doi:10.12068/j.issn.1005－3026.2026.20259030

面向重复性缺陷检测的钢材表面缺陷聚类方法

马博渊[1,2,3]，段智峰[1,4]，周 鹏[1,2,3]

（1.北京科技大学 北京材料基因工程高精尖创新中心，北京 100083；2.北京科技大学 人工智能学院，北京 100083；3.北京科技大学 顺德创新学院，广东 佛山 528399；4.北京科技大学 计算机与通信工程学院，北京 100083）

摘 要：在钢材制造过程中，设备磨损或工艺波动易引发表面重复性缺陷，亟须依托智能检测与聚类方法实现高效识别与故障溯源.然而，实际生产中存在同类缺陷形态差异显著，同时受漏检、误检、空间漂移及定位误差等影响，缺陷并不总是呈现严格或稳定的周期分布.为此，提出了一种基于层级分布对齐正则化的两阶段广义类别发现方法，以适应生产中出现的新形貌缺陷，该方法利用已知聚类簇监督模型作为知识锚点，通过隐式平衡批次中已知类与未知类群组的边缘概率，并显式促进各自群组内部预测的均匀性，实现更结构化的预测分布控制，有效提升了缺陷聚类精度.结果表明，所提方法在钢材表面重复性混合缺陷数据集的所有类别的聚类精度较SimGCD*方法提高了4.54%，在未知类别的聚类精度提高了5.87%，展现出在类别间有效分离方面的显著优势.

关 键 词：钢材表面缺陷聚类；重复性缺陷检测；深度学习；广义类别发现；层级分布对齐正则化

中图分类号：TP 391.4 **文献标志码：**A **文章编号：**1005-3026(2026)06-0143-12

Steel Surface Defect Clustering Method for Repetitive Defect Detection

MA Bo-yuan[1,2,3], *DUAN Zhi-feng*[1,4], *ZHOU Peng*[1,2,3]

(1. Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, Beijing 100083, China; 2. School of Artificial Intelligence , University of Science and Technology Beijing, Beijing 100083, China; 3. Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China; 4. School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China. Corresponding author: MA Bo-yuan, E-mail: mbytony@ustb.edu.cn)

Abstract: In the steel manufacturing process, surface repetitive defects are easily caused by equipment wear or process fluctuations, and intelligent detection and clustering methods are urgently required to achieve efficient recognition and fault source tracing. However, in practical production, significant morphological differences exist in the same type of defects. Meanwhile, affected by missed detection, false detection, spatial drift, and localization error, defects do not always present a strict or stable periodic distribution. Therefore, a two-stage generalized category discovery method based on hierarchical distribution alignment regularization was proposed to adapt to the new defect morphology appearing in production. In this method, the known cluster supervised model was utilized as a knowledge anchor. By implicitly balancing the marginal probabilities of known and unknown class groups in a batch and explicitly promoting the prediction uniformity within respective groups, a more structured prediction distribution control was achieved, and the defect clustering accuracy was effectively improved. The results show that compared with that of the SimGCD* method, the clustering accuracy of the proposed method on all categories of the steel surface repetitive mixed defect dataset is improved by 4.54%, and that on unknown categories is improved by 5.87%, which demonstrates a significant advantage in

收稿日期：2025-07-22
基金项目：国家科技重大专项(2024ZD0608200).
作者简介：马博渊(1992—)，男，山西太原人，北京科技大学副教授.
通信作者：马博渊，E-mail:mbytony@ustb.edu.cn.

effective separation between categories.

Key words: steel surface defect clustering; repetitive defect detection; deep learning; generalized category discovery; hierarchical distribution alignment regularization

钢材生产过程中,由于轧辊磨损、异物压入轧辊等因素,常会产生具有重复性规律的缺陷(如重复性划痕、凹坑、辊印或异常纹理)[1],这类缺陷在外观形貌上高度相似,并可能在生产过程中不断重复出现,一旦未被及时发现和聚类识别,可能引发整个批次的生产性质量问题,导致下游加工设备损耗甚至终端产品安全隐患[2].因此,在钢材表面缺陷检测的基础上,亟须构建基于图像表征的聚类方法,从外观相似性角度识别重复性缺陷簇,并据此辅助定位潜在设备异常源(如特定轧辊的磨损或杂质干扰),减少非计划停产带来的经济损失.

在真实生产环境下,缺陷的空间分布与理论情况下的周期分布不同,其不再呈现规则的空间分布.漏检、误检、速度扰动、板材尺寸变化、轧辊偏移、图像拼接误差等因素会导致缺陷的空间间隔呈现高度不确定性,使得难以依赖缺陷位置信息或周期距离推断缺陷性质,削弱了基于几何周期的周期性建模的可行性.因此,相比"周期性缺陷"这一理想化定义,重复性缺陷是更合理的表述,其核心特征在于外观形貌近似但不强求固定空间节距.

随着人工智能技术在材料科学领域的发展[3],通过高精度的机器视觉智能检测系统,已可实现钢材表面缺陷的实时智能检测[4-5].图1展示了钢材表面重复性缺陷的出现场景:随着时间推移,异物磨损、环境变化等因素会导致钢材表面形成凹坑.这些凹坑在视觉纹理上具有一定一致性,但形态存在开放性变化,具体表现为旧类凹坑与新类凹坑的差异,这一特点使得检测过程中容易出现漏检和误检现象.对于形貌突变或弱纹理缺陷,仅依赖时空信息难以有效划分簇结构,需从图像相似度角度出发进行重复缺陷类别聚类和新类识别,进而计算聚类簇的空间重复性质(如重复间隔的均值、方差与重复数目等).

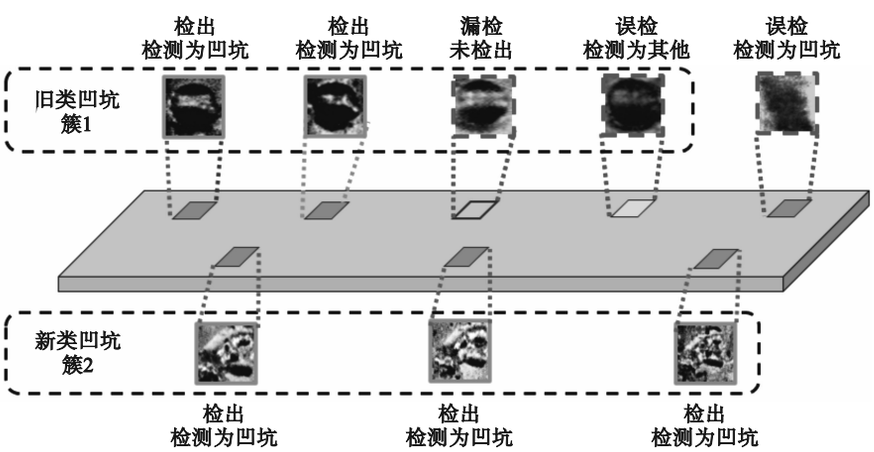


图1 钢材表面重复性缺陷簇示意图

Fig. 1 Schematic diagram of steel surface repetitive defect clusters

钢材表面重复性缺陷簇的发现与机器学习中聚类任务中的类别发现的设置相匹配.例如,在实际生产中出现与闭集中不同形态的凹坑簇,需要聚类识别成新的缺陷簇,在类别发现任务中可以视为潜在的新类别并对此进行聚类.本文重点探究广义类别发现(generalized category discovery, GCD),这是一个开放世界的任务,旨在用一组预先标注的已见类别的指导下,在未标注的数据集中发现新的类别[6].与传统的半监督学习(semi-supervised learning, SSL)和新类别发现(novel category discovery, NCD)任务相比,GCD在学习设定上更具通用性和挑战性.具体而言,SSL通常假设标注数据与未标注数据共享相同的类别空间[7-8],而NCD则假设标注数据的已知类别空间与未标注数据的类别空间完全不重叠[9-10].相比之下,GCD打破了这一限制,允许未标注数据同时包含已知类和未知类,要求模型在保持已知类识别能力的同时,从中自动发现未知类.与多数仅依赖单阶段无监督优化的GCD方法不同,持续广义类别发现方法[11]与知识蒸馏方

