

DOI: 10.19650/j.cnki.cjsi.J2514670

振动特征逻辑空间分组下的微凸点可靠性检测*

孙 钰^{1,2}, 宿 磊², 李 可², 曹 澍¹, 彭广盼¹

(1. 无锡职业技术大学机械工程学院 无锡 214121; 2. 江南大学智能制造学院 无锡 214122)

摘 要:倒装芯片微凸点振动信号存在微尺度特征、检测信号强干扰及多模式成分混淆等问题,导致缺陷特征映射关系复杂、缺陷识别率显著降低,严重制约缺陷识别精度与工业实用性,成为微纳电子制造缺陷检测领域亟待解决的关键瓶颈。为此,针对现有知识蒸馏方法未充分挖掘逻辑知识、类间关系利用薄弱,难以实现细粒度缺陷精准识别等问题,提出特征逻辑空间分组蒸馏(FLSGD)算法。该算法围绕知识传递逻辑解析显性与隐性知识,将交互信息投影至三概率空间,即目标空间 t 、类内空间 $o|t$ 及类外空间 $c|o$,解耦传统 Kullback-Leibler (KL) 散度实现逻辑特征定向学习;采用主客观权重法量化三空间权重,确定参数指导值,灵活调控知识传递侧重。为验证算法有效性,实验基于 PCB 基板与硅基芯片两组超声振动数据集,搭建单/双源超声激励检测平台开展验证,论证了 FLSGD 中目标空间起主导作用,类内、类外空间起辅助作用。实验结果表明,FLSGD 在多种教师-学生网络组合中,识别精度优于主流特征蒸馏方法,在两组数据集下的最高精度分别达到 96.47% 与 97.69%。FLSGD 能够在未增加计算模块与算力消耗的基础上有效挖掘振动信号模态特性,增强关键特征感知,提高缺陷识别精度,加速模型收敛,为工业端轻量级缺陷检测提供有效方案。

关键词: 倒装芯片; 振动信号; 知识蒸馏; 主客观权重法; 特征解耦

中图分类号: TH89 **文献标识码:** A **国家标准学科分类代码:** 510.40

Reliability detection of microbumps based on vibration feature logic space grouping

Sun Yu^{1,2}, Su Lei², Li Ke², Cao Shu¹, Peng Guangpan¹

(1. School of Mechanical Engineering, Wuxi University of Technology, Wuxi 214121, China;

2. School of Intelligent Manufacturing, Jiangnan University, Wuxi 214122, China)

Abstract: Flip-chip microbump vibration signals are characterized by microscale features, strong interference in acquired signals, and severe confusion of multi-modal components, which give rise to complicated mapping relationships of defect features and a remarkable drop in defect recognition accuracy. It severely restricts the accuracy of defect identification and industrial applicability, becoming a key bottleneck that urgently needs to be addressed in the field of micro-nano electronic manufacturing defect detection. To tackle the deficiencies of existing knowledge distillation, including insufficient exploitation of logical knowledge, weak utilization of inter-class relationships, and limited capacity to achieve accurate defect identification, this article proposes a feature logic space grouping distillation (FLSGD) algorithm. This algorithm analyzes explicit and implicit knowledge around the logic of knowledge transfer, and projects interactive information into three probability spaces, namely target space t , intra-class space $o|t$ and outer-class space $c|o$. It decouples the traditional Kullback-Leibler (KL) divergence to realize directed learning of logical features. The subjective and objective weighting method is adopted to quantify the weights of the three spaces and determine the parameter guidance values, so as to flexibly regulate the focus of knowledge transfer. To verify the effectiveness of the proposed algorithm, experiments are conducted on two ultrasonic vibration datasets derived from PCB substrate chip and silicon-based dummy chips. A single/dual-source ultrasonic excitation detection platform is established for validation, which verifies that the target space plays a dominant role in FLSGD, while the intra-class and outer-class spaces serve auxiliary functions. Experimental results demonstrate that FLSGD outperforms mainstream feature distillation

收稿日期: 2025-11-11 Received Date: 2025-11-11

* 基金项目: 国家自然科学基金项目 (U23B2044, 52375099)、国家重点研发项目 (2023YFB4404200) 资助

methods in terms of recognition accuracy across various teacher-student network combinations, achieving a maximum accuracy of 96.47% and 97.69% on the two datasets, respectively. FLSSGD can effectively exploit the modal characteristics of vibration signals, enhance the perception of key features, improve defect recognition accuracy and accelerate model convergence speed without introducing additional computational modules or increasing computing power consumption, thus providing an effective solution for lightweight defect detection in industrial applications.

Keywords: flip-chip; vibration signal; knowledge distillation; subjective and objective weighting method; feature decoupling

0 引言

电子封装是集成电路工序中的重要环节,对集成电路产品的成本、性能和可靠性等都有重要影响,封装水平的高低直接影响着我国与西方发达国家在芯片领域的竞争结果^[1]。据统计,集成电路制造成本中40%用于封装,而超过25%的失效也源于封装。缺陷检测一直是工业领域失效模式分析的重点,为了保证产品的高性能和可靠性,对缺陷进行精确分类是缺陷检测领域的一个重要课题。

近年来,国内外学术界与工程界围绕集成电路封装失效分析及缺陷检测方法开展了大量研究与探索。现有倒装芯片无损检测技术体系中,主流方法包括扫描声学显微镜检测、X射线检测及红外热像检测等^[2-4],但各类技术均存在明确应用局限,如X射线检测无法适配14 nm以下制程的高端芯片,自动化光学检测等手段无法实现内部缺陷的有效检出,人工目检效率低等。综上,传统检测方法普遍存在破坏性强、检测效率低、精度不足等问题,发展高性能集成电路封装无损检测技术的需求尤为迫切。在缺陷损伤识别与故障诊断领域,振动检测技术应用成熟且广泛。振动噪声信号是反映状态的核心信息载体,蕴含丰富特征参数,其信息承载量和特征维度直接影响倒装芯片失效状态评估准确性,如何通过先进信号处理算法挖掘信号特性、提取缺陷特征并提升数据价值密度,是国内外检测领域的核心研究方向与攻关焦点。

近年来,深度学习在学术界与工业界表现卓越,尤其在工程检测与诊断领域成效显著。然而,当前深度模型在工业检测应用中仍存在典型短板:模型为提高精度常增设模块,导致结构愈发臃肿,而这正是工业检测系统落地的核心约束。针对这一瓶颈,网络压缩技术成为研究热点,知识蒸馏(knowledge distillation, KD)是其中极具应用价值的方向,可有效降低资源消耗。KD作为一种学习范例,用于将知识从大型教师模型转移到浅层学生模型(teacher to student, T-S),并提高性能。2015年, Hinton首次提出了KD的概念,它通过最小化KL(Kullback-Leibler)散度来传递信息^[5]。随后, KD逐渐变得多样和丰富,出现了许多有代表性的方法,如注意转移(attention transfer, AT)^[6]和因子转移(factor transfer,

FT)^[7]。2019年后, KD进入繁荣期,并应用于各种任务,如目标分类、目标检测、目标分割等^[8-10]。在工业检测和诊断领域也涌现了许多优秀的研究和应用。Wu等^[11]提出了一种基于KD和神经架构搜索的风电齿轮箱边缘故障诊断新方法; Deng等^[12]针对数据极不平衡故障诊断问题,提出了一种轻量级故障诊断框架,并用两个经典轴承数据集验证了所提框架的有效性与优越性; Zhang等^[13]提出了一种基于遮挡的重要性分析方法来选择最相关的输入变量和学习特征,并设计了一种改进的KD方法来训练两个典型动态系统的学生模型; Wang等^[14]设计了一种用于机械智能检测的多尺度密集连接深度网络,并在多个数据集上验证了该方法的优良性能。

现有方法虽取得一定成效,但仍存在4个方面核心局限:1)多数算法依赖硬标签训练机制,此类标签将类别信息单一聚焦于目标类别,忽略了不同缺陷间的内在关联,难以刻画缺陷样本的真实分布特性;2)利用知识蒸馏体系进行缺陷程度分级时,信息传递不足的问题阻碍了模型对相似度较高的相邻等级缺陷的学习能力,导致决策边界样本的学习强度不足,无法实现精准分类判断;3)传统算法对决策边界样本的分类效能欠佳,根源在于未考量类别间的关系差异,限制了损失函数的自适应调整方向;4)逻辑知识应该具有比深层特征、正则化和参数优化更高的语义层次,然而现有方法多局限于对原有框架的正则化与参数优化,未能构建基于逻辑知识的全新体系,即对逻辑信息的挖掘与利用存在显著不足。针对上述问题,本研究基于逻辑蒸馏范式深入解析原始蒸馏内在机理,提出特征逻辑空间分组蒸馏(feature logic space grouping distillation, FLSSGD)。具体工作为:

