计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(8): 2111−2126, 2026

DiffAD: 基于差分卷积注意力的多类无监督异常检测

姚 健[1,2] 马 帅[1,2] 钱鹏江[1,2] 王 闯[1,2] 杨桂松[3] 何 源[4,5]

[1](江南大学人工智能与计算机学院(软件学院) 江苏无锡 214122)

[2](康养智能化技术教育部工程研究中心(江南大学) 江苏无锡 214122)

[3](上海理工大学光电信息与计算机工程学院 上海 200093)

[4](清华大学软件学院 北京 100084)

[5](北京信息科学与技术国家研究中心(清华大学) 北京 100084)

(jianyao_ai@jiangnan.edu.cn)

DiffAD: Difference Convolution Attention-Based Multi-Class Unsupervised Anomaly Detection

Yao Jian[1,2], Ma Shuai[1,2], Qian Pengjiang[1,2], Wang Chuang[1,2], Yang Guisong[3], and He Yuan[4,5]

[1] (*School of Artificial Intelligence and Computer Science* (*School of Software*), *Jiangnan University*, *Wuxi*, *Jiangsu* 214122)

[2] (*Engineering Research Center of Intelligent Technology for Healthcare* (*Jiangnan University*), *Ministry of Education*, *Wuxi*, *Jiangsu* 214122)

[3] (*School of Optical-Electrical and Computer Engineering*, *University of Shanghai for Science and Technology*, *Shanghai* 200093)

[4] (*School of Software*, *Tsinghua University*, *Beijing* 100084)

[5] (*Beijing National Research Center for Information Science and Technology* (*Tsinghua University*), *Beijing* 100084)

Abstract A novel unsupervised multi-class anomaly detection model, termed DiffAD, is proposed to tackle the critical challenges of annotation scarcity and insufficient detection accuracy plaguing visual anomaly detection tasks in complex industrial scenarios. The model employs a progressive feature reconstruction strategy, with its core innovation centered on a dedicated feature reconstruction component named DADE. Breaking new ground, DADE decomposes the feature reconstruction process into three sequential phases, namely chaos, pre-refinement and strong refinement, which synergistically enhance the quality of reconstructed features and ensure the stability of the entire reconstruction workflow. A key standout of DADE lies in its seamless integration of difference convolution attention and a detail enhancement mechanism. It combines difference convolution with multi-head self-attention and incorporates residual dense connections, thereby enabling the component to markedly strengthen the model's ability to capture subtle image variations and high-frequency details that are critical for anomaly identification and substantially elevate the precision of pixel-level anomaly localization. Extensive empirical evaluations are conducted on four representative benchmark datasets, including MVTec-AD, VisA, MVTec-3D and Uni-Medical. The experimental results consistently demonstrate that DiffAD achieves significant performance superiority over state-of-the-art mainstream models across both image-level and pixel-level anomaly detection metrics, fully underscoring its remarkable practical application value and great potential for widespread deployment in the domain of unsupervised visual anomaly detection.

Key words unsupervised; anomaly detection; computer vision; attention mechanism; feature reconstruction

收稿日期: 2025-09-21; **修回日期**: 2026-03-30

基金项目: 国家自然科学基金项目(62572218, 62172192); 江苏省基础研究项目(BK20252080)

This work was supported by the National Natural Science Foundation of China (62572218, 62172192) and the Basic Research Program of Jiangsu Province (BK20252080).

通信作者: 钱鹏江(qianpjiang@jiangnan.edu.cn)

摘 要 提出一种新颖的多类别无监督异常检测模型——DiffAD,以应对复杂工业场景下视觉异常检测所面临的标注匮乏与精度不足等挑战。该模型采用渐进式特征重建策略,其核心在于设计了一个特征重建组件 DADE,该组件创新性地将重建过程划分为混沌、预细化和强细化 3 个阶段,有效提升了重建质量与过程稳定性。DADE 的突出创新点在于融合差分卷积注意力与细节增强机制,通过整合差分卷积与多头自注意力机制,并借助残差密集连接,显著增强了对图像细微变化及高频信息的捕捉能力,进而提升了异常定位的精准度。在 MVTec-AD, VisA, MVTec-3D, Uni-Medical 这 4 个代表性数据集上的广泛实验表明,DiffAD 在图像级和像素级异常检测指标上整体表现显著优于现有主流模型,充分彰显了其在无监督视觉检测领域的实际应用价值与潜力。

关键词 无监督;异常检测;计算机视觉;注意力机制;特征重建

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202550706 **CSTR:** 32373.14.issn1000-1239.202550706

视觉异常检测(anomaly detection, AD)在智能制造领域具有举足轻重的地位,它能够及时识别生产过程中出现的异常现象,从而确保产品质量并提升生产效率。近年来,随着工业自动化水平的不断提升,该领域的研究也受到了广泛关注[1-2]。虽然监督学习模型[3]在众多视觉任务中表现优异,但在实际工业场景中应用于异常检测时仍面临以下 2 点主要挑战:其一,缺陷样本往往稀缺,难以支撑监督模型的充分训练;其二,监督模型通常对训练集未见过的“未知异常”缺乏检测能力。无监督异常检测模型[4-17]的出现则能有效解决上述问题。一方面,生产过程中通常可以获取到大量的正常样本,足以训练一个无需异常样本的异常检测模型。另一方面,通过正常样本训练的模型仅学习正常数据的分布特征,因此,在理想情况下可以检测出任何偏离正常分布的各类未知异常。

现有模型通常基于单类无监督异常检测进行设计[6-11],这意味着每个类别都需要训练一个独立的模型,从而显著增加了训练和存储成本。为缓解这一问题,UniAD[4]引入了多类设置,但在提升性能和降低训练成本等方面,仍然存在很大的改进空间。我们的工作着力于解决这个更具挑战性的多类无监督异常检测任务。

当前,无监督异常检测模型可大致归纳为三大类:基于重建的模型[8,17-18]、基于数据增强的模型[7,12-13]以及基于特征嵌入的模型[6,9]。其中亦有一些集成策略取得较好性能[8,14]。基于重建的模型因其实现简便且效果显著而被广泛采用,该类模型通过像素级别的比较来识别异常,从而实现精确定位。具体而言,模型仅在正常图像上进行训练,使其能够重建正常样本。当输入异常图像时,模型仍会尝试将其重建为正常图像。通过计算输入图像与重建图像之间的残差,即可获得异常区域的定位结果。然而,在处理具有复杂结构和纹理的图像时,重建过程易发生失败,导致网络产生“恒等映射”现象[12](即简单地复制输入)。为克服复杂重建的挑战,研究者们提出基于特征级别重建的模型[5,15]。虽然此类模型在分割精度方面可能不及像素级重建,但在性能和泛化能力方面具有显著优势。基于上述考虑,我们选择采用基于特征重建策略。

现有的特征重建网络结构设计各具特色。其中,基于卷积的 RD4AD[16]能够有效捕获局部上下文信息,但其在建立远程依赖关系方面存在局限性。UniAD[4]依赖于预训练的编码器和 Transformer[19]解码器架构,具备卓越的全局建模能力,但对局部细节的刻画不足。可变形注意力[20]虽能自适应采样局部特征,但计算复杂度较高,且缺乏对异常边界高频信息的强化。最新的视觉状态空间模型(Vmamba)[21]擅长远程依赖建模,但对局部细节(如微小划痕、边缘缺损)的捕捉能力同样较弱。因此,如何设计有效的特征融合机制,在保持全局上下文建模能力的同时增强对局部细节特征的感知,已成为推进多类无监督异常检测发展的关键问题。例如,Quan 等人[22]提出可学习令牌(token)注意力和卷积双分支结构,对不同细粒度特征进行重建。然而,现有工作未充分关注融合过程的稳定性,且缺乏对细节的进一步强化。

我们融合 Transformer 与卷积及其变体,提出了一种渐进式重建组件 DADE(difference attention and detail enhance reconstruction),以提升重建细粒度质量和过程稳定性,并将其应用于多类异常检测。具体而言,我们将重建过程划分为多个阶段:细化阶段与

混沌阶段。细化阶段进一步细分为预细化(或称过渡)和强细化阶段。在混沌阶段,我们采用全局与局部的多分支结构,以融合多样化的信息。我们引入了一种用于梯度增强的差分卷积[23](DEConv,详见1.4节),将其作为自注意力机制的前置模块,通过放大输入特征中的细微变化和高频信息来增强梯度,由此提出了DiffAttn(difference attention)的注意力机制,使模型更加关注边缘、拐点和纹理等关键细节,并进行了其他注意力机制的替换实验。鉴于简单的相加或拼接操作难以稳定地融合全局与局部分支的信息。因此,在进入强细化阶段之前,我们引入了2种简单高效的注意力机制进行过渡,分别为通道注意力和空间注意力。在强细化阶段,结合差分卷积、普通卷积和残差密集连接策略,促进多样化信息融合与复用,进而有效提高重建质量,整个细化模块命名为细节增强(detail enhance, DE)。通过堆叠和串联多个DADE,我们构建了完整的DADE解码器DiffAD。实验结果表明,DiffAD在4个具有代表性的异常检测数据集上实现了具有竞争力的性能。我们的贡献如下:

1)提出DiffAD,一种创新性的渐进式重建架构,通过多阶段细分特征重建过程解决多类无监督异常检测问题,为相关研究提供了新的解决思路。

2)结合普通卷积与差分卷积提取多样化特征,并将其整合到多头自注意力机制中,通过大量应用差分卷积强化细节,提出DiffAttn和DE这2个核心模块,显著增强了图像处理的特征表达能力,提升了模型捕捉复杂依赖关系的能力。

3)实验表明,DiffAD在7项图像级和像素级评估指标上均取得优异性能,同时保持较低的参数量和计算开销,证明了其在多类无监督异常检测任务中的实用性与通用性。

1 相关工作

1.1 无监督异常检测

视觉异常检测作为计算机视觉的重要问题,已衍生出监督[2]、半监督[24]与无监督等多种模型识别图像异常区域。鉴于工业场景中缺陷样本稀缺且异常类型难以枚举,我们聚焦无监督设置并回顾相关研究。

无监督异常检测模型仅依赖正常训练数据来区分测试过程中的异常区域,通常可以分为以下3类:

