

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-30

论文引用格式: Yang Yang, Wu Yiquan. Research progress on the robustness of multimodal surface defect detection systems[J/OL]. Journal of Image and Graphics, XXXX: 1-30. DOI: 10.11834/jig.260249. (杨洋, 吴一全. 多模态表面缺陷检测系统鲁棒性研究进展[J/OL]. 中国图象图形学报, XXXX: 1-30. DOI: 10.11834/jig.260249.) [DOI: 10.11834/jig.260249]

多模态表面缺陷检测系统鲁棒性研究进展

杨洋, 吴一全*

南京航空航天大学电子信息工程学院, 南京 211106

摘要: 多模态表面缺陷检测系统通过整合各类多模态数据, 能够有效突破单一模态的信息局限, 实现检测精度与泛化能力的显著提升, 但同时也引入了多模态数据噪声、模态质量退化、以及多模态对齐与融合偏差等鲁棒性扰动因素, 对多模态融合的可靠性提出了严峻挑战。本文聚焦于多模态表面缺陷检测系统鲁棒性这一核心研究对象, 围绕表面缺陷检测场景下的鲁棒性定义与评价、关键影响因素与典型鲁棒性挑战, 以及相应的鲁棒性优化路径, 系统梳理了近三年来国内外相关研究进展。首先, 阐述了多模态表面缺陷检测系统的架构及系统鲁棒性定义, 并进一步从外部环境工况与硬件设备、内部系统交互与数据处理两个维度对多模态表面缺陷检测系统的鲁棒性影响因素进行了分析。其次, 分别针对多模态表面缺陷检测系统的感知层鲁棒性、交互层鲁棒性以及大模型驱动的多模态系统鲁棒性等三个方面, 重点探讨了相关鲁棒性问题、挑战及制约因素, 并深入总结了对应的鲁棒性增强、应对策略及典型应用案例。随后, 围绕表面缺陷检测领域多模态鲁棒性研究的数据集与评价体系, 系统剖析了相关公开数据集的特性, 以及鲁棒性评价的核心参数与量化方法。最后, 结合当前研究现状与实际应用需求, 对面向表面缺陷检测领域的多模态鲁棒性技术的未来发展趋势进行了展望。

关键词: 表面缺陷检测; 多模态系统鲁棒性; 鲁棒性影响因素; 系统感知层鲁棒性; 系统交互层鲁棒性; 大模型鲁棒性

Research progress on the robustness of multimodal surface defect detection systems

Yang Yang, Wu Yiquan*

College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract: With the continuous advancement of industrial automation and intelligent manufacturing, multimodal surface defect detection systems have shown significant advantages in overcoming the information limitations of single-modality inspection by integrating heterogeneous data sources such as visible images, three-dimensional point clouds, depth data, infrared data, and textual information. Such integration enables substantial improvements in detection accuracy and generalization capability. However, while multimodal systems benefit from complementary information across modalities, they also introduce robustness-related disturbances, including multimodal data noise, modality quality degradation, modality missingness, and cross-modal alignment and fusion biases. These factors impose higher requirements on the stability and reliability of multimodal perception, interaction, and decision-making processes. Focusing on the robustness of multimodal

收稿日期: 2026-05-04; 修回日期: 2026-07-24

* 通信作者: 吴一全 nuuaimage@163.com

基金项目: 国家自然科学基金(61573183)项目资助

Supported by: National Natural Science Foundation of China (61573183)

© 中国图象图形学报版权所有

surface defect detection systems, this paper systematically reviews recent domestic and international research progress over the past three years in terms of robustness definition and evaluation, key influencing factors, typical robustness challenges, and corresponding optimization strategies in surface defect detection scenarios. This paper first analyzes the overall architecture of multimodal surface defect detection systems, summarizes their key technical components and core challenges, and further clarifies the connotation of robustness in such systems. Particular emphasis is placed on the ability of multimodal surface defect detection systems to maintain stable performance and reliable decisions under dynamic environments, uncertain inputs, and complex industrial conditions. On this basis, the key factors affecting system robustness are summarized from two perspectives: external operating conditions and hardware devices, and internal system interaction and data processing. External factors mainly include device calibration errors, sensor jitter, variations in UAV posture, changes in imaging angles, weather disturbances such as rain, snow, wind, and fog, production-line vibration, oil stains, water marks, surface reflection, lens contamination, and complex background interference. Internal factors are mainly related to the algorithmic mechanisms of multimodal models, including data training and adaptation mechanisms, cross-modal alignment methods, multimodal fusion mechanisms, and potential information conflicts among modalities. Based on the above analysis, this paper further investigates typical robustness problems such as multimodal data noise, modality missingness, and cross-modal interaction bias from three dimensions: perception-layer robustness, interaction-layer robustness, and robustness of large-model-driven multimodal surface defect detection systems. The technical routes and optimization directions of relevant studies are reviewed to provide a systematic reference for future research on robustness enhancement, performance evaluation, and engineering deployment of multimodal surface defect detection systems. For perception-layer robustness, this paper focuses on the system perception degradation caused by multimodal data noise and modality missingness. Data noise handling mainly concerns stable perception under data quality degradation, whereas modality missingness handling emphasizes reliable inference under incomplete modality inputs. Together, they constitute the core concerns of perception-layer robustness in multimodal surface defect detection systems. Regarding multimodal data noise, existing studies are categorized into two technical routes: denoising and noise-resistant learning. The denoising route focuses on suppressing or filtering noise at the data acquisition and preprocessing stages, thereby reducing the negative impact of noise on subsequent feature extraction, cross-modal alignment, and fusion decision-making. Representative methods include collaborative quality control of multisource acquisition, modality-specific denoising preprocessing, and coordinated normalization of multimodal data. The noise-resistant learning route aims to enhance the feature representation and task adaptation capability of models under noise perturbations, enabling models to capture defect features stably even when data quality deteriorates. Related strategies include joint noise-resistant learning based on multimodal complementarity, noise-aware data augmentation, modality feature enhancement, and noise disentanglement. For modality missingness, this paper further summarizes strategies such as modality-independent branch decoupling, missing-modality simulation training, multimodal knowledge distillation, and adaptive missing-modality adaptation, which improve inference stability and detection reliability under incomplete modality conditions. For interaction-layer robustness, this paper analyzes robustness challenges caused by cross-modal alignment instability and fusion bias from the perspectives of cross-modal alignment and cross-modal fusion. For cross-modal alignment instability, two major categories of strategies are discussed: feature alignment and spatial alignment. On the one hand, to address semantic shifts and detailed misalignments that may occur during feature alignment, deep networks can be used to extract and integrate high-level semantic information, thereby constructing interaction mechanisms based on feature content consistency. This helps alleviate the influence of shallow-level detail misalignment on multimodal collaborative representation in deep feature spaces. On the other hand, spatial alignment problems usually have more explicit physical causes. Sensor calibration errors, device vibration, and non-rigid deformation of measured surfaces may lead to spatial offsets or local misalignments among multimodal data, thereby disrupting the establishment of corresponding relationships in defect regions. To address this problem, existing studies mainly improve pixel-level or region-level cross-modal spatial alignment robustness through fault-tolerant alignment mechanisms, adaptive alignment, hierarchical cascaded alignment, and geometric constraints. For cross-modal fusion bias, this paper summarizes fusion-layer robustness enhancement mechanisms from the perspectives of adaptive fusion, bidirectional mapping constrained fusion, shared intermediate-space-driven cross-modal fusion, and other methods for suppressing fusion bias.

Regarding the robustness of large-model-driven multimodal surface defect detection systems, this paper first analyzes the basic architecture, technical characteristics, and potential advantages of large-model-driven detection systems. It highlights their value in cross-modal semantic understanding, defect semantic description, interactive analysis, and system generalization. Subsequently, the robustness constraints faced by large-model-driven systems are summarized, including insufficient modality compatibility, generative hallucination, sensitivity to prompt inputs, implicit dependence on general pretrained knowledge, and imbalance in fine-tuning. On this basis, this paper further summarizes major optimization routes for enhancing the robustness of large-model-driven systems, including modality adaptation optimization, reasoning enhancement and knowledge integration, cross-modal interaction balance optimization, multi-expert collaborative decision optimization, and structured generation constraint mechanisms. By incorporating representative application cases such as photovoltaic inspection, wafer defect detection, and power equipment inspection, this paper further analyzes the robustness enhancement effects and application value of these optimization routes in complex industrial scenarios. In addition, this paper systematically examines datasets and evaluation systems for multimodal robustness research in surface defect detection. It analyzes the modality composition, scenario characteristics, and robustness evaluation value of relevant public datasets, and further discusses core indicators and quantitative methods for robustness assessment. Finally, considering current research progress and practical engineering requirements, this paper provides an outlook on future development trends of multimodal robustness technologies for surface defect detection from the perspectives of noise-adaptive perception and robustness learning, robust detection mechanisms under the normalization of modality missingness, lightweight collaborative development driven by system robustness, and robust collaborative detection enabled by multimodal large models.

Key words: surface defect detection; multimodal system robustness; robustness influencing factors; system perception layer robustness; system interaction layer robustness; large model robustness

论文引用格式: DOI: 10.11834/jig.260249

0 引言

表面缺陷检测作为产品质量控制的核心环节,直接关系到产品经济效益与产品应用安全,其检测系统的稳定运行与可靠性是复杂缺陷检测场景下质量把控的关键前提。尽管多模态融合技术通过集成图像、点云及文本等异构数据,显著提升了表面缺陷检测场景下的缺陷识别精度,为表面缺陷检测的智能化发展提供了关键支撑(董禹彤等, 2026; 何毓芬等, 2025)。然而,表面缺陷检测场景下的非理想工况,如光照突变与波动、粉尘油污遮挡、设备振动干扰,以及多模态系统特有的异构数据冲突、模态缺失、多模态对齐偏移等问题,易导致多模态缺陷检测系统性能波动。上述问题所引发的多模态系统鲁棒性不足的问题,已成为制约多模态表面缺陷检测系统从实验室理想场景走向规模化工程应用的核心瓶颈。因此,系统梳理面向表面缺陷检测系统的多模态鲁棒性技术研究进展,明确其核心挑战与解决方案,不仅能弥补该领域缺乏系统性梳理的短板,为后续理论创新与算法优化提供清晰方向;而且更能为

多模态表面缺陷检测系统的鲁棒性设计与工程化部署提供方法论指导,从而推动高可靠、抗扰动的智能化多模态表面缺陷检测系统落地。

国内外学者基于多模态技术并围绕表面缺陷检测任务进行了大量研究,然而涉及该领域的相关研究方法或综述多聚焦于多模态系统架构优化与检测精度的提升,对鲁棒性的系统性探索仍显零散,缺乏对模态扰动类型、鲁棒性增强方法、评估标准等鲁棒性机制的系统性整合分析。如表1所示,从研究对象、模态类型、鲁棒性问题、数据集与评价指标、大模型覆盖情况以及与本文的主要区别等方面,对已有相关综述与本文进行了横向比较。

本文聚焦于多模态表面缺陷检测系统中的鲁棒性研究进展,以多模态表面缺陷检测系统的鲁棒性增强与保障为核心目标,重点强调了多模态检测系统在面对非理想输入、模态异常及多模态交互失稳等扰动时,仍能维持检测性能稳定与决策可靠的能力。围绕这一核心问题,本文从多模态表面缺陷检测系统架构与系统鲁棒性定义出发,对影响多模态检测系统稳定运行的关键因素进行了系统归纳,明确了外部环境、硬件设备及内部系统交互等维度对鲁棒性的作用机制。在此基础上,围绕多模态数据

噪声、模态缺失、多模态交互偏差等典型鲁棒性问题与挑战,分别从多模态表面缺陷检测系统感知层和交互层以及大模型驱动的系统鲁棒性等三个维度,重点探讨了国内外相关研究的技术路径、优化方向及应用案例,为后续针对性开展多模态表面缺陷检测系统鲁棒性(后续简称多模态缺陷检测鲁棒性)增强研究提供了清晰的研究脉络与思路借鉴。本文的主要架构如图1所示。

1 多模态表面缺陷检测系统及其鲁棒

性影响因素分析

1.1 多模态表面缺陷检测系统架构分析

多模态表面缺陷检测系统是指在统一检测框架下,通过协同利用来自不同传感器的异构数据,并结合多模态信息处理技术,实现对物体表面缺陷的高精度、高可靠性检测的综合性检测系统。其核心在于借助多源信息的互补性,弥补单一模态在复杂工况下感知能力的不足,使模型能够在复杂纹理背景、小目标、运动目标模糊等条件下实现更加稳定的缺陷识别、分类(白艳峰等, 2024)及量化分析。如

表1 现有相关综述多模态鲁棒性问题研究对比

Table 1 Comparative analysis of robustness issues in existing multimodal-related surveys

文献	研究对象	主要模态	鲁棒性问题	数据集与评价指标	大模型覆盖情况	与本文的区别
何毓芬等(2025)	工业缺陷检测大模型	图像、文本、视频、音频等	涉及数据质量、高可靠性、部署成本和统一测评标准等问题	主要支撑大模型研究与性能评估	核心覆盖	侧重工业缺陷检测大模型的发展历程、技术、应用和挑战,未专门讨论鲁棒性问题
Wang 等(2026e)	多模态工业诊断	1D时序/文本、2D、3D等多源工业数据	讨论多模态采集成本、标准化基准不足、融合方法泛化性与可解释性等	聚焦多模态工业诊断数据集与评价	不是核心内容	面向工业诊断大领域,不聚焦表面缺陷检测,也未专门讨论系统鲁棒性问题
Wu 等(2026c)	仅面向模态缺失场景	视觉、文本、音频	重点关注模态质量退化、不完整输入等模态缺失专有场景鲁棒性	聚焦缺失模态相关方法、应用场景和数据集	不是核心内容	聚焦模态缺失问题,涉及跨多个通用领域,未覆盖噪声、对齐、融合和大模型等问题
本文	多模态表面缺陷检测系统鲁棒性	RGB、深度/点云、红外、文本	重点覆盖噪声、模态缺失、对齐、融合及大模型等多方面的鲁棒性问题	面向表面缺陷检测领域,聚焦鲁棒性评价	独立章节重点讨论	以系统鲁棒性为主线,按照感知层、交互层和大模型三个层面梳理鲁棒性相关理论

图2所示是多模态表面缺陷检测系统架构图。从多模态数据构成来看,典型的多模态表面缺陷检测系统通常以视觉成像数据为核心模态,并结合三维点云、深度信息、红外数据以及文本语义等形成多源数据组合。其中,RGB图像作为视觉感知基础,用于捕捉颜色、纹理等表面信息;点云或深度数据提供表面三维缺陷与几何结构信息(冯明涛等, 2025);热红外数据则用于表征物体表面温度分布差异;文本模态则以非结构化形式注入工艺参数、缺陷语义等先验知识,为检测结果解释与工艺追溯提供支持。在关键技术路径上,多模态缺陷检测围绕多模态数据预处理、特征提取与强化、多模态对齐以及多模态融合等技术,形成了全流程的多源感知协同

检测方法,来实现稳健的多模态表面缺陷检测任务。然而,尽管多模态技术显著提升了表面缺陷检测性能,但面向实际应用时仍面临严峻的工程化挑战,尤其是鲁棒性问题,已成为多模态表面缺陷检测系统实际部署的关键挑战(Jiang 等, 2025b)。

1.2 多模态表面缺陷检测系统鲁棒性定义

对于多模态表面缺陷检测系统而言,其鲁棒性既延续了通用系统应对输入与环境等要素变化的系统鲁棒性基本特质(Tocchetti 等, 2025; McKinzie 等, 2023; Drenkow 等, 2024),又因多模态融合特性与复杂检测场景的独特需求及约束,展现出更具场景针对性的鲁棒性特点。作为整合多种感知模态的综合检测系统,其面临的输入与环境等要素变

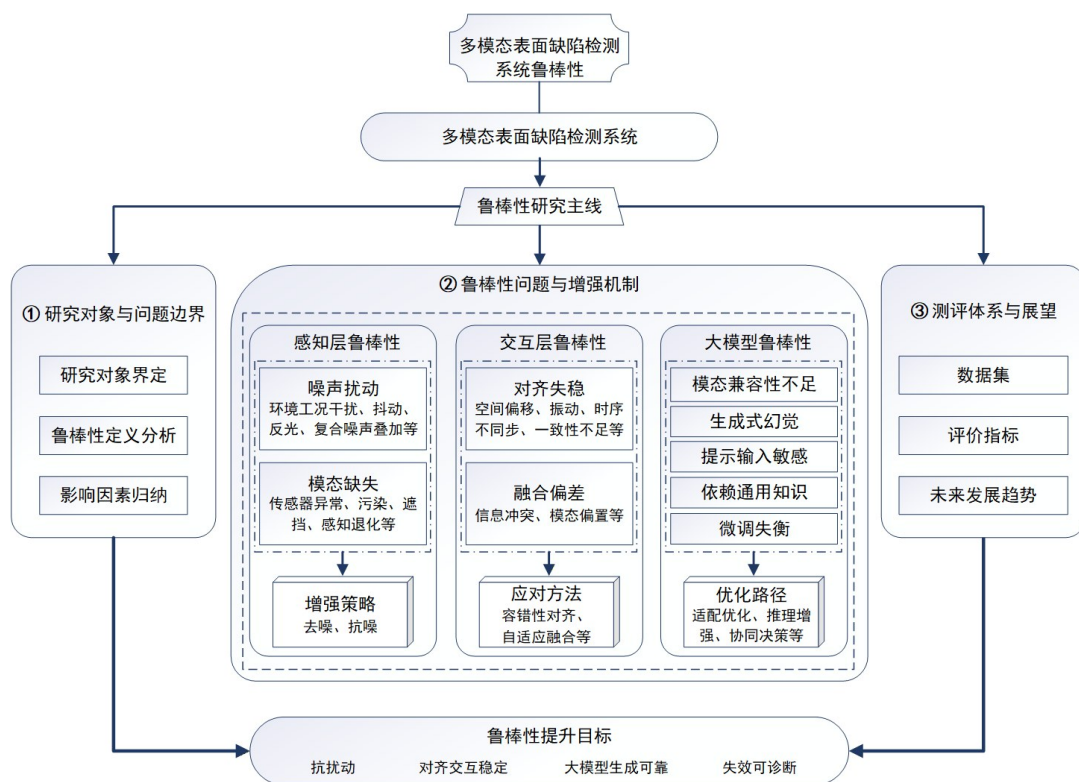


图1 多模态缺陷检测鲁棒性综述框架

Fig. 1 Review framework of robustness in multimodal surface defect detection systems