1) FLSSGD首次定义了三概率空间,即目标空间(t)、类内空间($o \setminus t$)及类外空间($c \setminus o$),通过解耦传统KL散度,实现对具有类边界可辨别性的逻辑特征表示定向学习,使知识传递过程兼具灵活性与可控性。

2)在两组倒装芯片振动信号数据集上,分别对FLSSGD中三概率空间的相互关系开展定性分析与定量建模;其中,定量分析阶段采用主客观权重分析方法,为三概率空间的权重系数确定提供量化指导。

3)本研究实现了知识蒸馏与倒装芯片微凸点缺陷检测任务的开创性结合,所提FLSSGD可实现振动信号中显性知识与隐性知识的精准对齐,有效克服信息传递空间中的

耦合矛盾;通过扩展对比实验,进一步验证了所提方法在倒装芯片微凸点缺陷检测任务中的有效性与稳健性。

1 相关工作

1.1 软标签的生成应用

采用硬标签(即标签类赋值为1、非标签类赋值为0的二元标记模式)训练深度网络时,模型易过度聚焦于类别间的绝对区分,忽略类间潜在关联与数据真实分布特性,进而引发过拟合问题。为缓解该问题,将硬标签转换为软标签以更精准反映类别差异,成为主流解决方案之一,现有软标签生成方法可归纳为以下两类:第1种为基于固定规则分配概率的方法。Szegedy等^[15]通过向硬标签引入恒定分布,将标签类的部分概率均匀分配至所有非标签类以生成软标签。该方法虽实现了标签软化,但对所有非标签类采取均匀处理方式,忽视了非标签类与标签类之间的内在差异及关联,难以匹配数据的真实分布特征。第2种为基于模型学习生成分布的方法。Wang等^[16]提出标签分布学习范式,以学习生成分布函数为核心目标,进而生成能够反映类别特征的可学习软标签。除此之外,Zhang等^[17]提出迭代式软标签生成方法,针对给定样本,利用与其同属一类的先前样本预测结果构建软标签;Wu等^[18]则设计了可生成偏向性软标签的程序,并通过实验验证了其有效性。尽管上述软标签分布模型在部分工业场景中具备应用价值,但仍存在两方面显著瓶颈:一方面,由于数据集的最佳目标分布无法预先知晓,模型需耗费大量调试试错时间以确定适配参数;另一方面,现有方法生成的软标签多以类别为单位构建,仅反映类别层面的整体差异,却忽略了同一类别内不同样本的个体间差异,难以适配复杂多变的缺陷特征分布。与上述方法不同,本研究所提方法以样本为独立单位生成软标签,能够精准捕捉单个样本的独特缺陷特征,更适配倒装芯片微凸点缺陷的复杂分布规律,为后续缺陷检测任务提供更具针对性的监督信号。

1.2 知识蒸馏

知识蒸馏 KD 构建了教师-学生的双模型学习范式,即通过性能更优的大尺度教师模型向轻量化学生模型传递知识,实现对学生模型训练过程的定向指导。根据知识传递载体的差异,KD可分为两类核心范式:基于标签的蒸馏^[19]与基于中间特征的蒸馏^[20]。如图1所示,基于中间特征的蒸馏以教师模型的中间层特征图激活值为监督标准,通过约束学生模型对应特征位置的激活分布,实现知识迁移。

与常规特征约束方法不同,Heo等^[21]提出创新思路,直接对特征映射的激活区域进行约束,旨在实现更精准

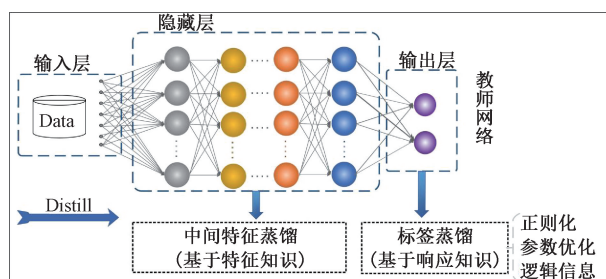


图1 知识蒸馏研究现状

Fig. 1 Research status of knowledge distillation

的边界信息传递,提升学生模型对关键特征的捕捉能力。基于标签的蒸馏(即基于逻辑的蒸馏)利用教师模型输出的逻辑信息(原始预测概率)为每个样本生成软标签,以更细粒度的类别概率分布为学生模型提供监督。得益于软标签包含的类别间差异信息,此类方法通常比传统硬标签更易获得优异性能。诊断证据蒸馏方法基于特征与逻辑的双蒸馏机制提升了缺陷识别精度,但引入了额外模块,导致计算复杂度增加,且未考虑缺陷类别间的相似性,难以适配类别边界模糊的场景^[22];投影集合特征蒸馏方法采用多分类器协同策略,降低了预测结果的基数冗余^[23];尺度解耦蒸馏方法在模型训练中引入多尺度设计,通过为样本不同区域分配差异化学习权重,优化模型学习效率^[24];逻辑标准化蒸馏允许学生模型关注教师模型中的逻辑基本关系,而不需要精确的量值匹配^[25];课程温度蒸馏方法引入动态温度调整方案,虽优化了软标签生成的适应性,但同时增加了蒸馏过程的调控难度^[26]。尽管上述方法在特定场景中取得显著效果,但仍存在两方面核心局限:1)非标签类处理同质化:对所有非标签类别采取无差别处理策略,忽略了当前样本与不同非标签类别间的差异程度,导致模型过度强调与标签类邻近的类特征,轻视远离标签类的类信息,难以全面捕捉类别分布规律;2)逻辑知识利用不充分:逻辑知识具有更高的语义层次,但其潜力尚未被充分释放,现有基于逻辑的蒸馏多聚焦于概率分布的表面匹配,未深入挖掘背后蕴含的类逻辑关系,导致性能提升受限。针对上述问题,本研究所提方法通过定向强调不同程度缺陷的空间特征差异,使模型能够更精准地捕捉类间层次关系,有效提升对边界模糊样本的识别能力,弥补了现有方法对非标签类差异忽视及逻辑知识利用不足的缺陷。

2 特征逻辑空间分组蒸馏

虽然神经网络极大地提高了检测精度,但也需更多的硬件资源,使得很难部署到实际应用终端设备上,亟需一个轻量、灵活匹配对象特性的高效检测网络。

2.1 T-S 知识传递

KD 是当前轻量技术的一种新兴方法,通过学生模仿老师以获得具有竞争力甚至更优秀的表现。常规深度网络的逻辑单元即 softmax 前一层,经过 softmax 函数后得到类概率:

$$p_i(z_i) = \frac{e(z_i)}{\sum_{j=1}^c e(z_j)} \quad (1)$$

式中: z_i 为逻辑单元值; p_i 为类概率; c 为类数量。

在 T-S 模型中,当使用逻辑单元来表示知识时易造

成噪声知识的误传输,使学生模型过拟合。因此, Hinton 引入参数 T 来平滑类概率,也称为温度系数(软化系数)。如图 2 所示,显示了网络中逻辑输出与类概率之间的关系,通过调整 T 教师模型可以将暗信息传递给学生模型。

$$p_i(z_i, T) = \frac{e\left(\frac{z_i}{T}\right)}{\sum_{j=1}^c e\left(\frac{z_j}{T}\right)} \quad (2)$$