1)基于重建的模型。该类模型通常由编码器和解码器组成,主要通过将异常图像重建为正常图像,并利用手工设置的测量函数[17]来获取输入数据与重建数据之间的差异,从而实现异常定位。衍生出多种技术路径,包括自编码器(autoencoder)[25-26]、生成对抗网络(GAN)[27]、Transformer[4,28-29]以及扩散模型(diffusion model)[30]。这些模型通常依赖RGB图像的像素级重建,例如Liu等人[25]从原始RGB图像的灰度边缘重建RGB图像,模型的跳跃连接保留了图像的高频信息,从而更好地指导重建。Zavrtanik等人[26]通过使用黑色漆块覆盖原图的方式使网络学习填补区域。同时,基于特征重建的检测模型[4,12]则关注更具表现力的深层特征,通常获得更好的性能。Zhang等人[12]引入自适应特征选择技术,减小深层特征维度,降低重建开销,提高了对异常的敏感度。

2)基于数据增强的模型。此类模型通过图像合成策略,生成伪造的异常区域,构建伪标签,从而实现自监督学习。NSA[13]利用泊松图像编辑进行无缝图像粘贴,以合成更自然的异常区域。最新的研究[12]推出了一种可控强度的扩散异常合成(SDAS)方法,可以生成更接近自然分布的缺陷样本,并提供了控制异常强度的灵活性。Zavrtanik等人[8]利用分布外数据进行异常合成的研究,综合纹理数据集DTD[31]模拟了各种纹理异常。

3)基于特征嵌入的模型。这类模型可以分为师生架构[15]、单类分类[11]和记忆库[9],主要专注于计算和分析网络的深层次特征。DeSTSeg[15]整合预训练教师网络、去噪学生编码器-解码器和分割网络,融合多级特征提升了对异常的适用性。P-SVDD[11]将图像裁剪为小块,并应用训练的超球边界决策图像块是否异常。

1.2 多类无监督异常检测

当前,多数异常检测模型[6-11]普遍采用单一类别设定模式,即针对每个类别均需独立训练并测试专属模型,且多围绕单一类别特性进行设计优化。这种设计在多类别应用场景下存在明显局限:随着类别数量增加,计算资源消耗急剧上升,且难以有效处理类内差异较大的情况。You等人[4]提出了一种统一框架UniAD,训练一个单一的模型用于检测所有类别物品的缺陷,同时提出了一种特征抖动策略,迫使模型即使在输入噪声的情况下也能恢复正确的信息。He等人[5]将Mamba[32]引入异常检测领域,结合级联混合状态空间(HSS)块(含5种扫描方式与8个方向)构建了一个重建解码器。DiAD[30]研究了一种基于扩散模型的异常检测框架,引入了语义引导网

络来保证重建图像语义的一致性。

1.3 特征重建架构

我们总结了 2 种效果出色的重建体系结构。如图 1(a)所示,第 1 类架构[5,28]在瓶颈层融合编码器不同层次的多尺度特征(多为中间层或全层融合)后逐步上采样解码,各架构在具体设计上略有差异。图 1(b)展示的第 2 类架构则对编码器每一层的特征进行独立重建,这在一定程度上有效避免了不同信息层的干扰。我们也认为,浅层特征与深层特征的融合是复杂过程。尽管第 2 类架构在信息处理上有优势,但在有效控制参数量和运算量方面相对较弱,文献 [12, 33] 均强调了特征选择的重要性并降低特征维度,提供了优化思路。基于图 1(a)架构在轻量化设计上有优势,我们也遵循该架构并探索如何有效重建浅层与深层特征的融合。

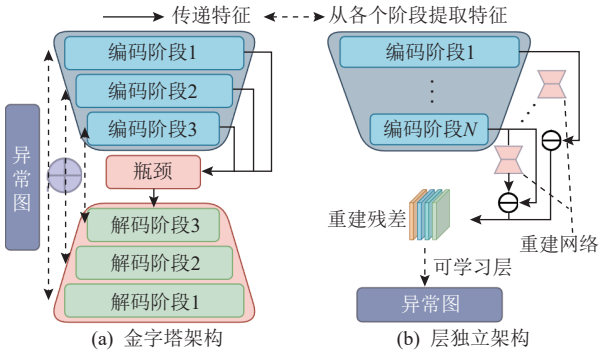


Fig. 1 Two different reconstruction architectures

图 1 2 类不同重建架构

1.4 差分卷积

差分卷积的起源可以追溯到 LBP(local binary pattern)[34],它将局部块中的像素差异编码为十进制数用于纹理分类。Yu 等人[35]提出了中央差分卷积(central difference convolution, CDC),通过关注像素间的差异,提升了光照不变性和细粒度特征提取能力,从而在面部活体检测任务中增强了模型对光照变化的鲁棒性和对伪造细节的识别能力,有效减少了传统卷积的局限。Chen 等人[23]在单图像中使用了差分卷积用于图像去雾任务,在 CDC 和角差分卷积(angular difference convolution, ADC)基础上还推导了水平差分卷积(horizontal difference convolution, HDC)和垂直差分卷积(vertical difference convolution, VDC),并结合了传统卷积提出了一个新的卷积结构 DEConv。这些差分卷积主要用于增强梯度级信息,而普通卷积用于获得强度级信息。我们将差分卷积应用于重建任务,打造了一个渐进式的解码器,有望精细重建质量,提升检测性能。

注意到差分卷积包含多个并行卷积,这并不会带来参数上的负担,事实上,在部署推理阶段,利用卷积的可加性将并行的多个卷积简化为单个标准卷积。计算如下:

$$DEConv(\boldsymbol{E})=\sum_{i=1}^{5}\boldsymbol{E}*\boldsymbol{K}_i=\boldsymbol{E}*\left(\sum_{i=1}^{5}\boldsymbol{K}_i\right)=\boldsymbol{E}*\boldsymbol{K}_{\text{tran}},\quad(1)$$

其中$\boldsymbol{E}$为输入特征,$\boldsymbol{K}_i$代表卷积,$*$为卷积操作,$\boldsymbol{K}_{\text{tran}}$表示将并行卷积组合成的转换核。参考普通卷积,可推演出分组和深度差分卷积。

2 本文方法

2.1 整体架构

DiffAD 模型的整体架构如图 2 所示,一个 Half-FPN 瓶颈层融合来自编码器前 3 阶段的多尺度特征,馈送到解码器进行多尺度特征重建。所提出的 DADE 解码器结构简单,是一个渐进式重建的过程。图 2 中,空间注意力中 CMP(channel-wise max pooling)和 CAP(channel-wise average pooling)分别为沿着通道维度的最大池化和平均池化,通道注意力中 GMP(global max pooling)和 GAP(global average pooling)分别为全局最大池化和全局平均池化,各种卷积块均包含实例归一化和 SiLU 激活函数。

1)编码器。尽管有许多新兴的网络架构(如 Vision Transformer[19], Vmamba[21] 等)在某些任务上表现更好,但 ResNet[36] 的稳健性和易用性使其在许多情况下仍然是一个可靠的选择。我们使用在 ImageNet[37] 上预训练的 ResNet34 作为编码器,并在训练过程中冻结其参数。

2)解码器。不同编码器阶段提取包含不同分辨率的特征,从早期阶段提取的特征包含局部细节信息,而从深层阶段提取的特征拥有更高层次的语义知识[38],同时也对应着不同大小的异常区域,融合这些语义信息对重建任务有很大帮助。

每个阶段的解码器由多个 DADE 堆叠,每个 DADE 均包含一个 GLA(global and local attention)和一个 DE。解码器的不同阶段重建不同尺度的特征,用于和编码器对应阶段生成多尺度异常图。

2.2 差分注意力

DiffAttn 通过将普通卷积和差分卷积提取的特征应用于多头自注意力机制,显著增强了模型的性能。在这一过程中,多头自注意力机制具有捕获全局上下文信息的能力,普通卷积提供的强度信息与差分卷积所提取的梯度信息相结合,为自注意力机

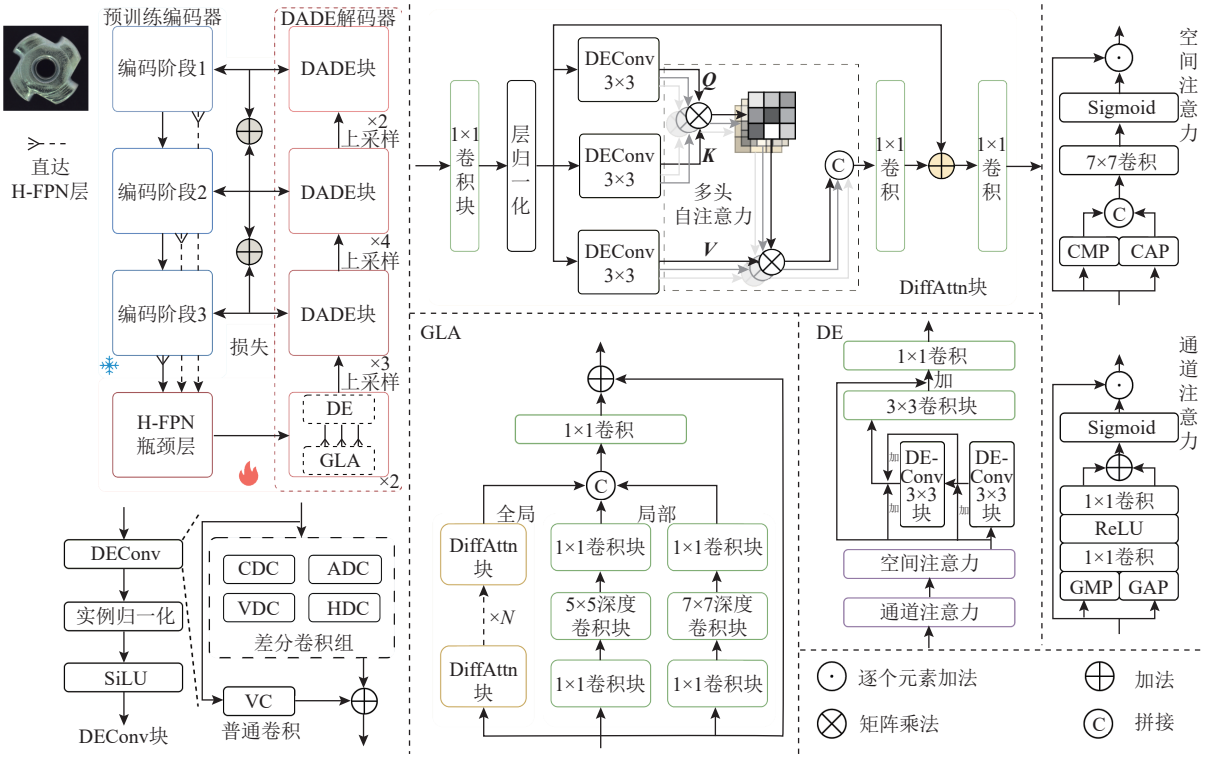


Fig. 2 The overall architecture of DiffAD

图 2 DiffAD 总体架构

制输入提供了更加全面的信息，从而有效提升特征表示的丰富性和细致度。这种方法使得模型在处理图像任务时，能够灵活地关注不同尺度和方向的特征变化，提高特征表达能力。

设 $\boldsymbol{X}_k \in \mathbb{R}^{\hat{C}\times\hat{H}\times\hat{W}}$ 为第 k 阶段需要重建的特征，注意力中查询（query, $\boldsymbol{Q}$）、键（key, $\boldsymbol{K}$）、值（value, $\boldsymbol{V}$）计算如下：

$$\boldsymbol{X}_k^{LN} = LN(Conv_{1\times1}(\boldsymbol{X}_k)), \tag{2}$$

$$\boldsymbol{X}_k^{\mathrm{QKV}} = DWDEConv_{3\times3}(\boldsymbol{X}_k^{LN}), \tag{3}$$

