

# 面向多类异常检测与定位的提示引导Mamba网络

林 太

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2025年3月7日; 录用日期: 2025年3月31日; 发布日期: 2025年4月9日

## 摘 要

现有的无监督异常检测方法在多类训练时往往受到类别间干扰的困扰, 导致在实际应用中检测性能显著下降。为了解决这个问题, 本文提出了一种新颖的基于提示引导的Mamba网络(PGM)用于多类异常检测和定位。首先, 通过采用UniRepLKNet作为预训练特征提取器, 模型有效压缩并记忆正常图像的大量信息。接着, 基于Mamba结构的解码器结合了全局和局部建模能力能够有效重建多尺度特征。最后, PGM采用了分层类感知提示模块, 动态地将类别特定信息编码到类别先验池中, 以减轻异常类别之间的干扰。在MVTec AD数据集上进行的实验表明, 所提出的PGM方法与大多数先进方法相比具有最佳性能, 在图像级AUROC方面达到98.7%、像素级AUROC方面达到98.2%。同时, 详尽的消融研究验证了模型每个组件的贡献。总体而言, 所提出的PGM方法在异常检测领域具有显著优势。

## 关键词

多类异常检测, 知识蒸馏, Mamba

# Prompt-Guided Mamba Network for Multi-Class Anomaly Detection and Localization

Tai Lin

School of Optical-Electrical and Computer Engineering, Shanghai University of Science and Technology, Shanghai

Received: Mar. 7<sup>th</sup>, 2025; accepted: Mar. 31<sup>st</sup>, 2025; published: Apr. 9<sup>th</sup>, 2025

## Abstract

Existing unsupervised anomaly detection methods often suffer from significant performance

degradation in practical applications due to inter-class interference during multi-class training. To address this issue, this paper proposes a novel Prompt-Guided Mamba Network (PGM) for multi-class anomaly detection and localization. Firstly, by employing UniRepLKNet as the pre-trained feature extractor, the model effectively compresses and memorizes a large amount of information from normal images. Subsequently, the decoder based on the Mamba structure combines global and local modeling capabilities to effectively reconstruct multi-scale features. Finally, PGM incorporates a hierarchical class-aware prompt module that dynamically encodes class-specific information into a class prior pool to mitigate interference between different anomaly classes. Experiments conducted on the MVTec AD dataset demonstrate that the proposed PGM method outperforms most state-of-the-art methods, achieving 98.7% in image-level AUROC and 98.2% in pixel-level AUROC. Additionally, extensive ablation studies validate the contribution of each component of the model. Overall, the proposed PGM method shows significant advantages in the field of anomaly detection.

## Keywords

Multi-Class Anomaly Detection, Knowledge Distillation, Mamba

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

异常检测(AD)作为工业缺陷检测[1]、医学影像分析[2]及安防监控[3]等领域的核心技术,致力于识别偏离正常模式的数据实例。智能制造的发展使工业视觉异常检测在生产流程中发挥关键作用,其通过替代人工质检显著提升效率、降低成本,同时保障产品质量与产线稳定性。传统单类别检测范式需为每个类别独立训练模型,存在训练资源消耗大、内存占用高等问题,难以适应大规模工业场景需求。近期研究重心已转向多类统一异常检测框架,通过共享模型参数实现跨类别检测,显著提升方法的可扩展性与实用价值。

无监督学习范式凭借其无需异常标注数据的特性,在 AD 领域占据主导地位。这类方法通过建模正常样本分布特征构建检测基准,近年来无监督异常检测方法[4]-[8]在主流工业数据集[1] [9]上展现出卓越性能,并逐步拓展至复杂应用场景。然而现有多类统一 AD 方法仍面临精度与效率的平衡挑战,尤其在部署中易受类别间特征干扰影响,导致性能显著衰减。其核心矛盾体现在两方面:网络有限的表征能力可能无法准确重建正常区域,却对异常区域保持过度泛化;同时现有方法在共享框架中处理多类别时,易造成语义空间纠缠,加剧灾难性遗忘问题——新类别引入时会破坏已学特征,且难以建立清晰的类别边界。

为了解决上述问题,本文提出了一种新的异常检测框架 PGM。首先,采用集成 GRN 单元的 UniRepLKNet [10]作为预训练主干网络,其大核卷积结构通过增强特征多样性突破网络容量瓶颈,有效提取深层次抽象模式。其次,基于 Mamba 架构的解码器结合并行多核卷积强化局部特征关联,在保持卓越建模能力的同时将线性复杂度降低,显著减少参数规模与计算开销。最后,本文引入了分层类感知提示模块,通过注入类别特异性提示信息构建先验知识池,引导模型解耦不同类别语义空间。这些提示提供了精确的上下文信息,推动不同类别的边界远离。通过集成类别特定的提示,本文的框架在最小化新类别干扰的同时,保留了先前学到的类别特征,从而提高了整体检测性能。

因此,基于以上讨论,本文主要贡献总结如下:

1) 采用了 UniRepLKNet 作为预训练骨干网络,通过大核卷积网络捕获多层次抽象特征,显著提升网络表征能力,为多类训练提供强判别性特征。

2) 基于 Mamba 架构的解码器,通过混合状态空间块与并行多核卷积操作实现全局信息捕获与局部建模的有机融合,同时显著降低了线性复杂度、参数量和计算需求。

3) 引入分层类感知提示模块,通过动态注入类别先验信息实现多类别语义空间解耦,有效减少类别干扰,提升模型检测性能。

## 2. 相关研究

基于嵌入的方法通常依赖于预训练网络,从在自然图像上训练的特征提取器中提取特征嵌入向量,将正常特征映射到一个紧凑的空间,通过区分正常和异常数据来识别异常。这些方法在单类别设置下表现出竞争力,但随着类别数量的增加,它们面临可扩展性问题。HGAD [11]通过避免将不同的类别分布映射到同一个高斯先验中,给归一化流的潜在空间带来更强的表示能力。MAAE [12]为了减轻由于不同类别的多样化分布模式导致的性能下降,它采用空间注意力和通道注意力来有效地捕获全局类别信息,并模拟多类的特征分布。

基于重构的方法通常依赖于自训练的编码器-解码器框架来重建输入图像。它们通过训练模型重构正常图像,依托于这样的假设:由于模型对正常图像已充分学习,异常样本在重构时会呈现出较差的质量。随着类别数量的增加,这将导致了时间和内存消耗的增加。同时,由于多类别训练中正常模式的多样性增加,异常定位不可避免地受到过度泛化的影响而降低。RKD [13]使用学生解码器从单类嵌入中恢复预训练教师编码器的多尺度表示。RLR [14]利用可学习的参考表示来强制模型显式学习正常特征模式,从而防止模型陷入“学习捷径”问题。UniAD [15]首先为异常检测构建了统一的重建框架。它利用向量量化框架以防止学习“相同的快捷方式”。为了使重建模型在保持高效率的同时增强其对统一异常检测的泛化能力,Gao [16]提出了一个简单而有效的异常检测框架,用一张正常图像提示来重建正常特征并恢复异常特征。DiAD [17]利用语义引导网络和空间感知特征融合模块来解决稳定扩散模型在多类异常检测中的类别和语义损失问题。

## 3. 方法

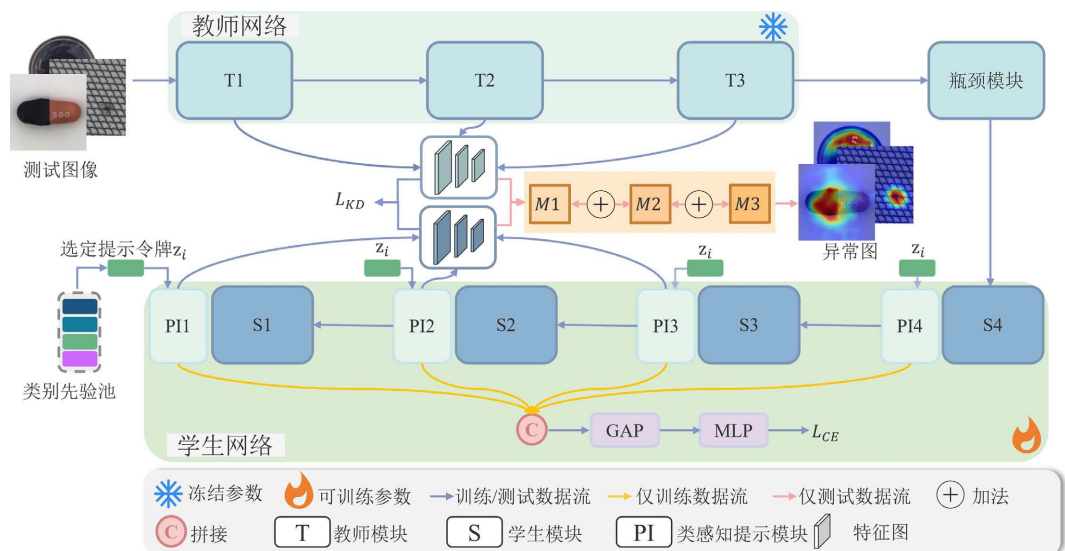
### 3.1. 模型框架

如图 1 所示,所提出的 PGM 模型基于反向蒸馏框架[13],包括基于 UniRepLKNet 的预训练编码器作为教师模块 T、瓶颈模块,基于 Mamba 的解码器作为学生模块 S,分层类感知提示模块 PI,类别先验池。教师编码器 T 从输入图像中提取多尺度空间特征,然后瓶颈模块将这些特征压缩以最小化冗余。Mamba 解码器重建编码器的输出以实现特征级别的对齐。PGM 模型仅使用正常样本进行训练,而在推理期间遇到异常样本时,解码器无法准确重建编码器的输出,从而生成异常图。训练目标是 minimized 正常图像每层对应输出特征之间的局部余弦距离损失。

