

多模态融合下的非对称师生网络图像异常检测^①



李昊, 张英俊, 谢斌红, 张睿

(太原科技大学 计算机科学与技术学院, 太原 030024)

通信作者: 李昊, E-mail: gutaiii@163.com

摘要: 现有基于多模态的图像异常检测方法存在以下不足: 在异常区域提取阶段易出现异常平滑现象, 同时在缺陷检测过程中面临细粒度感知能力不足与判别效率低下的问题, 最终导致整体检测性能下降. 为此, 提出多模态融合下的非对称师生网络图像异常检测模型 (multimodal image anomaly detection with asymmetric teacher-student network, MATS), 包括跨模态异常放大器 (CAA)、多扩张率局部注意力 (MDLA) 模块和 FastKAN 前馈网络. 首先, 跨模态放大器通过扩展与压缩辅助特征, 与目标特征融合后放大异常区域并减少噪声, 解决后续检测时的异常平滑问题. 随后, MDLA 模块通过不同扩张率卷积并结合局部注意力提取多尺度特征以提高异常区域细粒度感知能力, 并结合归一化流 (NF) 生成正常样本的条件概率分布; FastKAN 模块通过更轻量化的特征处理以实现高效判别异常, 生成与教师输出一致的特征图, 用于逐像素距离计算以评估异常程度. 在测试阶段, 教师与学生网络输出差异较大的区域被判断为异常. 在公开的工业图像数据集 MVTEC AD 和 MVTEC 3D-AD 上的实验结果表明, 该方法在多模态异常检测和定位方面具有先进的性能.

关键词: 图像异常检测; 多模态; 归一化流; 局部注意力; FastKAN

引用格式: 李昊, 张英俊, 谢斌红, 张睿. 多模态融合下的非对称师生网络图像异常检测. 计算机系统应用, 2026, 35(2): 53-64. <http://www.c-s-a.org.cn/1003-3254/10045.html>

Multimodal-fusion-based Asymmetric Teacher-student Network for Image Anomaly Detection

LI Hao, ZHANG Ying-Jun, XIE Bin-Hong, ZHANG Rui

(School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China)

Abstract: Existing multimodal-based image anomaly detection methods suffer from several limitations: anomaly smoothing during anomaly region extraction, insufficient fine-grained perception, and low discrimination efficiency in defect detection, leading to degraded overall performance. To address these issues, this study proposes a multimodal image anomaly detection model with an asymmetric teacher-student network (MATS), comprising three key components: a cross-modal anomaly amplifier (CAA), a multi-dilated local attention (MDLA) module, and a FastKAN feed-forward network. First, the CAA amplifies anomalous regions while reducing noise by expanding/compressing auxiliary features and fusing them with target features, thus alleviating anomaly smoothing in subsequent detection. Subsequently, the MDLA module enhances fine-grained perception of anomalies through multi-dilation-rate convolutions combined with local attention for multi-scale feature extraction, while integrating normalizing flow (NF) to generate the conditional probability distribution of normal samples. The FastKAN module enables efficient anomaly discrimination via lightweight feature processing, producing feature maps consistent with the teacher network's outputs for pixel-wise distance calculation to evaluate anomaly scores. During testing, regions with significant discrepancies between teacher and student network outputs are identified as anomalies. Experimental results on public industrial image datasets MVTEC AD and

① 基金项目: 山西省基础研究计划 (20210302123216); 山西省产教融合研究生联合培养示范基地项目 (2022JD11); 吕梁市引进高层次科技人才重点研发项目 (2022RC08)

收稿时间: 2025-07-02; 修改时间: 2025-07-23; 采用时间: 2025-08-13; csa 在线出版时间: 2025-11-04

CNKI 网络首发时间: 2025-11-05

MVTec 3D-AD demonstrate that the proposed method achieves state-of-the-art performance in multimodal anomaly detection and localization.

Key words: image anomaly detection; multimodal; normalizing flow (NF); local attention; FastKAN

图像异常检测旨在从大规模图像数据中识别偏离正常模式的图像或特定区域,异常包括但不限于形变、缺损、污渍、凹凸不平、纹理错位等不符合训练集中正常样本分布的局部特征^[1],如图1所示.在工业质检^[2]、医疗影像分析^[3]等领域有广泛的应用,通常分为有监督和无监督方法两大类.有监督方法依赖于标注好的异常样本进行训练,常见于分类模型,但在实际场景中,异常样本往往稀缺且难以获取,因此无监督方法得到了更多的关注.当前主流的无监督方法主要有基于自编码器(autoencoder)^[4]和生成对抗网络(GAN)^[5]框架,其核心思想是通过建立正常样本的特征分布,进而检测显著偏离该分布的异常样本.尽管图像异常检测方法在近年来取得了显著进展,但现有技术在处理多模态数据、捕捉局部异常特征以及高效特征处理方面仍存在诸多不足.

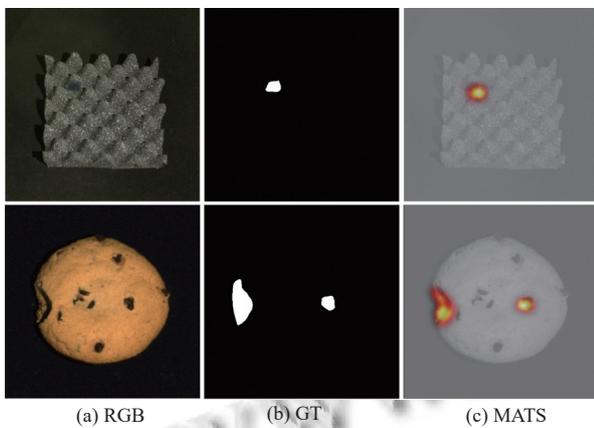


图1 图像异常检测示意图

首先,多模态数据融合阶段存在信息耦合不充分的问题.在工业检测和医疗影像分析中,图像数据通常与深度信息或其他模态数据结合使用.然而,现有方法往往忽视模态间的协同作用,仅简单地进行特征拼接或加权平均,无法充分利用多模态数据中的互补信息.此外,在多模态融合过程中,如何有效放大异常区域的特征并抑制噪声依然是未被充分解决的难题,尤其是在异常平滑问题的影响下,异常区域容易被正常区域或背景的相似特征所掩盖.因此,需要一种能够利用模

态间关联知识来增强异常特征的方法.

其次,现有方法缺乏对局部细节与多尺度特征的敏感建模.主流的图像异常检测技术,如自编码器和GAN,多以全局特征为核心,倾向于学习图像整体的分布规律.然而,在工业质检和医疗场景中,异常往往是微小而局部的,例如细微的裂纹、瑕疵或病灶,这种局部特征的检测依赖于对多尺度上下文信息的充分建模.但现有单一尺度的特征提取方法难以满足这一需求,而传统卷积神经网络对局部特征的建模能力也存在一定局限性.因此,需要引入能够捕捉多尺度信息并增强细粒度感知能力的模块,以提升模型对局部异常的敏感性.