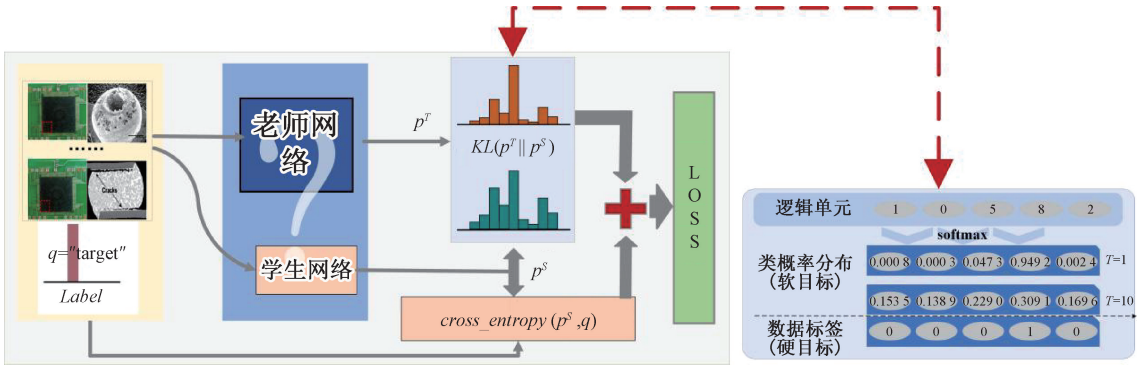


图2 知识蒸馏逻辑

Fig. 2 Knowledge distillation logic

当 $T=1$ 即普通的 softmax 输出, T 正无穷时转化为逻辑单元。目前, KD 中 T-S 之间的定义用 KL 散度来表示。

$$KD_1 = KL(p^T \| p^S) = \sum_{i=1}^c p_i^T \log\left(\frac{p_i^T}{p_i^S}\right) \quad (3)$$

式中: p^T 与 p^S 分别表示老师和学生类概率。通过调节 T , 老师模型将本身的“隐性知识”传输给学生模型。

2.2 特征逻辑空间分组

传统 KD 通过单一 KL 散度实现教师-学生模型的知识传递, 将目标类、类内非目标类、类外非目标类的概率信息高度耦合, 无法针对性挖掘不同概率空间的语义知识; 而在工业缺陷检测任务中, 不仅需要区分缺陷类别, 还需识别细粒度缺陷程度, 耦合的知识传递模式会导致细粒度缺陷的边界特征被淹没。因此, 以倒装芯片的缺陷识别为例, 将 KD 损失重新解耦为 3 组分加权和, 得到特征逻辑空间分组蒸馏算法 FLSSGD。FLSSGD 引入了特征逻辑概率空间, 分别为目标空间 t 、目标类内空间 o 及整体空间 c , 并定义以下符号 $b = [p_t, p_{o \setminus t}, p_{c \setminus o}]$ 。将整体概率空间 c (包含所有缺陷类别及程度) 按照“目标类别 → 同类内非目标 → 不同类非目标类别”的层级进行拆解, 满足互斥性与完备性, 即三空间无交集且空间并集为整体概率空间 c 。

$$\begin{cases} t \cup (o \setminus t) \cup (c \setminus o) = c \\ t \cap (o \setminus t) \cap (c \setminus o) = \emptyset \end{cases} \quad (4)$$

式中: 目标空间 t 表示待检测的具体目标缺陷; 类内空间 $o \setminus t$ 表示目标缺陷所属大类下的其他细粒度缺陷程度; 类外空间 $c \setminus o$ 表示非目标缺陷的所有其他类别。定义 p_t 表示目标 t 概率, $p_{o \setminus t}$ 表示在 o 空间内除目标 t 的概率, $p_{c \setminus o}$ 表示在整体空间 c 中排除 o 的概率。

$$\begin{cases} p_t = \frac{e(z_t)}{\sum_{j=1}^c e(z_j)} \\ p_{o \setminus t} = \frac{\sum_{k=1, (k \neq t)}^o e(z_k)}{\sum_{j=1}^c e(z_j)} \\ p_{c \setminus o} = \frac{\sum_{i=1, (i \neq o)}^c e(z_i)}{\sum_{j=1}^c e(z_j)} \end{cases} \quad (5)$$

另外, 为实现类内/类外空间的精细化知识蒸馏, 对两个空间的概率进行细粒度分解, 定义独立的概率向量以表征空间内各子类别/程度的概率分布。将 $o \setminus t$ 及 $c \setminus o$ 下的各类概率分别定义为 $\hat{p}$ 和 $\tilde{p}$ 。其中 $\hat{p} = [\hat{p}_1, \hat{p}_2, \dots, \hat{p}_{t-1}, \hat{p}_{t+1}, \dots, \hat{p}_o]$, 表示在类空间 o 内, 排除 t 的其他独立同类概率, 即同类下不同缺陷程度的概率分布特征, 提升模型对细粒度缺陷程度的区分能力。同理, $\tilde{p} = [\tilde{p}_{o+1}, \tilde{p}_{o+2}, \dots, \tilde{p}_c]$ 表示空间 c 内排除空间 o 后的其他独立概率, 即不

同缺陷类的概率分布特征,强化模型对不同类缺陷的边界区分能力。

$$\begin{cases} \hat{p}_k = \frac{e(z_i)}{\sum_{k=1, (k \neq i)}^o e(z_k)} \\ \hat{p}_i = \frac{e(z_i)}{\sum_{j=1, (j \neq o)}^c e(z_j)} \end{cases} \quad (6)$$

经典蒸馏学习中将 KL 散度作为损失函数,成为 T-S 知识传递的桥梁,即:

$$KD = KL(p^T \| p^S) = \sum_{i=1}^c p_i^T \log\left(\frac{p_i^T}{p_i^S}\right) = p_i^T \log\left(\frac{p_i^T}{p_i^S}\right) + \sum_{k=1, (k \neq i)}^o p_k^T \log\left(\frac{p_k^T}{p_k^S}\right) + \sum_{j=1, (j \neq o)}^c p_j^T \log\left(\frac{p_j^T}{p_j^S}\right) \quad (7)$$

该损失函数将目标类、类内非目标、类外非目标的概率信息统一融合,无法区分不同类知识贡献度:

1) 目标类的核心知识被非目标类的冗余信息稀释,导致模型对目标缺陷的识别精度下降;

2) 类内细粒度缺陷的程度特征被同质化求和,无法实现缺陷程度的精准分级;

3) 类外不同缺陷大类的边界特征未被针对性强化,导致模型对相似缺陷大类的区分能力不足。

综合以上分析,为解耦传统 KL 耦合性强问题,并结合式(5)~(7),得到:

$$\begin{cases} p_k = p_{o \setminus t} \odot \hat{p}_k \\ p_j = p_{c \setminus o} \odot \hat{p}_j \end{cases} \quad (8)$$

进一步分解可将经典 KD 损失重构为三分量和,解耦逻辑如图 3 所示。

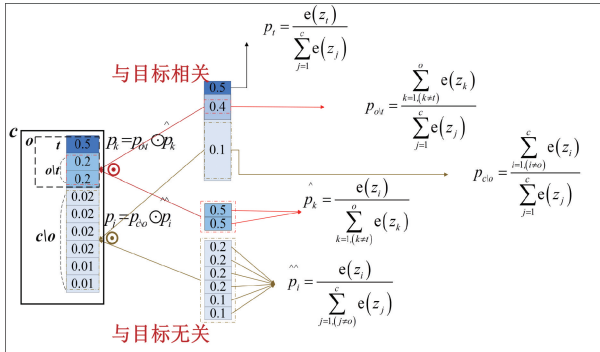


图3 FLSGD 解耦逻辑

Fig. 3 FLSGD decoupling logic

$$p_i^T \log\left(\frac{p_i^T}{p_i^S}\right) + \sum_{k=1, (k \neq i)}^o p_k^T \log\left(\frac{p_k^T}{p_k^S}\right) + \sum_{j=1, (j \neq o)}^c p_j^T \log\left(\frac{p_j^T}{p_j^S}\right) = p_i^T \log\left(\frac{p_i^T}{p_i^S}\right) + p_{o \setminus t}^T \log\left(\frac{p_{o \setminus t}^T}{p_{o \setminus t}^S}\right) + p_{c \setminus o}^T \log\left(\frac{p_{c \setminus o}^T}{p_{c \setminus o}^S}\right) +$$

$$p_{c \setminus o}^T \sum_{j=1, (j \neq o)}^c \hat{p}_j^T \left[\log\left(\frac{p_{c \setminus o}^T}{p_{c \setminus o}^S}\right) + \log\left(\frac{\hat{p}_j^T}{\hat{p}_j^S}\right) \right] = p_i^T \log\left(\frac{p_i^T}{p_i^S}\right) + p_{o \setminus t}^T \log\left(\frac{p_{o \setminus t}^T}{p_{o \setminus t}^S}\right) + p_{c \setminus o}^T \log\left(\frac{p_{c \setminus o}^T}{p_{c \setminus o}^S}\right) +$$

$$KL(X^T \| X^S) + (1 - p_i^T - p_{o \setminus t}^T) KL(Y^T \| Y^S) + (1 - p_i^T - p_{o \setminus t}^T) KL(Z^T \| Z^S) \quad (9)$$