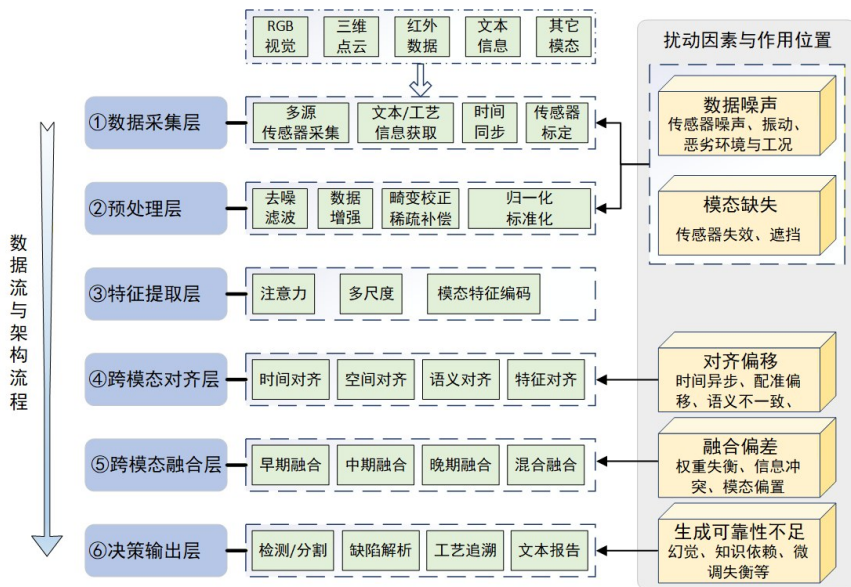


图2 多模态表面缺陷检测系统架构

Fig. 2 Architecture of multimodal surface defect detection systems

化不仅包括如噪声、遮挡等单一模态的数据质量退化,更涉及多模态间的信息异构、关联错位等特有问题。因此,结合通用系统鲁棒性定义、多模态表面缺陷检测系统的独特场景需求的综合分析,多模态表面缺陷检测系统的鲁棒性,特指其在表面缺陷检测

非理想环境下,面对单一或多种模态的输入噪声、数据缺失、模态冲突,以及光照、油污等环境因素扰动时,通过高效的去噪方法、多模态对齐及动态融合等策略,仍能稳定、准确的完成缺陷识别、定位与分类任务的能力。其核心在于依托多模态协同优势来抵

御单一模态的性能局限,在高噪声、动态变化的表面缺陷检测场景中仍能维持高检测率与低误报率(Wang等, 2023b; Jiang等, 2025b)。

在上述定义基础上,可将多模态缺陷检测鲁棒性进一步抽象为扰动集合约束下的性能保持问题。设模态输入为 $\mathcal{M} = \{M_1, M_2, \dots, M_N\}$, 其中 M_n 表示第 n 类模态输入, 如 RGB、深度/点云、红外、文本等, N 为模态总数。真实检测目标或标注记为 R , 模型预测结果记为 $\hat{R}$ 。多模态检测模型记为:

$$\hat{R} = f_{\theta}(\mathcal{M}) \quad (1)$$

式1中 θ 包括跨模态对齐、融合决策等模块中的可学习参数。而非理想工况下的未知扰动可统一表示为:

$$\mathcal{M}_{\delta} = \mathcal{T}_{\delta}(\mathcal{M}) \quad (2)$$

式2中 δ 可包含数据噪声、模态缺失、对齐偏移、融合偏差等因素。则多模态缺陷检测鲁棒性可定义为: 在给定扰动集合 Δ 内, 模型在各类扰动输入下的检测性能退化量仍保持在允许范围内, 即:

$$\Delta\Psi_{\delta} = \Psi(f_{\theta}, \mathcal{M}, R) - \Psi(f_{\theta}, \mathcal{T}_{\delta}(\mathcal{M}), R) \leq \eta \quad (3)$$

$\forall \delta \in \Delta$, 式3中 Ψ 表示性能评价函数, $\Delta\Psi_{\delta}$ 表示性能退化量, η 表示扰动条件下允许的性能退化上界, 且 $\eta \geq 0$ 。该数学模型表明, 多模态缺陷检测鲁棒性核心在于应对输入质量退化、模态结构不完整和跨模态交互受扰条件下的性能保持能力。

1.3 多模态表面缺陷检测系统鲁棒性影响因素

多模态缺陷检测鲁棒性, 本质是系统在如环境干扰、设备差异、数据不完整等复杂场景下仍保持稳定检测的能力, 其影响因素涉及模态特性、系统交互机制、环境条件等多个层面。

1) 外部干扰因素

外部干扰因素是影响多模态缺陷检测鲁棒性的重要因素之一, 也是噪声干扰的核心来源, 主要集中在传感器设备硬件运行情况、外部环境及工况条件、复杂纹理与背景干扰等三方面。

(1) 传感器设备硬件影响因素

多模态表面缺陷检测系统中, 传感器硬件的误差与动态扰动是常见的干扰源。例如, 传感器标定误差、相机视角变化、设备抖动等可能导致 RGB 图像与 3D 点云等模态之间的空间对齐失配(Wu等, 2026a), 甚至引起几十像素的平移与旋转误差(Hoang等, 2026)。Lai等人(2025)分析指出无人机

飞行姿态、相机拍摄角度的偏差具有随机性, 无法精准控制, 导致图像出现几何畸变、尺寸偏差等。又如 Almkhtar 等人(2026)分析认为多模态之间的像素级对应关系会受到传感器标定误差以及采集设备位姿偏移等因素影响, 易使得配准过程面临几何不确定性。此外, 由于传感器故障、遮挡、通信中断等因素, 可能会导致某些模态的信息缺失。例如, 3D 点云可能因设备故障丢失, RGB 图像可能因采集失败无法获得(Miao等, 2025; Lin等, 2025)。

(2) 外部环境及工况条件

光照变化、色温及室外检测场景下的风力、天气、云雾等环境因素在实际部署中无法完全控制(Han等, 2025; Zou等, 2025a), 且具有不规则的随机性; 此外, 在部分工况恶劣的工业检测现场, 产线机械振动(Zhao等, 2025a)、车间粉尘、水渍、油污(Wang等, 2023a)等工况因素, 也会对传感器采集工作的稳定性、测量精度、以及数据质量产生随机扰动。如 Xu 等人(2025)分析认为工业现场环境复杂且多变, 表面反光会造成 RGB 图像局部过曝并干扰深度传感器, 镜头污染也会遮挡关键表面特征, 从而加剧图像退化、点云稀疏等数据质量问题。

(3) 复杂背景干扰

对于表面具有随机纹理特征的物体表面缺陷检测, 如瓷砖等, 表面可能会因生产排产与釉面工艺的不同, 往往会出现训练数据中未知的凹凸性花纹纹理(Karimi等, 2024); 此外, 面对输电线路(吴华等, 2025)、光伏设备巡检场景, 复杂不确定性的自然背景也会对模型鲁棒性产生较大的挑战。

2) 内部系统影响因素

(1) 模态间的异构特征差异

RGB 图像与深度、热成像及文本信息等模态之间存在天然的数据特征差异。这种异构性使得不同模态的特征在对齐或融合过程中可能出现不匹配或冗余问题(Wu等, 2026a), 易导致模型在训练时倾向于信息更丰富或更易学习的模态, 造成模态偏差。

(2) 训练数据与模型的适配问题

在训练过程中, 数据集的质量与分布对系统鲁棒性至关重要。在实际检测任务中, 训练数据往往与实际生产线数据存在域偏移问题(Ha等, 2026)。例如, 训练数据可能是在实验室环境下采集的, 而实际生产中的光照变化、背景差异和缺陷类型可能与实验室场景存在差异, 从而导致模型适应性不足。

此外,模型在面对产线数据的逐步变化时,可能无法及时适应新环境,从而影响检测的稳定性。

(3) 系统对齐与融合策略

在多模态信息融合的过程中,如何合理分配各模态的权重和重要性是系统鲁棒性的关键。若融合策略设计不当,可能引发语义失配并扰乱多模态信息传递,进而影响系统稳定性(Jiang等, 2026a)。同时,融合过程中的信息冲突与不一致性,也可能会导致最终的决策不稳定,进而影响缺陷检测的准确性和鲁棒性(Chen等, 2026a; Tang等, 2026)。

如表2所示,总结梳理了多模态缺陷检测鲁棒性的相关影响因素。多模态缺陷检测鲁棒性受到外部扰动与内部机制的双重制约,深入理解这些影响因素及其耦合关系,是有效分析与强化多模态缺陷检测鲁棒性的理论前提。

此外,从系统链路角度看,多模态缺陷检测鲁棒性并非仅由单一数据质量问题决定,而是贯穿于多源感知、跨模态交互和高层语义决策的全流程。因此,可依据鲁棒性问题的主导发生环节、影响机制及优化路径等角度,将多模态缺陷检测鲁棒性划分为感知层鲁棒性、交互层鲁棒性和大模型驱动的系统鲁棒性三个层面。其中,感知层对应多源数据采集、预处理和特征编码阶段,重点关注噪声、遮挡、模态缺失和模态质量退化等输入端扰动对缺陷表征稳定性的影响;交互层关联跨模态对齐与融合阶段,聚焦不同模态在时间、空间、语义和特征分布上的对齐失稳,以及融合权重偏置、信息冲突等挑战;而大模型系统鲁棒性则面向多模态信息进入大模型后的语义理解、推理生成和结果解释过程。由此可见,感知层、交互层以及大模型驱动的系统鲁棒性分别指向输入感知、跨模态协同和语义决策等多个关键环节,在问题定义、作用位置和优化方法上具有明确差异,而且相关扰动并不会局限于单一环节,而是会沿着多模态数据获取、特征表达、跨模态交互、融合决策与结果生成等过程逐步传递,并共同影响系统整体鲁棒性。

2 多模态表面缺陷检测系统感知层鲁棒性问题与增强

在多模态表面缺陷检测系统中,由于外部环境干扰、传感器自身物理故障及设备部署偏差等因素,

给系统感知层带来数据噪声、模态缺失等诸多不稳定问题与影响。其中,数据噪声应对聚焦于数据质量退化下的稳定感知,模态缺失应对则着眼于模态数据不完备条件下的可靠推理,二者共同构成了多模态表面缺陷检测系统感知层鲁棒性的核心内容。

2.1 数据噪声与模态缺失下的感知层鲁棒性问题

2.1.1 多模态噪声影响下的感知扰动问题

相比较单模态检测场景,多模态表面缺陷检测系统的数据噪声扰动问题则从单一模态的信噪分离,演变为应对多种异构噪声源的并发干扰及耦合效应问题(曲海成等, 2025)。一方面,复杂工业环境中可能面临多种模态噪声的复合叠加,进一步限制抗噪模型的训练效果与泛化能力;另一方面,多模态协同去噪方法与差异化模态抗噪策略则可能会进一步引发噪声的多模态传递,例如,为增强可见光成像而采用的主动照明,可能改变物体表面热分布,从而在热成像通道中产生新的噪声,这种由一个模态的去噪而可能引发另一模态噪声的现象,加剧了多模态噪声应对的复杂性。因此,多模态表面缺陷检测系统中的数据噪声鲁棒性问题,本质上并不只是噪声来源和应用场景的简单扩展,而是在复杂缺陷特征易受扰动的基础上,进一步面临异构模态噪声的并发干扰、复合叠加等多重难题,最终体现为如何在多模态输入噪声并存的环境中,确保多模态模型的稳定特征表达与可靠检测性能。

2.1.2 模态缺失场景下的感知退化问题

相比较单模态检测方案,多模态表面缺陷检测系统的核心优势在于不同模态能够分别从外观纹理、色彩图案、空间几何、温度分布及文本语义等不同维度来对缺陷进行多维特征互补描述,从而构成全维度的缺陷特征体系。因此,一旦模态缺失或数据质量退化,则会导致原有的多维协同感知基础退化为局部模态甚至单维度感知,进而削弱对复杂缺陷的稳定识别能力。尤其在复杂缺陷检测场景中,表面缺陷具有复杂纹理干扰、小目标或跨尺度分布等挑战,模态缺失或模态数据质量退化易使系统损失多模态互补与交叉检测能力。

1) 完整输入依赖下的感知能力下降

多模态表面缺陷检测系统通常是基于完整模态输入条件构建的,其系统架构、特征提取等研究,基本都是以各模态能够稳定提供有效特征信息为前提假设。若当某一模态缺失时,原有完整输入结构被

表 2 多模态缺陷检测鲁棒性影响因素

Table 2 Influencing factors of robustness in multimodal surface defect detection systems

分类	影响因素	具体表现	可能失效影响
硬件设备		设备标定误差、抖动(Almukhtar等, 2026; Wu等, 2026a)	采集位置偏移、图像模糊,导致缺陷定位偏移、边界识别不稳定和跨模态配准误差
		无人机姿态、拍摄角度偏差(Lai等, 2025)	
		设备故障异常(Miao等, 2025; Lin等, 2025)	
外部干扰因素	环境工况	雨雪风、天气(Han等, 2025; Zou等, 2025a)	局部遮挡、纹理对比度下降或噪声较大,导致小目标漏检、伪缺陷误检和特征提取不稳定
		产线机械振动(Zhao等, 2025a);粉尘、水渍、油污(Wang等, 2023a)	
		表面反光、镜头污染(Xu等, 2025)	
	复杂背景	复杂纹理与花纹干扰(Karimi等, 2024)	削弱缺陷与背景之间的差异,导致背景纹理误检或低对比度缺陷被背景信息淹没
		复杂自然背景(吴华等, 2025)	
内部影响因素	多模态异构差异	模态之间存在天然的结构、数据特征差异(Wu等, 2026a)	造成跨模态特征尺度不一致、信息表达不匹配和互补信息利用不足等问题
	数据训练与适配	训练数据与实际生产线数据存在域偏移问题、系统的数据分布适应性不足(Ha等, 2026)	降低模型对真实产线扰动和复杂工况的泛化能力,导致检测性能波动
	对齐与融合	对齐或融合策略设计不当(Jiang等, 2026a)	引发跨模态特征错配、融合权重偏置和决策结果不稳定
		交互过程信息冲突与不一致(Chen等, 2026a)	

打破,系统由面向完整模态输入信息状态转向不完整的模态缺失运行状态。这种异常变化会削弱原有
多模态架构的适用性,使特征提取与传播过程偏离
原有预期目标,严重影响多模态感知层鲁棒性
(Maheshwari等, 2024)。而且不同模态在物理表
征、信息维度及与缺陷类型的关联强度上存在本质
差异,其缺失状态对系统鲁棒性的影响方式与程度
也各不相同。因此,模态缺失场景下,系统损失的并
非是可被其他模态替代的冗余信息,而是某一类特
定性质的关键缺陷特征缺失,从而进一步影响多模
态系统对缺陷特征的感知精度与范围等。

2) 模态缺失下的感知偏置问题

多模态系统的优势不仅在于信息源广泛,更在
于不同模态之间能够形成互补的感知机制。当关键
模态缺失后,这种互补关系则被明显削弱,系统虽仍
可依赖剩余模态继续运行,但原有的多模态协同互
补平衡机制已被削弱,系统推理可能会偏向现有无

损模态,造成模态偏差问题(Woo等, 2023)。

3) 模态缺失问题的传导与累积

由于多模态检测系统的信息传导特性,感知层
一旦在输入阶段形成信息缺失问题,可能会进一步
在后续特征处理与对齐融合环节中演化为对应关系
不稳、融合权重失衡以及决策输出偏差等问题,甚至
在信息传导链路中不断累积与放大(Wu等, 2026c)。

2.2 面向数据噪声扰动的感知层鲁棒性增强

围绕多模态表面缺陷检测系统的数据噪声应对
鲁棒性增强,现有方法主要聚焦于去噪与抗噪两大
核心技术路径。其中,去噪技术路径以数据采集源
头及预处理阶段抑制或过滤噪声干扰为出发点,核
心价值在于将噪声控制在数据生成与前期处理环
节,减少噪声对后续多模态特征提取、对齐及融合
的负面影响。而抗噪技术路径则以提升模型自身对
噪声的适应与抵抗能力为核心导向,其目标在于使模