此外,高效特征处理和精确异常评估仍是挑战.随着工业和医疗场景中数据复杂性的增加,模型不仅需要保证检测精度,还需要提高计算效率.然而,一些复杂的网络架构容易引入较高的计算成本,例如在师生网络的设计中,学生网络如何高效生成与教师网络输出一致的特征图,并实现对异常区域的精确评估,是一项重要的研究方向.同时,现有方法在逐像素计算中可能存在信息冗余,如何通过合理的网络设计实现更轻量化和高效的特征处理,仍需深入探索.

基于上述分析,本文提出跨模态异常放大器(cross-modal anomaly amplifier, CAA),旨在解决模态融合过程中异常特征放大的难题(本文所指多模态主要包括RGB图像与对应的深度信息,分别提供表面纹理信息与三维结构线索).多扩张率局部注意力(multi-dilated local attention, MDLA)模块专注于多尺度特征的捕捉和局部异常的建模.而FastKAN^[6]前馈网络则进一步提升了异常评估的效率与精度.这些方法的改进设计,不仅有效应对了现有技术中的局限性,也为图像异常检测领域提供了新的思路.

综上所述,本文的主要贡献如下.

(1) 提出一种作用于异常特征的跨模态放大器(CAA),RGB图像和深度信息各自作为目标模态,再利用辅助模态通过反向瓶颈投影扩展与压缩特征放大异常区域,减少噪声,并与目标模态特征融合,解决检测

时的异常平滑问题。

(2) 提出一种用于捕捉局部依赖关系和细微变化的多扩张率注意力方法, 通过结合膨胀卷积的多尺度建模能力与局部自注意力机制, 有效提升输入特征的局部语义表达, 弥补 CAA 模块在特征融合时可能忽略的局部细节。将归一化流 (normalizing flow, NF)、FastKAN 集成于非对称师生网络^[7]架构中, 解决像素级异常定位中的细粒度缺陷感知与高效判别问题。

(3) 在工业数据集 MVTEC AD^[2]和 MVTEC 3D-AD^[8]上的实验结果验证了本文模型的有效性和优越性。

1 相关工作

1.1 图像异常检测方法

图像异常检测的核心挑战在于如何构建正常样本的紧凑特征空间并识别偏离分布的模式, 主要以无监督方法为主, 基于特征嵌入差异或重构误差进行异常评分。大多数特征嵌入方法往往聚焦于全局语义一致性建模, 忽略了局部细节的表达, 尤其是在空间结构复杂或缺陷微弱的场景中, 定位能力显著下降。STPM^[9]和 STFPM^[10]等方法通过特征层对齐实现对异常区域的响应, 但缺乏对局部上下文和多尺度结构的敏感性, 容易在纹理复杂或缺陷微小的场景中产生定位偏差。RD4AD^[11]和 AST^[7]等改进方法引入特征增强机制或正则化建模以提升鲁棒性, 但代价是显著增加计算复杂度, 限制了在工业部署中的实时应用潜力。

重构方法如 DRAEM^[12]通过对抗训练合成异常样本, 利用 U-Net 架构实现像素级定位, 但在处理与正常样本高度相似的细微异常时容易产生误检。基于 Transformer 的重构方法 (如 VT-ADL^[13]和 ADTR^[14]) 利用全局注意力机制捕捉长程依赖, 在复杂纹理异常检测中表现突出, 但参数量大限制了工业部署。此外, 部分研究指出^[8,15], 多模态输入 (如 RGB 与深度图) 的融合若缺乏细粒度的语义对齐机制, 容易导致异常区域响应减弱或被背景信息掩盖, 从而降低检测精度。

为更精准解决像素级异常定位中的细粒度缺陷感知与高效判别问题, 本文在非对称师生网络框架中集成多扩张率局部注意力与 FastKAN^[6], 在继承 AST^[7]全局异常判别优势的同时, 显著提升复杂纹理缺陷的定位鲁棒性与计算效率。

1.2 多模态融合方法

在工业异常检测中, 多模态数据的有效融合是提

升模型鲁棒性的关键。传统方法主要依赖单一模态 (如 RGB 图像) 进行异常检测, 但实际工业场景中, 复杂光照、材质反光等因素可能导致单模态信息不充分, 为此, 研究者尝试引入点云和深度信息等多模态数据以增强特征表征能力。例如, Bergmann 等人^[8]提出基于 MVTEC 3D-AD 数据集构建 3D 异常检测框架, 通过点云与 RGB-D 数据融合捕捉空间几何特征, 但其方法依赖手工设计的局部描述符 (如 FPFH^[16]), 泛化能力受限。Horwitz 等人^[17]提出结合传统 3D 特征与深度学习方法 (如 PatchCore^[18]), 通过特征拼接实现多模态融合, 但未解决跨模态特征对齐问题。

针对多模态数据异质性挑战, AST 首次在教师-学生架构中引入深度信息, 通过非对称网络设计增强异常特征差异, 但未充分利用多模态间的互补性, 忽略了少部分异常特征。近期研究进一步探索跨模态生成模型的应用, 如 AnoDDPM^[19]将扩散模型引入工业异常检测, 通过模拟多模态噪声分布提升异常敏感性, 但其推理速度受迭代采样限制。此外, FYD^[20]提出由粗到细的对齐策略, 在图像级和特征级实现跨模态配准, 但存在动态生产环境中模态缺失的鲁棒性及异常平滑问题。

本文针对跨模态特征融合中的异常信号衰减问题, 提出跨模态放大器架构, 通过反向瓶颈投影实现特征扩展与压缩, 以目标模态特征为基准, 利用辅助模态特征动态放大异常区域响应, 同时抑制噪声干扰。

2 模型

2.1 总述

非对称师生架构的跨模态异常检测模型旨在通过跨模态特征增强与师生输出差异, 实现高精度的异常定位。MATS 框架如图 2 所示, 模型由 3 部分组成: 跨模态异常放大器、教师网络和学生网络。跨模态异常放大器 (CAA) 通过反向瓶颈投影机制, 从 RGB 图像和深度图像中提取并放大异常区域特征, 同时减少模态间的噪声干扰, 生成融合后的高质量特征, 为后续处理提供准确的信息。教师网络作为基准, 通过多扩张率局部注意力模块 (MDLA) 提取多尺度局部特征, 同时利用归一化流 (NF) 对特征分布进行建模, 输出高质量的概率分布, 为学生网络提供学习目标。学生网络采用轻量化的 FastKAN 模块, 通过逐像素距离计算异常位置, 生成与教师网络一致的特征图。非对称设计有效缓解传统师生网络中的过拟合问题, 同时保证推理效率。

