

基于无监督学习的工业图像异常检测研究综述

许建^{1,2}, 冯健聪¹, 侯富豪¹, 杨庚^{1,2}

(1. 南京邮电大学计算机学院, 江苏 南京 210023;

2. 南京邮电大学大数据安全与智能处理江苏省高校重点实验室, 江苏 南京 210023)

摘要: 工业4.0推动了智能制造的快速发展, 传统人工质检存在效率低、成本高和稳定性差等问题, 而基于监督学习的工业图像异常检测方法高度依赖精确标注的异常样本, 难以适应真实工业场景中缺陷稀少、多变的特点。相比之下, 基于无监督学习的工业图像异常检测方法仅需利用正常样本建模正常产品的分布特征与内在模式, 因而成为近年来学术界与工业界的研究热点。首先, 围绕单类学习设定对基于无监督学习的工业图像异常检测方法进行系统综述, 并依据异常分数来源将现有方法归纳为特征嵌入和图像重建两大范式, 对相关研究的发展脉络与核心思想进行梳理; 其次, 在上述分类框架下, 对两类方法的技术演进与研究趋势进行归纳, 总结不同方法在特征表示、异常评分机制和定位策略等方面的共性与差异, 并对现有研究成果的主要优势与局限进行综合分析, 从整体上梳理当前基于无监督学习的工业图像异常检测方法的发展特点与研究重点; 再次, 归纳了该领域常用的基准数据集与评价指标; 最后, 从实际工业应用需求出发, 对当前研究面临的主要问题进行了概括, 并对未来的发展方向进行了展望, 以期为该领域的后续研究与应用提供参考。

关键词: 工业图像异常检测; 无监督学习; 单类学习; 特征嵌入; 图像重建

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.AQ25287

Survey on industrial image anomaly detection based on unsupervised learning

Xu Jian^{1,2}, Feng Jiancong¹, Hou Fuhao¹, Yang Geng^{1,2}

1. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2. Jiangsu Key Laboratory of Big Data Security and Intelligent Processing,
Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract: With the rapid advancement of Industry 4.0, Intelligent manufacturing has become increasingly important. Traditional manual inspection still suffers from low efficiency, high cost, and limited stability, while supervised learning methods rely heavily on accurately labeled anomalous samples, which are scarce and diverse in real industrial scenarios. In contrast, unsupervised industrial image anomaly detection learns the distribution and intrinsic patterns of normal samples only, and has therefore attracted growing attention from both academia and industry. First, a systematic review of unsupervised industrial image anomaly detection under the one-class learning setting was provided. Existing methods were categorized into two main paradigms according to the source of anomaly scores: feature embedding and image reconstruction, and the development of representative approaches was summarized. Second, based on this taxonomy, the evolution and research trends of these two paradigms were analyzed.

收稿日期: 2025-12-28; 修回日期: 2026-05-15

通信作者: 许建, xuj@njupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62372244)

Foundation Item: The National Natural Science Foundation of China (No.62372244)

The common characteristics and differences among existing methods in terms of feature representation, anomaly scoring mechanisms, and localization strategies were summarized to provide a structured understanding of current research. Then, commonly used benchmark datasets and evaluation metrics in this field were reviewed. Finally, the main challenges in current research were discussed from the perspective of practical industrial applications, and potential future research directions were outlined to facilitate further development and deployment of unsupervised industrial image anomaly detection methods.

Key words: industrial image anomaly detection, unsupervised learning, one-class learning, feature embedding, image reconstruction

0 引言

在工业 4.0 推动制造业向智能化、数字化和互联化转型的背景下,工业图像异常检测^[1]作为融合物联网、大数据与人工智能技术的关键环节,正日益成为保障生产质量与效率的核心手段。传统的人工检测依赖目视判断,存在检测速度慢、人工成本高、检测人员易疲劳和主观性强等不足,从而导致威胁漏检与误检风险较高。相比之下,基于计算机视觉的图像异常检测技术能够快速、精准地识别产品表面的缺陷(如划痕、裂纹、色差)及设备运行中的异常状态,显著提升生产质量控制水平。目前,该技术已在纺织等行业成功应用于布料瑕疵检测,替代了低效的人工目检^[2]。该技术不仅能够及时发现生产过程中的质量隐患,还可以通过持续积累与分析检测数据,挖掘潜在缺陷的演化趋势,从而为预测性维护提供技术支撑。例如,通过监测设备运行图像识别其早期磨损迹象,提前发出预警以规避生产损失。工业图像异常检测通常需处理高分辨率图像,面临异常样本稀缺、缺陷形态多样、检测环境复杂及精度与实时性难以平衡等挑战。其技术路径涵盖传统图像处理与深度学习两大方向,包含有监督、无监督及半监督 3 种学习模式。目前,该技术已广泛应用于产品缺陷识别、生产质量检验与设备安全监控等工业场景。

然而,工业图像异常检测普遍存在异常样本匮乏、异常类型多变等问题。基于有监督与半监督学习的工业图像异常检测方法高度依赖大规模异常标注数据,数据标注成本高昂;同时,这类方法仅能识别训练集中已出现的异常类型,对未知异常缺乏检测能力,最终导致模型泛化性能较差。相比之下,基于无监督学习的工业图像异常检测方法无须依赖任何异常样本,仅通过对海量

正常样本进行建模,学习其固有的数据分布与特征模式。因此,此类方法省去了采集和标注异常样本的高昂成本。更为关键的是,该类方法仅对正常模式进行建模,任何偏离正常分布的样本(包括未知样本)均能被有效检出,从而确保了检测系统在真实工业场景下的高鲁棒性与泛化能力。基于现有研究成果,本文提出一种以异常分数计算为核心机制的二分范式,并对当前基于无监督学习的工业图像异常检测方法进行了系统性梳理。具体而言,本文将现有方法划分为特征嵌入与图像重建两大类。其中,特征嵌入范式方法在提取的特征空间中,通过距离度量、概率密度估计等方式,直接对正常样本的分布进行建模;异常分数由测试样本特征与正常特征之间的特征距离等偏差指标直接计算得到,整个过程不涉及显式的图像重建。图像重建范式方法的异常分数则通过计算重建图像与正常图像之间的像素残差、特征相似度等重建误差获得。该分类方法以异常分数的计算方式与来源为依据,有效避免了多维度交叉分类带来的复杂性。对于部分同时包含特征对齐与重建机制的混合方法(如逆向蒸馏、部分条件扩散模型等),本文依据其异常分数的来源进行归类。若最终的异常判定依赖重建误差,则归入图像重建范式;若最终的异常判定主要依靠特征空间中的直接分布比较,则归入特征嵌入范式。

基于上述分类框架,本文对现有主流基于无监督学习的工业图像异常检测方法进行了汇总,具体见表 1。特征嵌入范式方法通过预训练模型提取正常样本特征并构建参考模型,通过比较测试样本特征与正常特征的差异实现异常的检测和定位。该类方法高效提取高层次语义特征的优势,使其尤其适合需要快速推理或在边缘设备部

表1 基于无监督学习的工业图像异常检测方法汇总

检测思想	方法	技术原理	优点	缺点	适用场景
基于特征嵌入	知识蒸馏	使用预训练的教师模型指导学生模型学习对正常样本的特征提取	异常检测高效，模型轻量化	需设计合适的损失函数与训练策略，蒸馏过程复杂	算力有限的边缘设备
	单分类	通过寻找一个超球面，实现精准区分正常样本和异常样本	模型专注于正常样本的分布或特征，训练过程相对简单	在高维或者复杂图像上准确定义正常样本的边界具有挑战性	正常样本特征分布清晰（如金属表面缺陷检测）
	归一化流	学习正常样本的特征分布，并将测试样本和该分布进行比较	方法基于概率分布或者统计建模，具有较强的数学解释性	某些基于深度学习的分布映射方法训练和推理计算成本高	工业图像具有高维或者复杂特征
	记忆库	将测试图像和记忆库中的正常特征进行比较	存储的正常特征多样，为异常检测提供参考	存储空间需求大，高维特征比较计算开销大	产品类型单一的异常检测场景
	视觉语言模型	结合图像和语言信息强化检测能力	能够融合语言和图像，具备更丰富的语义理解能力	高度依赖语言提示的质量	正常样本稀少场景
基于图像重建	逆向蒸馏	教师编码器提取特征，单类瓶颈嵌入模块压缩信息，学生解码器重建特征	能够更好地捕捉语义信息，同时能够精确定位异常区域	高度依赖教师网络的特征提取能力，对环境噪声敏感	结构化产品和表面缺陷检测、多种产品场景
	自编码器	先通过编码器压缩图像特征，再使用解码器进行图像重建	通过提取图像的低维特征表示，捕捉正常样本的复杂模式	当异常特征空间与正常特征空间接近时，易造成误检	多阶段异常检测场景
	生成对抗网络	生成器与判别器通过对抗训练学习数据分布	能够生成与正常样本高度相似的图像、学习复杂的正常样本分布	对抗训练易出现模式崩溃或收敛困难	需要细粒度检测的场景
	Transformer	将Transformer用于编码—解码过程，利用自注意力机制捕捉全局特征与局部特征	Transformer的自注意力机制能够捕捉图像中的全局上下文和长距离依赖关系	全局注意力机制对微小或局部异常不够敏感	航空部件等具有复杂纹理、背景或全局结构的场景
	扩散模型	逐步添加噪声并学习逆向去噪过程，建模数据分布同时重建图像	能够生成高保真的正常样本图像、捕捉复杂的纹理和结构	需要进行多次迭代，计算复杂度高，推理时间长	高端制造等正常样本主导的生产场景

署的工业场景。图像重建范式方法则通过学习正常样本的重建规律，利用重建残差完成异常检测。该类方法对预训练模型的依赖度较低，且能够精确定位异常区域，适用于复杂纹理或结构化产品的异常检测。

与现有综述研究相比，本文采用更简洁、更贴合技术本质的二分分类法，将现有基于无监督学习的工业图像异常检测方法严格划分为特征嵌入与图像重建两类，并将其作为全文最高层的逻辑主线。该分类方式显著区别于同期综述采用的多维度、较分散的分类思路，使研究脉络更加清晰。同时，本文覆盖了近年来该领域的前沿技术，针对视觉语言模型、逆向蒸馏、最新扩散模型变种等新兴的研究方向进行了系统梳理，增强了技术演进的连贯性。综上，本文重点梳理了特

征嵌入与图像重建这两类基于无监督学习的工业图像异常检测方法的研究进展，并探讨了其未来的发展方向。

本文的主要贡献如下。

1) 将现有基于无监督学习的工业图像异常检测方法划分为特征嵌入和图像重建两大类，详细梳理了两类方法的技术原理、优缺点和适用场景，为相关研究者提供了清晰、直观的技术路线图。

2) 全面总结了工业图像异常检测领域常用的数据集，涵盖纹理、结构、微小缺陷等场景，并详细介绍了该领域的评价指标，为模型的性能评估和数据集的选择提供了标准化的参考依据。

3) 分析了现有基于无监督学习的工业图像异常检测方法面临的挑战（如跨产线数据分布差

异、像素级异常定位精度不足、伪异常样本合成逼真度低及工业领域真实异常数据匮乏等), 并提出了未来的研究方向(如多模态融合、开放集异常识别、轻量化模型架构设计和可控异常样本生成), 为工业图像异常检测的发展提供了指导思路。

1 无监督学习背景知识

无监督学习^[3]旨在从完全无标注的数据中自动挖掘隐藏的模式、数据结构或内在表示, 整个学习过程无须依赖任何人工提供的监督信号。这一特性使其高度契合数据标注成本高昂、异常样本稀缺且类型多变的工业图像异常检测场景。在实际的工业生产中, 异常样本通常极为稀少、形态多样且难以预先穷尽, 传统有监督或半监督学习方法因高度依赖精确标注的异常数据而难以大规模落地。相比之下, 无监督学习仅利用大量正常样本进行建模, 通过学习正常数据的固有分布与内在模式, 将任何显著偏离正常模式的样本判定为异常。这种单类学习思想赋予了无监督学习方法检测未知异常的开放集能力, 这也是无监督学习方法在工业质检领域获得广泛关注与应用的核心原因。

