的可修复性进行自动判定，记该方法为 Schmidt2-DT。具体来说，该方法计算每个元件的 AOI 条目数量，并训练关于该唯一特征的决策树。

通过上述实验，得到如图4.4所示的对比实验结果。图中使用不同颜色的柱状图表示不同方法的性能。同时，在实验结果可视化时，为使 F1 分数的数值范围与精确率、召回率保持一致，将 F1 分数乘以 100%，以百分比形式进行展示。

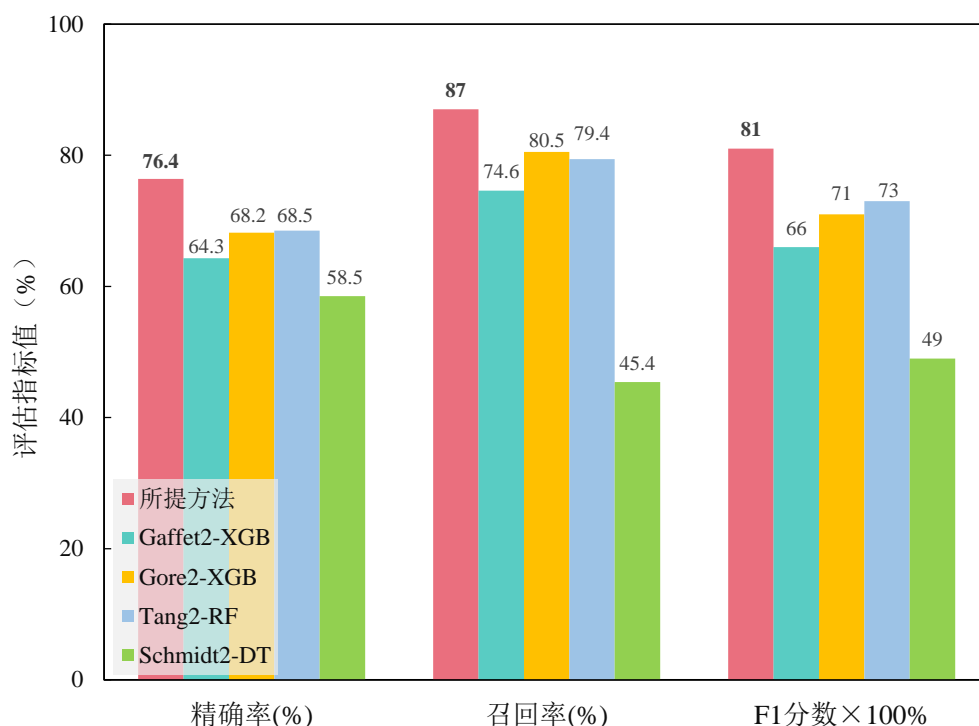


图 4.4 缺陷元件可修复性判定方法对比实验结果

通过图4.4中的实验结果可以看到，本章所提方法优于现有方法。在现有方法中，Tang2-RF 取得了基线方法中最优的精确率为 68.5% 和最优的 F1 分数为 0.73，Gore2-XGB 取得了基线方法中最优的召回率为 80.5%。而本章所提出的方法在精确率、召回率和 F1 分数上相比于基线方法均为最优。具体而言，相比于基线方法中的最优性能，本章所提方法在精确率上提升了 7.9%，在召回率上提升了 6.5%，在 F1 分数上提升了 0.08。该对比实验证明了本章所提方法的优越性。

2. 消融实验

在消融实验部分，为了分别证明元件分层特征提取方法和双阶段任务演进 TabNet 训练策略的有效性，进行了以下消融实验。

(1) 分层特征消融实验

本消融实验旨在验证分层特征提取方法的有效性。除所提方法外，其余消融实验方法如下：

- **Method_PinFeature**：仅保留分层特征中的引脚级特征部分，消融掉元件级特征部分，同时采用双阶段任务演进 TabNet 训练策略。记该方法为 Method_PinFeature。
- **Method_CompFeature**：仅保留分层特征中的元件级特征部分，消融掉引脚级特征部分，同时采用双阶段任务演进 TabNet 训练策略。记该方法为 Method_CompFeature。