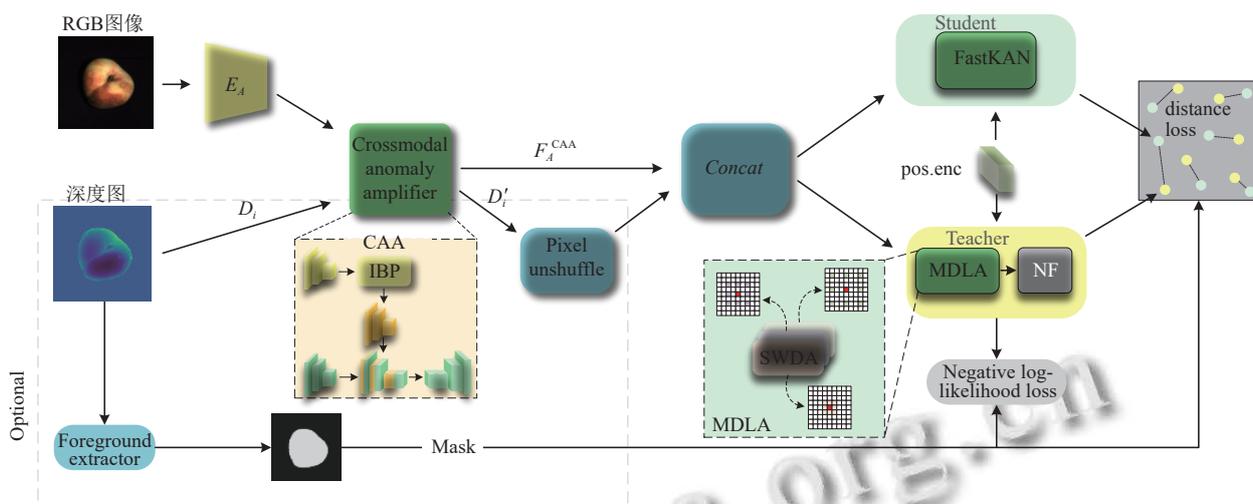


图2 模型整体架构

MATS 的初始输入由 RGB 图像及其对应的深度信息组成. RGB 图像通过 ImageNet 预训练网络提取语义特征, 生成低分辨率特征图, 深度信息则首先输入跨模态异常放大器 (CAA) 进行噪声抑制与异常增强, 随后通过像素重组操作 (pixel unshuffle) 实现与 RGB 特征空间分辨率对齐. 最终, 经 CAA 处理后的 RGB 特征与重组深度特征沿通道维度拼接, 形成师生网络的联合输入特征.

为了增强 MATS 对输入图像空间信息的感知能力, 在 NF 模块和 FastKAN 模块中分别为拼接后的 RGB 图像与深度图像引入正弦位置编码 (sinusoidal positional encoding). 位置编码 (pos. enc) 被添加到特征图中, 使特征与其空间位置相关联, 使得模型能够更好地感知和捕捉位置异常, 从而提高异常检测的精度. 位置编码的具体实现与其在归一化流 (NF) 中的应用将在第 2.3 节中详细讨论.

此外, 深度图像还用于前景提取, 通过生成前景掩码 (foreground mask) 聚焦于潜在的异常区域, 这种方法特别适合于实际应用中的静态或平面背景情况, 有效地减少了背景噪声对模型性能的影响.

为了进一步优化模型, 采用掩盖距离 (masked distance loss) 和负对数似然损失 (negative log-likelihood loss) 两种损失函数, 这些损失函数利用前景掩码屏蔽背景像素, 使模型专注于检测异常区域. 相关的损失函数设计与计算将在第 2.3、2.4 节中详细介绍.

2.2 跨模态异常放大器模块

跨模态异常放大器 (CAA) 通过跨模态交互机制增

强异常区域特征, 旨在有效利用模态间的互补信息, 提高多模态异常检测的性能. CAA 模块如图 3 所示, 由反向瓶颈投影 (inverted bottleneck projection, IBP) 和跨模态特征融合 (crossmodal feature fusion) 两部分组成, 分别负责异常特征的放大和模态特征的融合, 并通过解码生成最终的异常特征.

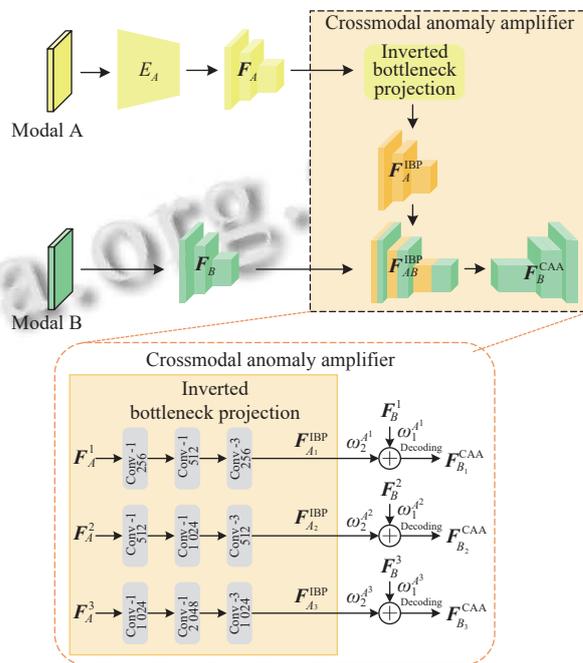


图3 跨模态异常放大器

反向瓶颈投影模块的核心任务是对辅助模态的特征进行空间扩展和压缩, 以放大异常特征并减少冗余信息. 以模态 A 辅助模态 B 为例, 辅助模态 A 的编码特征 F_A^i 首先通过 1×1 卷积扩展特征通道至原始维度

的两倍. 这一扩展操作增强了特征的表达能力, 特别是对异常区域的表征能力. 接着, 扩展后的特征通过 3×3 卷积进行压缩, 将通道数还原至原始维度, 同时抑制扩展过程中引入的噪声或冗余信息. 压缩后的特征 $F_{A_i}^{IBP}$ 被映射至与目标模态 B 特征 F_B^i 语义一致的特征空间, 为后续融合操作提供了高质量的辅助特征. 整个过程通过卷积核参数学习, 确保投影后的辅助模态特征能够对目标模态的异常检测提供有效支持.

跨模态特征融合模块用于将目标模态的编码特征 F_B^i 与辅助模态的投影特征 $F_{A_i}^{IBP}$ 进行融合, 并解码生成最终的异常特征图. 为了确保两种模态特征在融合过程中的有效性, 模块引入了动态加权机制, 具体而言, 两个模态特征分别通过加权系数 ω_1 和 ω_2 进行调节, 而动态加权系数 ω_1 和 ω_2 通过可学习的全连接层生成, 权重参数通过网络训练自适应地学习, 以控制目标模态和辅助模态在不同场景下的贡献比例. 其计算过程为:

$$\omega_1^{A_i} = \sigma(W_1 \cdot \text{ReLU}(F_B^i)) \quad (1)$$

$$\omega_2^{A_i} = \sigma(W_2 \cdot \text{ReLU}(F_{A_i}^{IBP})) \quad (2)$$

其中, σ 为 Sigmoid 函数, W_1 和 W_2 为权重矩阵, 确保辅助模态特征在异常区域响应中占据主导地位.

动态加权机制的设计确保了辅助模态在异常区域检测中的作用, 即使目标模态自身表现正常, 融合结果仍能够充分利用辅助模态的异常信号进行增强. 融合后的特征 $F_{B_i}^{CAA}$ 通过解码器进一步处理, 生成用于异常检测的输出特征图, 融合特征解码后的输出 $F_{B_i}^{CAA}$ 的计算公式如下:

$$F_{B_i}^{CAA} = \text{Decoder} \left(\frac{\exp(\omega_1^{A_i}) \cdot F_B^i + \exp(\omega_2^{A_i}) \cdot F_{A_i}^{IBP}}{\exp(\omega_1^{A_i}) + \exp(\omega_2^{A_i})} \right) \quad (3)$$

为确保不同模态特征在融合过程中的语义一致性, 本文从结构映射与语义调节两个角度设计了特征对齐策略. 一方面, IBP 模块通过卷积映射操作将辅助模态的原始特征压缩至目标模态语义空间, 确保空间维度和通道结构的一致性, 为语义对齐提供基础. 另一方面, 动态加权机制中的权重系数通过全连接层联合两种模态的全局特征学习获得, 并由 Sigmoid 函数归一化处理, 确保不同模态特征在融合前的数值尺度和响应趋势保持一致, 从而达到自适应语义对齐的效果. 该机制无需显式配准或外部标签指导, 能够在端到端训练过

程中自动调整两个模态的贡献比例, 增强异常区域响应的同时抑制无效冗余.