1.1 无监督学习技术的发展历程

无监督学习从早期的统计聚类^[4]、数据降维与概率密度估计, 逐步发展到神经网络表示学习(包括自组织映射、受限玻尔茨曼机、深度信念网络等模型), 目前已发展到深度生成模型主导的时代^[5]。变分自编码器(variational autoencoder, VAE)和生成对抗网络(generative adversarial network, GAN)的提出, 标志着生成式无监督学习实现了关键性突破; 而近年来兴起的扩散模型, 通过逐步加噪与逆向去噪过程, 在生成质量和复杂数据分布建模能力上显著超越了早期生成模型^[6]。此外, 自监督学习通过精心设计的代理任务从数据本身挖掘潜在的监督信号, 有效提升了表征学习的质量; 大规模预训练的视觉语言模型则将跨模态知识引入无监督表征学习, 成为近年来该领域的重要发展方向之一。这些技术层面的进步为工业图像异常检测提供了更强大的特征提取、分布建模和图像重建能力, 推动了该领域从传统统计方法向深度学习主导的范式转变。

1.2 工业图像异常检测中的无监督学习

在工业图像异常检测中, 无监督学习遵循单类学习的核心思想, 即仅从正常样本中学习其特征分布或生成规律, 将测试样本与学到的正常模式进行偏差比较, 任何不符合该模式的样本均被判定为异常。这一机制省去了稀有缺陷样本采集与标注带来的高昂成本。更加关键的是, 它对训练过程中未出现过的未知异常具备检测能力, 从而在真实工业环境中展现出更强的鲁棒性和泛化能力。本文所介绍的两大范式下的方法均具备异常检测与定位能力, 后续章节将围绕异常分数的来源与计算机制展开详细梳理, 并在分析过程中重点关注各类方法的异常定位机制与性能表现。

本文所介绍的两大范式下的方法均同时具备图像级异常检测与像素级异常定位能力。其中, 图像级异常检测的核心目标是判断整幅图像是否包含异常, 通常以异常分数或二分类结果(正常/异常)作为输出; 像素级异常定位则要求精确标识异常所在的像素区域, 通常以高分辨率热力图或分割掩码作为输出。该定位能力直接服务于工业质检中的缺陷严重程度分级、异常产品自动剔除决策、返工修复定位及质量问题根因追溯等下游环节, 因此其对空间分辨率保持、异常区域边界精确刻画及微弱异常敏感度的要求, 显著高于图像级异常检测。

本文基于异常分数的来源与计算机制, 将现有主流的基于无监督学习的工业图像异常检测方法划分为两大核心范式, 即特征嵌入范式与图像重建范式。其中, 特征嵌入范式主要依赖预训练模型或专门训练的特征提取器, 在高维特征空间中通过距离度量、概率密度估计或最近邻比较等方式, 直接量化测试样本与正常样本分布的偏差, 其异常分数源于特征空间的直接比较, 具有推理速度快、对预训练表征依赖度较高、适配边缘设备部署的特点。图像重建范式则通过学习正常样本的编码-解码或生成过程, 以重建图像与原始正常图像之间的像素残差、结构相似性或特征相似性作为异常分数的来源, 该范式的优势在于能够更精确地定位异常区域, 尤其适用于复杂纹理或结构化产品的表面缺陷检测, 但通常存在计算开销较大、对重建网络的泛化控制要求较高

的问题。

尽管两大范式在异常量化方式和侧重点上存在明显差异，但它们共同遵循无监督学习的基本原理：即在特征或图像空间中，尽可能地缩小正常样本的类内方差，同时增大正常样本与异常样本的类间距离。后续章节将围绕这两大范式分别展开代表性方法的技术原理、关键创新点、优缺点对比及适用场景的分析。

2 基于特征嵌入范式的工业图像异常检测

基于特征嵌入范式的工业图像异常检测方法通过预训练模型提取正常样本的特征，并构建正常特征的参考模型。如图1所示，这种方法通过量化测试样本与正常特征参考模型之间的特征差异，实现图像异常的判定与异常区域的定位^[7]。目前，采用特征嵌入范式实现图像异常检测的方法主要有基于知识蒸馏、单分类、归一化流、记忆库及视觉语言模型等。这种方法聚焦如何将正常特征通过更加准确、高效的方式提取出来，并改进特征比较策略，以提高模型对异常样本的检测精度及异常区域的定位能力。

2.1 基于知识蒸馏的图像异常检测方法

基于知识蒸馏的图像异常检测方法的核心思想是利用预训练的教师模型（通常为在 ImageNet 数据集上预训练的深度神经网络）提取正常样本的高层语义特征，进而指导结构简单的学生模型学习仅针对正常样本模式的特征表征。在推理阶段，通过比较教师模型与学生模型在多尺度特征上的偏差来计算异常分数，最终实现高效的异常检测。该方法特别适用于产品外观特征相对稳定、缺陷样本与正常样本在高层语义特征上差异显著的工业场景，部署在算力受限的边缘设备上时表现出色，其架构如图2所示。

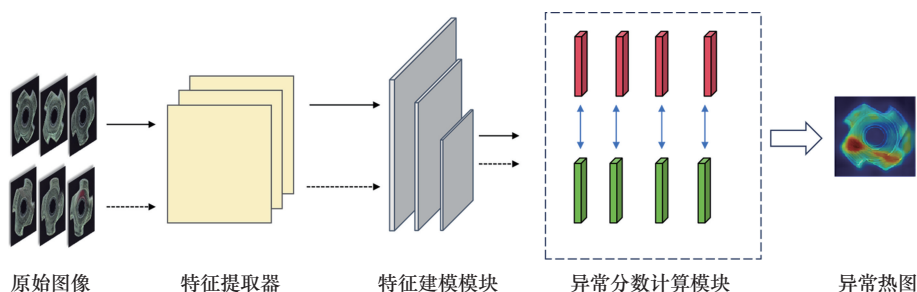


图1 基于特征嵌入范式的工业图像异常检测方法架构

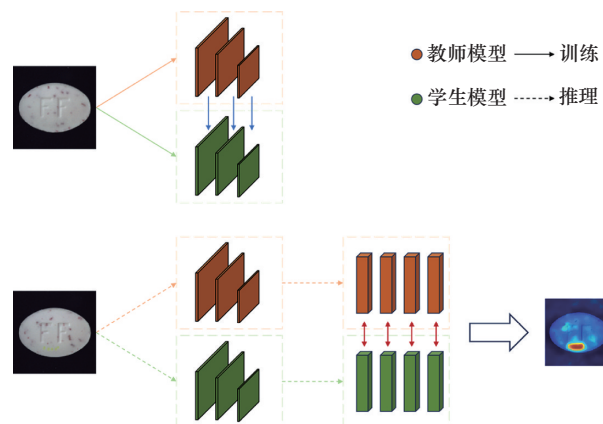


图2 基于知识蒸馏的图像异常检测方法架构

Salehi等^[8]首次将知识蒸馏引入工业图像异常检测领域，采用预训练视觉几何组（visual geometry group, VGG）网络作为教师模型，指导学生模型学习正常图像的特征表征。该方法在多个数据集上的性能表现优于当时的最佳方法。其核心逻辑在于，预训练教师模型具备强大的高层语义特征提取能力，可以引导学生模型聚焦于正常样本模式的学习，在推理阶段通过比较教师模型与学生模型输出特征的偏差来计算异常分数，实现高效、轻量的异常检测。随后，STPM^[9]引入特征金字塔结构聚合多层次蒸馏知识，提升了模型对多尺度异常的识别能力。IKD^[10]通过上下文相似性损失和自适应硬样本挖掘，缓解了过拟合问题。AST^[11]采用非对称教师—学生结构，显著提高了特征区分度。AST-MFR^[12]引入遮蔽特征再生（masking feature regeneration, MFR）机制增强学生模型的学习能力。Zhang等^[13]将注意力机制融入蒸馏框架，充分挖掘教师模型的语义信息，指导学生模型高效检测异常。UniNet^[14]结合对比学习和领域相关特征选择策略，提升了模型的跨领域泛化性能。

如表2所示，这些方法总体上依赖预训练教

表2 基于知识蒸馏的图像异常检测方法总结

方法（提出年份）	教师模型	学生模型	损失函数	辅助模块	特征曲线下面积（AUROC）（MVTec-AD）
STPM ^[9] （2021年）	ResNet-18	ResNet18	L2	金字塔特征嵌入	97.0%
IKD ^[10] （2022年）	HRNet	HRNet	上下文相似性	自适应硬样本挖掘（AHSM）	97.8%
AST ^[11] （2023年）	ReralNVP	前馈网络	负对数似然+L2	无	99.1%
AST-MFR ^[12] （2024年）	WideResNet50	ResNeXt50	L2+余弦相似性	遮蔽特征再生	98.7%
Zhang 等 ^[13] （2024年）	ViT	ViT	多视图一致约束+中心特征约束+编码一致性	中心特征约束+条件注意力掩码	98.9%
UniNet ^[14] （2025年）	WideResNet50	WResNet50	对比损失+边缘损失	多尺度嵌入（MEM）+相关领域特征选择（DFS）	99.9%

师模型的高层语义表征能力，通过优化教师—学生框架的特征不对称或网络结构，迫使学生模型专注于学习正常样本模式。这种方法在推理速度、模型轻量化和多尺度异常适应性方面具备显著优势，特别是通过多尺度特征偏差可以直接生成空间分辨力强的异常热力图，能够支持像素级异常定位，因此特别适用于实时性要求高、正常样本纹理相对一致但缺陷形态多样的工业质检场景。然而，当正常样本变异性极高，或异常样本与正常样本在高层语义特征上高度相似时，模型性能易受教师模型表征能力局限和蒸馏复杂度的影响。近年来，基于知识蒸馏的图像异常检测方法正向非对称结构、注意力增强和领域自适应演进，以更好地平衡精度与部署友好性。

2.2 基于单分类的图像异常检测方法

如图3所示，基于单分类的图像异常检测方法的核心思路是仅利用正常样本在特征空间中构建一个紧凑的决策边界（如最小超球面、超平面或概率密度阈值），将落在决策边界内部的样本判定为正常，任何偏离该边界的样本均被视为异常。这种方法的核心目标是最大化正常样本的类内紧密度，同时最小化决策边界的体积，从而赋予模型对未知异常的开放集检测能力。这种方法适用于正常模式定义清晰、样本表征一致性较强的工业质检场景（如规则几何形状零部件的表面缺陷检测）；但当正常样本自身内在变异性较高，或异常样本与正常样本在低层次特征上高度相似时，决策边界的定义难度会显著增大，进而导致模型检测性能下降。在技术演进方面，早期单分类方法的异常定位能力较弱，通常需要依赖

patch-level扩展策略或后处理热力图生成方法实现像素级定位；而后期发展的伪异常合成技术，有效提升了模型的像素级异常区分能力^[15]。

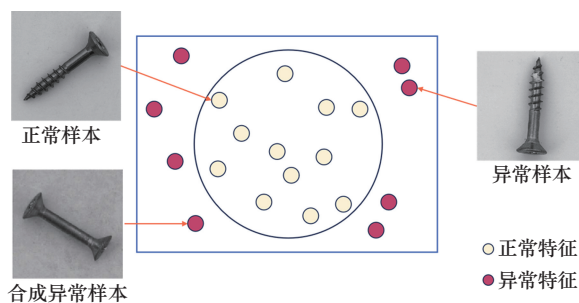


图3 基于单分类的图像异常检测方法架构

早期的单分类方法以Patch SVDD^[16]为代表，其将经典单类支持向量数据描述（SVDD）^[17]扩展至补丁级别，通过对深度特征施加局部紧凑约束，显著提升了异常检测精度。FCDD^[18]使用全卷积网络（fully convolutional network, FCN）保留图像空间信息，实现了更细粒度的异常热力图生成。PANDA^[19]通过引入预训练特征适应机制和异常打分策略，有效缓解了特征崩溃问题，提升了模型检测的稳定性。随着研究的深入，伪异常合成策略逐渐成为提升单分类方法泛化能力的主流技术路径。CutPaste^[20]通过简单高效的Cut-Paste操作合成局部异常样本，成为该方向的基石方法，其思路被后续研究广泛借鉴。SimpleNet^[21]通过在特征层面引入高斯噪声来模拟真实缺陷，既降低了计算开销，又提升了模型的异常检测性能。SuperSimpleNet^[22]不仅在特征层面引入高斯噪声来模拟真实缺陷，还引入Perlin噪声来增强伪异常的多样性，并优化损失函数设计，