通过进行上述实验，得到如图4.5所示的消融实验结果。

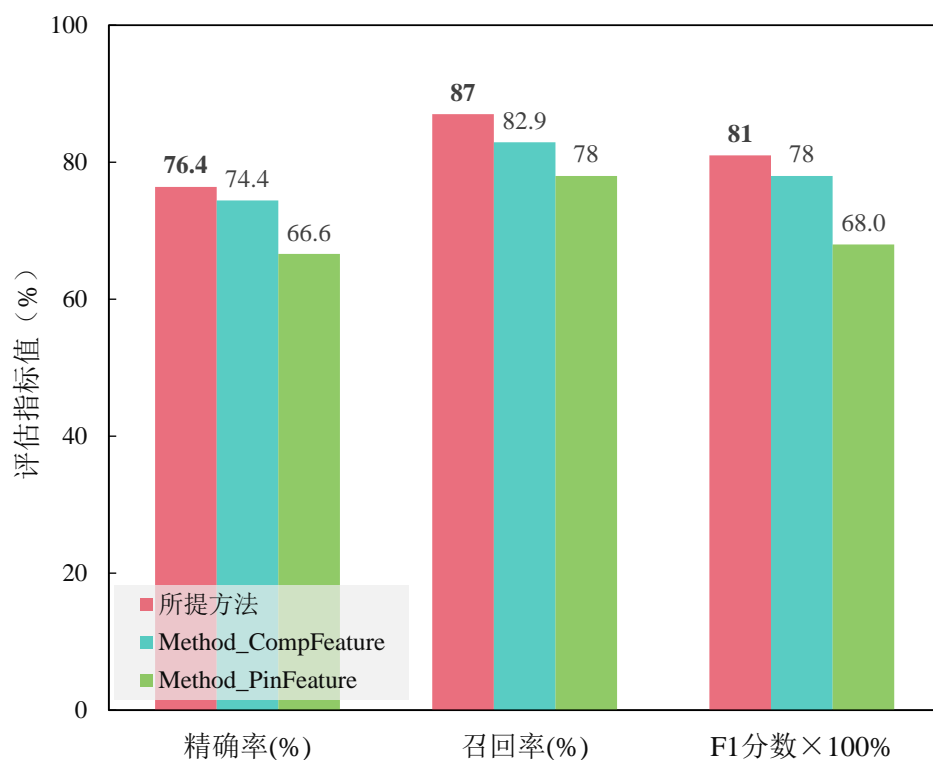


图 4.5 分层特征消融实验结果

在图4.5中,使用不同颜色的柱状图表示不同方法的性能。同时,在实验结果可视化时,为了使 F1 分数的数值范围与精确率、召回率保持一致,将 F1 分数乘以 100%,以百分比形式进行展示。

将分层特征中的引脚级特征和元件级特征分别消融时,模型性能出现下降,这验证了分层特征的有效性。引脚级特征可以较为精细地刻画两引脚元件的缺陷程度,但是在对其余引脚数量元件的缺陷程度进行刻画时,会损失部分有效信息。元件级特征可以从整体上刻画两引脚及多引脚元件的缺陷程度,但是会损失部分原始数据中的细节信息。而分层特征可以结合引脚级特征和元件级特征的优势,从引脚级和元件级全面表征元件的缺陷程度,从而便于模型判定其可修复性。

(2) 双阶段任务演进 TabNet 训练策略消融实验

本消融实验旨在验证双阶段任务演进 TabNet 训练策略的有效性。因此,本实验仅对双阶段任务演进 TabNet 训练策略进行消融。这里使用小样本复检缺陷元件数据直接训练 TabNet 模型,将其与采用双阶段任务演进训练策略的 TabNet 模型进行效果对比。

本消融实验的结果如图4.6所示。图中使用不同颜色的柱状图表示不同方法的性能。同时,在实验结果可视化时,为了使 F1 分数的数值范围与精确率、召回率保持一致,将 F1 分数乘以 100%,以百分比形式进行展示。

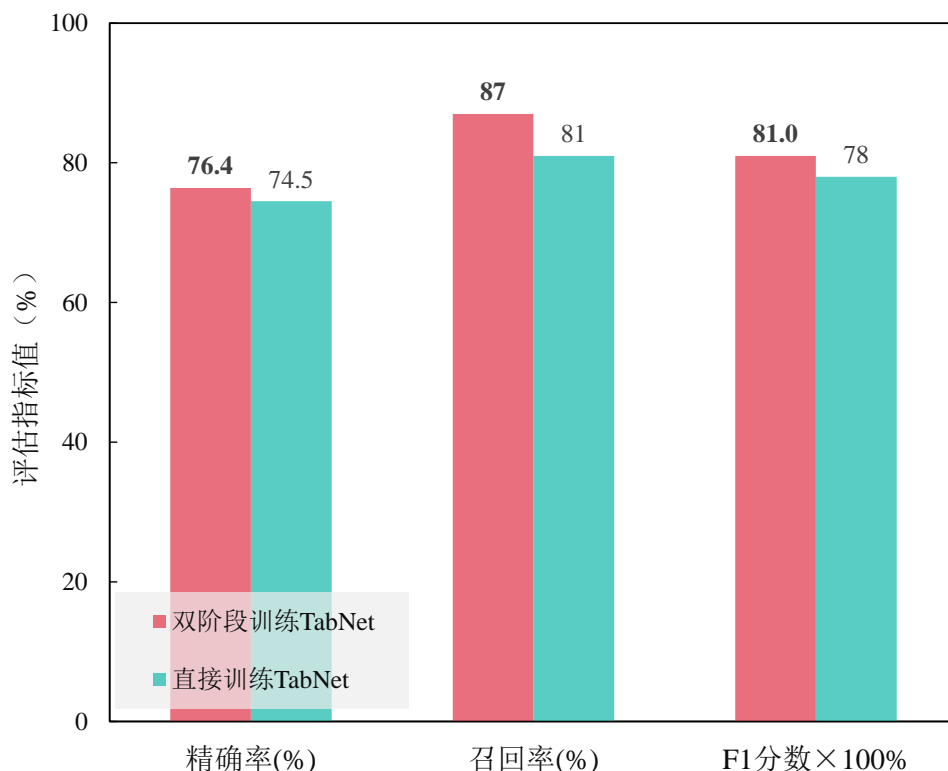


图 4.6 双阶段任务演进 TabNet 训练策略消融实验结果

通过图中结果可以看出,对于采用了双阶段任务演进 TabNet 训练策略的

TabNet 模型，其性能明显优于采用小样本复检缺陷元件数据直接训练的 TabNet 模型。具体而言，采用了双阶段任务演进 TabNet 训练策略的 TabNet 模型在精确率上提升了 1.9%，在召回率上提升了 6.0%，在 F1 分数上提升了 0.03。这验证了双阶段任务演进 TabNet 训练策略的有效性，说明所提训练策略可以提高模型在小样本数据上的性能。

