

基于混合融合记忆库的多模态 3D 点云异常检测 [17]

摘要

在现代工业生产中，产品质量检测是确保产品符合标准的关键环节，然而传统的工业异常检测方法主要依赖于单一模态的 2D 图像，这些方法没有考虑到产品的空间几何特性，在处理复杂形状和纹理的产品时存在局限性。随着 3D 传感器技术的发展，结合 3D 点云和 RGB 图像的多模态数据为工业异常检测提供了新的思路。本文介绍并复现了一种基于混合融合记忆库方案的多模态工业异常检测方法，名为 Multi-3D-Memory(M3DM)，该方法首先通过点特征对齐 (PFA) 模块将 RGB 图像和 3D 点云特征进行空间对齐；再通过无监督特征融合 (UFF) 模块使用对比学习增强模态间的特征交互，从而更好地捕捉产品的异常信息；最后通过决策层融合 (DLF) 模块利用多记忆库和单类支持向量机 (OCSVM) 来综合多模态信息，对异常进行检测和分割，从而提高异常检测的准确性和鲁棒性。实验结果表明，M3DM 在 MVTec 3D-AD 数据集上展现了强大的异常检测能力。

关键词：工业异常检测；混合融合；记忆库；多模态

1 引言

工业异常检测 (IAD) 旨在检测在制造过程中或者制造后产品的异常特征，以最小化潜在缺陷的风险。多模态工业异常检测在提升工业产品质量检查的精度和效率方面具有重要意义，尤其在面对日益复杂的工业生产环境和不断提高的质量要求时，传统的工业异常检测方法大多数依赖于 2D 的 RGB 图像，通过对大量正常样本的训练来进行异常检测 [11]。然而，在实际的工业产品质量检查中，人工检查不仅仅依赖于产品的颜色信息，还需要结合 3D 形状特征来判断产品是否合格。3D 形状信息，如表面积几何和深度信息，对检测的准确性至关重要，而单纯依靠 RGB 图像信息很难有效区分微小的表面缺陷和正常纹理，例如，某些具有细微表面缺陷的产品 (如 MVTec 3D-AD 数据集中的曲奇饼类数据) 仅通过 RGB 图像很可能无法准确判断其是否存在缺陷。

近年来，随着 3D 传感器技术的发展，结合 3D 点云和 RGB 图像的多模态数据 [2,4] 为工业异常检测提供了新的思路。3D 点云能够提供产品的空间几何信息，而 RGB 图像则提供了颜色和纹理信息，两者的结合能够更加全面地反映产品的特征，从而提高异常检测的准确性。现有的多模态工业异常检测方法 [9,15] 大多直接拼接多模态特征，很容易导致特征之间的强烈干扰，从而降低检测性能。因此，如何有效融合多模态特征，避免特征之间的干扰成为当前研究的一个重要方向。本文介绍的 M3DM [17] 方法通过无监督特征融合和决策层融合等技术，有效地解决了这一问题。

通过复现 M3DM 方法并对其性能进行深入分析,有助于更加深入地理解多模态数据融合的原理和方法,同时了解多模态数据在工业异常检测中的应用,为后续的研究和应用提供参考,同时该研究还可以推广到其他领域的多模态数据分析,如医学影像分析、自动驾驶等,具有广泛的应用前景和重要的研究价值。

2 相关工作

2.1 二维工业异常检测

早期的异常检测方法主要集中在单一模态的 2D 图像上,尤其是基于无监督学习的异常检测技术。在这些方法中,研究者们通常利用正常样本进行训练,然后通过推理阶段测试正常和异常样本的混合数据,其核心目标是提取能够有效表示正常样本的特征,从而在测试时实现对异常样本的区分。

常见的 2D 异常检测方法可以分为基于重建的方法和基于特征嵌入的方法 [11]。基于重建的方法 [16, 18, 19] 通常依赖于模型对正常样本图像的重建能力,例如自动编码器、生成对抗网络 (GAN)、修复模型或扩散模型。这些方法的基本思想是,模型经过训练后能够很好地还原正常样本,但在处理异常样本时,由于其与正常样本的特征分布不同,模型无法精确重建输入。因此,通过比较输入图像与重建图像的差异,可以生成像素级的异常图,并对异常区域进行定位。基于特征嵌入的方法 [1, 6, 8, 14] 则关注于深度神经网络定义的特征空间。这些方法通常利用预训练模型提取图像的特征表示,然后通过特征建模来检测异常。例如,基于记忆库的方法 [14] 过存储正常样本的特征分布,在推理阶段比较输入特征与记忆库中的特征以检测异常,这类方法由于其较强的灵活性,在 2D 图像异常检测中表现出了良好的性能。