实现了有监督与无监督场景下的高性能检测。GLASS^[23]通过同时生成全局与局部伪异常样本，扩大了异常模式的覆盖范围，提高了模型对不同尺寸缺陷的检测鲁棒性。PBAS^[24]采用渐进边界优化结合伪异常引导的策略，逐步细化正常与异常样本的决策边界，实现了更精准的异常区分。

如表3所示，这些方法总体上依赖高质量的伪异常合成来撑开正常类特征边界。这种方法具有训练流程相对简单、对预训练模型依赖度较低的优势，在正常样本高度标准化、生产环境稳定的工业场景中表现突出，尤其适用于金属零部件、电子元件等缺陷形态多样但正常纹理相对一致的产品检测。然而，当正常样本变异性较大或伪异常合成不够逼真时，这种方法的决策边界易出现过紧或过松的问题，进而导致模型误检率或漏检率升高。近年来，基于单分类的图像异常检

测方法正朝着更逼真、更多样的伪异常生成策略、渐进式决策边界优化，以及结合预训练特征的自适应调整方向演进，以期提升模型对复杂工业场景的泛化能力。其中，ABL指近似边界学习，AFS指异常特征合成，RBO指精细化边界优化。

2.3 基于归一化流的图像异常检测方法

如图4所示，归一化流^[25]方法将工业图像异常检测转化为精确的概率密度估计问题，其核心思路是通过可逆神经网络将高维特征空间中的复杂正常样本分布映射到简单先验分布（如高斯分布），从而直接计算测试样本在正常分布下的似然值，并将其作为异常分数。这种方法的核心优势在于，可逆映射机制不仅能够实现对正常样本分布的精确似然估计，还具备较强的数学可解释性，且对正常分布的轻微偏移具有极高的敏感性。基于这一特性，归一化流方法特别适用于需

表3 基于单分类的图像异常检测方法总结

方法（提出年份）	分类算法	方法架构	损失函数	AUROC (MVTec-AD)
Patch SVDD ^[17] （1999年）	Deep SVDD	层次化编码器+分类器对	交叉熵+紧凑性损失	92.1%
FCDD ^[18] （2020年）	可解释性单分类	全卷积网络	伪Huber损失	92.5%
PANDA ^[19] （2021年）	基于预训练特征的适应性调整	特征提取器+特征自适应+异常打分	弹性正则化紧凑性损失+离群曝光损失	86.5%
CutPaste ^[20] （2021年）	伪异常单分类	伪异常合成+单分类	基于交叉熵的CutPaste损失	93.5%
SimpleNet ^[21] （2023年）	伪异常单分类	特征提取+特征适配+异常生成+分类器	截断L1损失	97.6%
SuperSimpleNet ^[22] （2024年）	伪异常单分类	基于SimpleNet架构，改进了异常生成策略	截断L1损失+Focal损失	98.4%
GLASS ^[23] （2024年）	伪异常单分类	基于SimpleNet合成全局和局部伪异常样本	二分交叉熵+Focal损失+硬样本挖掘	99.7%
PBAS ^[24] （2024年）	通过伪异常优化正常样本与异常样本的边界	ABL损失+二分交叉熵损失	ABL+AFS+RBO	99.8%

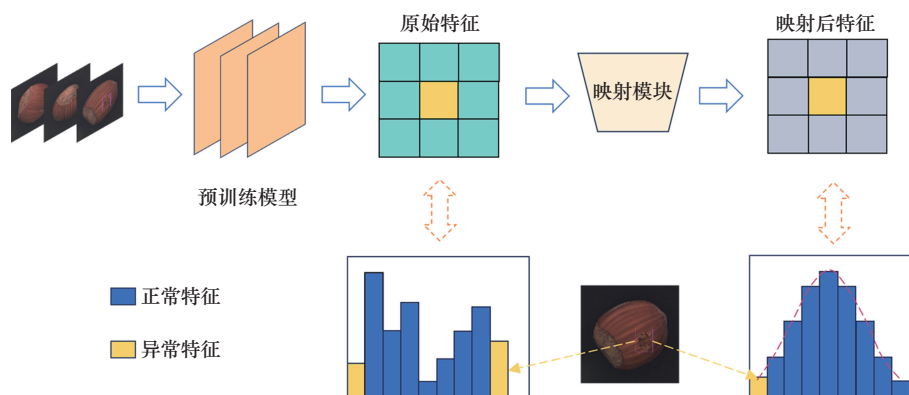


图4 基于归一化流的图像异常检测方法架构

要精确建模复杂纹理细节且正常模式高度一致的工业质检场景。例如，具有丰富表面纹理的产品检测任务，其对细微结构异常和分布偏移型缺陷具有出色的检测敏感性。然而，当正常样本存在显著类内变异，或异常样本与正常样本在全局统计特性上高度相似时，模型的检测性能会明显受限；同时，模型训练过程需要充分覆盖完整的正常样本分布，对训练数据的质量和图形处理单元（graphic processing unit, GPU）计算资源要求较高，因此更适用于精度优先、对实时性要求不高的工业质检场景。

DifferNet^[26]率先将实值非体积保持变化网络（Real-NVP）应用于多尺度特征评估，通过计算单幅图像多种变换形式的似然值实现精准异常评分，能够有效捕捉不同分辨率下的正常样本模式。CS-Flow^[27]在归一化的过程中集成了跨尺度全卷积耦合块，从而既能够充分利用多分辨率上下文信息，又可以保留图像的结构可解释性。CFLOW-AD^[28]基于条件归一化流框架，结合预训练编码器与多尺度生成解码器，显著提升了推理速度并实现了实时异常定位。FRAnomaly^[29]针对单一分辨率图像的检测需求设计了专用流模型，既可以减少特征提取过程中的信息遗漏，又能够加快检测效率。MSFlow^[30]引入了非对称并行耦合层，以增强模型对不同尺寸异常的适应性。Pyramid-Flow^[31]采用完全归一化流与金字塔结构相结合的设计，替代传统的预训练模型，以缩小正常样本的类内方差，并提升对高分辨率缺陷的对比和定位能力。AFlow^[32]通过注意力归一化流机制在细粒度层面融合多尺度特征，既提高了计算效率，又增强了模型对图像关键对象区域的关注。

如表 4 所示，这些方法总体上依赖可逆流模型对正常样本的特征分布进行精确建模，异常分数直接来源于测试样本的负对数似然值。这种方法在数学可解释性和对细微分布偏移的敏感性上具备突出的优势，尤其适用于纹理丰富、正常模式一致且对检测精度要求较高的工业质检场景。然而该类方法也存在明显局限。当训练数据覆盖不全或类内变异较大时，模型的检测性能会下降；同时，尽管其推理过程高效，但对 GPU 计算资源的依赖性较强。近年来，基于归一化流的图像异常检测方法正朝着多尺度耦合结构优化、注意力机制增强及条件流自适应调整等方向演进，以期更好地平衡检测精确性、推理效率与复杂工业场景的泛化能力。

2.4 基于记忆库的图像异常检测方法

基于记忆库的图像异常检测方法的架构如图 5 所示，其核心思路是将大量正常样本的特征原型存储到内存库中^[33]，在推理阶段，通过将测试图像的特征与记忆库中存储的正常特征原型进行最近邻匹配对比，进而量化二者差异以计算异常分数。这种方法本质上属于深度特征空间上的最近邻密度估计方法，其检测性能高度依赖记忆库中核心特征集的代表性和特征匹配度量方式的有效性。这种方法特别适用于正常样本模式多样但可以完全枚举的工业质检场景（如多品种电子元件检测、定制化产品质检及冷启动场景）。这种方法对训练过程中未见过的结构异常（如零部件缺失、异物混入等）具有较高的检测敏感度，同时因模型结构简单、无须额外的训练而具备突出的推理效率优势。然而，这种方法也存在若干局限。其一，由于内存消耗较大且难以完全

表 4 基于归一化流的图像异常检测方法总结

方法（提出年份）	方法架构	损失函数	AUROC（MVTec-AD）
DifferNet ^[26] （2021 年）	Real-NVP+软裁剪+多尺度提取和拼接	负对数似然	94.9%
CS-Flow ^[27] （2022 年）	Real-NVP+跨尺度全卷积耦合块	负对数似然	98.7%
CFLOW-AD ^[28] （2022 年）	基于条件归一化流的全卷积平移等变结构	负对数似然	98.6%
FRAnomaly ^[29] （2024 年）	Real-NVP+单尺度多层输入+全卷积耦合块	负对数似然	99.0%
MSFlow ^[30] （2024 年）	Real-NVP+非对称并行耦合层	负对数似然	99.6%
PyramidFlow ^[31] （2023 年）	基于金字塔融合和金字塔耦合块的完全归一化流	傅里叶损失函数	97.7%
AFlow ^[32] （2024 年）	Real-NVP+自注意力耦合层+堆叠 SE 块	负对数似然	99.7%

覆盖所有正常样本的变异模式，在产品频繁迭代的生产线上，记忆库的更新与维护成本较高；其二，这种方法对纹理渐变类缺陷或细微缺陷的检测效果有限，因此更适用于产品类型相对固定、正常样本数量充足的工业质检场景。

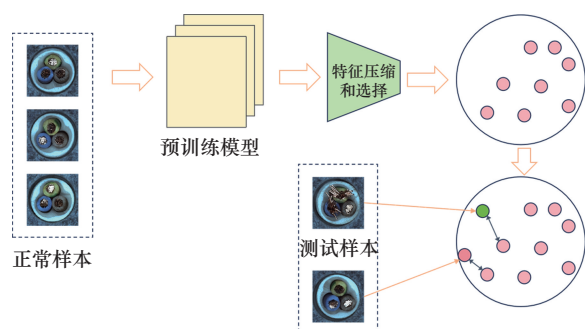


图5 基于记忆库的图像异常检测方法架构

工业图像异常检测中常面临冷启动问题，即正常样本数量不足，难以有效训练模型。PatchCore^[34]首次将能够完整概括正常图像特征的记忆库应用于工业图像异常检测领域，通过贪心核心集压缩策略，在保持竞争性推理速度的同时，实现了该领域领先的检测精度。针对传统记忆库将特征映射到离散空间易引发泛化性能不佳、模型学习身份捷径的问题，CRAD^[35]将空间特征转换为连续坐标并映射到网络中，有效提升了全局特征和局部特征的代表能力。FOCT^[36]聚焦少样本训练场景，通过条件传输度量方法计算查询图像与历史图像统计信息的相关性，并引入记忆库动态更新机制，以提高模型的整体泛化性能。由于记忆库本身难以完美表示多类正常特征的全部变体，研究者将记忆库作为辅助手段，间接提升异常检测效果。MemSeg^[37]从多分辨率特征中筛选最佳的差异信