型能够在噪声扰动下,依然能够精准捕捉缺陷本质特征、稳定完成检测任务。

2.2.1 数据去噪思想驱动下的鲁棒性增强策略

1) 面向模态差异的数据去噪预处理

在多模态表面缺陷检测系统中,不同模态由于各传感器原理与物理采样过程的差异,往往呈现出不同的噪声形态与统计特性,这种噪声分布的差异性决定了单一滤噪策略难以同时适配所有模态。针对各模态噪声分布特征设计专属的数据预处理与去噪算法,已成为多模态数据质量控制的基础环节。如针对可见光RGB图像,Wang(2025b)采用窗口大小自适应调整的中值滤波来去除激光飞溅噪声;Ren等人(2026)通过图像增强、降噪及基于特征匹配或图像配准的方法,来提高图像清晰度并抑制光照波动、成像模糊等噪声;Chen等人(2026b)结合惯性测量单元(inertial measurement unit, IMU)的运动先验进行定向去模糊补偿,从而提升多模态输入质量与后续检测稳定性;Rahmati等人(2025)则通过亮度调整、几何增强等训练阶段的数据扰动方式来提升模型对成像退化与光照波动的适应能力。对于红外数据,Wang(2025b)利用三层小波分解并结合软阈值处理,从而有效分离热信号与环境热噪声。针对点云或深度数据,Ren等人(2026)通过离群点移除、下采样、泊松表面重建等操作来消除扫描伪影、稀疏空洞等噪声;Chen等人(2026b)则对深度模态采用高斯滤波去噪并构造曲率特征,从而压制高反光表面带来的噪声干扰。对于文本数据去噪,Ren等人(2026)针对操作日志、维护记录等文本模态,通过正则解析、领域词典标准化等方法来解决语义模糊、标注偏差等语义噪声。而对于其他模态数据去噪,如Zhang等人(2026a)针对动态电阻信号采集过程中的随机微弱噪声,同样采用小波去噪算法进行针对性预处理,通过软阈值函数来滤除噪声,并保留了信号的核心物理特征;而Rahmati等人(2025)则针对声学模态,先将原始信号经短时傅里叶变换将时域信号转换至频域,再结合频谱自适应阈值来抑制瞬态噪声干扰。

2) 多模态数据协同噪声滤除方法

Wang等人(2025a)提出一种基于模态内对比与多模态加权的协同滤噪机制,通过分别比较RGB与3D点云特征和参考样本的相似性,同时结合双模态融合异常分数,利用多模态信息的互补判别优势来

协同滤除噪声样本。在表面缺陷检测场景中,不同模态往往受到不同类型环境扰动的影响,使得单模态独立数据预处理难以处理模态间耦合的复杂噪声干扰。多模态数据协同噪声滤除,依托多模态特征关联与协同优势,通过多模态数据间建立统计关联与信息互补机制,可实现噪声的协同抑制。又如Zhou等人(2025)通过将RGB特征与深度特征进行逐元素交互计算,以形成初步降噪后的深度特征,从而利用RGB图像的质量优势对深度数据进行引导式去噪,使得深度图像的噪声影响得到有效抑制。

2.2.2 数据抗噪思想驱动下的鲁棒性增强策略

1) 多模态互补协同驱动下的抗噪方法

多模态数据的互补特性为噪声应对提供了单模态系统所不具备的理论优势,多模态互补协同驱动的抗噪方法,充分利用不同模态间的信息冗余与互补特性,当某一模态受噪声严重污染时,依托其他低噪模态的可靠信息来实现缺陷特征的补偿与重建,可显著提升多模态检测系统的容错能力与噪声鲁棒性。Wang等人(2026c)提出了基于双模态互补融合的抗噪策略,通过利用热红外数据对温度异常特征的可靠提取,来弥补可见光模态在电弧火花及其飞溅下的强干扰问题,同时借助可见光模态对熔池形貌与空间细节的高分辨能力来补充热红外模态分辨率不足的问题,进而在动态融合过程中抑制单模态噪声影响并提升检测鲁棒性。Wang等人(2026d)针对复杂工业场景下光照波动、丝印、划痕及背景干扰等问题,采用RGB图像与三维高度信息的概率互补融合思路,分别构建RGB颜色概率与三维高度概率分布,并通过局部概率融合算子来联合建模两类信息,以抑制单一模态易受噪声干扰所导致的误分割传播。

2) 噪声感知的数据增强训练

Du等人(2026)和Zhang等人(2026a)在训练集中主动添加高斯、椒盐及泊松等噪声,模拟工业相机的随机噪声,构建包含不同信噪比的训练样本,使模型在学习过程中适应复杂的噪声扰动,提升对未知噪声的泛化能力。Lin等人(2025)在训练阶段引入翻转、旋转、颜色抖动与CutMix等扰动增强操作,通过模拟光照变化、局部遮挡及视角扰动等噪声产生条件,扩展训练样本分布并提升模型对复杂检测场景扰动下的适应能力。Zou(2025b)针对红外数据在采集与传输中易受高斯噪声等随机图像退化扰动的

问题,在训练阶段对每批次中的部分红外数据随机施加高斯噪声、模糊、亮度变化、低分辨率等扰动,使模型学习对噪声不变的特征表达,以此强化对相关噪声的适应能力。噪声感知的数据增强训练通过在训练阶段主动模拟各类可能出现的噪声扰动,构建包含不同信噪比水平的训练样本,使模型在学习过程中逐步适应多种复杂环境。

3) 特征增强驱动的抗噪方法

在复杂工业环境中,输入数据中的噪声往往具有随机性和不可预测性,仅依赖显式噪声建模或滤波策略难以全面覆盖各种干扰模式,尤其当噪声分布呈现非结构化特征或与缺陷形态相似时,单纯的去噪方法易导致缺陷信息的过度平滑与损失。针对这一去噪风险,可通过提升缺陷相关特征在特征空间中的表达强度与判别能力,设计针对性的特征增强机制,强化与缺陷相关的结构信息,从而拉大缺陷判别性特征与噪声干扰在特征空间的区分度,间接实现抗噪效果。如Cui等人(2026)提出了一种语义重组模块(semantic recomposition module, SRM),通过字典投影、语义检索、位置归一化与特征重构等方法,来增强缺陷相关特征,拉大有效特征与无效噪声的区分度,从而间接实现噪声抑制。Hoang等人(2026)提出一种面向多模态中各模态内部特征强化的状态空间建模方法,通过在各模态内部引入Mamba模块(per-modality mamba module, PMM),来对每个阶段的特征分别进行建模,并基于四向蛇形扫描方法(quadSnake scan, QSS),将二维特征以正向、反向及对角方向的蛇形路径展开到状态空间递推过程中,从而捕获长程依赖特征与局部连续结构,强化了划痕等细长缺陷的轮廓特征表达。刘婷婷等人(2025a)针对光伏电池板缺陷检测中可见光与红外数据易受设备噪声及复杂环境干扰影响的问题,一方面通过子块正交调制模块(sub-patch orthogonal modulation, SPOM)对可见光图像构建具有方向判别力的低秩子空间,在增强显著纹理响应的同时抑制冗余噪声干扰;另一方面则引入热梯度残差网络(thermal gradient residual network, TGRN)来强化红外数据中热异常区域的结构特征。

2.2.3 数据噪声应对鲁棒性增强策略剖析

如表3所示总结了面向数据噪声扰动的感知层鲁棒性增强方法。多模态噪声应对鲁棒性增强策略虽可有效应对来自各模态内与模态间的各类噪声,

但同时也存在部分场景适用性与局限性。

1) 场景适用性。数据专属去噪适用于噪声来源、采集流程较为明确,且缺陷特征与噪声具有一定可分性的场景;而多模态数据协同噪声应对则适用于

模态间互补关系明确,具有明确空间对应或语义关联的场景,且至少存在一个模态具备较高可信度;噪声感知的数据增强训练,通过构造多样化的退化样本来提升模型对噪声扰动的容忍能力,更适用于噪声来源复杂且训练阶段能够获得较充分样本,或能够合理模拟噪声分布的检测场景;特征增强抗噪方法则更强调模型内部的特征稳健性,适用于噪声形态复杂且缺陷特征具有一定结构规律的场景。

2) 局限性与应用风险。现有多模态噪声应对方法能够在一定程度上抑制感知噪声,但其效果通常依赖滤波参数、阈值或预处理流程,对复杂表面缺陷检测场景工况变化的适应能力有限,存在相关局限性。(1) 过度去噪。去噪强度设置不当可能削弱微弱缺陷特征表达或消除小目标。(2) 协同依赖性。协同应对方法依赖较准确的模态对齐和可靠的跨模态关联,对齐误差或模态质量失衡可能引入新的噪声或错误补偿问题。(3) 训练失衡。高斯、椒盐和泊松噪声等人工扰动难以完整表征真实工业环境中的复合噪声,且噪声增强方法的设计具有较强经验性,过弱或过强的噪声增强会导致模型在训练阶段学习到错误的噪声应对能力。(4) 特征同步放大。特征增强的间接抗噪,可能在复杂背景、低对比度缺陷和小目标场景中,因噪声与缺陷在特征层面上的高相似性,而造成噪声与缺陷响应被同步放大等问题。

针对上述场景适用性与局限性,须在多模态噪声应对中注意以下几个方面:首先是控制去噪强度,避免弱缺陷信息被过度平滑,并关注去噪后不同模态之间的空间、语义一致性,避免单模态质量提升破坏多模态对齐基础;其次,针对多模态协同,应建立模态质量评估机制,动态估计噪声水平与特征可信度,并通过门控或自适应权重控制各模态的交互贡献,避免低质量模态的无效协同;此外,对于特征增强方法,则应结合缺陷形态和模态特性实施选择性增强,并引入缺陷先验等约束条件,来避免对高频纹理背景、相关噪声等进行无差别放大。

2.3 面向模态缺失应对的感知层鲁棒性增强

针对上述不同模态缺失场景及其鲁棒性问题,

表3 面向数据噪声扰动的感知层鲁棒性增强方法
Table 3 Methods for enhancing perception-level robustness against data noise perturbations

分类	增强机制	具体方法
滤噪 机制	模态差异 滤噪方法	可见光: 中值滤波(Wang, 2025b)、图像增强与配准(Ren 等, 2026)、定向去模糊补偿(Chen 等, 2026b)、数据扰动训练(Rahmati 等, 2025)
		红外: 小波分解(Wang, 2025b)
		点云: 离群点移除、泊松重建(Ren 等, 2026)
		深度: 高斯滤波(Chen 等, 2026b)
		文本: 正则解析、领域词典(Ren 等, 2026)
		声学: 频域转换与处理(Rahmati 等, 2025)
滤噪 方法	模态协同	模态内对比与多模态加权协同(Wang 等, 2025a)
	滤噪方法	图像对深度引导式去噪(Zhou 等, 2025)
抗噪 机制	模态协同	可见光与红外互补融合(Wang 等, 2026c)
	抗噪方法	可见光与三维高度概率互补融合(Wang 等, 2026d)
	数据训练 抗噪方法	训练集添加高斯、椒盐及泊松等噪声(Du 等, 2026; Zhang 等, 2026a)
		训练集引入翻转、旋转、颜色抖动与 CutMix 等扰动(Lin 等, 2025)
抗噪 方法	数据训练 抗噪方法	训练集施加模糊、亮度变化、低分辨率等扰动(Zou, 2025b)
特征增强 抗噪方法	语义重组	语义重组(Cui 等, 2026)、状态空间建模、四向蛇形扫描(Hoang 等, 2026)、子块正交调制、热梯度残差网络(刘婷婷等, 2025a)
	抗噪方法	

国内外学者从架构设计、训练机制、知识迁移及决策融合等多个层面提出了一系列针对性强化策略。

1) 模态独立分支解耦策略

通过构建并行独立的模态处理分支,使每个分支具备独立完成检测任务的能力,无需依赖其他模态的信息即可独立完成任务,从而在任意模态缺失时都能提供基础的检测能力,保障系统的基础检测性能。如 Reich 等人(2025)和 Tang 等人(2025)采用 RGB 与点云双分支独立处理并支持灵活融合,使得仅单一模态可用时仍可输出检测结果。

2) 模态缺失模拟训练

Maeda 等人(2024)通过在数据集上人工生成不同缺失率的因果变量数据,并随机置零以模拟现实中的数据不完整状态,使模型在较高缺失率条件下仍具有较强的分类鲁棒性。Hoang 等人(2025)则直接在训练中引入模态丢弃增强,配合共享骨干网络,使模型无需重新训练即可快速适应传感器故障等缺

失场景。Hsu 等人(2024)在训练阶段引入掩码学习,随机模拟光学字符识别(optical character recognition, OCR)场景下的不稳定条件,引导模型学习在缺失模态信息下的适应能力。通过在训练阶段主动且随机地丢弃部分模态输入,引导模型学习不依赖于任何单一模态的鲁棒性特征,从而提升模型对缺失模式的适应能力。

3) 多模态知识蒸馏

基于知识迁移思想,来通过构建师生模型框架,将完整多模态数据知识迁移至单模态检测任务中,使学生模型在仅输入单一模态数据时,仍能达到接近多模态融合的检测性能,从而提升学生模型在缺失模态下的检测能力。如图 3 所示,Xu 等人(2025)提出了一种基于多模态知识蒸馏师生框架,通过构建双教师-单学生模型,在训练阶段利用完整的 RGB 与点云数据训练教师模型来生成较高质量异常标注,

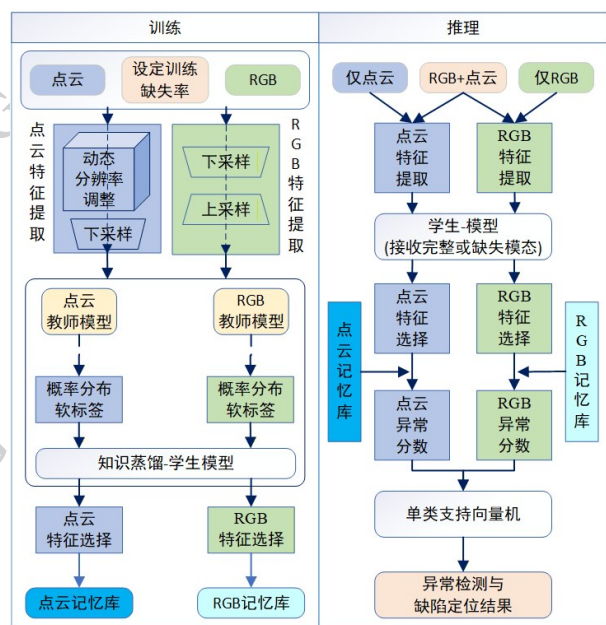


图3 基于知识蒸馏的多模态训练与模态缺失应对的推理框架

Fig. 3 A reasoning framework for multimodal training and modality missing response based on knowledge distillation

学生模型则通过结合交叉熵与KL散度(kullback leibler divergence)的蒸馏损失,在随机缺失某一模态的条件下学习多模态互补特征,从而在推理阶段即使仅接收单一模态输入也能实现鲁棒的异常检测。Lu等人(2025)针对工业异常检测中文本描述在推理时缺失的问题,设计了一个多模态融合网络作为教师模型,同时引入一个仅以图像模态为输入的学生模型,通过最小化学生模型输出与教师模型输出之间的KL散度,并结合交叉熵损失,将融合了图像与文本信息的丰富知识蒸馏至图像分支,使学生模型在仅有图像输入时仍能保持较高的检测精度,有效缓解了模态缺失带来的性能下降。

4) 模态缺失自适应适配

Lin等人(2025)基于不确定性量化方法,采用蒙特卡洛Dropout估计各模态的感知不确定性,并据此进行贝叶斯置信加权融合,利用多模态协方差建模来修正权重,这种动态融合权重适配机制通过降低缺失或退化模态的贡献度而强化高可靠模态的主导作用,提升了系统在缺失输入条件下的缺陷检测稳定性。Hsu等人(2024)引入门控函数动态加权机制,在推理阶段设计门控函数,动态评估各模态特征的置信度并自适应调整融合权重,选择性屏蔽失效模态输出,从而避免低质量模态来主导决策,实验结

果表明,在随机缺失60% OCR数据的极端条件下,F1分数相较于基线模型下降幅度较小,验证了其在模态缺失情况下的鲁棒性。Ren等人(2026)则提出基于综合置信度的动态融合机制,通过联合数据质量、缺陷敏感性与数据覆盖度等,来计算各模态综合置信度,动态确定主模态并自适应分配各模态融合权重。上述方法根据各模态输入数据的可靠性评估或置信度分数,动态调整其在融合决策中的贡献权重或实施模态切换,便于作出最优的融合或切换决策以应对模态缺失问题。

如表4所示,对比了模态缺失场景下的相关鲁棒性增强方法。现有模态缺失鲁棒性增强基本可归纳为结构冗余、训练扰动、知识迁移与动态适配等四类技术路径。其中,模态独立分支解耦通过为各模态设置独立推理路径以提升对突发模态缺失的鲁棒性,但同时也存在跨模态协同不足与模型计算量大等问题;缺失模拟训练通过模态丢弃、置零等方式进行扰动学习,以提升对随机缺失的适应能力,但其性能依赖模拟缺失与真实失效分布的一致性,否则易产生分布偏移;而知识蒸馏则利用完整模态教师模型向少模态学生模型迁移互补知识,使模型降低对完整输入的依赖,但可能存在误差传递或学生过拟合输出分布等问题。模态缺失自适应适配基

