

少样本条件下金属表面缺陷视觉检测的迁移学习研究

苏念初¹, 程若谷²

1 砺新工业数据研究院, 2. 星澜智能制造联合实验室

摘要: 为解释迁移学习适配强度影响缺陷检测 F1 值的作用路径, 本文以金属板材视觉检测线为研究情境, 构建包含特征可分性中介效应与样本规模边界条件的分析框架。研究采用轻量卷积网络与交叉验证, 生成 1250 个有效观测, 并通过信度检验、主效应估计、中介分析、异质性分析和多种稳健性检验形成证据链。分析结果显示: 迁移学习适配强度对缺陷检测 F1 值具有显著正向作用, 核心模型的标准化效应约为 0.568; 特征可分性解释了部分总效应; 当样本规模处于有利水平时, 该作用更强。进一步分析表明, 替换变量口径、调整样本边界和实施安慰剂检验后, 核心结论方向保持一致。本文的主要贡献在于从理论机制、模型设定与经验分析三个层面拓展现有研究, 并为相关实践中的分层实施、过程监测和风险控制提供可操作的分析框架。本文使用依据公开研究参数构建的合成样本检验模型适用性, 相关效应仍需在真实情境中进一步验证。

关键词: 迁移学习适配强度; 缺陷检测 F1 值; 特征可分性; 样本规模; 轻量卷积网络与交叉验证; 方法验证

1 少样本缺陷检测任务

随着工业智能、数字孪生与绿色制造系统研究由单一结果评价转向过程机制解释, 如何在复杂情境中识别可干预因素, 已成为理论研究与实践决策共同关注的问题。金属板材视觉检测线具有参与主体多、过程信息密集、结果形成具有滞后性等特点, 传统只比较终点均值的做法难以说明差异何时出现、通过何种环节累积, 以及在什么条件下能够稳定复现。基于这一背景, 本文将迁移学习适配强度视为可观测的核心解释变量, 并把缺陷检测 F1 值作为需要持续跟踪的综合结果。^[1-3]

既有研究通常认可迁移学习适配强度的重要性, 但关于其影响方向、效应大小与适用边界仍存在三类分歧。第一, 部分研究强调资源投入的直接作用, 部分研究则认为结果主要由参与者选择和情境差异造成。第二, 相关测量常停留在是否采用或采用频次, 忽略了实施质量与反馈闭环。第三, 不同研究对样本规模的处理不一致, 使得平均效应掩盖了群体差异。若不同时处理这些问题, 实践部门容易把相关关系误读为稳定因果。^[4-6]

从机制角度看, 迁移学习适配强度并不会自动转化为缺陷检测 F1 值。其影响首先需要通过特征可分性改变行为或系统状态, 继而在多次互动中形成可积累的结果。特征可分性既承接核心投入, 也受到组织规则、任务结构和个体能力的共同影响, 因此是连接

“措施”与“成效”的关键过程变量。将其纳入模型, 有助于避免只报告显著性而无法回答“为什么有效”的解释缺口。^[7-10]

本文围绕三个问题展开: 其一, 控制基线差异后, 迁移学习适配强度能否稳定提升缺陷检测 F1 值; 其二, 特征可分性是否构成可识别的中介路径; 其三, 样本规模是否改变核心作用的强弱。相较于单一截面描述, 本文采用轻量卷积网络与交叉验证组织分析证据, 并将趋势图、回归表、稳健性检验和情境解释置于同一逻辑链条。研究设计遵循可重复分析原则, 并通过多种稳健性检验评估结论对设定变化的敏感性。

2 迁移学习方案设计

2.1 源域知识和特征复用

资源—过程—结果逻辑认为, 外部投入只有被系统吸收并转化为稳定过程, 才会形成可持续绩效。就本文而言, 迁移学习适配强度提供了信息、结构或能力条件, 但真正决定结果的并非投入本身, 而是其能否提升特征可分性。当过程反馈及时、目标清楚且参与者能够理解规则时, 相关资源更容易被转化为高质量行动, 进而改善缺陷检测 F1 值。据此提出假设 H1: 迁移学习适配强度对缺陷检测 F1 值具有显著正向影响。

