

医学图像异常检测方法综述

1. 引言

1.1 研究背景和重要性

随着医学影像技术的迅猛发展，医学图像在疾病诊断、治疗规划和预后评估中发挥着至关重要的作用。从 X 光、计算机断层扫描 (CT)、磁共振成像 (MRI) 到超声和视网膜眼底成像，医学影像为临床医生提供了丰富的病灶信息。然而，医学图像中的异常区域检测一直是一个具有挑战性的任务，尤其是在识别罕见疾病和进行健康筛查时。异常检测 (Anomaly Detection, AD) 作为一种重要的机器学习方法，旨在识别偏离预期正常模式的异常样本。相比于传统的监督学习方法，AD 方法通常仅依赖于大量的正常样本进行训练，而无需依赖于异常样本，这使其在医学领域中具有显著的应用优势。

根据美国食品药品监督管理局的数据，超过 7000 种罕见疾病影响了超过 3000 万人，每种疾病单独影响的人数较少（约 0.06% 的人口）。由于这些疾病的稀有性和多样性，收集足够的异常样本进行训练成为不现实的任务，这进一步凸显了 AD 方法在医学图像分析中的重要性。传统的监督学习方法需要大量标注的异常样本进行训练，而在现实中，异常样本的获取和标注成本高且耗时，且许多异常类型的样本极为稀缺甚至不可用。因此，AD 方法在医学图像中的应用不仅具有理论意义，更具备实际价值，能够显著提升临床诊断的效率和准确性，尤其是在早期识别和诊断罕见疾病方面。

1.2 目的

本综述旨在系统性地总结和比较当前医学图像异常检测领域中的主要方法和技术。通过分析不同研究者提出的多种 AD 方法，评估其在各种医学图像模态上的表现，揭示现有方法的优缺点，并探讨未来研究的潜在方向。综述的主要目的是为研究人员提供一个全面的参考框架，促进医学 AD 领域的进一步发展和创新。

1.3 文献选择的标准和范围

本文的文献选择基于以下标准：首先，选取在医学 AD 领域具有代表性和影响力的研究工作，包括重构方法、自监督学习方法和特征参考方法等。其次，选取涵盖多种医学图像模态的研究，如胸部 X 光、脑部 MRI、视网膜眼底图像、皮肤镜图像和组织病理全片图像。最后，优先选择最近几年内发表的高质量研究，以确保综述内容的前沿性和实用性。

2. 问题引入

医学图像异常检测领域面临诸多挑战。首先，异常样本的稀缺性和多样性使得传统的监督学习方法难以应用，因其依赖大量标注的异常样本进行训练。其次，现有 AD 方法在不同数据集和模态上的性能差异显著，缺乏统一的基准和评估标准，导致结果的可比性和可重复性不足。此外，如何有效地处理高维度、复杂结构的医学图像，以及如何提升模型在实际临床环境中的鲁棒性和解释性，仍然是亟待解决的问题。

解决上述问题不仅可以提升医学图像分析的准确性和效率，还能在早期识别和诊断罕见疾病方面发挥关键作用，进而改善患者的预后和生活质量。同时，建立统一的基准和评估框架，有助于推动 AD 方法的公平比较和持续改进，促进该领域的整体进步。

3. 方法总结

3.1 方法分类



图 1 医学异常检测方法分类

医学异常检测方法分类如图 1 所示，3.2 部分详细介绍了各个方法与其优缺点。

3.2 方法详细介绍

3.2.1 重构方法（Reconstruction-based Methods）

重构方法在医学 AD 中占据主导地位，其核心思想是利用生成模型（如 GANs、AEs 等）学习正常图像的分布，通过重构误差来识别异常区域。以下详细介绍图像重构和特征重构两种子类方法。