于置信度或不确定性来动态调整融合权重,以应对模态质量波动与缺失,但较为依赖可靠性估计精度,一旦不确定性估计失准,可能引入融合偏差。

3 多模态表面缺陷检测系统交互层鲁棒性挑战与应对

多模态交互层鲁棒性强调在对齐、交互、融合及协同决策过程中所面临的不确定性交互挑战。从多模态系统交互层的组成机制与问题表现来看,其鲁棒性风险主要体现为多模态对齐失稳与多模态融合偏差两类问题:对齐失稳易导致多模态间信息难以准确对应,而融合偏差则可能造成模态贡献失衡。因此,交互层鲁棒性强化的关键,在于围绕对齐失稳与融合偏差等核心瓶颈,构建能够抑制模态交互扰动与偏差、提升多模态协同质量的应对机制。

3.1 多模态对齐与融合环节的交互层鲁棒性挑战

1) 多模态对齐失稳鲁棒性挑战

在多模态表面缺陷检测系统中,不同模态之间

表 4 模态缺失场景下的鲁棒性增强方法对比

Table 4 Comparison of robustness enhancement methods for missing-modality scenarios

文献	分类	方法	贡献	局限性
Tang等(2025)	模态独立分支	采用模态分支独立处理,仅单一模态可用时仍可输出检测结果	仅图像:I-AUROC: 96.7%, P-AUROC: 97.7% 仅深度:I-AUROC: 93.3%, P-AUROC: 92.6% 模态缺失时性能仅小幅衰减	网络结构复杂臃肿,增加计算与训练成本
Maeda等(2024)	模态缺失模拟训练	人工生成不同缺失率的因果变量数据并随机置零以模拟训练	单类别下缺失率低于 80% 时精度仍超 95%,两类叠加在 80% 缺失率下仍超 90%	缺失值仅通过随机置零模拟,未覆盖工业现场其他真实复杂缺失模式
Hoang等(2025)		通过模态丢弃模拟缺失输入,结合共享骨干学习多模态互补表征	仅图像:I-AUROC: 96.5%, P-AUROC: 96.7% 仅深度:I-AUROC: 95.5%, P-AUROC: 96.0%	缺失率模拟过高会削弱学习效果
Lu等(2025)	多模态知识蒸馏	构建图文融合教师与单图像学生分支进行知识蒸馏	准确率: +2.9%、灵敏度: +3.28%、 特异度: +2.83%、F1 分数: +3.68%	知识蒸馏迁移不充分,图像分支性能提升有限
Hsu等(2024)	动态权重适配	引入门控函数来动态评估各模态置信度并自适应调整融合权重	在随机缺失 60% 下,F1 分数仍有 83.07%	模型集成方式,可能增加存储与计算资源需求
Ren等(2026)		置信度驱动的自适应权重分配与缺失模态自动剔除	可容忍部分模态缺失,提升不完整输入下的决策稳定性,且无需重新训练	置信度计算的可靠性依赖先验知识
Xu等(2025)		双教师蒸馏训练学生,在随机缺失单模态下学习多模态互补特征	缓解不完整模态导致的性能退化,提升异常检测鲁棒性	模态严重缺失下信息不足,蒸馏补偿存在上限
Lin等(2025)		基于不确定性量化方法进行贝叶斯置信加权融合	模拟传感器缺失下,F1 分数: +4.6%; 光度模态缺失下,F1 分数: +5.2%	传感器失效可能存在相关性而非假设独立

的对应关系在传感器采集、数据处理与交互过程中往往难以长期保持稳定一致,可能会在空间、时序以及语义层面影响多模态对齐鲁棒性。一方面,多模态的空间对应关系易受传感器安装偏差、标定误差、镜头畸变以及设备振动等因素影响,使不同模态对同一缺陷区域的成像位置与局部结构产生偏移;另一方面,相关模态在时间对应关系上也可能受到传感器采样频率、触发及传输延迟等时序因素干扰,而导致同一缺陷在不同模态中的采集时序不一致;此外,RGB图像、三维点云、热红外及文本等模态在表征机制上的差异,也会使同一缺陷在不同模态下呈现出不一致的几何轮廓与语义特性,尤其在噪声、遮挡及小目标等条件下,这种对齐失稳会进一步加剧。如 Almkhtar 等人(2026)分析认为当外部扰动或硬

件设备姿态发生偏移时,不同模态间原本稳定的空间对应关系可能会发生波动和失准,进而影响后续交互与对齐的可靠性。Shi 等人(2025)则指出传统多模态映射主要依赖单向静态转换方法,而非建立模态间双向可验证、可还原的特征对应关系,难以刻画复杂工业场景下多模态之间的非线性关联,从而易导致多模态之间的对应关系不充分、表示一致性不足等对齐鲁棒性挑战。多模态对齐异常的产生,可能会在特征提取、注意力分配甚至后续融合决策过程中被逐层传播和放大,并最终表现为多模态系统的误检、漏检、定位不准以及决策可靠性下降等问题。

2) 多模态融合偏差鲁棒性挑战

在多模态表面缺陷检测系统中,当多个信息源
© 中国图象图形学报版权所有

同时输入时,如果缺乏有效的筛选和仲裁机制,信息之间可能相互冲突或产生冗余,最终导致模型决策偏差并输出错误结果(Chen等, 2026a)。在表面缺陷检测场景中,多模态融合偏差鲁棒性挑战主要表现为模态偏置固化与模态偏置敏感问题。一方面,在多模态融合过程中,由于不同模态在特征强度、学习难度以及训练稳定性等方面并不一致,模型可能会对信息量丰富或特征明显的模态产生过度依赖(Zheng等, 2025),从而抑制另一模态的有效贡献,导致融合失衡与决策偏差。另一方面,若系统对外部扰动变化过于敏感,模态贡献关系则缺乏必要的稳定性,一旦外部环境、工况或局部特征发生轻微波动,就可能引起模态偏向的无规律偏移,从而使得融合结果呈现出较强的不确定性(Bi等, 2025)。

3.2 多模态对齐失稳应对方法

针对特征对齐失稳风险,可利用深层网络对高层语义信息的提取与整合能力,构建基于特征内容一致性的交互机制,从而在深层特征空间中减弱浅层细节错位及语义偏移带来的影响。如Zhong等人(2025)考虑到多模态反向蒸馏过程中多模态特征直接全局对齐易造成模态有效信息被削弱,同时也考虑到各模态特征在频域空间中的低频、中频及高频部分的特征分布差异,提出了频谱特征对齐模块(frequency spectral feature alignment, FSFA),将多模态特征映射到统一频域特征空间并分解为低、中、高三个频段,再对描述物体全局与结构信息的低、中频成分进行对齐约束,而完全保留对应各模态特异性细节的高频成分不变,从而有效避免了模态特有信息的丢失。Shi等人(2025)则针对工业缺陷检测场景中振动干扰及单向映射易引发特征漂移问题,通过构建双向循环一致性约束,基于 $2D \rightarrow 3D \rightarrow 2D$ 与 $3D \rightarrow 2D \rightarrow 3D$ 的双向闭环可逆映射,引导特征在多模态转换中保持可逆对齐。

时序对齐失稳应对则致力于通过帧同步、动态时间规整或时间滞后补偿等方法,来确保时序维度上的协同一致性。相较于特征对齐与时序对齐,空间对齐问题往往具有更明确的物理诱因。传感器标定偏差、设备振动、被测表面的非刚性形变等因素,均会引起多模态数据在空间位置上的偏移或局部错位,并进一步干扰缺陷区域的对应关系建立。因此,国内外学者围绕如何应对多模态空间对齐失稳挑战,进行了较为广泛的研究,主要侧重于从容错性对

齐机制、自适应对齐、分层级联对齐以及几何约束等策略,来实现像素级的鲁棒性对齐方法。

1) 容错性空间对齐机制

为应对多模态空间配准中常见的微小错位和误差,可通过设计具备容错能力的对齐策略,使得系统在对齐阶段能有效克服细微的空间偏差,即使存在轻微变形或错位的情况下,仍能维持稳定的空间对应关系。如Hoang等人(2026)针对高层语义易丢失细节且低层特征空间对齐较为敏感的挑战,提出一种自顶向下的对齐感知融合模块(alignment-aware fusion, AAF),在保持空间分辨率的同时显式修正微小偏移,使得模型在多模态对齐层面能够一定程度上容忍微小空间错位并重建精确的低层定位特征,从而提高受试者工作特征曲线下面积(area under the receiver operating characteristic curve, AUROC)与逐区域重叠曲线下面积(area under the per-region overlap curve, AUPRO)指标,提高了定位可靠性。

2) 动态自适应空间配准

多模态数据之间的空间偏差通常会随着成像距离、观测视角、尺度变化等因素的变化而动态波动,而传统的静态对齐方法往往难以适应这种动态变化的多模态对齐失稳情况,动态自适应空间配准机制通过实时估计模态间的几何形变差异与变化,自动调整配准参数,能够在不同采集条件或场景变化下,维持稳定的空间对应关系。如Zhu等人(2025)在多模态定位阶段,针对不同模态在空间分辨率、视场对齐上的差异,提出了一种基于梯度特征的动态自适应对齐方法,首先从各模态图像中提取相对稳定的梯度信息,构建鲁棒特征对应关系以实现基础空间匹配;随后引入动态自适应机制,对模态间尺度、方向及信号强度变化进行自动补偿,从而实现空间一致、光度归一的多模态配准。动态空间配准机制不依赖固定的标定先验,而是依据各模态在当前输入中的空间结构信息与几何对应特征,对配准过程进行动态调整与补偿,从而提升多模态配准对尺度变化、旋转偏移等偏差的适应能力。

3) 分层递进式配准机制

Li等人(2023a)引入了粗配准-细对齐的递进式空间配准策略,通过全局单应性变换实现粗配准以消除几何偏移,从而实现全局空间粗对齐;并进一步结合域自适应与局部相似性约束完成细粒度对齐,从而提升跨模态空间配准鲁棒性。Li等人(2026b)

则基于级联式配准机制,先利用ORB特征检测与随机采样一致性算法(random sample consensus, RANSAC)实现全局粗配准,再结合高斯金字塔下的多尺度(bilateral total variation, BTV)正则化Demons算法,由粗到细地优化局部非刚性形变场,在抑制噪声引发畸变的同时保留边缘特征。Almukhtar等人(2026)针对多源传感器成像存在的几何畸变与视角差异问题,采用(scale-invariant feature transform, SIFT)特征匹配与RANSAC单应估计来完成粗粒度几何配准以剔除误匹配点,并对重投影误差超出阈值的样本基于相关性优化进行细粒度校正,弥补弱纹理区域的配准不足,将重投影误差均值由2.85像素降至1.57像素。分层递进式配准机制的核心在于将复杂的多模态对齐过程分解为由粗到细、逐层优化的多阶段配准过程,先利用全局配准消除尺度、旋转及位移等整体几何偏差,再针对局部形变与特征分布差异进行精细化对齐,从而有效降低单阶段配准对初始条件与匹配质量的依赖。

4) 基于几何约束的对齐机制

Na等人(2026)针对模态间在视角差异、模态异构及缺陷表征不一致等挑战,提出一种显式空间对齐模块(explicit alignment module, EAM),该方法基于正交投影交并比,来对双模态预测框在正交投影下的交叉扫描位置一致性进行约束,从而缩小两分支在输出层的空间偏差。通过直接构建多模态空间一致性约束,确保不同模态间的空间投影关系一致,可应对来自模态异构、视角差异、噪声等因素导致的多模态对齐失稳挑战。

如表5所示,从适用场景、鲁棒性应用价值、局限性,深入提炼与总结对比了多模态空间对齐交互鲁棒性挑战的应对方法,并给出了后续应用建议。

3.3 多模态融合偏差应对方法

在实际多模态融合过程中,受模态间固有差异、对齐偏差、特征冗余及噪声传递等因素影响,多模态融合过程中可能出现模态贡献失衡、冗余信息叠加等问题,从而使得融合结果偏离真实缺陷表达。因此,多模态融合偏差挑战应对的核心,在于构建信息互补、权重合理及抗干扰的鲁棒融合机制,通过抑制融合过程中的各类偏差、优化融合策略等,以提升多模态协同建模的稳定性与可靠性。

1) 自适应特征融合

由于不同模态在多模态融合过程中的重要性或

贡献度并不是固定不变的,而是会受噪声干扰、局部遮挡、对齐误差及特征冗余等因素影响而变化。若采用传统固定拼接或静态加权方式,则退化模态或冗余模态信息便可能被引入融合过程,并引发模

态贡献失衡与决策偏差(Wang等, 2026d)。而自适应特征融合强调依据当前输入在特征层面的实际贡献或响应强弱,动态调整不同模态的权重贡献与交互强度,从而降低模态偏差的影响(Li等, 2025b)。

(1) 注意力机制引导的动态融合

注意力机制引导的动态融合通过利用通道注意力或空间注意力等手段,自适应地捕获不同模态特征在全局或局部区域的相关性与重要性(Zhang等, 2026b),在抑制冗余噪声与模态偏差的同时,也能自适应增强关键缺陷特征。Wang等人(2026c)针对可见光与热红外特征在融合阶段易受模态信息偏差和局部干扰影响的问题,提出了一种多模态动态交互融合机制(cross-modal interaction module, CMIM),在双分支特征提取之后,通过引入联合通道注意力与多模态空间注意力,对两模态特征基于通道与区域进行自适应加权,从而抑制受电弧光与飞溅污染的可见光信息,并强化热红外模态中更稳定的热异常特征。He等人(2025a)则引入多模态注意力交互模块(cross-modal attention interaction module, CAIM),通过可学习的双向动态权重机制,自适应调节各模态在融合中的贡献比例,从而选择性保留缺陷关键信息并抑制无关噪声。

(2) 基于度量评估的权重自适应融合

Cheng等人(2025)通过度量单模态特征与多模态特征相对原始特征的差异,计算各自互补性权重并动态完成融合,从而引入差异引导的自适应融合机制,在保留核心特征的同时抑制多模态交互干扰。Liu等人(2025b)提出一种基于能量评估的动态融合机制,通过计算每个模态输出与预期输出之间的能量差异,动态分配权重给不同模态,其中,对于不确定性较低的模态分配更高的权重,从而避免了不确定性较高模态对决策结果的负面影响。Daci等人(2026)通过联合映射异常与重建误差来构建可靠性系数和置信权重,对多模态映射异常与各模态重建异常响应进行自适应加权融合,规避不可靠的模态对最终决策的影响。基于度量评估的权重自适应融合通过构造差异性、不确定性、重建误差等评价量,

表5 多模态空间对齐交互鲁棒性挑战应对策略提炼与总结对比

Table 5 Comparative summary of strategies for addressing multimodal spatial alignment interaction robustness challenges

文献	适用场景	对齐鲁棒性应对价值	局限性	应用建议
容错性对齐 (Hoang 等, 2026)	模态间存在轻微传感器错位、局部偏移或定位误差	修正局部偏移,缓解小幅错位对像素级定位的影响	对较大尺度的几何错位、明显视角差异等适应性较差	可结合不确定性估计,同时引入错位幅度感知方法,来明确容错范围
自适应对齐 (Zhu 等, 2025)	成像距离、观测视角、尺度、方向或光照响应随工况变化而动态波动	根据输入梯度、边缘或结构信息动态调整配准参数,降低对固定标定先验的依赖	依赖相对稳定的梯度、边缘与纹理特征,若相关要素变化,会影响对应关系可靠性	建立模态质量与可信度评估机制,以降低弱纹理、强反光或噪声干扰下的错误匹配风险
递进式对齐 (Li 等, 2023a; Li 等, 2026b; Almukhtar 等, 2026)	各传感器视角差异明显,存在全局几何偏移、局部形变或显微成像畸变	通过粗配准消除整体尺度、旋转和位移偏差,再通过细配准修正局部残余形变,提高像素级空间一致性	流程复杂、计算开销较大,粗配准阶段若出现错误匹配,后续细配准可能沿错误方向继续优化	在粗配准阶段加入匹配质量评估与失败回退机制,并在细配准阶段结合正则化、边缘保持或置信度约束,来兼顾精度与实时部署需求
几何约束对齐 (Na 等, 2026)	扫描路径、投影关系、检测框对应关系等几何空间关系较为明确	通过相关一致性、重投影误差等约束,增强空间对应关系的可解释性与可验证性	若不同模态对缺陷响应形态差异较大,可能削弱模态特有特征并导致过度对齐	将几何约束与模态特异性保留相结合,对不同模态响应差异较大的区域采用软约束或动态权重约束

来对各模态信息的可靠程度进行量化判断,并据此动态分配融合权重,更有针对性地削弱不确定性高、偏差大的模态信息对决策的干扰。

(3) 基于门控机制的自适应融合

Wang 等人(2026a)针对早期融合易引发模态主导抑制和交互失稳的问题,提出一种决策级自适应融合机制,通过构建多模态压缩激励门控融合模块(cross-modal squeeze-excitation gated fusion, CM-SEGF),利用多模态全局通道统计对各模态特征进行重标定,并联合 RGB 数据、深度模态信息及模态差异来生成像素级自适应权重,从而实现自适应决策融合,在 MVTEC 3D-AD(Bergmann 等, 2021)上取得 97.1% 的图像级 AUROC 与 98.8% 的像素级 PRO。门控机制通过对模态或通道信息进行选通与调制,以简单直接的方式实现多模态动态切换与融合,可有效应对单一模态退化或交互失稳。