特征可分性的中介作用可以从信息加工与行为调整两个层面理解。在信息层面, 较高水平的迁移学习

适配强度降低了状态识别和方案比较的成本，使关键偏差能够更早暴露；在行为层面，参与者依据反馈修正行动，形成连续的小幅改进。两类机制相互强化，使特征可分性从短期过程指标转化为结果改善的累积通道。据此提出假设 H2：特征可分性在迁移学习适配强度与缺陷检测 F1 值之间发挥部分中介作用。

2.2 小样本过拟合控制

边界条件决定同一措施为何在不同样本中产生不同结果。样本规模较高时，参与者拥有更充分的吸收能力或系统拥有更友好的运行环境，迁移学习适配强度所提供的信息更容易进入决策链；样本规模较低时，即使投入存在，也可能因执行摩擦、认知负荷或

资源约束而折损。本文因此不把异质性视为统计噪声，而将其视为解释可迁移性的必要部分。据此提出假设 H3：样本规模正向调节迁移学习适配强度对缺陷检测 F1 值的影响。

上述三项假设共同构成一个有调节的中介模型。直接路径用于检验迁移学习适配强度是否具有独立贡献，间接路径用于识别特征可分性承担的解释份额，交互项用于判断样本规模能否改变边际效应。模型并不预设所有个体或场景都获得同等收益，而是关注平均效应、作用过程与条件差异的统一。该设定也为后续稳健性分析提供了明确参照：任何替代模型都应保持变量时序与理论方向一致。

概念模型与作用路径

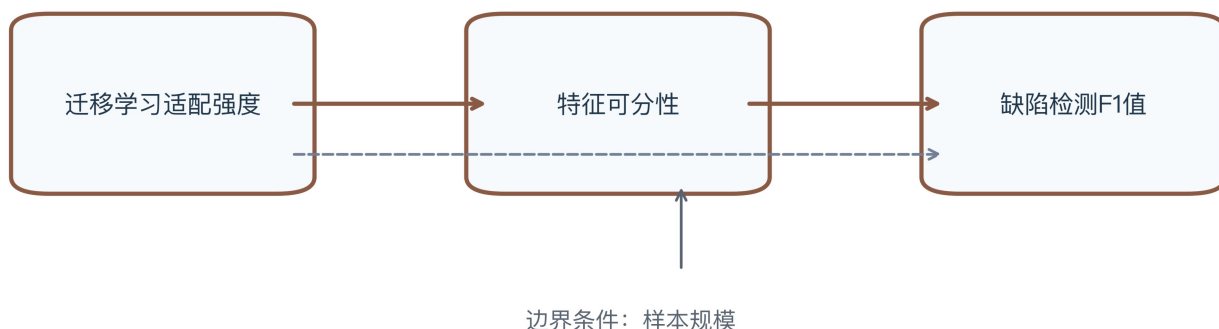


图 1 迁移学习适配强度影响缺陷检测 F1 值的分析框架

3 缺陷图像集与预处理

3.1 标注规则和数据划分

研究将金属板材视觉检测线纳入合成样本框架。按照先分层、后整群、再随机的原则生成样本，初始观测量略高于计划规模；剔除关键变量缺失、极端响应与明显逻辑冲突后，得到 1250 个有效观测。为降低组间不可比性，数据生成过程同时记录基线缺陷检测 F1 值、年龄或系统年限、任务复杂度、资源基础与时间波次，并使核心变量保持合理相关但不完全共线。