其中 $Conv_{1\times1}(\cdot)$ 是 1×1 卷积，$LN(\cdot)$ 为层归一化，$DWDEConv_{3\times3}(\cdot)$ 为深度差分卷积，$\boldsymbol{X}_k^{\boldsymbol{QKV}} \in \mathbb{R}^{(3\times\hat{C})\times\hat{H}\times\hat{W}}$。将其按通道分割并变形出 $\boldsymbol{Q}_k, \boldsymbol{K}_k, \boldsymbol{V}_k \in \mathbb{R}^{(n_\mathrm{h}\times(\hat{C}/n_\mathrm{h}))\times\hat{H}\times\hat{W}}$，$n_\mathrm{h}$ 是注意力头的数量，第 i 个头的注意力计算过程为：

$$\boldsymbol{h}_{i,k}=Attention(\boldsymbol{Q}_{i,k},\boldsymbol{K}_{i,k},\boldsymbol{V}_{i,k}) = softmax\left(\frac{\boldsymbol{Q}_{i,k}\boldsymbol{K}_{i,k}^{\mathrm{T}}}{\sqrt{d_{\boldsymbol{K}_k}}}\right)\boldsymbol{V}_{i,k}, \tag{4}$$

其中 $\boldsymbol{Q}_{i,k}, \boldsymbol{K}_{i,k}, \boldsymbol{V}_{i,k} \in \mathbb{R}^{(\hat{C}/n_\mathrm{h})\times(\hat{H}\times\hat{W})}$，$d_{\boldsymbol{K}_k}$ 是 $\boldsymbol{K}_k$ 的维度。一次的多头注意力计算的结果为：

$$MHSA(\boldsymbol{X}_k)=(\boldsymbol{h}_{1,k},\boldsymbol{h}_{2,k},\cdots,\boldsymbol{h}_{n_\mathrm{h},k}), \tag{5}$$

其中 [⋯] 是拼接操作。最终整个 DA 模块的计算流程如下：

$$DA(\boldsymbol{X}_k) = ConvB_{1\times1}(Conv_{1\times1}(MHSA(\boldsymbol{X}_k)) + \boldsymbol{X}_k^{LN}), \tag{6}$$

其中 $ConvB_{1\times1}(\cdot)$ 为 1×1 卷积块，我们堆叠了多个 DA，同时并行地扩展了大核深度可分离卷积分支，用于捕捉局部强度信息，如上构成了 GLA 模块。我们形象地将此重建阶段定义为混沌重建阶段。第 k 阶段的 GLA 计算为：

$$GLA(\boldsymbol{X}_k) = Conv_{1\times1}(DA_N(DA_{N-1}(\cdots DA_1(\boldsymbol{X}_k)\cdots)) + DWConvB_{5\times5}(\boldsymbol{X}_k) + DWConvB_{7\times7}(\boldsymbol{X}_k)) + \boldsymbol{X}_k, \tag{7}$$

其中 N 是可调的超参数，用于平衡全局语义建模与局部细节捕捉的权重分配与融合效能，$DWConvB(\cdot)$ 是深度卷积块。

2.3 细节增强

GLA 为混沌重建阶段，细节增强 DE 分为预细化（preliminary refinement, PR）和强细化（strong refinement, SR）。预细化引入了通道注意力和空间注意力，二者的引入能够显著提升重建准确性、模型鲁棒性和特征选择自适应性，促进对重要区域和特征的聚焦，使得重建结果更清晰和有效，为强细化的稳定进行打下基础。将 $GLA(\boldsymbol{X}_k)$ 记作 $\boldsymbol{X}_k^{\mathrm{gl}}$，预细化描述为：

$$PR(\boldsymbol{X}_k^{\mathrm{gl}}) = SA(CA(\boldsymbol{X}_k^{\mathrm{gl}})), \tag{8}$$

$PR(\boldsymbol{X}_k^{\mathrm{gl}})$ 记作 $\boldsymbol{X}_k^{\mathrm{pr}}$，其中 $CA(\cdot)$ 为通道注意力，$SA(\cdot)$ 为空间注意力。

强细化使用小尺寸差分卷积和普通卷积，残差

连接的引入使得每个输入可以“跳过”某些层直接传递到后面的层,网络中的某些特征可以更灵活地被重用。这样的特征重用有助于构建更加复杂的特征表示,进而提高模型的表现。强细化计算如下:

$$\begin{cases} \boldsymbol{X}_k^{t_1} = DWDEConvB_{3\times3}(\boldsymbol{X}_k^{\mathrm{pr}}), \\ \boldsymbol{X}_k^{t_2} = DWDEConvB_{3\times3}(\boldsymbol{X}_k^{\mathrm{pr}} + \boldsymbol{X}_k^{t_1}), \\ \boldsymbol{X}_k^{t_3} = ConvB_{3\times3}(\boldsymbol{X}_k^{t_1} + \boldsymbol{X}_k^{t_2} + \boldsymbol{X}_k^{\mathrm{pr}}), \\ SR(\boldsymbol{X}_k^{\mathrm{pr}}) = Conv_{1\times1}(\boldsymbol{X}_k^{\mathrm{pr}} + \boldsymbol{X}_k^{t_3}), \end{cases} \tag{9}$$

其中$\boldsymbol{X}_k^{t_1}$, $\boldsymbol{X}_k^{t_2}$, $\boldsymbol{X}_k^{t_3}$是为了方便计算设置的中间变量,用到的$DWDEConvB_{3\times3}(\cdot)$为深度差分卷积块。

2.4 DADE 解码器

DADE 模块由 GLA 模块与 DE 模块串联而成,其设计理念旨在通过层级化的结构逐步增强特征表达能力。每个重建组件都是一个循序渐进的过程,强调渐进式重建的重要性。在这一框架下,每个阶段的重建都堆叠了多个 DADE 模块,从而有效地提升了模型在特征提取和重建方面的能力。

GLA 模块负责全局和局部特征的捕捉,确保模型可以从全局视角理解输入数据的整体结构。而 DE 模块则在细化重建过程中发挥关键作用,针对局部特征做出精准的调节和优化,确保重建细节的丰富性。

DADE 具有如下的优点:

1)增强表达能力,通过堆叠多个 DADE 模块,模型能够有效捕捉到丰富的层级特征,因此在复杂任务处理上展现出更强的表达能力。这一设计使得模型具有较好的泛化能力,能够适应不同类型的输入数据。

2)逐步细化重建,采用循序渐进的重建策略,可以在每个阶段逐步优化输出结果,有助于减少信息的丢失。

2.5 训练约束

我们使用均方差(mean squared error, MSE)损失来优化多尺度特征重建,损失函数$\mathcal{L}_{\mathrm{rec}}$定义为:

$$\mathcal{L}_{\mathrm{rec}} = \sum_{i=1}^{3} \|\boldsymbol{I}_i - \boldsymbol{R}_i\|_2 。 \tag{10}$$

编码器仅提取 3 个阶段特征,整体网络在 3 个尺度上进行监督,其中$\boldsymbol{I}_i$和$\boldsymbol{R}_i$分别为编码器和重建网络各阶段对应特征。

3 实验结果与分析

3.1 实验设置

1)数据集。我们在 4 个具有挑战性的数据集上评估了所提出的模型,分别为 MVTec-AD[39], VisA[40], MVTec-3D[41] 和一个由 Zhang 等人[28] 综合整理的医学数据集 Uni-Medical。MVTec-AD 包含 5 种类型的纹理和 10 种类型的物体,总计 5 354 张高分辨率图像,其中 3 629 张正常图像被指定用于训练,剩下的 1 725 张图像保留用于测试,包括正常样本和异常样本。VisA 非常具有挑战性,有 12 个不同的对象,包括 3 种不同的类型:复杂结构、多实例和单实例,由 10 821 张图像组成,其中 9 621 张是正常样本,1 200 张是异常样本。MVTec-3D 是一个用于无监督异常检测和定位任务的综合 3D 数据集,包含由工业 3D 传感器获得的 4 000 多个高分辨率扫描,10 个不同的对象类别中的每一个都包含 1 组无缺陷的训练和验证样本,以及 1 组具有各种缺陷的样本,实验中使用其提供的 RGB 图像作为训练和验证。Uni-Medical 包含脑部 MRI、肝脏 CT 和视网膜 OCT 这 3 类样本,形状变化多样。

2)指标。对于异常检测和分割,我们采用了多个评估指标,包括接收者操作特征曲线下面积(area under the receiver operating characteristic curve, AUROC)、平均精度(average precision, AP)以及最大 $F1$ 分数($F1$_max)。此外,对于异常分割,我们还使用了每个区域重叠曲线下面积(area under the per-region overlap curve, AU-PRO)。为全面评估模型性能,我们计算了这 7 个评估指标的算术平均值,记为 mAD[28]。

3)实验细节。所有图像被调整为256×256的统一大小,没有进行数据增强,未引入异常合成策略,解码器中除了 DiffAttn 中,所有归一化均为实例归一化,我们的模型是面向多类重建任务,只期望关注单个类别的均值和方差,不宜使用批量归一化。GLA 中N设置为 2,解码器的深度(如图 2 所示的 4 组 DADE 堆叠情况)依次为 2, 3, 4, 2。采用 AdamW 优化器,学习率为 0.001,衰减率为 0.000 1,批次大小为 16,在单台 NVIDIA GeForce RTX™ 4090 24 GB GPU 上进行了 500 轮的多类设置训练,400 轮之后学习率调整为 0.000 1,上述实验设置在所有数据集上都是固定的,没有针对性的微调。测试过程中使用不同尺度上特征的余弦相似度之和作为异常图。

3.2 对比实验

我们主要对比了 6 个模型,分别为专注于多类异常检测的 UniAD[4] 和 DiAD[30],基于重建的 RD4AD[16] 和 Omni-AD[22] 以及基于深度特征嵌入的 SimpleNet[42] 和 DeSTSeg[15]。我们使用 3.1 节介绍的指标以及参数量和运算量进行了全方位的对比实验。