2) 双向映射约束融合

传统单向多模态映射虽能建立模态间的关联,但其交互过程主要依赖单一方向的信息传递,一旦受到噪声扰动、局部遮挡或模态偏差影响,误差便容

易沿该方向持续累积,进而削弱交互稳定性。相较之下,双向多模态映射通过同时构建正向与反向的映射关系,使各模态在相互关联的基础上进一步形成双侧约束与交叉校验机制,从而将多模态交互由单边推导转化为双向闭环反馈。如 Mai 等人(2025)通过 M2D 至 3D 与 M3D 至 2D 两个轻量级映射网络来构建多模态双向映射机制来实施逐像素双向预测,并引入随机特征掩码以引导模型从互补模态恢复被遮蔽区域,从而强化多模态互解释能力与交互稳定性。如图 4 所示,Asad 等人(2025)则通过反投影建立 RGB 特征与点云特征的点级对应,并利用双向多层感知机映射(multilayer perceptron, MLP)进行多模态特征适配,从而实现点级对齐与融合,减轻异构特征直接交互带来的干扰。

3) 中间表征空间驱动的多模态融合方法

传统多模态融合方法往往在异构模态之间直接建立刚性对应关系,存在一对多映射、局部冲突及外部因素扰动等问题,易导致交互不稳定。为此可引入中间表征空间机制,将原本直接的模态到模态映射转化为更稳定的中间交互空间,从而降低原有成

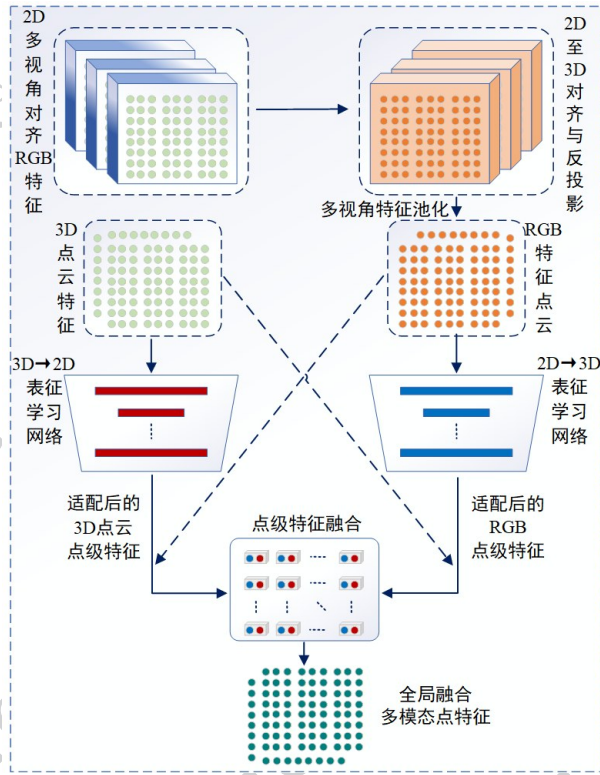


图4 多模态表征网络与点级融合示意图

Fig. 4 Schematic diagram of multimodal representation network and point-level fusion

像差异、局部扰动、光照变化等对原始映射的干扰,避免因直接映射偏差导致的融合不稳定问题。如Wang等人(2026d)提出了统一概率空间下的局部卷积式融合机制,将RGB图像模态与三维高度模态映射到统一的概率表达空间,避免了颜色与高度模态的简单组合在局部冲突场景下引发的融合偏差。Shangguan等人(2025)通过引入潜在桥联模态映射模块(latent bridged modal mapping module, LB3M),通过LB3M将原本模态到模态直接映射,优化为各模态与中间空间的双向转换,在中间潜在空间中完成信息交互,从而把受外部波动影响较大的直接对应关系,转化为更稳定的高层语义关联,缓解了一对多映射带来的映射不稳定性。

4) 其他模态融合偏差应对方法

(1) 多模态异常合成训练机制

Zhong等人(2025)针对多模态融合过程中因模态偏差引发的单一模态异常可能被另一正常模态掩盖的模态融合问题,提出频谱异常合成模块(frequency spectral anomaly synthesis, FSAS),在FSFA对齐优化机制基础上,通过将主动扰动生成的模态

异常特征,与另一模态的正常特征进行融合来输入学生网络,并利用损失函数来引导学生网络在融合特征中重构原始正常特征,从而使模型在多模态融合过程中减少受正常模态掩盖效应的影响,提升对模态间信息不一致的识别能力。通过主动构造单模态异常的多模态融合样本,依托异常合成训练机制,可强化模型对模态信息差异的感知能力。

(2) 双分支一致性约束融合

Wang等人(2026b)提出一种双分支多阶段融合与重映射机制(feature-level multi-stage fusion and remapping, FMFR),一方面通过构建融合-重构分支网络(fusion-reconstruction branch, FR)来学习信息完整的融合表征;另一方面则构建重映射-融合分支网络分支(remapping-fusion branch, RF),通过双向多模态重映射再融合,来形成第二条并行路径,并借助多阶一致性损失对两条路径的融合表示进行约束,从而建立表示层面的验证机制。在MVTec 3D-AD数据集上实现了98%的I-AUROC和97.9%的AUPRO@30%,提升了多模态交互的稳定性。

如表6所示,在公开数据集MVTec 3D-AD下对比了多模态融合交互鲁棒性挑战应对的相关方法及贡献。进一步的,如表7所示,从方法适应性、局限性及后续方法应用建议,深入提炼与总结对比了多模态融合交互鲁棒性挑战下的相关应对方法。

4 大模型驱动的多模态表面缺陷检测系统鲁棒性制约与优化

相比较而言,传统深度学习通常围绕具体任务来构建专用网络结构,其性能主要依赖数据构成与规模、算法架构的选择等。而大模型驱动的多模态缺陷检测系统则依托于大规模预训练形成的通用表征、语义理解与推理能力,并通过提示、微调等方式迁移到具体检测场景中。这种变化不仅使系统架构由专用模型主导逐步转向通用大模型支撑,也推动多模态缺陷检测系统由单一任务实现,拓展为识别、理解、分析等相结合的综合处理过程。但同时,其在实际应用中也面临着如模态兼容性、提示词敏感、生成可靠性和领域适配等一系列新的鲁棒性挑战。

4.1 大模型驱动的检测系统核心特性与优势

4.1.1 大模型驱动的多模态缺陷检测系统架构

表6 MVTec 3D-AD数据集下的多模态融合鲁棒性挑战应对策略对比

文献	分类	方法	贡献
Cheng 等(2025)	基于度量的 自适应融合	度量单模态与多模态特征相对源特征的差异,评估互补性并引导自适应融合	P-AUPRO@30%由97.1%提升至97.5%,P-AUPRO@1%由45.8%提升至47.1%
Daci 等(2026)		联合映射异常与重建误差来构建可靠性系数和置信权重以实现自适应加权融合	I-AUROC: 97.3%、P-AUROC: 99.6%、AUPRO@30%: 97.6%
Wang 等(2026a)	门控机制引导的 自适应融合	构建门控融合模块,以通道统计重标定特征,生成像素级权重来动态平衡双模态	I-AUROC: 97.1%、像素级 PRO: 98.8%
Mai 等(2025)	双向映射	通过 M2D 至 3D 与 M3D 至 2D 两个映射网络来构建多模态双向映射机制	I-AUROC: 95.3%、AUPRO@30%: 97.2%
Asad 等(2025)		反投影建立图像与点云特征的点级对应,利用双向 MLP 进行多模态特征适配	I-AUROC 与 AUPRO 均提升3个百分点,最终 I-AUROC: 96.6%、AUPRO: 97.5%
Shangguan 等(2025)	共享中间空间	基于 LB3M 实现各模态与中间空间的双向转换,在中间空间中完成信息交互	I-AUROC: 97.4%、AUPRO@30%: 97.8%
Zhong 等(2025)	异常合成训练	通过扰动来模拟模态一致异常,并进行多模态训练	I-AUROC 96.8%、P-AUROC 99.5%
Wang 等(2026b)	双路径一致性约束	构建 RF 与 FR 双分支来多角度学习融合表征	I-AUROC: 98%、AUPRO@30%: 97.9%

组成

大模型驱动的多模态检测系统主要建立在通用预训练底座之上,其整体架构一般由输入侧模块、模态编码器、任务适配层、大模型主体以及任务输出与决策层构成。如图 5 所示是大模型驱动的多模态缺陷检测系统架构图(董禹彤等, 2026; Gu 等, 2024;

Li 等, 2023b),视觉、深度、热红外及点云等模态数据首先分别进入对应编码器,并被转换为可计算的特征序列;随后经由任务适配层映射到统一语义空间,再与文本 token 一同输入大语言模型或视觉语言模型主体,在统一上下文中通过自注意力机制实现多模态交互与信息整合,最终输出缺陷描述、类别判断、定位结果或结构化检测报告。

4. 1. 2 大模型驱动的多模态表面缺陷检测系统

优势

传统多模态检测系统往往面向固定模态组合、固定输出形式和封闭任务流程进行设计,而大模型驱动的多模态检测系统则具有更强的知识迁移、任务扩展以及人机交互能力(Zhao 等, 2025b)。

1) 多模态学习与生成

相较于传统方法主要依赖多模态对齐、特征拼接或加权融合等方式来实现多模态信息整合,大模型则依托统一的标记序列与内置的 Transformer 自注意力来实现动态交互建模,尤其在视觉与文本等模态融合方面具有更强的先天基础,而且可围绕缺陷类型、检测工况等来生成相关描述信息与工艺解析,同时可利用其视觉理解与语言表达能力,对缺陷区域、类别属性等进行自动化或半自动化标注。

表7 多模态融合交互鲁棒性挑战应对策略提炼与总结对比

方法	适用条件	局限性	应用建议
注意力机制 (Wang等, 2026c; He等, 2025a;)	适用于不同模态具有互补性或模态响应差异大,且可通过空间、通道或语义差异进行动态筛选与加权的场景	在干扰、纹理相似或模态不平衡条件下,注意力权重可能偏向噪声、背景纹理或单一优势模态,影响融合稳定性	引入模态质量评估、缺陷区域约束和多模态一致性约束等,引导注意力权重更多依赖相关质量评估与机制约束
度量评估 (Cheng等, 2025; Liu等, 2025b; Daci等, 2026)	适用于模态可靠性差异明显,且能够通过不确定性、重建误差、映射差异或异常响应强度等度量值,来构建有效评价量的多模态缺陷检测场景	度量指标不一定等同于真实缺陷贡献,噪声、弱纹理、深度稀疏或异常本身等可能引起高误差响应,导致可靠性判断与缺陷判别混淆	可将单一度量扩展为多指标联合评估,并结合任务损失、缺陷先验和扰动验证来校准权重分配的依据
门控机制 (Wang等, 2026a)	适用于模态贡献差异显著,或需要低复杂度动态融合决策的部署场景	门控机制可能过度依赖优势模态,且无法应对模态缺失或退化场景	引入软门控、置信度约束等,避免硬性选通造成互补信息损失
双向映射约束 (Mai等, 2025; Asad等,2025)	适用于模态间具有明确空间对应、点级关系或结构关联,需要缓解单向映射误差传播的场景	模态间初始空间对齐若存在偏差,易导致双向映射出现同步错误;一致性约束过强也可能削弱模态间互补性	优先通过精准空间标定保障对齐精度,并同时结合空间校准、局部置信度筛选和缺陷区域约束来提升检测可靠性
中间表征空间驱动 (Wang等, 2026d; Shangguan等, 2025)	适用于RGB、高度、深度、点云等模态物理属性差异较大,直接拼接或直接映射易失稳的检测任务	对数据分布规律或物理先验条件依赖较强,面对复杂场景中的未知缺陷与工况扰动时,中间空间难以有效表征多模态内在关联,导致融合精度下降	引入缺陷判别与模态一致性约束等,并建立在线检验校准机制,以调整中间空间的构建方式,增强模型对已知模式和未知异常的判别鲁棒性

2) 交互式缺陷分析与决策

通过自然语言接口接收文本指令、专家规则、及上下文约束,并以问答、解释或结构化报告形式来实现交互式分析,增强系统的人机交互能力。

3) 通用数据基础与领域知识迁移

依托于大规模、高质量的通用数据基础与预训练知识,使得大模型具备更强的迁移学习能力,能够在小样本甚至零样本条件下(Deng等, 2025; Jiang等, 2025a; Li等, 2026a),通过提示学习、参数高效微调或领域适配实现对新任务的快速迁移。

4) 统一任务接口与多任务执行能力

传统多模态表面缺陷检测通常需针对分类、定位和分割等任务构建专用模型,系统集成与部署成

本较高,且难以灵活扩展。相比之下,大模型驱动的多模态检测系统则通过统一的序列到序列架构,可将不同任务转化为对文本或坐标的生成问题,实现缺陷分类、定位、像素分割、语义描述、异常解释及交互问答等多任务协同处理(Li等, 2025c)。

4.2 大模型驱动的检测系统鲁棒性制约因素

大模型驱动的多模态表面缺陷检测系统虽然在知识迁移、多模态协同和任务灵活性等方面展现出显著优势,但由于受到模态兼容性、幻觉问题、提示输入敏感性、预训练知识依赖以及大模型微调失衡等多重因素的影响,对系统在复杂检测环境中的稳定可靠运行构成了新的鲁棒性挑战。

1) 模态兼容性不足

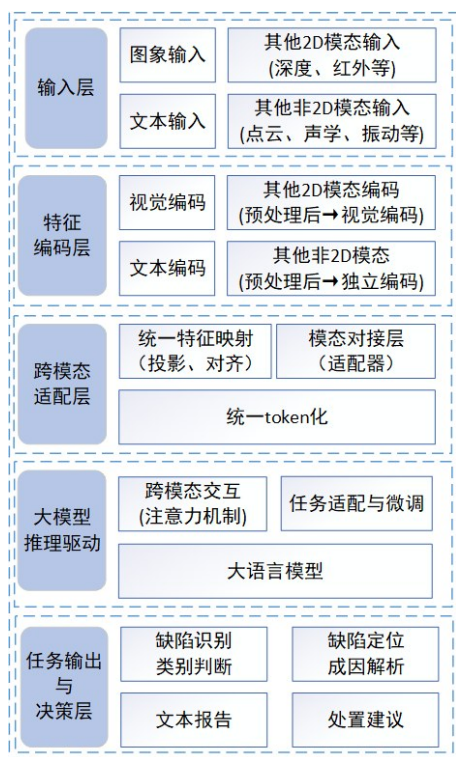


图5 大模型驱动的多模态缺陷检测系统架构图

Fig. 5 Architecture of large-model-driven multimodal surface defect detection systems

当前大模型驱动的多模态表面缺陷检测系统虽然在视觉与文本协同方面表现出较强能力,但对深度、点云、声学等非视觉文本模态的支持仍相对有限(Yin等, 2024)。如Salah等人(2026)分析认为现有热成像模态与视觉语言模型(vision-language model, VLM)所依赖的自然图像分布之间存在域偏移,难以直接形成与基础VLM相对齐的输入表示。针对这种问题,现有系统往往需借助专用编码器、投影层或数据转换策略才能接入统一语义空间,而模态信息在编码、压缩和映射过程中可能出现语义表达失真等问题,一旦模态特征在输入过程中发生信息损耗,就会影响系统后续的多模态理解与推理稳定性。

2) 生成式幻觉问题

幻觉问题是大模型作为生成式模型所特有的鲁棒性缺陷(Wang等, 2024b; Zhong等, 2024)。传统深度学习模型在面对不确定输入时通常表现为置信度降低或预测偏差,而大模型则倾向于生成看似合理但实际错误的输出。Chen等人(2026a)指出预训练大模型在工业异常检测中,容易在输入查询图像和问题时生成错误或虚假的输出,特别是在面对复杂缺陷和细微异常时,模型可能会遗漏关键细节或

作出过于笼统的判断。Lee等人(2026)分析认为通用VLM因缺乏基于图像的上下文信息支撑,在复杂纹理或缺陷特征不足时,不仅会对无缺陷的复杂纹理图像,生成看似合理却错误的判断,而且对真实缺陷图像,无法给出精确的类型和坐标,仅提供模糊的位置描述。吴华等人(2025)则指出现有方法的性能均依赖于预训练的多模态模型,这些模型往往会集中于对显著对象的关注,当面对有复杂背景的缺陷检测任务时,可能会产生幻觉问题,从而导致误检。

3) 提示输入敏感性

传统多模态缺陷检测系统在固定输入和固定网络结构下具有较稳定的推理输出流程,其输出结果仅取决于各模态数据质量。而大模型则将用户提供的文本提示作为输入的重要组成部分,系统行为高度依赖于提示文本的质量与措辞。输入提示词的微小变化以及语义差异性可能导致输出结果的不确定变化(Ismithdeen等, 2025)。在需要长期稳定运行的工业系统中,这种对自然语言指令的过度敏感,使得系统行为难以保持一致性,影响系统在连续复杂检测场景中的稳定可靠运行。

4) 对通用预训练知识的过度依赖

由于通用预训练数据多来源于自然场景和开放语义环境,其知识分布、视觉偏好与表面缺陷检测所要求的细粒度缺陷特征识别和生产工艺理解等方面存在偏差(Guo等, 2026; Li等, 2025e)。这就可能导致模型在宏观语义理解上表现较强,却在具体缺陷特征、类别解析及专业领域术语理解上存在偏差。即使经过模型微调,大模型仍然倾向于用通用知识而非检测任务下的领域视角理解(Chen等, 2025)。

5) 大模型微调失衡

大模型驱动的多模态表面缺陷检测系统,通常需要通过领域微调来完成具体检测任务场景下的知识适配与迁移,微调过程中的参数更新与知识融合会直接影响系统的整体鲁棒性。微调方法不当会导致模型原有通用认知能力、跨任务理解与推理能力受到抑制,进而出现模态关注偏差及查询响应下降等问题。如Jiang等人(2026b)指出,现有大模型在晶圆缺陷领域微调时,易过度聚焦视觉模态输入,未能有效判断用户问题与视觉内容的相关性,从而忽视文本查询语义,削弱对原有通用知识的理解能力。