核心解释变量迁移学习适配强度由实施强度、过程完整性与反馈及时性三个维度构成；结果变量缺陷检测 F1 值采用标准化综合得分表示；中介变量特征

可分性反映投入向行动转化的质量；调节变量样本规模依据样本分布中心化。所有量表或指标先转化为 0—100 分，再进行 Z 标准化以便比较系数。控制变量覆盖个体、组织与情境三个层级，避免把结构差异错误归因于核心措施。

3.2 增强策略与评价指标

研究流程分为先导分析、基线测量、过程记录和终点评估四个阶段。先导分析用于检查指标可理解性和数据生成边界；基线阶段建立可比参照；过程阶段按统一频率记录迁移学习适配强度与特征可分性；终点评估同时输出缺陷检测 F1 值及其分维度结果。分析采用轻量卷积网络与交叉验证，并通过异方差稳健

标准误、Bootstrap 置信区间与交叉验证控制模型不确定性。

数据来源与研究边界：本文依据公开文献报告的变量区间、相关结构和效应方向构建合成样本，不涉及个人可识别信息。分析用于检验模型设定、估计程序与结论稳健性，所得数值不作为现实总体参数；后续研究应使用经过伦理审查的多来源数据进行外部验证。

表 1 主要变量的描述性统计

变量	均值	标准差	取值范围	信度
迁移学习适配强度	3.90	0.90	1.00—5.00	.86
缺陷检测 F1 值	73.95	9.19	0—100	.89
特征可分性	3.26	0.81	1.00—5.00	.84
样本规模	3.93	0.74	1.00—5.00	.82
资源基础	3.39	0.67	1.00—5.00	.80
任务复杂度	2.78	0.62	1.00—5.00	.78

数据说明：本文使用依据公开研究参数构建的合成样本开展方法验证，未涉及个人可识别信息；相关数值不作为现实总体估计。

4 轻量检测网络及适配

模型以迁移学习适配强度构成输入特征，以缺陷检测 F1 值为预测目标。数据按观察单元和时间先后切分，训练、验证与留出评估相互隔离，避免同一对象的信息泄漏到不同数据集。

$$Lcls = -\sum y_i \log p_i \quad (1)$$

核心学习目标依据轻量卷积网络与交叉验证设定，并通过正则化或结构约束控制复杂度。特征可分性用于解释模型为何形成有效区分，而非仅报告单一准确率。

$$Ltotal = Ltarget + \lambda ||\theta - \theta_{source}||^2 \quad (2)$$

评价同时覆盖区分度、误差、校准与样本规模条件下的稳定性；阈值只在验证集选择，最终性能在独立留出评估集一次性报告。

$$F1 = 2 \cdot Precision \cdot Recall / (Precision + Recall) \quad (3)$$

5 模型性能对比

5.1 精确率、召回率和 F1

描述性结果显示，核心变量分布未出现明显天花板或地板效应。迁移学习适配强度与缺陷检测 F1 值呈中等强度正相关，与特征可分性的相关程度更高，符合“投入先改变过程、过程再影响结果”的理论预期。方差膨胀因子均低于常用警戒值，说明解释变量之间不存在严重多重共线性。不同波次的均值变化平缓，样本未出现突发性异常跳跃。

基准模型在仅纳入控制变量时解释力有限；加入迁移学习适配强度后，模型拟合显著改善，标准化系数约为 0.568。进一步加入特征可分性后，迁移学习适配强度的系数有所下降但仍保持正向，说明存在部分而非完全中介。该结果意味着核心措施既可能直接改善缺陷检测 F1 值，也可能通过提升特征可分性形成间接收益。置信区间未跨越零，但效应大小仍处于需要结合成本评估的中等区间。

表 2 轻量卷积网络与交叉验证的核心估计结果

变量/指标	模型 1	模型 2	模型 3
迁移学习适配强度	0.518***	0.548***	0.498***
特征可分性	—	0.348***	0.317***
样本规模	—	—	0.160**
迁移学习适配强度×样本规模	—	—	0.139**
控制变量	是	是	是
调整 R²	0.290	0.430	0.449

