医学影像无监督异常检测技术综述

赵映程[1,2,3]，宋霄罡[1,2,3]，石争浩[1,2,3]，尤珍臻[1,2,3]，黑新宏[1,2,3]
（1.西安理工大学计算机科学与工程学院，陕西 西安 710048；2.陕西省网络计算与安全技术重点实验室，陕西 西安 710048；
3.人机共融智能机器人陕西省高校工程研究中心，陕西 西安 710048）

摘　要：无监督异常检测技术能够在仅使用正常样本训练的前提下识别偏离分布的异常样本，因此在医学影像领域中能够用于病变的自动检测与区域定位，对辅助疾病筛查，尤其是罕见病的诊断具有重要意义。针对医学影像无监督异常检测领域的研究进展，首先，依据核心检测机制将现有异常检测方法划分为4个主要类别，并详细阐述了每类方法的机理、特性及其代表性工作；其次，汇总了常用的医学影像公开数据集，并归纳了主要的异常评分计算方式以及模型评估基准指标；进而，通过在多模态医学影像数据集上的系统性定量实验，对比了各类代表性方法的检测性能与计算效率，并对其结果进行了深入分析；最后，讨论了该领域在检测性能及临床适用性等方面存在的挑战，并对未来的研究方向进行了展望。

关键词：医学影像分析；异常检测；深度学习；无监督学习
中图分类号：TP391.41
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2026046

Review of unsupervised anomaly detection techniques for medical imaging

Zhao Yingcheng[1,2,3], Song Xiaogang[1,2,3], Shi Zhenghao[1,2,3], You Zhenzhen[1,2,3], Hei Xinhong[1,2,3]
1. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China
2. Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an 710048, China
3. Human Machine Integration Intelligent Robot Shaanxi Provincial University Engineering Research Center, Xi'an 710048, China

Abstract: Unsupervised anomaly detection technology enabled the identification of out-of-distribution anomalous samples using only normal samples for training. Thus, it could be applied in the field of medical imaging for automated lesion detection and localization, playing a significant role in assisting disease screening, particularly in the diagnosis of rare diseases. Research progress on unsupervised anomaly detection in medical imaging. First, existing anomaly detection methods were categorized into four main types based on their core detection mechanisms, and the principles, characteristics, and representative works of each category were elaborated in detail. Second, commonly used public medical imaging datasets were summarized, and major anomaly scoring methods as well as benchmark evaluation metrics were introduced. Furthermore, through systematic quantitative experiments on multi-modal medical image datasets, the detection performance and computational efficiency of various representative methods were compared, and an in-depth analysis was provided based on the results. Finally, challenges in detection performance and clinical applicability were discussed, and future research directions were outlined.

Keywords: medical imaging analysis, anomaly detection, deep learning, unsupervised learning

收稿日期：2025–11–01；修回日期：2026–02–16
通信作者：黑新宏，heixinhong@xaut.edu.cn
基金项目：国家自然科学基金资助项目（No.62502377）；陕西省自然科学基础研究计划基金资助项目（No.2025JC-YBMS-755, No.2025JC-YBQN-934）
Foundation Items：The National Natural Science Foundation of China (No.62502377), The Natural Science Basic Research Program of Shaanxi Province (No.2025JC-YBMS-755, No.2025JC-YBQN-934)

0　引言

异常检测（anomaly detection, AD）是机器学习的重要研究领域之一，其核心目标为通过学习正常样本的模式，以无监督的方式识别和定位偏离该模式的“异常”样本。由于其对先验信息极低的依赖性，该技术在工业缺陷检测[1-2]、时间序列检测[3]、视频监控[4-5]、网络安全检测[6]等异常样本稀缺任务中得到了广泛应用。

目前，影像学检查已成为临床诊疗中的重要辅助手段，为疾病诊断、治疗方案制定和预后评估提供了全周期的决策依据。并且，医学影像技术涵盖多种成像模态，如超声成像（ultrasound, US）、计算机断层扫描（computed tomography, CT）、磁共振成像（magnetic resonance imaging, MRI）等，能够从不同角度直观地展现人体内部结构及其病变情况，为疾病筛查提供多元化的诊断信息。

医学影像分析是寻找病灶与挖掘病理信息的关键环节。然而，传统的人工分析方式往往受制于判读效率和稳定性等问题，常见的有监督机器学习方法虽然在分析速度方面具有显著优势，但其性能高度依赖于大量高质量的标注数据。在临床实践中，高质量医学影像数据的收集和标注通常是一项艰巨的任务，尤其对于罕见病而言，难度更为突出。此外，疾病类型复杂、表现多样，构建能够覆盖所有可能情形的疾病信息库也难以实现。相比之下，在多数医学影像检查中，正常（健康）受试者的数量通常远多于异常（患病）受试者，因此正常图像的获取相对容易，这一数据分布特点使不依赖于标注数据的无监督异常检测（unsupervised anomaly detection, UAD）技术在医学影像分析领域展现出显著的优势。

基于UAD的医学影像分析旨在利用大量易得的正常样本进行模型训练，学习健康影像的分布特征。当输入新的影像样本时，模型会判断其与“正常”模式之间的差异，并将具有显著偏差的样本判定为“异常”。通过该方式，UAD模型便能够从影像中检测出组织病变或解剖结构形变等多种类型的异常，从而为健康评估、疾病识别、病灶定位等任务提供辅助。

相较于工业品图像、监控视频等场景，医学影像具有异常边界模糊、数据复杂性高且维度大、图像风格多变、个体差异性强等特性。因此，UAD技术在医学影像分析中面临的核心挑战在于：如何从无标签数据中学习稳健的正常模式，如何避免将正常组织差异误判为异常，以及如何捕捉早期病变等情形下的细微异常。本文旨在通过对医学影像UAD任务的研究现状进行全面的整理与分析，为上述挑战的解决思路提供翔实可靠的参考。

本文首先根据现有检测方法的原理类别论述各方法的核心思想及其特性。随后详细介绍了常见的医学影像病理数据集、异常评分计算方式及异常检测任务的定量指标，作为对方法性能的评价基准。最后对该领域的现存挑战及未来探索方向进行了讨论。

1　无监督异常检测方法综述

UAD任务可以被描述为一个离群值检测问题，多数情形下的目标为对样本进行单类别分类（one-class classification, OCC）[7]，以区分规范分布中的内样本和外样本。在医学影像病理检测背景下，一组样本通常可指代整张图像或图像中的单个像素（体素），健康样本由“正常”分布构成，而包含病灶的样本则由于偏离该正常分布而被检测为离群值。具体而言，给定一个包含N个正常样本的正常训练集$D_{\text{train}}=\{\boldsymbol{x}_i\}_{i=1}^{N}$，目标是学习一个由输入样本$\boldsymbol{x}_i$参数化的异常评分函数$\mathcal{A}(\boldsymbol{x}_i)$，该函数为每组输入样本计算一个分数。该分数应满足对于任何正常样本$\boldsymbol{x}_n$和异常样本$\boldsymbol{x}_a$，有$\mathcal{A}(\boldsymbol{x}_n)<\mathcal{A}(\boldsymbol{x}_a)$。针对不同的任务需求，计算图像级样本的异常评分，可以实现异常分类。而通过逐像素计算像素级样本的异常评分，则可以进一步实现异常分割。

早期的异常检测方法大多基于数据的分布信息或几何性质寻找离群点，如基于统计量的方法[8-9]、基于距离的方法[10-11]、基于密度的方法[12]和基于聚类的方法[13-14]等。近年来，随着深度学习方法的快速发展，最新的医学UAD研究均利用深度神经网络构建异常评分模型，根据评分机制的不同，这些研究可被划分为4个主要类别，如表1所示。

1.1　基于重建的异常检测方法

目前，最主流的一类医学异常检测技术被称为基于重建的方法[29]。一般而言，此类方法通过训练一个重建模型对输入图像进行重建，然后在推理过程中利用重建结果与原始样本间的差异作为异常评分。其核心思想在于仅通过对正常样本进行学

表1　医学影像无监督异常检测方法对比

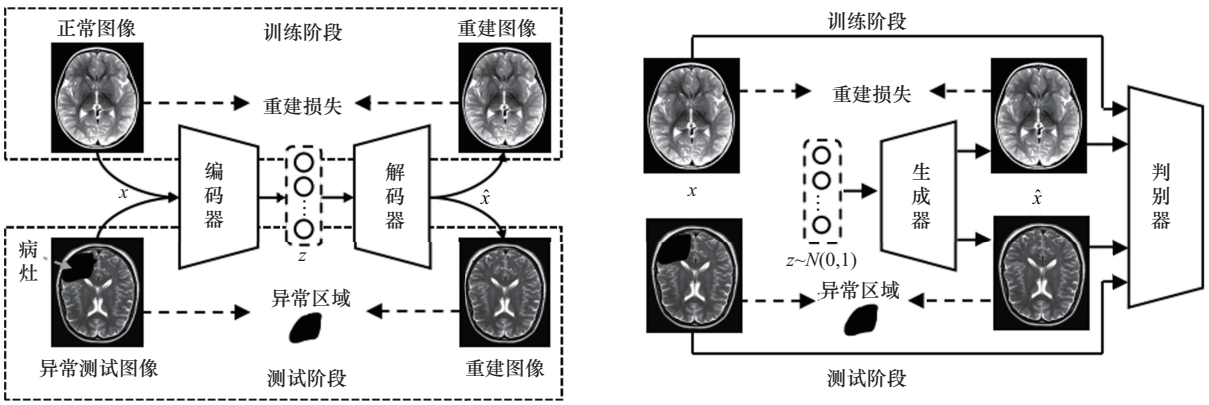
范式	代表性方法	原理	优势	局限性
基于重建的方法	GANomaly[15]、f-AnoGAN[16]、RD4AD[17]、EA2D[18]	训练模型重建正常样本，以重建误差为异常评分	实现简便，训练稳定，可解释性强，能够充分进行多尺度学习	可能出现过度重建问题导致方法失效，对误差度量方法敏感
基于合成的方法	CutPaste[19]、PII[20]、DRAEM[21]、SimpleNet[22]	向正常样本加入人工异常，训练模型区分正常与合成异常	可用于罕见场景的数据集扩充，异常模式可控，可与自监督学习结合进一步提升特征判别能力	受合成策略影响较大，模型泛化性受限
基于自监督学习的方法	Sohn等[23]提出的两阶段框架、SALAD[24]、PMSACL[25]	设计自监督任务学习图像的核心表征，通过测试样本的任务完成能力判定异常	能够深入挖掘正常样本表示、充分学习判别性特征，泛化能力强	关注全局信息，对细微异常不够敏感，难以实现异常定位
基于记忆库匹配的方法	MemAE[26]、PaDiM[27]、P-VQ[28]	存储正常特征原型，通过测试样本匹配度判定异常	原型特征能够进一步凸显异常样本匹配的差异性，有效防止“恒等捷径”出现	时空复杂度过高，缺乏统一的原型特征选取策略

习，以获取通用的特征分布，并基于异常数据难以有效重建的假设，仅学习了“正常”模式的模型将难以准确重建未见过的异常区域，将重建过程中产生高误差的区域判定为病理异常[30]。基于重建的方法可进一步被划分为两种子类：基于数据重建和特征重建的方法。

1.1.1　基于数据重建的方法

基于数据重建的方法主要关注图像空间中的重建结果及异常评分计算过程，如图1所示。此类方法对输入图像执行端到端的重建，并直接利用原始图像与重建结果间的差异来驱动模型训练和基于评分的异常检测。此外，通常采用生成模型作为主干结构，如自编码器（autoencoder, AE）[31]、生成对抗网络（generative adversarial network, GAN）[32]及其变体。

AE能够通过编码器网络对输入图像进行非线性变换以映射到低维流形的潜在空间，并利用解码器网络逐步恢复至原始图像，从而对原始图像进行重建，并根据重建结果的差异程度识别和定位异常。基于这一原则，许多工作探索了基础AE模型及其变体在UAD任务中的性能，如变分AE（variational autoencoder, VAE）[33-34]通过学习数据的概率分布，从而更好地捕捉样本特征。对抗AE（adversarial autoencoder, AAE）[35-36]利用对抗训练迫使编码器输出的潜在空间分布严格匹配预设的先验分布，从而更敏感地识别不符合正常数据分布的异常样本。这些早期方法具有较为简单的模型结构，因而训练速度较快且稳定性高，但缺乏足够的重建约束，导致检测性能欠佳，主要体现在两个方面：一是模型可能具有过强的泛化能力，导致病灶异常被同步重建，这一问题被称为“恒等捷径”，即模型倾向于简单复制输入而非学习有意义的正常模式特征，导致异常区域被同步重建而无法被检测[26]；二是模型判别能力的固有限制可能会将组