法[12-13]利用完全监督训练所获得的已知类模型，在精确编码已知类别的判别性知识的同时，还能在未知类别发现阶段将显著偏离已知类分布结构的预测特征作为潜在的未知类识别信号．此外，为了在新类别探索中偏向于某些特定类别，通常需要引入正则化项来鼓励预测的均匀性．其中熵正则化[14]通常用于避免聚类模型过早收敛到局部最优解．在大多数GCD工作[15-17]中，训练范式涉及使用一种无监督均值熵最大化方法来提升批次整体分布的不确定性，从而增强对未知类的发现．这一策略在无标签条件下构建了“弱监督”信号，但其对所有类别采用同质化处理，往往无法精细地调控所有类别群体间的学习动态以及各群体内部的类别区分度．

针对上述缺陷聚类的难题，本文提出了一种基于层级分布对齐正则化（hierarchical distribution alignment regularization，HDAR）的两阶段广义类别发现方法，以图像特征为基础对钢材表面重复性缺陷进行聚类．该方法不依赖缺陷的时空位置信息，而是利用已知聚类簇监督模型作为知识锚点，通过隐式地平衡批次中已知类与未知类群组的边缘概率、显式地促进各自群组内部预测的均匀性，辅助对已知簇和未知簇的发现，有效维持了全局类别平衡，在缓解旧类遗忘的同时强化了未知类辨识度，进而显著提升了缺陷聚类精度．本文的代码与示例数据集已公开至https://github.com/zhifeng-d/HDAR.

1　基于HDAR的两阶段广义类别发现方法

1.1　问题表述

在钢材缺陷聚类任务中，使用已标注数据集$D^{\mathrm{l}}=\left\{\left(x_i,y_i\right)\right\}\in X_{\mathrm{l}}\times Y_{\mathrm{l}}$在初始化阶段进行监督训练，标注数据集中仅包含已知类的样本，其中Y_{l}是标注样本X_{l}的标签空间，即已知类别集合．在类别发现阶段中，给定已标注数据集并额外使用未标注数据集$D^{\mathrm{u}}=\left\{\left(x_i,y_i\right)\right\}\in X_{\mathrm{u}}\times Y_{\mathrm{u}}$进行GCD设置的训练．在GCD设置中，$D^{\mathrm{u}}$包括已知和未知类样本，其中$Y_{\mathrm{u}}$是未标注示例$X_{\mathrm{u}}$的标签空间，且$Y_{\mathrm{l}}\in Y_{\mathrm{u}}$，$n_{Y_{\mathrm{l}}}$与$n_{Y_{\mathrm{u}}}$分别为$Y_{\mathrm{l}}$和$Y_{\mathrm{u}}$的类别数，所有类别的总数为$K=n_{Y_{\mathrm{u}}}$，总数据集表示为$D=D^{\mathrm{l}}\cup D^{\mathrm{u}}$．GCD旨在利用训练过程$D^{\mathrm{l}}$中的知识去预测$D^{\mathrm{u}}$中样本的标签．此外，分类参数的一个常见假设是标注类别数目K_{l}和未知类别数目K_{n}是已知的[9,18]，或者可以使用文献[6,19]的方法进行估计．为了将钢材表面重复性缺陷簇的发现任务对齐GCD设置，本文设置已知类（或旧类）来表述已知簇，使用未知类（或新类）来表述未知簇．

1.2　网络整体架构图

本文的网络整体架构如图2所示．本文在SimGCD[14]基础上提出了两阶段广义类别发现框架．图2a展示了初始阶段的网络结构：该阶段采用完全监督的方式，在已知类别的标注样本上训练模型．该阶段的网络由1个编码器f和2个并行的头部组成：参数化分类头与投影头．编码器f使用经过DINO[20]预训练的ViT-B/16模型[21]，训练时仅微调编码器的最后一个块；参数化分类头由K_{l}个可学习的已知类别原型向量$\boldsymbol{c}_{\mathrm{old}}=\left[c_1,c_2,\cdots,c_{K_{\mathrm{l}}}\right]$构成，用于对已标注样本进行分类；投影头由一个2层多层感知机（multilayer perceptron，MLP）构成．该阶段的核心学习机制细节如图2d所示，通过增强视图间的特征一致性提升表示能力．

在类别发现阶段，如图2b所示，网络结构基本与初始阶段保持一致，区别在于分类头进行了维度扩展（图2b中扩展分类头的右侧区域），引入了K_{n}个可学习的未知类别原型向量$\boldsymbol{c}_{\mathrm{new}}=\left[c_{K_{\mathrm{l}}+1},c_{K_{\mathrm{l}}+2},\cdots,c_{K_{\mathrm{l}}+K_{\mathrm{n}}}\right]$，同时使用层级分布对齐正则化方法（第1.4节）平衡预测分布，其核心优化目标如图2c所示．除此之外，该阶段的参数初始化策略为：除$\boldsymbol{c}_{\mathrm{new}}$外的网络权重参数均继承自初始阶段的监督模型，而$\boldsymbol{c}_{\mathrm{new}}$采用聚类引导的方式进行初始化，其初始化方式可参见1.5节．

1.3　两阶段广义类别发现

不同于SimGCD以单阶段方式直接进入类别发现阶段，本文提出的两阶段广义类别发现框架包括初始阶段与新类发现阶段，并分别在D^{l}与D训练集中进行表征学习与分类学习，以充分建模新旧类别的内部表示结构．

表征学习旨在学习一种能有效区分各类别的特征表示．其中自监督对比损失可增强样本的特征表示一致性，其目标函数$L_{\mathrm{rep}}^{\mathrm{u}}$为

$$L_{\mathrm{rep}}^{\mathrm{u}}=\frac{1}{n_B}\sum_{x_i\in B}-\ln\frac{\exp\left(\hat{\boldsymbol{z}}_i^{\top}\tilde{\boldsymbol{z}}_i/t_{\mathrm{u}}\right)}{\sum_{x_j\in B}\exp\left(\hat{\boldsymbol{z}}_j^{\top}\tilde{\boldsymbol{z}}_i/t_{\mathrm{u}}\right)}.\tag{1}$$

其中：B为给定训练集的批次训练数据；n_B为B的样本数目；设图像样本x_i经主干编码器f得到表示特征$v_i=f(x_i)$，再通过投影头映射为表征特征

$z_i=g(v_i)$;对每个样本x_i构造两种数据增强视图$\tilde{x}_i$和$\hat{x}_i$,并分别编码得到$\tilde{z}_i$与$\hat{z}_i$;$\hat{z}_j$是与样本x_i同批次下样本的表征特征;t_u为温度,用于控制视图间分布的平滑度.