式中:①表示 T-S 间在所定义三逻辑空间的知识蒸馏 (knowledge distillation in the three logic space, TKD); ②表示 T-S 在类内非目标 $o \setminus t$ 的知识蒸馏 (knowledge distillation for $o \setminus t$, OKD); ③表示 T-S 在其他类间 $c \setminus o$ 的知识蒸馏 (knowledge distillation for $c \setminus o$, CKD)。不同分量在信息传递时的权重直接影响老师的教学水平,为了能够灵活可控的将老师信息传递给学生网络,可以将上式简化为:

$$KD' = \alpha \times TKD + \beta \times OKD + \gamma \times CKD \quad (10)$$

TKD 为模型提供核心的目标缺陷识别能力,OKD 和 CKD 在 TKD 的基础上,分别从细粒度程度和跨大类边界两个维度对模型能力进行补充;通过调整 α 、 β 、 γ 的权重,可根据不同的缺陷检测任务需求灵活调控各分支的贡献度,实现方法的任务适配性。

3 实验结果与分析

为验证所提方法,现利用所采数据进行各部分实施细节阐述。研究均在 Window11 专业版操作系统下进行,编程语言为 Python 3.7,框架为 PyTorch-GPU。工作站处理器为 11th Gen Intel(R) Core(TM) i9-11900K @ 3.50 GHz,显卡为 NVIDIA GeForce RTX A5000,显存 24 GB。GPU 通过 CUDA 的加速运算可提升深度学习模型运算速率。

3.1 FLSGD 在 PCB 基板数据下的有效性分析

1) PCB 基板倒装芯片单源超声激励信号数据采集

单源超声振动检测系统如图 4 所示,其中芯片载体通过胶带固定在隔振台上。信号发生器产生一个范围为 40~110 kHz 的周期性扫描电压信号,并通过功率放大器放大,增强信号能量,充分触发倒装芯片。经过放大的信号通过空气耦合电容式超声波换能器转换为超声波,以入射角 45°发射到倒装芯片表面的中心,使其充分激励,产生振动。超声波在倒装芯片表面产生辐射力,激发芯片振动。利用激光扫描振计 (PSV-400, Polytec) 测量芯片振动,并以 1.28 MHz 的频率进行采样。固定在振动隔离台上的手动 X-Y 运动台用于调整电容式空气耦合超声

换能器点的位置,集成在激光扫描头上的 CCD 摄像机可以捕捉基板上的标记,以便精确对准。数据采集和处理由工业控制计算机控制,用于倒装芯片检测程序。

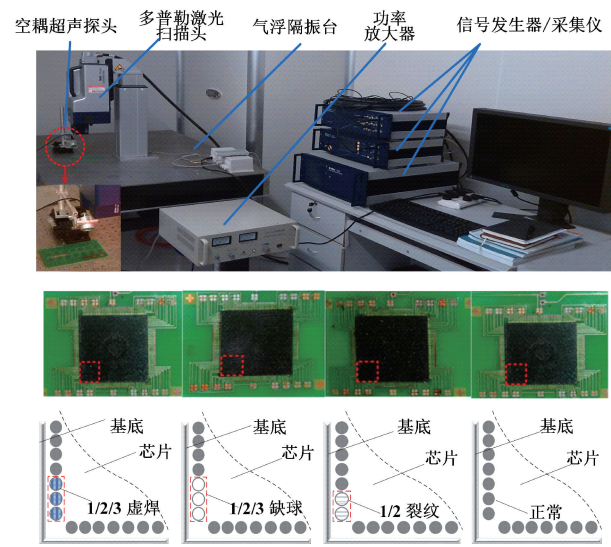


图 4 单源超声振动检测系统

Fig. 4 Single source ultrasonic vibration detection system

实验数据所收集的振动信号载体为来自真实组件的 PCB 基板菊花链式倒装芯片 PAC2.1。每个倒装焊芯片尺寸为 10 mm×10 mm,共 184 个微凸点分布在芯片周边。所有的微凸点都由 SnAg3Cu0.5 无铅材料组成,微凸点直径为 120 μm,微凸点间距为 200 μm。实验共获得 4 种缺陷倒装芯片,包括良好的芯片、微凸点缺失、微凸点虚焊及微凸点裂纹。另外的,具体细分的缺陷类型有 9 小种,图 4 显示了芯片缺陷的细节,微凸点缺失包含缺失 1 个焊球、2 个焊球、3 个焊球;微凸点虚焊包含虚焊 1 个焊球、2 个焊球、3 个焊球;同样微凸点裂纹包含裂纹 1 个焊球、2 个焊球。

2) FLSGD3 组分定性/定量分析

(1) 定性分析

利用所提方法对 PCB 基板数据进行三空间逻辑解耦,得到三分量即:TKD 分量、OKD 分量与 CKD 分量。方便学生模型从老师模型习得有效知识。为准确辨析所提方法中各分量的影响,本研究在同一学生网络上通过调整各分量系数进行知识传递分析。使用的学生网络仅包含 1 个卷积层与 2 个全连接层。不同分量组合下知识传递结果如表 1 所示。

表 1 不同分量组合下的结果

Table 1 Results for different ingredient combinations

行号	分量组合			学生网络		
	α	β	γ	Densenet	Alexnet	Resnet18
				97.37% ($\pm 0.19\%$)	96.79% ($\pm 0.09\%$)	97.23% ($\pm 0.17\%$)
①	1			80.73% ($\pm 0.18\%$)	87.69% ($\pm 0.12\%$)	87.60% ($\pm 0.06\%$)
②		1		10.19% ($\pm 0.00\%$)	10.19% ($\pm 0.00\%$)	10.19% ($\pm 0.00\%$)
③			1	23.55% ($\pm 1.78\%$)	15.47% ($\pm 2.27\%$)	15.26% ($\pm 2.90\%$)
④	1	1		84.56% ($\pm 0.05\%$)	81.91% ($\pm 0.05\%$)	88.24% ($\pm 0.12\%$)
⑤	1		1	88.05% ($\pm 0.15\%$)	82.51% ($\pm 0.11\%$)	87.87% ($\pm 0.04\%$)
⑥	1	1	1	86.77% ($\pm 0.18\%$)	81.97% ($\pm 0.17\%$)	86.94% ($\pm 0.14\%$)

根据表 1 前 3 行:逻辑空间中知识传递有效的分量为 TKD 分量,剩余 2 个分量在知识传递过程中不包含核心信息;单 OKD 分量在准确率贡献上几近于 0;单 CKD 分量在准确率贡献上相当不稳定,导致误差波动范围较大。根据表中第④、⑤行:在没有 TKD 分量辅助下的 OKD 与 CKD 在知识蒸馏中毫无意义;在 TKD 的基础上,OKD 与 CKD 能够进行知识的辅助传递,且准确率较单一 TKD 分量高。因此,可以初步说明 OKD 与 CKD 的隐含知识传递能力需要结合 TKD 才能表现。最后进行三分量知识同时传递,结果准确率普遍下降。表明 OKD 与 CKD 之间存在潜在的合作竞争,分量系数关系需进一步研究确定。

综上所述,TKD 关注于与目标标签相关的知识,它涵盖了三概率空间知识(粗糙知识/显性知识)的粗略表示。粗信息可以简单地解释为网络先行层在类内空间和类外空间的影响下识别目标空间的难度。相反,OKD 和 CKD 分别关注非目标类之间的知识(精细知识/隐性知识),它们分别是类内知识和类外知识。本研究的核心是重组原始 KL 算法,通过调整 α 、 β 和 γ 来平衡 TKD、OKD 和 CKD 的重要性,从而实现各逻辑空间信息的更灵活控制。

(2) 分量定量分析

在逻辑空间知识传递解耦中引入了新的超参数 (α 、 β 和 γ),由上节内容分析可知该超参数在 T-S 间知识传递过程中影响巨大。以准确率最高为目标,直接利

用树形估计器 (tree-structured parzen estimator, TPE) 算法进行分量系数的有效寻优。为了探讨 α 、 β 和 γ 这 3 个系数在数量上的逻辑关系,通过多次实验发现,这些参数可以设置在一定范围内:将温度 T 固定为 2,使用 TPE 搜寻空间 $\{\alpha:[0, 2] \in \mathbb{R}, \beta:[0, 1] \in \mathbb{R}, \gamma:[0, 1] \in \mathbb{R}\}$ 。结果如图 5 所示,准确率范围为 40.00% ~ 98.46%。在 100 个超参数组合中,近七成准确率 > 80.00%。