(a) 基于自编码器的UAD框架　　(b) 基于生成对抗网络的UAD框架

图1　基于数据重建的异常检测

织间的正常变化误判为病理异常[37]。

为了突破 AE 模型的固有限制，后续研究陆续在重建过程中引入多种优化策略，如探索利用不同的距离函数约束重建过程，包括采用结构相似性（structural similarity, SSIM）指数[38-39]、感知损失（perceptual loss, PL）等[40-41]，旨在通过改善图像空间保真度和感知一致性来提高检测效果。或是构建额外的约束模块，强化重建过程对异常界限的理解能力，如 Gong 等[26]提出了一种记忆增强 AE 模型（MemAE），通过引入记忆模块，基于注意力将编码器输出映射为查询检索记忆矩阵中最相关的项，从而迫使模型记录代表正常数据的原型元素。Zimmerer 等[42]提出了一种结合上下文编码的 VAE（CeVAE），利用掩码重建机制强化模型对正常样本语义信息的学习。Mao 等[43]研究发现，误检测通常出现在边界等高频率区域周围，因此引入不确定性预测自适应地度量边缘变异与真实异常。Bercea 等[41]探索了密集变形场在重建多样性度量中的作用，利用其作为对重建误差的额外约束，用于改善检测准确性。Zhou 等[44]针对视网膜光学相干断层扫描（optical coherence tomography, OCT）影像模态提出了空间上下文变分自编码器（spatial-contextual variational autoencoder, SCVAE）方法，通过在 VAE 的潜在空间中引入空间和上下文约束来放大异常重构误差。

作为另一类主要的生成模型，Schlegl 等[45]最先将 GAN 模型引入医学 UAD 任务，提出了 AnoGAN 方法，该方法采用深度卷积 GAN（deep convolutional GAN, DCGAN）和特征匹配机制，训练生成器来学习正常解剖学变异的流形，并利用判别器评价重建图像与隐空间中正常样本分布的差异。其异常评分由判别器损失和像素级重建损失构成，用于衡量生成与输入图像在低维和高维上的差异。然而，该方法的训练效率较低，且对复杂图像的重建能力有限。为改进这一不足，Baur 等[46]提出了与变分自编码器网络相结合的 AnoVAEGAN 方法，通过强化对全局空间特征的学习来提升模型在高分辨率图像中的检测能力。Schlegl 等[16]针对 AnoGAN 的训练效率问题，在原模型的基础上提出了 f-AnoGAN 方法，该方法采用 Wasserstein GAN[47] 生成更平滑的特征表示，并通过编码器逆映射策略替代原模型的迭代优化过程，从而大幅提高训练速度。但后续 Akcay 等[15]指出，f-AnoGAN 方法的迭代优化过程需要数百次反向传播，导致方法的计算复杂度极高，同时其两阶段训练过程较为烦琐，因此提出了一种端到端的联合学习方法 GANomaly，同时对生成器和判别器进行训练。此外，基于 GAN 模型的方法还有 Han 等[48]提出的医学异常检测生成对抗网络（medical anomaly detection GAN, MADGAN）等，针对多相邻切片类型的数据进行检测。Li 等[49]提出的 PAttl-GAN 着眼于解决像素级异常定位的挑战。其核心在于通过多级高效融合与像素级定位模块的协同设计，在利用未标注与伪异常数据的同时，显著提升了模型对细微异常的分辨与定位能力。总体而言，基于 GAN 模型的方法能够生成高质量的重建图像，但难以保留与原始图像一致的解剖结构，往往表现出过拟合的迹象，且训练难度较高[29]。

此外，一些研究探索了去噪算法对重建及检测过程的影响，如 Kascenas 等[50-51]在训练过程中对正常图像添加特定结构的噪声，通过优化噪声的空间分辨率和强度来提升去噪自编码器（denoising autoencoder, DAE）和扩散模型在医学图像异常检测中的性能。Ho 等[52]利用去噪扩散概率模型，通过对输入图像进行扩散以破坏异常区域，随后经去噪过程获取相应的正常重建图像。Behrendt 等[53]在此基础上进一步探索了基于图像块（patch）的扩散重建过程，利用噪声块周围的图像背景信息，规避病灶组织被过度重建的问题。然而，其在医学影像应用中面临的关键挑战为，施加过多噪声能有效破坏异常区域便于后续修复，但也会损失图像细节，而噪声过少则可能导致异常无法被有效移除。针对该问题，Bi 等[54]提出了一种 Synomaly 噪声与多阶段扩散框架，通过在训练时向正常图像注入模拟真实病变形态的合成噪声，使模型学习针对性的异常修复能力；推理阶段则采用多阶段迭代去噪与掩码融合的方式，在有效去除异常的同时更好地保留了健康组织的结构细节。

除上述直接关注重建结果的研究外，还有一些研究着眼于如何将显著性可视化技术，如将类激活映射（class activation mapping, CAM）[55]、梯度加权类激活映射（Grad-CAM）[56]等嵌入重建过程中，以提升异常检测的可视化定位效果，包括聚焦异常核心、勾勒异常组织轮廓等[57-59]。

1.1.2　基于特征重建的方法

与数据重建方法不同，另一类重建范式通常更关注于重建过程中输出的多尺度特征，而非重建前后的图像。如图2所示，此类方法旨在利用大规模数据集，如基于ImageNet[60]的预训练模型，将图像映射到特征空间，从而提供信息丰富的特征建模。尽管其图像重建结果在视觉上可能失真，但深层特征蕴含的高级语义信息往往能够更有效地捕捉到与异常模式相关的判别性模式。

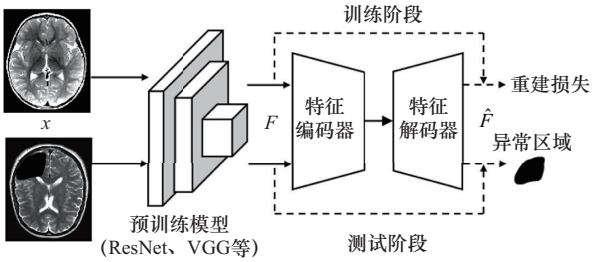


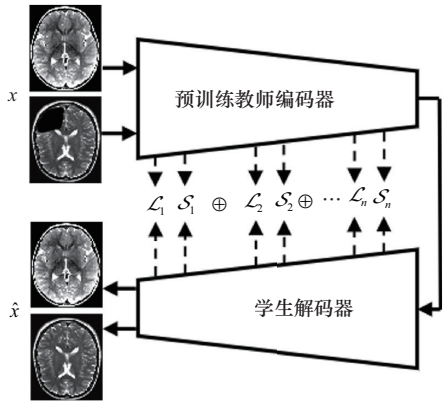
图2　基于特征重建的异常检测

由于预训练模型能够有效简化特征提取网络从头训练的过程，并且提供了具有判别k性和高信息量的通用图像表征作为数据重建的前置条件，近年来多位研究者也围绕如何将该技术引入医学UAD领域展开了探索。早期的一些相关研究指出，预训练模型为判别式模型，而特征重建方法需要对称的编码器–解码器结构，直接连接预训练编码器会破坏重建能力[61]。因此，初期研究多为探索如何对预训练模型进行间接利用，如Silva-Rodríguez等[59]利用预训练网络提取的正常样本特征进行多元高斯分布拟合，并根据测试样本与该分布的马氏距离作为异常评分的评判标准。同时，该研究还暗示了语义判别信息主要存在于正常数据的低方差成分中。随着Transformer架构在视觉任务中的广泛应用，多项研究开始探索其在医学UAD任务中的潜力。You等[62]针对预训练卷积神经网络（convolutional neural network, CNN）存在的“恒等捷径”问题，提出了一种基于Transformer架构的ADTR方法，引入与正常样本高度相关的辅助查询嵌入。该方法首先利用预训练CNN模型提取特征并转换为token，随后利用Transformer模型进行重建，从而解决CNN重建器在语义区分和过度重建方面的问题。然而，ADTR方法所采用的标准Transformer架构存在计算复杂度高与多尺度感知较弱的局限性。为此，该研究进一步提出了UTRAD框架[63]，对图像进行分块处理，通过U形金字塔分层Transformer降低计算复杂度，并提出了双阶段位置编码策略以提升异常定位的空间精度。文献[64]提出了一种多类别异常检测框架UniAD，将预训练特征提取模型与分层查询解码器、邻居掩码注意力模块相结合，并采用特征抖动策略，进一步解决“恒等捷径”问题。与ADTR方法类似，Lu等[65]提出了一种异构自编码器Hetero-AE，其采用预训练的ResNet结构，解码器则设计为混合CNN-Transformer网络，通过异构结构使模型学习正常数据的内在信息，并扩大异常样本之间的差异以抑制过度重建问题，同时利用多尺度稀疏Transformer模块对长程依赖建模与计算复杂度进行平衡。

上述方法虽成功将预训练模型嵌入UAD任务，但这种多阶段的利用方式仍存在一些关键瓶颈。究其原因，由于自然图像与多模态的医学影像之间存在显著的跨领域语义差异，预训练模型提取特征往往难以匹配医学数据的正常模式，因而对复杂异常的敏感度较弱。针对这一问题，多项研究尝试利用迁移学习方法，将预训练模型的知识适配至医学UAD任务中。其中最典型的实现手段为知识蒸馏技术[66-69]，如Wang等[69]采用与预训练教师网络同构的学生网络，通过匹配教师网络的多层级特征构建隐式特征金字塔，并利用分层监督进行多尺度特征匹配，逐层生成异常图像并进行融合。Salehi等[70]将预训练专家网络输出的多层次特征表示迁移到一个更简单的克隆网络以实现轻量化蒸馏，并通过克隆与专家网络对样本提取特征的总体差异来判别异常。

在上述研究的基础上，研究者尝试对预训练模型的迁移特征进行端到端重建。其中一项称为AE-Flow的方法由Zhao等[71]提出，该方法采用预训练编码器并锁定其模型权重（冻结），联合可训练的对称解码器模型进行重建，其特点在于在编码器–解码器之间构建了归一化流瓶颈块，通过对特征的分布进行建模，将预训练模型提取的特征转化为可量化的似然指标，从而通过似然度约束最大程度地保留与正常模式相关的语义信息，进而提升重建质量。Deng等[17]联合知识蒸馏方法与数据重建框架，提出了一种RD4AD方法。该方法同样将冻结参数的预训练模型作为教师编码器，并通过教师编码器提取多尺度特征图聚合为一个密集嵌入，随后以该

嵌入特征作为学生解码器的输入以逐级重建对应尺度的特征，训练阶段仅更新嵌入层与学生解码器的权重，从而以“反向知识蒸馏”的方式直接将预训练网络嵌入自编码器重建模型。如图 3 所示，RD4AD 模型构建的反转型教师–学生模型能够通过多层次的压缩和恢复过程使预训练与医学语义信息渐进地对称匹配，同时依据数据与特征重建误差共同实现异常检测，提供了一种新颖的 UAD 范式。



$\mathcal{L}_n$：第n层损失值　　$\mathcal{S}_n$：第n层异常得分

图3　基于反向知识蒸馏的异常检测

由于 RD4AD 模型能够兼顾图像结构信息提取与预训练语义迁移，展现出较高的异常敏感性，多位研究者已在其基础上针对医学领域的数据特性加以改进，以进一步适应病理检测任务。在此基础上，Liu 等[72]提出了一种带有跳跃连接的师生模型 Skip-TS，在反向蒸馏模型中引入跳跃连接机制传递多尺度特征，并设计了多尺度异常一致性损失以提升定位精度。为进一步解决模型在医学图像上的过泛化问题，Ge 等[73]提出了一种 ESC-DRKD 方法，该方法引入了投影层与增强的跳跃连接机制，通过将伪异常特征映射至正常特征空间，有效提升了模型对正常模式的重建能力。但后续研究发现，尽管知识蒸馏方法在一定程度上实现了预训练特征的跨域迁移，但其核心局限在于作为教师编码器的预训练模型被完全冻结。这种策略本质上是为了避免在端到端微调编码器和解码器时可能出现的“模式崩溃”，即模型在微调过程中判别能力退化，倾向于将所有样本映射到相似的特征表示，导致异常检测性能急剧下降这一问题[7]，但它也使编码器无法根据目标医学图像域的特点进行自适应优化，仅能够通过解码器训练进行特征重建。因此，解码器面对的是一个未经目标域优化的、缺乏域特异性的特征表示，极易学习到一个简单的通用映射函数，这使模型难以学习到高度判别性的特征表示，从而限制了其在医学异常检测任务上的性能上限。针对如何安全地微调预训练编码器，使其适应医学图像域的问题，Rahmaniar 等[74]对 RD4AD 模型进行扩展，提出了 Multi-AD 方法，通过解耦域通用与域专用特征，并引入对抗性判别器进行特征分布对齐，从而提升模型在工业图像、医学影像等多个异构领域的泛化能力。Guo 等[75]提出了一种编码器–解码器对比（encoder-decoder contrast, EDC）方法，通过联合优化教师编码器和学生解码器，缓解预训练特征与医学图像之间的语义鸿沟。通过引入停止梯度操作，利用编码–解码特征对之间的对比学习引导特征重建，有效防止了编码器微调时的“模式崩溃”。同时设计了一种全局余弦距离损失取代传统的逐点对比，显著提升了训练稳定性。Tang 等[18]针对 EDC 方法不通过反向传播，而是隐式训练编码器导致的优化不充分问题，提出了 EA2D 方法，在 RD4AD 模型的基础上引入转换一致性对比学习（transformation-consistency contrastive learning, TCCL）方法，联合显式编码器优化、自注意力增强重建和多视角融合等多种优化策略，进一步解决预训练模型在医学 UAD 领域的域适应与模式崩溃问题。