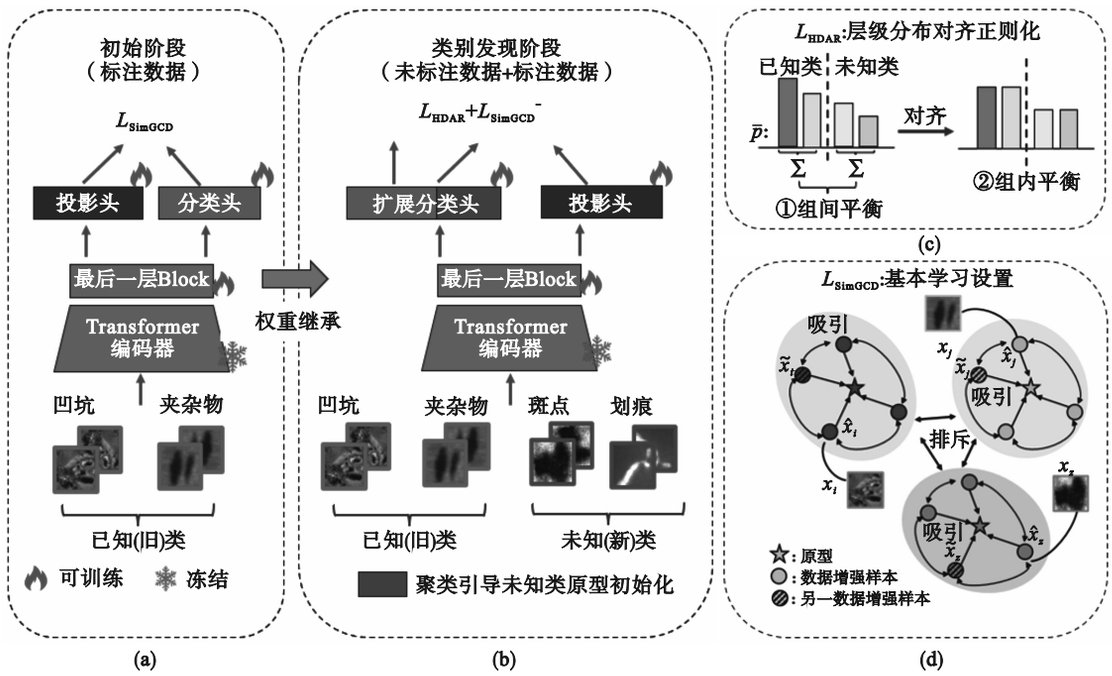


图2 本文网络结构

Fig. 2 Network structure in this paper

(a)—初始阶段;(b)—类别发现阶段;(c)—层级分布对齐正则化;(d)—基本学习设置.

为了进一步利用已标注数据集D^l,引入监督对比损失,将同类样本视为正对,不同类样本视为负对.利用监督对比损失,通过最大化同一类样本之间的相似性,同时最小化不同类样本之间的相似性,使得模型学习到的特征表示更具判别性.其目标函数为

$$L_{\text{rep}}^{\text{s}}=\frac{1}{n_{B^l}}\sum_{x_i\in B^l}\frac{1}{n_{N_i}}\sum_{p\in N_i}-\ln\frac{\exp\left(\hat{z}_i^{\top}\tilde{z}_{\text{p}}/t_{\text{c}}\right)}{\sum_{n\neq i}\exp\left(\hat{z}_i^{\top}\tilde{z}_n/t_{\text{c}}\right)}. \quad (2)$$

其中:N_i是小批量中所有与x_i具有相同标签的图像的索引;n_{N_i}是N_i的数目;B^l表示当前批次中的标注数据;n_{B^l}是B^l的数目;t_{c}为一个额外的温度;$\tilde{z}_{\text{p}}$与$\tilde{z}_n$分别代表与样本x_i同类表征特征和除x_i外的样本表征特征.最终的表征学习目标为两种损失的加权组合:$L_{\text{rep}}=\lambda L_{\text{rep}}^{\text{s}}+(1-\lambda)L_{\text{rep}}^{\text{u}}$,$\lambda$是加权超参数.此阶段仅更新编码器与投影头参数,并为后续分类学习提供高质量特征表征.

分类学习的目的是学习一种可以为所有未标注数据分配标签的分类头.与SimGCD的参数化原型分类结构保持一致,本文同样为每个类别设置可学习原型$[c_1,c_2,\cdots,c_K]$.但由于初始阶段已通过监督学习得到稳定的已知类原型,后续的类别发现便无需从随机初始化开始,而是直接在已有语义空间中进行新类结构的发掘与扩建,从而减少伪聚类扰动并提升开放类可分性.

对于每个样本的一个数据增强视图$\hat{x}_i$,分类头的输出是一个概率向量$\hat{\boldsymbol{p}}_i$.更准确地说,$\hat{\boldsymbol{p}}_i$是通过使用Softmax函数对原型分类头的输出进行归一化而产生的,其在每个可学习原型c_k对应的类别k上的概率$\hat{p}_{i,k}$可表述为

$$\hat{p}_{i,k}=\frac{\exp\left[\frac{1}{t_{\text{s}}}\left(\hat{\boldsymbol{v}}_i/\|\hat{\boldsymbol{v}}_i\|\right)^{\text{T}}\left(c_k/\|c_k\|\right)\right]}{\sum_{j=1}^{K}\exp\left[\frac{1}{t_{\text{s}}}\left(\hat{\boldsymbol{v}}_i/\|\hat{\boldsymbol{v}}_i\|\right)^{\text{T}}\left(c_j/\|c_j\|\right)\right]}. \quad (3)$$

其中:t_{s}为温度,控制预测分布是否尖锐;$\hat{\boldsymbol{v}}_i$表示样本$\hat{x}_i$经过编码器f的输出表征.对于标注数据,分类监督目标为真实标签与模型预测概率之间的交叉熵损失,旨在最小化二者之间的差异,从而有效提升模型对已知类别的判别能力.该目标函数$L_{\text{cls}}^{\text{s}}$表示为

$$L_{\text{cls}}^{\text{s}}=\frac{1}{n_{B^l}}\sum_{i\in B^l}\sum_{k}-y_i^k\ln\hat{p}_{i,k}. \quad (4)$$

式中,y_i^k为第i个样本在第k个类别上的真实标签.

在无监督分类学习中,本文采用自蒸馏方法生成软伪标签$\tilde{\boldsymbol{p}}_i$,并将其作为"类概率分布"形式的监督信号,其中$\tilde{\boldsymbol{p}}_i$是由样本的另一个数据增强视图$\tilde{x}_i$在式(3)中以更低的温度t_{t}(即更尖锐的预测分布)计算得到,与预测输出的概率向量$\hat{\boldsymbol{p}}_i$进

行交叉熵学习.随后,使用交叉熵损失项$L_{\mathrm{cls}}^{\mathrm{u(ce)}}$与熵正则化项$H(\bar{p})$组合构建无监督学习目标$L_{\mathrm{cls}}^{\mathrm{u}}$.其中,交叉熵损失促使模型在不同视图间保持一致性;而熵正则化项通过最大化预测分布的熵值,鼓励模型学习更均匀且分散的类别划分,从而有助于增强未知类别的可分性.

$$L_{\mathrm{cls}}^{\mathrm{u(ce)}}=\frac{1}{n_B}\sum_{x_i\in B}\sum_{k}-\tilde{p}_{i,k}\ln\hat{p}_{i,k}, \tag{5}$$

$$H(\bar{p})=\epsilon\sum_{k}\bar{p}_k\ln\bar{p}_k, \tag{6}$$

$$L_{\mathrm{cls}}^{\mathrm{u}}=L_{\mathrm{cls}}^{\mathrm{u(ce)}}+H(\bar{p}). \tag{7}$$

其中:ϵ是熵正则化项的权重;下标k表示类别索引;$\bar{p}=\frac{1}{2n_B}\sum_{i\in B}(\hat{\boldsymbol{p}}_i+\tilde{\boldsymbol{p}}_i)$,为批次中每个类的平均预测概率.分类学习目标组合为$L_{\mathrm{cls}}=\lambda L_{\mathrm{cls}}^{\mathrm{s}}+(1-\lambda)L_{\mathrm{cls}}^{\mathrm{u}}$.

表征学习目标L_{rep}与分类学习目标L_{cls}联合构成基础训练目标:$L_{\mathrm{SimGCD}}=L_{\mathrm{rep}}+L_{\mathrm{cls}}$.为便于介绍本文的层级分布对齐正则化目标$L_{\mathrm{HDAR}}$,本文将熵正则化项单独记为$L_{\mathrm{reg}}$,并将除熵正则化外的学习目标记为$L_{\mathrm{SimGCD}^-}$.

$$L_{\mathrm{reg}}=(1-\lambda)H(\bar{p}), \tag{8}$$

$$L_{\mathrm{SimGCD}^-}=L_{\mathrm{SimGCD}}-L_{\mathrm{reg}}. \tag{9}$$

两阶段广义类别发现的总学习目标如式(10)所示,L_{init}为初始阶段的学习目标,L_{cd}为类别发现阶段的学习目标.当初始阶段训练完成后,类别发现阶段模型通过严格权重继承的方式,复制初始阶段监督模型的全部参数,并在此基础上按照既定策略继续训练.后续消融实验表明,初始阶段习得的精确旧类表征可实现高效迁移,显著提升类别发现任务的性能.

$$\left.\begin{aligned}L_{\mathrm{init}}&=L_{\mathrm{SimGCD}^-}+L_{\mathrm{reg}}, &&\text{初始阶段;}\\ L_{\mathrm{cd}}&=L_{\mathrm{SimGCD}^-}+L_{\mathrm{HDAR}}, &&\text{类别发现阶段.}\end{aligned}\right\} \tag{10}$$

与单阶段SimGCD相比,此结构化训练流程对先验原型约束进行简单初始化,显著减少了初始特征分布波动对新类拟合的干扰,并为后续引入层级分布对齐正则化奠定了可迁移的语义原型空间.