尽管这些方法在 2D 异常检测中取得了较好的效果,但在工业检查中,特别是在面对微小缺陷或复杂纹理的情境时,2D 图像常常不足以准确识别异常。

2.2 三维工业异常检测

与传统的 2D 异常检测方法类似,目前的 3D 异常检测方法也主要分为基于特征嵌入和基于重建两类,如图 1 所示。基于特征嵌入的方法 [3, 5, 7, 10] 通过使用预训练编码器提取特征,并将这些特征映射到一个正态分布上进行学习。若输入的分布不在预定义范围内则会被判定为异常。目前,大多数 3D 异常检测方法采用记忆库机制,即在训练阶段存储一些具有代表性的特征以隐式构建特征分布,然后在测试阶段通过计算输入测试对象与记忆库中存储的所有模板点云之间的欧氏距离来判断是否存在异常。基于重建的方法 [21] 则通过训练一个能够准确重建正常点云的网络来实现检测。这类方法假设,由于异常点云在训练阶段未被学习,因此无法被有效地重建。异常检测通过比较输入点云与其重建结果之间的差异生成异常图。IMRNet [20] 使用 PointMAE [13] 对输入点云进行多次迭代重建,分别计算点云的显式空间坐标差异和隐式深度特征差异,以最终生成异常图。基于重建的方法对于单模态比较容易实现,但是对于多模态输入很难找到重建的目标,而直接使用预训练特征提取器的特征嵌入方法不需要设计多模态重构目标,是多模态任务的更好选择。

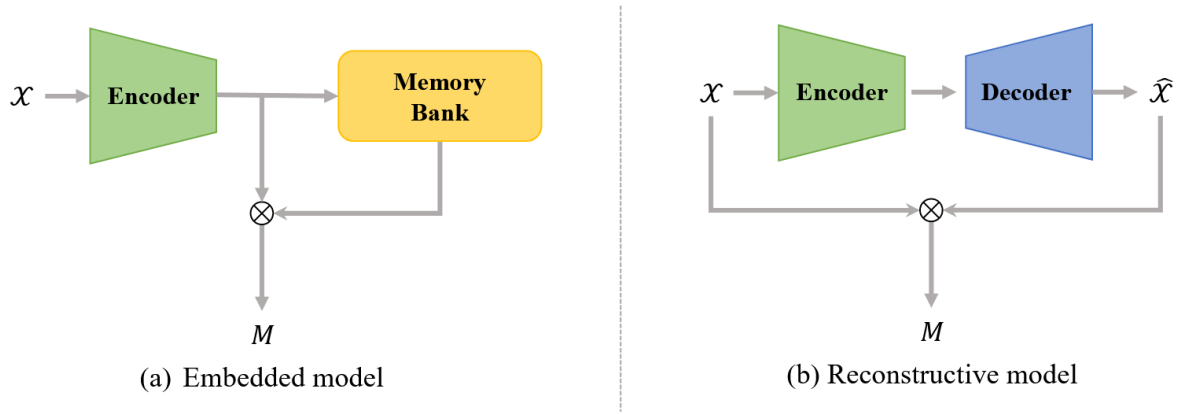


图 1. 基于特征嵌入 (左) 与基于重建 (右) 的异常检测框架

2.3 基于记忆库的异常检测

基于记忆库的方法通常不需要损失函数进行训练,可以快速构建模型,其性能由强大的预训练网络和额外的内存空间保障,这种方法目前已经在二维图像的异常检测中取得了最有效的成果 [11]。如图 2 所示,基于记忆库的异常检测方法只需要最小的网络训练,并且仅需要采样或映射已收集的正常图像的特征进行推理,在推理过程中,测试图像的特征会与记忆库中的特征进行比较,异常概率等于其与记忆库中正常特征的空间距离。PatchCore [14] 是工业异常检测领域的重大进展,其有两个特殊之处: 1) PatchCore 的记忆库通过核心子采样 (coreset-subsampling) 来降低推理成本,同时最大化性能; 2) PatchCore 通过测试样本与记忆库中最近邻特征及其他特征之间的距离来判断样本是否异常,大大增强了其鲁棒性。自 PatchCore 提出以来,许多改进方法都在其基础上发展,并逐渐从 2D 图像扩展到 3D 中,本文所使用的 M3DM 方法中所使用到的记忆库也是借鉴了 PatchCore 的方法。