异常放大后的深度图像要进行像素重组操作, 具体来说, 就是将深度图像中每个 $d \times d$ 像素块重组为单一像素, 并沿通道维度扩展为 d^2 个通道, 使其与 RGB 特征图的空间维度一致, 然后与 RGB 特征进行通道维度上的融合, 进入师生网络.

2.3 教师网络

教师网络由多扩张率局部注意力 (MDLA) 模块和归一化流 (NF) 组成. MDLA 模块通过捕捉输入特征在不同尺度下的局部信息来增强特征表示能力, 而 NF 模块则用于建模和变换特征的概率分布, 以更好地区分正常和异常样本. 结合两者, 教师网络在异常检测中表现出更强的敏感性和特征建模能力.

2.3.1 多扩张率局部注意力模块

多扩张率局部注意力 (MDLA) 模块通过结合膨胀卷积的多尺度建模能力与局部自注意力机制, 有效提升输入特征的局部语义表达. MDLA 结构图如图 4 所示.

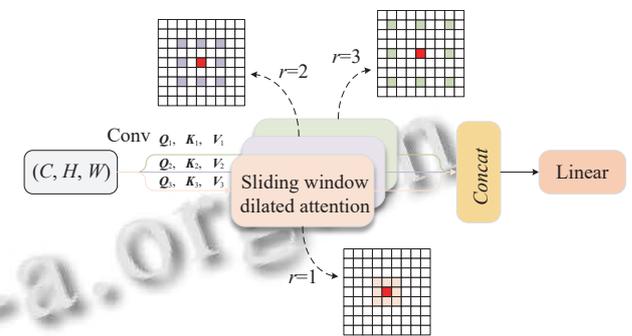


图4 多扩张率局部注意力

针对解码器输出的特征图 $X \in \mathbb{R}^{C \times H \times W}$, 首先通过线性投影操作将其映射为 Q, K, V , 这一映射使得原始特征能够以标准自注意力框架进行处理, 同时引入了必要的参数变换以适配多尺度局部建模的需求. 具体来说, 给定输入特征图 X , 采用 3 个独立的 1×1 卷积层分别生成:

$$Q = XW_Q, K = XW_K, V = XW_V \quad (4)$$

其中, W_Q, W_K 和 W_V 为可学习的权重矩阵.

对于每个在特征图上的查询位置 (i, j) , 定义一个大小为 $\omega \times \omega$ 的局部滑动注意力窗口 (SWDA), 在该窗口内进行局部注意力计算. 结合不同膨胀率的采样策

略(即膨胀注意力),使得局部窗口能够感知更大范围的上下文信息,从而在局部区域内建立长距离依赖,并在多尺度下有效捕捉特征模式.采用膨胀率 $r \in \mathbb{N}^+$ 对窗口内的键和值进行稀疏采样,采样坐标满足:

$$(i', j') = \left\{ (i + p \cdot r, j + q \cdot r) \mid -\frac{w}{2} \leq p, q \leq \frac{w}{2} \right\} \quad (5)$$

其中, p 、 q 为整数偏移量.通过调节膨胀率 r ,滑动窗口可以在保持固定大小的同时扩大感受野,例如,当 $\omega = 3$ 且 $r = 2$ 时,有效覆盖的邻域范围为 5×5 .边缘位置的查询点通过零填充策略维持特征图分辨率,稀疏采样后的键值矩阵通过标准注意力机制计算最终输出.

$$x_{ij} = \text{Attention}(q_{ij}, \mathbf{K}_r, \mathbf{V}_r) = \text{Softmax}\left(\frac{q_{ij} \mathbf{K}_r^T}{\sqrt{d_k}}\right) \mathbf{V}_r \quad (6)$$

其中, \mathbf{K}_r 和 \mathbf{V}_r 分别为键、值矩阵, d_k 为特征维度.

为了充分利用多尺度特征,MDLA模块将特征图的通道划分为多组注意力头,每组头内采用不同的膨胀率执行SWDA操作.假设总头数为 n ,每组的膨胀率集合为 r_1, r_2, \dots, r_n ,则每组子特征的注意力计算如下:

$$h_i = \text{SWDA}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i, r_i), 1 \leq i \leq n \quad (7)$$

其中, \mathbf{Q}_i 、 \mathbf{K}_i 、 \mathbf{V}_i 为第 i 组的查询、键和值特征, r_i 为对应的膨胀率.默认设置下, $r = [1, 2, 3]$,感受野大小 3×3 、 5×5 、 7×7 .这种多尺度设计使得模块能够自适应地捕获不同尺度的异常模式.

所有注意力头的输出沿通道维度拼接后,经过线性层进一步融合多尺度特征,生成最终输出特征.

$$\mathbf{X} = \text{Linear}(\text{Concat}(h_1, h_2, \dots, h_n)) \quad (8)$$

2.3.2 归一化流模块

归一化流(NF)作为教师网络的重要组成部分,旨在通过一系列双射映射操作,将输入特征空间转换为标准正态分布 $\mathcal{N}(0, I)$,以建模正常数据的概率密度分布并为异常检测提供依据.NF模块的核心任务是利用正常数据的全局分布特性,为学生网络提供精准的目标指导,同时捕捉异常区域的显著特征.NF模块如图5所示.

NF模块的初始输入为特征图 $\mathbf{X} \in \mathbb{R}^{n_{\text{feat}} \times h \times w}$,在固定排列并偶信道分割后,与正弦位置编码进行拼接,确保特征的语义信息与其空间位置信息相结合.其过程如下,特征图 $\mathbf{X} \in \mathbb{R}^{n_{\text{feat}} \times h \times w}$ 和正弦位置编码 $\mathbf{P} \in \mathbb{R}^{n_{\text{pos}} \times h \times w}$ 在通道维度进行拼接,则输入NF的特征为:

$$x = \text{Concat}(\mathbf{X}, \mathbf{P}) \in \mathbb{R}^{(n_{\text{feat}} + n_{\text{pos}}) \times h \times w} \quad (9)$$

其中, h 、 w 表示特征图的高度和宽度, n_{feat} 和 n_{pos} 分别表示特征通道数与位置编码通道数.拼接后的特征既包含异常区域的语义表征,也与位置相关的几何特性相适配,有助于检测与位置错位相关的异常.