### 3.2. 基于 UniRepLKNet 的预训练编码器

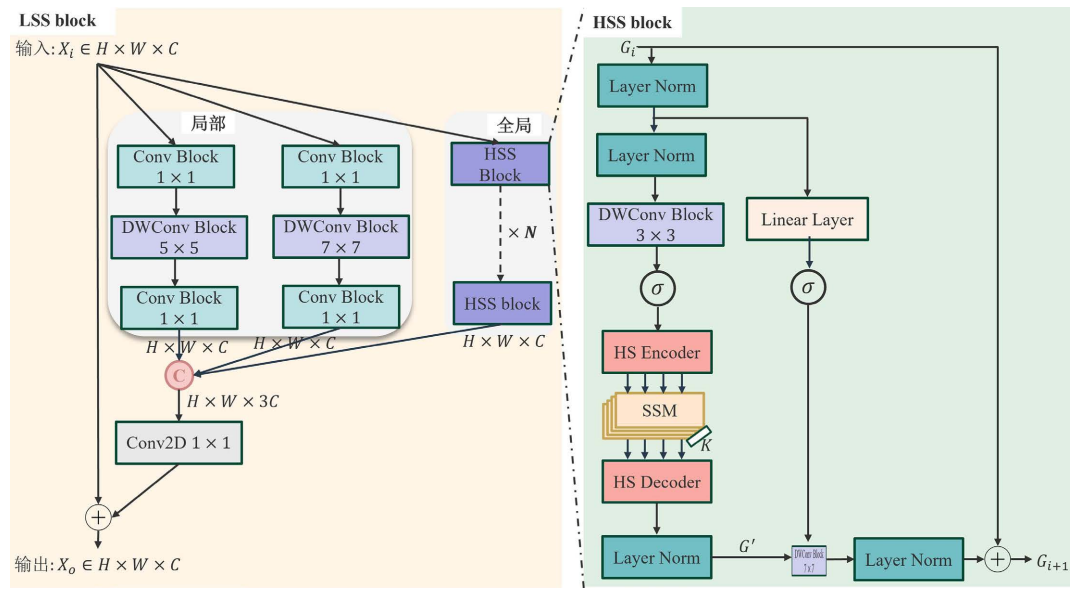
受 MinMaxAD [18]的启发,本文引入 UniRepLKNet 作为预训练编码器。该网络深度融合全局响应归一化(GRN)单元[19],通过建立特征通道间的动态抑制机制,显著提升特征表征多样性,突破传统卷积网络的容量瓶颈。其核心创新在于大核卷积架构设计——利用超宽感受野实现多尺度语义特征融合,完成高保真信息压缩与结构化编码。网络采用多级特征金字塔架构,通过堆叠下采样模块与异质卷积组合(LarK 大核模块/SmaK 小核模块)实现协同优化空间模式层次和表示能力。其核心是通过大核实现广域感

知，小核深化局部模式。块内的残差连接架构增强了空间特征聚合与通道间信息交互，在确保梯度流稳定性的同时，构建了高判别性特征空间。



**Figure 1.** An overview of the proposed PGM model  
**图 1.** 提出的 PGM 模型概览

**3.3. Mamba 解码器**



**Figure 2.** Schematic diagram of the HSS Module and LSS Module  
**图 2.** HSS 模块与 LSS 模块的结构示意图

受 MambAD [20]启发，本文使用了基于 Mamba 结构的解码器，通过融合状态空间模型(SSM)的长程建模优势与 CNN 的局部感知特性，实现多类无监督异常检测的高效协同优化。该架构将 SSM 的全局建模能力与 CNN 的细粒度特征提取深度耦合，形成兼具线性计算复杂度与多尺度感知优势的解码范式。

如图 2 所示, Mamba 解码器由局部增强状态空间模块(LSS)与混合状态空间模块(HSS)构成核心计算单元。LSS 模块采用  $n \in \{2, 3\}$  个 HSS 模块堆叠的配置方案, 对输入特征  $X_i$  多分支处理: 全局分支通过级联 HSS 模块捕获跨区域语义关联, 局部分支采用双路径深度可分离卷积架构(含  $1 \times 1$  卷积块、 $k \times k$  深度卷积块及另一个  $1 \times 1$  卷积块)提取细粒度特征。特征融合阶段通过通道拼接与  $1 \times 1$  卷积实现维度对齐以匹配输入进行残差连接。HSS 模块创新设计混合扫描编解码机制, 采用 Hilbert 扫描策略建立全局特征关联——编码阶段通过多向扫描路径生成方向敏感特征序列, 经 SSM 建模后由解码器重构至原始空间布局。该机制有效缓解长程依赖建模困境, 在保持线性计算复杂度的同时实现多尺度特征建模能力的突破。