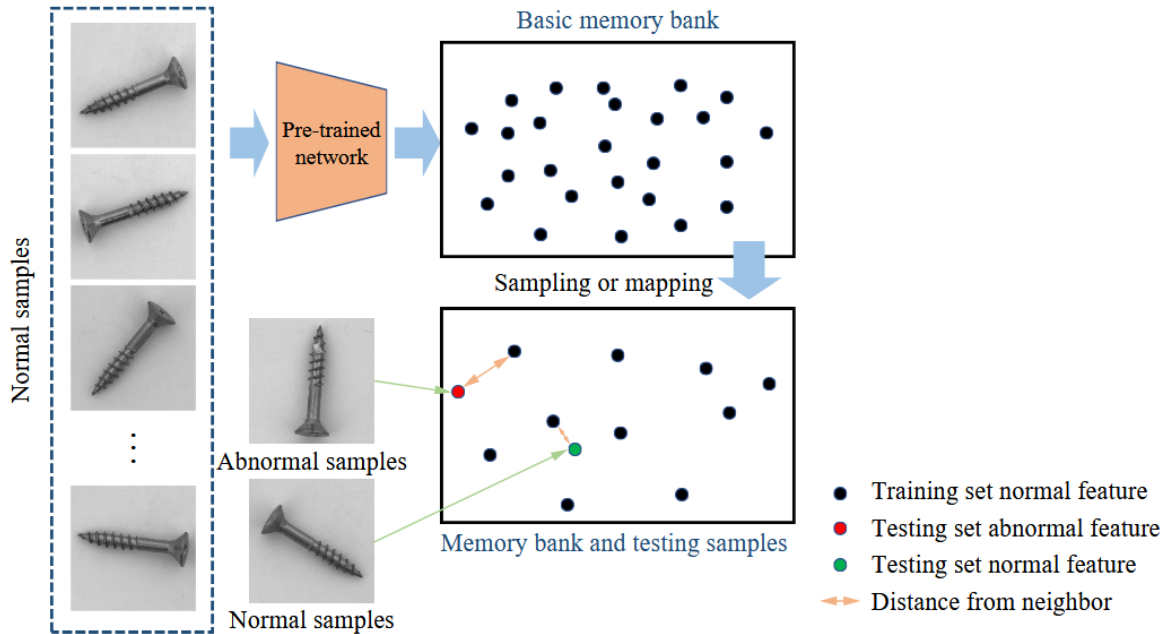


图 2. 基于 Memory Bank 的异常检测框架

3 本文方法

工业异常检测是工业生产中确保质量的重要任务，当前的方法大多基于单模态数据（如 RGB 图像），无法充分利用多模态数据（如 RGB 和 3D 点云）的互补信息。此外，简单地对多模态特征进行拼接可能导致模态间的干扰，降低检测性能。针对这些问题，M3DM [17] 设计了一种混合融合策略，通过对齐和融合 RGB 图像与 3D 点云的特征，提升多模态异常检测的准确性。

3.1 本文方法概述

M3DM 方法以三维点云和 RGB 图像作为输入进行三维异常检测和分割，提出混合融合以促进跨域信息交互，同时保持每个单独域的原始信息。使用两个预训练的特征提取器：DINO 用于提取 RGB 特征，PointMAE 用于提取点云特征。如图 3 所示，其整体结构主要包含三个部分：1) 点特征对齐模块 (PFA)：将 3D 点云特征投影到 2D 平面，与 RGB 图像特征对齐，从而实现模态间空间一致性。2) 无监督特征融合模块 (UFF)：通过基于对比学习的特征融合策略，增强 RGB 和 3D 点云特征的交互，同时减少模态间的干扰。3) 决策层特征融合模块 (DLF)：通过为 RGB、3D 点云和融合特征构建独立的记忆库，并结合单类支持向量机 (OCSVM)，实现异常检测和分割任务。尽管 UFF 可以提高检测性能，但是一些信息丢失是不可避免的，故进一步提出 DLF。

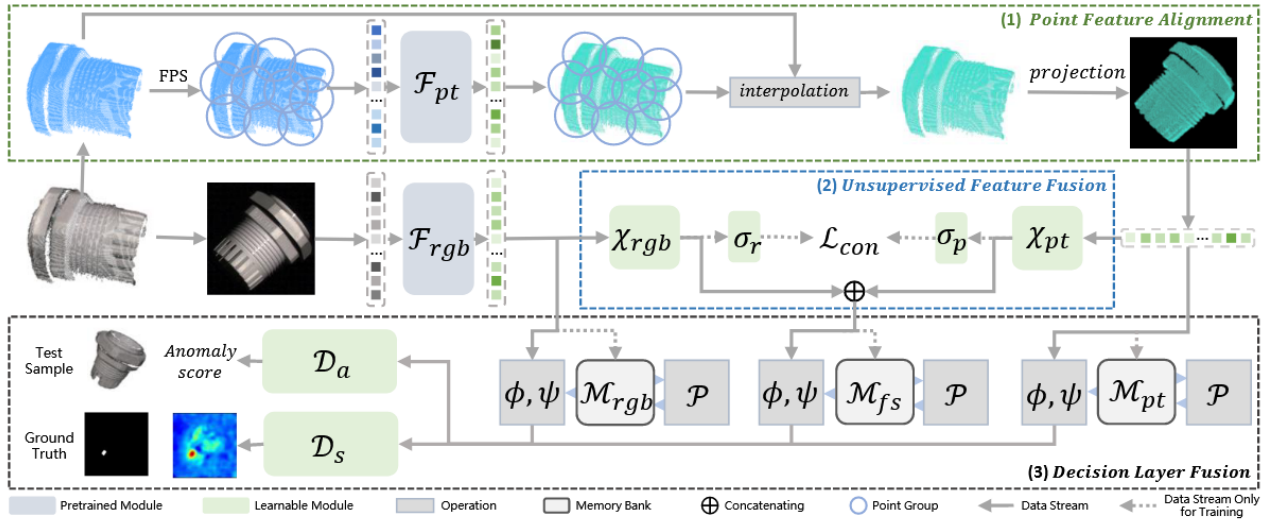


图 3. M3DM 整体框架 (来自原文)

3.2 点特征对齐模块

PFA 旨在实现 3D 点云特征与 RGB 图像特征的空间对齐，通过特征提取、插值和投影等操作，将稀疏的 3D 点云特征映射到二维平面，与 RGB 图像特征对齐，从而确保多模态数据在空间上的一致性。PFA 首先使用预训练的 Point Transformer 提取 3D 点云的局部和全局特征，Point Transformer 通过注意力机制捕获点云的几何信息和空间关系，其输出是多个点组的高维特征表示，这些点组由最远点采样 (FPS) 生成，FPS 是一种高效的点云降采样方法，能够选取覆盖整个点云空间的代表性点作为中心点，减少计算开销并保留点云的结构特