1)定量对比。如表 1 和图 3 所示，所提出的 DiffAD 在 4 个数据集上优于所有评估方案，在应用最广泛的 MVTec-AD 数据集上，DiffAD 优于 DiAD，在图像级 AU-ROC/AP/*F*1_max 上的结果为 98.7%/99.4%/97.5%，实现了 DiAD 1.5, 0.4, 1.0 个百分点的性能增益，在像素级 AU-ROC/AP/*F*1_max/AU-PRO 指标上，DiffAD 获得了 97.8%, 56.1%, 59.2, 93.1% 的优秀结果，实现了 1.0, 3.5, 3.7, 2.4 个百分点的性能增益，在平均指标 mAD 实现了 1.9 个百分点的性能增益。Omni-AD 基于可学习令牌注意力机制，侧重全局语义建模，在前景对象明显、异常轮廓清晰的类别中图像级指标略优，而 DiffAD 在纹理结构复杂类别中像素级指标优势显著，平均提升约 1.8 个百分点。在其他 3 个数据集上 DiffAD 的增益同样明显，这表明 DiffAD 具有很强的跨不同类型数据集的泛化能力以及出色的性能。

Table 1 Quantitative Comparison of Different Models on Four Datasets

表 1 在 4 个数据集上与不同模型的定量对比 %

数据集	模型	图像级			像素级				mAD
		AU-ROC	AP	*F*1_max	AU-ROC	AP	*F*1_max	AU-PRO	
MVTec-AD	RD4AD	94.6	96.5	95.2	96.1	48.6	53.8	91.1	82.3
	UniAD	96.5	98.8	96.2	96.8	43.4	49.5	90.7	81.7
	SimpleNet	95.3	98.4	95.8	96.9	45.9	49.7	86.5	81.2
	DeSTSeg	89.2	95.5	91.6	93.1	54.3	50.9	64.8	77.1
	DiAD	97.2	99.0	96.5	96.8	52.6	55.5	90.7	84.0
	Omni-AD	**98.7**	**99.5**	**98.1**	**97.8**	54.7	58.6	**93.2**	85.8
	DiffAD（本文）	**98.7**	99.4	97.5	**97.8**	**56.1**	**59.2**	93.1	**86.0**
VisA	RD4AD	92.4	92.4	89.6	98.1	38.0	42.6	91.8	77.8
	UniAD	88.8	90.8	85.8	98.3	33.7	39.0	85.5	74.6
	SimpleNet	87.2	87.0	81.8	96.8	34.7	37.8	81.4	72.4
	DeSTSeg	88.9	89.0	85.2	96.1	39.6	43.4	67.4	72.8
	DiAD	86.8	88.3	85.1	96.0	26.1	33.0	75.2	70.1
	Omni-AD	94.7	95.5	90.3	**99.0**	40.8	44.9	92.1	79.6
	DiffAD（本文）	**95.2**	**95.9**	**90.9**	**99.0**	**42.9**	**46.1**	**91.9**	**80.3**
MVTec-3D	RD4AD	77.9	92.4	91.4	98.4	29.8	36.4	93.5	74.3
	UniAD	78.9	93.4	91.4	96.5	21.2	28.0	88.1	71.1
	SimpleNet	72.5	91.0	90.3	93.6	18.3	25.3	77.6	66.9
	DeSTSeg	79.6	94.1	90.6	95.1	**38.1**	**39.9**	46.4	69.1
	DiAD	**84.6**	94.8	**95.6**	96.4	25.3	32.3	87.8	73.8
	Omni-AD	81.9	94.2	92.4	**98.7**	34.4	38.8	94.6	76.4
	DiffAD（本文）	84.5	**95.1**	92.3	**98.7**	34.7	39.2	**94.9**	**77.1**
Uni-Medical	RD4AD	75.6	75.8	78.0	96.5	38.7	40.1	86.4	70.2
	UniAD	78.5	75.2	76.6	96.4	37.6	40.2	85.0	69.9
	SimpleNet	75.6	76.9	76.8	95.9	38.3	39.6	80.5	69.1
	DeSTSeg	80.7	79.8	78.7	86.6	38.0	37.5	25.1	60.9
	DiAD	80.4	80.1	77.8	95.9	38.0	35.6	85.4	70.5
	Omni-AD	80.6	77.2	79.7	97.1	43.0	44.4	**88.6**	72.9
	DiffAD（本文）	**83.6**	**80.6**	**80.8**	**97.3**	**44.5**	**47.0**	**88.6**	**74.6**

注：黑体数值表示最优结果。

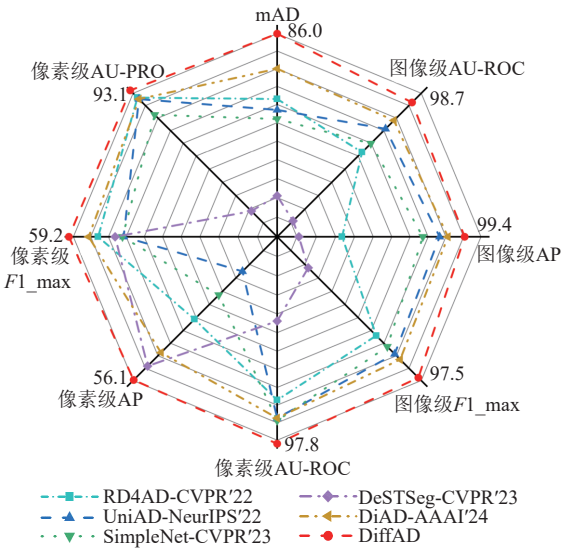


Fig. 3 Radar chart for performance comparison

图 3 性能对比雷达图

当前模型在 Uni-Medical 医学异常检测数据集上仍存在提升空间。该数据集包含多种具有挑战性的模态，存在病变区域与正常组织灰度重叠、噪声干扰显著等问题。DiffAD 通过差分卷积放大梯度信息，能够有效定位微小病变区域，在该数据集的各项评估指标上均实现了显著提升，验证了其在医学异常检测任务中的应用潜力。

在表 2 中，我们对比了各类模型的参数量与计算量。其中，DiAD 因涵盖扩散模型的完整流程（包括多步采样与去噪迭代过程），导致其参数量与计算量

Table 2 Comparison of Parameter Quantity and Computational Power

表 2 参数量和算力对比

模型	参数量	计算量/GFLOPS	mAD/%
RD4AD	80.6×10^6	28.4	82.3
UniAD	$\mathbf{24.5\times10^6}$	**3.6**	81.7
SimpleNet	72.8×10^6	16.1	81.2
DeSTSeg	35.2×10^6	122.7	77.1
DiAD	$1\,331.3\times10^6$	451.5	84.0
Omni-AD	29.9×10^6	11.2	85.8
DiffAD（本文）	35.8×10^6	11.3	**86.0**

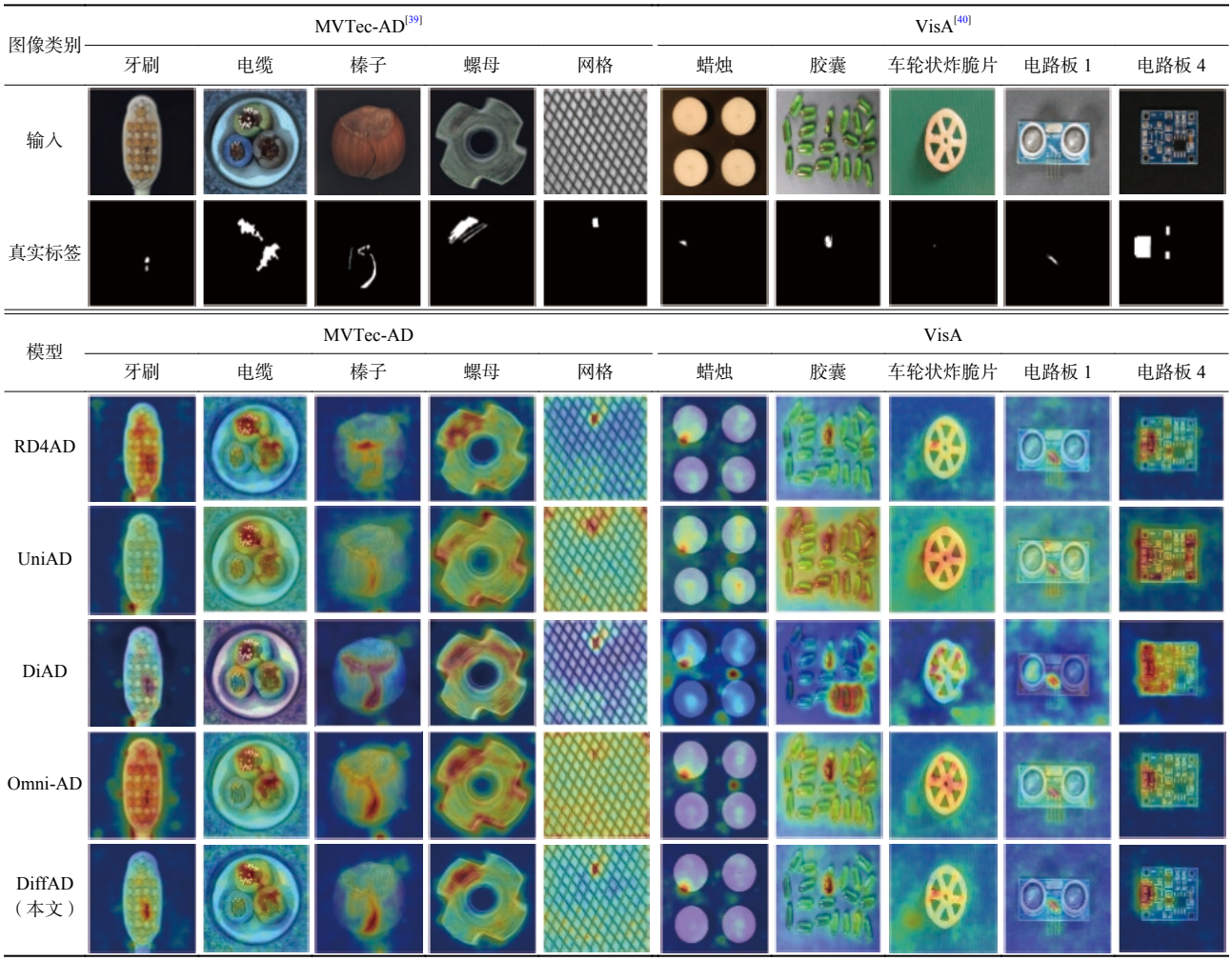
注：黑体数值表示最优值。

显著偏高；而实验结果表明，DiffAD 在参数量控制与运算效率优化 2 方面均表现十分优异，且同时实现了卓越的性能水平。

2)定性对比。我们在 MVTec-AD 和 VisA 数据集上进行了定性实验，表 3 表明，DiffAD 具有更精确的异常分割能力，证实了其在异常分割方面的准确性。与 DiAD 相比，DiffAD 能够更准确、更紧凑地定位异常区域，在使用具有复杂结构的 Cable 和 Pcb 物体的示例时，DiffAD 不仅能更精准地分割异常区域，还能有效降低误报率。