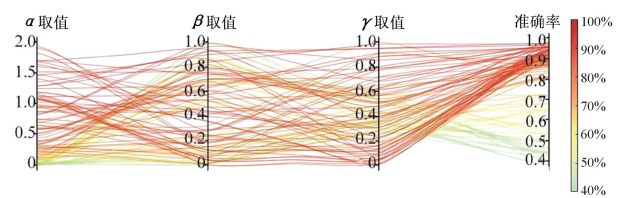


图 5 不同参数配置下的识别精度

Fig. 5 Recognition accuracy under different parameter configurations

本研究提出的 FLSGD 在 T-S 之间具有有效的信息传输能力,各分量系数排列都有一定的规律,与上述总结一致,结果由表 2 所示: α 处于领先地位,即 TKD 比 OKD 与 CKD 更重要,OKD 与 CKD 权值之间存在一定的模糊。

表 2 主客观权重分析

Table 2 Subjective and objective weight analysis			
权重分析方法	TKD (α)	OKD (β)	CKD (γ)
CRITIC	51.65%	24.84%	23.51%
变异系数法	35.65%	35.32%	29.03%
熵评价法	35.33%	34.78%	29.89%
AHP	42.48%	31.25%	26.27%
主客观权重	41.36%	31.53%	27.11%
指导值	1.00	0.76	0.66

下面,通过权重分析,对各分量系数权重进行量化。将精度 > 80% 的系数作为对象,采用主客观权重分析法对 FLSGD 中三分量系数进行量化。客观表达系数权重关系的方法主要包括内部标准重要性评价法 (criteria importance through inter-criteria correlation, CRITIC)、变异系数法、熵评价法,主观权重来自于分析层次结构过程 (analyze hierarchical process, AHP)^[27-29]。权重评价对象来源于上述 TPE 超参数搜索过程数据。由于 TPE 已经是一种相对成熟的方法,本研究没有过度描述。但不可否认的是,由于处理黑盒函数,在执行超参数优化时会存在一定的随机特征。因此,本研究通过均匀地调用每种权重评价方法来平衡基于 TPE 结果的模糊性。根据专家经验,将 0.3 和 0.7 比例的主/客观方法相结合得到最终系数权重,可以在很大程度上弥补单独确定权重时的

主客观权重方法的不足。

在上述权重评价方法的支持下,对每个 T-S 组合的系数进行了统计学分析,并定义了百分比,分别为 41.36%、32.53% 和 27.11%。可以初步得出结论,TKD 通过量化的权重包含了最有用的信息。OKD 的权重略高于 CKD,说明类内信息比类外信息能提供更多的知识,但类外信息是必不可少的。将 TKD 的系数设为 1,根据其数值计算 3 种组分的重量百分比,形成 FLSGD 固化指数,分别为 $\{\alpha=1.00, \beta=0.85, \gamma=0.66\}$ 。

3) FLSGD 效果对比分析

选取 3 个识别精度较高、模型规模较大的模型作为教师网络,3 个识别精度较低、模型规模较小的模型作为学生网络,所有实验都在同一个水平基准上进行。对比方法中,前两种方法属于特征知识蒸馏,后 4 种方法属于逻辑知识传递方法。与经典的 KD 相比,FLSGD 在几乎所有的 T-S 组合中都取得了持续的进步,FLSGD 实现了与特征蒸馏方法相当甚至更好的性能,取得了最优的改善 (加粗体部分),显著提高了蒸馏性能和训练效率之间的权衡,有力地支持了 FLSGD 在不同 T-S 间信息传递的鲁棒性。通过垂直分析,FLSGD1 得到的精度不是最高,但可以与现有的 KD 方法齐头,证明权重分析得到的系数关系符合 FLSGD 逻辑;同时 FLSGD2 在几乎所有的 KD 方法中都处于领先地位,结果如表 3 所示。

与现有的 KD 方法相比,FLSGD 对倒装芯片振动信号具有很高的竞争力,能有效传递教师模型信息,使学生模型获得比自身更突出的缺陷识别能力。通过水平比较,可以发现教师网络在知识传递过程中控制着信息源,并在很大程度上决定了蒸馏精度上限。值得注意的是,以 Shufflenet 为学生网络的蒸馏框架结果相对于其他两组 T-S 并不理想,初步判断为此网络不适用于该倒装芯片振动信号的特征提取学习。因此,设计一个有效的网络也是提高蒸馏性能的有效方法,实现知识转移过程的特征协调。

3.2 FLSGD 在硅基数据下的有效性分析

1) 硅基 dummy 片双超声激励信号数据采集

双超声激励振动检测平台如图 6 所示。检测平台选用优利德公司生产的 UTG-2025A 信号发生器,将周期变化的正弦波作为激励信号。信号发生器产生一个范围为 50~350 kHz 的周期性扫描电压信号,并通过 PINTTECH 公司生产的 HA-8202A 功率放大器放大,增强信号能量,充分触发倒装芯片。放大的信号通过使用 ULTRAN GROUP 公司设计的 NCG200-D25-P76 空气耦合压电式超声换能器转换为超声波,在芯片上方两侧以入射角 45° 对称发射到芯片表面中心,使其充分激励并产生振动。超声波在倒装芯片表面产生辐射力,激发芯片振动,扫频循环周期为 2 ms。舜宇公司 LV-SC400-Probe 激光扫描

表 3 不同 T-S 组合下的 KD 检测结果
Table 3 KD detection result under different T-S

蒸馏师生 T-S 设置				特征蒸馏	
T size/M	S size/M	老师网络	学生网络	AT	CRD
27. 1	5. 95	Densenet 97. 37% (±0. 19%)	MobilenetV2 95. 70% (±0. 36%)	96. 09% (±0. 1%)	96. 12% (±0. 24%)
217. 0	8. 76	Alexnet 96. 79% (±0. 09%)	MobilenetV3 94. 40% (±0. 42%)	94. 94% (±0. 30%)	94. 54% (±0. 06%)
42. 7	6. 03	Resnet18 97. 23% (±0. 17%)	Shufflenet 94. 72% (±0. 15%)	92. 65% (±0. 35%)	93. 62% (±0. 27%)
逻辑蒸馏					
KD	DKD	RKD	FLSGD1	FLSGD2	
95. 94% (±0. 11%)	96. 21% (±0. 01%)	95. 78% (±0. 37%)	95. 58% (±0. 30%)	96. 20% (±0. 05%)	
93. 40% (±0. 27%)	95. 28% (±0. 29%)	96. 35% (±0. 28%)	95. 93% (±0. 05%)	96. 47% (±0. 24%)	
92. 65% (±0. 37%)	93. 69% (±0. 57%)	93. 04% (±0. 29%)	93. 89% (±0. 14%)	94. 16% (±0. 04%)	

注:FLSGD1 表示超参数依据主客观权重法: { $\alpha=1.00, \beta=0.76, \gamma=0.66$ }。FLSGD2 表示超参数来自 TPE。

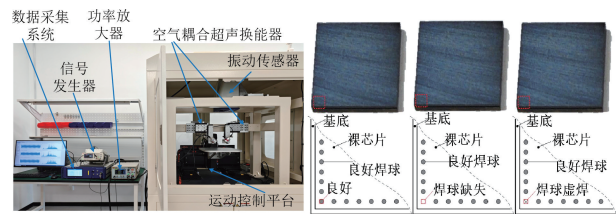


图 6 实验数据采集

Fig. 6 Experimental data collection

振动计测量芯片振动,并以 3.84 MHz 的频率进行采样,采样时间为 1 s。固定在振动隔离台上的手动 X-Y 运动台用于调整电容式空气耦合超声换能器点的位置,集成在激光扫描头上的 CCD 摄像机可以捕捉基板上的标记,以便精确对准。数据采集和处理由工业控制计算机控制,用于倒装芯片检测程序。

实验数据所收集的振动信号载体来自中国电科集团第五十八研究所产线设计制成的硅基板 dummy 芯片,尺寸为 7.5 mm×7.5 mm。每个 dummy 片有 52 个微凸点分布在周边。所有的微凸点都由 SnAg3Cu0.5 无铅材料组成,微凸点直径为 200 μm ,间距为 300 μm 。实验共获得 3 种倒装芯片,包括良好、微凸点缺失及微凸点虚焊。另外,微凸点缺失包含缺失 1 个焊球、5 个焊球;微凸点虚焊包含虚焊 1 个焊球、5 个焊球。因此,具体细分的缺陷类型可以分为 5 小种,图 6 显示了用于无损诊断的芯片细节。