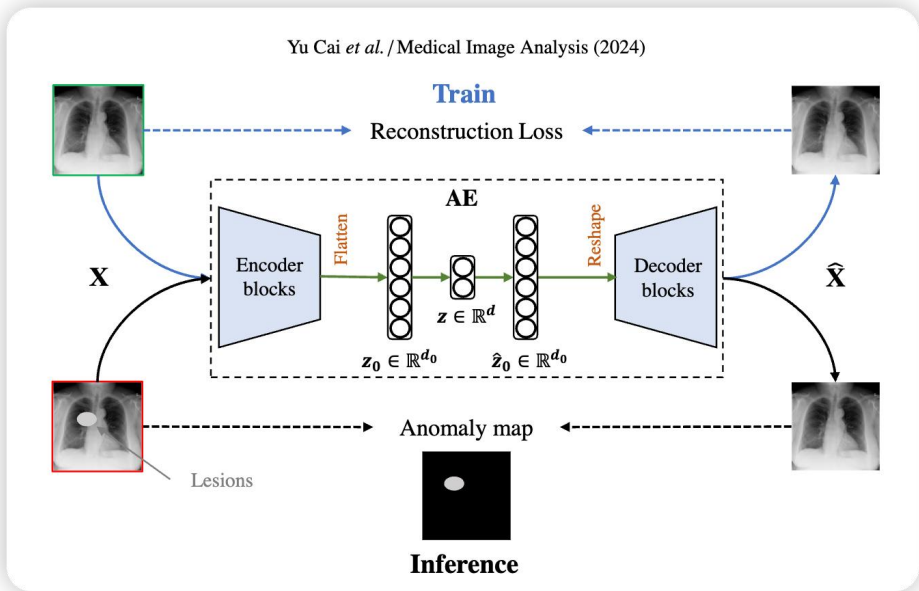


图 2 基于重构方法的异常检测流程

图像重构（Image-reconstruction）

图像重构方法直接在图像空间中进行重构和异常评分计算。典型的图像重构方法包括：

- ✧ **生成对抗网络（GANs）**：Schlegl 等人（2017, 2019）训练 GANs 来学习正常图像的流形，并通过重构误差和判别器中间特征的差异来检测异常。然而，GANs 的训练过程复杂且易受模式崩溃（mode collapse）的影响。

- ✧ **自编码器 (AEs) 及其变种**: 包括标准 AE、变分自编码器 (VAEs)、对抗自编码器 (AAE) 和向量量化 VAE (VQ-VAE)。这些方法通常具有稳定的训练过程, 但重构质量可能不尽如人意。为了提升重构质量, 一些研究引入了对抗训练和不同的距离函数 (如结构相似性指数测量 SSIM 和感知损失 Perceptual Loss)
- ✧ **对抗训练**: 通过引入判别器, 提升 AE 的重构质量。例如, Baur 等人 (2019) 和 Akcay 等人 (2019) 在 AE 中引入对抗训练, 改善重构结果的细节和一致性。
- ✧ **距离函数优化**: 采用 SSIM 或感知损失来替代传统的 L2 损失, 以更好地捕捉图像的结构和语义信息, 从而提升异常检测的效果。例如, Bergmann 等人 (2018) 使用 SSIM 度量重构误差, Shvetsova 等人 (2021) 使用感知损失来优化 AE 的重构过程。

然而, 图像重构方法存在两个主要问题: 一是模型可能对某些未见的异常区域进行良好重构, 导致假阴性; 二是由于网络容量有限, 正常区域的重构误差可能较高, 导致假阳性。为解决这些问题, 研究者提出了多种改进策略, 如记录正常训练样本的原型、引入噪声以限制模型对异常的重构能力等。

特征重构 (Feature-reconstruction)

与图像重构方法不同, 特征重构方法在高层特征空间中进行重构, 而非直接在图像空间中操作。具体步骤如下:

1. **特征提取**: 使用预训练的深度神经网络 (通常在 ImageNet 上预训练) 将输入图像映射到高层特征空间。
2. **特征重构**: 在特征空间中训练一个重构模型, 尽可能准确地重构正常图像的特征。
3. **异常检测**: 通过计算输入图像特征与重构特征之间的差异来评估异常程度。

特征重构方法的优势在于高层特征具有更强的判别能力, 能够更有效地捕捉异常特征。然而, 由于高层特征通常缺乏低级细节信息, 特征重构方法在局部异常的定位上可能存在一定的局限性。

代表性方法:

- ✧ **FAE-SSIM 和 FAE-MSE** (Meissen 等人, 2022a): 基于预训练的 ResNet18, 采用 SSIM 和 MSE 作为特征重构的损失函数, 分别评估重构特征的结构相似性和均方误差。

- ✧ **ReContrast** (Guo 等人, 2023a) : 通过对比学习进一步增强特征重构模型的判别能力, 提升异常检测的准确性。

3.2.2 自监督学习方法 (Self-supervised Learning-based Methods)