3. 特征重要性分析

TabNet 模型可以评估决策过程中不同特征的重要性。因此，本章在此利用 TabNet 评估了缺陷元件可修复性判定任务中不同特征的相对重要性，并在图4.7中展示了相对重要性排名前 10 的特征。

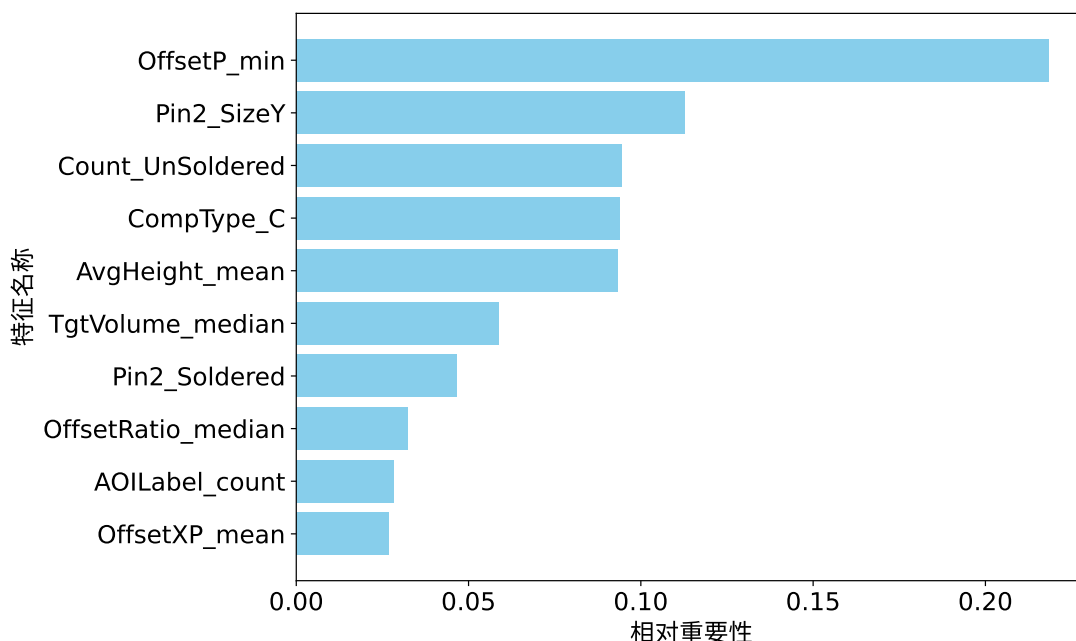


图 4.7 特征相对重要性示意图

通过图4.7可以观察到，特征重要性排名前 10 的特征中，有 2 个引脚级特征，即 Pin2_SizeY 和 Pin2_Soldered，剩余 8 个为元件级特征。同时，在重要性排名前 10 的特征中，有 4 个特征是在本章生成的引脚级特征上提取的元件级统计特征，分别为：OffsetP_min、AvgHeight_mean、TgtVolume_median 和 OffsetRatio_median。这些引脚级生成特征是本章在理解原始引脚级特征物理意义的基础上，由原始引脚级特征计算而来。这说明分层特征是有效的，其中的引脚级特征和元件级特征均起到作用。

特征相对重要性排名前 10 的特征中，有 3 个特征反映的是锡膏偏移的程度，分别是 OffsetP_min、OffsetRatio_median 和 OffsetXP_mean。这说明锡膏的偏移对分类结果有重要影响。CompType_C 表示元件的类型，这说明元件的类型也很重要。AOILabel_count 反映的是元件被 AOI 机器报告的缺陷条目数量。直观上，被报告的缺陷条目数量越多，反映该元件的缺陷程度越大，其修复可能性越低。

4. 模型在实际产线中的应用可行性分析

为评估所提模型在实际产线中的部署与应用可行性，本文对模型的计算复杂度与实时性进行了测试与分析。考虑到工业现场通常不具备 GPU 或高性能计算资源，测试在一台搭载 Intel Core i5 处理器（16GB 内存，无 GPU 加速）的通用设备上进行。结果表明，所训练模型在推理阶段的内存占用约为 0.07MB，远低于当前常见处理器可用内存（通常为 GB 级别）；同时，模型对单个元件进行可修复性判定的平均推理耗时为 2.01ms，在无需硬件加速的条件下也能实现毫秒级响应。相比之下，人工判定元件可修复性的过程通常耗时较长，时间单位多以秒甚至分钟计算。因此，本文方法具备良好的资源适配性与实时性，能够在不增加额外硬件成本的前提下部署于现有产线设备中，满足实际工业应用需求。