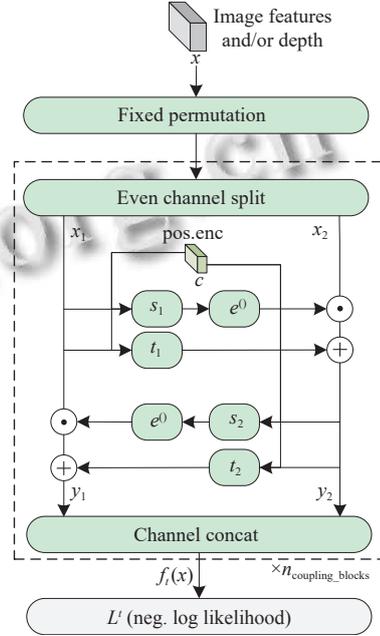


图5 归一化流

NF模块基于Real-NVP(实值非体积保持)架构,通过多个仿射耦合层(affine coupling layer)完成密度变换.每个仿射耦合层将输入特征通道划分为两部分 x_1 和 x_2 ,并利用以下公式对其进行双射映射:

$$y_1 = x_1 \odot \exp(s_2(x_2)) + t_2(x_2) \quad (10)$$

$$y_2 = x_2 \odot \exp(s_1(x_1)) + t_1(x_1) \quad (11)$$

其中, s_i 和 t_i 是子网络生成的缩放与平移参数函数,采用轻量化的卷积网络实现.由于拼接后的位置编码信息被直接嵌入到输入特征中, s_i 和 t_i 函数不再显式接受位置编码作为条件输入.

在仿射变换过程中, NF模块通过控制变换的缩放系数范围来确保训练过程的稳定性,例如采用Alpha-clamping技术限制缩放系数,避免数值溢出;利用Jacobian行列式进行高效计算.根据变量变换公式, NF模块通过最大化输入特征的对数似然实现优化,其目标函数为:

$$\log p_X(x) = \log p_Z(z) + \log \left| \det \frac{\partial z}{\partial x} \right| \quad (12)$$

其中, z 为通过仿射耦合层变换后的特征, $p_Z(z)$ 是标准正态分布的概率密度函数. 损失函数被定义为:

$$L_t = -\log p_X(x) = \frac{\|z\|_2^2}{2} - \log \left| \det \frac{\partial z}{\partial x} \right| \quad (13)$$

通过优化, NF 模块能够学习正常样本的密度分布. 输出既可作为异常检测的概率参考, 又可作为学生网络的学习目标, 为其提供高质量的目标特征, 用于指导异常区域的逐像素训练与评估.

2.4 学生网络

学生网络的输入由 RGB 和深度信息融合后的特征图与基于正弦函数的空间位置编码拼接而成. 这些拼接特征随后经过学生网络的一系列特征处理模块, 生成最终输出特征图, 与教师网络输出对齐, 为像素级异常计算提供基础.

如图 6 所示, 学生网络的结构由多个残差块组成, 每个残差块包含两层 FastKAN 变换、批量归一化 (batch normalization) 和 LeakyReLU 激活函数. 相比于传统卷积, FastKAN 采用径向基函数 (RBF) 进行特征变换, 具备更强的非线性建模能力, 同时避免了固定权重共享的局限性, 使特征映射更具自适应性. 此外, 其计算方式无需滑动窗口操作, 减少了冗余计算, 提高了计算效率. 输入特征首先通过首层 FastKAN 模块映射到更高维度的隐空间, 以提升网络的特征提取能力. 中间层块通过逐步提取与融合特征, 捕获异常区域的细粒度信息, 同时抑制冗余特征. 最终, 末层 FastKAN 模块将特征图维度降到与教师网络输出一致, 以便后续逐像素的异常差异计算.

学生网络的训练目标是 minimized 其输出与教师网络输出之间的像素级平方欧氏距离. 对于给定的训练样本 x , 在像素位置 (i, j) 的损失函数 (masked distance loss) 定义为:

$$L_s^{ij} = \|f_s(x)_{ij} - f_t(x)_{ij}\|_2^2 \quad (14)$$

其中, $f_s(x)$ 和 $f_t(x)$ 分别表示学生网络和教师网络的输出特征图. 在训练过程中, 损失仅针对前景像素计算, 并对所有前景像素的损失取平均, 作为优化目标. 通过这一过程, 学生网络能够高效拟合正常数据的分布特性.

在测试阶段, 异常检测通过计算学生网络输出与教师网络输出之间的逐像素差异实现. 对于每个测试样本, 计算像素位置的平方距离, 得到像素级异常分数, 并进一步通过聚合方法生成样本级异常分数. 为更稳

定地反映异常程度, 本文采用 top- $k\%$ 聚合策略对像素级热图进行图像级评分.

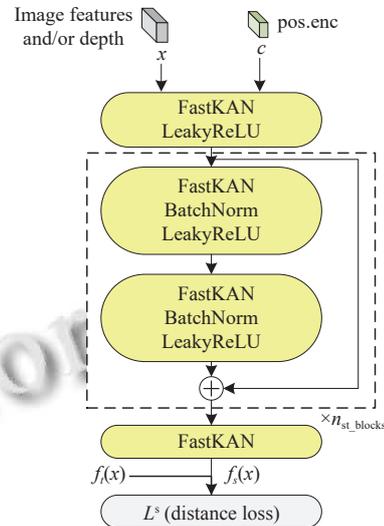


图 6 学生网络

3 实验

3.1 数据集与评价指标

本研究的实验部分在两个数据集上进行: MVTEC AD 和 MVTEC 3D-AD. MVTEC AD 数据集包含 10 个物体类别和 5 个纹理类别, 总共 73 种缺陷类型, 图像尺寸在 700–1024 像素之间, 训练集中仅包含无缺陷样本, 每个类别的样本数量在 60–320 之间; 而 MVTEC 3D-AD 数据集则包含 10 个类别, 除 RGB 图像外还提供与之对应的深度图数据, 图像边长范围为 400–800 像素, 训练样本数量在 210–300 之间, 测试集则同时包含正常样本和含缺陷样本. 由于两种数据集的特点各异, 本研究在仅含 RGB 数据以及 3D 数据融合的场景下均进行了全面评估.

为实现异常区域的二值化判定, 本文在像素级采用最大 F1 准则设定热图阈值, 图像级异常得分则基于 top- $k\%$ 聚合策略, 均参考现有工作 (如 AST、PatchCore) 中的通用做法, 具体参数在验证集上选取.

评价指标上, 采用了常用的接收者操作特性曲线下面积 (AUROC) 作为异常检测性能的衡量标准. 对于图像级别检测, AUROC 可以有效反映不同阈值下的真阳性率和假阳性率的变化, 且与类别不平衡无关; 而在缺陷定位的评估中, 像素级 AUROC 被用于量化模型对缺陷区域的准确分割.

3.2 参数设置

所有输入图像均调整至 768×768 像素,并通过预训练的 EfficientNet-B5 提取特征,生成尺寸为 24×24、通道数为 304 的特征图.对于 3D 数据,先进行像素重组以匹配 RGB 特征图的分辨率,同时利用 3D 信息生成前景掩码,用于屏蔽背景区域.

教师网络采用基于 Real-NVP 的归一化流架构,共包含 4 个耦合块.在耦合块内部,针对 MVTEC AD 数据集,子网络隐藏通道数设为 1 024,而在 MVTEC 3D-AD 数据集中设为 64;同时,Alpha-Clamping 参数在 MVTEC AD 上取 3,在 MVTEC 3D-AD 上取 1.9.教师网络的训练采用 Adam 优化器,参数设置为 $\beta_1=0.9$ 、 $\beta_2=0.999$,学习率为 2×10^{-4} ,权重衰减为 10^{-5} ;训练周期分别为 240 个 epoch (MVTEC AD) 和 72 个 epoch (MVTEC 3D-AD).