1.4　层级分布对齐正则化

在钢材表面重复性缺陷簇发现任务中,样本数量分布不均与类别间特征表达的高度相似性是两大显著挑战.一方面,不同缺陷类别在实际采集过程中样本数量悬殊,训练阶段模型更容易偏向于频次较高的已知类别,从而削弱对低频或未知类的建模能力[22].另一方面,钢材表面缺陷的形成常受环境干扰及材料特性等因素共同影响,即使是不同类别的缺陷,在图像表现上亦可能呈现出高度相似的边缘结构、纹理模式和局部形态[23].这种跨类相似性进一步加剧了模型在类别判断上的难度,尤其对未知类别的识别与分离提出了更高要求.如式(8)所示,这种对所有类别施加全局约束的策略在样本不均衡与类别间差异小的情况下反而会加剧强类主导的问题.为克服这些局限性,本文提出一种新的正则化框架——层级分布对齐正则化.具体而言,HDAR在两个不同的层级上引导模型的预测分布:在宏观层级上使用组间平衡,在已知类组与未知类组之间通过约束它们的边缘概率实现两者在模型整体关注度上的平衡;在微观层级上使用组内对齐策略,分别在已知类组内部和未知类组内部独立地促进预测概率的均匀化,以维持已知类的区分度并鼓励对未知类的充分探索.

通过这种层级化的设计,HDAR能够对模型的预测行为施加更精细、更具针对性的控制.本文提出的HDAR学习目标L_{HDAR}由分别对应上述两个层级目标的组件构成,并仅在类别发现阶段执行,具体如下所述:

1)组间平衡损失.为了解决类别发现阶段广义类别发现中模型可能出现的预测偏差和性能退化问题,需要对模型的预测分布施加明确的约束.考虑到模型在训练初期对已知类别具有较强的判别能力,将所有已知类别视为一个已知类组,并将所有潜在的未知类别视为未知类组.这样,可以在宏观层面将类别平衡问题简化为一个概念上的二分类问题.具体来说,计算批次中样本的边缘概率$\bar{p}\in\mathbf{R}^K=\frac{1}{2n_B}\sum_{i\in B}(\hat{\boldsymbol{p}}_i+\tilde{\boldsymbol{p}}_i)$,$\bar{p}_{\mathrm{old}}^{\mathrm{in}}\in\mathbf{R}=\sum_{k\in Y_{\mathrm{l}}}\bar{p}_k$和$\bar{p}_{\mathrm{new}}^{\mathrm{in}}\in\mathbf{R}=\sum_{k\in Y_{\mathrm{n}}}\bar{p}_k$是标量,分别表示已知、未知类别群组内部的边缘概率和,并且$\bar{p}_{\mathrm{old}}^{\mathrm{in}}+\bar{p}_{\mathrm{new}}^{\mathrm{in}}=1$.本文定义组间平衡损失$L_{\mathrm{group}}$如下:

$$L_{\mathrm{group}}=\bar{p}_{\mathrm{old}}^{\mathrm{in}}\ln\bar{p}_{\mathrm{old}}^{\mathrm{in}}+\bar{p}_{\mathrm{new}}^{\mathrm{in}}\ln\bar{p}_{\mathrm{new}}^{\mathrm{in}}+\ln 2. \tag{11}$$

该损失函数基于二元类别的熵形式构建,常数项ln2的加入确保了损失值非负.通过最小化该项,模型被鼓励在已知类别组与未知类别组之间做出更明确的区分,从而在利用已知知识的同时,保持对新类别潜在存在的敏感性.该设计在已知类别指导下实现了对新类别探索与分类边界清晰度之间的平衡.

2)组内均匀损失.对于已知类别,期望模型

不仅能区分它们,而且能稳定地识别所有已知类别,避免模型仅对少数几个容易或常见的已知类产生高置信度预测,而忽略或遗忘其他已知类(即“类内坍塌”或过拟合).保持已知类内部预测的相对均匀性,有助于维持模型对整个已知类别空间的判别能力.形式上,考虑在批次边缘概率$\bar{p}$仅属于K_{l}个已知类别的输出神经元对应的平均预测概率向量$\bar{\boldsymbol{p}}_{\mathrm{old}}=\left[\bar{p}_1,\bar{p}_2,\cdots,\bar{p}_{K_{\mathrm{l}}}\right]$,本文通过惩罚归一化后的旧类概率分布$\bar{p}_{\mathrm{old,norm}}=\bar{p}_{\mathrm{old}}/\bar{p}^{\mathrm{in}}_{\mathrm{old}}$与理想的均匀分布$\bar{p}^{\mathrm{u}}_{\mathrm{old}}$(每个旧类的概率为$1/K_{\mathrm{l}}$)之间的显著偏离来实现这一目标.本文采用平方欧氏距离作为衡量这种偏离程度的度量,因为它能有效且计算简便地惩罚任何远离均匀状态的概率尖峰.本文定义已知类群组内平衡损失L_{old}如下:

$$L_{\mathrm{old}}=\sum_{k=1}^{K_{\mathrm{l}}}\left(\bar{p}_{k,\mathrm{old,norm}}-1/K_{\mathrm{l}}\right)^2. \tag{12}$$

其中,$\bar{p}_{k,\mathrm{old,norm}}$是归一化后的旧类概率分布$\bar{p}_{\mathrm{old,norm}}$的第$k$类的概率值.

在处理潜在的未知类别时,需要鼓励模型进行广泛、均匀的探索,而不是过早地将所有新颖性信号汇聚到少数几个假设的新簇上.类似地,考虑在批次边缘概率$\bar{p}$仅属于K_{n}个未知类别的输出神经元对应的平均预测概率向量$\bar{\boldsymbol{p}}_{\mathrm{new}}=\left(\bar{p}^{K_{\mathrm{l}}+1},\bar{p}^{K_{\mathrm{l}}+2},\cdots,\bar{p}^{K_{\mathrm{l}}+K_{\mathrm{n}}}\right)$,并将其归一化得到条件概率分布$\bar{p}_{\mathrm{new,norm}}=\bar{p}_{\mathrm{new}}/\bar{p}^{\mathrm{in}}_{\mathrm{new}}$.通过最小化该分布与理想均匀分布$\bar{p}^{\mathrm{u}}_{\mathrm{new}}$(每个未知类的概率为$1/K_{\mathrm{n}}$)之间的平方欧氏距离,能够有效防止新类别的概率质量过度集中,鼓励模型在新类别空间中保持更广泛的感受野,为发现多样化的新类别模式奠定基础.本文定义未知类群组内平衡损失L_{new}如下:

$$L_{\mathrm{new}}=\sum_{k=K_{\mathrm{l}}+1}^{K_{\mathrm{l}}+K_{\mathrm{n}}}\left(\bar{p}_{k,\mathrm{new,norm}}-1/K_{\mathrm{n}}\right)^2. \tag{13}$$

最终的层级分布对齐正则化学习目标是上述3个分量的加权和:

$$L_{\mathrm{reg}}=(1-\lambda)\left(w_{\mathrm{group}}L_{\mathrm{group}}+w_{\mathrm{old}}L_{\mathrm{old}}+w_{\mathrm{new}}L_{\mathrm{new}}\right). \tag{14}$$

其中,w_{group},w_{old}与w_{new}为超参数,用于调整不同优化目标的相对重要性.后续实验展示了本文提出的层级分布对齐正则化方法替换熵正则化方法所带来的性能提升.