征。由于 3D 点云通常是稀疏且不规则分布的，仅利用点组的中心点特征可能导致信息不足。因此，PFA 模块引入了逆距离加权插值方法，将点组特征传播到点云中的所有点上，生成一个平滑的特征分布。插值过程中，特定点的特征通过其与周围中心点的距离加权计算，距离越近的中心点贡献越大。最后，为了实现点云特征与 RGB 图像特征的对齐，PFA 模块利用相机投影矩阵，将三维点云特征投影到二维平面上，即将插值后的点云特征通过相机内外参数映射到像素空间，从而生成一个与 RGB 图像分辨率一致的二维特征图。该二维特征图不仅保留了点云的几何和空间信息，还实现了与 RGB 图像在空间位置上的对齐。这种特征对齐操作作为后续的多模态特征融合奠定了基础，使得 RGB 图像和 3D 点云可以在同一特征空间中进行交互和联合分析。

3.3 无监督特征融合模块

UFF 旨在通过无监督学习的方法有效地融合 RGB 图像和 3D 点云特征，从而充分挖掘这两种模态间的互补信息。由于 RGB 和 3D 点云模态在表示形式和数据特性上的显著差异，直接拼接特征可能导致模态间信息的干扰，从而影响检测性能。UFF 模块通过基于位置的对比学习，确保两种模态在相同位置上的特征表示更加一致，同时减少不同位置间的相似性，从而提升多模态融合的效果。UFF 首先利用预训练的 Vision Transformer(ViT) 提取 RGB 图像的高维特征表示，并通过 Point Transformer 提取 3D 点云经过 PFA 模块对齐后的二维特征图，提取出的 RGB 和 3D 点云特征分别表示了物体的纹理和形状信息，是互补的关键来源。为了进一步增强两种模态的交互，UFF 模块引入了多层感知器 (MLP) 对每种模态的特征进行映射，将其变换为统一的表示空间，生成对应的查询 (query) 和键 (key) 向量，用于计算模态间的相关性。具体来说，RGB 特征通过 MLP 映射为查询向量，点云特征映射为键向量，从而建立起模态间的交互基础。然后使用对比学习的方式，通过 InfoNCE 损失对同一位置上的 RGB 和 3D 点云特征进行对比学习，如图 4 所示。目标是在同一位置上最大化两种模态特征的相似性，同时最小化不同位置特征间的相似性。InfoNCE 损失通过正样本和负样本的对比关系定义，确保模型能够捕捉到模态间的协同信息。对比损失函数如公式 1 所示，其中 $h_{\text{rgb}}^{(i,j)}$ 和 $h_{\text{pt}}^{(i,j)}$ 是 RGB 和点云特征在位置的查询向量和键向量。

$$L_{\text{con}} = -\log \frac{\exp(h_{\text{rgb}}^{(i,j)} \cdot h_{\text{pt}}^{(i,j)})}{\sum_{(k,l)} \exp(h_{\text{rgb}}^{(k,l)} \cdot h_{\text{pt}}^{(k,l)})} \quad (1)$$

在推理阶段，UFF 模块将对比学习后的 RGB 和 3D 点云特征进行拼接，生成多模态融合特征表示。处理后的 RGB 和点云特征通过拼接操作组合为一个高维特征向量，既保留了每种模态的独特信息，又通过对比学习强化了模态间的协同关系。这些融合特征为后续的异常检测和分割任务提供了更丰富的表示能力。