学生网络采用全卷积结构,由 4 个残差块构成,每个残差块包含两个 FastKAN 变换、批量归一化和 LeakyReLU 激活函数,首尾变换层用于调整特征维度以匹配教师网络输出.学生网络的训练目标为最小化像素级平方欧氏距离,同样使用 Adam 优化器,参数设置与教师网络保持一致.

3.3 实验结果与分析

实验在配备 NVIDIA 3090 显卡、8V CPU、30 GiB 内存和 24 GB 显存的服务器上进行,开发环境基于

PyTorch 框架,操作系统为 Ubuntu 22.04.所有数据预处理、模型训练及评估均在统一的硬件与软件平台下完成,确保结果的可重复性和公平性.

3.3.1 MVTEC AD 对比实验

本节基于 MVTEC AD 数据集开展了对比实验,涵盖 15 类工业对象及 10 种主流无监督异常检测方法,以图像级 AUROC (I-AUROC) 作为性能指标,结果如表 1 所示,最佳结果用加粗显示. MATS 在 Grid 类别上以 99.9% 的精度超越 ARNet (88.3%) 和 GAN (70.8%),较 PatchCore (98.2%) 提升 1.7 个百分点,验证了动态特征选择机制对细微纹理异常的捕捉能力;在 MetalNut 上以 98.8% 的精度优于 AST (98.5%),表明跨尺度特征融合策略有效抑制了光噪干扰;在 Capsule 上以 98.1% 的精度与 PatchCore (98.1%) 持平,但较 GAN (73.2%) 提升 24.9 个百分点,较 CFlow (92.5%) 提升 5.6 个百分点,证明多层次特征提取架构能更精准定位局部异常区域;在 Pill 上以 99.2% 的精度超越所有模型. MATS 在多数检测类别中达到或接近最优水平,平均精度达 99.0%,展现出卓越的泛化能力与稳定性.值得指出的是,本文主要的检测对象是 3D 图像,面向的是 RGB 与深度信息的多模态融合场景,为此,模型中加入了跨模态异常放大器模块,在 RGB 单模态下可能引入轻微特征偏移,这也是在检测 RGB 图像时性能略有波动的主要原因.

表 1 不同模型在 MVTEC AD 数据集上的性能对比 (%)

Method	Grid	Leather	Tile	Carpet	Wood	Bottle	Capsule	Pill	Transistor	Zipper	Cable	Hazelnut	MetalNut	Screw	Toothbrush	Avg
ARNet ^[21]	88.3	86.2	73.5	70.6	92.3	94.1	68.1	78.6	84.3	87.6	83.2	85.5	66.7	100	100	83.9
GAN ^[5]	70.8	84.2	79.4	69.9	83.4	89.2	73.2	74.3	79.2	74.5	75.7	78.5	70.0	74.6	65.3	76.2
Rippel ^[22]	93.7	100	100	99.6	99.2	99.0	96.3	91.4	98.2	98.8	99.1	100	97.4	94.5	94.1	97.5
PatchCore ^[18]	98.2	100	98.7	98.7	98.8	100	98.1	96.6	100	99.4	99.5	100	100	98.1	100	99.1
DifferNet ^[23]	84.0	97.1	99.4	92.9	92.9	99.0	86.9	88.8	91.1	95.1	95.5	99.3	96.1	96.3	98.6	94.7
CFlow ^[24]	99.6	100	99.9	98.7	98.7	100	97.7	96.8	95.2	98.5	97.6	100	99.3	91.9	99.7	98.3
CS-Flow ^[25]	99.0	99.9	100	100	100	99.8	97.1	98.6	99.3	99.7	99.1	99.6	99.1	97.6	91.9	98.7
Uninf.Stud. ^[26]	98.1	94.7	94.7	99.9	99.1	99.0	92.5	92.2	79.4	94.4	78.7	99.1	89.1	86.0	100	93.2
STFPM ^[10]	100	95.5	95.5	98.9	99.2	100	88.0	93.8	93.7	93.6	92.3	100	100	88.2	87.7	95.4
AST ^[7]	99.1	100	100	97.5	100	100	99.7	99.1	99.3	99.1	98.5	100	98.5	99.7	96.6	99.2
MATS	99.9	100	99.7	99.0	99.4	99.9	98.1	99.2	99.0	99.4	98.1	99.5	98.8	99.2	96.1	99.0

3.3.2 MVTEC 3D-AD 对比实验

为验证 MATS 在复杂 3D 工业异常检测任务中的有效性,本节基于 MVTEC 3D-AD 数据集开展了系统性对比实验,涵盖 10 类工业对象及 9 种主流无监督异常检测方法.实验采用广泛认可的 I-AUROC (%) 和

PRO (%) 作为量化指标,表 2 详细展示了各方法在单类别及整体均值上的性能对比(最后一列左为 I-AUROC,右为 PRO,其余均为 I-AUROC).

从表 2 中可以看出, MATS 在 Cable Gland、Carrot 和 Foam 这 3 个类别上分别获得了 95.1%、98.5% 和

94.8% 的检测精度, 在多数方法中表现突出. 与 DepthGAN 相比, Carrot 类别的性能提升达 40.5 个百分点 (98.5% vs. 58.0%), Foam 类别提升 41.4 个百分点 (94.8% vs. 53.4%). 在 Rope 类别上, MATS 达到 100% 的检测精度, 与多种方法持平. 其在 Cookie、Dowel、Peach 等类别中也表现稳定, 检测精度均在 93% 以上. 在 Bagel、Potato 和 Tire 等类别上, MATS 的性能亦保

持较高水平, 例如在 Bagel 类别达到 98.2%, 与其他方法差距较小; 在 Potato 上为 96.0%, 优于多种传统深度模型. 从整体平均性能来看, MATS 在 I-AUROC 和 PRO 指标上分别达到 95.5% 和 97.7%. 该结果表明, MATS 在图像级检测与像素级区域定位任务中都具备良好的泛化能力, 能够较好地适应不同物体结构和缺陷类型, 展现出一定的稳定性与实用性.

表 2 不同模型在 MVTec 3D-AD 数据集上的性能对比 (%)

Method	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Avg
DepthGAN ^[8]	53.8	37.2	58.0	60.3	43.0	53.4	64.2	60.1	44.3	57.7	53.2, 47.4
DepthAE ^[8]	64.8	50.2	65.0	48.8	80.5	52.2	71.2	52.9	54.0	55.2	59.5, 48.1
DepthVM ^[8]	51.3	55.1	47.7	58.1	61.7	71.6	45.0	42.1	59.8	62.3	55.5, 33.5
BTF ^[17]	91.8	74.8	96.7	88.3	93.2	58.2	89.6	91.2	92.1	88.6	86.5, 95.9
M3DM ^[27]	99.4	90.9	97.2	97.6	96.0	94.2	97.3	89.9	97.2	85.0	94.5, 96.4
Shape-guided ^[28]	98.6	89.4	98.3	99.1	97.6	85.7	99.0	96.5	96.0	86.9	94.7, 97.6
CFM ^[29]	99.4	88.8	98.4	99.3	98.0	88.8	94.1	94.3	98.0	95.3	95.4, 97.1
AST ^[7]	98.3	87.3	97.6	97.1	93.2	88.5	97.4	98.1	100	79.7	93.7, 94.4
MMRD ^[30]	99.9	94.3	96.4	94.3	99.2	91.2	94.9	90.1	99.4	90.1	95.0, 97.6
MATS	98.2	95.1	98.5	95.1	93.7	94.8	97.4	96.0	100	86.2	95.5, 97.7

3.3.3 消融实验

为系统评估 CAA 模块与 MDLA 模块的有效性, 本节在 MVTec AD 和 MVTec 3D-AD 数据集上开展消融实验, 通过逐级叠加模块的方式对基线模型 AST 进行优化, 以 I-AUROC 作为量化指标, 如表 3 所示.