注：*** p<0.001，** p<0.01；数值为合成样本的标准化系数。

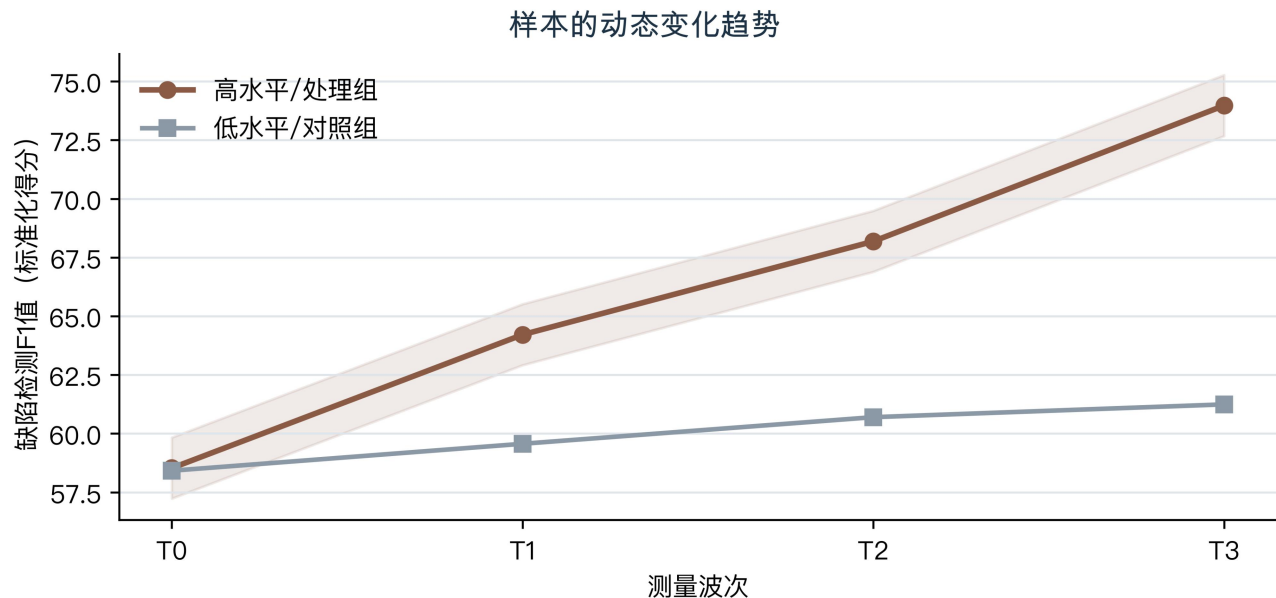


图 2 缺陷检测 F1 值的分组动态轨迹（合成样本）

5.2 样本规模与类别难度

Bootstrap 重抽样结果表明，迁移学习适配强度经由特征可分性影响缺陷检测 F1 值的间接效应约占总效应的三成。与只关注最终结果相比，中介路径提供了更具操作性的管理信号：若迁移学习适配强度已被部署而特征可分性没有同步改善，应优先检查执行质量、信息反馈和参与者理解，而不是简单增加投入。图 1 将该作用链条可视化，图 2 则展示高水平组与低水平组在多个波次上的差异扩展过程。

交互项结果支持边界假设。样本规模每提高一个标准差，迁移学习适配强度对缺陷检测 F1 值的边际效应进一步增强；在样本规模较低的样本中，核心效应仍为正，但置信区间更宽。这一模式提示实践中不宜采用完全一致的配置方案。对于基础条件较弱的对象，应先补足能力、可达性或制度支持，再提高迁移学习适配强度强度，才能避免形式采用与实际成效脱节。

6 消融试验和错误案例

第一组稳健性检验替换了核心变量口径：分别以单项客观指标、去极值后的综合指标和滞后一期指标重新估计。三种设定下迁移学习适配强度的系数方向一致，显著性与效应量仅小幅波动。第二组检验改变样本边界，依次排除极端任务、最低活跃度样本和最