4.5 本章小结

针对缺陷元件可修复性判定问题，考虑到数据样本量小的难点，本章提出了一种基于任务演进 TabNet 训练策略的缺陷元件可修复性判定方法，实现了比现有方法更优的自动判定性能。首先，通过设计分层特征提取方法，在引脚层级和元件层级分别构建并提取特征，从而充分利用原始数据信息，并从不同层次全面表征元件的缺陷程度，以便模型更准确地判定复检缺陷元件的可修复性。其次，基于这些分层特征，提出双阶段任务演进 TabNet 训练策略。在缺陷感知预训练阶段，利用大量 AOI 初检缺陷元件数据对 TabNet 进行训练，使其实现元件缺陷的自动复检；在可修复性认知迁移阶段，调整 TabNet 的输出层结构，并结合少量复检缺陷元件数据对 TabNet 进行再训练，从而进一步提升其在可修复性判定任务中的性能。最后，在 PHME2022 数据集上进行了对比实验和消融实验。对比实验的结果表明，与现有方法的最佳性能相比，所提方法的 F1 分数提高了 0.08，精确率提高了 7.9%，召回率提高了 6.5%，这验证了所提方法的优越性。此外，消融实验进一步验证了所提出的元件分层特征提取方法和双阶段任务演进 TabNet 训练策略各自的有效性。

第 5 章 总结与展望

5.1 总结

现有 SMT 产线在缺陷元件复检和缺陷元件可修复性判定环节高度依赖人工，这会耗费较大的人力成本，并影响生产效率。而实现缺陷元件的自动复检和缺陷元件可修复性的自动判定，可以节省人力成本，并提高产线的自动化水平及生产效率。

但是，在实现自动的缺陷元件复检及其可修复性判定的过程中，面临以下难点：在缺陷元件自动复检任务中，复检无缺陷元件的数量远超复检缺陷元件的数量，存在数据类别高度不平衡的难点；在缺陷元件可修复性自动判定任务中，复检缺陷元件的数量少，存在数据样本量小的难点。目前相关研究较少，尚处于初步探索阶段，未充分关注并深入解决上述难点，所以取得的效果有限。

因此，本文对缺陷元件自动复检任务和缺陷元件可修复性自动判定任务展开研究，并针对上述难点分别提出了有效解决方案，相比于现有研究，提升了缺陷元件自动复检和缺陷元件可修复性自动判定的性能。

本文的具体研究工作和创新点总结如下：

(1) 考虑到数据类别高度不平衡的难点，提出了一种基于深度自编码 SMOTE 的 PCB 缺陷元件复检方法，实现了比现有方法更优的自动复检性能。首先，结合深度学习与 SMOTE 方法，提出了一种针对 PCB 元件数据生成的深度自编码 SMOTE 数据增强方法。该方法通过引脚行重组的方式扩充原始不平衡数据集，并利用扩充的不平衡数据集训练一个由编码器和解码器构成的数据重构模型。深度自编码 SMOTE 利用已训练编码器将原始数据集中的复检缺陷元件数据编码到低维特征空间，并利用 SMOTE 方法在编码特征空间内生成新特征，之后利用已训练解码器将 SMOTE 生成的新特征解码为新的复检缺陷元件数据，从而平衡数据集。其次，为充分挖掘数据中的有效信息，并进一步缓解不平衡数据带来的不利影响，提出了一种层次特征融合分类模型。该模型通过神经网络在引脚层级和元件层级分别进行深度特征融合，并进行元件缺陷评估，输出复检分类结果。最后，在 PHME2022 数据集上进行了对比实验及消融实验，验证了所提方法的有效性和优越性。相比于现有方法的最佳性能，所提方法的 F1 分数提高了 0.09，精确率提高了 6.9%，召回率提高了 8.3%。

(2) 考虑到数据样本量小的难点，提出了一种基于任务演进 TabNet 训练策略的 PCB 缺陷元件可修复性判定方法，实现了比现有方法更优的自动判定性能。首先，通过设计分层特征提取方法，在引脚层级和元件层级分别构建并提取特征，从而充分利用原始数据信息，并从不同层次全面刻画元件的缺陷程度，以便

模型更准确地判定缺陷元件的可修复性。其次,采用双阶段任务演进训练策略对 TabNet 进行训练。在缺陷感知预训练阶段,利用大量的 AOI 初检缺陷元件数据对 TabNet 进行预训练,使 TabNet 实现对 AOI 初检缺陷元件的复检分类。通过这一训练过程,TabNet 得以从大量的 AOI 初检缺陷元件数据中学习如何评估元件的缺陷程度,从而提升 TabNet 对元件缺陷的感知能力。在可修复性认知迁移阶段,调整 TabNet 的输出层结构,并结合少量复检缺陷元件数据对 TabNet 进行再训练,进一步提升其在可修复性判定任务中的性能,使其适配目标任务。最后,在 PHME2022 数据集上进行了对比实验及消融实验,验证了所提方法的有效性和优越性。相比于现有方法的最佳性能,所提方法的 F1 分数提高了 0.08,精确率提高了 7.9%,召回率提高了 6.5%。

5.2 展望

本文针对 SMT 实际产线中的缺陷元件复检任务和缺陷元件可修复性判定任务开展数据驱动算法研究,取得了一定的进展。然而,本文的研究仍存在一些不足之处,可进一步改进:

(1) 在缺陷元件自动复检任务中,提出了一种层次特征融合分类模型。尽管该模型能够在不同层级进行深度特征融合,但其主要依赖 MLP 或卷积神经网络进行特征提取,没有显式建模元件各引脚之间的物理相邻或电气连接关系。而图神经网络能够通过构建图结构,充分捕捉引脚间的拓扑关系与信号传输特性。因此,未来可以考虑将图神经网络引入到层次特征融合模型中,从而更好地从整体评估元件的缺陷程度,提高模型的自动复检性能。

(2) 在缺陷元件可修复性自动判定任务中,提出了一种元件分层特征提取方法。该方法共提取了 192 个特征,以全面表征元件的缺陷程度,进而便于模型进行可修复性判定。然而,其中可能包含部分无用或冗余特征,这些特征可能影响模型对元件可修复性的准确判定。因此,未来的研究可以进一步探索特征选择方法,通过递归特征消除或主成分分析等方法去除无用特征或冗余特征,以降低特征维度,提高模型的自动判定性能。

(3) 本文仅依赖于 SPI 和 AOI 数据进行缺陷元件自动复检及其可修复性自动判定,未来可以考虑引入 SMT 产线中的更多生产过程数据,如回流炉的温度曲线、PCB 图像数据等,获取更多有效信息,从而进一步提高模型性能。

参考文献

- [1] ULGER F, YUKSEL S E, YILMAZ A, et al. Solder joint inspection on printed circuit boards: A survey and a dataset[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-21.
- [2] 郭静. SMT 产线质量因素关联分析及参数优化方法研究[D]. 西安: 西安电子科技大学, 2022.
- [3] DWIVEDI G, PENSIA L, LOHCHAB V, et al. Nondestructive inspection and quantification of soldering defects in PCB using an autofocusing digital holographic camera[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-8.
- [4] CAO N, WON D, YOON S W. PADS: Predictive anomaly detection for SMT solder joints using novel features from SPI and pre-AOI data[J]. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2024, 14(3): 501-509.
- [5] LAU J H, KO C T, LIN C, et al. Fan-out panel-level packaging of mini-LED RGB display[J]. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2021, 11(5): 739-747.
- [6] WELLS E W, SIGMARSSON H H, MCDANIEL J W. A surface-mountable suspended integrated strip-line technology using castellated vias[J]. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2023, 13(7): 1067-1069.
- [7] SHAMKHALICHENAR H, BUECHE C J, CHOI J W. Printed circuit board (PCB) technology for electrochemical sensors and sensing platforms[J]. Biosensors, 2020, 10(11): 159.
- [8] 刘潇龙, 李鑫, 朱孟达, 等. 基于知识图谱的 SMT 产线工艺推荐与优化[J]. 机械设计与制造工程, 2024, 53(6): 122-126.
- [9] 白硕. 基于人工智能算法的 SMT 产线质量预测与优化方法研究[D]. 西安: 西安电子科技大学, 2024.
- [10] ROKACH L, HUTTER D. Automatic discovery of the root causes for quality drift in high dimensionality manufacturing processes[J]. Journal of Intelligent Manufacturing, 2012, 23: 1915-1930.
- [11] XIE J, GUO Y, LIU D, et al. A multimodal fusion method for soldering quality online inspection[J]. Journal of Intelligent Manufacturing, 2024, 34: 1-14.
- [12] LIU B, KANG Y, ZHAO Y B, et al. PCB layout-based spatiotemporal graph convolution network for anomaly prediction in solder paste printing[J]. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2025, 15(1): 214-223.
- [13] LAI Y, HA J H, DEO K A, et al. Reflow recipe establishment based on CFD-Informed ma-

- chine learning model[J]. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2023, 13(1): 127-134.
- [14] 卞佰成, 陈田, 吴入军, 等. 基于改进 YOLOv3 的印刷电路板缺陷检测算法[J]. 浙江大学学报 (工学版), 2023, 57(4): 735-743.
- [15] FARRAG A, KATAOKA J, YOON S W, et al. SRP-PINN: A physics-informed neural network model for simulating thermal profile of soldering reflow process[J]. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2024, 14(6): 1098-1105.
- [16] WU H, LEI R, PENG Y. PCBNet: A lightweight convolutional neural network for defect inspection in surface mount technology[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-14.
- [17] GORE P, MINAMI T, KUNDU P, et al. A novel methodology for health assessment in printed circuit boards[C]//PHM Society European Conference: Vol. 7. PHM Society, 2022: 556-562.
- [18] TANG H, TIAN Y, DAI J, et al. Prediction of production line status for printed circuit boards [C]//PHM Society European Conference: Vol. 7. PHM Society, 2022: 563-570.
- [19] GAFFET A, ROA N B, RIBOT P, et al. A hierarchical xgboost early detection method for quality and productivity improvement of electronics manufacturing systems[C]//PHM Society European Conference: Vol. 7. PHM Society, 2022: 541-549.
- [20] SCHMIDT I, DINGELDEIN L, HÜNEMOHR D, et al. Application of machine learning methods to predict the quality of electric circuit boards of a production line[C]//PHM Society European Conference: Vol. 7. PHM Society, 2022: 550-555.
- [21] 张永清, 卢荣钊, 乔少杰, 等. 一种基于样本空间的类别不平衡数据采样方法[J]. 自动化学报, 2022, 48(10): 2549-2563.
- [22] 朱诗能, 韩萌, 杨书蓉, 等. 不平衡数据流的集成分类方法综述[J]. 计算机工程与应用, 2025, 61(2): 59-72.
- [23] LIN C, TSAI C F, LIN W C. Towards hybrid over-and under-sampling combination methods for class imbalanced datasets: an experimental study[J]. Artificial Intelligence Review, 2023, 56(2): 845-863.
- [24] BAI D, LI G, JIANG D, et al. Surface defect detection methods for industrial products with imbalanced samples: A review of progress in the 2020s[J]. Engineering Applications of Artificial Intelligence, 2024, 130: 107697.
- [25] SHAMSUDIN H, YUSOF U K, JAYALAKSHMI A, et al. Combining oversampling and undersampling techniques for imbalanced classification: A comparative study using credit card fraudulent transaction dataset[C]//IEEE 16th International Conference on Control & Automation (ICCA). IEEE, 2020: 803-808.
- [26] BANSAL A, JAIN A. Analysis of focussed under-sampling techniques with machine learning