1.5 未知类原型初始化

随机初始化分类头的权重会导致未知类原型在初始阶段即偏离可行区域,引发训练震荡.为此,本文利用聚类结果为未知类别提供确定性起点.具体来说,利用初始阶段的监督模型对D^{u}提取样本表征,在所有样本表征上执行K-means聚类算法计算K个归一化簇中心$\boldsymbol{c}_i(i=1,2,\cdots,K)$.考虑到部分聚类中心可能与已知类别分布高度重合,通过余弦相似度筛选出最具“未知类代表性”的簇中心.具体而言,对于每个簇中心$\boldsymbol{c}_j$,计算其与初始阶段的已知类分类头权重$\left(\boldsymbol{\phi}^{\mathrm{old}}_m,m=1,2,\cdots,K_{\mathrm{l}}\right)$的最大余弦相似度,并选取其中与已知类最不相似的$K_{\mathrm{n}}$个簇中心索引,满足:

$$t_j=\mathrm{top}_{K_{\mathrm{n}}}(-\max_m \boldsymbol{c}_j^{\top}\boldsymbol{\phi}^{\mathrm{old}}_m),j=1,\cdots,K_{\mathrm{n}}. \tag{15}$$

其中:t_j为第j个被选中的簇中心索引;$\mathrm{top}_{K_{\mathrm{n}}}(\cdot)$为取前$K_{\mathrm{n}}$大的值所对应索引的操作,用于从所有簇中心中筛选出与已知类最不相似的K_{n}个;$-\max_m \boldsymbol{c}_j^{\top}\boldsymbol{\phi}^{\mathrm{old}}_m$表示簇中心$\boldsymbol{c}_j$与所有已知类原型$\boldsymbol{\phi}^{\mathrm{old}}_m$的余弦相似度取最大值后再取负,值越大代表该簇中心与已知类的相似度越低.最终,将这些索引对应的簇中心$\boldsymbol{c}_{t_j}$(即索引为t_j的原始簇中心)作为分类头中未知类别原型$\boldsymbol{c}^j_{\mathrm{new}}$的初始化:

$$\boldsymbol{c}^j_{\mathrm{new}}=\boldsymbol{c}_{t_j},j=1,2,\cdots,K_{\mathrm{n}}. \tag{16}$$

1.6 类别数量估计

在GCD任务中,未标注数据集D^{u}的真实类别数量通常未知.本文采用文献[6]中的策略,通过已标注数据集D^{l}反向估计整体类别数量,并在不改变其核心思想的前提下将该方案应用于钢材表面缺陷聚类任务.由于拥有真实标签,可将其作为监督信号,用于评价不同聚类数的优劣.为对齐“聚类类别索引”与“真实类别标签”,本文采用Hungarian算法进行最优匹配.当候选所有类别总数K大于真实类别数量时,多余类别会被映射为“空标签集合”;若K小于真实类别数量,则部分真实类别也会被映射为空集合.在这两种情况下,模型的聚类精度都会下降.因此,文献[6]将聚类精度$\mathrm{ACC}(K)$视作关于K的黑盒函数:$\mathrm{ACC}(K)=f(K,D)$,并假设其在正确类别数附近取得峰值.

不同于穷举遍历所有K的方法[9],本文使用Brent算法对$\mathrm{ACC}(K)$进行一维黑盒优化搜索.具体实现如下,首先将全部样本组成数据集$D=D^{\mathrm{l}}\cup D^{\mathrm{u}}$;其次在对应特征空间中,在类别数的约束区间$\left[K_{\min},K_{\max}\right]$内,对每个$K$多次运行$K$-means并在$D^{\mathrm{l}}$上计算映射后的聚类精度ACC;最后,以$\mathrm{ACC}(K)$作为目标函数,使用Brent单峰优化寻找最优聚类数$\hat{K}$.该方法无需额外标签参与训练,仅通过D^{l}的部分标签结构引导类别数估计,具有较高的效率与可扩展性.

2　实验结果与讨论

本文通过客观和主观评估实验证明所提方法的有效性.

2.1　数据集介绍

为了验证本文所提方法的有效性,实验使用2个数据集:一是11类混合缺陷数据集,其中5类采集自某钢厂热轧棒材生产线产生的重复性缺陷,其余6类选自NEU-DET[24]热轧带钢数据;二是GC10-DET[25]数据集,收录的是钢板表面的10类缺陷图像,用于验证跨材质、跨工艺的泛化能力.选用上述钢材体系主要基于两点:第一,热轧工艺在工业生产中广泛存在,且易产生由轧辊磨损、氧化、物料嵌入等引发的重复性缺陷(如辊印、划痕、凹坑);第二,热轧棒材、热轧带钢与钢板三类缺陷数据在表面纹理、缺陷尺度和成因上既有共性也存在差异,可用于验证层级分布对齐方法在跨钢种表面图像中的鲁棒性与泛化能力.本文遵循SimGCD[14]的数据集设置方式:将所有类的50%作为标注(已知)类别,这些标注类别中50%的图像用于构建已标注数据集D^{l},而数据集中剩余的图像被视为未标注数据集D^{u}.表1为两种数据集的统计数据,分别介绍了标注类别数目n_{Y_l}、已标注样本数目n_{D^l}、无标注类别数目n_{Y_u}与无标注样本数目n_{D^u}.

表1　数据集介绍

Table 1　Dataset description

数据集	n_{Y_l}	n_{D^l}	n_{Y_u}	n_{D^u}
混合缺陷数据集	6	1 106	11	3 966
GC10-DET	5	776	10	1 967

因篇幅有限,本研究仅在混合缺陷数据集上进行了可视化展示,结果如图3所示,共展示11类缺陷的50幅图像.

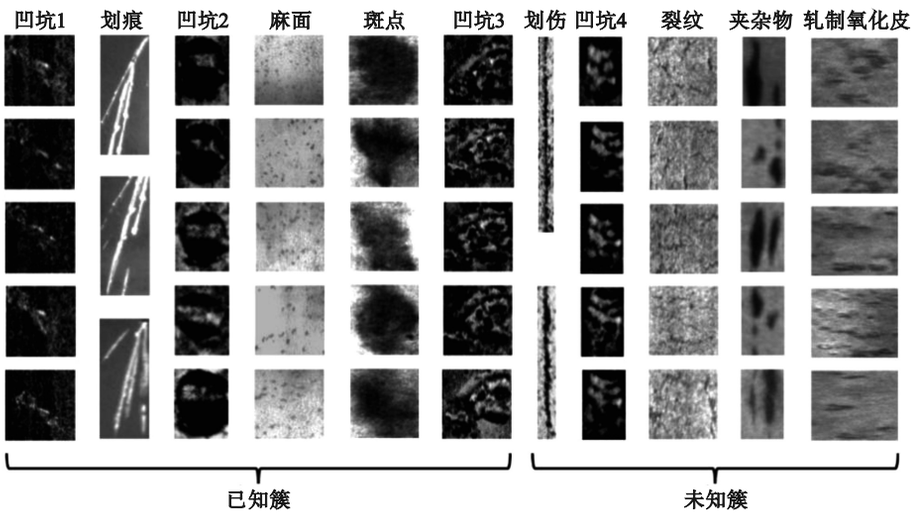


图3　钢材表面重复性混合缺陷数据集缺陷可视化

Fig. 3　Defect visualization of steel surface repetitive mixed defect dataset

2.2　实验参数设置

本文的实验参数与SimGCD[14]保持一致,使用768维的[CLS]token输出作为图像特征.训练时采用128的批尺寸(batch size),初始阶段进行30轮次完整训练,类别发现阶段进行200轮次完整训练.学习率从0.1开始,并使用余弦调度器(cosine scheduler)进行衰减.在类别发现阶段,平衡因子λ设置为0.35,温度t_u和t_c分别设置为0.07和1.0.对于分类学习目标,将t_s设置为0.1,并将t_t初始化为0.07,然后在类别发现阶段起始30轮次完整训练中使用余弦调度器将其预热到0.04.层级分布对齐正则化方法权重w_{group},w_{old}与w_{new}分别设置为3,1与1.为保证公平性,所有实验均使用NVIDIA GeForce RTX 3090 GPU完成训练.为了估计未标注数据集中的聚类数目,本文提取总数据集D中样本的表征特征,在聚类数目的约束域上运行类别估计方法[3,6],其中约束域中的最小值$K_{\min}=n_{Y_l}$,最大值$K_{\max}=200$.在类别发现的鲁棒性评估中:局部高斯模糊的模糊核半径设为3像素,模糊区域在整幅图像中随机选取以模拟服役过程中的局部磨损、污染模糊现象;在腐蚀类斑点扰动中,每张图像随机生成2~3处伪腐蚀区域,斑点半径4~8像素随机采样,为体现金属表面氧化、剥落或点蚀等服役缺陷的暗化特征,斑点灰度值从区间[40,120]内随机选取.