### 3.4. 分层类感知提示模块

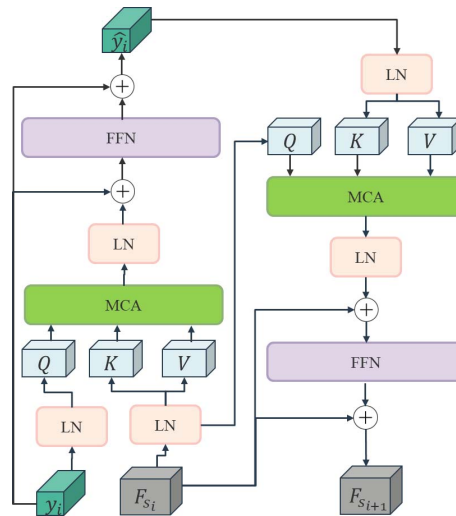


Figure 3. Schematic diagram of the Hierarchical Class-Aware Prompt Module

图 3. 分层类感知提示模块的结构示意图

面向多类别异常检测中普遍存在类别间干扰, 受 ROADS [21]启发, 本文引入分层类感知提示模块, 如图 3 所示。该模块能够动态学习并直接从多样化的正常类别中提取类别特定的提示令牌, 并将其整合到我们的类别先验池中。该模块通过建立提示向量集成机制, 从不同正常样本中提取类别特异性特征并注入类别先验池中。通过将类别提示向量作为约束条件, 引导解码器执行类条件特征重构, 该策略有效解耦跨类别语义纠缠, 实现多类异常检测精度与泛化能力的协同提升。

在我们的数据集中, 每个类别的特定先验信息被编码为可学习的提示令牌。我们使用 Xavier 初始化 [22]初始化一个类别先验池  $Y \in \mathbb{R}^{N \times l \times M_l}$ , 其中  $l$  表示令牌长度,  $M_l$  是每个令牌的维度,  $N$  表示类别数量。数据集中的每个类别都与多个提示令牌相关联, 这些提示令牌封装了特定于该类别的先验信息。当处理具有特定类别  $i$  的图像  $x$  时, 通过从类别先验池  $Y$  中选择相应的类别特定提示令牌  $y_i \in \mathbb{R}^{l \times M_l}$ 。然后, 学生解码器  $S$  使用该令牌从中间表示中重构特征。这种设计使解码器能够自适应地选择并整合类别特定的先验信息, 通过参数化类别先验池实现跨类别知识隔离, 有效减轻了类别间相互耦合的影响。

类别集成机制通过交叉注意力将这些提示令牌中封装的知识与学生解码器的特征表示进行融合。通过将提示令牌视为查询向量, 并将特征表示视为键向量和值向量, 该模块促进了类别特定提示令牌与学生解码器特征表示之间的高效全局信息交换, 同时保持了计算效率。给定重构后的中间表示  $F_{S_i} \in \mathbb{R}^{n \times M_l}$  (其中  $n = H \times W$ ) 和提示令牌  $y_i \in \mathbb{R}^{l \times M_l}$ , 该机制首先从学生解码器中聚合图像特定信息以生成后验令牌  $\hat{y}_i$ 。

$$\hat{y}_i = \text{FFN}\left(\text{LN}\left(\text{MCA}\left(\text{LN}(z_i), \text{LN}(F_{S_i})\right)\right) + y_i\right) + y_i \quad (1)$$

其中, FFN 表示前馈神经网络, MCA 代表多头交叉注意力机制, LN 表示层归一化。公式(1)中的 MCA 模块将提示符  $y_i$  作为查询向量  $Q$ , 并将特征表示  $F_{S_i}$  作为键向量  $K$  和值向量  $V$ 。MCA 模块按如下方式计算注意力机制:

$$\text{MCA}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

其中,  $Q = y_i W_Q, K = F_{S_i} W_K, V = F_{S_i} W_V$ , 而  $d_k$  是键向量(Key vectors)的维度。多个注意力头的输出随后被拼接并投影, 形成最终的交叉注意力输出。接着, 这些后验令牌被重新整合到  $F_{S_i}$  中, 生成一个提示增强的特征图  $F_{S_{i+1}}$ , 具体如下:

$$F_{S_{i+1}} = \text{FFN}\left(\text{LN}\left(\text{MCA}\left(\text{LN}(F_{S_i}), \text{LN}(\hat{y}_i)\right) + F_{S_i}\right)\right) + F_{S_i} \quad (3)$$

其中,  $F_{S_i}$  作为查询向量, 而后验提示令牌  $\hat{y}_i$ ,  $i$  作为键向量和值向量。来自所有尺度的后验令牌  $\{y_i\}_{i=1}^M$ , 以及由类别先验池生成的后验特征向量  $\hat{y}_i$ , 被拼接成  $\hat{y}_{cat}$ 。该张量依次通过全局平均池化层(GAP)与线性层完成异常概率预测:  $\hat{y}_i = \text{Linear}(\text{GAP}(\hat{y}_{cat}))$ 。模型随后使用交叉熵损失进行参数更新, 定义为  $L_{CE} = -\sum_{i=1}^N \varphi_i \log(\hat{\varphi}_i)$ , 其中  $\varphi_i$  是真实标签,  $\hat{\varphi}_i$  是类别  $i$  的预测概率。我们的分层、上下文感知方法促进了在多尺度上学习和应用类别特定的提示令牌, 显著提高了跨多样类别的异常检测能力。这种分层上下文感知机制与多尺度特征知识蒸馏相结合, 实现多类别异常特征的准确辨识。