- classifiers[C]//IEEE 19th International Conference on Software Engineering Research, Management and Applications (SERA). IEEE, 2021: 91-96.
- [27] 颜学明, 黄翰, 金耀初, 等. 面向不平衡短文本情感多分类的三阶语义图数据增广方法[J]. 计算机学报, 2024, 47(12): 2742-2759.
- [28] AZLIM KHAN A K, AHAMED HASSAIN MALIM N H. Comparative studies on resampling techniques in machine learning and deep learning models for drug-target interaction prediction [J]. *Molecules*, 2023, 28(4): 1663.
- [29] AZHAR N A, POZI M S M, DIN A M, et al. An investigation of SMOTE based methods for imbalanced datasets with data complexity analysis[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(7): 6651-6672.
- [30] KHAIRY M, MAHMOUD T M, ABD-EL-HAFEEZ T. The effect of rebalancing techniques on the classification performance in cyberbullying datasets[J]. *Neural Computing and Applications*, 2024, 36(3): 1049-1065.
- [31] ELREEDY D, ATIYA A F, KAMALOV F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning[J]. *Machine Learning*, 2024, 113(7): 4903-4923.
- [32] WONGVORACHAN T, HE S, BULUT O. A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining [J]. *Information*, 2023, 14(1): 54.
- [33] 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.
- [34] MUKHERJEE M, KHUSHI M. SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features[J]. *Applied System Innovation*, 2021, 4(1): 18.
- [35] ZHOU H, YU K M, CHEN Y C, et al. A hybrid feature selection method RFSTL for manufacturing quality prediction based on a high dimensional imbalanced dataset[J]. *IEEE Access*, 2021, 9: 29719-29735.
- [36] FERNÁNDEZ A, GARCÍA S, GALAR M, et al. *Learning from imbalanced data sets*[M]. Springer, 2018.
- [37] TAN H, WANG L, ZHU D, et al. Intrusion detection based on adaptive sample distribution dual-experience replay reinforcement learning[J]. *Mathematics*, 2024, 12(7): 948.
- [38] JAISWAL A, BABU A R, ZADEH M Z, et al. A survey on contrastive self-supervised learning [J]. *Technologies*, 2020, 9(1): 2.
- [39] 李志鹏, 国雍, 陈耀佛, 等. 基于数据生成的类别均衡联邦学习[J]. 计算机学报, 2023, 46(3): 609-625.

- [40] FAJARDO V A, FINDLAY D, JAISWAL C, et al. On oversampling imbalanced data with deep conditional generative models[J]. *Expert Systems with Applications*, 2021, 169: 114463.
- [41] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [42] YANG Z, XU B, LUO W, et al. Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review[J]. *Measurement*, 2022, 189: 110460.
- [43] LUO W, YANG W, HE J, et al. Fault diagnosis method based on two-stage GAN for data imbalance[J]. *IEEE Sensors Journal*, 2022, 22(22): 21961-21973.
- [44] CHEN C, WANG T, LU K, et al. Compact convolutional transformers-generative adversarial network for compound fault diagnosis of industrial robot[J]. *Engineering Applications of Artificial Intelligence*, 2024, 138: 109315.
- [45] 吴美君, 杨新, 潘超凡, 等. 自编码器结合持续学习: 现状、挑战与展望[J]. *计算机学报*, 2025, 48(2): 317-357.
- [46] DABLAIN D, KRAWCZYK B, CHAWLA N V. DeepSMOTE: Fusing deep learning and SMOTE for imbalanced data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(9): 6390-6404.
- [47] DING M. The road from MLE to EM to VAE: A brief tutorial[J]. *AI Open*, 2022, 3: 29-34.
- [48] FAN S K S, TSAI D M, YEH P C. Effective variational-autoencoder-based generative models for highly imbalanced fault detection data in semiconductor manufacturing[J]. *IEEE Transactions on Semiconductor Manufacturing*, 2023, 36(2): 205-214.
- [49] ABDULGANIYU O H, TCHAKOUCHE T A, SAHEED Y K, et al. XIDINTFL-VAE: Xgboost-based intrusion detection of imbalance network traffic via class-wise focal loss variational autoencoder[J]. *The Journal of Supercomputing*, 2025, 81(1): 1-38.
- [50] RAZEGHI B, CALMON F P, GUNDUZ D, et al. Bottlenecks CLUB: Unifying information-theoretic trade-offs among complexity, leakage, and utility[J]. *IEEE Transactions on Information Forensics and Security*, 2023, 18: 2060-2075.
- [51] PEI Z, JIANG H, LI X, et al. Data augmentation for rolling bearing fault diagnosis using an enhanced few-shot wasserstein auto-encoder with meta-learning[J]. *Measurement Science and Technology*, 2021, 32(8): 84007.
- [52] HARIKRISHNAN C, DHANYA N. Improving text classifiers through controlled text generation using transformer wasserstein autoencoder[C]//*Inventive Communication and Computational Technologies*. Springer, 2022: 97-105.
- [53] ROSS T Y, DOLLÁR G. Focal loss for dense object detection[C]//*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017: 2980-2988.
- [54] ROY D, PRAMANIK R, SARKAR R. Margin-aware adaptive-weighted-loss for deep learning