2.3 钢材缺陷数据集实验分析和讨论

2.3.1 客观评估指标

本文采用聚类精度(ACC)[6]评估模型对不同缺陷的聚类能力,聚类精度高表明模型能够更有效地识别不同重复下的缺陷表征结构,并将相似的结构分组在一起.在评估过程中,分别假设$\hat{y}_i$与y_i作为未标注数据集D^{u}中样本x_i的预测类别和真实类别,$\mathrm{ACC}=\frac{1}{N}\sum_{i=1}^{N}I\left(y_i=\delta\left(\hat{y}_i\right)\right)$,其中:$N=n_{D^{\mathrm{u}}}$;$\delta(\cdot)$为匹配预测聚类分配到真实类标签的最优排列;$I(\cdot)$为统计匹配成功的对数.此外,本文参照文献[6],列出了所有类别、已知类别和未知类别的聚类精度.

2.3.2 与先进方法比较

为了评估层级分布对齐正则化的性能,将本文方法与8个先进的GCD方法、2个传统聚类方法和2个正则化方法进行了比较.传统聚类方法通过HOG[26]提取整图梯度方向直方图作为全局特征,该方法通过在固定网格上统计局部梯度幅值与方向分布,并对块进行归一化拼接,从而形成定长的结构化向量,可反映缺陷的形貌结构.为提取缺陷形态的高阶特征,引入DINO特征向量,两种特征分别进行聚类并作为非深度学习(non-deep learning)方法基线;8个GCD方法包括4种非参数分类方法(GCD[6],GPC[27],DCCL[28]和CMS[29])和4种参数分类方法(SimGCD[14],PPC-DCR[15],LegoGCD[16]和BPCL-GCD[17]).在正则化方法中,本文提取了SimGCD[14]和Happy[11]的正则化方法,在两阶段广义类别发现框架下进行应用,该方法命名为SimGCD*与Happy*,以此构成不同正则化方法的核心验证基准.

需要指出的是,本文方法定位于"基于特征相似度的重复性缺陷检测"范式.表2显示,尽管非深度学习方法基线(HOG+K-means,DINO+K-means)已具备一定辨别力,但其仅依赖朴素聚类,缺乏对语义结构的显式建模,难以充分挖掘缺陷形态的高阶相似性;相比之下,GCD框架通过联合表征学习与伪标签自训练,能够更有效地提取并归纳同类缺陷的细粒度特征,从而在各评价维度上均显著超越传统聚类方案.

表2 本文方法与SOTA方法在混合缺陷数据集及GC10-DET聚类精度上的比较结果

Table 2 Comparison results of clustering accuracy between proposed method and SOTA methods on mixed defect dataset and GC10-DET

类型	方法	会议	混合缺陷数据集聚类精度/%			GC10-DET聚类精度/%		
			所有类别	已知类别	未知类别	所有类别	已知类别	未知类别
传统聚类	HOG+K-means	CVPR'2005	23.68	25.75	22.64	21.32	19.45	22.25
	DINO+K-means	ICCV'2021	58.85	73.23	53.28	54.43	72.84	42.41
	DINO+semi-K-means	ICCV'2021	67.12	76.22	63.60	63.72	73.25	57.51
非参数化	GCD	CVPR'2022	88.88	98.92	85.00	69.62	95.99	52.42
	DCCL	CVPR'2023	63.54	83.09	55.98	67.38	74.07	63.02
	GPC	ICCV'2023	77.11	60.22	83.64	66.25	75.31	60.34
	CMS	CVPR'2024	73.95	98.28	64.55	74.86	94.14	62.28
参数化	SimGCD	ICCV'2023	87.01	85.08	87.76	69.42	89.92	56.04
	PPC-DCR	NN'2024	87.95	97.47	84.27	64.82	81.06	54.22
	LegoGCD	CVPR'2024	81.69	95.39	76.40	70.51	88.89	58.52
	BPCL-GCD	ICME'2024	86.79	91.68	87.27	69.13	88.99	56.17
正则化	SimGCD*	ICCV'2023	90.34	96.56	87.94	74.25	80.66	70.07
	Happy*	NeurIPS'2024	91.83	97.74	89.55	75.71	87.96	67.72
	HDAR(本文)	—	94.88	97.65	93.81	80.42	92.08	72.82

注:"*"代表该方法以两阶段广义类别发现设置进行了适配;"NN"代表neural networks.

在正则化方法中,SimGCD*是在平均熵最大化的基础上发现未知类别;Happy*旨在通过结构化的方式约束模型的平均预测概率分布,以实现组间平衡和组内均匀,但其核心方案依旧基于信息熵最大化原理,模型更容易在训练中对未知类识别上出现偏好或过度集中.本文提出的HDAR方法改善了其预测偏差与强类主导问题,如表2所示,本文HDAR方法相比SimGCD*在混合缺陷数据集与跨工艺钢板缺陷数据集GC10-DET的所有类别的聚类精度分别提升了4.54%与6.17%,已知类别的聚类精度分别提升了1.09%与11.42%,未知类别的聚类精度分别提升了5.87%与2.75%.HDAR与Happy*都致力于通过分层策略解决类别平衡和均匀性问题,并在两个数据集中所有类聚类精度上分别取得最优与次优表现.本文在实现组内均匀损失时对模型偏好的

强抑制能力,使其在促进群组内部有效分离方面更具优势,HDAR 相比 Happy* 在混合缺陷数据集的未知类别聚类性能上提升了 4.26%,在 GC10-DET 的已知类别聚类性能上提升了 4.12%.

结果表明,在混合缺陷数据集与跨工艺钢板缺陷数据集 GC10-DET 上的所有类别聚类精度上,本文 HDAR 方法分别为 94.88% 与 80.42%,优于其他先进方法.由于训练阶段仅使用表面图像,未引入任何冶金学先验知识,该结果说明 HDAR 方法对热轧棒材、带钢及钢板缺陷均具有可迁移的视觉表征能力.

两阶段广义类别发现相对于单阶段广义类别发现展现出更多优势.例如,在混合缺陷数据集中,SimGCD* 相较 SimGCD 方法,在所有类别的聚类精度上实现了 3.33% 的提升,并超越了当前 8 种 GCD 方法.改进结果得益于初始阶段监督模型所提供的精确旧类知识.

2.3.3　类别发现鲁棒性评估

本文所采用的数据主要来源于轧制生产阶段形成的表面缺陷,而实际服役环境(如腐蚀介质、疲劳载荷、冲刷磨损)也会引发具有不同形貌特征的非生产性缺陷.由于现有数据集(包括现场棒材、NEU-DET 与 GC10-DET)并不包含工况信息,本文未对生产性缺陷与服役性缺陷进行显式区分.为此,本文在 GC10-DET 数据集预处理中叠加局部模糊、暗色腐蚀斑点扰动,以模拟腐蚀、冲刷磨损导致的形貌退化;同时随机选取折痕、冲孔、水斑、夹杂物 4 类典型缺陷样本进行扰动前后的可视化展示.其中,折痕为钢板轧制过程中形成的线性形变缺陷,冲孔为外力冲压导致的孔洞型缺陷,水斑由生产中的干燥产生,夹杂物通常表现为小斑点,不规则地分布在带钢表面,4 类缺陷在形态、成因与视觉特征上存在显著差异.具体扰动效果如图 4 所示.