息，结合空间注意力模块，显著增强了模型检测的灵活性与效率。PNI^[38]借助记忆库存储局部感知补丁特征，利用邻域信息建模条件分布，并通过微调网络防止特征丢失。ReConPatch^[39]通过对比学习生成目标导向的特征表示，将上下文相似性与成对相似性作为伪标签，使正常特征与异常特征更易分离。

如表5所示，这些方法总体依赖记忆库对正常特征的多样性存储，异常分数来源于测试特征与记忆库中最近邻特征的偏差度量。该类方法在推理速度和对多样正常模式的覆盖能力上优势明显，尤其适合产品类型单一、正常样本可枚举且需快速响应的工业场景。近年来，基于记忆库的图像异常检测方法正朝着连续空间映射、动态更新机制及与注意力或对比学习相融合的方向演进，以更好地平衡存储效率、泛化能力和对复杂工业场景的适应性，其中MSE指均方误差。

2.5 基于视觉语言模型的图像异常检测方法

针对正常样本极少的工业生产场景（如柔性产线或新产品上线），基于视觉语言模型的图像异常检测方法依托大规模跨模态预训练模型（如CLIP^[40]），将文本提示与图像特征映射到共享嵌入空间，在无须或仅需要少量正常样本参与训练的情况下，实现零样本/少样本工业图像异常检测。这种方法的核心优势在于依托预训练模型的强大语义理解能力，实现对高层语义类异常（如组件错装、标签错误或产品类型混杂等）优异的检测能力，但其实际检测效果高度依赖提示工程的设计水平，同时对细微纹理缺陷或微观结构异常的检测敏感度较低。这种方法特别适用于产品型号频繁变更、难以采集充足的正常样本的生产

表5 基于记忆库的图像异常检测方法总结

方法（提出年份）	存储内容	压缩方式	损失函数	AUROC (MVTec-AD)
PatchCore ^[34] （2022年）	邻域聚合的特征	贪心核心集压缩	L2 损失	92.6%
CRAD ^[35] （2024年）	连续坐标转换	精炼假阳性区域	MSE+余弦相似度	97.8%
FOCT ^[36] （2024年）	当前查询图像和历史图像统计信息整合	最小—最大设施核心集选择	条件传输损失+最小—最大设施核心集选择	93.2%
MemSeg ^[37] （2023年）	不同分辨率的特征	最佳差异信息选择+注意力机制	L1 损失+Focal 损失	99.56%
PNI ^[38] （2023年）	局部感知补丁特征	贪心核心集提取	L2 损失+梯度损失	99.56%
ReConPatch ^[39] （2024年）	处理后的目标导向特征	贪心近似算法压缩特征	松弛对比损失	99.56%

场景；但由于其需要依托稳定的网络环境调用大型预训练模型的特性，在算力受限的边缘设备上难以直接部署。

针对传统视觉语言模型方法难以获取多类别文本提示问题，Zhou 等^[41]提出的 AnomalyCLIP 方法率先探索学习对象无关的通用文本提示，引导模型聚焦图像中异常区域而非对象本身的语义信息，实现了通用化的正常与异常识别。随后，PromptAD^[42]采用语义级联方式将正常样本对应的文本提示与异常相关后缀相结合，生成大量负样本，并通过超参数控制正负样本比例，提升了异常边界的建模精度。为解决文本描述细粒度不足导致异常定位不准的问题，FiLo^[43]引入大语言模型（LLM）自适应生成细粒度描述（FG-Des），融合位置信息，实现了高质量的异常定位。LAFT^[44]则将用户知识和偏好直接融入文本提示设计，利用视觉—语言共享嵌入空间实现视觉特征与用户需求的对齐，精准建模正态性边界。面对视觉语言模型普遍缺乏工业领域知识的难题，AnomalyGPT^[45]通过生成伪异常样本及对应文本描述构建专属训练数据集，并设计提示学习器对视觉语言模型进行微调，大幅提升模型在工业场景中的适配能力。Li 等^[46]提出一对一个性化对比策略，融合个性化图像、异常自由样本与文本提示，通过三元组对比学习细化异常分布图谱，实现了更精确的细粒度异常定位与特征比较。

如表 6 所示，这些方法总体上依赖视觉语言模型的跨模态嵌入能力，异常分数来源于图像—文本特征的对齐偏差，或提示引导下的特征相似性度量。视觉语言模型方法在少样本乃至零样本条件下展现出强大的语义理解优势，在高层语义异常检测任务中表现突出，尤其适用于高层语义

异常检测和产品类型混杂的柔性生产线。然而，该类方法也存在明显局限。当提示工程设计不足或低层次纹理缺陷占主导时，模型的检测性能易受限制，同时大型预训练模型的部署门槛较高。近年来，基于视觉语言模型的图像异常检测方法正朝着细粒度提示自适应生成、用户知识与领域知识深度融入、个性化对比机制优化的方向演进，以期更好地平衡模型的语义理解能力与实际工业部署的友好性。其中，CLIP 训练损失指对比语言—图像预训练损失，SSIM 损失指结构相似性损失。

3 基于图像重建的工业图像异常检测方法

与基于特征嵌入范式的工业图像异常检测方法相比，基于图像重建范式的工业图像异常检测方法对预训练模型的依赖程度更低，且具备更优的异常定位能力，但难以精确提取高层语义信息，导致图像分类能力较弱。如图 6 所示，基于图像重建范式的工业图像异常检测方法在训练过程中仅将正常图像送入重建网络，利用损失函数指导网络学习对正常样本的重建。在推理阶段，依靠训练好的重建网络对测试图像进行重建还原。由于训练时从未见异常区域，网络无法良好地重建这些区域，从而通过重建误差实现异常定位。大多数基于图像重建的方法均遵循上述范式，不同方法的核心差异体现在重建网络的结构设计及数据增强策略的选择上。根据重建网络所采用的结构，可以将该类方法进一步划分为 5 类，即基于逆向蒸馏的图像异常检测方法、基于自编码器的图像异常检测方法、基于 GAN 的图像异常检测方法、基于 Transformer 的图像异常检测方法和基于扩散模型的图像异常检测方法。由于不同的网络结构存在固有的优势与局

表 6 基于视觉语言模型的图像异常检测方法总结

方法（提出年份）	文本提示生成方法	损失函数	AUROC（MVTec-AD）
AnomalyCLIP ^[41] （2026 年）	学习通用的文本提示	全局损失+局部损失	91.5%
PromptAD ^[42] （2024 年）	级联正常样本和异常后缀，将正常样本转换为异常样本	准确异常边界损失+CLIP 训练损失	96.6%
FiLo ^[43] （2024 年）	利用 LLM 提升文本提示的细粒度	交叉熵损失+Focal 损失+Dice 损失	91.2%
LAFT ^[44] （2025 年）	在文本提示中融入用户知识和偏好	无	96.5%
AnomalyGPT ^[45] （2024 年）	使用 LLM 对图像进行精准的文本描述	交叉熵损失+Focal 损失+Dice 损失	97.4%
Li 等 ^[46] （2024 年）	结合状态和物理两个维度生成文本提示	去噪损失+SSIM 损失	96.2%

限,研究者开始探索如何通过优化这些网络结构来提升其在图像异常检测任务中的性能。

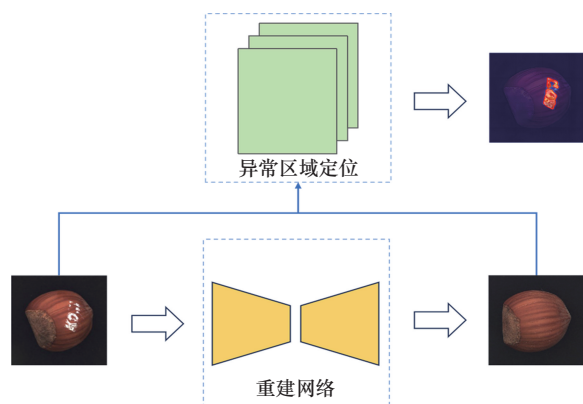


图6 基于图像重建范式的工业图像异常检测方法架构

3.1 基于逆向蒸馏的图像异常检测方法

基于逆向蒸馏的图像异常检测方法通过改进传统知识蒸馏的网络结构,使用教师编码器提取特征、单类瓶颈嵌入层(one-class bottleneck embedding, OCBE)强制压缩正常样本的语义信息、学生解码器重建特征,从而在特征层面和图像层面均产生可以显著区分的残差。逆向蒸馏方法同时具备特征嵌入范式的高层语义捕捉能力和图像重建范式的精确定位优势,特别适用于表面纹理复杂且正常模式高度一致的工业场景(如精密光学元件、结构化产品细微缺陷识别),对微小划痕、局部纹理畸变等低层次异常具有较高的敏感度。然而,这种方法对教师网络的特征提取能力依赖较大,且对环境噪声和光照变化较敏感,同时计算复杂度较高,更适合具备稳定GPU资源的工业质检系统。

Deng等^[47]于2022年首次提出逆向蒸馏思想,并将其应用于工业图像异常检测领域,通过

最小化特征相似性损失,引导学生解码器仅重建正常样本模式。这种方法在MVTec-AD数据集上取得了优异的检测性能。针对异常差异性不足问题,EMMFRKD^[48]对单类瓶颈嵌入层(OCBE)模块进行了改进,实现了多层次特征的不均匀融合,并引入坐标注意力机制以增强正常样本与异常样本的表征差异。REASON^[49]聚焦训练过程中的异常泄露问题,通过在推理阶段增加潜在的异常压制机制,有效阻止学生解码器学习并接受异常模式。面对学生模型过度泛化导致关键表征丢失的挑战,RD-E^[50]引入了专家网络以增强学生模型的表征能力,并注入引导信息以添加图像细节信息,从而提升图像细节的重建效果。为优化推理速度,RD++^[51]在教师网络中间层直接连接投影层,为学生解码器提供紧凑且无异常的特征表示,同时融合多种损失函数对蒸馏过程进行整体优化。

如表7所示,这些方法总体上依赖教师—学生框架的逆向设计,通过OCBE强制压缩正常样本的语义信息,异常分数来源于重建特征与教师提取特征之间的相似性偏差。这种方法在高层语义捕捉与像素级定位的平衡上具备突出优势,尤其适用于复杂纹理产品的表面缺陷检测。然而,当教师网络的特征表征能力不足或环境噪声较大时,模型的检测性能易受限制,且较高的计算开销限制了其部署范围。近年来,基于逆向蒸馏的图像异常检测方法正朝着多尺度特征融合、异常泄露抑制及推理效率优化等方向演进,以期提升模型在真实工业场景下的鲁棒性和实用性。

3.2 基于自编码器的图像异常检测方法

基于自编码器的图像异常检测方法通过“编

表7 基于逆向蒸馏的图像异常检测方法总结

方法(提出年份)	教师编码器	学生解码器	瓶颈层	损失函数	AUROC (MVTec-AD)
RD ^[47] (2022年)	WResNet50	反向对称WResNet50	多尺度特征融合块+一类嵌入块	余弦相似性	97.8%
EMMFRKD ^[48] (2024年)	WResNet50	反向对称WResNet50	协调注意力+多尺度特征融合块+记忆库	余弦相似性+结构相似性+梯度相似性	98.3%
REASON ^[49] (2025年)	孪生网络	反向对称WResNet50	多尺度特征融合块+一类嵌入块	潜在异常压制损失+多粒度知识蒸馏损失	99.44%
RD-E ^[50] (2025年)	WResNet50	反向对称WResNet50	多尺度特征融合块+一类嵌入块	余弦相似性+L1	99.2%
RD++ ^[51] (2023年)	WResNet50	反向对称WResNet50	多尺度特征融合块+一类嵌入块	自监督优化传输损失+重建损失+对比损失	99.0%