- based imbalanced data classification[J]. IEEE Transactions on Artificial Intelligence, 2023, 5(2): 776-785.
- [55] 李长洪, 郑凯, 林博宇. 针对不平衡数据分类的改进 GBDT 算法[J]. 计算机与数字工程, 2024, 52(7): 1932-1937.
- [56] YANG Y, LV H, CHEN N. A survey on ensemble learning under the era of deep learning[J]. Artificial Intelligence Review, 2023, 56(6): 5545-5589.
- [57] 翟利志, 任一夫, 白洁, 等. 基于传递式领域自适应的异构样本增强方法[J]. 计算机技术与发展, 2024, 34(1): 17-22.
- [58] ZENG W, XIAO Z Y. Few-shot learning based on deep learning: A survey[J]. Mathematical Biosciences and Engineering, 2024, 21(1): 679-711.
- [59] 史燕燕, 史殿习, 乔子腾, 等. 小样本目标检测研究综述[J]. 计算机学报, 2023, 46(8): 1753-1780.
- [60] VILALTA R, DRISSI Y. A perspective view and survey of meta-learning[J]. Artificial Intelligence Review, 2002, 18: 77-95.
- [61] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述[J]. 软件学报, 2021, 32(2): 349-369.
- [62] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta-learning with memory-augmented neural networks[C]//International Conference on Machine Learning (ICML). PMLR, 2016: 1842-1850.
- [63] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//International Conference on Machine Learning (ICML). PMLR, 2017: 1126-1135.
- [64] JAMAL M A, QI G J. Task agnostic meta-learning for few-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 11719-11727.
- [65] LIU X, LUAN X, XIE Y, et al. Transfer learning research and algorithm review[J]. Journal of Changsha University, 2018, 32(5): 29-36.
- [66] 赵嘉晖, 温杰, 蔡星娟, 等. 基于高维多目标优化的少样本皮肤癌检测[J]. 控制与决策, 2024, 39(11): 3597-3606.
- [67] 安胜彪, 郭昱岐, 白宇, 等. 小样本图像分类研究综述[J]. 计算机科学与探索, 2023, 17(3): 511-532.
- [68] NAKAMURA A, HARADA T. Revisiting fine-tuning for few-shot learning[A]. 2019. arXiv: 1910.00216.
- [69] KAN S, HE Z, CEN Y, et al. Contrastive bayesian analysis for deep metric learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(6): 7220-7238.
- [70] ZHAO W, RAO Y, ZHOU J, et al. DIML: Deep interpretable metric learning via structural

- matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(4): 2518-2532.
- [71] 郭礼华, 王广飞. 基于任务感知关系网络的少样本图像分类[J]. 电子与信息学报, 2024, 46(3): 977-985.
- [72] KOCH G, ZEMEL R, SALAKHUTDINOV R, et al. Siamese neural networks for one-shot image recognition[C]//ICML Deep Learning Workshop: Vol. 2. Lille, 2015: 1-30.
- [73] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning [J]. Advances in Neural Information Processing Systems, 2016, 29: 1-9.
- [74] LI W, XU J, HUO J, et al. Distribution consistency based covariance metric networks for few-shot learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 33. AAAI, 2019: 8642-8649.
- [75] 陈良臣, 傅德印. 面向小样本数据的机器学习方法研究综述[J]. 计算机工程, 2022, 48 (11): 1-13.
- [76] 王振东, 徐振宇, 李大海, 等. 面向入侵检测的元图神经网络构建与分析[J]. 自动化学报, 2023, 49(7): 1530-1548.
- [77] GIDARIS S, KOMODAKIS N. Generating classification weights with GNN denoising autoencoders for few-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 21-30.
- [78] 李凡, 贾东立, 姚昱旻, 等. 结合残差与自注意力机制的图卷积小样本图像分类网络[J]. 计算机科学, 2023, 50(S1): 376-380.
- [79] 刘颖, 雷研博, 范九伦, 等. 基于小样本学习的图像分类技术综述[J]. 自动化学报, 2021, 47(2): 297-315.
- [80] 刘杨. 基于自编码器的深度聚类算法研究[D]. 无锡: 江南大学, 2024.
- [81] 陶文彬, 钱育蓉, 张伊扬, 等. 基于自编码器的深度聚类算法综述[J]. 计算机工程与应用, 2022, 58(18): 16-25.
- [82] 田瑞端. 基于自编码器的电力能耗异常监测研究[J]. 自动化应用, 2025, 66(2): 102-104+108.
- [83] 袁振, 刘进锋. 基于有损压缩编码的降噪自编码器[J]. 计算机科学, 2024, 51(S1): 472-478.
- [84] 蒋宗礼, 史倩月. 面向不平衡数据的分类算法[J]. 计算机系统应用, 2019, 28(8): 120-128.
- [85] LI X, XIE K, WANG X, et al. A light-weight and robust tensor convolutional autoencoder for anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(9): 4346-4360.
- [86] ARIK S Ö, PFISTER T. TabNet: Attentive interpretable tabular learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 35. AAAI, 2021: 6679-6687.
- [87] MASELLO L, SHEEHAN B, CASTIGNANI G, et al. Predictive modeling for driver insurance