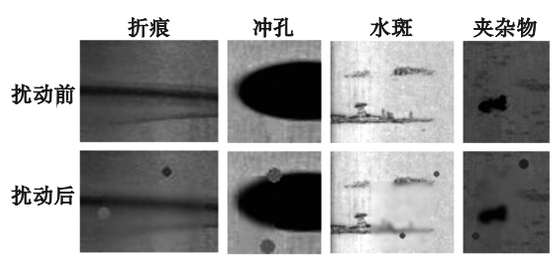


图 4　原始缺陷图像与服役损伤仿真扰动后的对比示例

Fig. 4　Comparison example between original defect images and those after service damage simulation disturbance

本文在不改变训练流程的前提下,对预处理后的 GC10-DET 数据集进行了类别发现鲁棒性验证.如表 3 所示,在正则化类型中,即使在服役缺陷形貌扰动下,本文 HDAR 方法仍保持最优性能,相比 SimGCD*,在所有类别的聚类精度提升了 2.70%,已知类别的聚类精度提升了 6.70%,未知类别的聚类精度提升了 0.09%,说明其对非生产性缺陷的特征扰动具有一定稳健性.

表 3　面向服役损伤仿真数据的聚类鲁棒性结果

Table 3　Clustering robustness results for service damage simulation data

类型	方法	GC10-DET 聚类精度/%		
		所有类别	已知类别	未知类别
正则化	SimGCD*	66.95	86.74	54.03
	Happy*	68.99	91.89	54.03
	HDAR(本文)	69.65	93.44	54.12

2.3.4　主观评估结果

本文提出的层级分布对齐正则化相比于平均熵最大化方法输出了更结构化的预测分布.为了更好地理解层级分布对齐正则化方法所带来的特征分布改进,本文利用 t-SNE[30] 进行了特征集群分布可视化.对混合缺陷数据集所有类别的评估结果如图 5 所示,整体上 SimGCD* 与本文提出的方法可以有效进行簇间划分.如图 5a 所示,可以观察到 SimGCD* 在未知类与未知类 $\{(8,10),(6,7)\}$、已知类与未知类 $\{(1,9),(2,6)\}$ 之间发生了纠缠;如图 5b 所示,本文 HDAR 方法生成了更紧凑的输出,有效地对属于同一类别的样本进行分组,形成了更清晰的类间决策边界.

2.3.5　估计聚类数目

本文采用文献[6]对未标注图像数据集进行类别数量估计,具体的估计方法可参见 1.6 节.表 4 为混合缺陷数据集与 GC10-DET 中类别数量的估计结果,可以观察到本文估计的类别数量非常接近未标注数据集中的真实类别数量,最大误差率为 9%,GC10-DET 能够准确地识别聚类数目.关于误差率,文献[3,6]指出,预测类别总数量受已知标注定义的影响,因此不同标注标准会导致不同的聚类类别数量估计结果.

2.3.6　训练开销与精度增益分析

为进一步验证两阶段策略在效率与性能上的实用价值,本文对比了单阶段 SimGCD 与两阶段的 SimGCD*,HDAR 方法的训练时间与聚类精度.其中 SimGCD* 采用与 SimGCD 一致的正则化

熵损失函数,仅将训练流程改为两阶段.尽管两阶段方法在初始阶段需额外进行已知类的监督预训练,但由于该阶段训练数据规模较小且训练轮次有限(具体可参见2.1节数据集介绍与2.2节实验参数设置),其时间开销占总训练时间的比例极低.如表5所示,两阶段方法在两个数据集中额外引入的初始阶段仅耗时约0.1 h,相对增幅小于4%.然而,在混合缺陷数据集与GC10-DET数据集上,SimGCD*在所有类别聚类精度相较于SimGCD分别提升了3.33%与4.83%,HDAR方法在所有类别的聚类精度相较于SimGCD分别提升了7.87%与11.00%.可以认为,在两阶段方法中,初始阶段步骤以可忽略的时间成本换取显著精度提升,验证了两阶段策略在钢材重复性缺陷簇发现场景下的高效性与实用性.

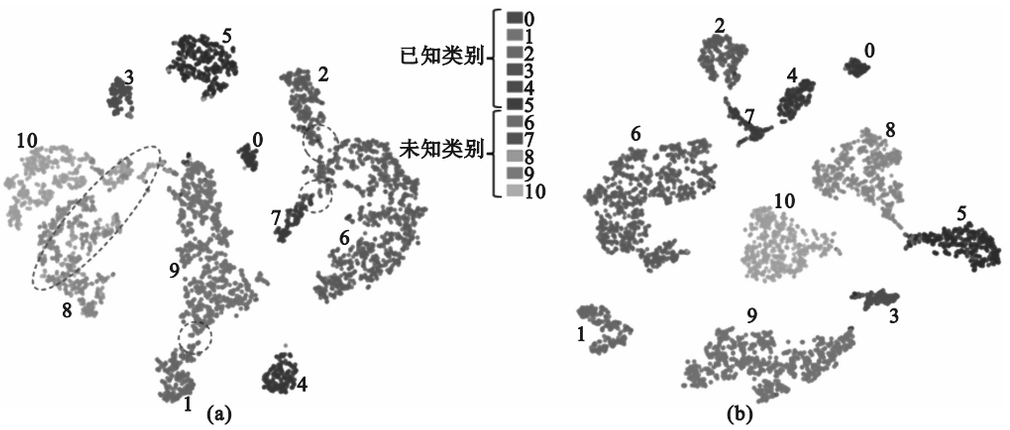


图5 混合缺陷数据集所有类别t-SNE可视化

Fig. 5 t-SNE visualization of all categories in mixed defect dataset

(a)—SimGCD*;(b)—本文HDAR方法.

表4 未标注数据集中类别数量的估计

Table 4 Category number estimation in unlabeled dataset

参数	混合缺陷数据集	GC10-DET
真实类别数量	11	10
预测类别数量	12	10
误差率/%	9	0

2.4 消融实验分析和讨论

针对混合缺陷数据集的损失函数组件消融实验结果如表6所示,本文研究了层级分布对齐正则化方法的各种组件的有效性,包括组件平衡损失L_{group}、已知类群组内平衡损失L_{old}和未知类群组内平衡损失L_{new}.本文将基础组件定义为分类学习与表征学习的结合(不含熵正则化),即L_{SimGCD}.实验1~5的结果有力地支持了本文提出组件的有效性.

表5 两阶段方法与单阶段SimGCD的训练时间与所有类别聚类精度对比

Table 5 Comparison of training time and all-category clustering accuracy between two-stage method and single-stage SimGCD

方法	混合缺陷数据集		GC10-DET	
	训练时间/h	所有类别聚类精度/%	训练时间/h	所有类别聚类精度/%
SimGCD	4.33	87.01	2.39	69.42
SimGCD*	4.43(+2.31%)	90.34(+3.33%)	2.47(+3.35%)	74.25(+4.83%)
HDAR(本文)	4.45(+2.77%)	94.88(+7.87%)	2.48(+3.76%)	80.42(+11.00%)

注:括号中数据为增幅.

实验1中,在缺乏任何显式平衡约束的情况下,模型表现出较差的整体性能,无法有效保持对已知知识的判别能力,甚至发生了灾难性遗忘.额外引入组间平衡损失L_{group}后,如实验2所示,已知类别精度大幅提高,证明了其在平衡新旧类整体关注度、防止遗忘的关键作用.比较实验2和实验4时发现,模型对已知类和未知类的样本识别能力出现显著提升,这表明L_{new}对于促进未知类内部的均匀探索、防止模式坍塌以及提升新类别的可分性至关重要.在广义类别发现任务中,有效区分未知类别通常是主要挑战,因此L_{new}对整体性能的贡献尤为突出.同时,比较实验2和实验3显示,L_{old}有助于增强已知类别内部的区分度,揭示了其稳定已知类内部表示的积极影

响.最终如实验5所示,在所有组件协同工作时,　　模型达到了最佳性能.