总体而言，得益于预训练特征蕴含的丰富判别信息，近年来多数基于预训练模型的重建方法均展现出了优异的检测性能。经过医学领域的持续迭代，此类方法在医学影像 UAD 领域的能力已被充分验证，逐渐取代从头训练的数据重建模型，成为一类新兴的主流 UAD 范式。在特征空间中，异常样本因大幅偏离正常模式而更难以被重建。因此，在数据重建类方法中普遍存在的过度重建问题在一定程度上得以缓解。

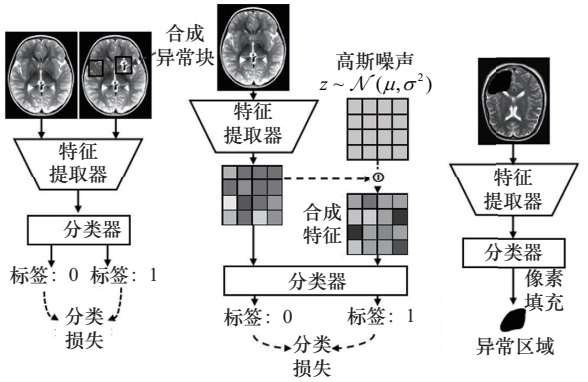
尽管基于预训练模型的重建方法具有显著优势，但因多尺度特征提取过程中固有的低级信息丢失，其在微小病变的精准定位上仍面临挑战。虽然具有更丰富语义信息的特征图可以突出图像强度较为细微的异常区域，但却难以完全保留异常区域的形状和边界等低级细节。因此，现有方法通常聚合来自多个中间层的特征图作为折中方案[64, 71, 76-77]。然而，如何针对特定大小的异常模式对层的数量和

深度进行选择，仍是特征重建类方法需要探索的方向。

综上所述，数据与特征重建的核心差异在于其优化目标与信息层级。数据重建直接操作于像素空间，其优势在于能够保留丰富的低级细节，因此对于小病灶的精确定位具有理论优势。但其劣势在于易受图像噪声干扰，且模型可能因“过泛化”将异常区域同步重建，导致检测失效。特征重建则操作于深层语义空间，其优势在于对高级语义异常更为敏感，并能有效缓解过度重建问题。但其劣势在于因特征图分辨率降低导致的空间信息丢失，可能影响细小或边界模糊异常的定位精度。两类方法在定位精度与语义敏感性方面各有优劣，因此在实际应用中需根据目标任务的数据特性进行选择。

1.2 基于合成的异常检测方法

作为另一类学习范式，基于合成的UAD方法旨在将UAD任务转换为一个全监督的二分类问题。如图4所示，此类方法通过向正常样本加入人工合成的异常，构造伪异常检测任务来训练一个分类或分割模型。该模型尝试学习区分正常与合成异常的数据模式，并利用在该任务中学到的判别能力检测真实异常。异常合成类方法的原则为设计尽可能逼真的伪异常，以覆盖多样化的异常模式，从而提升模型在各种未知真实场景中的泛化性能。



(a) 基于图像级异常合成的训练流程 (b) 基于特征级异常合成的训练流程 (c) 测试流程

图4 基于合成的异常检测

根据异常合成层次的不同，此类方法可分为图像级和特征级异常合成两种。前者在图像水平上伪造异常，是最为常见的一种类型。此类方法大多通过patch的形式在图像中合成异常，如Li等[19]提出的CutPaste方法，从正常训练图像中裁出一小块可变大小和纵横比的矩形区域，通过旋转patch或者抖动patch的像素值并将其粘贴到原始图像的随机位置来合成异常，随后构造一个自监督表征学习任务，将正常样本与CutPaste样本进行分类以学习图像表征，最后在学习到的图像表征上构建高斯密度估计器。Sato等[78]基于CutPaste方法提出了一种解剖感知粘贴（anatomy-aware pasting, AnalPaste）策略，利用基于阈值的肺部分割掩膜来指导胸部X光片的异常合成，仅在被分割的肺部区域内生成逼真的异常阴影，从而利用解剖学信息强化方法在医学任务中的表现。Tan等[79]提出了一种外部块插值（foreign patch interpolation, FPI）方法，从两个独立正常样本中提取相同位置的图像块，用这两个图像块的线性插值替换其中一个样本的对应块，从而合成异常区域。但这种方法合成的异常过于显著，会影响模型的细微异常识别能力，因此后续研究提出了基于泊松图像插值（poisson image interpolation, PII）的合成方法[20]，通过融合图像块的梯度而非原始强度值，从而能够产生更加真实的异常，并在多种不同类型的医学数据集中进行验证。后续的多项研究在PII方法的基础上进行了改进，如Müller等[80]通过正态分布控制圆形异常生成的半径和位置。Schlüter等[81]将数据增强、形状采样和背景约束等策略与PII方法进行集成，以合成更加多样化且与任务相关的伪异常。Baugh等[82]针对三维医学影像对PII方法进行扩展，解决了传统方法在非凸区域和高维数据中的局限性，并引入基于合成任务的交叉验证方法，防止模型对合成异常过拟合。此外，还有一些研究采用像素水平的异常合成策略，如Zhang等[83]提出的DeSTSeg方法，通过Perlin噪声生成不规则掩码并嵌入外部图像内容以构造逼真异常。该方法利用预训练的教师网络提取干净图像特征，并训练一个去噪学生网络从合成异常图像中重建正常特征表示，迫使两者在异常区域产生显著特征差异，最终利用合成数据的像素级标注监督分割网络，自适应融合多层级特征差异，实现端到端的异常检测与定位。Cai等[84]为了解决不同复杂程度合成策略对检测鲁棒性的影响提出了差异感知框架（discrepancy aware framework, DAF）方法，通过对特定像素区域进行异常合成生成训练数据，并采用教师–学生网络提取异常区域的外观无关差异图，引导分割网络识别异常区域，有效缓

解对合成策略的过拟合。

特征级异常合成方法旨在从特征水平上伪造异常，如在正常特征图中添加高斯噪声等。此类方法的代表如 Liu 等[22]提出的 SimpleNet 方法，使用预训练主干网络提取多层级特征，并对正常特征添加高斯噪声生成合成异常作为负样本，通过简单的多层感知机判别器区分正常特征与合成异常特征，以实现异常检测。Zhang 等[85]提出的 RealNet 方法，基于扩散模型生成逼真、可控且多样化的异常模式，并通过异常感知特征选择和重建残差选择策略自适应筛选最具判别性的特征，将异常合成与数据重建机制相结合以提升模型的检测能力。

一些研究者也尝试结合了图像级和特征级的异常合成策略，如 Chen 等[86]提出的 GLASS 方法，包括 3 条分支路径：正常样本处理分支，用于提取适应性的正常特征；全局异常合成分支，通过正常特征合成全局异常特征；局部异常合成分支，用于在图像级别上合成局部异常，通过纹理叠加的方法生成更多样化的异常。这 3 条分支路径通过一个共享的判别器进行联合训练，从而在特征级和图像级上协同工作，使模型能同时精准检测微弱和明显的缺陷，以可控的方式合成更广泛的异常样本并提升检测效果。

除上述研究外，将异常合成策略与重建框架相结合的方法也得到了深入探索，Zavrtanik 等[21]提出的 DRAEM 方法通过异常模拟器合成异常样本，并构建重建与判别双网络结构，其中重建网络修复合成的异常区域，判别网络则对比原始与重建图像。通过端到端训练，模型可直接输出像素级异常分割图。但 Cai 等[87]指出 DRAEM 方法的判别网络容易对人工合成的异常模式产生过拟合，因此提出了一种 ASR-Net 方法，不以合成异常作为直接检测的目标，而是学习将原始异常分数图映射到更准确的异常区域，从而减少对人工合成异常模式的依赖，避免过拟合并提升泛化能力。

1.3　基于自监督学习的异常检测方法

自监督学习（self-supervised learning, SSL）在图像处理与分析领域中已被广泛应用，通常被用于预文本或代理任务。SSL 在不依赖于大量人工标注数据的情况下，能够让模型学习到高质量的图像特征表示，并且该特征对下游任务具有很强的可迁移性[88]。如图 5 所示，目前基于 SSL 的 UAD 方法通常由两个阶段构成，首先在正常训练数据上构造预文本/代理任务以自监督地学习样本的特征表示，随后基于该特征表示构建单类分类器以进行检测。

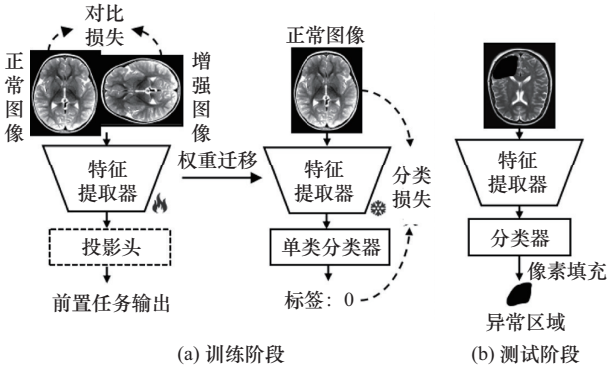


图 5　基于自监督学习的异常检测

医学 UAD 领域中最主要的一类 SSL 预文本实现范式是对比学习方法。UAD 任务通常需要模型学习正常样本的分布，而异常样本往往在局部区域表现出与该正常分布模式的显著差异。对比学习方法通过拉近输入和输出中相近的语义特征，同时疏远异常样本在特征空间中的距离，确保重建的局部语义信息与正常样本一致，这种机制能够有效用于 UAD 任务中正常样本的重建和异常样本的判别。

许多先进的对比学习方法能够通过数据增强来构造正负样本对，并设计预文本/代理任务进行图像表示学习，从而为无监督学习中的一致性样本表示学习过程提供支持，并将提取到的特征用于包括 UAD 任务在内的下游任务。这些方法包括使用判别性损失区分正负样本对的 LSTM 方法[89]、以最大化全局与局部特征之间的互信息为目标的 Deep InfoMax 方法[90]、使用自回归模型的对比预测编码（contrastive predictive coding, CPC）方法[91]、使用大规模数据增强和批处理构建样本对的 SimCLR 方法[92]、采用动态字典和动量编码器的 MoCo 方法[93]以及不需要显式构建正负样本对，而是采用在线–目标网络双路径架构、通过增强视图进行自预测的 BYOL 方法[94]等。

表示学习–单类别分类的两阶段 UAD 框架首先由 Sohn 等[23]提出，其构建的检测模型包括一个自监督学习模块，通过分布增强对比学习扩展训练数据分布，减少对比表示的一致性，使同类样本的表示更加紧凑；以及一个表示驱动的单类分类器，如单类支持向量机（one-class support vector machine,

OC-SVM）[95]、支持向量数据描述（support vector data description, SVDD）[96]或核密度估计（kernel density estimation, KDE）[12]等。Tian等[97]在该框架基础上将基于预文本任务约束的对比分布学习与异常合成相结合，针对医学UAD任务中的图像小病变挑战引入块级对比学习，同时对医学正常样本分布单一的问题采用强增强合成伪异常，优化特征表示并提升模型泛化性。该研究后续还提出了一种基于对比学习方法的伪多类强增强（pseudo multi-class strong augmentation via contrastive learning, PMSACL）方法[25]，分别引入多中心损失和温度缩放策略优化对比学习过程，以及引入医学图像数据增强方法MedMix，通过切割、变形和粘贴图像块来合成不同数量、大小和外观的异常病变。从而使模型进一步适应不同医学图像模态和多样化的病灶类型。Lu等[98]提出了一种基于块级对比学习的自编码器PatchCL-AE，将重建与原始图像对应的局部块与非对应块分别作为正/负样本，通过最大化正样本对的语义相似性并最小化负样本对的语义相似性，增强了模型对正常医学图像局部语义一致性的建模能力。

由于预训练模型提供的通用性表征能力在医学UAD领域已展现出显著优势，部分研究也尝试将其与SSL进行结合。Reiss等[99]探索了标准对比损失在利用预训练特征执行医学UAD任务时表现不佳的问题，表明这一问题源于标准对比损失导致特征均匀性增加，并针对性地提出了一种均值偏移对比（mean-shifted contrastive, MSC）损失，通过将对比损失的计算从原点转移到正常样本特征的中心，提升预训练模型的迁移性能，解决了标准对比损失在跨域特征上的适应性问题。