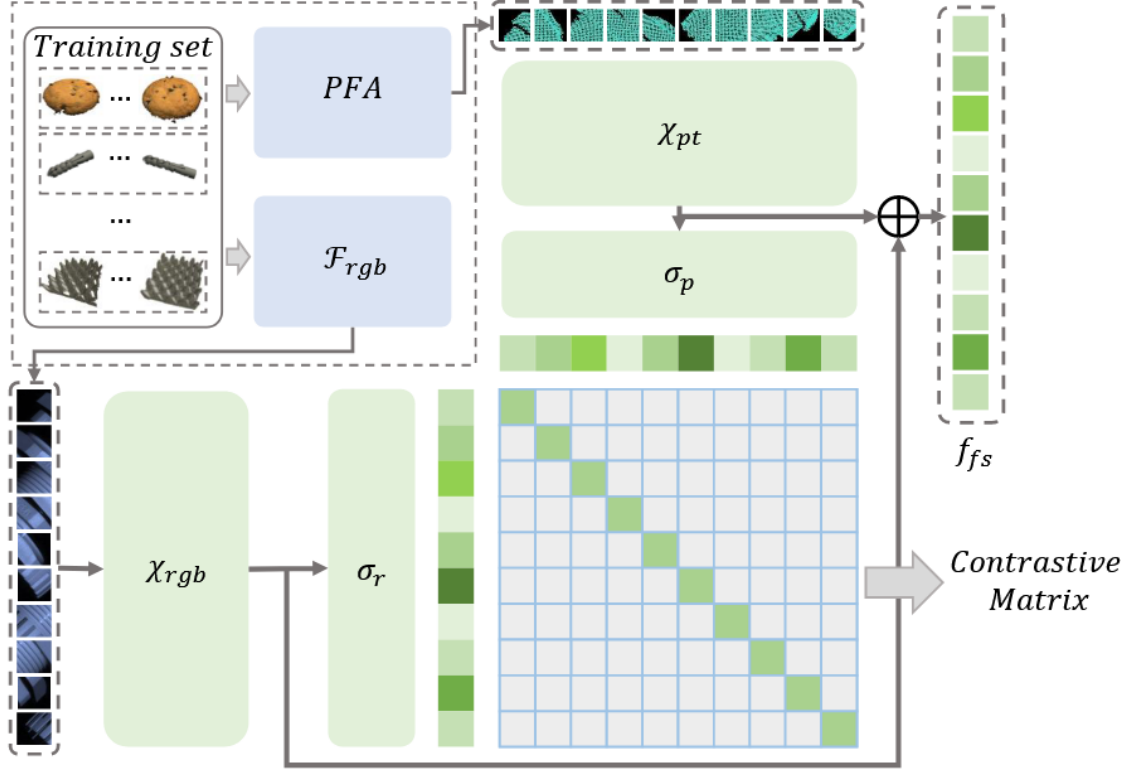


图 4. UFF 整体框架 (来自原文)

3.4 决策层特征融合

DLF 负责利用多模态特征的综合信息，实现稳健的异常检测和分割。DLF 模块首先为 RGB 特征、3D 点云特征以及融合特征分别建立三个记忆库 M_{rgb} 、 M_{pt} 和 M_{fs} ，这些记忆库以训练集中的正常样本特征为基础，存储了不同模态下正常数据的典型表示，用于在推理阶段衡量测试样本与正常分布的偏离程度。在与推理阶段，测试样本的每个点或像素特征会记忆库中的正常特征进行匹配，计算其异常得分。异常得分的计算基于特征与记忆库中最近邻特征之间的距离，如公式 2，其中 $f(p')$ 是测试样本中某点或像素的特征，是记忆库中的正常特征。这个距离衡量了测试特征与正常分布的相似性，距离越大，得分越高，表明该点或像素可能是异常的。

$$s(p') = \min_{m \in M} \|f(p') - m\|_2^2 \quad (2)$$

为了进一步优化检测性能，DLF 模块引入了单类支持向量机 (OCSVM) 对异常得分进行融合分析。OCSVM 利用训练数据学习到正常分布的边界，通过测试样本的偏离程度生成最终的异常检测结果。此外，DLF 模块通过对所有点或像素的异常得分生成异常分割图，结合设定的阈值将得分转化为二值分割结果，从而精确定位异常区域。

4 复现细节

4.1 与已有开源代码对比

论文作者已开源了其大部分功能的代码，但是对于多模态特征融合部分，其源码中只进行了简单的拼接操作，本实验在源代码的基础上对其特征融合部分进行了完善，在特征融合时使用训练好的 UFF 模型进行融合。开源代码来自 <https://github.com/nomewang/M3DM>。

4.2 数据集

MVTec 3D-AD [2] 数据集是专为无监督 3D 异常检测和定位任务而设计的首个综合性数据集。它由 MVTec Software GmbH 联合德国慕尼黑工业大学和卡尔斯鲁厄理工学院共同推出，旨在模拟现实世界中的视觉检查场景，例如制造业中对产品缺陷的检测。该数据集包含了 10 个不同物体类别的 4147 个 3D 点云扫描，3D 扫描数据通过结构光传感器采集，其中包含物体的 RGB 信息和 3D 位置信息，每个样本都提供一个彩色的 RGB 图像以及一个 3D 点云图像，便于 RGB 和 3D 点云的多模态分析。如图 5 所示，该数据集中的类别涵盖了从食品（如百吉饼和土豆）到工业部件（如电缆接头和轮胎）的多种对象。每个类别的数据集都提供了训练集和验证集，训练集仅包含正常样本，而测试集则包含了展示各种缺陷（如划痕、凹痕、孔洞、污染或变形）的样本。对于测试集中的每个异常样本，都提供了精确的像素级标注，这为评估异常检测算法的性能提供了重要依据。