### 3.5. 训练和推理

本框架中以建立教师-学生网络正样本特征一致性为核心优化目标, 确保通过特征对齐实现精确的异常检测。为了实现这一目标, 我们采用了逐向量余弦相似度损失  $L_{KD}$ , 该损失量化了教师编码器  $T$  和学生解码器  $S$  生成的特征图在通道维度上的对齐程度。损失函数定义如下:

$$L_{KD} = \sum_{i=1}^M 1 - \left( \frac{F_{T_i}^T \cdot F_{S_i}}{\|F_{T_i}\|_2 \cdot \|F_{S_i}\|_2} \right) \quad (4)$$

其中,  $F_{T_i}$  和  $F_{S_i}$  分别表示教师模型和学生模型在第  $i$  层的特征图。符号  $\|\cdot\|_2$  表示  $L_2$  范数, 用于对特征向量进行归一化。该损失函数通过逐层累加  $M$  个层级特征差异, 驱动学生网络多尺度特征空间与教师网络全面知识传递。

最后, 我们的训练优化中的整体目标函数整合了多个损失函数, 以实现准确的异常检测。总损失函数  $L_{total}$  由知识蒸馏损失  $L_{KD}$ 、用于训练提示令牌交叉熵损失  $L_{CE}$ 。整体损失函数定义如下:

$$L_{total} = \alpha L_{KD} + \beta L_{CE} \quad (5)$$

其中,  $\alpha$  和  $\beta$  是用于平衡不同损失函数贡献的权重超参数。在训练过程中, 本文框架利用正常样本驱动学生解码器精确重构教师网络的多尺度特征表征。二者的协同优化确保模型在多类别场景下实现判别性特征学习与稳定泛化能力。

## 4. 实验结果与分析

### 4.1. 数据集介绍与实验配置

MVTec AD 数据集是一个专注于工业异常检测的公开基准数据集, 包含 15 类工业产品(如电子元件、

纹理材料)的高分辨率图像,其中每个类别由仅含正常样本的训练样本和包含各种类型缺陷的异常样本构成,广泛应用于工业质检算法研究与性能评估。

**实验配置:** 本文的研究工作是在使用 PyTorch 和 NVIDIA GeForce GTX 3070 和 Intel(R)Core(TM)i7-10700F CPU@2.90 GHz 上实施的。我们将 MVTEC 图像的进行归一化,并调整大小为  $256 \times 256$ ,用于训练和测试。教师网络是在 ImageNet 上预训练的网络 UniRepLKNet,在训练过程中被冻结。学生网络是一个随机初始化的网络,由 AdamW 算法优化。

**评估指标:** 本实验采用了图像级和像素级的 AUROC [1]、区域重叠下的面积(AUPRO) [23]、平均精度[24] (AP)和 F1\_max [25]作为评价指标。图像级 AUROC 用于评估模型在整体图像中检测异常的能力。像素级 AUROC 则用于评估模型在精确定位图像中异常区域的效果。AUPRO 通过计算模型预测的异常区域与真实异常区域之间的重叠程度来评估模型性能。AP 是精确率-召回率曲线下的面积,用于衡量模型在不同召回率下的平均精确率。F1\_max 是精确率和召回率的调和平均数在不同阈值下的最大值,用于衡量模型在最优阈值下的综合性能。

## 4.2. 消融实验结果与分析

### 4.2.1. 架构消融

PGM 的三个重要组成部分是基于 UniRepLKNet 的预训练特征提取器、Mamba 解码器和分层类感知提示模块。本实验探索了这三个改进各自的贡献。实验结果如表 1 所示,与传统的 ResNet 作为编码器不同,UniRepLKNet 通过大核卷积的全局建模、GRN 增强的特征多样性、结构重参数化的高效设计,为模型提供了高容量、轻量化的特征提取基础。此外,Mamba 解码器将局部和全局信息融合,增强了特征序列的建模能力。这在一定程度上增强了模型对正常特征的泛化,阻碍了异常模式的恢复。最后引入的分层感知提示模块极大提高了模型的定位性能。这是因为在 multi-class 异常检测中,模型不仅需要区分正常与异常样本,还需要能够区分多种正常图像类别,这要求模型能够同时处理类别间的差异与相似性,并有效地减少类间干扰。它们共同发挥作用,将模型的性能提升到先进的水平,检测精度的 AUROC 达到了 98.7%,定位精度的 AUROC 达到了 98.2%。

**Table 1.** The impact of different architectures on model accuracy

**表 1.** 不同架构对模型精度的影响

| 基准 | UniRepLKNet | Mamba 解码器 | 分层类感知提示模块 | 检测精度 | 定位精度 |
|----|-------------|-----------|-----------|------|------|
| ✓  |             |           |           | 95.3 | 96.5 |
| ✓  | ✓           |           |           | 97.4 | 97.0 |
| ✓  |             | ✓         |           | 97.6 | 97.2 |
| ✓  |             |           | ✓         | 97.1 | 96.7 |
| ✓  | ✓           | ✓         |           | 98.4 | 97.8 |
| ✓  | ✓           |           | ✓         | 98.2 | 97.4 |
| ✓  |             | ✓         | ✓         | 98.5 | 97.7 |
| ✓  | ✓           | ✓         | ✓         | 98.7 | 98.2 |