- premium calculation using advanced driver assistance systems and contextual information[J].
IEEE Transactions on Intelligent Transportation Systems, 2025, 26(2): 2202-2211.
- [88] 李洪剑. 基于 TabNet 模型的糖尿病分类方法研究[D]. 成都: 西南交通大学, 2023.
- [89] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep
convolutional generative adversarial networks[A]. 2016. arXiv: 1511.06434.

致谢

时光荏苒，三年的研究生生涯即将画上句点，这也意味着我长达二十余载的学生时代即将落下帷幕。在整理论文、准备踏上新征程的此刻，心中满是感慨，也更加清晰地意识到，这段旅程的顺利完成离不开太多人的帮助与支持，在此向所有关心和帮助过我的人，献上最诚挚的谢意。

感谢导师赵云波教授。赵老师严谨的治学态度与前瞻性的学术视野，为我打开了科研世界的大门。正是赵老师的严格教导与悉心栽培，让我一步步地学会了发现问题、定义问题和解决问题的方法，同时培养了对科研工作的深刻品味。这种严谨务实的学术精神，将成为我一生的财富。

感谢刘斌琨师兄。在研究的初始阶段，我一直与师兄进行讨论交流，逐步建立起科学的研究思维。在论文撰写环节，师兄凭借丰富的科研与写作经验，多次与我就论文结构、逻辑框架、内容表述等方面进行深入探讨，对论文中的专业术语运用、数据图表呈现等细节提出诸多宝贵建议。

感谢李鹏飞研究员、许镇义研究员和陈绍冯博后。李鹏飞师兄、许镇义师兄和陈绍冯师姐多次在我面对研究困境时，为我提供全新的思考角度，引导我突破思维瓶颈，并在日常研究中给予了我诸多实际帮助。

感谢求学期间结识的所有朋友。在实验室忙碌的间隙，在图书馆伏案学习的疲惫时刻，是你们的陪伴让我的研究生生活变得充实而多彩。我们曾一起为了攻克课程难题挑灯夜战，分享学习心得；也曾在课余时间相约运动、聚餐，释放学习压力。你们的鼓励和支持，让我在面对困难和挫折时不再感到孤单，让这段求学时光充满温暖与欢乐。

感恩我的父母和女朋友。父母多年来含辛茹苦的养育，始终是最坚实的后盾。而我的女朋友李帅，在精神上始终坚定地支持我，让我能够心无旁骛地投入到学习和研究中。

最后，也要感谢坚持到现在的自己。求学之路并非一帆风顺，面对科研上的重重困难、学业与生活的压力，我曾迷茫、焦虑，但最终凭借自身的努力与坚持，一步一个脚印地完成了这段旅程。

再次向所有帮助过我的人致以最衷心的感谢。这份情谊我将永远铭记于心，未来我也会带着这份感恩与力量，在新的人生阶段继续努力前行，在专业领域不断探索进取，以实际行动回报大家的关爱与支持。

在读期间取得的科研成果

已发表论文

1. **Longxin Chen**, Yunbo Zhao, Binkun Liu, Shaojie Dong, Huijuan Zhu, and Peng Bai. A Reliable Ensemble Model based on Hierarchical Component Features for Repair Label Prediction of Soldering Defects[C]//2024 14th Asian Control Conference (ASCC). IEEE, 2024: 1563-1568.
2. Yunbo Zhao, Shaojie Dong, Yu Kang, Kangcheng Wang, **Longxin Chen**, and Peng Bai. Prediction of Yield in Functional Testing of Motherboards in Laptop Manufacturing[C]//2024 14th Asian Control Conference (ASCC). IEEE, 2024: 1-5.

已公开专利

1. 赵云波, **陈龙鑫**, 刘斌琨, 朱慧娟, 许镇义, 柏鹏. 故障诊断模型的训练方法、装置、电子设备及存储介质 [P]. 安徽省:CN202310901745.4,2023-09-19.

已授权专利

1. 赵云波, 刘斌琨, 康宇, 曹洋, **陈龙鑫**, 李佳玉. 锡膏印刷质量异常预测方法、系统、设备及存储介质 [P]. 安徽省:CN202411566368.4,2025-01-21.
2. 赵云波, 刘斌琨, 康宇, 曹洋, 余碧桢, **陈龙鑫**. 液压减震器无监督故障检测方法、系统、设备与存储介质 [P]. 安徽省:CN202411593785.8,2025-01-28.

已登记软件著作权

1. 赵云波, **陈龙鑫**, 朱慧娟, 康宇, 许镇义. 锡膏印刷机在线故障预测软件 (V1.0).2022.