Table 3 Visual Comparison of Localization Effectiveness

表 3 定位效果可视化对比



3.3 消融研究

所有消融实验均在 MVTec-AD 数据集上完成，除非特别说明，所有实验设置均和 3.1 节中的实验细节保持一致。

1)不同预训练骨干和解码器深度的有效性比较。在研究中，我们比较了 2 个预训练特征提取网络，以了解它们在特定任务中的表现，重点关注了解码器的深度配置以及 GLA 模块中全局块 N 的数量，具体如表 4 所示。ResNet18 参数量与计算量较小，但同解码器配置下性能欠佳，仅当解码器深度设为 [2, 3, 4, 2] 时，其与 ResNet34 的性能差距缩小，不过 ResNet18 在其他数据集上表现未达预期，存在任务局限性。不同层级特征的信息密度与冗余度差异显著，我们采用“浅层深堆、深层浅堆”的非对称解码器策略。编码器早期的浅层特征对应 DADE 模块堆叠 4 层以挖掘细粒度线索(通道数少，故开销可控)，编码器后期的深层特征则压缩 DADE 层数，避免通道数增长引发的参数与计算成本激增。单纯增加网络深度和 N 的数量并未能显著提升模型性能，甚至在一些情况下导致了性能的下降。这表明在深度学习网络设计中，增加参数量并不能确保性能的提升，反而可能引入复杂性和过拟合问题。经过全面的比较和分析，

Table 4 Ablation Study on Different Pre-Trained Backbones and Decoder Depths

表 4 不同预训练骨干和解码器深度的消融研究

骨干网络	解码器深度	N	参数量	计算量/GFLOPS	mAD/%
Resnet18	[1, 1, 1, 1]	2	17.23×10^6	5.72	82.2
	[2, 2, 2, 2]	2	27.44×10^6	7.72	84.9
	[2, 2, 2, 2]	3	31.65×10^6	8.53	85.1
	[2, 3, 4, 2]	2	30.37×10^6	9.21	85.7
	[2, 3, 4, 2]	3	35.15×10^6	10.32	85.7
	[3, 4, 6, 3]	1	34.12×10^6	10.09	84.9
	[3, 4, 6, 3]	2	41.09×10^6	11.71	82.5
	[3, 4, 6, 3]	3	48.07×10^6	13.33	84.9
Resnet34	[1, 1, 1, 1]	2	22.61×10^6	7.84	81.2
	[2, 2, 2, 2]	2	32.85×10^6	9.84	85.3
	[2, 2, 2, 2]	3	37.04×10^6	10.65	83.1
	[2, 3, 4, 2]	2	35.76×10^6	11.33	**86.0**
	[2, 3, 4, 2]	3	40.54×10^6	12.44	85.6
	[3, 4, 6, 3]	1	39.50×10^6	12.21	84.8
	[3, 4, 6, 3]	2	46.48×10^6	13.83	85.6
	[3, 4, 6, 3]	3	53.45×10^6	15.45	85.1

注：黑体数值表示最优结果，解码器深度代表 4 组 DADE 堆叠情况（如图 2 所示）。

最终决定选择 ResNet34 作为主干网络，并将解码器深度设置为 [2, 3, 4, 2]，N设置为 2。

实际应用场景中，可考虑根据数据特性动态配置 DADE 深度，对于细节丰富的细粒度异常（如微小划痕），可适当增加浅层 DADE 堆叠数以增强局部重建能力，对于大范围语义异常（如结构缺损），则可加强深层 DADE。

2）GLA 中局部块卷积不同尺寸组合的有效性比较。为了更好地捕捉局部语义信息，我们在 GLA 模块中并行引入了局部深度卷积。具体做法是将这些深度卷积的输出与 DiffAttn 的输出进行拼接，然后再通过一个 1×1 卷积进行调制。此方法使得模型能够更有效地整合不同来源的特征，从而提升重建质量。如表 5 所示，我们对这部分的局部卷积进行了全面的组合设计，尝试了不同核大小（包括 3, 5, 7）的深度卷积。我们的实验表明，使用较小的卷积核（例如 3）时，会导致性能下降，尤其是在没有 1×1 卷积包裹进行通道调制和特征交互的情况下，模型的性能明显下滑。通过对不同卷积配置的实验进行对比，我们发现，使用 5 和 7 的深度卷积组合，并在其前后添加 1×1 卷积的方式，能够更好地提升模型性能。这种设计不仅增强了对局部语义信息的捕捉能力，同时

Table 5 Ablation Study on Different Size Combinations of Local Block Convolutions in GLA

表 5 GLA 中局部块卷积不同尺寸组合的消融研究 %

局部块卷积尺寸组合	图像级			像素级				mAD
	AU-ROC	AP	$F1_max$	AU-ROC	AP	$F1_max$	AU-PRO	
[1, 3]	97.4	98.9	97.0	97.7	53.3	57.4	92.7	84.9
[1, 5]	97.5	99.1	97.0	97.6	55.1	58.1	92.6	85.3
[1, 3, 5]	97.4	98.9	97.1	**97.8**	54.9	58.9	92.8	85.4
[1, 5, 7]	**98.7**	**99.4**	97.5	**97.8**	**56.1**	**59.2**	**93.1**	**86.0**
[1, 3, 5, 7]	98.1	99.3	**97.6**	**97.8**	55.7	59.1	92.9	85.8
[3, 5]	97.5	99.1	96.7	97.5	54.5	58.0	92.3	85.1
[5, 7]	97.5	98.8	97.1	97.6	55.4	57.1	92.3	85.1
[3, 5, 7]	97.9	99.2	97.0	97.5	55.3	58.3	92.6	85.4

注：最好结果用黑体数值表示。“1”代表普通的 1×1 卷积，分别包裹在深度卷积的前后。

也优化了特征融合和通道调制，最终在 GLA 模块的局部分支中取得了显著的效果提升，为进一步提升解码器的性能奠定了坚实的基础。

3）不同组件有效性研究。在 DADE 解码器中，我们大量使用了差分卷积，以增强梯度并提高重建质量。为深入研究差分卷积在这一过程的有效性，我们进行了对比实验，具体做法是将 DiffAttn 和 DE 中的差分卷积替换为普通卷积，这一替换使我们能够更清晰地观察到差分卷积对解码性能的影响。正如前文所述，每个解码阶段都是一个循序渐进的过程。在这一过程中，空间注意力（SA）和通道注意力（CA）起到了至关重要的作用，它们分别在混沌阶段和细化阶段之间实现了有效的过渡。空间注意力帮助网络更好地聚焦于输入特征中的重要区域，而通道注意力则增强了对不同通道特征的敏感性，从而为重建质量提供了保障。为验证对含高频噪声图像的泛化性，我们在训练中添加高斯（标准差为 0.01~0.1）、椒盐噪声（密度为 0.01~0.05）做数据增强，并统计样本重建前后特征 2 范数均值以衡量重建质量。实验结果如表 6 所示，差分卷积在提升梯度传递和细化重建细节方面表现优越，而空间和通道注意力的结合则进一步提升了重建效果。

为了验证重建组件在训练中的稳定性，针对表 6 中所列的不同情况开展了 200 轮的训练实验，并以每 25 轮为 1 个统计周期进行数据收集与分析。鉴于较小的轮次间隔更易导致效果出现波动，故选择此统计频率以便精准捕捉训练动态。实验结果如图 4 所示（为使图示表达更简洁清晰，选择情况 1~6 进行绘制展示），DiffAD 的最优组件设计（情况 1）在除像

Table 6 Ablation Study on the Effectiveness of Different Components

表 6 不同组件有效性的消融研究

情况	CA	SA	DiffAttn	DE	mAD/%	L2 距离	变量说明
1	✓	✓	差分卷积	差分卷积	**86.0**	**9.443 1±0.926 8**	DADE 全组件
2	✗	✓	差分卷积	差分卷积	85.5	9.819 8±0.978 8	仅移除 CA（贡献下降 0.5%）
3	✓	✗	差分卷积	差分卷积	85.5	9.773 4±0.990 2	仅移除 SA（贡献下降 0.5%）
4	✗	✗	差分卷积	差分卷积	85.3	10.003 1±0.997 3	移除 CA+SA（贡献下降 0.7%）
5	✓	✓	普通卷积	差分卷积	85.1	10.225 4±1.020 9	DiffAttn 换普通（贡献下降 0.9%）
6	✓	✓	差分卷积	普通卷积	85.2	10.300 8±1.026 7	DE 换普通（贡献下降 0.8%）
7	✓	✓	普通卷积	普通卷积	84.4	11.543 5±1.836 5	双模块换普通（贡献下降 1.6%）