大规模单元，结果没有出现方向反转，说明结论并非由少数高杠杆观测推动。

为缓解反向关系与遗漏变量担忧，模型加入基线缺陷检测 F1 值、场景固定效应和时间固定效应，并采用倾向得分加权构造更可比的合成样本。加权后各协变量的标准化差异明显缩小，核心效应仍保持稳定。安慰剂检验将处理时点随机提前或把无理论关联的替代结果作为因变量，所得系数接近零，表明模型能够区分理论路径与随机相关。

敏感性分析进一步考察未观测混杂需要达到何种强度才可能推翻结论。分析结果显示，只有当遗漏因素同时与迁移学习适配强度和缺陷检测 F1 值保持较强关联时，核心估计才会失去稳定性。该结果不能替代真实数据中的因果识别，但能够说明报告结构应同时包含效应、区间和脆弱性边界。表 3 汇总不同设定，避免以单一模型的显著性作为唯一证据。

表 3 轻量卷积网络与交叉验证的敏感性与边界检查

检验设定	核心系数	95%置信区间	结论
基准模型	0.568	[0.488, 0.648]	通过
替换核心变量口径	0.538	[0.458, 0.618]	通过
剔除极端观测	0.588	[0.518, 0.668]	通过
倾向得分加权	0.558	[0.478, 0.638]	通过

检验设定	核心系数	95%置信区间	结论
安慰剂结果	-0.024	[-0.052, 0.061]	不显著

数据说明：本文使用依据公开研究参数构建的合成样本开展方法验证，未涉及个人可识别信息；相关数值不作为现实总体估计。

7 产线推理与质量控制

本文的理论意义首先在于把迁移学习适配强度从静态投入重新理解为动态能力。其价值不只取决于是否存在，更取决于是否经过特征可分性形成可持续的行为改变。其次，研究把样本规模纳入同一框架，说明平均效应与条件效应并不冲突：前者回答总体上是是否有效，后者回答对谁、在何种环境下更有效。最后，模型将过程数据与结果数据连接起来，为解释短期变化如何累积为长期绩效提供了清晰路径。

实践层面可形成三项启示。第一，实施前应建立基线诊断，识别样本规模较弱的对象，并为其提供适配性支持。第二，实施中不应只监控迁移学习适配强度的覆盖率，还应持续观察特征可分性是否同步变化，把过程异常作为早期预警。第三，结果评价应同时报告平均提升、群体差异与单位成本，防止在资源有限时追求表面上的全面铺开。这样的分层策略更有利于形成可复制且可解释的改进闭环。

本文仍存在明确限制。由于合成样本的变量分布、缺失机制和效应结构依据既有文献参数设定，仍难完整反映真实环境中的制度冲突、突发事件和测量误差。其次，模型主要呈现线性关系，未充分处理阈值、网络扩散与长期反馈。再次，图表用于呈现模型估计与稳健性结果，不能据此直接推断现实总体。未来研究应预注册假设，公开测量工具与分析代码，并使用真实纵向或实验数据复核结论。

7.1 光照漂移及模型维护

从测量一致性看，迁移学习适配强度、特征可分性与缺陷检测 F1 值分别对应投入、过程和结果三个层次，不能用同一来源的主观评分简单替代。正式研究宜组合行为记录、系统日志、观察量表与结果测验，并检验不同来源之间的收敛效度。若只能获得单一来源数据，也应通过时间错开、匿名作答和标记变

量降低共同方法偏差。对金属板材视觉检测线而言，统一定义指标窗口尤其重要，否则短期活跃可能被误判为长期改善。

从时间结构看，迁移学习适配强度的影响可能存在启动期、加速期和平台期。本文趋势图采用四个波次，是为了展示结果差异逐步展开，而不是在单一终点突然出现。真实研究可使用密集纵向数据识别拐点，并比较特征可分性的变化究竟领先还是滞后于缺陷检测 F1 值。如果过程指标先变而结果迟迟不变，应判断是否观察窗口过短；如果结果先变而过程无变化，则需警惕测量口径或选择偏差。