### 4.2.2. 损失消融

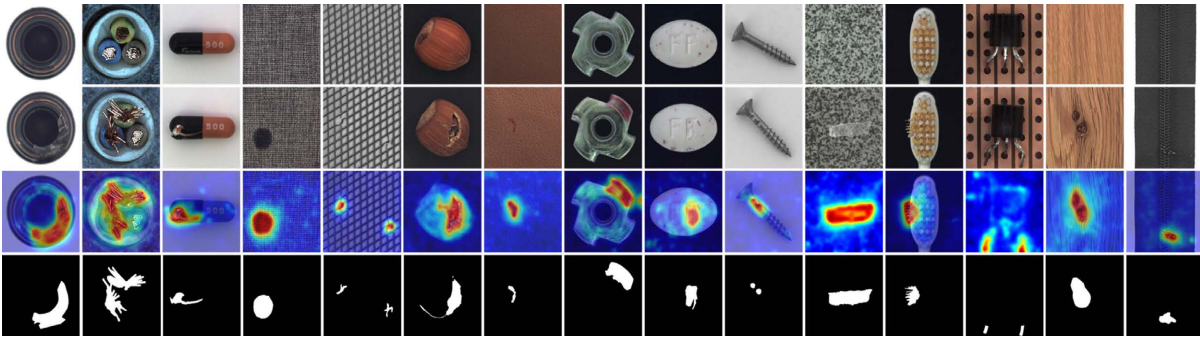
本节探索了交叉熵损失和两种衡量特征差异的损失对性能的影响。结果如表 2 所示,在仅使用 L2 损失或 COSINE 损失时,模型的检测精度和定位精度表现不佳。因为在 multi-class 异常检测中,不同异常类别之

间在特征空间中存在重叠或相似性，从而使得模型难以有效地区分各个类别的异常。而交叉熵损失能够优化不同类别的提示令牌，显式分离语义空间，避免特征重建时的跨类别混淆。另一方面，基准实验同时使用 L2 损失和交叉熵损失时，模型在异常检测的 AUROC 为 98.5%，定位精度的 AUROC 为 97.9%。当使用 COSINE 损失取代 L2 损失时，模型的检测精度提升至 98.7%，定位精度提升至 98.2%。可以看出，COSINE 在 multi-class 异常检测中展现出更强的判别能力。因为 L2 损失仅关注数值差异，而 COSINE 损失更关注特征之间的方向差异。在高维空间中，方向差异更能体现语义上的区别，而 L2 损失可能因特征维度过多而丢失关键信息。

**Table 2.** The effect of different loss functions on model performance  
**表 2.** 不同的损失对模型性能的影响

| 基准 | L2 损失 | COSINE 损失 | 交叉熵损失 | 检测精度 | 定位精度 |
|----|-------|-----------|-------|------|------|
| √  | √     |           |       | 98.1 | 97.3 |
| √  |       | √         |       | 98.3 | 97.5 |
| √  | √     |           | √     | 98.5 | 97.9 |
| √  |       | √         | √     | 98.7 | 98.2 |

4.2.3. 模型对比实验与分析



**Figure 4.** Visualization of detection results on the MVTec AD dataset  
**图 4.** 在 MVTec AD 数据集上的可视化检测结果

如表 3 所示，所提方法在模型检测精度与定位精度方面与当前先进模型进行了比较，并取得了显著的表现。在检测精度方面，该模型的 AUROC 达到了 98.7%，相较于 transformer 模型如 UniAD 高出了 2.2%，并且比扩散模型如 DiAD 高出 1.5%。一方面，得益于 UniRepLKNet 预训练模型有效提高了网络容量的上限，使得更加丰富的特征多样性更好地支持了复杂的异常检测；另一方面，LSS 模块的全局建模和局部特征提取，有效捕获长距离依赖与细节异常，减少漏检。然而，值得注意的是，所提方法在 AP 数值上不具备优势，因此在检测纹理复杂、背景干扰多的场景泛化能力仍有改进空间。

在定位精度方面，该模型的 AUROC 达到了 98.2%，AUPRO 达到了 94.2%。相较于现有的 MambaAD 模型，表现出更加先进的水平。这主要得益于分层类感知提示模块，通过动态集成类别特定的提示令牌，明确区分不同异常类别的语义空间，显著减少了 multi-class 异常检测中的类间干扰。此外，大核心卷积和并行的小核心卷积相辅相成，模型处理压缩信息的能力有效检测到了细微异常。在 MVTec AD 数据集上的可视化检测结果如图 4 所示。