注：黑体数值表示最优结果。“±”前数值表示平均值；“±”后数值表示标准差。

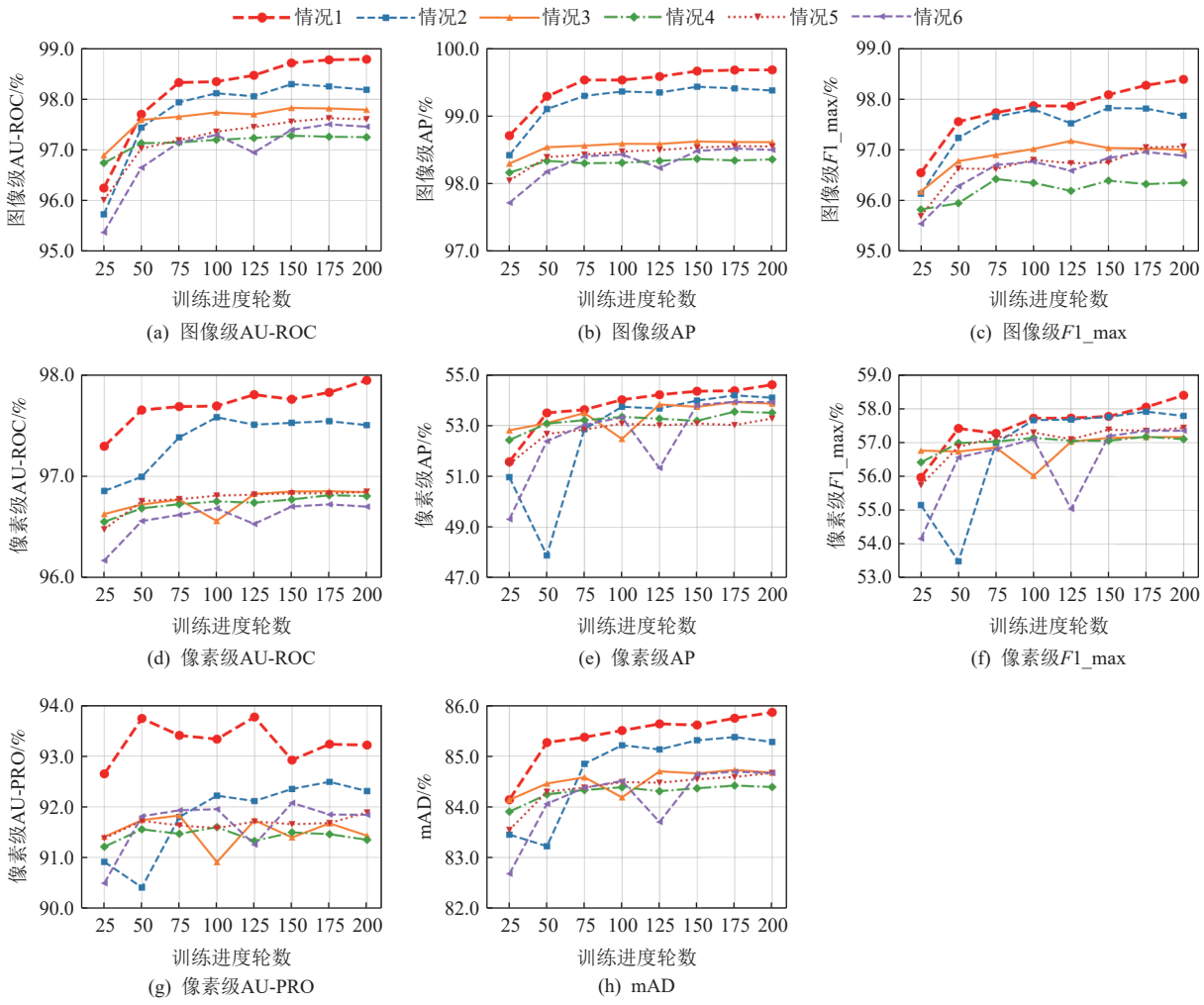


Fig. 4 Comparison of training stability among different components

图 4 不同组件训练稳定性的对比

素级 AU-PRO 指标之外的其他各项指标上，均呈现出稳定且逐步上升的良好态势，这表明该组件设计在训练过程中具备较高的稳定性和有效性。与此同时，对于完全未采用差分卷积的组件设计（情况 7），其训练效果较为不理想，各项指标上升较为缓慢且不稳定，这充分凸显了差分卷积在提升训练效果方面的重要作用。

如表 7 所示，我们对 DiffAttn 进行了多项替换实验，包括与 Transformer 及其变体（ViT[19]，Swin ViT[43]，MAViT[44]，Differential Transformer[45]）、可变形注意力

Table 7 Performance Comparison of DiffAttn and Different Attention Mechanisms

表 7 DiffAttn 和不同注意力机制的性能表现对比

注意力机制	类型	参数量	计算量/GFLOPS	mAD/%
ViT	Transformer	35.9×10^6	10.2	84.6
Swin ViT	Transformer	36.0×10^6	10.3	79.9
DCNv3	可变形	**33.3×10^6**	**9.8**	82.4
Vmamba	Mamba	40.0×10^6	11.2	84.8
MAViT	线性	37.7×10^6	10.8	84.6
Differential Transformer	Transformer	36.1×10^6	11.5	85.1
DiffAttn（本文）	Transformer	35.8×10^6	11.3	**86.0**

注：黑体数值表示最优结果。

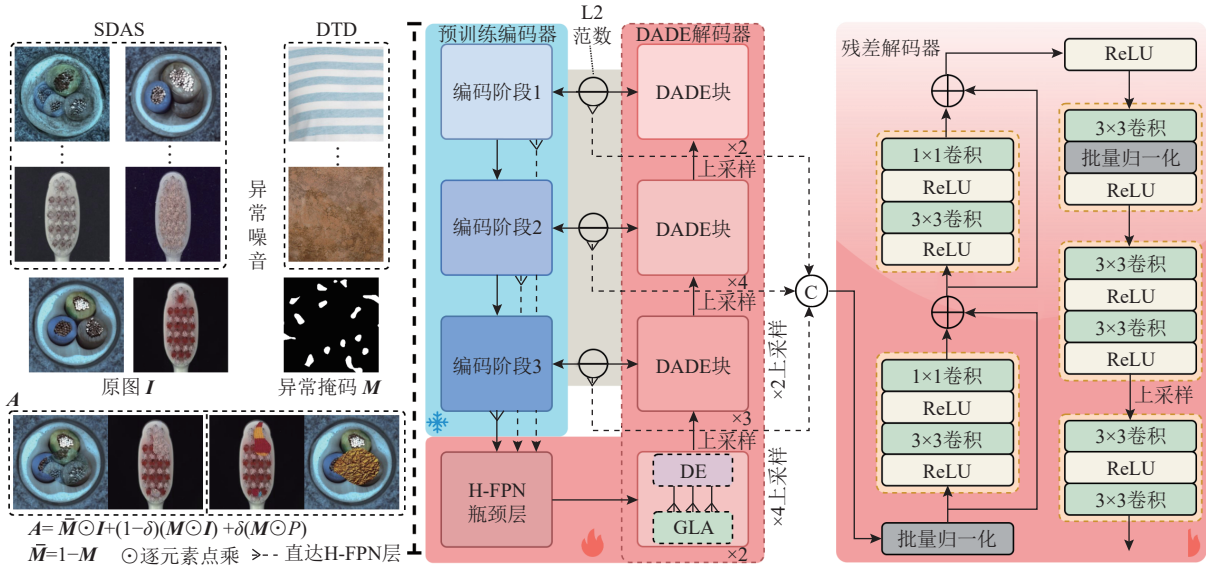
(DCNv3[20])以及 Mamba(Vmamba[21])等模块对比，并在前后包裹$ConvB_{1\times1}(\cdot)$来保持与局部分支的一致。结果表明，在当前架构设计中，DiffAttn 取得了最优性能。与现有差分注意力研究(如 Differential Transformer)相比，DiffAttn 的核心在于通过差分卷积增强细节感知，而 Differential Transformer 主要通过注意力差分优化全局依赖，二者的技术路径不同但共同印证了差分思想的有效性。未来可探索将局部细节增强与全局噪声抑制相融合，形成双重优势。如何将这些先进注意力机制更有效地与我们的框架融合，仍需进一步探索与研究。

3.4 引入异常合成策略

1)网络更新和损失函数变化。基于数据增强的模型通过在无异常图像中合成缺陷并完成自监督学习，以增强网络主动修复异常区域的能力，该策略已在诸多研究中[8,12,14]被证明是有效的。实验中我们引入了应用最为广泛的 DTD[31]和最先进的 SDAS[12]的异常合成策略，并与完全不引入数据增强的情况进行对比。由于手工制作了缺陷标注，可以完成自监督学习，因此在原有 DiffAD 基础上，将重建特征和输入特征的残差输入到一个解码器中构成 DiffAD-RD，如图 5 所示。我们使用交叉熵损失函数$\mathcal{L}_{seg}(\boldsymbol{A},\boldsymbol{M})$来监督残差解码器的训练，DiffAD-RD 总体损失函数为

$$\mathcal{L}=\mathcal{L}_{rec}+\mathcal{L}_{seg}\ 。\tag{11}$$

2)前后性能对比。如表 8 和表 9 所示，2 种缺陷合成方法显著提升了模型像素级检测性能，优化了缺陷分割边界精度并集中了预测置信度，验证了模拟异常生成策略对边界优化与置信度校准的有效性。但该策略仍存在局限性，Cable 等类别出现的高置信度伪阳性预测直接导致 AU-PRO 指标明显下滑。核心原因在于自监督学习依赖的模拟异常(含噪声注入与异常掩码生成)，其分布难以完全匹配真实场景中异常的多样化形态与纹理特征，工业缺陷产生机制复杂且不确定，模拟异常无法穷尽所有真实异常特征。尽管该策略强化了异常与正常样本的区分度，使预测边界更锐利，但易导致模型对正常区域过度敏感，产生高置信度伪阳性预测，这类区域会破坏预测与真实异常的区域重叠一致性，而 AU-PRO 指标对区域匹配度高度敏感，最终引发指标下滑。在 mAD 方面，基于 DTD 增强的方法取得 88.3% 的算术



注：通过左侧的异常合成策略，制作的标签可以完成端到端的像素级监督训练。

Fig. 5 DiffAD-RD with a residual decoder

图 5 带有残差解码器的 DiffAD-RD

Table 8 Analysis of Results Under Different Application Proportions of Synthesis Strategies

表 8 不同合成策略应用比例下的结果分析 %

比例	图像级			像素级				mAD
	AU-ROC	AP	$F1_max$	AU-ROC	AP	$F1_max$	AU-PRO	
[1, 0, 0]	**98.7**	**99.4**	**97.5**	97.8	56.1	59.2	**93.1**	86.0
[1, 1, 0]	95.8	98.3	96.0	97.9	69.2	65.3	78.4	85.8
[1, 0, 1]	98.0	99.3	97.2	98.5	73.8	69.4	81.6	**88.3**
[2, 1, 1]	97.1	98.9	97.1	**98.8**	**75.3**	**71.2**	73.9	87.4

注：黑体数值表示最优结果。括号内数值表示正常、SDAS、DTD 的比例。

平均，较原始 DiffAD 实现 2.3% 的性能增益。需说明的是，该增强策略使模型参数量（提升约 4.6×10^6）与计算量（提升约 18.7 GFLOPS），为了确保对比的公平性，主体实验仅聚焦原始 DiffAD。