除了上述结合分类任务的二阶段范式外，SSL还可以通过联合重建任务来学习特征表示，并直接构建端到端的异常检测模型。这类方法通常将SSL的预文本任务与主任务深度融合。例如，Zhao等[24]提出的SALAD框架将SSL作为一种强大的特征增强器，该方法在图像和特征空间构建双向重建路径，并引入了修复、局部像素打乱和非线性强度变换等多种预文本任务来扰动输入图像，迫使编码器学习对内容破坏具有鲁棒性的特征表示。其核心创新在于通过特征层面的重建一致性来放大异常差异，同时利用中心约束损失在特征空间构建紧凑的“正常”原型，从而不需要构造显式的正负样本对，而是通过生成式重建任务隐式地学习正常模式的一致性，为基于SSL的医学UAD提供了另一种有效的实现思路。

1.4 基于记忆库匹配的异常检测方法

基于记忆库匹配的异常检测方法通过对正常模式进行原型学习来约束特征表达，从而避免模型泛化性过高导致的恒等映射问题。如图6所示，此类方法的核心机理在于通过编码器提取训练阶段正常样本的特征，并构建一个存储正常模式原型特征向量的记忆库。在推理阶段，该方法通过度量测试样本特征与记忆库中所有原型之间的匹配程度来判定异常，若样本特征能够与记忆库中的某个正常模式有效匹配，则视为正常样本；反之，若无法找到足够匹配的原型，则依据显著偏离正常分布的匹配结果将其判定为异常。该方法通过显式约束对正常模式的高效记忆，有效避免了模型过拟合。

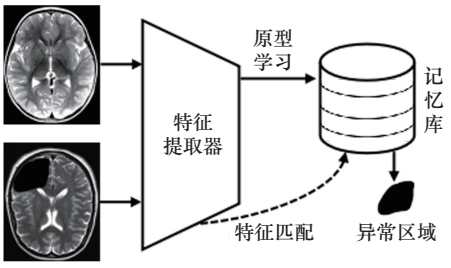


图6 基于记忆库匹配的异常检测

目前，由于存储大量特征嵌入向量造成的空间开销及对所有测试样本进行检索匹配带来的高时间复杂度，记忆匹配方法在医学UAD领域中的应用相对较少。在相关研究中，Roth等[100]提出的PatchCore方法基于ImageNet预训练网络提取中间层特征，构建包含局部块特征的记忆库，通过邻域聚合操作增强特征的鲁棒性，并采用贪心核心集降采样技术，在优化存储开销的同时保持特征覆盖的完整性。在测试阶段，通过计算样本特征与记忆库的最近邻距离实现异常检测。Zhou等[101]提出了一种ProxyAno方法，利用带有中间记忆库的自编码器重构超像素而非原始像素。该方法首先将输入图像转换为超像素代理图像，从而对图像表示进行简化，并通过记忆模块匹配正常模式重建图像以进行检测。Xiang等[102]提出了一种名为SQUID的方法，专注于胸部X射线图像中的异常检测，同时发布了一个合成异常数据集DigitAnatomy。该方法通过将

输入图像分割为不重叠的图像块，利用编码器提取特征，并引入一个空间感知的记忆队列来动态存储和检索正常的解剖模式。通过教师–学生生成器结构和特征层面的修复机制，异常特征能够被修复为最接近的正常模式，最终通过判别器实现异常检测。Zhou 等[103]提出了一种 MemSTC-Net 方法，通过记忆正常图像的结构–纹理对应关系来实现异常检测。该方法使用两个结构提取器分别提取语义和低级结构，并通过一个结构–纹理记忆模块（structure-texture correspondence module, STCM）将结构特征映射到对应的纹理特征以重建图像，继而根据重建误差进行检测。此外，该方法还引入注意力融合模块和结构一致性约束，以进一步提升重建质量和检测鲁棒性。作为对传统记忆库方法的深化，Kim 等[28]提出的基于块状向量量化的方法 P-VQ 则构建了一种将记忆原型从特征向量升级为特征图块的新思路。该方法通过块级向量量化，学习并记忆正常图像块的整体空间上下文信息，从而扩大了每个记忆单元的感受野，能更有效地防止异常模式被临近的正常模式所协同重建。此外，该研究还针对“码本坍塌”，即记忆库中仅有少数原型被频繁使用而丧失多样性，导致模型表达能力下降这一向量量化中的常见问题，提出了 top-k%抛出策略，通过动态掩码高使用率的码本项，强制模型更均衡地利用所有记忆单元，从而提升码本利用效率和模型检测鲁棒性。

此外，Defard 等[27]提出的 PaDiM 方法通过利用预训练 CNN 提取图像块的多层特征嵌入，建立参数化的多元高斯分布来建模正常样本的统计特征，并通过马哈拉诺比斯距离度量测试与正常样本分布的差异，从而同时实现高精度异常检测和定位。该方法构建的高斯模型近似于对嵌入特征原型的压缩归纳，而基于马哈拉诺比斯距离的度量操作则充分体现出特征空间中基于匹配的异常检测思想。类似的研究还有 Xiao 等[104]提出的 MS-NF 方法，通过一系列可逆的归一化流网络，将图像的多尺度特征映射到一个结构化的潜在空间，并直接计算特征在该空间下的精确似然值作为匹配度量的依据。与 PaDiM 方法利用预定义的统计距离不同，MS-NF 方法通过最大似然估计端到端地学习从复杂特征分布到简单先验分布的映射函数，实现了对正常模式更灵活、更强大的参数化建模。因此，PaDiM 与 MS-NF 方法均可被视为对传统记忆库匹配思想的一种参数化重构，并针对传统记忆库存在的时空复杂度问题进行了优化。

2　基准数据集及评价指标

2.1　医学影像检测常用数据集

北美放射学会（RSNA）数据集是胸部 X 射线（CXR）数据集[105]，其样本表现为 DICOM 格式的灰度图像，共包含 8 851 张健康 X 光片及 6 012 张存在肺混浊的异常 X 光片。该数据集由北美放射学会发布于 Kaggle 机器学习挑战赛，主要用于胸部肺炎 X 光图像的分类及病灶区域定位任务。在 UAD 领域可用于检验模型在胸部异常情形中的图像级异常检测能力。

VinDr-CXR 同样为胸部 X 射线数据集[106]，包含 10 606 张正常胸部 X 光片和 4 394 张异常胸部 X 光片，其异常类型包括主动脉增宽、肺不张、钙化、心脏肥大、肺实变、间质性肺病、浸润、肺部混浊影、结节/肿块、胸腔积液、胸膜增厚、气胸、肺纤维化以及其他病变，共计 14 个类别。该数据集由越南 108 医院和河内医科大学医院共同收集，由于其异常类型丰富，能更全面地评估模型对不同胸部异常的分类性能，为胸部 X 光影像 UAD 研究提供了更广泛的场景选择。

Br35H 是脑部 MRI 数据集[107]，发布于 Kaggle 机器学习挑战赛。该数据集由从 3D 模态图像中截取的 2D 灰度图像切片构成，总共包含 1 500 张健康图像和 1 500 张脑肿瘤图像，用于肿瘤图像级检测和分类任务，该数据集中的肿瘤通常较明显，作为 UAD 任务相对简单。

Brain Tumor 是脑部 MRI 数据集[108]，包含 2 000 张正常脑部 MRI 切片、1 621 张含胶质瘤的切片以及 1 645 张含脑膜瘤的切片，其中胶质瘤和脑膜瘤均被视为异常情形。该数据集由多个来源的数据集共同构成，以解决 Br35H 数据集挑战性过低的问题。其中无肿瘤的图像来自 Br35H 数据集和 Saleh 等的收集，含胶质瘤和脑膜瘤的图像则来自 Saleh 等和 Cheng 等的收集，用于评估模型对不同类型脑部肿瘤 MRI 的图像级异常检测性能。

OCT2017 是视网膜光学相干断层扫描数据集，由加利福尼亚大学圣地亚哥分校（UCSD）的 Zhang 等整理发布[109]，用于评估算法在检测和区

分不同类型视网膜病变方面的性能。该数据集由从3D模态图像中截取的2D灰度图像切片构成，包含26 565张正常图像和三类异常图像，包括250张玻璃疣（drusen）病变图像、250张糖尿病黄斑水肿（diabetic macular edema, DME）病变图像和250张脉络膜新生血管（choroidal neovascularization, CNV）病变图像。

LAG是视网膜眼底图像数据集[110]，其样本表现为RGB图像，包含6 882张正常视网膜眼底图像和4 878张含青光眼的异常图像。该数据集由文献[110]收集，用于早期青光眼图像分类与异常检测等领域的研究。

APTOS也是一种视网膜病变数据集[111]，其样本表现为RGB图像，包含1 805张正常图像和1 857张糖尿病视网膜病变图像，病变图像依照严重程度由轻到重被标注为1~4级，共同作为异常样本。该数据集由2019年APTOS盲症检测挑战赛官方发布，用于糖尿病视网膜病变图像级检测和病变程度分类，作为UAD任务其特点为轻、中度病变图像中病灶极不明显，检测难度较高。

ISIC2018是皮肤病变数据集[112]，其样本表现为RGB图像，其中色素痣皮肤图像被作为正常样本，共包含10 015张正常图像。异常数据集共包括3 310个样本，由6个类别的疾病皮肤图像组成，分别为327张光化性角化病图像、514张基底细胞癌图像、1 099张良性角化病图像、115张皮肤纤维瘤图像、1 113张黑色素瘤图像和142张血管性皮肤病变图像。该数据集由ISIC2018挑战赛发布，主要用于皮肤病图像的病灶分割、疾病分类及像素级UAD任务，可评估模型在多类型皮肤病变识别上的性能，以及对色素痣与皮肤病变的区分能力。

Camelyon16是淋巴结苏木精&伊红（H&E）数字病理图像数据集[113]，其样本表现为RGB图像，包含130张正常淋巴组织切片图像和110张作为异常样本的乳腺癌淋巴组织切片图像，图像数据以“金字塔”的形式进行存储，在40倍放大倍数下的最大分辨率图像矩阵大小能达到300 000×150 000，因此在实际任务中需首先通过裁剪等预处理方式划分成较小的图像块。该数据集由医学顶级会议ISBI在2016年开展的Camelyon16竞赛中发布，旨在对乳腺癌在淋巴结中的转移场景进行病理切片的分类、病灶定位与异常检测。

FastMRI+是脑部MRI病理数据集[114]，其样本表现为T1加权脑部磁共振影像，包含来自健康个体的131张训练图像和15张验证图像作为正常样本。异常数据集包括171个独特病理样本，涵盖13种疾病类别（如水肿、肿块、病变、中隔缺失等），共643个标注病理区域。该数据集由FastMRI团队扩展发布，用于脑部影像的异常检测与定位，其特点在于涵盖了广泛且多样的脑部病理类型，可作为评估模型在复杂多病种场景下的泛化能力。

ATLAS v2.0是脑卒中病灶数据集[115]，其样本表现为脑部磁共振影像，包含655例带有详细卒中病灶标注的训练样本。测试集按病灶大小分为小（<71像素）、中、大（≥570像素）三类，共420例受试者数据。该数据集由ATLAS挑战赛发布，专门用于脑卒中病灶的检测、分割与异常定位，其特点在于病灶大小、形状和强度差异显著，可有效评估模型对不同尺度病变的敏感性与恢复能力。

图7展示了部分医学UAD公开数据集的图像示例，其中包括来自不同模态、具有不同色彩类型的样本。

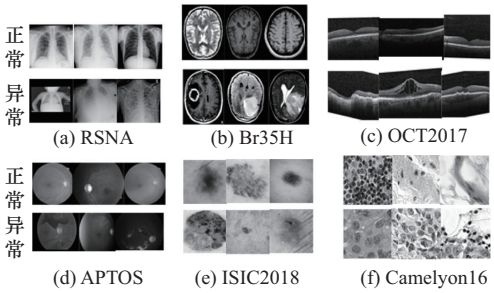


图7　部分医学UAD公开数据集图像示例

2.2　异常评分及评价指标

2.2.1　异常评分计算方法

异常评分是UAD任务的核心输出，为一个连续标量值，用于量化测试样本与学习到的正常模式之间的偏差程度。该分数直接决定了模型的最终输出，在图像级分类中，异常评分计算方法通过聚合操作，如全图计算或逐像素计算后取最大值或平均值获得，并与阈值比较以判断图像是否异常。在像素级定位中，分数直接构成异常热力图，经阈值化后得到异常区域的分割掩码。其核心价值在于提供了灵活的决策支持，使用者可通过调整阈值在灵敏度和误报率之间取得平衡，并支持通过排序类指标作为模型性能的评估准则，但同时也面临着分数本