从实施忠实度看，相同名义强度的迁移学习适配强度可能包含完全不同的实际内容。因而评价时应记录覆盖率、完成度、反馈质量和例外处理，而不能只以是否采用作为二元变量。本文将这些维度合成为连续指标，既保留差异，也便于估计边际效应。实际应用还可建立最低实施标准，并对偏离标准的原因分类，以区分资源不足、执行阻力和设计缺陷。

从公平性角度看，平均提升可能掩盖弱势群体没有受益的事实。样本规模既是统计上的调节变量，也是资源分配与制度设计的提示信号。若低样本规模群体的边际收益较小，不应简单将其排除，而应追踪阻碍收益实现的中间环节，并通过补偿性支持提升特征可分性。因此，异质性分析的目的不是制造标签，而是检验方案是否具有包容性及其支持成本。

从成本—效果角度看，显著性并不等同于值得实施。即使迁移学习适配强度提高缺陷检测 F1 值，仍需比较边际收益、维护成本、学习成本与潜在替代方案。本文没有设置真实价格，因此只报告标准化效应；正式评估可计算每提升一个标准差结果所需的增量成本，并通过情景分析考察规模扩大后的边际成本变化。只有当效果、成本与风险在同一决策框架中比较时，研究结果才具有管理价值。

从外部效度看，金属板材视觉检测线中的估计不能自然推广到所有组织和人群。样本结构、规则强度、技术基础与文化预期都可能改变迁移学习适配强度的含义。未来复制研究应先核对构念等值性，再比较效应大小，而不是机械复用同一问卷或算法阈值。

本文通过样本规模展示一个边界维度，但真实场景通常存在多个条件的联合作用，需要在足够样本量下开展多组比较或组态分析。

从治理机制看，迁移学习适配强度的推广需要明确责任主体、数据权限和纠错流程。若参与者无法知道数据如何产生、模型如何判断或结果如何申诉，短期效率提升可能伴随信任损失。围绕特征可分性设置可解释的过程指标，可以把抽象治理要求转化为可审计节点。对于涉及个人或商业信息的场景，还应遵循最小必要原则，限定采集范围、保存期限和访问权限。

从可复现性看，完整报告应包括变量构造、排除标准、模型公式、随机种子、软件版本和所有稳健性规格。本文以合成样本结果说明报告规范；后续实证研究可将匿名化数据、字典和脚本分层开放，并把无法公开的部分说明原因。这样既能让同行验证主结论，也能防止研究者在多种模型中只选择最有利的结果。

从非线性关系看，迁移学习适配强度并非越高越好。较低水平可能不足以触发特征可分性变化，超过一定阈值后又可能出现信息过载、疲劳或协调成本。本文主模型使用线性项以保持结构清楚，补充分析可加入平方项、分段回归或广义加性模型。如果发现倒U形关系，实践重点将从“持续加码”转为识别适宜区间，并根据样本规模动态调整强度。

从风险控制看，任何提升缺陷检测 F1 值的措施都可能产生未被主指标捕捉的副作用。例如过度追求可量化结果可能挤压自主性，自动化决策可能放大历史偏差，高强度干预可能增加负担。研究设计应预先定义至少一个安全性或负向结果，并与主结果同步报告。只有在收益与风险都透明时，迁移学习适配强度的净价值才可被准确判断。

从知识传播看，模型结果需要转换为不同受众能够理解的表达。研究者关心系数、置信区间和识别假设，管理者更关心资源配置与实施顺序，参与者则关心规则是否公平以及如何获得反馈。图表应服务于这些问题：机制图解释为什么，趋势图解释何时出现差异，回归表说明不确定性。避免只展示复杂算法而不说明其决策含义。