2) FLSGD 效果对比分析

3.1 节中对 FLSGD 空间分量权重的指导来自于 PCB 基板倒装芯片 PAC2.1 振动信号,不一定适用于硅基 dummy 芯片缺陷检测应用,因此这里重新按照 0.3/0.7 比例的主客观方法进行分量权重换算,结果如图 7 所示。

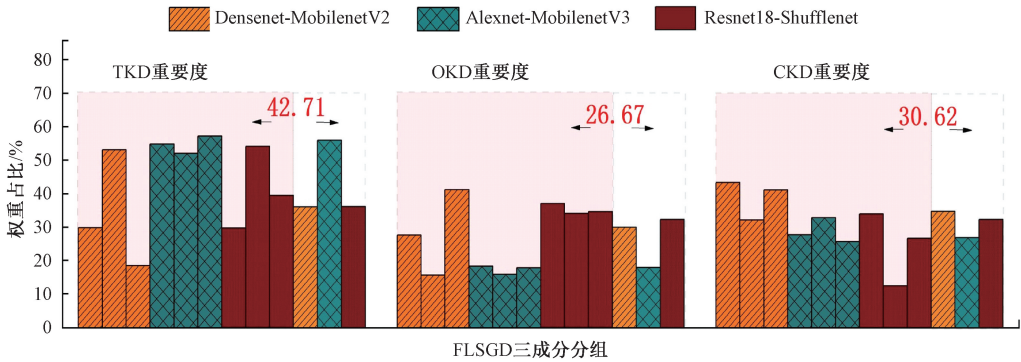


图 7 FLSGD 系数评估结果

Fig. 7 FLSGD coefficient evaluation results

各 T-S 组合下采用不同权重分析方法计算的系数分布状态如图 7 所示,不同 T-S 组合在各类权重方法下的系数分布存在差异,结果如表 4 所示。

对每个 T-S 组合的 3 个系数开展统计分析并确定百分比,精准得到 42.714 2%、26.664 8%和 30.621 0%,所

得初步结论与 PCB 基板倒装芯片 PAC2.1 结果趋势相同:TKD 通过量化权重涵盖最具价值信息,OKD 与 CKD 差值较小,意味着类内信息与类外信息均能有效提供知识。将 TKD 系数设为 1,并计算 3 组分量权重,分别得到 { $\alpha=1.000\ 0, \beta=0.624\ 3, \gamma=0.716\ 9$ }。这一结果也表

表 4 权重分析结果

Table 4 Weight analysis result

主/客观 评价法	TKD 分量			OKD 分量			CKD 分量		
	A	B	C	A	B	C	A	B	C
CRITIC	29.56	54.32	29.49	27.42	18.18	36.81	43.02	27.49	33.70
变异系数法	52.63	51.62	53.74	15.52	15.79	33.82	31.85	32.60	12.44
熵评价法	18.38	56.78	39.12	40.82	17.72	34.40	40.80	25.50	26.47
AHP	35.82	55.49	35.89	29.71	17.84	32.05	34.46	26.66	32.06

注:T-S 组合 A;Densenet-MobilenetV2;T-S 组合 B;Alexnet-MobilenetV3;T-S 组合 C;Resnet18-Shufflenet。

明,不同数据集环境中,尽管各分量对结果影响的作用方向是明确的,但适配的具体权重配比需结合更多实验、从更多维度深入探索。

以 AlexNet、DenseNet、ResNet18 为教师模型, MobilenetV2、MobilenetV3、ShuffleNet 为轻量级学生模型,交叉构建 9 组教师-学生网络蒸馏组合,如表 5 展示了 3 组老师模型与 3 组学生模型的各轻量化指标,对所提 FLSGD 算法的性能与泛化能力开展系统性对比验证。

所有实验均在同一水平基准上进行测试,计算结果如图 8 所示。

表 5 不同 T-S 组合下轻量化指标

Table 5 Lightweight indicators under different T-S combinations

模型选择/量化指标	准确率/%	模型大小/M	参数量/M	
老师模型	Alexnet	97.34	217.00	61.10
	Densenet	97.94	27.10	7.98
	Resnet18	97.11	42.70	11.69
学生模型	MobilenetV2	93.42	5.95	3.50
	MobilenetV3	93.09	8.76	5.48
	Shufflenet	95.83	6.03	7.39

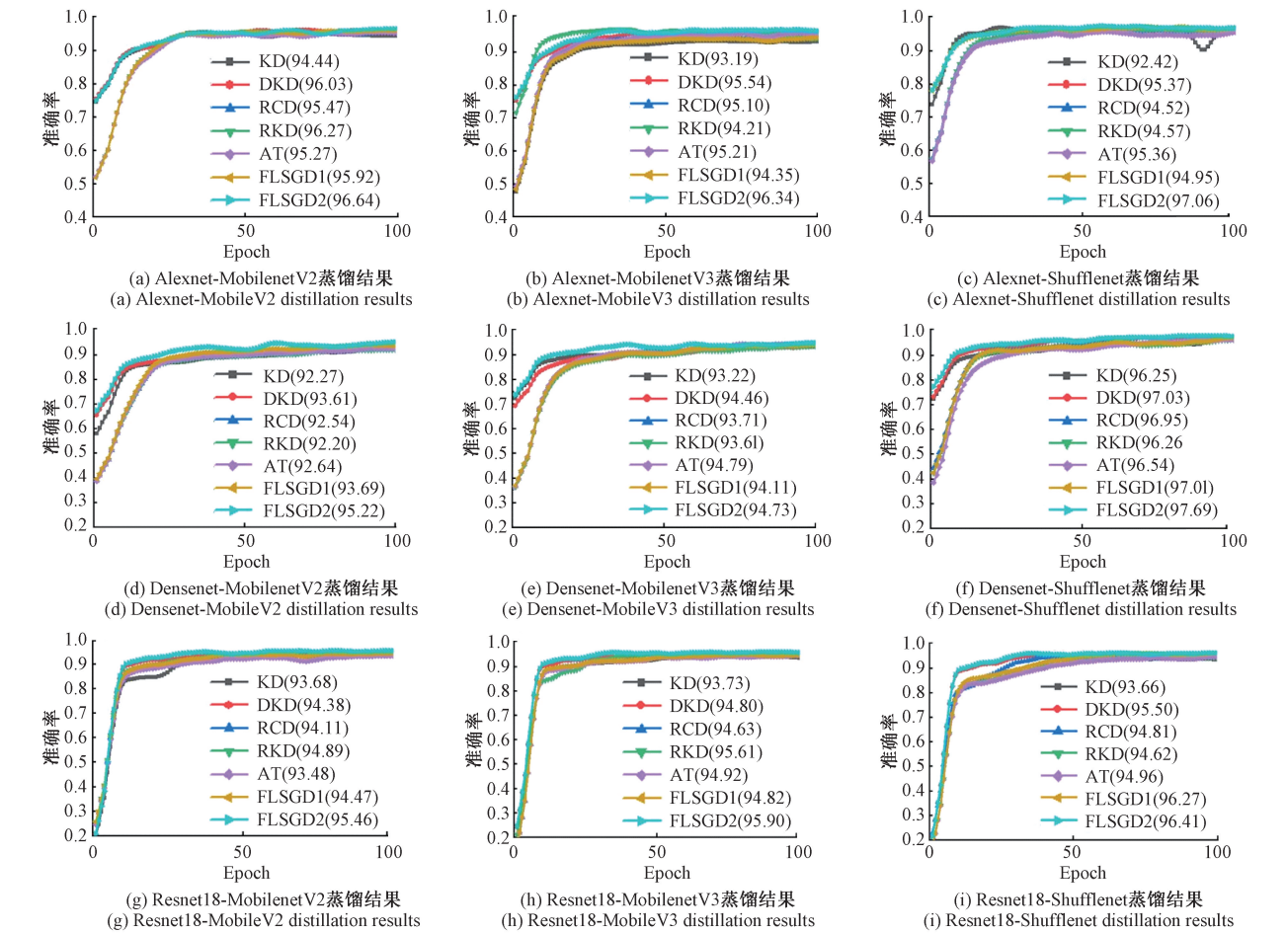


图 8 不同 T-S 组合下蒸馏结果

Fig. 8 Distillation results under different T-S combinations

同样的,FLSGD1 表示该系数来自于主观/客观权重分析得到的指导值,FLSGD2 表示使用 TPE 获得的具体寻优值。本研究在不同 T-S 组合下进行了 9 次交叉实验。