身无绝对意义以及阈值难以客观选取的固有挑战，因此现有UAD方法也基于模型特性及性能侧重建立了多种不同的异常评分计算方式，总结如下。

1) 基于像素重建误差的方法

基于像素的重建误差是最直观的方法，通过计算输入图像x与重建图像$\hat{x}$在像素空间的差异来定位异常。对于图像中的第i个像素，图像的异常评分计算方式包括绝对值误差$\mathcal{A}(x)=\sum|x_i-\hat{x}_i|$、平方误差$\mathcal{A}(x)=\sum|x_i-\hat{x}_i|$和结构相似性指数$\mathcal{A}(x)=\sum\left(1-\mathrm{SSIM}(x,\hat{x})_i\right)$。尽管计算方式较简单，但该方法对高频细节和图像噪声极为敏感，易在正常组织边界处产生高误差，往往导致较高的假阳性率。

2) 基于特征重建误差的方法

对于对称型自编码器模型或预训练知识蒸馏模型，其异常评分计算过程可完全在网络提取的特征空间中进行。给定具有多个尺度的特征$\boldsymbol{F}$，其重建误差为每层的提取特征$\boldsymbol{F}$与对称层重建特征$\hat{\boldsymbol{F}}$间的距离总和，对于其中的第k层特征，计算公式与像素重建误差方法类似，表示为$\mathcal{A}(x)=\sum(\boldsymbol{F}_{k,i}-\hat{\boldsymbol{F}}_{k,i})^2$、$\mathcal{A}(x)=\sum\left(1-\mathrm{SSIM}(\boldsymbol{F},\hat{\boldsymbol{F}})_{k,i}\right)$。近年来，有研究表明余弦相似度度量能更好地捕捉语义方向的差异，且具有幅度不变性和训练稳定性，基于图像级或像素级余弦相似度的异常评分计算方法也逐渐被多项研究采用，表示为$\mathcal{A}(x)=\sum\left(1-\dfrac{\boldsymbol{F}_k\cdot\widehat{\boldsymbol{F}_k}}{\|\boldsymbol{F}_k\|\|\widehat{\boldsymbol{F}_k}\|}\right)$或$\mathcal{A}(x)=\sum\left(1-\dfrac{\boldsymbol{F}_{k,i}\cdot\hat{\boldsymbol{F}}_{k,i}}{\|\boldsymbol{F}_{k,i}\|\|\hat{\boldsymbol{F}}_{k,i}\|}\right)$。总体而言，由于异常与正常特征在高层语义上偏差更大，基于特征重建误差的方法能有效缓解“过度重建”问题。然而，特征图分辨率的降低会导致低层信息丢失，可能影响小病灶的精确定位，因此通常需要通过上采样操作聚合多尺度特征来弥补。

3) 基于分类器输出的方法

基于异常合成的UAD方法可直接利用训练好的分类器输出作为异常评分，表示为$\mathcal{A}(\boldsymbol{x})=\sigma(g(f(\boldsymbol{x})))$，其中$f$为特征提取器，$g$为分类头，$\sigma$为激活函数。该方法的性能高度依赖于合成异常的真实性。若合成数据与真实异常外观差异显著，模型易过拟合至合成伪影而非学习真正的异常模式。

4) 基于概率密度估计的方法

此类计算方式首先获取图像的高级特征表示，然后以此特征为基础，在特征空间中构建密度估计器$\mathcal{A}(x)=-\ln p_\psi(f(x))$，其中$p_\psi$是基于学习特征$f(x)$估计的概率密度函数，随后对数据的分布计算概率密度基于负相关原则转化为异常评分。该方法能够较好地避免传统分类器易过拟合的问题，具有更强的泛化能力。

5) 基于感知损失重建误差的方法

此类计算方式适用于直接采用预训练模型的数据重建方法，其核心思想是将原始与重建图像通过预训练模型映射至高维特征空间ϕ计算差异，计算方式为$\mathcal{A}(x)=\dfrac{\|\phi_l(x)-\phi_l(\hat{x})\|_1}{\|\phi_l(x)\|_1}$，其中$\phi_l$表示预训练网络第$l$层的计算过程。该方法能够避免无关像素波动引起的伪异常，更好地捕捉与语义相关的异常模式。

6) 其他方法

除上述相对通用的异常评分计算方法外，一些研究也针对其模型特性设计了对应的评分方法，主要包括以下几种。

基于特征差异的方法。通过比较测试样本特征与正常参考特征之间的差异来计算异常评分，参考特征可来自知识蒸馏的教师模型或基于特征原型提取构建的记忆库。基于教师–学生模型的异常评分计算方法为$\mathcal{A}(x)=\|f_{\text{teacher}}(x)-f_{\text{student}}(x)\|_2$，其中通过学生模型提取测试样本特征$f_{\text{student}}(x)$；基于记忆库的异常评分计算方法为$\mathcal{A}(x)=\arg\min_{m\in\mathcal{M}}\|f(x)-m\|_2$，其通过测试样本特征$f(x)$与其在记忆库中$\mathcal{M}$最近邻特征$m$的距离进行计算。

梯度类方法。此类方法计算重建误差相对于输入图像的梯度作为异常评分，表示为$\mathcal{A}(x)=\sum\left(\dfrac{\partial(x_i-\hat{x}_i)^2}{\partial x_i}\right)$，一些基于VAE模型的方法在其基础上还引入了KL（Kullback-Leibler）散度的梯度项，表示为$\mathcal{A}(x)=\sum\left((x_i-\hat{x}_i)^2\dfrac{\partial\ell_{\mathrm{KL}}}{\partial x_i}\right)$。

融合判别误差的方法。此类方法利用输入与重建图像在GAN判别器特征空间中的差异进行高级语义校验，作为对传统重建误差的补充，其表达式为$\mathcal{A}(x)=\sum\left(\left\|x_i-\hat{x}_i\right\|^2+\kappa\left\|f(x_i)-f(\hat{x}_i)\right\|^2\right)$，其

中$f(x)$表示判别器的特征提取结果，$\|x_i - \hat{x}_i\|^2$和$\|f(x_i) - f(\hat{x}_i)\|^2$两项分别代表生成器重建误差和判别误差。

2.2.2 模型检测性能量化评价指标

为了将异常评分转换为可评估的分类决策，需要建立一个将连续分数映射为离散标签的决策规则，其核心在于选择一个合适的阈值τ。该阈值通常通过在测试集上排序所有异常评分并选取最佳分位点来确定，从而生成二值的异常预测标签，常见的准则包括最大化F1-Score、最大化Dice系数等。随后，将这些模型预测结果与真实标签进行比对，构建混淆矩阵，并计算出一系列依赖于阈值的指标，用于从多个维度准确评估模型在特定决策边界下的分类性能。混淆矩阵定义如表2所示。

表2 根据模型预测结果和真实标签生成的混淆矩阵

真实值	预测为正常	预测为异常
真实正常样本	真阳性（TP）	假阴性（FN）
真实异常样本	假阳性（FP）	真阴性（TN）

表2中的真阳性（true positive, TP）指真实异常样本被模型正确识别为异常的数量，代表了成功的检测；假阳性（false positive, FP）指真实正常样本被模型误判为异常的数量，即误报案例；真阴性（true negative, TN）指真实正常样本被模型正确判断为正常的数量，体现了模型排除干扰的能力；而假阴性（false negative, FN）则指真实异常样本被模型错误归类为正常的数量，即漏报案例。这些基础统计量为多项量化评价指标的计算提供了根本依据，常见评估指标主要包括。

1) 准确率（accuracy, ACC）

所有正确分类样本的比例，综合反映了模型的分类性能，其计算式为

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

2) 精确率（precision, PRE）

预测为异常的样本中真正异常的比例，用于衡量模型避免误报的能力，其计算式为

$$\mathrm{PRE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

3) 特异度（specificity, SPE）

预测为正常的样本中为真实正常的比例，聚焦于模型精准识别负例的能力，其计算式为

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{3}$$

4) 召回率（recall）

该指标又称灵敏性（sensitivity, SEN）、真阳性率（true positive rate, TPR），指在所有真正的阳性（异常）样本中，被模型成功预测出来的比例，用于衡量模型避免漏报的能力，其计算式为

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

5) 受试者操作特征（receiver operating characteristic, ROC）曲线的曲线下面积（area under the curve, AUC）

ROC曲线是一条描绘二分类模型在所有可能的分类阈值下性能的曲线，其纵轴为TPR，横轴为假阳性率（false positive rate, FPR），可视化地展示二分类模型在所有可能的分类阈值下的性能权衡，其中FPR计算式为

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{5}$$

AUC则代表ROC曲线下的面积，其数值大小直接衡量模型的整体分类性能，AUC越接近于1表明模型区分正负样本的能力越强。在异常检测中，AUC因其不受阈值选择和类别不平衡影响的特性，成为评估模型性能的核心指标，其计算式为

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FPR})\,\mathrm{d}(\mathrm{FPR}) \tag{6}$$

6) F1分数（F1-Score）

F1-Score为PRE和SEN指标的调和平均数，用于综合衡量这两项指标，尤其适用于评估正负样本比例高度不平衡的二分类问题，其计算式为

$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{7}$$

7) Dice系数

Dice系数是一种用于衡量两个集合相似度的统计指标，像素级UAD任务可采用该指标来衡量模型预测出的异常区域与真实异常区域之间的像素重叠程度，能够客观衡量模型定位异常、勾勒异常轮廓的能力，其计算式为

$$\mathrm{Dice} = \frac{2|P \cap G|}{|P| + |G|} = \frac{2\mathrm{TP}_{\mathrm{px}}}{2\mathrm{TP}_{\mathrm{px}} + \mathrm{FP}_{\mathrm{px}} + \mathrm{FN}_{\mathrm{px}}} \tag{8}$$

其中，P与G分别表示预测为异常的像素集合与真实异常像素集合，px表示像素级样本的统计量。

在UAD任务中，评价指标的选择主要取决于任务类型与医学图像模态的特性。首先，由于异常检测模型通常输出连续的异常评分，而如何确定一个适当的二值化阈值是一个需要权衡的问题，其根源在于难以找到一个普适的阈值以平衡灵敏度与特异性。因此，在通常情况下该领域优先参照的指标为与阈值无关的AUC，以确保评估的公平性和稳健性。其次，针对不同层级的检测任务，评价指标的侧重点也应有所区分。

在图像级异常检测中，模型需对整张图像作出是否存在异常的判断。针对不同模态的图像特性，首先需采用不同的像素级评分聚合策略。对于病灶弥漫或全局性病变的图像，采用全像素的平均异常分数能更好地反映整体异常程度；对于病灶局灶且显著的任务，则通常取像素级异常分数的最大值，以捕捉图像中最可疑的区域。在模型评估阶段，AUC因其对类别不平衡不敏感而被广泛采用，其值可作为直观的综合检测性能指标，而F1分数则在正负样本比例悬殊时更具参考价值，尤其适用于罕见病筛查等高风险漏诊场景。此外，精确率和召回率的组合分析有助于理解模型在误报与漏报之间的权衡，适用于对假阳性容忍度较低的临床初筛任务。

在像素级异常分割任务中，评估需以像素为基本统计单位。除采用上述图像级指标外，Dice系数和交并比（intersection over union, IoU）也常作为衡量异常区域重叠程度的关键指标，能够直观反映模型对病灶边界的勾勒能力。在病灶细小、边界模糊的模态（如OCT与MRI）中，Dice系数对局部误差更为敏感；在异常区域较大、形态规则的任务（如X光影像）中，IoU也可作为补充评估手段。

3　实验与分析

3.1　定量对比实验

为系统性评估各类无监督异常检测方法在医学影像任务中的性能，本文根据公开代码，或论文中的相关描述对部分代表性方法进行复现，涵盖上文总结的四大类范式，并选用了涵盖不同成像模态与异常类型的多个公开医学影像数据集进行实验，包括OCT2017、APTOS、ISIC2018、Br35H、Brain Tumor与RSNA，并进行了图像级异常检测任务评估，评估指标包括AUC与ACC，以全面反映模型的分类性能与鲁棒性，结果取5次实验的平均值。具体的实验设置如下。

在数据划分方面，对于OCT2017数据集，遵循官方划分方式，其中训练集包含26 315张正常图像，测试集包含1 000张图像，其中drusen病变、DME病变、CNV病变及剩余正常图像各250张；对于APTOS数据集，随机选择其中的1 000张正常图像作为训练集，剩下的2 662张图像（805张正常图像，1 857张异常图像）作为测试集；对于ISIC2018数据集，遵循官方划分方式，其中训练集包含6 705张正常图像，测试集包括193张图像（123张正常图像，70张异常图像）；对于Br35H数据集，随机选取其中的1 000张正常图像作为训练集，其余2 000张图像（500张正常图像，1 500张异常图像）则作为测试集；对于Brain Tumor数据集，随机选取其中的1 000张正常图像作为训练集，600张正常图像与600张异常图像（300张胶质瘤，300张脑膜瘤）作为测试集；对于RSNA数据集，随机选取其中的3 851张正常图像作为训练集，1 000张正常图像和1 000张异常图像作为测试集。在数据预处理方面，所有图像首先统一调整至256×256像素尺寸，并将像素值归一化至[0,1]区间，随后针对RGB图像（如APTOS、ISIC2018）采用ImageNet数据集的均值和标准差进行通道标准化，对于灰度图像（如Br35H、RSNA）则采用单通道均值和标准差进行标准化。在阈值选取方法上，二值分类阈值通过在验证集上最大化F1分数来确定，即在验证集上以每个样本的异常评分作为候选阈值，计算对应F1分数，选取使F1分数最大化的评分作为最优阈值，并将其应用于测试集的异常评分以生成最终分类结果。在具体的模型参数，包括学习率、批大小、训练轮次等超参数方面，均遵循各对比方法原文中所描述的设置进行配置，以确保对比的公平性。所有的实验结果如表3所示，其中，加粗代表同数据集同指标下的最优方法，加下划线代表同数据集同指标下的次优方法。