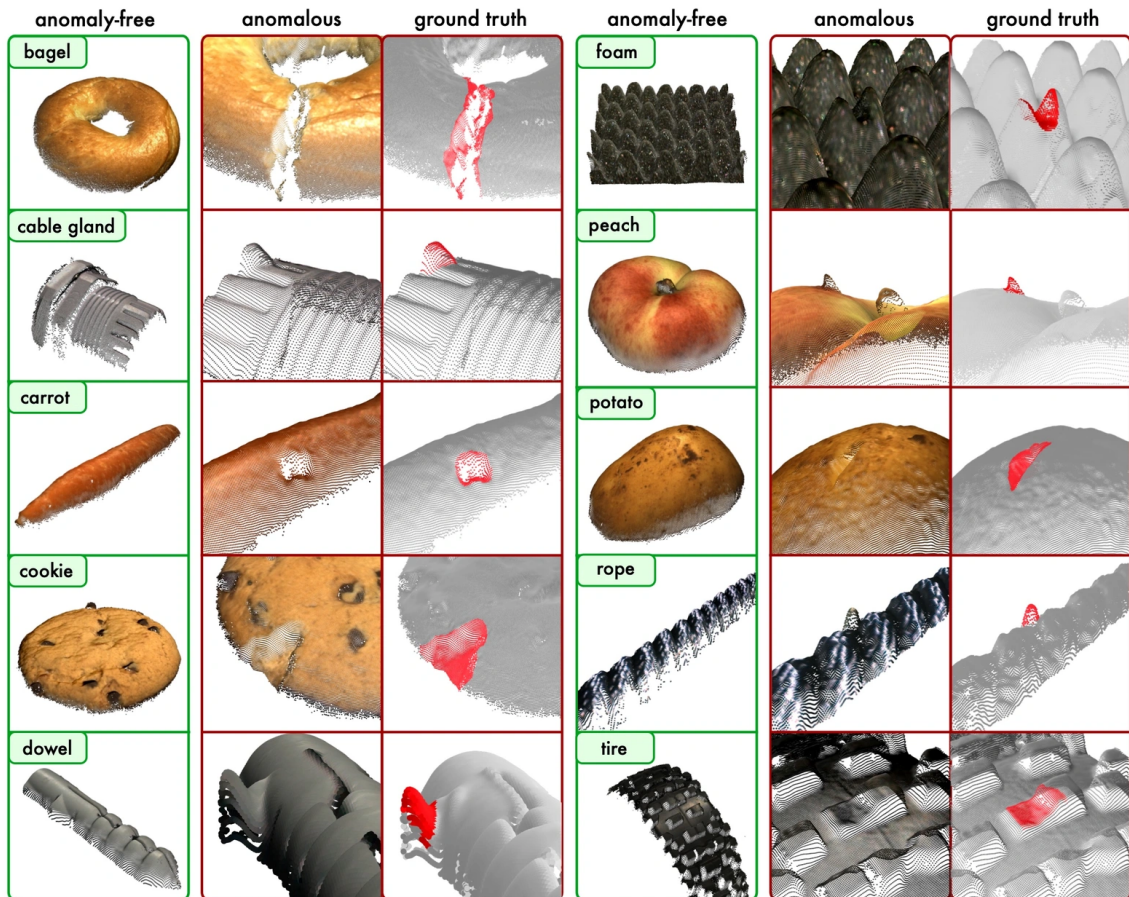


图 5. MVTec 3D-AD 数据集类别展示示例图

4.3 实验细节设置

4.3.1 实验环境搭建

实验环境为 Ubuntu16.04, 使用 GPU: NVIDIA Tesla P100 16G, 代码基于 Pytorch 框架, 使用 Python 编写, 使用 CUDA 加速 GPU 计算, TensorBoard 可视化训练过程中的各种指标, 帮助监控模型的训练状态。

4.3.2 数据预处理

数据预处理部分主要针对 RGB 图像和 3D 点云数据进行一系列操作, 以保证输入数据的质量和一致性。对于 3D 点云数据, 首先使用随机采样一致性 (RANSAC) 算法估计背景平面, 并移除与该平面距离在 0.005 以内的点, 同时, 在 RGB 图像中将这此移除点对应的像素值设置为零。接着将点云位置信息和 RGB 图像都调整为 224×224 的尺寸, 以适配后续特征提取器的输入大小, 该过程可以加速训练与推理的过程, 减少背景对异常检测的干扰。

4.3.3 特征提取

在实验中, 特征提取部分对 RGB 图像和 3D 点云数据分别进行处理, 以提取有意义的特征用于后续的异常检测任务。对于 RGB 图像, 使用 DINOv2 [12] 进行特征提取, 通过将图像切分为多个固定大小的图像块 (patch) 并将这些图像块作为输入送入自注意力机制以学习图像的全局上下文信息。该方法能够有效捕捉图像中的细节信息, 为每个图像块生成高维特征表示, 作为后续融合和异常检测的基础。

对于 3D 点云数据, 本文采用 Point MAE [13] 来提取点云的特征, 其通过自注意力机制和位置编码, 能够处理稀疏和不规则分布的点云数据, 并捕捉其空间结构和几何特征。在处理 3D 点云时, 首先对点云数据进行最远点采样 (FPS), 选取具有代表性的点作为输入特征, 接着 Point MAE 从这些点中提取局部和全局特征, 并通过注意力机制捕捉点与点之间的关系, 生成高维特征表示。这样, RGB 图像和 3D 点云的特征就能够分别被提取出来, 作为后续特征融合的输入。

4.3.4 可学习模块

M3DM 包括两个可学习模块: 无监督特征融合 (UFF) 模块和决策层融合 (DLF) 模块。UFF 模块使用多层感知机将 RGB 和 3D 点云特征分别映射到隐藏空间。优化器采用 AdamW, 学习率为 0.003, 并在 250 步内进行余弦退火, 最佳的异常检测结果是在 700 步训练后得到的。DLF 模块中的 OCSVM 使用随机梯度下降 (SGD) 优化器, 学习率为 1×10^{-4} , 每个类别训练 1000 步。