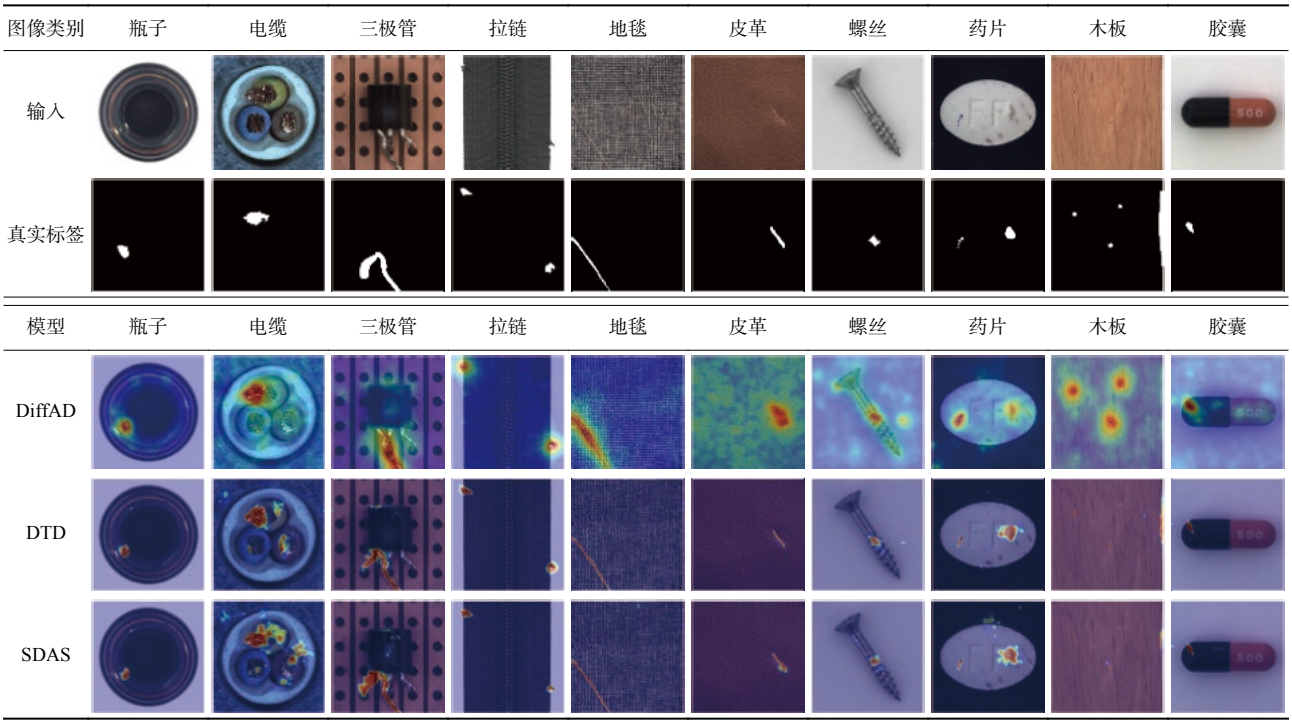
3.5 单一类别设置实验结果

我们评估了 DiffAD 在 MVTec-AD 和 VisA 数据集上的单一类别无监督异常检测性能，使用图像级和像素级上的 AU-ROC 这 2 个指标，并将其与其他先进模型进行比较，如表 10 和表 11 所示，DiffAD 在单一类别设置下的表现尤为突出，其性能相较于其他先进模型呈现出显著的优势，但是也面临一个类别训练一个单独模型的缺点。

在训练开销方面，单类模型与多类模型的训练耗时差异较小，其核心影响因素为训练数据集的规模。而在部署与推理阶段，2 类模型的开销差异显著，在单线程处理批量推理任务的场景下（目标硬件环境固定），如图 6 所示，不同模型部署时的内存与显存占用量存在显著差异（5 次实验平均结果）；在推理耗时维度，多类模型支持批量推理模式可一次性完成对整批样本的处理，而单类模型需以串行方式逐一处理样本，且需额外引入人工干预或前置分类模型，以筛选适配当前样本类别的目标推理模型。

Table 9 Comparison of Fully Unsupervised and Self-supervised Models on Localization Effectiveness

表 9 完全无监督模型和自监督模型的定位效果对比



4 结 论

本研究提出了一种名为 DiffAD 的多类无监督异常检测模型。该模型创新性地将重建过程细分为多个阶段，旨在通过渐进式特征重建策略解决复杂工业场景中的视觉异常检测问题。DiffAD 的核心技术在于融合 Transformer 与卷积神经网络的优势，设计了一种名为 DADE 的渐进式重建组件。该组件通过差分卷积增强特征表达能力，并结合 CA 机制和 SA 机制优化特征融合与细节重建。实验结果表明，DiffAD 在多个具有挑战性的数据集上取得了优异的性能，同时保持了较低的模型参数量和计算量，展现出良好的泛化能力与实用性。

Table 10 Comparative Results of AU-ROC for DiffAD on MVTec-AD Dataset Under Single-Class Setting

表 10 单一类别设置下 DiffAD 在 MVTec-AD 数据集上的 AU-ROC 对比结果 %

类别		FYD[6]		DRAEM[8]		CutPaste[7]		P-SVDD[11]		DiffAD（本文）	
		图像级	像素级	图像级	像素级	图像级	像素级	图像级	像素级	图像级	像素级
纹理	地毯	98.8	98.5	97.0	95.5	92.9	92.6	93.1	98.3	**99.88**	**99.16**
	网格	98.9	96.8	99.9	**99.7**	94.6	96.2	99.9	97.5	**100.00**	98.96
	皮革	**100.0**	**99.2**	**100.0**	98.6	90.9	97.4	**100.0**	99.5	**100.00**	99.18
	地砖	98.8	96.8	99.6	**99.2**	97.8	91.4	93.4	90.5	**99.89**	96.91
	木板	**99.4**	**99.6**	99.1	96.4	96.5	90.8	98.6	95.5	99.21	96.58
	平均	99.2	98.2	99.1	97.9	94.5	93.7	97.0	96.3	**99.80**	98.16
对象	瓶子	**100.0**	98.3	99.2	**99.1**	98.6	98.1	98.3	97.6	99.68	98.36
	电缆	95.3	97.5	91.8	94.7	90.3	96.8	80.6	90.0	**98.54**	**97.97**
	胶囊	92.5	98.6	**98.5**	94.3	76.7	95.8	96.2	97.4	94.70	**98.69**
	榛子	99.9	98.7	**100.0**	**99.7**	92.0	97.5	97.3	97.3	**100.00**	99.01
	螺母	**99.9**	98.2	98.7	**99.5**	94.0	98.0	99.3	93.1	98.05	97.08
	药片	94.5	97.3	**98.9**	**97.6**	86.1	95.1	92.4	95.7	97.00	97.54
	螺丝	90.1	98.7	93.9	97.6	81.3	95.7	86.3	96.7	**95.06**	**99.33**
	牙刷	**100.0**	98.9	**100.0**	98.1	**100.0**	98.1	98.3	98.1	96.78	**98.94**
	三极管	99.2	**98.1**	93.1	90.9	91.5	97.0	95.5	93.0	**99.63**	97.84
	拉链	97.5	98.2	**100.0**	98.8	97.9	95.1	99.4	**99.3**	95.12	98.57
	平均	97.0	98.3	97.4	97.0	90.8	96.7	94.3	95.8	**97.45**	**98.33**
总平均值		97.7	98.2	98.0	97.3	95.2	96.0	92.1	95.7	**98.24**	**98.27**

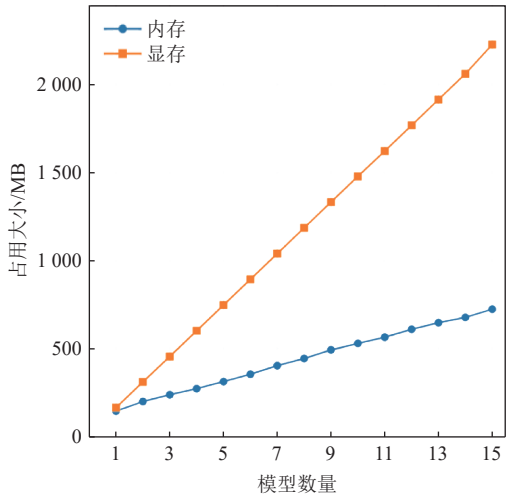
注：黑体数值表示最优结果。

Table 11 Comparative Results of AU-ROC for DiffAD on VisA Dataset Under Single-Class Setting

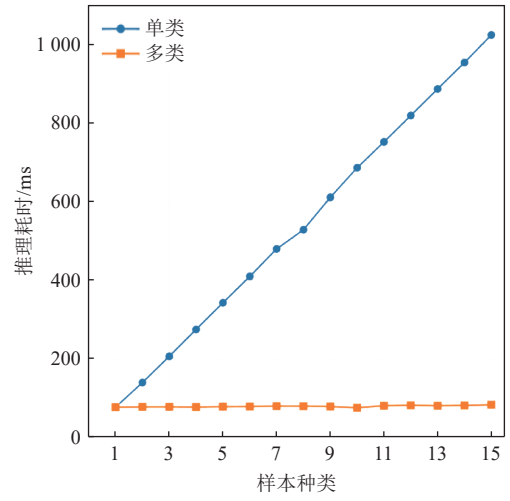
表 11 单一类别设置下 DiffAD 在 VisA 数据集上的 AU-ROC 对比结果 %

类别	SPADE[10]		DRAEM[8]		FAPM[9]		DiffAD（本文）	
	图像级	像素级	图像级	像素级	图像级	像素级	图像级	像素级
蜡烛	91.0	97.9	91.8	96.6	92.8	94.9	**92.87**	**99.00**
胶囊	61.4	60.7	74.7	98.5	71.2	75.3	**88.62**	**99.29**
腰果	**97.8**	86.4	95.1	83.5	91.0	91.4	95.86	**98.43**
口香糖	85.8	98.6	94.8	96.8	91.4	98.6	**99.35**	**98.77**
车轮状炸脆片	88.6	96.7	**97.4**	87.2	88.6	97.3	94.42	**94.45**
通心面 1	95.2	96.2	97.2	99.9	**98.3**	97.3	90.52	**99.68**
通心面 2	**87.9**	87.5	85.0	99.2	86.3	89.2	87.28	**99.59**
电路板 1	72.1	66.9	47.6	88.7	77.4	75.2	**93.29**	**99.75**
电路板 2	50.7	71.1	89.8	91.3	61.9	67.3	**96.25**	**98.86**
电路板 3	90.5	95.1	92.0	98.0	74.3	94.8	**97.27**	**99.16**
电路板 4	83.1	89.0	98.6	96.8	80.9	89.9	**99.38**	**97.47**
管状炸脆片	81.1	81.8	**100.0**	85.8	72.0	87.3	99.34	**99.29**
平均	82.1	85.6	88.7	93.5	82.2	88.2	**94.54**	**98.65**

注：用黑体数值表示最优结果。



(a) 内存和显存占用情况



(b) 推理耗时情况

Fig. 6 Comparison of resource consumption and inference time-consuming of the model in single-class and multi-class scenarios