**Table 3.** Comparison of detection and localization accuracy on the MVTec AD dataset  
**表 3.** 在 MVTec AD 数据集上的检测精度与定位精度比较

| Method         | 检测精度  |      |        | 定位精度  |      |        |       |
|----------------|-------|------|--------|-------|------|--------|-------|
|                | AUROC | AP   | F1_max | AUROC | AP   | F1_max | AUPRO |
| RKD [13]       | 94.6  | 96.5 | 95.2   | 96.1  | 48.6 | 53.8   | 91.1  |
| UniAD [15]     | 96.5  | 98.8 | 96.2   | 96.8  | 43.4 | 49.5   | 90.7  |
| SimpleNet [26] | 95.3  | 98.4 | 95.8   | 96.9  | 45.9 | 49.7   | 86.5  |
| DeSTSeg [27]   | 89.2  | 95.5 | 91.6   | 93.1  | 54.3 | 50.9   | 64.8  |
| DiAD [17]      | 97.2  | 99.0 | 96.5   | 96.8  | 52.6 | 55.5   | 90.7  |
| MambaAD [20]   | 98.6  | 99.6 | 97.8   | 97.7  | 56.3 | 59.2   | 93.1  |
| Ours           | 98.7  | 99.5 | 97.8   | 98.2  | 56.8 | 59.8   | 94.2  |

## 5. 结论

由于多类异常检测的数据集中存在大量正常样本的类内多样性，本文提出了 PGM 框架，显著扩展了模型的信息容量且最大化信息压缩，很大程度上缓解了不同正常样本互相耦合的影响。该研究通过“特征提取 - 信息融合 - 语义解耦”的全链路优化，三者协同推动多类检测在准确性与精度的突破，为多类检测任务提供了高效可靠的解决方案。在具有挑战性的无监督异常检测数据集 MVTec AD 上的广泛实验证实了 PGM 较主流方法如 DiAD 等展现出更优的异常定位精度，验证了其优越的鲁棒性与泛化性。

局限性。当前框架仍存在长距离依赖建模引发的语义干扰风险，尤其在处理工业场景中的高复杂度背景(如金属反光表面)时，易受非关键区域信息干扰。未来需要考虑设计空间显著性权重过滤器，集成预训练显著性检测模块抑制无关区域交互，同时引入稀疏激活机制优化显存消耗，以实现全局 - 局部建模的动态平衡与工业级部署需求。此外，本研究将进一步拓展 PGM 在视频工业异常检测(如流水线零件脱落监测)与多模态数据协同(如红外 - 可见光融合)中的应用潜力，推动工业异常检测与定位领域的发展。

## 参考文献

- [1] Bergmann, P., Fauser, M., Sattlegger, D. and Steger, C. (2019) MVTec AD—A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 9584-9592. <https://doi.org/10.1109/cvpr.2019.00982>
- [2] Zhang, X., Wu, Y., Angelini, E., et al. (2023) MAPSeg: Unified Unsupervised Domain Adaptation for Heterogeneous Medical Image Segmentation Based on 3D Masked Autoencoding and Pseudo-Labeling. arXiv: 2303.09373. <https://doi.org/10.48550/arXiv.2303.09373>
- [3] Wu, P., Zhou, X., Pang, G., Zhou, L., Yan, Q., Wang, P., et al. (2024) VadCLIP: Adapting Vision-Language Models for Weakly Supervised Video Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 6074-6082. <https://doi.org/10.1609/aaai.v38i6.28423>
- [4] Roth, K., Pemula, L., Zepeda, J., Scholkopf, B., Brox, T. and Gehler, P. (2022) Towards Total Recall in Industrial Anomaly Detection. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 14298-14308. <https://doi.org/10.1109/cvpr52688.2022.01392>
- [5] Rudolph, M., Wehrbein, T., Rosenhahn, B. and Wandt, B. (2023) Asymmetric Student-Teacher Networks for Industrial Anomaly Detection. 2023 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 2-7 January 2023, 2591-2601. <https://doi.org/10.1109/wacv56688.2023.00262>
- [6] Cao, Y., Wan, Q., Shen, W. and Gao, L. (2022) Informative Knowledge Distillation for Image Anomaly Segmentation. *Knowledge-Based Systems*, **248**, Article 108846. <https://doi.org/10.1016/j.knosys.2022.108846>
- [7] Batzner, K., Heckler, L. and König, R. (2024) EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies. 2024 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2024, 127-137. <https://doi.org/10.1109/wacv57701.2024.00020>