4.4 评价指标

本文在实验中使用了三个主要的评价指标来全面评估 M3DM 框架的性能: I-AUROC(图像级别接收者操作特性曲线下面积), 用于衡量模型在区分正常和异常图像的能力; AUPRO(像素级异常分割平均精度), 评估模型在异常区域分割精度方面的表现; 以及 P-AUROC(像素级 AUROC), 用于衡量模型在像素级别上定位异常区域的准确性。这些指标共同反映了 M3DM

在异常检测和分割任务中的综合能力，帮助验证了模型在不同层次（图像级和像素级）上的检测和分割精度。

5 实验结果分析

为了验证 M3DM 框架的有效性，实验设计了多个对比实验，包括仅使用 DINO 提取 RGB 特征、仅使用 Point_MAE 提取点云特征、仅使用 Fusion 模块提取融合特征、结合 DINO 和 Point_MAE 分别提取 RGB 和点云特征，以及完整地使用 DINO+Point_MAE+Fusion 模块来处理 RGB 图像、点云数据及其融合特征。这些实验旨在评估不同组件对异常检测性能的影响，并全面了解各个部分对整体模型贡献。从表1中可以看出，当结合 DINO 和 Point_MAE 并引入 Fusion 模块时，几乎所有类别的 AUPRO 得分都有所提升，尤其在 Bagel、Cable_Gland 等类别上表现尤为明显，说明多模态信息的结合有助于提高异常分割的准确性。如表2所示，对于图像级的 AUROC 而言，同样观察到结合 DINO、Point_MAE 及 Fusion 模块的方法显著提高了图像级别的异常检测性能，特别是在 Carrot、Cookie 等类别上，表明多模态融合能够增强模型对复杂异常情况的识别能力。如表3所示，即使在像素级（点级）定位方面，加入 Fusion 模块后，模型依然保持了较高的准确率，特别是针对复杂的几何形状或纹理细节，这证明了多模态融合对于精确定位异常的重要性。

表 1. AUPRO Result

Method	Bagel	Cable_Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
DINO	0.952	0.972	0.973	0.891	0.932	0.843	0.970	0.956	0.968	0.966	0.942
Point_MAE	0.946	0.818	0.977	0.871	0.871	0.784	0.959	0.976	0.946	0.945	0.909
Fusion	0.896	0.847	0.952	0.877	0.886	0.758	0.907	0.941	0.942	0.889	0.890
DINO+Point_MAE	0.967	0.968	0.978	0.934	0.931	0.924	0.976	0.967	0.970	0.971	0.959
DINO+Point_MAE+Fusion	0.970	0.974	0.976	0.931	0.932	0.926	0.972	0.981	0.975	0.971	0.961

表 2. I-AUROC Result

Method	Bagel	Cable_Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
DINO	0.944	0.918	0.896	0.749	0.959	0.767	0.919	0.648	0.936	0.767	0.850
Point_MAE	0.976	0.626	0.957	0.942	0.871	0.788	0.878	0.922	0.901	0.856	0.872
Fusion	0.851	0.747	0.899	0.815	0.885	0.721	0.806	0.791	0.891	0.755	0.816
DINO+Point_MAE	0.990	0.808	0.979	0.978	0.950	0.859	0.935	0.955	0.934	0.905	0.929
DINO+Point_MAE+Fusion	0.994	0.890	0.972	0.968	0.979	0.937	0.926	0.897	0.976	0.883	0.942

表 3. P-AUROC Result

Method	Bagel	Cable_Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
DINO	0.992	0.993	0.994	0.977	0.984	0.956	0.994	0.990	0.995	0.994	0.987
Point_MAE	0.982	0.947	0.996	0.933	0.960	0.946	0.988	0.994	0.994	0.983	0.972
Fusion	0.935	0.922	0.986	0.927	0.940	0.901	0.970	0.983	0.990	0.954	0.951
DINO+Point_MAE	0.993	0.992	0.996	0.978	0.981	0.982	0.996	0.994	0.996	0.995	0.990
DINO+Point_MAE+Fusion	0.995	0.991	0.996	0.917	0.982	0.985	0.991	0.990	0.994	0.982	0.982

图 6展示了 MVTEC 3D-AD 数据集上的异常检测和分割结果，通过对比不同方法在各个类别上的表现来评估模型的性能。图中以 Bagel 和 Rope 为例，其中第一列为原始图像，第二列为 Ground Truth，后五列分别为五种特征提取及记忆库存储决策方式的结果。可以看到，在一些情况下，单独使用 RGB 或点云信息可能不足以正确检测出所有的异常，而结合两者的信息则可以更全面地捕捉异常特征，从而实现更好的检测效果。