(1) 检测精度分析

FLSGD1 分类精度虽非最高,却能紧跟其他现有 KD 方法节奏;FLSGD2 在几乎所有 KD 方法里均处领先。在图 8 中,使用模型大小 Model size 和参数量 Params 两个指标来反映 T-S 之间的容量差异,并验证 FLSGD 对振动信号数据集相比其他 KD 方法具有更高的精度及收敛速度,助力学生模型获得比自身更突出的缺陷识别能力。经过水平方向分析,可发现知识传递过程中信息来源由教师网络控制,在很大程度上决定了蒸馏精度的上限,故设计有效教师网络是提升蒸馏性能的有效路径。通过纵向观察,可发现学生网络很大程度决定准确率波形的收敛状态。此外,虽然 Shufflenet 在 PCB 基板倒装芯片 PAC2.1 的缺陷检测任务上表现欠佳,但是在硅基 dummy 片数据上却能保持与其他网络均衡的识别精度,说明数据特性相匹配的网络结构能起到一定的信息传递促进作用。

(2) 收敛性分析

定义收敛商 k_M 评估 FLSGD 检测精度及收敛速度,即:

$$index = \begin{cases} sort(k_M) \\ k_M = \frac{M_{accuracy}}{M_{convergenceepoch}} \end{cases} \quad (11)$$

式中: $M_{accuracy}$ 表示不同蒸馏算法在各 T-S 组合下识别精度均值; $M_{convergenceepoch}$ 表示不同蒸馏算法在 T-S 组合下的收敛点均值; k_M 综合了各 T-S 组合下知识蒸馏算法的准确性和收敛性并进行定量排名。

本研究提出的算法 FLSGD 是基于逻辑信息解耦的标签蒸馏,旨在改变知识的信息表示维度。中间特征蒸馏是通过添加模块来捕获网络中的不同特征。大多数特征蒸馏方法可以比标签蒸馏方法获得更好的性能,但这是以高昂的计算和存储成本为代价的。因此,本研究的初衷是寻找一种逻辑蒸馏方式,在不增加计算能力要求的情况下,可以获得媲美甚至超越特征蒸馏方法的效果。如图 9 所示,FLSGD 将低维语义信息转换到高维空间,在不增加计算模块的情况下获得更有意义的逻辑信息,从而实现了网络分类任务的高效执行。具体的,如表 6 所示,FLSGD2 准确率高达 96.16%,为各方法之首,且收敛步仅 63,收敛速度较快,精度与收敛效率双优,知识蒸馏效果突出;FLSGD1 准确率 95.07%,虽稍逊于 FLSGD2,但收敛点 69 能跟上主流 KD 方法节奏,可适配不同蒸馏需求;传统 KD 准确率为 93.65%最低,收敛点为 73 速度较慢,检测精度与收敛速度均弱,蒸馏性能欠佳;DKD 准

确率达到 95.19%,虽未达最优,但收敛商表现优异排名第 2;CRD 准确率 94.65%、收敛点 69,精度和收敛速度均不如 FLSGD 系列,综合表现一般;RKD 准确率 94.58%,收敛点 68,精度、收敛速度均有提升空间;AT 准确率 94.80%、收敛点 71,在精度与收敛效率上,整体表现不突出,不具有竞争性。综上,通过与其他方法的对比分析,证实了 FLSGD 能够将低维语义转换到高维空间,在不增加计算模块的情况下获得更有意义的逻辑信息,从而实现网络分类任务的高效执行。

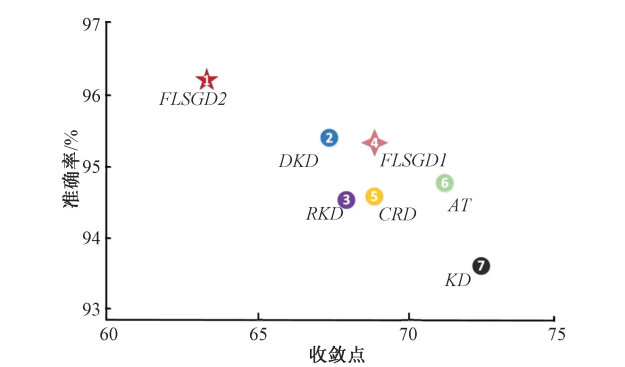


图 9 收敛商 k_M 结果
Fig. 9 Convergence k_M result

表 6 不同 T-S 组合下的收敛商结果			
Table 6 Convergence quotient under different T-S			
指标	$M_{acc}/\%$	$M_{epoch}/\text{轮}$	k
KD	93.65	73	1.28
DKD	95.19	67	1.42
CRD	94.65	69	1.37
RKD	94.58	68	1.39
AT	94.80	71	1.34
FLSGD1	95.07	69	1.38
FLSGD2	96.16	63	1.53

3.3 讨 论

根据上述实验过程分析,可以得出结论:蒸馏效果主要受教师网络、教学方法和学生网络的影响。教师网络提供了原始信息的关键知识。教学方法,即 KD 方法,通过教师网络将表征信息与抽象信息传递到学生网络。学生网络根据自己的学习接受抽象知识,最终表示为缺陷识别的准确性。运用 KD 方法实现教师模型教学、学生模型学习的前提是教师网络、教学方法、学生网络具有协调性。不同数据的独特性分布可能会影响教学质量,例如本研究应用中针对 PCB 基板数据在使用 Shufflenet 作为学生网络时的教学情况,而来自 CWRU 的公共轴承振动信号也会出现这种现象,这里不做细述。

从蒸馏效果来看,传统的蒸馏策略是对所有蒸馏点或整个蒸馏过程施加蒸馏约束,容易导致学生网络过度正则化,学习成绩下降。这一不足也反映在现有的几种KD方法中。与大多数方法相比,FLSGD方法更有效地改善了学生性能:在不同的概率空间中采用不同的蒸馏侧重点,即通过调整3个系数使教学方法具有一定的适应性和容错性,有重点的进行信息和知识的传递。在FLSGD的辅助下,提取教师网络的高层语义逻辑空间知识(以 K 表示)用于学生网络学习,学生网络学习数据特征以获得其认知(以 k 表示),实现超越教师网络知识 K 的特征融合 $K+k>K$ 。从理论上讲,FLSGD的设计是为了包含更多的语义逻辑信息,使小型网络在降低算力的前提下,有更高的概率向 $K+k$ 的目标发展。

从计算复杂度来看,FLSGD的核心是对传统知识蒸馏的KL散度逻辑进行解耦重构,而非引入新的网络结构或复杂操作,与中间特征蒸馏方法(如AT、CRD)相比,它属于逻辑蒸馏(标签蒸馏)范式,无需对教师网络的中间层特征图进行提取,减少了中间特征计算成本。中间特征蒸馏方法在训练成本上过高,在训练期间引入了用于提取深层特征的計算和存储使用,如网络模块和复杂操作。随着教师模型的网络结构越来越大,需要提取的中间特征知识越多,算法模块调用次数越多,基于中间特征提取的方法计算复杂度将急剧增加。相反,基于逻辑的蒸馏方式没有这样的担忧,本研究所提FLSGD的实现是基于教师-学生网络的原始逻辑输出(Softmax前逻辑张量),由原KL散度重构而来,通过掩码、概率分布計算及KL散度损失求解等基础操作完成空间知识解耦传递,未增加任何额外的卷积层、注意力模块或特征提取分支,避免了因模块叠加导致的算力消耗激增,且其性能可以与中间特征精馏相媲美甚至超越。因此,通过综合评估算法的計算效率和发展前景,FLSGD具有相当大的竞争优势。

4 结 论

为满足工业领域缺陷检测的准确性和速度要求,更为精确地激活感兴趣的检测对象,亟需探索一个能够有效提取信号特性的轻量级检测网络是目前倒装芯片缺陷检测领域的重要课题之一。本研究所提FLSGD为解释知识蒸馏的信息传递路径提供了一个新的视角。该方法打破了知识蒸馏中T-S之间的知识转移逻辑,将传统知识蒸馏的逻辑空间解耦为三概率空间。从目标空间 t 、类内空间 $o|t$ 和类外空间 $c|o$ 这3个角度对KD中的逻辑概率空间进行了新的解释,揭示了现有方法T-S之间信