4.3 大模型驱动的缺陷检测系统鲁棒性优化路径

与应用案例

大模型驱动的多模态缺陷检测鲁棒性风险不再主要集中于某一专用模块的局部失效,而更多体现为大模型底座能力、任务接口、生成过程以及领域知识之间的系统性耦合。因此,针对大模型系统的鲁棒性提升,应更多从模态接入、推理增强、多模态交互平衡、多专家协同决策、结构化生成约束等层面展开,在保持大模型通用能力优势的同时,增强其在复杂缺陷检测场景下的稳定性与可控性。

1) 模态适配优化

针对深度、点云、热红外和声学等非视觉文本模态在接入大模型时易出现信息压缩和语义失真的问题,需构建专用编码器或适配模块,以保留点云几何结构、热红外温度分布和声学时频特征等关键物理信息,并提升其向统一语义空间映射的一致性。如 Salah 等人(2026)面向航空航天复合材料检测,引入主动热红外数据与视觉语言模型适配器(active infrared thermography and vision-Language model, AIRT-VLM),通过标准化、掩码编码与潜表示聚合等操作,将原始热红外数据序列转换为兼具高信噪比与域对齐特性的单幅图像特征,从而在保留缺陷显著性的同时提升了热红外模态对视觉语言模型的接入适配能力。Wu 等人(2026b)则引入动态专家路由低秩适配机制(dynamic expert routing with LoRA, DyLoRA),将传统低秩适配(low-rank adaptation, LoRA)扩展为共享低秩矩阵与多个专家适配分支组成的 DyLoRA 结构,并依据输入模态与语义内容动态分配专家权重,使大模型能够对不同传感模态进行差异化表征学习,从而有效应对了光伏故障诊断场景中可见光 RGB、近红外、热成像数据等异构模态,难以通过统一静态参数实现稳定适配的问题。

2) 推理增强与知识融合

Chen 等人(2026a)为应对生成式幻觉问题,首先引入参考图像检索机制来提供与查询图像相似的正常样本作为对比,避免依赖参数记忆而产生虚假判断;其次,基于领域知识引导方法来集成行业特定的缺陷标准与领域判别规则,来为模型生成真实可靠依据;最后采用零样本链式推理策略,并配合门控混合专家路由机制来动态筛选有效模块,避免冗余信息干扰,从而结合上述方法的协同作用,提出了一种推理增强与知识融合技术路径来增强大模型在应对复杂、细微缺陷时的鲁棒性,减少了因信息缺失或

领域知识不足导致的幻觉问题。Lee 等人(2026)则提出一种动态检索增强策略,通过约束语义分类将缺陷类型限定在预定义工业词表内,之后基于特征匹配的异常先验图为模型提供空间证据,并将语义证据与位置先验共同注入增强提示,同时借助质量保证(quality assurance, QA)模块来筛选高质量模板,以降低生成过程中的自由漂移,从而有效抑制无缺陷误报并消除假阳性幻觉。Mistry 等人(2026)针对光伏组件缺陷检测中通用大模型缺乏领域推理逻辑、易产生幻觉的问题,提出了检测框架 REVL-PV(reasoning-boosted vision-language for photovoltaics),构建了由视觉证据、故障成因和处置建议组成的结构化推理链条,在提升分类准确率的同时,增强了诊断结果的可解释性和可靠性。通过构建结构化的推理链条,重构缺陷判别过程中的各模态信息组织方式与推理约束机制,并结合领域词典、专家规则或参考样本等外部知识,降低模型仅凭通用预训练知识来进行开放式生成所带来的不稳定性,能够弥补大模型在逻辑推导上的固有局限。

3) 多模态交互平衡优化

针对大模型微调失衡导致的模态偏差问题,Jiang 等人(2026b)在晶圆缺陷检测场景中,基于调制模块与交互式语料库协同设计来构建应对方案,具体而言,通过引入注意力交互、可学习校正和动态系数更新,并结合交互式语料库训练策略,在图像、文本、缺陷掩码和问题信息间建立更平衡的融合关系,从而缓解模型对单一视觉特征的过度依赖,提升小目标和低对比度缺陷条件下的识别稳定性。

4) 多专家协同决策优化

为提升系统在不确定场景中的稳定识别能力,国内外学者通过构建多个具有差异化能力的专家分支,使系统能够从不同视角对输入样本进行并行建模,并在后续通过一致性约束、动态加权、门控路由等方式来整合多专家意见,从而降低传统单专家偏置带来的检测不确定性。如吴华等人(2025)面对电力巡检影像中复杂背景、光照视角动态扰动易引发模型误检的幻觉问题,提出双专家协同判别方法,对多个中间层视觉特征分别引入双专家模块,并分别基于两类不同的数据集对各专家模块进行独立训练以获取多视角的差异化判别能力,并最终基于逐元素相乘融合方式完成协同专家决策。

5) 结构化生成约束机制

Ouda 等人(2026)考虑到大模型开放式生成结果往往存在稳定性差、可控性不足的问题,提出一种三段式结构化生成约束机制。具体而言,将热成像与光学图像按既定顺序输入模型,并通过结构化提示将输出约束为热成像结果分析、真实性评估、缺陷位置与成因等三部分,以标准化三段式模板来限定输出框架;同时要求模型区分直接观测与推测性判断,并在证据不足时显式表达其不确定性。通过上述约束,将传统大模型较为开放的自由生成过程转化为受证据边界约束的结构化报告生成过程,从生成端来抑制虚假输出。区别于模态适配、推理增强等侧重于输入建模与判别过程优化,该控制机制更强调对输出行为本身的约束,通过结构化模板、内容边界限定和不确定性显式表达等控制策略,有效约束了大模型固有的生成式发散性局限,从而以较低的计算成本实现了输出稳定性的提升。

上述方法分别作用于缺陷检测系统的模态输入、异常证据获取、跨模态语义交互、协同决策和结果生成等环节。其中,模态适配侧重于降低非视觉模态转换过程中的信息损失和语义偏移;推理增强与知识融合聚焦于提升复杂背景下的缺陷定位、类别判断和成因解析的可靠性;多模态交互平衡用于缓解不同信息源之间的偏置;多专家协同决策侧重于降低单一模型在复杂干扰下的误检风险;结构化生成约束机制则用于抑制开放式生成中的幻觉、语义漂移等问题。由此可见,大模型鲁棒性优化并非通用能力的简单迁移,而是贯穿于缺陷感知、定位、解释以及输出生成的全流程任务适配机制。如表8所示,分析了大模型驱动的多模态缺陷检测鲁棒性优化的相关典型应用案例,从应用场景、模态组合、问题表现以及优化方法等角度进行了归纳。

5 数据集与评价指标

5.1 数据集

当前专门面向多模态缺陷检测鲁棒性研究的公开数据集相对匮乏,尚未形成独立、系统的多模态鲁棒性数据支撑体系。如 MVTEC 3D-AD、MulSen-AD (Li 等, 2025a) 和 Eyecandies (Bonfiglioli 等, 2022) 等现有公开数据集,主要面向多模态缺陷检测性能验证,对噪声扰动、模态缺失及多模态交互偏差等鲁棒性问题的显式关联较少。然而,这些数据集所包含

的多模态配对信息,也为后续构建鲁棒性测试场景提供了基础。具体而言, MVTEC 3D-AD 提供了 RGB 图像与三维点云的配对信息,为多模态表面缺陷检测中纹理异常与几何异常的协同分析奠定了数据基础,如表9所示,分析了围绕 MVTEC 3D-AD,在噪声干扰、模态缺失等扰动因素下的重构或改进措施; MulSen-AD 通过融合 RGB 图像、红外热成像和三维点云数据,进一步扩展了外观纹理、热辐射特征和几何结构之间的互补关系,可用于分析多传感器系统在不同模态退化或局部失效条件下的稳定性;此外, Eyecandies 虽然提供了 RGB 图像、深度图和法向图等多模态信息,但其检测对象为程序化

渲染生成的糖果类目标,属于合成多模态异常检测数据集,更适合作为可控扰动构造和方法验证的数据基础。此外, Miao 等人(2025)也基于 MVTEC-3D AD 与 Eyecandies 构建了面向模态不完整工业异常检测的基准数据集 MIAD Bench,来模拟真实场景中的模态缺失问题。其核心在于评估了多模态缺陷检测系统对输入模态缺失的鲁棒性,涵盖了点云缺失、RGB 模态缺失及双模态缺失三类场景,并设置了 30%、50%、70% 三种缺失率,基本复现了模态缺失的随机性与不确定性,为测试模型在非理想输入下的性能稳定性提供了标准化支撑。如表10所示,对比探讨了上述数据集的鲁棒性相关适用边界。

5.2 评价指标

多模态缺陷检测鲁棒性的评价核心,在于考察系统在噪声扰动、模态缺失、多模态对齐失稳以及融合偏差等非理想条件下,能否依然保持稳定可靠的检测能力。如 Wang 等人(2026c)基于实验标准差 Std 与扩展不确定度 U 来刻画重复实验中的性能离散程度,用以评估模型在随机初始化与训练扰动下的稳定性,从而为鲁棒性分析提供了定量依据。Du 等人(2026)创新性地引入噪声灵敏度指标(noise sensitivity, NS)、光照适应性指标(illumination adaptability, IA)与模态缺失适应性指标(modality missing adaptability, MMA),分别用于评估模型在随机噪声干扰、环境光照变化及单模态输入故障下的性能稳定性。然而针对多模态缺陷检测鲁棒性评估的专用指标,目前还是较为缺乏的,一方面,多模态表面缺陷检测系统的鲁棒性成因具有显著差异性,各类系统或任务在模态组合、退化形式、扰动来源及任

表 8 大模型驱动的多模态缺陷检测鲁棒性优化的典型应用案例

Table 8 Typical application cases of robustness optimization in large model-driven multimodal surface defect detection systems

应用场景	文献	鲁棒性问题	应对机制	技术路径	鲁棒性作用
复合材料近表面检测 (关联表面缺陷检测)	Salah 等(2026)	热成像特征难以直接适配传统视觉语言大模型	模态适配优化	AIRT-VLM 适配器、特征标准化与潜表示聚合	提升热红外模态接入稳定性与多模态一致性
光伏组件设备故障诊断 (关联表面缺陷检测)	Wu 等(2026b)	静态 LoRA 难以适配多源异构数据	模态适配优化	动态专家路由低秩适配; 共享低秩矩阵、多专家分支	提升异构模态接入适配能力与诊断生成稳定性
光伏组件表面缺陷检测 (聚焦表面缺陷检测)	Mistry 等(2026)	诊断推理不稳定、结果不可解释、易产生幻觉	推理增强与知识融合	结构化推理链、两阶段推理增强(2PRE)、领域知识注入	分类准确率达 93%, 且噪声扰动下准确率仍达 84%
通用工业表面缺陷检测 (聚焦表面缺陷检测)	Chen 等(2026a)	幻觉生成、领域知识缺失、复杂问题推理不稳	推理增强与知识融合	参考图像检索、领域知识引导、链式推理、多专家路由	在 MVTec-AD 与 VisA 上平均准确率提升明显
通用工业表面缺陷检测 (聚焦表面缺陷检测)	Lee 等(2026)	幻觉误报、定位模糊、缺乏图像证据支撑	推理增强与知识融合	检索增强生成、异常先验图、模板筛选	抑制无缺陷误报并消除假阳性幻觉
晶圆表面缺陷检测 (聚焦表面缺陷检测)	Jiang 等(2026b)	模态偏置、视觉与文本信息失衡	多模态交互平衡优化	调制模块、交互语料训练、动态权重更新	平衡视觉与语义信息, 提高细粒度识别稳定性
电力设备表面缺陷检测 (聚焦表面缺陷检测)	吴华等(2025)	背景复杂、光照与视角变化影响、幻觉问题	多专家协同决策优化	双专家模块、多层特征建模	像素级 F1-max: 33.1%、图像级 AUROC: 68.2%
文化产品表面缺陷检测 (可迁移借鉴)	Ouda 等(2026)	生成结果不稳定、证据边界不清、解释不一致	结构化生成约束机制	三段式报告模板、不确定性显式表达、约束证据判定	提升生成结果可控性与可复现性

务目标上差异较大,使得鲁棒性难以由单一指标统一刻画;另一方面,该领域目前仍以鲁棒性增强或保障机制设计、退化场景下的性能验证等为主,尚未沉淀出具有统一的、可解释的专门评价体系。因而,采用如 I-AUROC、P-AUROC、AUPRO、F1 分数等传统性能指标来间接反映系统鲁棒性,仍是当前主要且广泛的系统鲁棒性评价方式。具体而言, I-AUROC 主要从图像级角度来刻画模型区分正常与异常样本的能力,可用于反映模型整体异常判别能力的稳定

性;P-AUROC 从像素级层面来体现异常响应图对缺陷区域与正常区域的区分能力,适合反映噪声、模态退化或缺失条件下的异常定位鲁棒性;而 AUPRO 则从区域级角度来分析异常区域与真实缺陷区域间的重叠关系,能够进一步反映缺陷区域的空间完整性、边界一致性和连通区域保持能力,对空间对齐偏移、局部遮挡和融合偏差等较为敏感。例如,在噪声应对鲁棒性评估方面, Wang 等人(2025a)通过 I-AUROC、P-AUROC、AUPRO 这三类性能指标来比较

表9 基于MVTec 3D-AD数据集的扰动与缺失改进

Table 9 Perturbation and missing-modality extensions based on the MVTec 3D-AD dataset

文献	分类	针对鲁棒性的数据集重构或改进措施
Jiang等(2025b)	模态缺失	在训练或推理阶段对各模态进行随机丢弃,并设置缺失率为30%、50%、70%
Hoang等(2025)		在训练阶段对各模态数据进行随机单一模态丢弃,在推理阶段构造单模态场景
Xu等(2025)		在训练时随机丢弃30%样本的RGB或3D模态,来模拟模态缺失与数据不完整
Wang等(2025a)	数据噪声	训练集注入10%的真实异常样本作为噪声,构建重叠与非重叠两类噪声训练集
Lee等(2026)		对测试图像动态施加光照、模糊、位移等合成扰动,以模拟实际干扰
Liao等(2025)	分布偏移	在原始数据集上合成了高斯噪声、散焦模糊、对比度、亮度变化等四种分布偏移

表10 典型多模态缺陷检测数据集的鲁棒性适用边界对比

Table 10 Comparative analysis of robustness applicability boundaries for representative multimodal defect detection datasets

数据集	模态类型	鲁棒性研究适用性	局限性
MVTec 3D-AD	RGB图像、三维点云	RGB模态退化、三维模态缺失、点云稀疏、RGB与三维信息协同失效等问题研究	模态数据验证类型主要集中于数据集内部模态,多模态数据的验证泛化性不足,此外由于数据集本身并非专门为鲁棒性扰动评价设计,故仍需要根据鲁棒性研究需求与数据集特性手动设定扰动类型与扰动强度
MulSen-AD	RGB图像、红外热成像、三维点云	视觉、红外数据或点云退化、多传感器融合失衡、局部传感器失效等问题研究	
Eyecandies	RGB图像、深度图、法向图	外观扰动、深度退化、法向图退化、模态缺失等问题研究	除上述局限性外,检测对象为程序生成糖果,属于合成数据;工业场景属性弱
MIIAD Bench	RGB图像、三维点云	专门面向模态不完整任务构建,用于评估模态缺失下的性能下降与鲁棒性保持能力	聚焦于模态缺失问题研究,对噪声扰动、空间对齐偏移、融合偏差等问题仍需进一步二次构造

模型在含噪多模态场景下的检测与分割表现,从而间接反映系统对噪声扰动的鲁棒性。针对多模态交互鲁棒性挑战,Hoang等人(2026)分析认为,AUPRO因基于连通区域计算,对空间错位导致的对齐问题最为敏感,在平移扰动实验中采用AUPRO来验证系统在多模态空间错位下的稳定对齐能力。此外,在模态缺失应对鲁棒性层面,Wang等人(2024a)以平均缺陷点精度(mean defect point accuracy, MDPA)、平均缺陷交并比(mean defect intersection over union, MDIoU)为核心指标,通过测试不同点云丢失率下指标的变化来评估系统鲁棒性。

为进一步分析多模态缺陷检测鲁棒性评价的基本环节及其适用边界,如图6所示,从测试集构建、扰动条件设计、模型对照测试、指标变化分析和失效

案例归因等环节进行归纳。同时,为增强鲁棒性评价内容的系统性,表11从鲁棒性问题、扰动设置、评价指标和评价重点四个维度,将感知层输入退化、交互层对齐与融合失效以及大模型生成可靠性问题纳入统一分析框架,从而为后续开展可控扰动测试、跨模态一致性分析和性能退化评估提供参考。

6 发展趋势

1) 噪声自适应感知与应对鲁棒性学习

复杂表面缺陷检测场景中,如光照变化、反光干扰、点云稀疏、温度漂移、运动模糊等噪声,因各模态物理特性的不同而产生的影响程度也有区别,未来针对系统噪声应对鲁棒性的发展,应从被动抗噪转