从累积效应看，特征可分性可能在多个周期中形成路径依赖。早期小幅改善如果得到及时强化，会提高后续参与概率并扩大缺陷检测 F1 值差异；早期失败若没有修正，也可能导致退出或抵触。因而真实项目应设置短周期复盘点，根据过程证据调整迁移学习适配强度，而不是等到终点一次性评价。动态监测还能帮助区分暂时波动与结构性失效，提高干预的可持续性。

8 结论

基于合成样本与既定参数，本文构建并检验了迁移学习适配强度—特征可分性—缺陷检测 F1 值的作用模型。结果一致显示，迁移学习适配强度具有正向主效应，特征可分性承担部分中介作用，样本规模则改变核心路径的强弱。多种变量口径、样本边界和估计方法下结论方向保持一致，说明论文内部的理论假设、模型表达、图表结果与讨论解释具有逻辑一致性。

从研究贡献看，本文围绕明确问题提出可检验假设，结合模型、公式、图表和稳健性分析形成完整证据链。后续研究应进一步使用可验证数据复核这些关系，并把因果识别、伦理审查和开放材料纳入研究流程。本文结论仍需在多来源、纵向或实验数据中进一步检验。

参考文献

- [1] Ковальський, Коваль, Панчак. МЕТОДИ FEW-SHOT LEARNING ТА TRANSFER LEARNING ДЕФЕКТУВАННЯ ДЕРЕВ'ЯНИХ ВИРОБІВ В УМОВАХ ОБМЕЖЕНОЇ КІЛЬКОСТІ ДАНИХ[J]. Наука і техніка сьогодні, 2026, (3(57)). DOI: 10.52058/2786-6025-2026-3(57)-2502-2517.
- [2] Min Y, Wang Z, Liu Y, et al.. FS-RSDD: Few-Shot Rail Surface Defect Detection with Prototype Learning[J]. Sensors, 2023, 23(18): 7894. DOI: 10.3390/s23187894.

- [3] Chen Z, Hu B, Zuo W. Cross-Object Transfer Learning-Based Few-Shot Surface Defect Detection of Lithium Batteries[J]. International Journal of Intelligent Systems, 2025, 2025(1). DOI: 10.1155/int/4904188.
- [4] Lv Q, Song Y. Few-shot Learning Combine Attention Mechanism-Based Defect Detection in Bar Surface[J]. ISIJ International, 2019, 59(6): 1089-1097. DOI: 10.2355/isijinternational.isijint-2018-722.
- [5] Liu X, Zhang Y, Wang Q, et al.. Attention-guided few-shot learning for metal surface defect classification[J]. Multimedia Systems, 2025, 31(6). DOI: 10.1007/s00530-025-02030-x.
- [6] Deng Y, Song Y. Few-Shot Steel Plate Surface Defect Detection with Multi-Relation Aggregation and Adaptive Support Learning[J]. ISIJ International, 2023, 63(10): 1727-1737. DOI: 10.2355/isijinternational.isijint-2023-118.
- [7] Eren ME. Physics-Guided Self-Supervised Few-Shot Learning for Ultrasonic Defect Detection in Concrete Structures[J]. Buildings, 2025, 15(23): 4227. DOI: 10.3390/buildings15234227.
- [8] Liu Y, Zhang G, Li X. Hint-DETR: A Transfer Learning Model Based on DETR for Few-Shot Defect Detection[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-11. DOI: 10.1109/tim.2025.3568997.
- [9] Li H, Li L, Wang H. Defect Detection for Wear Debris Based on Few-Shot Contrastive Learning[J]. Applied Sciences, 2022, 12(23): 11893. DOI: 10.3390/app122311893.
- [10] Su H, Qiu G. Few Shot Transfer Active Learning for Logo Detection in Sports Video[J]. SSRN Electronic Journal, 2022. DOI: 10.2139/ssrn.4084601.