息转移中的局限性与不灵活性,实现T-S之间响应层的语义分布信息对齐,学习具有类边界可辨别的逻辑特征表示,促进特征迁移。在两组超声激励实验倒装芯片振动信号的支持下通过大量的实验分析对FLSGD进行了验证,表明该方法能够在倒装芯片缺陷检测中取得比其他KD方法更先进的性能,且更具鲁棒性及灵活性。当然,需要正视的是,在本研究中3个成分之间的定量关系是通过主客观赋权法获得,虽给出了参考值,仍有待进一步探索。

参考文献

- [1] 李可,王翀,明雪飞,等. 基于改进多路径匹配追踪的芯片超声信号去噪方法[J]. 仪器仪表学报, 2023, 44(1): 93-100.
LI K, WANG CH, MING X F, et al. Denoising method of chip ultrasonic detection signals based on the improved multipath matching pursuit[J]. Chinese Journal of Scientific Instrument, 2023, 44(1): 93-100.
- [2] 王辰辰,杨梦冉,姚贞建. 噪声频谱混叠干扰下超声检测信号高质量提取方法[J]. 仪器仪表学报, 2024, 45(10): 244-252.
WANG CH CH, YANG M R, YAO ZH J. High quality extraction method of ultrasonic detection signals under noise spectrum aliasing interference[J]. Chinese Journal of Scientific Instrument, 2024, 45(10): 244-252.
- [3] WANG X Y, WANG ZH, XU Z K, et al. High-resolution X-ray imaging technologies for non-destructive testing of integrated circuits[J]. Nano Materials Science, 2025.
- [4] LI J Q, XIA X L, ZHENG ZH W, et al. Inverse analysis on non-uniform temperature and emissivity fields based on multispectral infrared thermal image data[J]. Applied Thermal Engineering, 2025, 267: 125592.
- [5] QU ZH, YAO X W, YANG X G, et al. Relaxed knowledge distillation [J]. International Journal of Computer Vision, 2025, 134(101): 1-18.
- [6] ZHANG X, YAN R Y, FAN Y SH, et al. Cross-domain attention transfer network for recommendation [J]. Advanced Engineering Informatics, 2025, 68: 103667.
- [7] KIM J H, PARK S, KWAK N. Paraphrasing complex network: Network compression via factor transfer [J]. ArXiv preprint arXiv: 1802.04977, 2018.
- [8] 吴振宇,屈先锋,王慧,等. 基于电磁场特征蒸馏的无刷直流电机匝间短路智能辨识[J]. 仪器仪表学

- 报, 2025, 46(10): 107-119.
- WU ZH Y, QU X F, WANG H, et al. Intelligent identification of inter-turn short circuits in brushless DC motor based on feature distillation of multi-source electromagnetic field signals [J]. Chinese Journal of Scientific Instrument, 2025, 46(10): 107-119.
- [9] 杨帆, 张海东, 霍军周, 等. 结合边缘信息提取增强的裂纹检测模型压缩方法[J]. 光学学报, 2026, 46(4): 159-173.
- YANG F, ZHANG H D, HUO J ZH, et al. A crack detection model compression method combined with edge information extraction enhancement [J]. Acta Optica Sinica, 2026, 46(4): 159-173.
- [10] 孟祥福, 张智超, 俞纯林, 等. 结合边界自知识蒸馏的结肠镜息肉图像分割方法[J]. 中国图象图形学报, 2025, 30(2): 589-600.
- MENG X F, ZHANG ZH CH, YU CH L, et al. Colonoscopy polyp image segmentation method with boundary self-knowledge distillation[J]. Journal of Image and Graphics, 2025, 30(2): 589-600.
- [11] WU Y L, TANG B P, DENG L, et al. Distillation-enhanced fast neural architecture search method for edge-side fault diagnosis of wind turbine gearboxes[J]. Expert Systems with Applications, 2022, 208: 118049.
- [12] DENG J, JIANG W J, ZHANG Y, et al. HS-KDNet: A lightweight network based on hierarchical-split block and knowledge distillation for fault diagnosis with extremely imbalanced data [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3091498.
- [13] ZHANG W F, BISWAS G, ZHAO Q, et al. Knowledge distilling based model compression and feature learning in fault diagnosis[J]. Applied Soft Computing, 2020, 88: 105958.
- [14] WANG X CH, CHEN AI G, ZHANG L, et al. Distilling the knowledge of multiscale densely connected deep networks in mechanical intelligent diagnosis [J]. Wireless Communications & Mobile Computing, 2021: 4319074.
- [15] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [16] WANG J, GENG X. Theoretical analysis of label distribution learning [C]. 33rd AAAI Conference on Artificial Intelligence, 2019: 5256-5263.
- [17] ZHANG CH B, JIANG P T, HOU Q B, et al. Delving deep into label smoothing [J]. IEEE Transactions on Image Processing, 2021, 30: 5984-5996.
- [18] WU M, LI Q M, YANG F, et al. Learning from biased crowdsourced labeling with deep clustering [J]. Expert Systems with Applications, 2024, 211: 118608.
- [19] SHUAI J W, XU C, JIANG H Y, et al. Meta-label-corrected knowledge distillation for partial multi-label learning[J]. Electronics, 2026, 15(6): 1233.
- [20] 童小钟, 赵宗庆, 苏绍璟, 等. 基于知识蒸馏自适应 DenseNet 的无人机对地目标可见光与红外图像融合[J]. 仪器仪表学报, 2024, 45(5): 20-32.
- TONG X ZH, ZHAO Z Q, SU SH J, et al. Fusion of visible and infrared images of ground targets by unmanned aerial vehicles based on knowledge distillation adaptive DenseNet[J]. Chinese Journal of Scientific Instrument, 2024, 45(5): 20-32.
- [21] HEO B, LEE M, YUN S, et al. Knowledge transfer via distillation of activation boundaries formed by hidden neurons [C]. 33rd AAAI Conference on Artificial Intelligence, 2019: 3779-3787.
- [22] LIN Y, JIANG J CH, CHEN D X, et al. Ded: Diagnostic evidence distillation for acne severity grading on face images[J]. Expert Systems with Applications, 2023, 228: 120312.
- [23] CHEN Y D, WANG S, LIU J J, et al. Improved feature distillation via projector ensemble [J]. ArXiv preprint arXiv: 2210.15274, 2022.
- [24] WEI SH C, LUO CH B, LUO Y. Scale decoupled distillation [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 15975-15983.
- [25] SUN SH Q, REN W Q, LI J ZH, et al. Logit standardization in knowledge distillation [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 15731-15740.
- [26] LI ZH, LI X. YANG L F, et al. Curriculum temperature for knowledge distillation[C]. 37th AAAI Conference on Artificial Intelligence, 2023: 1504-1512.
- [27] LIU Y, ECKERT C M, EARL C. A review of fuzzy AHP methods for decision-making with subjective judge-

ments[J]. Expert Systems with Applications, 2020, 161: 113738.

[28] LI X Y, LI CH CH, GUO F CH, et al. Coefficient of variation method combined with XGboost ensemble model for wheat growth monitoring [J]. Frontiers in Plant Science, 2024, 14: 1267108.

[29] CAO ZH W, ZHANG T L, TONG X K. Quality evaluation of chicken soup based on entropy weight method and grey correlation degree method[J]. Scientific Reports, 2024, 14(1): 13038.

作者简介



E-mail:15961184597@ 163. com

孙钰,2017 年于南京工程学院获得学士学位,2020 年于浙江海洋大学获得硕士学位,2025 年于江南大学获得博士学位,现为无锡职业技术大学讲师,主要研究方向包括信号处理与缺陷检测。

Sun Yu received his B. Sc. degree from Nanjing Institute of Technology in 2017, his M. Sc. degree from the Zhejiang Ocean University in 2020, and his Ph. D. degree from Jiangnan University in 2025. He is currently a lecturer at Wuxi University of Technology. His main research interests include defect detection and signal processing.



李可(通信作者),1999 年于长春工业大学获得学士学位,2012 年于日本三重大学获得博士学位,现为江南大学智能制造学院教授,主要研究方向包括计算视觉、信号处理、故障诊断及结构健康监测。
E-mail:like@ jiangnan. edu. cn

Li Ke (Corresponding author) received his B. Sc. degree from Changchun University of Technology in 1999, and his Ph. D. degree from Mie University, Japan, in 2012. He is currently a professor at the College of Mechanical Engineering at Jiangnan University. His main research interests include computer vision, signal processing, fault diagnosis, and structural health monitoring.