表6　损失函数组件消融实验结果

Table 6　Ablation experiment results of loss function components

实验序号	损失函数组件				混合缺陷数据集聚类精度/%		
	L_{SimGCD^-}	L_{group}	L_{old}	L_{new}	所有类别	已知类别	未知类别
1	√	×	×	×	64.80	29.57	78.43
2	√	√	×	×	75.42	94.58	68.01
3	√	√	√	×	83.71	97.56	78.36
4	√	√	×	√	91.70	97.29	89.55
5	√	√	√	√	94.88	97.65	93.81

该消融研究清晰地验证了本文提出的层级分布对齐正则化方法中每一个损失函数组件的必要性和有效性.L_{group}对于建立新旧类别的基础平衡、防止对已知类别的遗忘起着决定性作用;L_{old}有助于稳定和精炼对已知类别的判别;L_{new}则显著促进了对多样化新类别的探索和区分,对提升整体性能特别是新类别发现能力贡献巨大.实验结果表明,只有将这些组件结合起来,才能实现最佳的广义类别发现性能.

3　结　　语

本文面向钢材表面重复性缺陷簇的识别与发现任务,提出了一种基于层级分布对齐正则化的两阶段广义类别发现方法.针对重复性缺陷簇中普遍存在的开放性缺陷形态、误检与漏检导致的分布不稳定,以及仅依赖传统时空信息造成的类别识别偏差、样本聚类不均和特征相似性高等问题,本文方法引入组间平衡与组内对齐相结合的正则化策略,有效提升了模型对已知类别的判别能力及对未知结构的探索能力.实验结果表明,该方法在钢材表面混合缺陷数据集上表现优异,实现了94.88%的整体聚类精度与93.81%的未知类别聚类精度.相较于采用全局熵调控的SimGCD*方法,上述两项指标分别提高了4.54%和5.87%,充分验证了该方法在实现缺陷类别有效分离方面的显著优势.

参考文献:

[1] 吴昆鹏,石杰.基于孪生网络的带钢表面周期性缺陷检测方法[J].冶金自动化,2020,44(6):93-98.
(Wu Kun-peng, Shi Jie. Periodic defect detection method for strip steel surface based on Siamese network [J]. *Metallurgical Industry Automation*, 2020, 44(6): 93-98.)

[2] 李丰范,匡健隆,季佳浩,等.机器学习在金属材料服役性能预测中的应用[J].工程科学学报,2024,46(1):120-136.
(Li Feng-fan, Kuang Jian-long, Ji Jia-hao, et al. Application of machine learning in predicting the service performance of metallic materials [J]. *Chinese Journal of Engineering*, 2024, 46(1): 120-136.)

[3] Jiang Y, Li J S, Yang X, et al. Applications of generative adversarial networks in materials science [J]. *Materials Genome Engineering Advances*, 2024, 2(1): e30.

[4] Xie Y F, Hu W T, Xie S W, et al. Surface defect detection algorithm based on feature-enhanced YOLO [J]. *Cognitive Computation*, 2023, 15(2): 565-579.

[5] 马鸽,李洪伟,严梓维,等.基于多注意力的改进YOLOv5s小目标检测算法[J].工程科学学报,2024,46(9):1647-1658.
(Ma Ge, Li Hong-wei, Yan Zi-wei, et al. Improved small target detection algorithm based on multi-attention and YOLOv5s for traffic sign recognition [J]. *Chinese Journal of Engineering*, 2024, 46(9): 1647-1658.)

[6] Vaze S, Han K, Vedaldi A, et al. Generalized category discovery [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, 2022: 7492-7501.

[7] Zheng M K, You S, Huang L, et al. SimMatch: semi-supervised learning with similarity matching [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, 2022: 14451-14461.

[8] Sohn K, Berthelot D, Carlini N, et al. FixMatch: simplifying semi-supervised learning with consistency and confidence [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 596-608.

[9] Han K, Rebuffi S A, Ehrhardt S, et al. AutoNovel: automatically discovering and learning novel visual categories [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(10): 6767-6781.

[10] Zhong Z, Fini E, Roy S, et al. Neighborhood contrastive learning for novel class discovery [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, 2021: 10862-10870.

[11] Ma S J, Zhu F, Zhong Z, et al. Happy: a debiased learning framework for continual generalized category discovery [J]. *Advances in Neural Information Processing Systems*, 2024, 37: 50850-50875.

[12] Gu P Y, Zhang C Y, Xu R J, et al. Class-relation knowledge distillation for novel class discovery [C] //2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, 2024: 16428-16437.

[13] Wang Y Z, Chen Z Y, Yang D K, et al. Self-cooperation knowledge distillation for novel class discovery [C]// Computer Vision—ECCV 2024. Cham: Springer, 2025: 459-476.

[14] Wen X, Zhao B C, Qi X J. Parametric classification for generalized category discovery: a baseline study [C]//2023

IEEE/CVF International Conference on Computer Vision (ICCV). Paris,2024: 16544-16554.

[15] Hu Z Y, Duan Y, Zhang Y M, et al. Prototypical classifier with distribution consistency regularization for generalized category discovery: a strong baseline[J]. *Neural Networks*, 2025, 182: 106908.

[16] Cao X Z, Zheng X Y, Wang G H, et al. Solving the catastrophic forgetting problem in generalized category discovery [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2024: 16880-16889.

[17] Wang Y, Zhao B C, Lu Y C, et al. Debiased prototypical learning improves generalized category discovery[C]//2024 IEEE International Conference on Multimedia and Expo (ICME).Niagara Falls, 2024: 1-6.

[18] Fini E, Sangineto E, Lathuilière S, et al. A unified objective for novel class discovery [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, 2021: 9284-9292.

[19] Han K, Vedaldi A, Zisserman A. Learning to discover novel visual categories via deep transfer clustering [C]// IEEE/CVF International Conference on Computer Vision (ICCV).Seoul,2019: 8401-8409.

[20] Caron M, Touvron H, Misra I, et al. Emerging properties in self-supervised vision Transformers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal,2021: 9630-9640.

[21] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2020-10-22)[2025-11-30]. https://arxiv.org/abs/2010.11929.

[22] 白云鹍,张昊宇,邢宇翔.基于无监督学习的工业图像异常检测研究综述[J].中国体视学与图像分析,2025,30(1):102-125.
(Bai Yun-kun, Zhang Hao-yu, Xing Yu-xiang. A review of industrial image anomaly detection based on unsupervised deep learning [J]. *Chinese Journal of Stereology and Image Analysis*, 2025,30(1):102-125.)

[23] 吴俊飞,顾新福.Ti2AlNb热轧板中反常轧制组织成因分析[J].中国体视学与图像分析,2024,29(1):19-28.
(Wu Jun-fei, Gu Xin-fu. Mechanism of abnormal rolling microstructure in Ti2AlNb alloy hot-rolled plate [J]. *Chinese Journal of Stereology and Image Analysis*, 2024,29(1):19-28.)

[24] Bao Y Q, Song K C, Liu J, et al. Triplet-graph reasoning network for few-shot metal generic surface defect segmentation [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5011111.

[25] Lyu X M, Duan F J, Jiang J J, et al. Deep metallic surface defect detection: the new benchmark and detection network [J]. *Sensors*, 2020, 20(6): 1562.

[26] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). San Diego, 2005: 886-893.

[27] Zhao B C, Wen X, Han K. Learning semi-supervised gaussian mixture models for generalized category discovery [C]//2023 IEEE/CVF International Conference on Computer Vision(ICCV).Paris, 2023: 16577-16587.

[28] Pu N, Zhong Z, Sebe N. Dynamic conceptional contrastive learning for generalized category discovery[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).Vancouver, 2023: 7579-7588.

[29] Choi S, Kang D, Cho M. Contrastive mean-shift learning for generalized category discovery [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Seattle,2024: 23094-23104.

[30] Van der Maaten L, Hinton G. Visualizing data using t-SNE [J]. *Journal of Machine Learning Research*, 2008, 9(11): 2579-2605.