3.2　实验总结与分析

3.2.1　定量实验结果分析

基于表3所示的实验结果，本文从检测性能维

表3 医学影像异常检测方法定量性能对比

方法	所属类型	来源	OCT2017		APTOS		ISIC2018		Br35H		Brain Tumor		RSNA	
			AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
AE	基于重建的方法	—	86.84%	79.44%	81.45%	73.48%	71.48%	68.87%	97.22%	93.48%	82.44%	76.84%	66.44%	61.64%
VAE		—	92.19%	85.95%	80.49%	74.96%	72.58%	70.48%	97.41%	93.67%	72.83%	67.86%	68.64%	53.13%
GANomaly		ACCV′18	89.60%	84.93%	80.05%	77.71%	70.39%	67.36%	89.28%	88.28%	85.86%	79.05%	81.37%	74.63%
f-AnoGAN		MIA′19	87.89%	83.06%	76.78%	77.22%	72.22%	63.90%	82.05%	81.80%	75.65%	71.58%	76.96%	69.35%
RD4AD		CVPR′22	99.40%	97.20%	94.68%	88.77%	87.23%	80.65%	99.77%	99.24%	97.01%	93.25%	90.05%	82.56%
AE-FLOW		ICLR′23	98.02%	96.24%	91.47%	88.17%	87.82%	85.05%	99.11%	98.67%	97.14%	**96.48%**	81.94%	79.49%
EDC		TMI′24	99.56%	97.90%	95.41%	90.08%	89.33%	84.28%	**99.85%**	99.35%	97.21%	95.08%	87.20%	74.40%
EA2D		TMI′25	**99.84%**	98.42%	**97.05%**	**93.89%**	**90.88%**	**86.87%**	99.83%	**99.44%**	97.66%	95.89%	**91.55%**	85.48%
SALAD	基于自监督学习的方法	TMI′21	95.58%	89.88%	77.38%	76.84%	74.57%	69.88%	97.82%	95.24%	94.47%	92.14%	79.47%	75.60%
PMSACL		MIA′23	95.84%	91.48%	86.97%	82.78%	81.48%	77.96%	96.48%	92.48%	93.88%	89.54%	87.45%	82.41%
PatchCL		CMIG′24	99.29%	96.76%	92.48%	88.49%	86.44%	82.45%	99.56%	99.01%	**97.68%**	94.98%	89.42%	85.44%
SimpleNet	基于合成的方法	CVPR′23	98.45%	94.99%	90.48%	86.48%	83.05%	76.58%	98.88%	98.25%	96.42%	95.43%	78.18%	72.40%
RealNet		CVPR′24	98.86%	97.66%	86.45%	83.49%	85.46%	82.11%	99.11%	98.60%	97.49%	95.89%	79.34%	76.49%
GLASS		ECCV′24	97.75%	96.48%	86.49%	82.49%	86.94%	84.19%	98.64%	96.54%	96.10%	94.59%	81.19%	78.98%
PaDiM	基于记忆库匹配的方法	ICPR′21	96.88%	92.80%	79.00%	78.50%	82.13%	73.06%	98.90%	97.40%	96.78%	95.87%	81.48%	78.16%
SQUID		CVPR′23	97.55%	93.70%	89.61%	86.20%	85.76%	81.76%	99.04%	98.48%	97.12%	96.03%	90.48%	**88.41%**
P-VQ		PRL′25	99.20%	**98.94%**	86.49%	83.48%	79.70%	76.45%	97.66%	93.74%	95.14%	92.48%	78.66%	76.58%

度对各类方法进行系统性分析，以讨论其在不同临床场景中的适用性与表现差异。

在6种医学影像数据集上，各类异常检测方法展现出显著的性能差异。基于重建的方法（RD4AD、AE-FLOW、EDC和EA2D）在多数数据集上保持领先地位。其中，EA2D作为该范式的最新代表，在OCT2017、APTOS、Br35H和Brain Tumor数据集上均取得了最优或接近最优的检测性能，AUC分别达到99.84%、97.05%、99.83%和97.66%，ACC也相应维持在较高水平。RD4AD和EDC同样表现优异，在OCT2017和Br35H数据集上AUC均超过99%，表明该类方法得益于预训练模型提供的丰富语义特征，能够有效捕捉医学图像中的异常模式。然而，在更具挑战性的胸部X光片（RSNA）检测任务中，该类方法的性能出现显著下降，EA2D方法的AUC从OCT2017的99.84%降至91.55%，下降幅度达8.29个百分点；AE-FLOW的AUC更是从98.02%降至81.94%，这一现象说明该方法对具有复杂背景噪声和组织结构重叠的影像仍然敏感度不足，难以在密集的正常解剖结构中准确识别出细微的病理改变。

基于自监督学习的方法在多个数据集上表现出稳健的检测能力。PatchCL作为该范式的最新代表，在OCT2017、Br35H和Brain Tumor数据集上AUC分别达到99.29%、99.56%和97.68%，与基于重建的方法性能相当。在RSNA数据集上，PatchCL的AUC达到89.42%，显著优于多数其他方法，显示出良好的泛化能力。PMSACL和SALAD虽然在部分数据集上性能略逊，但在复杂场景下仍保持相对稳定的表现，表明自监督学习通过预文本任务学习到的特征表示具有较强的迁移性和鲁棒性。

基于合成的方法在不同模态间表现出较大的性能波动。SimpleNet作为该类方法的代表，在Br35H和Brain Tumor数据集上AUC分别达到98.88%和96.42%，展现出优异的检测能力；但在RSNA数据集上AUC仅为78.18%，与最优方法相差超过13个百分点。RealNet和GLASS同样呈现出类似的性能波动特征，在OCT2017和Br35H数据集上表现优异，但在APTOS和RSNA数据集上性

能明显下降。这表明基于合成的方法高度依赖于合成与真实异常之间的分布匹配程度，当合成策略与目标域异常模式存在差异时，模型性能将受到显著影响。

基于记忆库匹配的方法在特定场景下展现出独特优势。SQUID 方法在 RSNA 数据集上 AUC 达到 90.48%，ACC 为 88.41%，在胸部 X 光片这类复杂影像上表现优于多数其他方法，体现了记忆库机制通过存储正常模式的原型特征，能够有效防止过度重建问题，并在复杂背景中保持较高的检测精度。PaDiM 和 P-VQ 方法在 OCT2017 和 Br35H 等数据集上也取得了良好性能，其中 P-VQ 在 OCT2017 数据集上 AUC 达 99.20%，ACC 达 98.94%，表明记忆库匹配方法在特定模态下具备与主流方法竞争的能力。

从数据集特性角度分析，各类方法在 Br35H 和 OCT2017 这类病灶特征相对明显的数据集上普遍表现优异，多数方法 AUC 超过 95%；而在 ISIC2018 和 RSNA 这类具有复杂背景、病灶边界模糊或组织结构重叠的数据集上，检测性能普遍下降，方法间差异也更为显著。特别是在 RSNA 数据集上，EA2D、PatchCL 和 SQUID 分别以 91.55%、89.42% 和 90.48% 的 AUC 位居前列，传统方法如 AE 和 VAE 的 AUC 仅为 66.44% 和 68.64%，凸显了先进方法在复杂临床场景中的必要性。

综合来看，基于重建的方法凭借其强大的语义表征能力，在多数场景下保持性能领先；自监督学习方法以其良好的泛化性，在复杂数据集上展现出稳定表现；异常合成方法虽在部分模态表现优异，但跨域泛化能力仍需提升；记忆库匹配方法则展现出在特定场景下的价值。这一性能分布特征表明，方法的选择需综合考虑目标数据集的模态特性、病灶特征及临床需求，而非单一范式的简单套用。

3.2.2　医学异常检测方法分析

本文依据核心检测机制将现有方法划分为重建、合成、自监督学习与记忆库匹配 4 个主要类别，旨在从逻辑上提供一个清晰的结构化技术认知框架。然而值得注意的是，随着该领域研究的逐渐深入，前沿方法往往不再严格遵循单一范式的设计思路，而是在实现层面呈现出明显的交叉互融趋势。其核心原因在于各类方法固有的优势与局限性，促使研究者通过机制组合以实现性能互补。基于重建的范式因其架构的灵活性和优异的像素级定位能力，常被视为一个强大的基础框架；合成、自监督学习与记忆库匹配等方法，在高级语义表征学习与判别性特征约束方面的独特优势，往往在构建更具泛化能力的复杂特征提取与语义理解系统中发挥着关键作用。

具体而言，当前研究主要体现出以下几种主要的融合模式。重建范式与记忆库匹配范式的集成旨在解决过度重建问题，其技术核心是利用记忆库对潜空间进行离散化约束。代表性方法 MemAE 通过在自编码器瓶颈层引入可寻址记忆模块，迫使解码器仅能基于存储的正常原型进行重建；PatchCore 则通过构建图像级特征记忆库，将异常检测转化为特征匹配问题。重建与自监督范式的融合则以自监督作为特征表示终端，如 SALAD 框架通过多种图像扰动预文本任务增强编码器的语义建模能力；PatchCL-AE 通过局部块对比学习强化自编码器对正常语义一致性的理解；EDC 方法则引入编码器–解码器对比学习以安全微调预训练模型。合成范式与自监督/重建范式的结合利用异常合成构建代理任务，如 CutPaste 通过图像块裁剪粘贴合成异常并构建自监督分类任务；PMSACL 结合多中心损失与医学图像增强策略，在特征空间实现伪多类异常识别。此外，合成、自监督学习与记忆库匹配三类方法之间亦存在直接协同，如 PaDiM 方法将自监督预训练特征与参数化记忆库结合，SQUID 方法通过合成异常优化记忆匹配机制。

从上述范例来看，医学影像无监督异常检测方法已呈现出从单一机制探索向多机制协同设计的演进路径，这一趋势也说明了未来的性能突破将更依赖于对不同机制协同作用的深入理解与系统化整合。

4　挑战与展望

4.1　研究领域主要挑战

近年来，基于深度学习的医学影像异常检测方法已取得了显著进展，涌现出多种有效的检测范式。然而，现阶段该领域仍存在一些关键挑战，值得此后的研究进一步探索以突破现有方法的性能瓶颈。

基于重建的方法由于其端到端的检测特性、模型稳定性及多尺度特征学习带来的异常定位能力，在医学领域具备较为显著的优势，但在重建误差度

量准则的问题上仍存在争议。数据重建方法虽能保留丰富的低级细节，利于小病灶定位，但易受图像噪声干扰且存在过泛化风险；特征重建方法虽能有效缓解过度重建问题，但因特征图分辨率降低可能导致细小异常定位能力下降。现有距离函数如均方误差、感知误差、SSIM等，分别适用于不同数据集但通用性不足。这表明未来的研究仍需开发可针对不同任务进行自适应调整的度量方法，从而减少异常分布模式不确定性带来的影响。

大规模数据集预训练模型目前在基于自监督学习与特征重建的UAD方法中已得到广泛应用，但由于自然图像与多模态医学图像间存在显著的语义差异，预训练模型的性能仍受其跨域适应能力的制约。基于自监督学习的方法虽能学习到鲁棒的特征表示，但其对全局信息的偏重使其在面对细微病变时敏感性不足，需通过局部对比学习等策略加以强化。冻结参数这一常见手段往往难以最大程度地发挥预训练模型的表征能力，而进行全面微调则极易引发模型的灾难性遗忘问题，即模型在学习新任务的过程中完全丧失先前学习到的通用视觉知识，导致异常判别性能急剧下降甚至完全失效，继而导致模型对异常模式的判别性能完全失效。因此，如何针对目标数据集的样本分布稳定迁移预训练特征仍然是一个亟待解决的问题。

在医学任务中，基于合成与记忆库匹配的方法则因其固有的高时空复杂度而存在局限性。基于合成的方法高度依赖合成与真实异常之间的分布匹配程度，当合成策略与目标域异常模式存在差异时，模型性能将受到显著影响。基于记忆库匹配的方法虽能通过存储正常原型有效防止过度重建，但其时空复杂度高、原型选取策略缺乏统一标准，且泛化能力在不同模态间波动较大。此外，两类方法的模型结构复杂、推断流程烦琐，记忆库的原型匹配过程还需多次迭代采样，导致推理时延高，难以满足临床实时性需求，且高度依赖训练数据的质量与分布完整性。这些局限性使它们在临床场景中面临可靠性不足与诊断效率受限等问题。