表 3 模块消融实验 (%)

Model	MVTec AD	MVTec 3D-AD
AST	99.2	93.7
AST+CAA	97.6	94.8
AST+MDLA	99.3	93.9
AST+CAA+MDLA	99.0	95.5

在 MVTec 3D-AD 数据集上, 单独引入 CAA 模块将 I-AUROC 从 93.7% 提升至 94.8%, 表明在 RGB 与深度图像协同输入的条件下, CAA 能有效放大异常区域, 增强检测敏感性. MDLA 模块引入后, 性能略有提升 (93.9%), 显示其在多尺度局部建模方面具备一定补充作用. 当同时使用 CAA 与 MDLA 两个模块时, 检测精度进一步提升至 95.5%, 在 3D 场景下展现出良好的协同增益效果. 在 MVTec AD 数据集 (RGB 图像) 上, MDLA 模块对基线模型具有一定增强效果, I-AUROC 从 99.2% 提升至 99.3%. 相比之下, 单独加入 CAA 模块后性能有所下降 (97.6%), 这是由于该模块依赖多模态交互机制, 在单模态条件下难以充分发挥作用. 但当两个模块联合使用时, I-AUROC 恢复至 99.0%, 接近原

始性能, 说明 MDLA 在一定程度上弥补了 CAA 在单模态场景中的性能损失.

表 4 为评估位置编码 (pos.enc) 和前景掩码 (mask) 在 3D+RGB 输入场景下的检测性能. 当位置编码与前景掩码同时使用时, 教师网络和 MATS 分别达到 91.4% 与 95.5%, 取得了最优性能. 由此可见, 位置编码与前景掩码在跨模态异常检测中互为补充, 二者的结合能够在增强空间位置感知的同时有效排除背景干扰, 从而获得更高的检测精度.

表 4 3D+RGB 输入场景下位置编码和前景掩码消融实验 (%)

pos.enc	mask	Teacher	MATS
×	√	91.3	94.1
√	×	66.7	84.6
√	√	91.4	95.5

表 5 展示了 CNN 与 FastKAN 在计算效率与模型轻量化方面的消融实验结果. 通过对比 MATS 模型分别采用 CNN 与 FastKAN 模块的性能指标可知, 当使用 FastKAN 替换 CNN 后, 模型每秒处理的图像数量 (FPS) 从 4.8 f/s 提升至 6.2 f/s, 增幅达 29.2%; 计算复杂度 (FLOPs) 由 10.2G 降至 8.1G, 减少 20.6%; 模型参数量从 21.9 MB 压缩至 18.6 MB, 降低 15.1%. 实验结果表明, FastKAN 在显著提升推理速度的同时, 有效降低了计算资源消耗与模型存储需求, 验证了其在轻量化设计上的优势.

表5 CNN与FastKAN在MATS模型中的消融实验

CNN	FastKAN	FPS (f/s)	FLOPs (G)	Params (MB)
√	×	4.8	10.2	21.9
×	√	6.2	8.1	18.6

3.4 可视化

异常检测结果如图7所示, 每行分别展示了RGB图像、深度信息、对应的真实异常区域 (ground truth, GT) 以及基线模型 AST 和本文方法 (MATS) 输出的异常热力图。

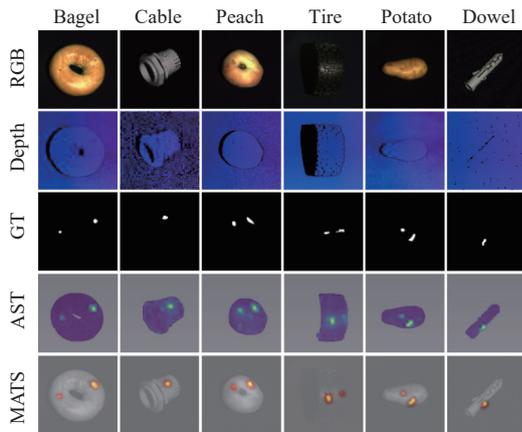


图7 检测结果可视化对比

可以观察到, MATS 相较于 AST 在多个场景中都展现出更高的检测准确度和更低的背景噪声。对于 Bagel 和 Tire 等目标, MATS 的热力图不仅与真实标注高度吻合, 而且有效抑制了背景和正常表面纹理的干扰; 而 AST 在检测大范围缺陷时易出现背景高响应, 干扰了对真实异常区域的判定。对于表面细微凹陷或较大缺损, MATS 也能在热图上呈现明显的高响应区域, 进一步说明模型在捕捉局部几何变形和纹理差异方面具备较高的敏感度。综上所述, MATS 与 AST 相比, 跨模态特征融合策略在视觉特征和几何信息之间建立了更有效的互补关系, 实现了对复杂工业场景中异常区域的精细检测与定位。

为验证所提出模型在复杂场景下的鲁棒性, 进一步选取了包含光照变化和局部遮挡条件的典型实例进行可视化分析, 如图8所示。以 Rope 类别为例, 其背景纹理结构复杂, 且目标呈线状分布, 易受图像对比度及边缘干扰影响, MATS 在此场景中仍然能够准确定位被遮挡的缺陷区域, 并有效抑制背景区域的干扰响应, 展现出良好的异常边界判别能力。对于 Carrot 类别, 由于表面存在显著的明暗不均与反射变化, 容易导致异

常特征在特征空间中表现模糊, 然而 MATS 依然在多个微小缺陷处生成高响应热区, 说明其在非理想光照条件下仍具有较强的细粒度感知能力与跨模态一致性建模能力。

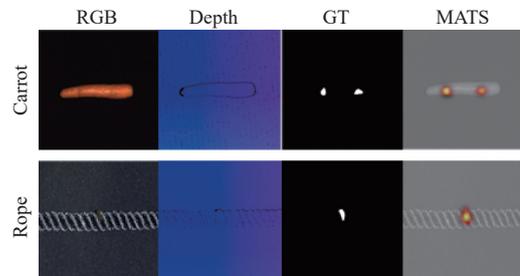


图8 检测结果可视化

图9展示了在 Potato 这个类别中, 异常区域和正常区域的师生分布距离。正常区域的特征较为一致, 且与模型的预测结果相符, 异常区域的距离分布明显偏离正常区域, 显示出较大且分散的距离值, 表明模型能够有效区分正常和异常样本, 而异常区域的特征和分布存在显著差异, 能够被模型识别并反映为较大的距离值。