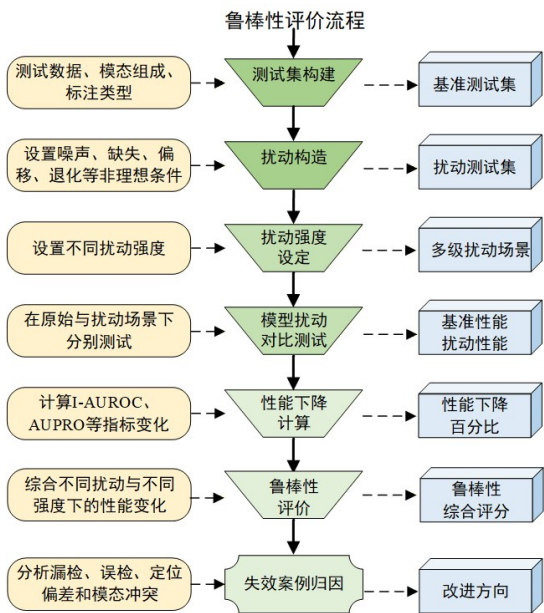


图6 多模态缺陷检测鲁棒性测试与评价分析流程图

Fig. 6 Robustness testing and evaluation workflow for multi-modal surface defect detection

向主动识别噪声并进行差异化处理,通过引入噪声强度估计、模态质量评分与置信度调整等机制,来发展噪声自适应鲁棒性技术,依托多模态数据的互补特性,设计动态去噪融合策略(Shi等,2025),使系统能够自动预判相关模态对某类噪声的敏感度,并进一步自适应抑制噪声传播路径,保留相对可靠的模态特征信息,提升系统噪声应对鲁棒性。

2) 面向模态缺失常态化趋势的鲁棒检测机制

在实际缺陷检测场景中,传感器失效、点云遮挡、热成像不稳定等问题十分常见,尤其是如基于无人机检测的室外场景或水渍粉尘干扰的传统工业检

测现场。模态缺失并不是多模态检测系统的偶发异常,而应该成为未来多模态表面缺陷检测系统设计

的常态化假设(Miao等,2025)。基于此,设计鲁棒的多模态表面缺陷检测系统应将模态缺失应对作为一种常规化算法应对模块,在模型设计阶段与训练阶段就应被充分考虑,从而更符合多模态表面缺陷检测系统工程化部署的需求。

3) 系统鲁棒性驱动下的轻量化协同发展趋势

为提升系统对噪声扰动、模态退化、缺失及交互偏差的鲁棒性,通常需引入更复杂的对齐与融合机制,但这也易导致模型结构臃肿与参数规模增加,进而削弱实时性与边缘部署能力。然而,表面缺陷检测通常面向在线生产与资源受限场景,对检测精度、稳定性及实时响应均提出较高要求。因此,未来多模态鲁棒性系统设计需协同考虑轻量化约束,通过动态计算、模型压缩与高效融合等手段,在控制模型复杂度的同时保障系统性能,实现可靠性、实时性以及可部署性的统一。

4) 多模态大模型驱动的鲁棒协同检测

现阶段,大模型驱动的多模态检测模型在表面缺陷检测中的应用仍较有限,普遍面临领域先验不足、模型复杂度高等问题。未来研究应注重多模态大模型与表面缺陷检测任务的深度适配,增强模型对真实复杂工况的理解能力。同时结合参数高效微调、迁移学习与领域对齐等机制,提升模型对跨场景变化的适应能力,并借助大模型统一语义建模能力来强化图像、三维信息、热红外及文本数据之间的协同表征,实现多模态表面缺陷检测系统向更高鲁棒

表 11 面向多模态表面缺陷检测系统的分层鲁棒性评价

Table 11 Hierarchical robustness evaluation framework for multimodal surface defect detection systems			
鲁棒性问题	扰动设置	评价指标	评价重点
感知层-数据噪声	模拟高斯噪声、椒盐噪声等,对数据集动态施加光照、模糊等合成扰动,以模拟实际干扰	I-AUROC、P-AUROC、AUPRO、F1、mAP等	评价模型在输入质量退化条件下的检测稳定性
感知层-模态缺失	模拟单模态缺失、随机模态丢弃、局部遮挡、设置缺失率		评价系统对模态不完整输入的容错能力
交互层-对齐失稳	人为引入平移、旋转、尺度扰动、局部错位、时间延迟	AUPRO、定位误差、配准误差	评价对齐误差对定位、注意力与决策的影响
交互层-融合偏差	降低某一模态质量、改变融合权重、遮蔽关键模态、构造模态冲突样本	模态贡献度、置信度校准误差	评价模态贡献失衡、信息冲突与融合稳定性
大模型-生成可靠性	提示词改写、语义噪声、干扰性上下文、缺陷类别相近描述	分类准确率、回答一致性、幻觉率、事实一致性	评价大模型在缺陷识别、定位解释和生成描述中的可靠性

性、更强泛化能力的智能检测框架发展。

5) 面向多模态缺陷检测鲁棒性的专用数据集与评价基准构建

现有多模态表面缺陷检测数据集尚未形成面向系统鲁棒性的统一基准,普遍缺乏针对噪声干扰、模态缺失、对齐失稳、融合偏差以及大模型生成可靠性下降等问题的标准化扰动设计与测试协议等。未来可通过引入可控噪声退化、模态缺失率及模态冲突等扰动机制,构建覆盖感知退化、交互失稳与生成可靠性下降等失效场景的鲁棒性测试数据集。同时,还应建立与扰动类型相匹配的评价标准,如在 I-AUROC、P-AUROC、F1、mAP 等指标基础上,引入性能下降率、跨模态一致性误差、配准误差、模态贡献稳定性等指标,从而推动多模态缺陷检测鲁棒性评价由单一性能比较向专业性、系统级的可靠性评估方向转变。

7 结 论

多模态表面缺陷检测系统中,多源异构数据的引入增强了表面缺陷检测系统的感知能力,但同时也面临多模态数据噪声、模态质量退化、多模态对齐失稳等鲁棒性问题,成为制约多模态表面缺陷检测工程化应用的重要因素。围绕这一问题,本文系统综述了近年来相关研究成果,构建了多模态缺陷检测鲁棒性完整研究框架。本文首先明确了多模态表面缺陷检测系统的架构及鲁棒性的核心定义,分析了影响系统鲁棒性的关键因素。在此基础上,一方面,针对多模态数据噪声、模态质量退化、多模态交互偏差等鲁棒性挑战,从感知层与交互层两个维度深入剖析了各类鲁棒性问题与挑战,并梳理出不同问题与挑战下的强化策略与应对机制,为实际场景中鲁棒性问题解决提供了可落地的方法论指导;另一方面,围绕大模型驱动的多模态表面缺陷检测系统,探讨了其鲁棒性制约因素、优化路径及典型场景应用。此外,本文归纳了相关数据集与评价指标,为鲁棒性技术的验证、对比和优化提供参考。综上,本文立足于多模态缺陷检测的工程化与可靠性应用需求,聚焦多模态系统鲁棒性这一核心目标,系统总结了该领域鲁棒性研究的关键问题、技术路径与评价方法,为后续多模态系统鲁棒性技术的深入研究提供了较为全面的理论参考与思路借鉴,为制造业质

量控制水平的智能化提升提供有力支撑。

参考文献 (References)

- Almukhtar N and Kurnaz S. 2026. Multimodal fusions for defect detection of photovoltaic panels by mask R-CNN and hawkfish optimization algorithm. *Frontiers in Earth Science*, 13: 1702396 [DOI: 10.3389/feart.2025.1702396]
- Asad M, Azeem W, Jiang H, Mustafa H T, Yang J and Liu W. 2025. 2M3DF: Advancing 3D industrial defect detection with multi perspective multimodal fusion network. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(7): 6803-6815 [DOI: 10.1109/TCSVT.2025.3536475]
- Bai Y, Wang L, Gao W and Ma Y. 2024. Multi-modal hierarchical classification for power equipment defect detection. *Journal of Image and Graphics*, 29(7): 2011-2023 (白艳峰, 王立彪, 高卫东, 马应龙. 2024. 面向电力设备缺陷检测的多模态层次化分类. *中国图象图形学报*, 29(7): 2011-2023) [DOI: 10.11834/jig.230269]
- Bergmann P, Jin X, Sattlegger D and Steger C. 2021. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization [E/OL]. [2021-12-16]. <https://arxiv.org/pdf/2112.09045.pdf>
- Bi L, Zhang Y, Wang L, Niu Y and Zhao H. 2025. Two challenges, one solution: Robust multimodal learning through dynamic modality recognition and enhancement. *Findings of the Association for Computational Linguistics: EMNLP 2025. Association for Computational Linguistics (ACL)*: 12855-12867 [DOI: 10.18653/v1/2025.findings-emnlp.689]
- Bonfiglioli L, Toschi M, Silvestri D, Fioraio N and Gregorio D D. 2022. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. *Proceedings of the Asian Conference on Computer Vision*: 3586-3602 [DOI: 10.1007/978-3-031-26348-4_27]
- Chen Z, Chen H, Imani M and Imani F. 2025. Can multimodal large language models be guided to improve industrial anomaly detection [EB/OL]? [2025-01-27]. <https://arxiv.org/pdf/2501.15795.pdf>
- Chen Z and Imani F. 2026a. A multi-expert framework for enhancing multimodal large language models in industrial anomaly detection. *Pattern Recognition*, 172: 112752 [DOI: 10.1016/j.patcog.2025.112752]
- Chen Z, Xie Q, Zhang L, Si B, Shen Y, Miao J and Hong J. 2026b. Dynamic visual defect detection system for disc castings based on multimodal fusion strategy. *Knowledge-Based Systems*, 336: 115325. [DOI: 10.1016/j.knosys.2026.115325]
- Cheng H, Luo J and Zhang X. 2025. Multimodal industrial anomaly detection via uni-modal and cross-modal fusion. *IEEE Transactions on Industrial Informatics*, 21(6): 5000-5010 [DOI: 10.1109/TII.2025.3552723]

- Cui J, Li W, Zhou S, Li W, Qin R and Shen F. 2026. Cross-modal alignment and fusion for RGB-D transmission-line defect detection [EB/OL]. [2026-02-21].
<https://arxiv.org/pdf/2602.01696.pdf>
- Daci R, Reno V, Patruno C, Cardellicchio A, Taleb-Ahmed A, Leo M and Distanto C. 2026. Cross-modal mapping and dual-branch reconstruction for 2D-3D multimodal industrial anomaly detection [EB/OL]. [2026-03-04]
<https://arxiv.org/pdf/2603.03939.pdf>
- Deng H, Luo H, Zhai W, Guo Y, Cao Y and Kang Y. 2025. Vmad: Visual-enhanced multimodal large language model for zero-shot anomaly detection. *IEEE Transactions on Automation Science and Engineering*, 23: 3607-3618 [DOI: 10.1109/TASE. 2025. 3591656]
- Dong Y, Hou H, Gong M and Chen Z. 2026. Multimodal industrial defect detection and intelligent decision-making driven by large vision-language models. *Computer Engineering and Application*, 62 (8): 176-188 (董禹彤, 侯惠芳, 龚明明, 陈自康. 2026. 视觉语言大模型驱动的多模态工业缺陷检测与智能决策. *计算机工程与应用*, 62 (8): 176-188) [DOI: 10.3778/j. issn. 1002-8331. 2507-0389]
- Drenkow N and Unberath M. 2024. RobustCLEVR: A benchmark and framework for evaluating robustness in object-centric learning. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*: 4518-4527 [DOI: 10.1109/WACV57701.2024.00446]
- Du Z, Ma Z, Yang Y, Zhang D, Song H, Zhang X and Zhang Y. 2026. Research on a unified multi-type defect detection method for lithium batteries throughout their entire lifecycle based on multimodal fusion and attention-enhanced YOLOv8. *Sensors*, 26 (2): 635 [DOI: 10.3390/s26020635]
- Feng M, Shen J, Wu Z, Peng W, Zhong H, Guo Y, Shu X, Zhang H, Dong W and Wang Y. 2025. Advancements in 3D vision understanding using multimodal large language models. *Journal of Image and Graphics*, 30(6): 1744-1791 (冯明涛, 沈军豪, 武子杰, 彭伟星, 钟杭, 郭裕兰, 舒祥波, 张辉, 董伟生, 王耀南. 2025. 多模态大模型驱动的三维视觉理解技术前沿进展. *中国图象图形学报*, 30(6): 1744-1791) [DOI: 10.11834/jig.240588]
- Gu Z, Zhu B, Zhu G, Chen Y, Tang M and Wang J. 2024. Anomalygpt: Detecting industrial anomalies using large vision-language models. *Proceedings of the AAAI conference on artificial intelligence*, 38 (3): 1932-1940 [DOI: 10.1609/aaai.v38i3.27963]
- Guo X, Li Y, Cui S, Xu Y and Zhong R Y. 2026. Knowledge-guided multimodal LLM with mixture-of-experts for interpretable defect detection and fault diagnosis: An industrial AI copilot. *Journal of Manufacturing Systems*, 85: 662-677 [DOI: 10.1016/j.jmsy.2026. 02.017]
- Ha T, Hwang C and Jeong J. 2026. CLIP-MDC: CLIP encoder based multimodal defect classification with synthetic anomaly generation for real-time surface defect detection[EB/OL]. [2026-01-17].
Journal of Intelligent Manufacturing. [DOI: 10.1007/s10845-025-02773-4]
- Han Y, Liu T and Li K. 2025. Fault diagnosis of wind turbine blades under wide-weather multi-operating conditions based on multimodal information fusion and deep learning [EB/OL]. [2025-05-05].
Structural Health Monitoring. [DOI: 10.1177/1475921725133761]
- He X, Tan J, Hu T and Zhu L. 2025a. Multimodal segmentation for photovoltaic module defect detection. *IEEE Access*, 13: 208652-208665 [DOI: 10.1109/ACCESS.2025.3640579]
- He Y, He Y, Yin Y, Deng B and Wang Y. 2025b. Application status and prospects of large models in the field of industrial defect detection. *Chinese Journal of Scientific Instrument*, 46(10): 22-41. (何毓芬, 何赞泽, 尹勇, 邓堡元, 王耀南. 2025. 大模型在工业缺陷检测领域的应用现状与展望. *仪器仪表学报*, 46(10): 22-41) [DOI: 10.19650/j.cnki.cjsi.J2513981]
- Hoang D C, Tan P X, Nguyen A N, Duong T H A, Huynh T M, Nguyen D M, Cao M D, Ngo D H Nguyen T H, Phan K T, Do M Q, Dinh X T, Duong V H, Hoang N A and Nguyen V T. 2025. Accurate industrial anomaly detection with efficient multimodal fusion. *Array*, 28: 100512 [DOI: 10.1016/j.array.2025.100512]
- Hoang D C, Tan P X, Nguyen A N, Ngo D H, Cao M D, Vu M Q, Duong H N, Nguyen S H, Le T H, Dang V V, Dinh X T, Nguyen M A, Do M Q, Giap V K and Duong V H. 2026. MambaAlign: Alignment-aware state-space fusion for RGB-X industrial anomaly detection. *Journal of Computational Design and Engineering*, 13 (1): 514-537 [DOI: 10.1093/jcde/qwaf143]
- Hsu C C, Lee C M, Sun C H and Wu K M. 2024. OCR is All you need: Importing multi-modality into image-based defect detection system [EB/OL]. [2024-03-18].
<https://arxiv.org/pdf/2403.11536.pdf>
- Ismithdeen M I, Khattak M U, and Khan S. 2025. Promptception: How sensitive are large multimodal models to prompts [EB/OL]? [2025-09-04].
<https://arxiv.org/pdf/2509.03986.pdf>
- Jiang J, He Z, Wan A, Al-Bukhaiti K, Wang K, Zhu P and Cheng X. 2025a. Zero-shot industrial anomaly detection via CLIP-DINOv2 multimodal fusion and stabilized attention pooling. *Electronics*, 14 (24): 4785 [DOI: 10.3390/electronics14244785]
- Jiang S, Ma Y, Zhou J, Bian Y, Wang Y and Liu M. 2025b. Resilient multimodal industrial surface defect detection with uncertain sensors availability. *IEEE/ ASME Transactions on Mechatronics*, 30 (6): 4261-4271 [DOI: 10.1109/TMECH.2025.3607147]
- Jiang Y, Cao Y, Cheng Y, Zhang Y and Shen W. 2026a. VTFusion: A vision-text multimodal fusion network for few-shot anomaly detection[EB/OL]. [2026-01-21].
IEEE Transactions on Cybernetics. [DOI: 10.1109/TCYB.2026. 3651630]