码—解码”结构在潜在空间对模型施加重建约束，迫使模型仅学习正常样本的紧致特征表示，在推理阶段利用重建图像与原始测试图像之间的残差实现异常的检测与定位。这种方法对预训练模型的依赖度较低，计算复杂度适中，可以部署在中等算力的边缘设备上，特别适用于检测与正常模式在视觉特征上差异明显的异常（如表面裂纹、污渍附着或结构变形等缺陷检测）。然而，当正常样本本身具有高度多样性或异常区域与正常区域在局部纹理、边缘结构等低层视觉特征上的差异较小时，模型的泛化能力过强易导致异常区域被“修复”，使得重建残差缩小，从而降低检测精度和定位准确性。

2014年，Sakurada等^[52]首次将自编码器引入异常检测领域，并验证了其在工业图像异常检测中的应用潜力。随后，针对传统自编码器重建方法缺乏语义表示的问题，Dual-Siamese^[53]通过Siamese网络捕获合成异常与正常样本的判别性特征，结合密集特征融合模块提升重建质量。SSM方法^[54]采用随机掩膜自监督训练策略，引导模型学习不同尺度掩膜区域的重建过程，有效提高了对局部异常和小尺寸异常的检测能力。由于图像重建方法在训练过程中普遍需要引入伪异

常合成策略以增强模型的泛化能力，相关研究围绕伪异常生成与网络优化展开。DSR^[55]通过双解码器量化特征空间架构，在特征层面可控生成伪异常样本，避免了图像级伪异常合成带来的数据分布偏差。DRAEM^[56]将Perlin噪声与外部纹理融合生成伪异常，并通过重建子网络与判别子网络协同训练，提升了模型的检测鲁棒性。RealNet^[57]利用扩散模型生成更逼真的伪异常数据集，并引入异常感知特征选择与残差选取模块，显著减少了冗余信息干扰。IFgNet^[58]采用多任务注意力共享子网络，同时实现前景提取与图像重建，提升了复杂背景下的异常检测性能。为解决自编码器瓶颈层泛化能力难以控制的问题，Zhang等^[59]提出了一种双约束自编码器，通过对抗学习与记忆库约束增大正常样本与异常样本的特征差异。FeatDAE方法^[60]通过在潜在空间引入梯度直方图与特征旋转移动策略，增强模型的全局注意力，并扩大伪异常的生成范围。INP-Former^[61]从测试图像中直接提取内在正常原型，以此指导模型重建过程，突破了训练集特征对齐的限制，使得模型在陌生类别产品的异常检测任务中仍能保持较高的检测精度。

基于自编码器的图像异常检测方法总结如表8

表8 基于自编码器的图像异常检测方法总结

方法（提出年份）	伪异常合成策略	方法架构	损失函数	AUROC (MVTec-AD)
Dual-Siamese ^[53] （2022年）	通过DRG模块生成多样化伪异常样本	Siamese网络+DRG+DFF	基于L2损失的重建和修复损失	96.7%
SSM ^[54] （2022年）	通过随机遮蔽正常图像生成伪异常样本	条件自编码器+掩码自注意力	MSE损失+GMS损失+SSIM损失+Mask损失	92.0%
DSR ^[55] （2022年）	将图像映射到量化空间，采样生成伪异常样本	编码器+从事不同外观解码的双解码器	L2损失+Focal损失	98.2%
DRAEM ^[56] （2021年）	使用Perlin噪声结合外部纹理生成伪异常样本	重建子网络+判别子网络	SSIM损失+Focal损失	98.0%
RealNet ^[57] （2024年）	使用扩散模型结合超参数生成可以控制的伪异常样本	SDAS+AFS+RRS	负对数似然损失+AFS损失+重建损失+交叉熵损失	99.6%
IFgNet ^[58] （2024年）	DRAEM的异常合成策略	多注意力子网络+判别子网络	L2损失+SSIM损失+Focal损失	98.4%
Zhang等 ^[59] （2024年）	DRAEM的异常合成策略	对抗学习与记忆库约束的自编码器结构	对抗损失+L2损失+SSIM损失+Focal损失	99.2%
FeatDAE ^[60] （2025年）	在图像中添加Perlin噪声，在潜在空间特征图上进行径向移动	特征级降噪自编码器	余弦相似度损失+Focal损失+L1损失	99.4%
INP-Former ^[61] （2025年）	无	INP提取器+INP指导的自编码器	余弦相似度损失+软挖掘损失	99.7%

所示, 其中, DRG 指缺陷随机生成, DFF 指密集特征融合, DRAEM 指判别式训练的重建异常嵌入模型, INP 指内在正常原型, GMS 指梯度幅值相似性损失。这些方法总体上依赖自编码器生成的重建残差作为异常分数的来源, 通过潜在空间约束和伪异常增强策略, 实现对正常样本模式的学习。这种方法在定位精度和计算效率的平衡上具备显著优势, 尤其适用于生产线环境稳定、产品外观一致性高的表面缺陷检测场景。然而, 当正常样本多样性高或细微缺陷与正常特征接近时, 重建误差易被掩盖, 导致模型的检测性能下降。近年来, 基于自编码器的图像异常检测方法正朝着更逼真的伪异常合成、更高效的泛化约束机制及测试时原型自适应提取等方向演进, 以期提升模型在复杂工业场景下的鲁棒性和泛化能力。

3.3 基于生成对抗网络的图像异常检测方法

基于 GAN 的图像异常检测方法架构如图 7 所示, 其核心原理是通过生成器与判别器的对抗训练, 使生成器逐步学习并掌握正常样本的分布规律, 进而将输入图像重建为符合正常样本分布的图像。在推理阶段, 该方法通过量化重建图像与原始图像之间的残差, 实现异常的检测与定位。该方法能够生成与正常样本高度相似的图像, 对复杂纹理和中等尺度异常具备较强的建模能力, 特别适用于表面纹理复杂、对检测细节要求较高的工业场景 (如电子元件检测)。然而, 该方法对抗训练过程中易出现模式崩溃、收敛困难等问题, 对训练数据的质量和数量要求较高, 且模型训练稳定性较差, 因此适用于具备充足正常样本且对图像重建保真度有较高需求的工业质检场景。

2019 年, Schlegl 等^[62]提出的 f-AnoGAN 为后续工业应用奠定了基础, 其核心思想是通过生成器学习正常数据的分布规律, 进而检测偏离该分布的异常样本。I3AD^[63]采用条件对抗网络与

动态调整的掩码矩阵, 通过迭代修复异常区域, 结合结构相似性指标优化重建效果, 提升了对图像微小细节的重建能力。SCADN^[64]针对传统方法中预训练特征忽略多尺度语义信息的不足, 通过在正常样本上添加多尺度条纹掩码生成多样化重建样本, 引导模型强化上下文语义学习。针对小异常区域重建困难问题, CKAAD^[65]引入少量粗糙异常标签辅助对抗训练, 同时设计基于能量的判别器以抑制生成器对异常区域的重建行为, 从而强化局部异常识别。OCR-GAN^[66]从频率角度出发, 通过频率解耦模块与自适应通道选择机制, 实现多频率特征的并行编码与解码, 提升了重建方法的整体检测准确率。为提高生成图像的真实分布符合度, GAN-Inversion^[67]与 RGI^[68]引入了稳健的逆向映射策略, 通过优化潜在向量和掩码设计, 并引入稀疏性约束, 实现了无掩膜条件下的高质量图像修复。

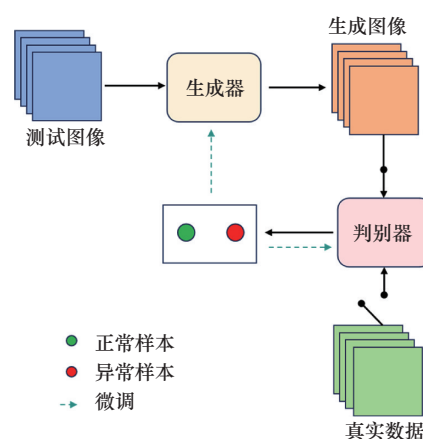


图 7 基于 GAN 的图像异常检测方法架构

基于 GAN 的图像异常检测方法总结如表 9 所示, 其中 RGI 损失指相对梯度信息损失。这些方法总体上依赖生成器与判别器的对抗博弈学习正常样本的数据分布, 异常分数来源于重建图像

表 9 基于 GAN 的图像异常检测方法总结

方法 (提出年份)	整体结构	损失函数	AUROC (MVTec-AD)
I3AD ^[63] (2021 年)	SN-Patch GAN+动态调整的掩码矩阵	Hinge 损失	86.3%
SCADN ^[64] (2021 年)	多尺度语义上下文学习+GAN+多尺度条纹掩码	L1 损失+对抗损失	89.8%
CKAAD ^[65] (2025 年)	自编码器生成器+基于能量的判别器	余弦损失+对抗损失	99.5%
OCR-GAN ^[66] (2023 年)	频率解耦器+编码器+通道选择器+解码器	重建损失+对抗损失+潜在损失	98.3%
RGI ^[68] (2025 年)	生成器+判别器	RGI 损失	97.7%

与原始图像之间的残差度量。这种方法在复杂纹理的高保真重建与细粒度异常检测能力上具备突出优势，尤其适用于电子元器件等存在中等尺度异常的工业质检场景。然而，该方法模型训练不稳定、模式崩溃问题仍较突出，且对训练数据的质量要求较高。近年来，基于 GAN 的图像异常检测方法正朝着频率解耦、多尺度上下文增强及稳健 GAN 逆向映射等方向演进，以期提升模型训练的稳定性及工业部署的实用性。

3.4 基于 Transformer 的图像异常检测方法

基于 Transformer 的图像异常检测方法^[69]将自注意力机制引入编码—解码结构，凭借其强大的全局上下文建模和长距离依赖捕捉能力，在图像重建过程中能够同时兼顾图像的全局结构与局部细节，从而生成更贴合正常样本模式的图像。该方法特别适用于需要全局理解的复杂工业检测场景（如航空部件检测、大面积表面缺陷检测，以及具有复杂纹理与背景的工业产品检测），其对长距离结构异常的识别能力显著优于传统的卷积神经网络方法。然而，该方法的全局注意力机制对微小或局部异常的检测敏感度相对较低，且高度依赖充足的正常样本与稳定的成像条件，同时计算复杂度较高，因此更适合对全局上下文理解要求高而对实时性要求适中的工业质检系统。

InTra^[70]首次将 Transformer 作为骨干网络应用于工业图像异常检测领域，通过图像 Patch 划分与多头自注意力机制，直接将图像特征映射到像素空间完成重建任务，完全摒弃传统卷积操作，从而更高效地处理大范围异常区域。VT-ADL^[71]在 Transformer 编码器中保留图像 Patch 的

空间信息，结合高斯混合密度网络建模正常样本分布，实现对细微异常的精确检测。为解决工业场景中背景干扰与异常知识缺失问题，MSFR^[72]引入可训练先验增强视觉 Transformer（PViT），通过掩码视觉编码器专注捕捉正常样本的关键特征知识，并与特征提取器的输出进行对比，确定异常区域。U-ViT^[73]针对单一类型样本训练不稳定、卷积重建定位精度不高的问题，对 ViT 的自注意力机制进行窗口限制，并融入深度卷积聚合策略，提升了对局部信息的利用能力与计算效率。在重建过程中，模型常出现恒等映射问题，导致异常无法被识别，对此 Luo 等^[74]提出了 TFA-Net，通过基于模板的特征聚合机制将正常样本的特征融合到模板中，并利用特征细节细化模块生成可区分的特征嵌入，有效避免了恒等映射。PNPT^[75]采用双流重建框架，引入特定类别先验正常提示，通过语义对齐耦合编码模块与上下文语义条件解码模块，实现正常先验知识与异常特征的深度融合，为基于先验知识的异常检测提供了新的高效技术路径。