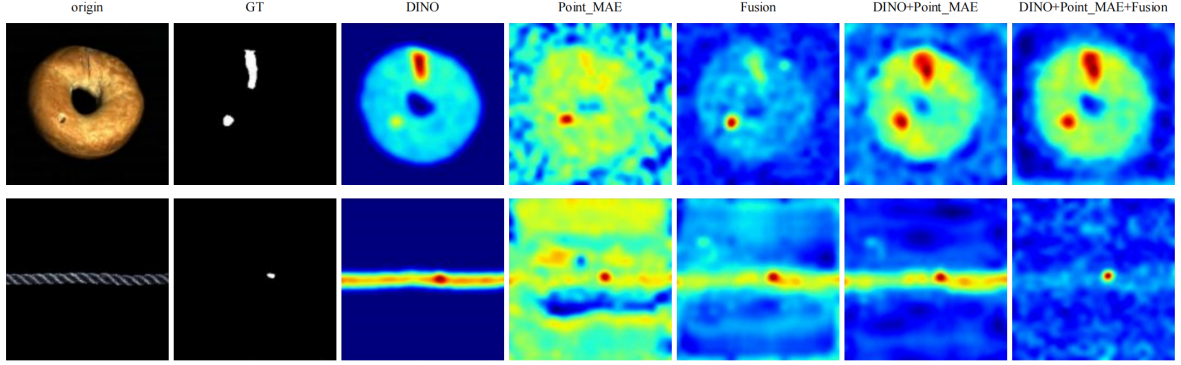


图 6. 实验结果示意

综上所述，该实验表明 M3DM 方法通过有效利用 RGB 图像和 3D 点云之间的互补信息，能够在工业异常检测任务中取得优于现有方法的表现。

6 总结与展望

本文复现了一种基于混合融合方案的多模态工业异常检测 Multi-3D-Memory (M3DM) 方法，通过对 M3DM 方法的深入分析和实验验证发现该方法在 MVTEC-3D AD 数据集上的检测和分割性能都比较优越。M3DM 通过无监督特征融合模块有效地促进了不同模态特征之间的交互，并利用决策层融合模块综合多模态信息，提高了异常检测的准确性和鲁棒性。此外，点特征对齐操作进一步增强了 3D 和 2D 特征的对齐效果，提升了检测性能。这些研究成果为工业异常检测领域提供了新的思路和方法，具有重要的实际应用价值。

尽管 M3DM 方法在多模态工业异常检测中取得了显著的成果，但仍有一些方面有待进一步研究和改进。首先，当前方法在处理大规模数据集时的计算效率仍有提升空间，未来可以探索更高效特征提取和融合算法，以适应更大规模的工业数据。其次，对于更复杂的工业场景，如动态变化的生产线和多类型产品的混合检测，现有方法的适应性有待提高，未来可以研究如何将 M3DM 方法扩展到更广泛的工业应用场景中，提高其泛化能力。此外，随着深度学习技术的不断发展，可以尝试将更多的先进技术和模型应用于多模态工业异常检测中，如注意力机制、生成对抗网络等，以进一步提升检测性能和鲁棒性。总之，多模态工业异常检测是一个具有广阔前景的研究领域，未来的研究将为工业生产质量控制提供更加强有力的技术支持。

参考文献

- [1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Pro-*

- ceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4183–4192, 2020.
- [2] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. *arXiv preprint arXiv:2112.09045*, 2021.
 - [3] Ankan Bhunia, Changjian Li, and Hakan Bilen. Looking 3d: Anomaly detection with 2d-3d alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17263–17272, 2024.
 - [4] Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the Asian Conference on Computer Vision*, pages 3586–3602, 2022.
 - [5] Yunkang Cao, Xiaohao Xu, and Weiming Shen. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognition*, 156:110761, 2024.
 - [6] Li-Ling Chiu and Shang-Hong Lai. Self-supervised normalizing flows for image anomaly detection and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2927–2936, 2023.
 - [7] Alex Costanzino, Pierluigi Zama Ramirez, Giuseppe Lisanti, and Luigi Di Stefano. Multi-modal industrial anomaly detection by crossmodal feature mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17234–17243, 2024.
 - [8] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9737–9746, 2022.
 - [9] Eliahu Horwitz and Yedid Hoshen. An empirical investigation of 3d anomaly detection and segmentation. *arXiv preprint arXiv:2203.05550*, 2(3):5, 2022.
 - [10] Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2968–2977, 2023.
 - [11] Jiaqi Liu, Guoyang Xie, Jinbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, 21(1):104–135, 2024.
 - [12] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

- [13] Yatian Pang, Wenxiao Wang, Francis EH Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning. In *European conference on computer vision*, pages 604–621. Springer, 2022.
- [14] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2022.
- [15] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2592–2602, 2023.
- [16] Xian Tao, Dapeng Zhang, Wenzhi Ma, Zhanxin Hou, ZhenFeng Lu, and Chandranath Adak. Unsupervised anomaly detection for surface defects with dual-siamese network. *IEEE Transactions on Industrial Informatics*, 18(11):7707–7717, 2022.
- [17] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*, pages 8032–8041, 2023.
- [18] Yi Yan, Deming Wang, Guangliang Zhou, and Qijun Chen. Unsupervised anomaly segmentation via multilevel image reconstruction and adaptive attention-level transition. *IEEE Transactions on Instrumentation and Measurement*, 70:1–12, 2021.
- [19] Jie Yang, Yong Shi, and Zhiquan Qi. Dfr: Deep feature reconstruction for unsupervised anomaly segmentation. *arXiv preprint arXiv:2012.07122*, 2020.
- [20] Xin Yang and Chunling Yang. Imrnet: An iterative motion compensation and residual reconstruction network for video compressed sensing. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2350–2354. IEEE, 2021.
- [21] Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3d-ad: Reconstruction via diffusion for 3d anomaly detection. In *European Conference on Computer Vision*, pages 91–107. Springer, 2025.