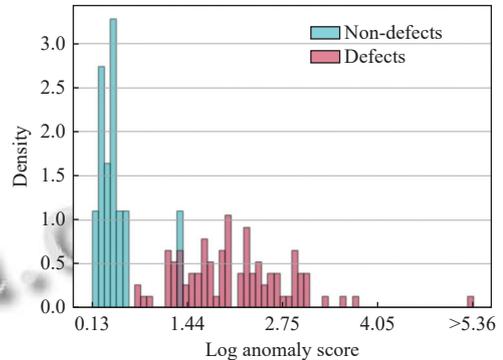


图9 异常分布直方图

4 结语

本文提出了一种基于多模态融合的非对称师生网络图像异常检测模型 (MATS), 通过融合 RGB 图像与深度信息的互补特征, 有效解决了传统方法中异常区域提取时的平滑现象与细粒度感知不足问题。模型通过跨模态异常放大器 (CAA) 实现特征扩展与压缩的动态融合, 显著增强异常区域的显著性表达并抑制背景噪声; 结合多扩张率局部注意力 (MDLA) 模块, 从多尺度视角提升局部异常特征的判别能力, 并通过归一化流 (NF) 构建正常样本的概率分布模型, 进一步强化检

测鲁棒性。此外, FastKAN 模块的轻量化设计也降低计算资源消耗。实验结果表明, MATS 在 MVTEC AD 与 MVTEC 3D-AD 数据集上的异常检测与定位性能优于现有方法, 验证了多模态融合策略与模块化设计的有效性。

参考文献

- 1 Liu JQ, Xie GY, Wang JB, *et al.* Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, 2024, 21(1): 104–135. [doi: [10.1007/s11633-023-1459-z](https://doi.org/10.1007/s11633-023-1459-z)]
- 2 Bergmann P, Fauser M, Sattlegger D, *et al.* MVTEC AD—A comprehensive real-world dataset for unsupervised anomaly detection. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 9584–9592.
- 3 Fernando T, Gammulle H, Denman S, *et al.* Deep learning for medical anomaly detection—A survey. *ACM Computing Surveys (CSUR)*, 2022, 54(7): 141.
- 4 Madan N, Ristea NC, Ionescu RT, *et al.* Self-supervised masked convolutional Transformer block for anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(1): 525–542. [doi: [10.1109/TPAMI.2023.3322604](https://doi.org/10.1109/TPAMI.2023.3322604)]
- 5 Akcay S, Atapour-Abarghouei A, Breckon TP. GANomaly: Semi-supervised anomaly detection via adversarial training. *Proceedings of the 14th Asian Conference on Computer Vision (ACCV 2018)*. Perth: Springer, 2019. 622–637.
- 6 Li ZY. Kolmogorov-Arnold networks are radial basis function networks. *arXiv:2405.06721*, 2024.
- 7 Rudolph M, Wehrbein T, Rosenhahn B, *et al.* Asymmetric student-teacher networks for industrial anomaly detection. *Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa: IEEE, 2023. 2591–2601.
- 8 Bergmann P, Jin X, Sattlegger D, *et al.* The MVTEC 3D-AD dataset for unsupervised 3D anomaly detection and localization. *arXiv:2112.09045*, 2021.
- 9 Yamada S, Hotta K. Reconstruction student with attention for student-teacher pyramid matching. *arXiv:2111.15376*, 2021.
- 10 Wang GD, Han SM, Ding ER, *et al.* Student-teacher feature pyramid matching for anomaly detection. *arXiv:2103.04257*, 2021.
- 11 Deng HQ, Li XY. Anomaly detection via reverse distillation from one-class embedding. *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE, 2022. 9727–9736.
- 12 Zavrtanik V, Kristan M, Skočaj D. DRÆM—A discriminatively trained reconstruction embedding for surface anomaly detection. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021. 8310–8319.
- 13 Mishra P, Verk R, Fornasier D, *et al.* VT-ADL: A vision Transformer network for image anomaly detection and localization. *Proceedings of the 30th IEEE International Symposium on Industrial Electronics (ISIE)*. Kyoto: IEEE, 2021. 1–6.
- 14 You ZY, Yang K, Luo WH, *et al.* ADTR: Anomaly detection Transformer with feature reconstruction. *Proceedings of the 29th International Conference on Neural Information Processing*. New Delhi: Springer, 2022. 298–310.
- 15 Lee S, Lee S, Song BC. CFA: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access*, 2022, 10: 78446–78454. [doi: [10.1109/ACCESS.2022.3193699](https://doi.org/10.1109/ACCESS.2022.3193699)]
- 16 Rusu RB, Blodow N, Beetz M. Fast point feature histograms (FPFH) for 3D registration. *Proceedings of the 2009 IEEE International Conference on Robotics and Automation*. Kobe: IEEE, 2009. 3212–3217.
- 17 Horwitz E, Hoshen Y. Back to the feature: Classical 3D features are (almost) all you need for 3D anomaly detection. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Vancouver: IEEE, 2023. 2968–2977.
- 18 Roth K, Pemula L, Zepeda J, *et al.* Towards total recall in industrial anomaly detection. *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE, 2022. 14298–14308.
- 19 Wyatt J, Leach A, Schmon SM, *et al.* AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise. *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. New Orleans: IEEE, 2022. 649–655.
- 20 Zheng Y, Wang X, Deng R, *et al.* Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization. *Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME)*. Taipei: IEEE, 2022. 1–6.
- 21 Ye F, Huang CQ, Cao JK, *et al.* Attribute restoration framework for anomaly detection. *IEEE Transactions on Multimedia*, 2022, 24: 116–127. [doi: [10.1109/TMM.2020](https://doi.org/10.1109/TMM.2020)]

- 3046884]
- 22 Rippel O, Mertens P, Merhof D. Modeling the distribution of normal data in pre-trained deep features for anomaly detection. Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan: IEEE, 2021. 6726–6733.
- 23 Rudolph M, Wandt B, Rosenhahn B. Same same but DifferNet: Semi-supervised defect detection with normalizing flows. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 1906–1915.
- 24 Gudovskiy D, Ishizaka S, Kozuka K. CFLOW-AD: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022. 1819–1828.
- 25 Rudolph M, Wehrbein T, Rosenhahn B, *et al.* Fully convolutional cross-scale-flows for image-based defect detection. Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022. 1829–1838.
- 26 Bergmann P, Fauser M, Sattlegger D, *et al.* Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 4182–4191.
- 27 Wang Y, Peng JL, Zhang JN, *et al.* Multimodal industrial anomaly detection via hybrid fusion. Proceedings of the 2023 the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 8032–8041.
- 28 Chu YM, Liu C, Hsieh TI, *et al.* Shape-guided dual-memory learning for 3D anomaly detection. Proceedings of the 40th International Conference on Machine Learning. Honolulu: PMLR, 2023. 6185–6194.
- 29 Costanzino A, Ramirez PZ, Lisanti G, *et al.* Multimodal industrial anomaly detection by crossmodal feature mapping. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 17234–17243.
- 30 Gu ZH, Zhang JN, Liu L, *et al.* Rethinking reverse distillation for multi-modal anomaly detection. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2024. 8445–8453.

(校对责编: 李慧鑫)