- Jiang Y, Jin Q, Lu X, Lyu X, He Z and Yin X. 2026b. Fabgpt: An efficient large multimodal model for complex wafer defect knowledge queries[EB/OL]. [2026-02-12].
- IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. [DOI: 10.1109/TCAD.2026.3664304]
- Karimi N, Mishra M and Lourenco P B. 2024. Deep learning-based automated tile defect detection system for Portuguese cultural heritage buildings. *Journal of Cultural Heritage*, 68: 86-98 [DOI: 10.1016/j.culher.2024.05.009]
- Lai Y S, Hsieh C C, Liao T W, Huang C Y and Jeffrey Kuo C F. 2025. Deep learning-based automatic defect detection of photovoltaic modules in infrared, electroluminescence, and red - green - blue images. *Energy Conversion and Management*, 332: 119783 [DOI: 10.1016/j.enconman.2025.119783]
- Lee M and Choi J. 2026. ID-RAG: industrial defect retrieval-augmented generation for industrial surface defect detection. *Machine Vision and Applications*, 37 (2) : 36 [DOI: 10.1007/s00138-026-01797-x]
- Li J, Gao B, Woo W L, Xu J, Liu L and Zeng Y. 2023a. A novel multi-spectral fusion defect detection framework with coarse-to-fine multi-spectral registration. *IEEE Transactions on Instrumentation and Measurement*, 73: 5005313 [DOI: 10.1109/TIM.2023.3344145]
- Li J, Li D, Savarese S, and Hoi S. 2023b. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *International conference on machine learning*. PMLR, 202: 19730-19742 [DOI: 10.5555/3618408.3619222]
- Li S, Sun W, Liang Q, Sun J and Liu C. 2025a. Multimodal fabric defect classification using channel switching and multiscale feature fusion. *IEEE Transactions on Instrumentation and Measurement*, 74: 2501812 [DOI: 10.1109/TIM.2024.3470964]
- Li W, Zheng B, Xu X, Gan J, Lu F, Li X, Ni N, Tian Z, Huang X, Gao S and Wu Y. 2025a. Multi-sensor object anomaly detection: Unifying appearance, geometry, and internal properties. *Proceedings of the Computer Vision and Pattern Recognition Conference* : 9984-9993 [DOI: 10.1109/CVPR52734.2025.00933]
- Li X, Jing J, Wu T, Zhang X and Liu W. 2026a. Supad: a superordinary zero-shot industrial anomaly detection network based on gated-agnostic multimodal adaptive learning prompts[EB/OL]. [2026-01-08]. *Journal of Intelligent Manufacturing* [DOI : 10.1007/s10845-025-02757-4]
- Li Y, Wang H, Yuan S, Liu M, Zhao D, Guo Y, Xu C, Shi G and Zuo W. 2025b. Myriad: Large multimodal model by applying vision experts for industrial anomaly detection[EB/OL]. [2025-01-17]. <https://arxiv.org/pdf/2310.19070.pdf>
- Li Y, Wang Z, Li Y, Liu C, Liu J, Hu X and You X. 2026b. Multimodal optical microscopy defect detection method for indium bump arrays based on multi-scale Demons deformation field difference. *Journal of Microscopy*, 302 (2) : 160-170 [DOI: 10.1111/jmi.70061]
- Li Y, Yuan S, Wang H, Li Q, Liu M, Xu C, Shi G and Zuo W. 2025c. Triad: Empowering LMM-based anomaly detection with expert-guided region-of-interest tokenizer and manufacturing process. *Proceedings of the IEEE/CVF International Conference on Computer Vision*: 21917-21926 [DOI: 10.48550/arXiv.2503.13184]
- Liao J, Xu X, Su Y, Tu R C, Liu Y, Tao D and Yang X. 2025. Robust distribution alignment for industrial anomaly detection under distribution shift[EB/OL]. [2025-03-19] <https://arxiv.org/pdf/2503.14910.pdf>
- Lin Y, Lin W and Lin Y. 2025. Real-time multimodal defect detection and mes feedback on edge devices via bayesian fusion and causal adaptation. *IEEE Access*, 13: 162226-162237 [DOI: 10.1109/ACCESS.2025.3610341]
- Liu T and Song J. 2025a. Defect detection for photovoltaic panel components integrating multimodal feature and improved YOLOv12. *Computer Engineering and Applications*, 62(7) : 107- 120 (刘婷婷, 宋家友. 2025a. 融合多模态特征与改进 YOLOv12 的光伏电池板组件缺陷检测. *计算机工程与应用*, 62(7): 107- 120) [DOI: 10.3778/j.issn.1002-8331.2506-0144]
- Liu Y and Gao Y. 2025b. Multi-modal dynamic fusion for defect detection in electronic products: a novel approach based on energy and deep learning. *IEEE Access*, 13: 118565-118573 [DOI: 10.1109/ACCESS.2025.3584551]
- Lu M, Chai Y, Xu K, Chen W, Ao F and Ji W. 2025. Multimodal fusion and knowledge distillation for improved anomaly detection. *The Visual Computer*, 41(8) : 5311-5322 [DOI: 10.1007/s00371-024- 03723-6]
- Maeda T, Takada D, Kimpar S, Itou H and Arima S. 2024. Multimodal data analysis for superposition defect classification: case study of semiconductor defect patterns. *International Conference on Sustainable Design and Manufacturing*. Singapore: Springer Nature Singapore, 112: 21-34 [DOI: 10.1007/978-981-96-4459-9_3]
- Maheshwari H, Liu Y C and Kira Z. 2024. Missing Modality Robustness in Semi-Supervised Multi-Modal Semantic Segmentation. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV): 1020-1030 [DOI: 10.1109/WACV57701.2024.00106]
- Mai L, Dai C, Ma H, Lin X, Guo S and Yan G. 2025. Masked cross-modal reconstruction network (MCR-Net) for multi-modal industrial anomaly detection. 2025 IEEE 3rd International Conference on Computer, Vision and Intelligent Technology (ICCVIT). IEEE: 1-6 [DOI: 10.1109/ICCVIT67848.2025.11391423]
- McKinzie B, Shankar V, Cheng J, Yang Y, Shlens J and Toshev A. 2023. Robustness in multimodal learning under train-test modality mismatch. *Proceedings of the 40th International Conference on Machine Learning (PMLR)*, 202: 24291-24303 [DOI: 10.5555/3618408.3619419]
- Miao B, Zhang W, Li J, Wu W, Tang S, Li Z, Shi H, Xiao J and Zhuang Y. 2025. Robust modality-incomplete anomaly detection: a modality-Instructive framework with benchmark. *Proceedings of the*

- 33rd ACM International Conference on Multimedia: 7317-7326 [DOI: 10.1145/3746027.3754766]
- Mistry D, Qiu F, Chen B, Liu F, Chen C, Shahidehpour M and Wang R. 2026. From prediction to diagnosis: Reasoning-aware AI for photovoltaic defect inspection[EB/OL]. [2026-03-24]. <https://arxiv.org/pdf/2603.26776.pdf>
- Na Y, He Y, Deng B, Yang C, Li Q, Wang L and Cao Y. 2026. A deep mutual learning-based framework for wind turbine blade defect detection in multimodal phased array ultrasonic data. *Ultrasonics*, 164: 108035 [DOI: 10.1016/j.ultras.2026.108035]
- Ouda E, Salah M, Chulkov A O, Gargiulo G, Tartaglia G L, Sfarra S and Abdulrahman Y. 2026. Vision-language based expert reporting for painting authentication and defect detection [EB/OL]. [2026-03-13]. <https://arxiv.org/pdf/2603.13437.pdf>
- Qu H and Lin J. 2025. Multimodal multiscale industrial anomaly detection via flows. *Journal of Image and Graphics*, 30(2): 451-466 (曲海成, 林俊杰. 基于归一化流的多模态多尺度工业场景缺陷检测. 中国图象图形学报, 2025, 30(2): 451-466) [DOI: 10.11834/jig.240183]
- Rahmati M and Rahmati N. 2025. A multimodal deep learning framework for real-time defect recognition in industrial components using visual, acoustic and vibration signals. *Journal of Intelligent Manufacturing and Special Equipment*, 6(3): 273-292 [DOI: 10.1108/JIMSE-07-2025-0015]
- Reich B, Kunda M, Zolotarev F, Eerola T, Zemčík P and Kauppi T. 2025. Multimodal surface defect detection from wooden logs for sawing optimization[EB/OL]. [2025-03-27]. <https://arxiv.org/pdf/2503.21367.pdf>
- Ren C, Li M, Chen C, Guan X and Huang G Q. 2026. Multi-modal digital twins for industrial anomaly detection: framework, method, and application. *Robotics and Computer-Integrated Manufacturing*, 97: 103068 [DOI: 10.1016/j.rcim.2025.103068]
- Salah M, Ouda E, Dell'avvocato G, Sarasini F, D'Accardi E, Dias J, Svetinovic D, Sfarra S and Abdulrahman Y. 2026. Towards cognitive defect analysis in active infrared thermography with vision-text cues[EB/OL]. [2026-03-11]. <https://arxiv.org/pdf/2603.10549.pdf>
- Shangguan W, Wu H, Niu Y, Yin H, Yu J, Chen B and Huang B. 2025. CIPR: Multimodal industrial anomaly detection via latent bridged cross-modal prediction and intra-modal reconstruction. *Advanced Engineering Informatics*, 65: 103240 [DOI: 10.1016/j.aei.2025.103240]
- Shi Y, Zhan X, Li Y, Wu Z, Zhang W and Li H. 2025. Cycle-CFM: An unsupervised framework for robust multimodal anomaly detection in industrial settings. *Expert Systems with Applications*, 298: 129745 [DOI: 10.1016/j.eswa.2025.129745]
- Tang K, Li Y, Zi B and Feng K. 2025. A multimodal unsupervised coating defect detection method based on dual-branch hybrid CNN-Mamba UNet. *IEEE Transactions on Instrumentation and Measurement*, 74: 5026617 [DOI: 10.1109/TIM.2025.3556907]
- Tang S, Fei F, Zhang L and Yu J. 2026. Multimodal data fusion for welding defect detection using ensemble deep learning. *Automation in Construction*, 182: 106694 [DOI: 10.1016/j.autcon.2025.106694]
- Tocchetti A, Corti L, Balayn A, Yurrita M, Lippmann P, Brambilla M and Yang J. 2025. AI robustness: a human-centered perspective on technological challenges and opportunities. *ACM Computing Surveys*, 57(6): 1-38 [DOI: 10.1145/3665926]
- Wang C, Chen J and Zhang H. 2026a. DCRDF-Net: A dual-channel reverse-distillation fusion network for 3D industrial anomaly detection. *Sensors*, 26(2): 412 [DOI: 10.3390/s26020412]
- Wang C and Zhang H. 2026b. FMFR: Feature-level multistage fusion and remapping for multimodal industrial anomaly detection. *Journal of Computational Design and Engineering*, 13(3): 233-245 [DOI: doi.org/10.1093/jcde/qwag016]
- Wang C, Zhu H, Peng J, Wang Y, Yi R and Wu Y. 2025a. M3dm-nr: Rgb-3d noisy-resistant industrial anomaly detection via multimodal denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11): 9981-9993 [DOI: 10.1109/TPAMI.2025.3592089]
- Wang J. 2025b. Development of an online defect detection system for additive manufacturing based on multi-sensor fusion. 2025 2nd International Conference on Advanced Computer Applications and Artificial Intelligence (ACAAI 2025), 80: 03007 [DOI: 10.1051/itmconf/20258003007]
- Wang Q, Wang X, He Q, Huang J, Huang H, Wang P, Yu T and Zhang M. 2024a. 3D tensor-based point cloud and image fusion for robust detection and measurement of rail surface defects. *Automation in Construction*, 161: 105342 [DOI: 10.1016/j.autcon.2024.105342]
- Wang S, Xu K, Li B and Cao X. 2023a. Online micro defects detection for ductile cast iron pipes based on twin light photometric stereo. *Case Studies in Construction Materials*, 19: e02561 [DOI: 10.1016/j.cscm.2023.e02561]
- Wang W, Gao P, Chen D, Yu R, Kang H and Zhao Z. 2026c. Visible-infrared dual-modal monitoring system for overlap defects in wire arc additive manufacturing. *Materials*, 19(5): 899 [DOI: 10.3390/ma19050899]
- Wang X, Pan J, Ding L and Biemann C. 2024b. Mitigating hallucinations in large vision-language models with instruction contrastive decoding. *Findings of the Association for Computational Linguistics: ACL 2024*. 2024: 15840-15853. [DOI: 10.18653/v1/2024.findings-acl.937]
- Wang Y, Peng J, Zhang J, Yi R, Wang Y and Wang C. 2023b. Multimodal industrial anomaly detection via hybrid fusion. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*: 8032-8041 [DOI: 10.1109/CVPR52729.2023.00776]

- Wang Y, Yang X, Yuan C, Sun H, Yu X and Rodriguez-Andina J J. 2026d. Solder paste segmentation method based on RGB and 3-D height modality feature fusion[EB/OL]. [2026-02-12]. IEEE Transactions on Industrial Informatics. [DOI: 10.1109/TII.2026.3659269]
- Wang Y, Yu W, Yu C, Shi H and Li W. 2026e. A survey on multimodal learning for industrial diagnostics: a data dimensionality perspective. *Artificial Intelligence Review*, 59: 126. [DOI: 10.1007/s10462-026-11527-4]
- Woo S, Lee S, Park Y, Nugroho M A and Kim C. 2023. Towards good practices for missing modality robust action recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37 (3): 2776-2784 [DOI: 10.1609/aaai.v37i3.25378]
- Wu F, Wang Z and Li L. 2026a. DFRF-MIAD: Multimodal industrial anomaly detection via feature reconstruction and fusion. *International Conference on Multimedia Modeling*. Singapore: Springer Nature Singapore: 119-133 [DOI:10.1007/978-981-95-6957-1_9]
- Wu H, Jia D, Zhang T, Bai X, Sun L and Pu M. 2025. Multimodal zero-shot anomaly detection using dual-experts for electrical power equipment inspection images. *Journal of Image and Graphics*, 30 (3): 672-682. (吴华, 贾栋豪, 张婷婷, 白晓静, 孙笠, 蒲梦杨. 2025. 基于双专家的巡检影像多模态零样本缺陷检测. *中国图象图形学报*, 30(3): 672-682) [DOI: 10.11834/jig.240246.]
- Wu J, Chen Y, Min Q, Chen M, Zhao J and Ye M. 2026b. Domain-adaptive multimodal large language models for photovoltaic fault diagnosis via dynamic LoRA routing. *Processes*, 2026, 14 (4): 653 [DOI: 10.3390/pr14040653]
- Wu R, Wang H, Chen H T and Carneiro G. 2026c. Deep multimodal learning with missing modality: A survey[EB/OL]. [2026-02-04]. <https://arxiv.org/pdf/2409.07825.pdf>
- Xu J, Yuan J, Yang M and Yan W. 2025. Modality-resilient multimodal industrial anomaly detection via cross-modal knowledge transfer and dynamic edge-preserving voxelization. *Sensors*, 25(21): 6529 [DOI: 10.3390/s25216529]
- Yin S, Fu C, Zhao S, Li K, Sun X, Xu T and Chen E. 2024. A survey on multimodal large language models. *National Science Review*, 11 (12): nwae403 [DOI: 10.1093/nsr/nwae403]
- Zhang A, Chen K, Li B, Song X, Li N and Xie M. 2026a. Design of defect detection system based on multi-modal cascade for laser metal wire deposition additive manufacturing. *Measurement*, 267: 120518 [DOI: 10.1016/j.measurement.2026.120518]
- Zhang Q, Zhao Z, Wang Z, Wan Z, Wang H and Li G. 2026b. Deep multimodal fusion of spectral and visual data for laser welding defect classification. *Engineering Applications of Artificial Intelligence*, 165: 113473 [DOI: 10.1016/j.engappai.2025.113473]
- Zhao C, Sun J, Ma Z, Yang Z and Qiu X. 2025a. A multimodal information fusion method for real-time monitoring of glass cutting with laser-induced thermal-crack propagation. *Optics & Laser Technology*, 180: 111457 [DOI: 10.1016/j.optlastec.2024.111457]
- Zhao S, Lin Y, Han L, Zhao Y and Wei Y. 2025b. Omniad: Detect and understand industrial anomaly via multimodal reasoning[EB/OL]. [2025-05-28] <https://arxiv.org/pdf/2505.22039.pdf>
- Zheng X, Lyu Y, Jiang L, Paudel D P, Gool L V and Hu X. 2025. Reducing unimodal bias in multi-modal semantic segmentation with multi-scale functional entropy regularization. *Proceedings of the IEEE/CVF International Conference on Computer Vision*: 21166-21176 [DOI: 10.48550/arXiv.2505.06635]
- Zhong Q, Qiu X, Wang X, Zhang Z, Liu G and Xiao J. 2025. FAMRD: Frequency-aware multimodal reverse distillation for industrial anomaly detection. *Proceedings of the 33rd ACM International Conference on Multimedia*: 1862-1871 [DOI: 10.1145/3746027.3755473]
- Zhong W, Feng X, Zhao L, Li Q, Huang L, Gu Y, Ma W, Xu Y and Qin B. 2024. Investigating and mitigating the multimodal hallucination snowballing in large vision-language models. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 1: 11991-12011 [DOI: 10.18653/v1/2024.acl-long.648]
- Zhou W, Wu Y, Qiang F and Yan W. 2025. Lightweight scope integration network for rail surface defect detection. *IEEE Transactions on Big Data*, 12 (2): 321-329 [DOI: 10.1109/TBDATA.2025.3618444]
- Zhu H, Zhang X, Qian Y, Chen J and Huang M. 2025. A hybrid multimodal algorithm for defect inspection in compound semiconductors. *Second International Conference on Optical Communication and Optoelectronic Technology (OCOT 2025)*. SPIE, 13985: 81-87 [DOI: 10.1117/12.3077783]
- Zou L, Chen A, Yang X and Sun Y. 2025a. An improved method of AUD-YOLO for surface damage detection of wind turbine blades. *Scientific reports*, 15 (1): 5833 [DOI: 10.1038/s41598-025-89864-7]
- Zou P. 2025b. Robust multi-modal infrared fault diagnosis for power equipment with cross-device generalization and corruption robustness[EB/OL]. [2025-12-10]. <http://dx.doi.org/10.2139/ssrn.5899048>

作者简介

杨洋,男,博士研究生,主要研究方向为图像处理与机器视觉。E-mail: 1716481354@qq.com

吴一全,通信作者,男,教授,主要研究方向为视觉检测与图像测量、遥感图像处理与理解、视频处理与智能分析等。E-mail: nuaaimage@163.com