如表 10 所示，这些方法总体上依赖 Transformer 的自注意力机制进行全局—局部特征的协同建模，异常分数来源于重建图像与原始图像之间的残差度量或特征层面的相似性对比。这种方法在复杂纹理与全局结构理解能力上具有突出优势，尤其适用于航空部件等具有长距离依赖关系的工业质检场景。然而，该方法对微小局部异常的检测敏感度不高、恒等映射问题仍需持续优化。近年来，基于 Transformer 的图像异常检测方法正朝着先验知识融入、双流架构设计及窗口

表 10 基于 Transformer 的图像异常检测方法总结

方法（提出年份）	使用的 Transformer 类型	整体结构	损失函数	AUROC (MVTec-AD)
InTra ^[70] （2022 年）	一般 Transformer	多头特征自注意力机制+长距离残差连接	L2 损失+GMS 损失+SSIM 损失	96.6%
VT-ADL ^[71] （2021 年）	一般 Transformer	Transformer 编码器+解码器+高斯混合密度网络	负对数似然损失+MSE 损失+SSIM 损失	98.4%
MSFR ^[72] （2024 年）	PViT	掩码视觉编码器+多尺度金字塔解码器+特征提取器	余弦相似度损失	98.4%
U-ViT ^[73] （2024 年）	ViT	重建子网+判别子网+异常生成模块+掩码生成模块	均方误差损失+焦点损失	97.7%
TFA-Net ^[74] （2024 年）	ViT	基于模板的特征聚合机制+特征细节细化模块	欧氏距离损失+余弦相似度损失	98.7%
PNPT ^[75] （2024 年）	ViT	CS-NPP+HPE+SACE+CSCD	余弦相似度损失	98.3%

注意力与局部卷积融合等方向演进，以期更好地平衡全局建模能力、定位精度与工业部署效率。其中，CS-NPP指类别特定正常性提示池，HPE指层次化补丁嵌入，SACE指语义对齐耦合编码，CSCD指上下文语义条件解码。

3.5 基于扩散模型的图像异常检测方法

基于扩散模型的DDPM^[76]通过逐步加噪与逆向去噪过程，学习从任意噪声状态恢复至正常样本流形的条件概率路径。由于异常区域不符合正常样本的固有去噪轨迹，在多步逆向去噪迭代中会累计显著的误差，从而实现高保真图像重建与异常定位。这种方法天然具备对复杂纹理和微小结构异常的高敏感性，且能够生成与正常样本高度贴合的高质量图像，特别适用于高价值精密制造场景。然而，其推理过程需多次迭代，计算复杂度高、单样本检测耗时长，难以满足高速生产线的实时检测需求，且模型训练对数据质量敏感，因此更适合离线高精度分析或低速质检场景。

AnoDDPM^[77]率先将去噪扩散概率模型(DDPM)引入工业图像异常检测领域，通过设计部分扩散策略和多尺度单纯形噪声，针对性提升高分辨率工业图像的检测效率。针对标准扩散模型推理速度过慢的问题，DiffusionAD^[78]基于范式指导的一步去噪重建框架，结合语义指导与细节保留机制，直接生成无异常图像，显著提升检测效率。GLAD^[79]通过逐时间步对比加噪与降噪过程的特征差异，动态调整去噪迭代步数，同时融合正常区域与异常区域的特征信息，突破高

斯分布的限制，提升异常区域的重建水平。为解决结构重构不准难题，DDAD^[80]采用条件去噪机制，利用贝叶斯规则分解后验分数以精准调整噪声项，实现了在保留正常样本名义模式的同时去除异常区域的重建；DiAD^[81]基于潜在空间扩散模型，引入语义引导网络与空间感知特征融合器，提高多尺度重建精度。DeCo-Diff^[82]在训练阶段采用随机掩码策略，仅对图像中的噪声区域进行加噪处理，同时结合偏差方向预测模块，引导模型学习从异常到正常的修正方向，提升模型的上下文修正能力。Fučka等^[83]提出了基于透明度的扩散模型TransFusion，该方法将异常图像建模为正常外观与异常外观的线性插值，通过透明度因子连续控制异常区域的可见性，将传统“重建一分割”两阶段检测流程统一为单阶段迭代框架。

如表11所示，这些方法总体上依赖扩散模型的逐步去噪机制，异常分数来源于多步重建误差或噪声预测偏差。该类方法在高保真图像重建与复杂纹理异常敏感性上具备突出优势，适用于高端制造中微小缺陷的高精度检测场景。然而，该方法的推理时间长、计算开销大。近年来，基于扩散模型的图像异常检测方法正朝着一步去噪/条件去噪加速、掩码与透明度控制策略优化及语义与空间特征融合机制创新等方向演进，以期更好地平衡重建质量、定位精度与工业实时部署需求。其中，LDM指潜空间扩散，SG指语义引导，SFF指空间感知特征融合，DoD指双层扩散架构。

表11 基于扩散模型的图像异常检测方法总结

方法（提出年份）	扩散模型	整体架构	损失函数	AUROC (MVTec-AD)
AnoDDPM ^[77] （2022年）	一般扩散模型	DDPM+部分扩散策略+多尺度单纯形噪声	DDPM变分下界损失	86.3%
DiffusionAD ^[78] （2025年）	单步去噪扩散模型	范式指导的一步去噪重建子网络+分割子网络	均方误差损失+L1损失+焦点损失	99.7%
GLAD ^[79] （2024年）	传统扩散模型	传统扩散模型+自适应降噪时间步决策+空间自适应特征融合	一般扩散模型损失+均方误差损失	99.3%
DDAD ^[80] （2024年）	条件去噪扩散模型	隐式去噪扩散模型+条件去噪过程	L1损失+余弦相似度损失	99.8%
DiAD ^[81] （2024年）	潜空间扩散模型	LDM模型+SG网络+SG降噪网络+SFF模块	均方误差损失	96.8%
DeCo-Diff ^[82] （2025年）	隐式去噪扩散模型	变分自编码器+隐式去噪扩散模型+随机掩码策略+DoD模块	DDPM变分下界损失+均方误差损失+偏差方向预测损失	99.3%
TransFusion ^[83] （2024年）	基于透明度的扩散模型	掩码生成模块+合成异常生成模块+基于透明度扩散模型	SSIM损失+L1损失+Focal损失+L2损失	99.2%

4 数据集和衡量指标

4.1 数据集

本节将系统介绍工业图像异常检测领域常用的基准数据集，这些数据集覆盖不同的场景、针对各类检测挑战，为模型性能评估提供了标准化基准。

在表面纹理异常检测领域，MVTec-AD 数据集提供了丰富的纹理样本，涵盖织物、线材、皮革等材料上的划痕、污渍、纹理等常见异常，这类异常主要表现产品表面的细微瑕疵。在结构性异常检测方面，BTAD 和 MPDD 数据集囊括了组件缺失、错位和变形等物体结构层面的缺陷，这类异常直接影响产品的结构完整性与功能稳定性。在微小缺陷检测领域则由 MTD 数据集覆盖，该数据集中样本的缺陷虽尺寸细微（占整体图像极小比例），却可能严重影响产品的正常使用。人工合成数据集 DAGM 和 Real-IAD 模拟了多样化的异常场景和缺陷类型来提高模型的鲁棒性。MVTec LOCO-AD 数据集则专注于逻辑上下文异常检测，此类异常并不会表现为产品功能或表面的缺陷，而是在功能逻辑或语义层面存在不合理之处。VisA 数据集侧重于多视角属性分析，探索模型在不同视角下对异常的一致性识别能力。上述工业图像异常检测领域主要数据集的关键信息汇总见表 12，具体包括数据集的样本总数、类别总数、正常样本数、异常样本数和标注级别（像素级或图

像级）。通过汇总这些数据集的关键信息，能够为研究人员根据自身研究方向、检测任务需求选择合适的数据集开展实验提供指导。

表 12 工业图像异常检测主要数据集关键信息汇总

数据集名称	样本总数/幅	类别总数/类	正常样本数/幅	异常样本数/幅	标注级别
MVTec-AD	5 354	15	4 096	1 258	像素级
BTAD	2 830	3	2 250	580	像素级
MPDD	1 346	6	1 064	282	像素级
MTD	1 344	1	952	392	像素级
DAGM	16 100	10	14 000	2 100	图像级
Real-IAD	151 050	30	99 721	51 329	像素级
MVTec LOCO-AD	3 644	5	2 651	993	像素级
VisA	10 821	12	9 621	1 200	像素级

4.2 衡量指标

在工业图像异常检测领域，模型性能评估是一项多维度的系统性工程，必须通过多视角衡量指标全面、精准地反映模型在真实工业环境中的应用价值。当前主流评估体系主要围绕两个核心维度展开：图像级分类聚焦于判别单幅图像是否存在异常；像素级定位则致力于精准确定缺陷的位置。

在图像级分类维度，评估核心是衡量模型对整个样本进行二元判断的宏观能力，这一能力直接关系到自动化产线的“放行/拦截”决策。在此维度下，评价指标如表 13 所示。其中，TP 指

表 13 评价指标

指标名称	公式	含义
精度	$P = \frac{TP}{TP + FP}$	检测为异常中真正异常的比例
召回率	$R = \frac{TP}{TP + FN}$	真实异常样本中被正确检测出的比例
误检率	$FPR = \frac{FP}{FP + TN}$	正常样本中被误判为异常的比例
F1	$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$	精度和召回率的调和平均数
AUROC	$AUROC = \int_0^1 TPRd(FPR)$	ROC 曲线下面积，体现模型在不平衡数据集上的性能
PRO	$PRO = \frac{1}{N} \sum_i \sum_k \frac{P_i \cap G_{i,k}}{G_{i,k}}$	检测出的缺陷区域与真实缺陷区域的重叠程度
AP	$AP = \frac{1}{N} \sum_{k=1}^N P(k)$	检测为异常中真正异常的平均比例

真正例 (true positive): 实际为正类, 模型也预测为正类的数量; FP 指假正例 (false positive): 实际为负类, 但模型错误预测为正类的数量; TN 指真负例 (true negative): 实际为负类, 模型也预测为负类的数量; FN 指假负例 (false positive): 实际为正类, 但模型错误预测为负类的数量。主流技术主要关注3个相互关联的基础指标, 即精度 (precision)、召回率 (recall) 和误检率 (false positive rate, FPR)。由于基础指标之间存在相互制衡关系, 单一指标难以全面反映模型性能, 因此更具综合性的复合指标被广泛采用。F1 通过计算精度和召回率的调和平均值, 能够有效评估模型在工业场景常见的极度不平衡数据下的综合性能; 受试者操作特征曲线下面积 (area under the receiver operating characteristic curve, AUROC) 则通过绘制召回率与误检率在不同决策阈值下的变化曲线, 提供了一个与阈值选择无关的、衡量模型区分正常与异常样本整体能力的稳健分数。AUROC 值越接近 1, 表明模型的整体能力越强。

在完成图像级异常判定后, 评估重心将转向像素级定位, 其目的是精确识别缺陷的位置、形状和大小, 为后续的缺陷分级与根因分析提供依据。在该维度下, 像素级 AUROC 与平均精度 (average precision, AP) 被用于评估模型对每个像素点分类的准确性与稳定性, 其中 AP 通过计算不同置信度下像素级精度-召回率曲线的面积, 能够有效评估模型在多种条件下的检测性能。像素级别区域重叠度 (pixel-level region overlap, PRO) 指标则侧重于应用层面, 通过计算预测异常区域与真实标注区域在结构上的一致性反映模型定位缺陷的精度。

5 关键问题与挑战

近年来, 工业图像异常检测作为保障智能制造质量与生产效率的关键技术, 受到学术界与工业界的广泛关注。随着深度学习的快速发展, 尤其是 GAN、自编码器和扩散模型的迭代升级, 基于无监督学习的异常检测方法在工业图像异常检测领域实现了快速发展。从现有研究成果来看, 通过在训练过程中仅使用无标注的正常样本, 近年来陆续提出的无监督学习模型在 MVTec-AD