图 6 模型在单类和多类场景下的资源占用与推理耗时对比

作者贡献声明：姚健和马帅提出论文整体思路、技术架构，并负责技术实现；姚健撰写论文；马帅编写算法和协助开展实验验证；钱鹏江审阅技术创新性和论文结构与内容，提出指导意见；王闯负责部分实验验证、论文图表制作及参考文献整理；杨桂松负责部分对比算法整理和部分论文内容撰写；何源负责论文审阅并提出指导意见。

参 考 文 献

[1] Wang Wenpeng, Qin Yinchang, Shi Wenxuan. A review of unsupervised deep learning methods for industrial defect detection[J]. Journal of Computer Applications, 2025, 45(5): 1658−1670 (in Chinese)
(王文鹏, 秦寅畅, 师文轩. 工业缺陷检测无监督深度学习方法综述[J]. 计算机应用, 2025, 45(5): 1658−1670)

[2] Nie Hui, Wang Ruiping, Chen Xilin. Open world object recognition and detection systems: Landscapes, challenges and prospects[J]. Journal of Computer Research and Development, 2024, 61(9): 2128−2141 (in Chinese)
(聂晖, 王瑞平, 陈熙霖. 开放世界物体识别与检测系统: 现状、挑战与展望[J]. 计算机研究与发展, 2024, 61(9): 2128−2141)

[3] Tabernik D, Šela S, Skvarč J, et al. Segmentation-based deep-learning approach for surface-defect detection[J]. Journal of Intelligent Manufacturing, 2020, 31(3): 759−776

[4] You Zhiyuan, Cui Lei, Shen Yujun, et al. A unified model for multi-class anomaly detection[J]. Advances in Neural Information Processing Systems, 2022, 35: 4571−4584

[5] He Haoyang, Bai Yuhu, Zhang Jiangning, et al. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 71162−71187

[6] Zheng Ye, Wang Xiang, Deng Rui, et al. Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization[C]//Proc of the 23rd IEEE Int Conf on Multimedia and Expo (ICME). Piscataway, NJ: IEEE, 2022: 1−6

[7] Li Chunliang, Sohn K, Yoon J, et al. CutPaste: Self-supervised learning for anomaly detection and localization[C]//Proc of the 34th IEEE/CVF Conf on Computer Vision And Pattern Recognition. Piscataway, NJ: IEEE, 2021: 9664−9674

[8] Zavrtanik V, Kristan M, Skočaj D. Draem—A discriminatively trained reconstruction embedding for surface anomaly detection[C]//Proc of the 18th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 8330−8339

[9] Kim D, Park C, Cho S, et al. Fapm: Fast adaptive patch memory for real-time industrial anomaly detection[C]//Proc of the 48th IEEE Int Conf on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ: IEEE, 2023: 1−5

[10] Cohen N, Hoshen Y. Sub-image anomaly detection with deep pyramid correspondences[J]. arXiv preprint, arXiv: 2005.02357, 2020

[11] Yi Jihun, Yoon S. Patch SVDD: Patch-level SVDD for anomaly detection and segmentation[C]//Proc of the 15th Asian Conf on Computer Vision. Berlin: Springer, 2020: 396–412

[12] Zhang Ximiao, Xu Min, Zhou Xiuzhuang. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection[C]//Proc of the 37th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 16699−16708

[13] Schlüter H M, Tan J, Hou B, et al. Natural synthetic anomalies for self-supervised anomaly detection and localization[C]//Proc of the 17th European Conf on Computer Vision. Berlin: Springer, 2022: 474−489

[14] Yang Minghui, Wu Peng, Feng Hui. MemSeg: A semi-supervised method for image surface defect detection using differences and commonalities[J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105835

[15] Zhang Xuan, Li Shiyu, Li Xi, et al. DeSTSeg: Segmentation guided denoising student-teacher for anomaly detection[C]//Proc of the 36th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 3914−3923

[16] Deng Hanqiu, Li Xingyu. Anomaly detection via reverse distillation

from one-class embedding[C]//Proc of the 35th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 9737−9746

[17] Ma Zeqi, Li Jiaxing, Wong W K. Patch distance based auto-encoder for industrial anomaly detection[J]. Expert Systems with Applications, 2025, 270: 126537

[18] Yu Wenyong, Zhang Yang, Yao Haiming, et al. Visual inspection of surface defects based on a lightweight reconstruction network[J]. Acta Automatica Sinica, 2022, 48(9): 2175−2186 (in Chinese)
(余文勇, 张阳, 姚海明, 等. 基于轻量化重构网络的表面缺陷视觉检测[J]. 自动化学报, 2022, 48(9): 2175−2186)

[19] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proc of the 31st Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates, Inc, 2017: 6000−6010

[20] Wang Wenhai, Dai Jifeng, Chen Zhe, et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions[C]//Proc of the 36th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 14408−14419

[21] Liu Yue, Tian Yunjie, Zhao Yuzhong, et al. Vmamba: Visual state space model[J]. Advances in Neural Information Processing Systems, 2024, 37: 103031−103063

[22] Quan Jiajie, Tong Ao, Cai Yuxuan, et al. Omni-AD: Learning to reconstruct global and local features for multi-class anomaly detection[J]. arXiv preprint, arXiv: 2503.21125, 2025

[23] Chen Zixuan, He Zewei, Lu Zheming. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 33: 1002−1015

[24] Cheng Yudong, Zhou Fang. Semi-supervised learning based method for unknown anomaly detection[J]. Journal of Computer Research and Development, 2024, 61(7): 1670−1680 (in Chinese)
(程渝栋, 周昉. 基于半监督学习的未知异常检测方法[J]. 计算机研究与发展, 2024, 61(7): 1670−1680)

[25] Liu Tongkun, Li Bing, Zhao Zhuo, et al. Reconstruction from edge image combined with color and gradient difference for industrial surface anomaly detection[J]. arXiv preprint, arXiv: 2210.14485, 2022

[26] Zavrtanik V, Kristan M, Skočaj D. Reconstruction by inpainting for visual anomaly detection[J]. Pattern Recognition, 2021, 112: 107706

[27] Song J, Kong K, Park Y I, et al. AnoSeg: Anomaly segmentation network using self-supervised learning[J]. arXiv preprint, arXiv: 2110.03396, 2021

[28] Zhang Jiangning, Chen Xuhai, Wang Yabiao, et al. Exploring plain ViT reconstruction for multi-class unsupervised anomaly detection[J]. Computer Vision and Image Understanding, 2025, 253: 104308

[29] Xing Peng, Jiang Xin, Pan Yonghua, et al. Visual anomaly detection based on feature-constrained distillation learning[J]. Journal of Software, 2023, 34(9): 4378−4391 (in Chinese)
(邢鹏, 蒋鑫, 潘永华, 等. 基于特征约束蒸馏学习的视觉异常检测. 软件学报, 2023, 34(9): 4378−4391)

[30] He Haoyang, Zhang Jiangning, Chen Hongxu, et al. A diffusion-based framework for multi-class anomaly detection[C]//Proc of the 38th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2024: 8472−8480

[31] Cimpoi M, Maji S, Kokkinos I, et al. Describing textures in the wild[C]//Proc of the 27th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 3606−3613

[32] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces[J]. arXiv preprint, arXiv: 2312.00752, 2023

[33] Heckler L, König R, Bergmann P. Exploring the importance of pretrained feature extractors for unsupervised anomaly detection and localization[C]//Proc of the 36th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 2917−2926

[34] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971−987

[35] Yu Zitong, Zhao Chenxu, Wang Zezheng, et al. Searching central difference convolutional networks for face anti-spoofing[C]//Proc of the 33rd IEEE/CVF Conf on Computer Vision And Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5295−5305

[36] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//Proc of the 29th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770−778

[37] Deng Jia, Dong Wei, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]//Proc of the 22nd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 248−255

[38] Hou Qibin, Cheng Mingming, Hu Xiaowei, et al. Deeply supervised salient object detection with short connections[C]//Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 3203−3212

[39] Bergmann P, Fauser M, Sattlegger D, et al. MVTec-AD —A comprehensive real-world dataset for unsupervised anomaly detection[C]//Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 9592−9600

[40] Zou Yang, Jeong J, Pemula L, et al. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation[C]//Proc of the 17th European Conf on Computer Vision. Berlin: Springer, 2022: 392−408

[41] Bergmann P, Jin Xin, Sattlegger D, et al. The MVTec 3D-AD dataset for unsupervised 3D anomaly detection and localization[J]. arXiv preprint, arXiv: 2112.09045, 2021

[42] Liu Zhikang, Zhou Yiming, Xu Yuansheng, et al. SimpleNet: A simple network for image anomaly detection and localization[C]//Proc of the 36th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 20402−20411

[43] Liu Ze, Lin Yutong, Cao Yue, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows[C]//Proc of the 18th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 10012−10022

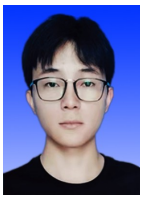
[44] Fan Qihang, Huang Huaibo, Ai Yuang, et al. Rectifying magnitude neglect in linear attention[C]//Proc of the 20th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2025: 21505−21514

[45] Ye Tianzhu, Dong Li, Xia Yuqing, et al. Differential Transformer[J]. arXiv preprint, arXiv: 2410.05258, 2024



Yao Jian, born in 1984. Master, senior engineer. His main research interests include lightweight edge learning, explainable artificial intelligence, and smart healthcare.

姚　健, 1984 年生。硕士，高级工程师。主要研究方向为轻量化边端学习、可解释人工智能、智能康养。



Ma Shuai, born in 2001. Bachelor. His main research interests include deep learning and product defect detection.

马　帅, 2001 年生。学士。主要研究方向为深度学习、工业产品缺陷检测。



Qian Pengjiang, born in 1979. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include machine learning, deep learning, intelligent healthcare, and smart city.

钱鹏江, 1979 年生。博士，教授，博士生导师。CCF 高级会员。主要研究方向为机器学习、深度学习、智能康养、智慧城市。



Wang Chuang, born in 1998. Master. His main research interests include machine learning, deep learning, and explainable artificial intelligence.

王　闯, 1998 年生。硕士。主要研究方向为机器学习、深度学习、可解释人工智能。



Yang Guisong, born in 1982. PhD, associate professor, master supervisor. His main research interests include internet of things and ubiquitous computing, mobile crowd-sensing, and edge computing.

杨桂松, 1982 年生。博士，副教授，硕士生导师。主要研究方向为物联网与普适计算、移动群智感知、边缘计算。



He Yuan, born in 1980. PhD, professor, PhD supervisor. Distinguished member of CCF. His main research interests include internet of things, wireless networks, and mobile and ubiquitous computing.

何　源, 1980 年生。博士，教授，博士生导师。CCF 杰出会员。主要研究方向为物联网、无线网络、移动和普适计算。