现阶段的研究多集中于图像级异常检测任务，而对像素级异常分割的探索仍相对有限。该局限主要源于以下几方面因素：首先，像素级标注数据稀缺，尤其对于分辨率高、异常边界模糊的医学异常而言；其次，异常区域往往在整幅图像中占比极小且分布稀疏，导致模型难以捕捉其细微特征。此外，当前的多种主流范式多依赖于全局特征编码与分类，此类学习模式在密集预测任务中的表现通常不及以重建为代表的多尺度学习方法，后者能够更好地保留细节信息与空间关系，故而更适用于像素级定位。因此，未来的探索方向可扩展到对多样化和细微异常的分割，以增强UAD技术在真实临床场景中的适用性。

4.2 未来研究方向与展望

除上述挑战所指示的探索方向外，未来的研究重点还可能包括如下方向。

1) 视觉语言模型。VLM技术[116]通过大规模图像–文本预训练，具备跨模态语义感知能力，为解决医学UAD中标注稀缺、异常多样等挑战提供了新思路。CLIP等模型通过构建“语义偏离”判别边界，实现了不依赖精确分布拟合的异常判定，展现出更强的泛化能力。尽管CLIP在医学领域仍面临跨域语义差异问题，但已有研究尝试通过扩散模型进行域适应转换[117]或通过局部对齐机制增强医学语义感知[118]来缓解这一问题。如何突破预训练模型的域迁移瓶颈，充分发挥VLM技术在医学异常检测中的潜力，是值得深入探索的方向。

2) 元学习。元学习通过从大量相关任务中提取共享知识，使模型获得快速适应新任务的能力。在异常检测领域，已有研究将其应用于解决数据分布偏移和新类型异常识别等挑战，如工业图像中的AnoDual[119]和农业数据中的MTL框架[120]。然而，面向医学影像特化的元学习应用仍较欠缺。未来可着力构建适配医学特点的元学习框架，包括发展跨模态元学习以利用多模态数据的互补信息，以及设计面向像素级分割的元训练机制以解决异常定位中的细节保留与空间建模难题。

3) 3D影像异常检测。现有方法多针对2D数据设计，从3D体积中截取切片分析会丢失丰富的解剖上下文信息，无法实现精准的体素级定位。此外，3D检测还面临计算复杂度高、网络深度受限、标注数据集稀缺等技术瓶颈。目前已有直接架构扩展[121]、联合2D-3D表示学习[122]、隐式场学习[123]等初步探索。未来需构建支持2D与3D任务对比的基准数据集，开发低复杂度新型3D网络架构，探索体积扩散模型、跨平面注意力等前沿技术。

5　结束语

本文系统性地梳理了无监督异常检测技术在医学影像领域的发展现状，首先按照检测机制将现有方法划分为4类范式，并详细阐述了每个类别的基本原理、代表性方法及其特点；其次，本文从医学异常检测数据集、异常评分计算方法及量化指标3个方面为检测模型的性能评价提供基准参照；此外，本文通过在多模态数据集上的定量对比实验，系统评估了各类方法的图像级异常检测性能与计算效率，揭示了不同范式在临床适用场景上的优势与局限；最后，本文对现有方法存在的挑战进行了分析，并对未来的发展方向进行了展望。

总体而言，本文建立的方法分类体系、性能分析与未来展望能够为医学异常检测研究人员提供清晰的技术脉络与发展方向，助力开发出更高效、鲁棒且具备临床实用价值的异常检测系统，最终为疾病早期筛查、精准诊断及健康管理提供可靠的智能辅助支持。

参考文献：

[1] Bergmann P, Fauser M, Sattlegger D, et al. MVTec AD: a comprehensive real-world dataset for unsupervised anomaly detection[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 9584-9592.

[2] 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.
Wang W P, Qin Y C, Shi W X. Review of unsupervised deep learning methods for industrial defect detection[J]. Journal of Computer Applications, 2025, 45(5): 1658-1670.

[3] 尚书一, 李宏佳, 宋晨, 等. 互联网服务场景下基于机器学习的KPI异常检测综述[J]. 计算机研究与发展, 2025, 62(1): 207-231.
Shang S Y, Li H J, Song C, et al. Survey of machine learning-based KPI anomaly detection on Internet-based services[J]. Journal of Computer Research and Development, 2025, 62(1): 207-231.

[4] Liu W, Luo W X, Lian D Z, et al. Future frame prediction for anomaly detection-a new baseline[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6536-6545.

[5] 张振宇, 綦小龙, 肖琪, 等. 视频异常检测技术综述[J]. 激光杂志, 2026, 47(2): 1-15.
Zhang Z Y, Qi X L, Xiao Q, et al. A survey on video anomaly detection techniques[J]. Laser Journal, 2026, 47(2): 1-15.

[6] 张赛男, 孙彪. 基于机器学习的网络异常检测方法综述[J]. 吉林大学学报(信息科学版), 2021, 39(6): 732-742.
Zhang S N, Sun B. Research on network anomaly detection method basedon machine learning[J]. Journal of Jilin University (Information Science Edition), 2021, 39(6): 732-742.

[7] Ruff L, Vandermeulen R A, Goernitz N, et al. Deep one-class classification[C]//Proceedings of the 35th International Conference on Machine Learning. New York: PMLR, 2018: 4393-4402.

[8] Hido S, Tsuboi Y, Kashima H, et al. Statistical outlier detection using direct density ratio estimation[J]. Knowledge and Information Systems, 2011, 26(2): 309-336.

[9] Rousseeuw P J, Hubert M. Robust statistics for outlier detection[J]. WIREs Data Mining and Knowledge Discovery, 2011, 1(1): 73-79.

[10] Knorr E M, Ng R T, Tucakov V. Distance-based outliers: algorithms and applications[J]. The VLDB Journal the International Journal on Very Large Data Bases, 2000, 8(3/4): 237-253.

[11] Angiulli F, Basta S, Pizzuti C. Distance-based detection and prediction of outliers[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(2): 145-160.

[12] Breunig M M, Kriegel H P, Ng R T, et al. LOF: identifying density-based local outliers[C]//Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2000: 93-104.

[13] Yang X W, Latecki L J, Pokrajac D. Outlier detection with globally optimal exemplar-based GMM[C]//Proceedings of the 2009 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2009: 145-154.

[14] Al-Zoubi M B. An effective clustering-based approach for outlier detection[J]. European Journal of Scientific Research, 2009, 28(2): 310-316.

[15] Akcay S, Atapour-Abarghouei A, Breckon T P. GANomaly: semi-supervised anomaly detection via adversarial training[C]//Computer Vision-ACCV 2018. Berlin: Springer, 2019: 622-637.

[16] Schlegl T, Seeböck P, Waldstein S M, et al. F-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks[J]. Medical Image Analysis, 2019, 54: 30-44.

[17] Deng H Q, Li X Y. Anomaly detection via reverse distillation from one-class embedding[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 9727-9736.

[18] Tang P, Yan X X, Hu X B, et al. Anomaly detection in medical images using encoder-attention-2Decoders reconstruction[J]. IEEE Transactions on Medical Imaging, 2025, 44(8): 3370-3382.

[19] Li C L, Sohn K, Yoon J, et al. CutPaste: self-supervised learning for anomaly detection and localization[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 9659-9669.

[20] Tan J, Hou B, Day T, et al. Detecting outliers with Poisson image interpolation[C]//Proceedings of the Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. New York: ACM Press, 2021: 581-591.

[21] Zavrtanik V, Kristan M, Skočaj D. DRÆM-A discriminatively trained reconstruction embedding for surface anomaly detection[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 8310-8319.

[22] Liu Z K, Zhou Y M, Xu Y S, et al. SimpleNet: a simple network for im-

age anomaly detection and localization[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 20402-20411.

[23] Sohn K, Li C L, Yoon J, et al. Learning and evaluating representations for deep one-class classification[PP]. arXiv (2020-11-04) [2025-11-01]. arXiv: arXiv.2011.02578.

[24] Zhao H, Li Y X, He N J, et al. Anomaly detection for medical images using self-supervised and translation-consistent features[J]. IEEE Transactions on Medical Imaging, 2021, 40(12): 3641-3651.

[25] Tian Y, Liu F B, Pang G S, et al. Self-supervised pseudo multi-class pre-training for unsupervised anomaly detection and segmentation in medical images[J]. Medical Image Analysis, 2023, 90: 102930.

[26] Gong D, Liu L Q, Le V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 1705-1714.

[27] Defard T, Setkov A, Loesch A, et al. PaDiM: a patch distribution modeling framework for anomaly detection and localization[C]//Pattern Recognition. ICPR International Workshops and Challenges. Berlin: Springer, 2021: 475-489.

[28] Kim T, Lee Y G, Jeong I, et al. Patch-wise vector quantization for unsupervised medical anomaly detection[J]. Pattern Recognition Letters, 2024, 184: 205-211.

[29] Baur C, Denner S, Wiestler B, et al. Autoencoders for unsupervised anomaly segmentation in brain MR images: a comparative study[J]. Medical Image Analysis, 2021, 69: 101952.

[30] 侯金磊. 基于卷积自编码器的脑MRI无监督异常检测算法研究[D]. 杭州: 浙江大学, 2021.
Hou J L. Research on brain MRI Unsupervised anomaly detection algorithm based on convolutional autoencoder[D]. Hangzhou: Zhejiang University, 2021.

[31] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[32] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 2: 2672-2680.

[33] Kingma D P, Welling M. Auto-encoding variational Bayes[PP]. arXiv (2013-12-20) [2025-11-01]. arXiv:arXiv.1312.6114.

[34] Milković F, Filipović B, Subasić M, et al. Ultrasound anomaly detection based on variational autoencoders[C]//Proceedings of the 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA). Piscataway: IEEE Press, 2021: 225-229.

[35] Makhzani A, Shlens J, Jaitly N, et al. Adversarial autoencoders[PP]. arXiv (2015-11-18) [2025-11-01]. arXiv:arXiv.1511.05644.

[36] Zhang H B, Guo W P, Zhang S Q, et al. Unsupervised deep anomaly detection for medical images using an improved adversarial autoencoder[J]. Journal of Digital Imaging, 2022, 35(2): 153-161.

[37] Meissen F, Wiestler B, Kaissis G, et al. On the pitfalls of using the residual error as anomaly score[C]//Proceedings of the 5th International Conference on Medical Imaging with Deep Learning. New York: PMLR, 2022: 914-928.

[38] Bergmann P, Löwe S, Fauser M, et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders[PP]. arXiv (2018-07-05) [2025-11-01]. arXiv:arXiv.1807.02011.

[39] Behrendt F, Bhattacharya D, Maack L, et al. Diffusion models with ensembled structure-based anomaly scoring for unsupervised anomaly detection[C]//Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI). Piscataway: IEEE Press, 2024: 1-4.

[40] Shvetsova N, Bakker B, Fedulova I, et al. Anomaly detection in medical imaging with deep perceptual autoencoders[J]. IEEE Access, 2021, 9: 118571-118583.

[41] Bercea C I, Rueckert D, Schnabel J A. What do AEs learn? Challenging common assumptions in unsupervised anomaly detection[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2023. Berlin: Springer, 2023: 304-314.

[42] Zimmerer D, Kohl S A A, Petersen J, et al. Context-encoding variational autoencoder for unsupervised anomaly detection[PP]. arXiv (2018-12-14) [2025-11-01]. arXiv:arXiv.1812.05941.

[43] Mao Y F, Xue F F, Wang R X, et al. Abnormality detection in chest X-ray images using uncertainty prediction autoencoders[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Berlin: Springer, 2020: 529-538.

[44] Zhou X Y, Niu S J, Li X H, et al. Spatial-contextual variational autoencoder with attention correction for anomaly detection in retinal OCT images[J]. Computers in Biology and Medicine, 2023, 152: 106328.

[45] Schlegl T, Seeböck P, Waldstein S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[C]//Information Processing in Medical Imaging. Berlin: Springer, 2017: 146-157.

[46] Baur C, Wiestler B, Albarqouni S, et al. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Berlin: Springer, 2019: 161-169.

[47] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning -Volume 70. New York: ACM Press, 2017: 214-223.

[48] Han C, Rundo L, Murao K, et al. MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction[J]. BMC Bioinformatics, 2021, 22(S2): 31.

[49] Li G L, Zou Y X, Liu B Y, et al. PAttL-GAN: pixel attention localization generative adversarial network for unsupervised anomaly detection in medical images[J]. Neurocomputing, 2025, 654: 131345.