等数据集上的检测精度逐渐接近 100%。此外, 此类模型能够检测出更多类型、不同尺寸的工业异常, 并且在一定程度上解决了深度学习方法应用中不可避免的恒等映射问题。然而, 基于无监督学习的工业图像异常检测方法在理论研究和实际应用中依旧面临诸多挑战, 具体表现在以下几个方面。

1) 复杂背景和变化光照条件下的检测。现有研究所采用的训练数据集通常具有背景简单、光照均匀的特点, 这与真实工业场景存在显著差异。在实际工业生产环境中, 产品往往处于结构多变的背景中 (如织物上固有的花纹), 这使得模型难以区分“正常但复杂的背景模式”与“真正的异常区域”, 从而导致正常纹理被误判异常, 或者异常区域被复杂背景掩盖而无法识别。此外, 工业现场的光照条件难以精准控制, 光照变化会显著改变图像的像素分布, 引起输入数据分布的偏移, 最终使模型的检测性能急剧下降。因此, 如何设计一种对背景和光照变化具有鲁棒性, 同时能够精准捕捉细微异常的检测算法, 仍是推动无监督学习图像异常检测方法在实际中落地的核心挑战之一。

2) 逻辑异常识别。先前的无监督学习图像异常检测方法主要关注图像外观及结构层面的缺陷检测, 而忽视了部件之间在位置或者配置上的逻辑异常。这类逻辑异常在局部像素或纹理层面往往难以察觉, 必须依赖模型对全局语义的理解、对部件间关联关系的挖掘及对生产逻辑的推理才能实现有效识别, 这对模型的认知与推理能力提出了更高的要求。现有方法在逻辑异常检测方向上仍处于探索阶段, 如何结合视觉语言模型、图神经网络或知识图谱等技术是突破该瓶颈的关键路径。

3) 轻量化网络的设计与部署。尽管深度学习模型的检测性能不断提升, 但其对计算和存储资源、能耗的高需求, 与工业现场的实际部署条件之间存在显著矛盾。许多工业检测场景需要依赖边缘设备进行实时处理, 而这类设备通常算力较弱、内存有限且功耗受限, 难以运行参数量大、计算密集的复杂模型。此外, 工业质检通常要求毫秒级的响应速度, 以匹配高速运转的生产线节

奏。若模型推理时间过长,则会影响生产节拍,造成数据积压,从而无法实现实时预警。因此,在追求检测高精度的同时,必须兼顾模型的运行效率与实用性,未来如何实现精度、速度、资源消耗之间的最优平衡,已成为无监督学习工业图像异常检测领域的核心课题之一。

4) 跨领域方法的设计。工业场景具有高度多样性、专业性和非标准化的特征,现有的无监督学习图像异常检测方法仅在特定领域或数据集上表现优异,难以将习得的知识有效地迁移到全新领域。不同工业领域的图像在纹理结构、缺陷类型及背景干扰方面存在根本性差异,若将某一领域训练的模型直接迁移到新领域,数据分布的偏移会导致检测性能急剧下降,甚至完全失效。尽管已有研究尝试构建领域通用的检测框架,但这些框架普遍存在需在新领域进行微调、结构设计复杂且难以轻量化部署的问题。未来需突破模型在特征泛化、知识迁移、小样本学习等方面的瓶颈。

5) 工业领域实际数据匮乏。尽管现有仿真数据集(如 MVTEC-AD、BTAD、Real-IAD 等)已在实验室环境下支撑模型实现了极高的检测准确率(多数代表性方法的 AUROC 已接近或超过 98%),但这些数据集大多基于受控环境下的缺陷模拟,难以真实再现工业生产中复杂的背景噪声、光照变化、设备老化、产线动态干扰和极低频的异常模式。这种仿真环境与真实场景之间的差异,导致大量先进的无监督学习图像异常检测方法在仿真数据上表现优异,却在实际产线部署时泛化性能显著下降。因此,研究者必须更加关注领域自适应学习、少样本/零样本迁移及真实工业多模态数据集的构建^[84],以突破仿真与真实之间的差距带来的性能瓶颈,为构建真正可部署的工业质检系统奠定基础。

6) 像素级异常定位准确度低。像素级异常定位作为工业图像异常检测中的独立关键子任务,要求模型精确输出异常发生的像素区域,其定位结果直接为缺陷分级、质量溯源等下游工业应用提供支撑。然而,当前无监督学习方法在该任务上仍存在明显局限。其一,基于重建范式方法容易因过度泛化而修复部分异常区域,导致重建边

界残差明显衰减和热力图对比度不高;其二,基于特征嵌入范式方法受骨干网络下采样操作的限制,难以保留图像精细的空间细节,对微小或低对比度异常的边界刻画效果较差;其三,在复杂纹理、光照变化或产线噪声干扰下,定位结果常出现边界模糊、假阳性增多及真实异常被掩盖等问题。这些局限使得像素级指标虽在基准数据集上有提升,但在真实工业场景中的定位可靠性和实用性明显不足,成为制约无监督学习图像异常检测系统落地的核心瓶颈。为推动像素级异常定位的工业可用性,未来研究可重点探索多尺度高分辨率特征融合机制,提升模型对微小和低对比度异常的空间敏感性,同时引入更精细的边界感知约束或伪异常合成策略来放大异常与正常区域在边界处的差异,结合多模态信息来缓解产线扰动引起的泛化退化。

6 结束语

近年来,基于无监督学习的工业图像异常检测已形成以特征嵌入和图像重建两大范式为主导的研究框架。特征嵌入范式依托预训练模型优异的特征表征能力,通过度量测试样本与正常特征的偏差实现异常检测,具备推理速度快、边缘部署开销低的优势;图像重建范式借助自身的图像重建机制,实现了对复杂纹理特征的准确表征。随着知识蒸馏、归一化流、扩散模型及视觉语言模型等技术的飞速迭代,多数无监督学习图像异常检测方法在部分数据集上的图像级分类精度接近饱和,像素级定位性能也随之显著提升。尽管如此,在真实工业场景下部署应用仍面临诸多挑战,包括跨域泛化能力不足、对微小和低对比度缺陷的定位精度受限、高性能模型难以兼顾实时性与资源开销。更加关键的是,模型对高层逻辑与语义层面异常的检测能力依旧明显缺失。

未来的研究应着力于解决基于无监督学习的工业图像异常检测方法难以落地的问题。其一,借助多模态融合与跨模态先验建模技术的快速发展,优化模型特征学习能力,降低模型对光照变化、背景波动和各类干扰的敏感度;其二,引入增量学习机制,使模型能够持续适应数据扰动带来的分布偏移,不断学习未知异常与正常模式之

间的偏差；其三，为满足产线部署需求，应强化模型轻量化与自适应架构的设计，在检测精度、推理时延与能耗之间实现符合工业实际需求的平衡；其四，构建高度可解释的检测框架与人机协同机制，以提高模型在复杂场景下的运行可靠性与检测可信度。总体而言，工业图像异常检测的突破不应仅着眼于基准指标的提升，而应着力于设计能够协同满足上述多方面需求的模型，这些模型需在多样化、分布未知及算力受限的真实产线环境中保持稳定运行，从而真正支撑起高质量智能制造的落地。

参考文献：

- [1] Tao X, Gong X, Zhang X, et al. Deep learning for unsupervised anomaly localization in industrial images: a survey[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1-21.
- [2] Cui Y, Liu Z, Lian S. A survey on unsupervised anomaly detection algorithms for industrial images[J]. *IEEE Access*, 2023, 11: 55297-55315.
- [3] Naeem S, Ali A, Anam S, et al. An unsupervised machine learning algorithms: comprehensive review[J]. *International Journal of Computing and Digital Systems*, 2023, 13(1):911-921.
- [4] Gui J, Chen T, Zhang J, et al. A survey on self-supervised learning: algorithms, applications, and future trends[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(12): 9052-9071.
- [5] Zhang J, Yang L, Mohammadabadi S M S, et al. A survey on self-supervised learning: recent advances and open problems[J]. *Neurocomputing*, 2025: 131409.
- [6] Khan A, Sohail A, Fiaz M, et al. A survey of the self supervised learning mechanisms for vision transformers[PP]. *arXiv preprint*, 2025, arXiv: 2408.17059.
- [7] Liu J, Xie G, Wang J, et al. Deep industrial image anomaly detection: a survey[J]. *Machine Intelligence Research*, 2024, 21(1): 104-135.
- [8] Salehi M, Sadjadi N, Baselizadeh S, et al. Multiresolution knowledge distillation for anomaly detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021: 14902-14912.
- [9] Wang G, Han S, Ding E, et al. Student-teacher feature pyramid matching for anomaly detection[PP]. *arXiv preprint*, 2021, arXiv: 2103.04257.
- [10] Cao Y, Wan Q, Shen W, et al. Informative knowledge distillation for image anomaly segmentation[J]. *Knowledge-Based Systems*, 2022, 248: 108846.
- [11] Rudolph M, Wehrbein T, Rosenhahn B, et al. Asymmetric student-teacher networks for industrial anomaly detection[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Piscataway:IEEE Press, 2023: 2592-2602.
- [12] Gu H, Li G, Liu Z. Masked feature regeneration based asymmetric student - teacher network for anomaly detection[J]. *Multimedia Tools and Applications*, 2024, 83(42): 90573-90594.
- [13] Zhang S, Liu J. Feature-constrained and attentionconditioned distillation learning for visual anomaly detection[C]//*Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Piscataway: IEEE Press, 2024: 2945-2949.
- [14] Wei S, Jiang J, Xu X. UniNet: a contrastive learning-guided unified framework with feature selection for anomaly detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway:IEEE Press, 2025: 9994-10003.
- [15] 黄诗瑀, 叶锋, 黄添强, 等. 人脸伪造与检测中的对抗攻防综述[J]. *网络与信息安全学报*, 2023,9(4):1-15.
Huang S Y, Ye F, Huang T Q, et al. Survey on adversarial attacks and defense of face forgery and detection[J]. *Chinese Journal of Network and Information Security*, 2023, 9(4): 1-15.
- [16] Yi J H, Yoon S. Patch SVDD: patch-level SVDD for anomaly detection and segmentation[C]//*Proceedings of the 15th the Asian Conference on Computer Vision*. Berlin:Springer, 2020: 375-390.
- [17] Tax D M J, Duin R P W. Support vector domain description[J]. *Pattern Recognition Letters*, 1999, 20(11): 1191-1199.
- [18] Liznerski P, Ruff L, Vandermeulen R A, et al. Explainable deep one-class classification[PP]. *arXiv preprint*, 2021, arXiv: 2007.01760.
- [19] Reiss T, Cohen N, Bergman L, et al. PANDA: adapting pretrained features for anomaly detection and segmentation[C]//*Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway:IEEE Press, 2021: 2805-2813.
- [20] Li C L, Sohn K, Yoon J, et al. CutPaste: self-supervised learning for anomaly detection and localization[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway:IEEE Press, 2021: 9664-9674.
- [21] Liu Z K, Zhou Y M, Xu Y S, et al. SimpleNet: a simple network for image anomaly detection and localization[C]//*Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway:IEEE Press, 2023: 20402-20411.
- [22] Rolih B, Fučka M, Skočaj D. SuperSimpleNet: unifying unsupervised and supervised learning for fast and reliable surface defect detection[C]//*Proceedings of the International Conference on Pattern Recognition*. Cham: Springer Nature Switzerland, 2024: 47-65.
- [23] Chen Q Y, Luo H Y, Lv C K, et al. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization[C]//*Proceedings of the 2024 European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024: 37-54.
- [24] Chen Q Y, Luo H Y, Gao H, et al. Progressive boundary guided anomaly synthesis for industrial anomaly detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 35(2): 1193-1208.
- [25] Dinh L, Sohl-Dickstein J, Bengio S. Density estimation using real nvp[PP]. *arXiv preprint*, 2017, arXiv: 1605.08803.
- [26] Rudolph M, Wandt B, Rosenhahn B. Same same but different:

- Semi-supervised defect detection with normalizing flows[C]//Proceedings of the 2021 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2021: 1906-1915.
- [27] Rudolph M, Wehrbein T, Rosenhahn B, et al. Fully convolutional crossscale-flows for image-based defect detection[C]//Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2022: 1829-1838.
- [28] Gudovskiy D, Ishizaka S, Kozuka K. CFLOW-AD: realtime unsupervised anomaly detection with localization via conditional normalizing flows[C]//Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2022: 1819-1828.
- [29] Milković F, Posilović L, Medak D, et al. FRAnomaly: flow-based rapid anomaly detection from images[J]. *Applied Intelligence*, 2024, 54(4): 3502-3515.
- [30] Zhou Y X, Xu X, Song J K, et al. MSFlow: multiscale flow-based framework for unsupervised anomaly detection[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 36(2): 2437-2450.
- [31] Lei J R, Hu X B, Wang Y, et al. PyramidFlow: high-resolution defect contrastive localization using pyramid normalizing flow[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 14143-14152.
- [32] Wu X Z, Mao G J, Xing S L. Unsupervised anomaly detection in images using attentional normalizing flows[J]. *Engineering Applications of Artificial Intelligence*, 2024, 127: 107369.
- [33] Park H, Noh J, Ham B. Learning memory-guided normality for anomaly detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 14360-14369.
- [34] Roth K, Pemula L, Zepeda J, et al. Towards total recall in industrial anomaly detection[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 14298-14308.
- [35] Lee J C, Kim T, Park E, et al. Continuous memory representation for anomaly detection[C]//Proceedings of the 2024 European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 438-454.
- [36] Tian L, Zhao H Y, Lu R Y, et al. FOCT: few-shot industrial anomaly detection with foreground-aware online conditional transport[C]//Proceedings of the 32nd ACM International Conference on Multimedia. New York: ACM, 2024: 6241-6249.
- [37] Yang M H, Wu P, Feng H. MemSeg: a semi-supervised method for image surface defect detection using differences and commonalities[J]. *Engineering Applications of Artificial Intelligence*, 2023, 119: 105835.
- [38] Bae J, Lee J H, Kim S. PNI: industrial anomaly detection using position and neighborhood information[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 6350-6360.
- [39] Hyun J, Kim S, Jeon G, et al. ReConPatch: contrastive patch representation learning for industrial anomaly detection[C]//Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2024: 2041-2050.
- [40] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the International Conference on Machine Learning. London: PMLR, 2021: 8748-8763.
- [41] Zhou Q H, Pang G S, Tian Y, et al. AnomalyCLIP: object-agnostic prompt learning for zero-shot anomaly detection[PP]. arXiv preprint, 2026, arXiv: 2310.18961.
- [42] Li X F, Zhang Z Z, Tan X, et al. PromptAD: learning prompts with only normal samples for few-shot anomaly detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 16848-16858.
- [43] Gu Z P, Zhu B K, Zhu G B, et al. FiLo: zero-shot anomaly detection by fine-grained description and high-quality localization[C]//Proceedings of the 32nd ACM International Conference on Multimedia. New York: ACM, 2024: 2041-2049.
- [44] Yun E G, Ha H, Nam Y, et al. Language-assisted feature transformation for anomaly detection[PP]. arXiv preprint, 2025, arXiv: 2503.01184.
- [45] Gu Z P, Zhu B K, Zhu G B, et al. AnomalyGPT: detecting industrial anomalies using large vision-language models[J]//Proceedings of the AAAI conference on artificial intelligence, 2024, 38(3): 1932-1940.
- [46] Li Y Y, Zhang S T, Li K, et al. One-to-normal: anomaly personalization for few-shot anomaly detection[C]//Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024). Vancouver: NeurIPS, 2024: 78371-78393.
- [47] Deng H Q, Li X Y. Anomaly detection via reverse distillation from one-class embedding[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 9727-9736.
- [48] Tong G X, Li Q Q, Song Y. Enhanced multi-scale features mutual mapping fusion based on reverse knowledge distillation for industrial anomaly detection and localization[J]. *IEEE Transactions on Big Data*, 2024, 10(4): 498-513.
- [49] Wang G, Zou Y S, He S L, et al. Anomaly detection and localization via reverse distillation with latent anomaly suppression[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025, 35(10): 9592-9607.
- [50] Liu X Y, Wang J Y, Leng B, et al. Unlocking the potential of reverse distillation for anomaly detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2025, 39(6): 5640-5648.
- [51] Tien T D, Nguyen A T, Tran N H, et al. Revisiting reverse distillation for anomaly detection[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 24511-24520.
- [52] Sakurada M, Yairi T. Anomaly detection using autoencoders with nonlinear dimensionality reduction[C]//Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis. New York: ACM, 2014: 4-11.
- [53] Tao X, Zhang D P, Ma W Z, et al. Unsupervised anomaly detection

- for surface defects with dual-siamese network[J]. IEEE Transactions on Industrial Informatics, 2022, 18(11): 7707-7717.
- [54] Huang C Q, Xu Q W, Wang Y F, et al. Self-supervised masking for unsupervised anomaly detection and localization[J]. IEEE Transactions on Multimedia, 2022, 25: 4426-4438.
- [55] Zavrtanik V, Kristan M, Skočaj D. DSR: a dual subspace reprojection network for surface anomaly detection[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 539-554.
- [56] Zavrtanik V, Kristan M, Skočaj D. DRAEM - a discriminatively trained reconstruction embedding for surface anomaly detection[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 8310-8319.
- [57] Zhang X M, Xu M, Zhou X Z. RealNet: a feature selection network with realistic synthetic anomaly for anomaly detection[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 16699-16708.
- [58] Chen X L, Xu H T, Deng C H, et al. Implicit foreground-guided network for anomaly detection and localization[C]//Proceedings of the ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2024: 2970-2974.
- [59] Zhang R F, Wang H, Feng M Y, et al. Dual-constraint autoencoder and adaptive weighted similarity spatial attention for unsupervised anomaly detection[J]. IEEE Transactions on Industrial Informatics, 2024, 20(7): 9393-9403.
- [60] Zhou Z Y, Wang J C, Yu Z A, et al. FeatDAE: introducing features with denoising autoencoder for anomaly detection[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-14.
- [61] Luo W, Cao Y K, Yao H M, et al. Exploring intrinsic normal prototypes within a single image for universal anomaly detection[C]//Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2025: 9974-9983.
- [62] Schlegl T, Seeboeck P, Waldstein S M, et al. F-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks[J]. Medical Image Analysis, 2019, 54: 30-44.
- [63] Nakanishi H, Suzuki M, Matsuo Y. Iterative image inpainting with structural similarity mask for anomaly detection[C]//the 9th International Conference on Learning Representations. 2021.
- [64] Yan X D, Zhang H D, Xu X M, et al. Learning semantic context from normal samples for unsupervised anomaly detection[C]//Proceedings of the AAAI conference on artificial intelligence. Vancouver: AAAI Press, 2021, 35(4): 3110-3118.
- [65] Fang Q Q, Su Q L, Lyu W X, et al. Boosting fine-grained visual anomaly detection with coarse-knowledge-aware adversarial learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2025: 16532-16540.
- [66] Liang Y F, Zhang J N, Zhao S W, et al. Omni-frequency channel-selection representations for unsupervised anomaly detection[J]. IEEE Transactions on Image Processing, 2023, 32: 4327-4340.
- [67] Creswell A, Bharath A A. Inverting the generator of a generative adversarial network[J]. IEEE transactions on neural networks and learning systems, 2019, 30(7): 1967-1974.
- [68] Mou S C, Gu X Y, Cao M, et al. RGI: robust GAN-inversion for mask-free image inpainting and unsupervised pixel-wise anomaly detection[PP]. arXiv preprint, 2025, arXiv: 2302. 12464.
- [69] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems. New York: ACM, 2017: 5998-6008.
- [70] Pirnay J, Chai K. Inpainting transformer for anomaly detection[C]//Proceedings of the International Conference on Image Analysis and Processing. Cham: Springer, 2022: 394-406.
- [71] Mishra P, Verk R, Fornasier D, et al. VT-ADL: a vision transformer network for image anomaly detection and localization[C]//Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE). New York: IEEE, 2021: 1-6.
- [72] Iqbal E, Khan S U, Javed S, et al. Multi-scale feature reconstruction network for industrial anomaly detection[J]. Knowledge-Based Systems, 2024, 305: 112650.
- [73] Tang J L, Lv Z, Liu Y, et al. An improved transformer U-Net-based reconstruction method for unsupervised anomaly detection[C]//Proceedings of the 2024 14th Asian Control Conference (ASCC). Piscataway: IEEE Press, 2024: 153-158.
- [74] Luo W, Yao H M, Yu W Y. Template-based feature aggregation network for industrial anomaly detection[J]. Engineering Applications of Artificial Intelligence, 2024, 131: 107810.
- [75] Yao H M, Cao Y K, Luo W, et al. Prior normality prompt transformer for multiclass industrial image anomaly detection[J]. IEEE Transactions on Industrial Informatics, 2024, 20(10): 11866-11876.
- [76] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.
- [77] Wyatt J, Leach A, Schmon S M, et al. AnoDDPM: anomaly detection with denoising diffusion probabilistic models using simplex noise[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2022: 649-655.
- [78] Zhang H, Wang Z, Zeng D, et al. DiffusionAD: norm-guided one-step denoising diffusion for anomaly detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(8): 7140-7152.
- [79] Yao H, Liu M, Yin Z C, et al. GLAD: towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection[C]//Proceedings of 2024 the European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 1-17.
- [80] Mousakhan A, Brox T, Tayyub J. Anomaly detection with conditioned denoising diffusion models[C]//Proceedings of the 46th DAGM German Conference on Pattern Recognition. Cham: Springer Nature Switzerland, 2024: 181-195.
- [81] He H Y, Zhang J N, Chen H X, et al. A diffusion-based framework for multi-class anomaly detection[C]//Proceedings of the AAAI conference on artificial intelligence. Vancouver: AAAI Press, 2024, 38: 8472-8480.

[82] Beizae F, Lodyginsky G A, Desrosiers C, et al. Correcting deviations from normality: a reformulated diffusion model for multiclass unsupervised anomaly detection[C]//Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2025: 19088-19097.

[83] Fučka M, Zavrtanik V, Skočaj D. TransFusion: a transparency-based diffusion model for anomaly detection[C]//Proceedings of the 18th European Conference on Computer Vision. New York: ACM, 2024: 91-108.

[84] 侯壹凡, 苗壮, 王俊, 等. 融合色彩空间特征与频域特征的多源伪造图像检测方法[J]. 网络与信息安全学报, 2026, 12(1): 141-150.
Hou Y F, Miao Z, Wang J, et al. Multi-source forgery detection method based on the fusion of color space features and frequency domain features[J]. Chinese Journal of Network and Information Security, 2026, 12(1): 141-150.

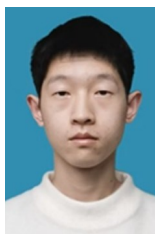
[作者简介]



许建 (1980-), 男, 博士, 南京邮电大学副教授, 主要研究方向为深度学习目标检测、网络空间安全等。



冯健聪 (2002-), 男, 南京邮电大学硕士生, 主要研究方向为工业图像异常检测。



侯富豪 (2002-), 男, 南京邮电大学硕士生, 主要研究方向为深度学习目标检测。



杨庚 (1961-), 男, 博士, 南京邮电大学教授, 主要研究方向为计算机通信与网络、并行与分布式计算、信息安全等。