- [8] Li, C., Sohn, K., Yoon, J. and Pfister, T. (2021) CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 9659-9669. <https://doi.org/10.1109/cvpr46437.2021.00954>
- [9] Mishra, P., Verk, R., Fornasier, D., Piciarelli, C. and Foresti, G.L. (2021) VT-ADL: A Vision Transformer Network for Image Anomaly Detection and Localization. 2021 *IEEE 30th International Symposium on Industrial Electronics (ISIE)*, Kyoto, 20-23 June 2021, 1-6. <https://doi.org/10.1109/isie45552.2021.9576231>
- [10] Ding, X., Zhang, Y., Ge, Y., Zhao, S., Song, L., Yue, X., *et al.* (2024) UniRepLKNet: A Universal Perception Large-Kernel Convnet for Audio, Video, Point Cloud, Time-Series and Image Recognition. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 5513-5524. <https://doi.org/10.1109/cvpr52733.2024.00527>
- [11] Yao, X., Li, R., Qian, Z., Wang, L. and Zhang, C. (2024) Hierarchical Gaussian Mixture Normalizing Flow Modeling for Unified Anomaly Detection. *Computer Vision—ECCV 2024*, Milan, 29 September-4 October 2024, 92-108. [https://doi.org/10.1007/978-3-031-73411-3\\_6](https://doi.org/10.1007/978-3-031-73411-3_6)
- [12] Liu, J. and Wang, F. (2024) Mixed-Attention Auto Encoder for Multi-Class Industrial Anomaly Detection. *ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, 14-19 April 2024, 4120-4124. <https://doi.org/10.1109/icassp48485.2024.10446794>
- [13] Deng, H. and Li, X. (2022) Anomaly Detection via Reverse Distillation from One-Class Embedding. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 9727-9736. <https://doi.org/10.1109/cvpr52688.2022.00951>
- [14] He, L., Jiang, Z., Peng, J., Zhu, W., Liu, L., Du, Q., *et al.* (2024) Learning Unified Reference Representation for Unsupervised Multi-Class Anomaly Detection. *Computer Vision—ECCV 2024*, Milan, 29 September-4 October 2024, 216-232. [https://doi.org/10.1007/978-3-031-72855-6\\_13](https://doi.org/10.1007/978-3-031-72855-6_13)
- [15] You, Z., Cui, L., Shen, Y., *et al.* (2022) A Unified Model for Multi-Class Anomaly Detection. *Advances in Neural Information Processing Systems*, **35**, 4571-4584.
- [16] Gao, B. (2024) Learning to Detect Multi-Class Anomalies with Just One Normal Image Prompt. *Computer Vision—ECCV 2024*, Milan, 29 September-4 October 2024, 454-470. [https://doi.org/10.1007/978-3-031-72855-6\\_26](https://doi.org/10.1007/978-3-031-72855-6_26)
- [17] He, H., Zhang, J., Chen, H., Chen, X., Li, Z., Chen, X., *et al.* (2024) A Diffusion-Based Framework for Multi-Class Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 8472-8480. <https://doi.org/10.1609/aaai.v38i8.28690>
- [18] Wang, F., Liu, C., Shi, L., *et al.* (2024) MiniMaxAD: A Lightweight Autoencoder for Feature-Rich Anomaly Detection. arXiv: 2405.09933. <https://doi.org/10.48550/arXiv.2405.09933>
- [19] Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., *et al.* (2023) ConvNeXt V2: Co-Designing and Scaling Convnets with Masked Autoencoders. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 16133-16142. <https://doi.org/10.1109/cvpr52729.2023.01548>
- [20] He, H., Bai, Y., Zhang, J., *et al.* (2024) MambaAD: Exploring State Space Models for Multi-Class Unsupervised Anomaly Detection. arXiv: 2404.06564. <https://doi.org/10.48550/arXiv.2404.06564>
- [21] Kashiani, H., Talemi, N.A. and Afghah, F. (2024) ROADS: Robust Prompt-Driven Multi-Class Anomaly Detection under Domain Shift. arXiv: 2411.16049. <https://doi.org/10.48550/arXiv.2411.16049>
- [22] Glorot, X. and Bengio, Y. (2010) Understanding the Difficulty of Training Deep Feedforward Neural Networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia, 13-15 May 2010, 249-256.
- [23] Bergmann, P., Fauser, M., Sattlegger, D. and Steger, C. (2020) Uninformed Students: Student-Teacher Anomaly Detection with Discriminative Latent Embeddings. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 4182-4191. <https://doi.org/10.1109/cvpr42600.2020.00424>
- [24] Zavrtanik, V., Kristan, M. and Skocaj, D. (2021) DRÆM—A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 8310-8319. <https://doi.org/10.1109/iccv48922.2021.00822>
- [25] Zou, Y., Jeong, J., Pemula, L., Zhang, D. and Dabeer, O. (2022) Spot-the-Difference Self-Supervised Pre-Training for Anomaly Detection and Segmentation. *Computer Vision—ECCV 2022*, Tel Aviv, 23-27 October 2022, 392-408. [https://doi.org/10.1007/978-3-031-20056-4\\_23](https://doi.org/10.1007/978-3-031-20056-4_23)
- [26] Liu, Z., Zhou, Y., Xu, Y. and Wang, Z. (2023) SimpleNet: A Simple Network for Image Anomaly Detection and Localization. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 20402-20411. <https://doi.org/10.1109/cvpr52729.2023.01954>
- [27] Zhang, X., Li, S., Li, X., Huang, P., Shan, J. and Chen, T. (2023) DeSTSeg: Segmentation Guided Denoising Student-Teacher for Anomaly Detection. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 3914-3923. <https://doi.org/10.1109/cvpr52729.2023.00381>