[50] Kascenas A, Pugeault N, O'Neil A Q. Denoising autoencoders for unsupervised anomaly detection in brain MRI[C]//Proceedings of the International Conference on Medical Imaging with Deep Learning. New York: PMLR, 2022: 653-664.

[51] Kascenas A, Sanchez P, Schrempf P, et al. The role of noise in denoising models for anomaly detection in medical images[J]. Medical Image Analysis, 2023, 90: 102963.

[52] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.

[53] Behrendt F, Bhattacharya D, Krüger J, et al. Patched diffusion models

for unsupervised anomaly detection in brain MRI[C]//Proceedings of the Medical Imaging with Deep Learning. New York: PMLR, 2024: 1019-1032.

[54] Bi Y, Huang L, Clarenbach R, et al. Synomaly noise and multi-stage diffusion: a novel approach for unsupervised anomaly detection in medical images[PP]. arXiv (2024-11-06) [2025-11-01]. arXiv: arXiv.2411.04004.

[55] Zhou B L, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 2921-2929.

[56] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 618-626.

[57] Zimmerer D, Isensee F, Petersen J, et al. Unsupervised anomaly localization using variational auto-encoders[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2019. Berlin: Springer, 2019: 289-297.

[58] Liu W Q, Li R Z, Zheng M, et al. Towards visually explaining variational autoencoders[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 8639-8648.

[59] Silva-Rodríguez J, Naranjo V, Dolz J. Constrained unsupervised anomaly segmentation[J]. Medical Image Analysis, 2022, 80: 102526.

[60] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[61] Rippel O, Mertens P, Merhof D. Modeling the distribution of normal data in pre-trained deep features for anomaly detection[C]//Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE Press, 2021: 6726-6733.

[62] You Z Y, Yang K, Luo W H, et al. ADTR: anomaly detection transformer with feature reconstruction[C]//Neural Information Processing. Berlin: Springer, 2023: 298-310.

[63] Chen L Y, You Z Y, Zhang N, et al. UTRAD: anomaly detection and localization with U-Transformer[J]. Neural Networks, 2022, 147: 53-62.

[64] You Z Y, Cui L, Shen Y J, et al. A unified model for multi-class anomaly detection[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM Press, 2022: 4571-4584.

[65] Lu S, Zhang W H, Zhao H, et al. Anomaly detection for medical images using heterogeneous auto-encoder[J]. IEEE Transactions on Image Processing, 2024, 33: 2770-2782.

[66] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[PP]. arXiv (2015-03-09) [2025-11-01]. arXiv:arXiv.1503.02531.

[67] Tien T D, Nguyen A T, Tran N H, et al. Revisiting reverse distillation for anomaly detection[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 24511-24520.

[68] Bergmann P, Fauser M, Sattlegger D, et al. Uninformed students: student-teacher anomaly detection with discriminative latent embeddings[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 4182-4191.

[69] Wang G D, Han S M, Ding E R, et al. Student-teacher feature pyramid matching for anomaly detection[PP]. arXiv (2021-03-07) [2025-11-01]. arXiv:arXiv.2103.04257.

[70] Salehi M, Sadjadi N, Baselizadeh S, et al. Multiresolution knowledge distillation for anomaly detection[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 14897-14907.

[71] Zhao Y Z, Ding Q Q, Zhang X Q. AE-FLOW: autoencoders with normalizing flows for medical images anomaly detection[C]//The Eleventh International Conference on Learning Representations. Kigali: 2023.

[72] Liu M X, Jiao Y R, Lu J Q, et al. Anomaly detection for medical images using teacher-student model with skip connections and multiscale anomaly consistency[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 2520415.

[73] Ge C K, Yu X J, Zheng H, et al. ESC-DRKD: enhanced skip connection-based direct reverse knowledge distillation for medical image anomaly detection[J]. Neurocomputing, 2025, 651: 130994.

[74] Rahmaniar W, Suzuki K. Multi-AD: cross-domain unsupervised anomaly detection for medical and industrial applications[J]. Pattern Recognition, 2026, 172: 112486.

[75] Guo J, Lu S, Jia L Z, et al. Encoder-decoder contrast for unsupervised anomaly detection in medical images[J]. IEEE Transactions on Medical Imaging, 2024, 43(3): 1102-1112.

[76] Shi Y, Yang J, Qi Z Q. Unsupervised anomaly segmentation via deep feature reconstruction[J]. Neurocomputing, 2021, 424: 9-22.

[77] Meissen F, Paetzold J, Kaissis G, et al. Unsupervised anomaly localization with structural feature-autoencoders[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Berlin: Springer, 2023: 14-24.

[78] Sato J, Suzuki Y, Wataya T, et al. Anatomy-aware self-supervised learning for anomaly detection in chest radiographs[J]. IScience, 2023, 26(7): 107086.

[79] Tan J, Hou B, Batten J, et al. Detecting outliers with foreign patch interpolation[PP]. arXiv (2020-11-09) [2025-11-01]. arXiv: arXiv.2011.04197.

[80] Müller J P, Baugh M, Tan J, et al. Confidence-aware and self-supervised image anomaly localisation[C]//Uncertainty for Safe Utilization of Machine Learning in Medical Imaging. Berlin: Springer, 2023: 177-187.

[81] Schlüter H M, Tan J, Hou B, et al. Natural synthetic anomalies for self-supervised anomaly detection and localization[C]//Computer Vision-ECCV 2022. Berlin: Springer, 2022: 474-489.

[82] Baugh M, Tan J, Müller J P, et al. Many tasks make light work: learning to localise medical anomalies from multiple synthetic tasks[C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2023. Berlin: Springer, 2023: 162-172.

[83] Zhang X, Li S Y, Li X, et al. DeSTSeg: segmentation guided denoising student-teacher for anomaly detection[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 3914-3923.

[84] Cai Y X, Liang D K, Luo D L, et al. A discrepancy aware framework for robust anomaly detection[J]. IEEE Transactions on Industrial Informatics, 2024, 20(3): 3986-3995.

[85] Zhang X M, Xu M, Zhou X Z. RealNet: a feature selection network with realistic synthetic anomaly for anomaly detection[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 16699-16708.

[86] Chen Q Y, Luo H Y, Lv C K, et al. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization[C]//Computer Vision-ECCV 2024. Berlin: Springer, 2024: 37-54.

[87] Cai Y, Chen H, Yang X, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images[J]. Medical Image Analysis, 2023, 86: 102794.

[88] Chen X L, He K M. Exploring simple Siamese representation learning[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 15745-15753.

[89] Srivastava N, Mansimov E, Salakhutdinov R. Unsupervised learning of video representations using LSTMs[C]//Proceedings of the 32nd International Conference on Machine Learning-Volume 37. New York: ACM Press, 2015: 843-852.

[90] Hjelm R D, Fedorov A, Lavoie-Marchildon S, et al. Learning deep representations by mutual information estimation and maximization[PP]. arXiv (2018-08-20) [2025-11-01]. arXiv:arXiv.1808.06670.

[91] Oord A V D, Li Y Z, Vinyals O. Representation learning with contrastive predictive coding[PP]. arXiv (2018-07-11) [2025-11-01]. arXiv: arXiv.1807.03748.

[92] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//International Conference on Machine Learning. New York: PMLR, 2020: 1597-1607.

[93] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 9726-9735.

[94] Grill J B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning[J]. Advances in Neural Information Processing Systems, 2020, 33: 21271-21284.

[95] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution[J]. Neural Computation, 2001, 13(7): 1443-1471.

[96] Tax D M J, Duin R P W. Support vector data description[J]. Machine Learning, 2004, 54(1): 45-66.

[97] Tian Y, Pang G S, Liu F B, et al. Constrained contrastive distribution learning for unsupervised anomaly detection and localisation in medical images[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Berlin: Springer, 2021: 128-140.

[98] Lu S, Zhang W H, Guo J, et al. PatchCL-AE: anomaly detection for medical images using patch-wise contrastive learning-based auto-encoder[J]. Computerized Medical Imaging and Graphics, 2024, 114: 102366.

[99] Reiss T, Hoshen Y. Mean-shifted contrastive loss for anomaly detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(2): 2155-2162.

[100] Roth K, Pemula L, Zepeda J, et al. Towards total recall in industrial anomaly detection[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 14298-14308.

[101] Zhou K, Li J, Luo W X, et al. Proxy-bridged image reconstruction network for anomaly detection in medical images[J]. IEEE Transactions on Medical Imaging, 2022, 41(3): 582-594.

[102] Xiang T G, Zhang Y X, Lu Y Y, et al. SQUID: deep feature inpainting for unsupervised anomaly detection[PP]. arXiv (2021-11-26) [2025-11-01]. arXiv:arXiv.2111.13495.

[103] Zhou K, Li J, Xiao Y T, et al. Memorizing structure-texture correspondence for image anomaly detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(6): 2335-2349.

[104] Xiao Y F, Huang X T, Liang W, et al. Medical images anomaly detection for imbalanced datasets with multi-scale normalizing flow[J]. Computer Science and Information Systems, 2025, 22(1): 219-238.

[105] Stein A, Wu C, Carr C, et al. RSNA pneumonia detection challenge[DS]. [2025-11-01].

[106] Nguyen H Q, Lam K, Le L T, et al. VinDr-CXR: an open dataset of chest X-rays with radiologist's annotations[J]. Scientific Data, 2022, 9: 429.

[107] Hamada A. Br35H: brain tumor detection 2020 kaggle[DS]. [2025-11-01].

[108] Nickparvar M. Brain tumor MRI dataset[DS]. [2025-11-01].

[109] Kermany D S, Goldbaum M, Cai W J, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning[J]. Cell, 2018, 172(5): 1122-1131.

[110] Li L, Xu M, Wang X F, et al. Attention based glaucoma detection: a large-scale database and CNN model[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 10563-10572.

[111] Karthik, Maggie, Dane S. APTOS 2019 blindness detection[DS]. [2025-11-01].

[112] Codella N, Rotemberg V, Tschandl P, et al. Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC) [PP]. arXiv (2019-02-10) [2025-11-01]. arXiv:arXiv.1902.03368.

[113] Bejnordi B E, Veta M, Diest P J V, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer[J]. Jama, 2017, 318(22): 2199-2210.

[114] Zhao R Y, Yaman B, Zhang Y X, et al. FastMRI+, clinical pathology annotations for knee and brain fully sampled magnetic resonance imaging data[J]. Scientific Data, 2022, 9: 152.

[115] Liew S L, Lo B P, Donnelly M R, et al. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms[J]. Scientific Data, 2022, 9: 320.

[116] Gao H H, Jiang W Y, Ran Q, et al. Vision-language interaction via contrastive learning for surface anomaly detection in consumer electronics manufacturing[J]. IEEE Transactions on Consumer Electronics, 2024, 70(3): 6119-6130.

[117] Chen Y H, Tao H K, Yang Z, et al. Diffusion-based vision-language model for zero-shot anomaly detection in medical images[J]. Engineering Applications of Artificial Intelligence, 2025, 161: 112181.

[118] Liu Y Y, Li Q Y, Wang Z H, et al. LECLIP: boosting zero-shot anomaly detection with local enhanced CLIP[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 5034111.

[119] Yao M Y, Tao D, Qi P, et al. Rethinking discrepancy analysis: anomaly detection via meta-learning powered dual-source representation differentiation[J]. IEEE Transactions on Automation Science and Engineering, 2025, 22: 8579-8592.

[120] García R, Aguilar J. A meta-learning approach in a cattle weight identification system for anomaly detection[J]. Computers and Electronics in Agriculture, 2024, 217: 108572.

[121] Viana J S, Rosa E D L, Vyvere T V, et al. Unsupervised 3D brain anomaly detection[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Berlin: Springer, 2021: 133-142.

[122] Kang I, Park J. Joint embedding of 2D and 3D networks for medical image anomaly detection[PP]. arXiv (2022-12-21) [2025-11-01]. arXiv:arXiv.2212.10939.

[123] Marimont S N, Tarroni G. Implicit field learning for unsupervised anomaly detection in medical images[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Berlin: Springer, 2021: 189-198.

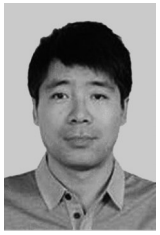
[作者简介]

赵映程（1993– ），男，甘肃平凉人，博士，西安理工大学讲师，主要研究方向为计算机视觉、医学影像处理。

宋霄罡（1987– ），男，河南漯河人，博士，西安理工大学教授、博士生导师，主要研究方向为计算机视觉、无人系统自主导航。

石争浩（1968– ），男，陕西渭南人，博士，西安理工大学教授、博士生导师，主要研究方向为机器视觉、医学图像处理及机器学习。

尤珍臻（1989– ），女，陕西西安人，博士，西安理工大学讲师，主要研究方向为生物医学图像处理、机器学习。



黑新宏（1976– ），男，陕西延安人，博士，西安理工大学教授、博士生导师，主要研究方向为智能系统与安全关键系统、人工智能、大数据及其在轨道交通等系统中的应用。