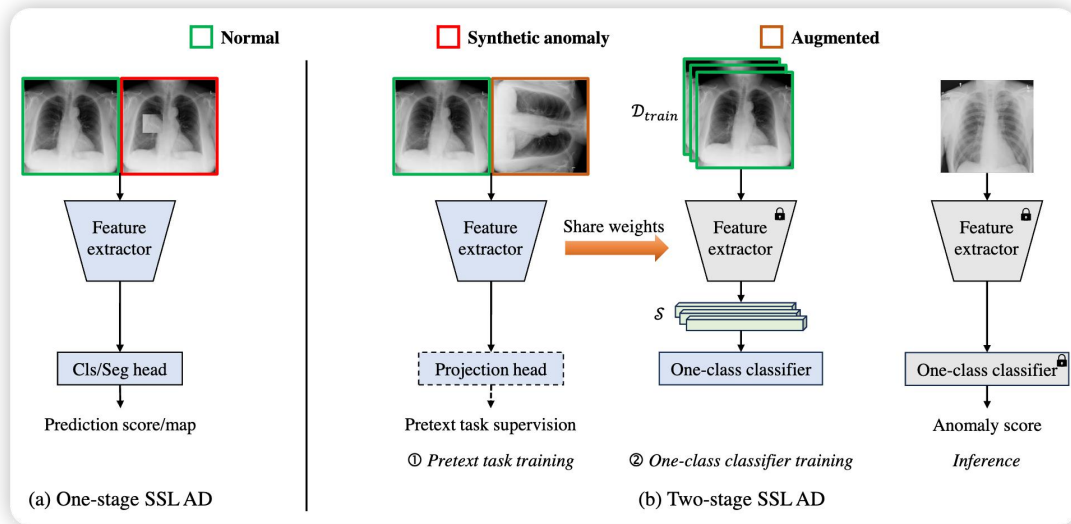


图 3 自监督异常检测的两种范式概述

自监督学习 (Self-supervised Learning, SSL) 通过设计预任务利用大量无标签数据进行训练, 已在各种机器学习任务中取得显著成果。在医学 AD 中, SSL 方法主要分为单阶段和双阶段两种范式。

单阶段方法 (One-stage Approach)

单阶段方法通过设计合成异常样本进行训练, 直接将训练好的模型应用于实际异常检测。其核心思想是设计逼真的伪异常, 使得模型在训练时能够学会识别这些异常, 从而在测试时能够泛化到真实异常样本。

代表性方法:

- ✧ **FPI (Foreign Patch Interpolation)** (Tan 等人, 2020) : 通过在正常图像中插入其他正常图像的随机补丁来生成合成异常样本。然而, 这种方法生成的异常与原图像存在较大不一致, 可能不利于模型泛化。

- ✧ **PII (Poisson Image Interpolation)** (Tan 等人, 2021) : 改进了 FPI, 通过使用泊松图像编辑技术, 使合成异常区域与原图像更为无缝融合, 提升了异常的真实性。
- ✧ **AnatPaste** (Sato 等人, 2023) : 针对胸部 X 光图像设计的解剖感知粘贴策略, 利用基于阈值的肺部分割掩码指导合成异常区域, 使得合成异常更符合解剖结构。

方法优劣:

- ✧ **优点:** 通过设计合成异常, 充分利用正常样本进行训练, 避免了对异常样本的依赖。
- ✧ **缺点:** 合成异常与真实异常可能存在差异, 尤其是在异常类型多样且复杂的医学图像中, 可能导致模型在实际应用中的泛化能力不足。

双阶段方法 (Two-stage Approach)

双阶段方法首先通过预任务学习自监督表示, 然后在学到的表示上构建单类分类器。该方法旨在通过预任务提升模型的特征表达能力, 使其更适合异常检测任务。

代表性方法:

- ✧ **PANDA** (Reiss 等人, 2021) : 使用预训练的 ResNet152, 通过分布增强的单类对比学习来学习紧凑的内围分布, 并在此基础上构建单类分类器。
- ✧ **MSC (Mean-Shift Contrastive)** (Reiss 和 Hoshen, 2023) : 采用基于均值偏移的对比损失, 旨在减少特征的均匀性, 提高异常检测的性能。
- ✧ **CutPaste** (Li 等人, 2021) : 通过剪切并粘贴图像补丁生成合成异常样本, 并训练模型区分正常样本与合成异常样本, 随后在学到的特征空间上应用高斯密度估计器 (GDE) 进行异常检测。

优缺点:

- ✧ **优点:** 通过预任务学习更具判别力的特征表示, 通常在多个数据集上表现优异; 双阶段方法能够更好地适应不同类型的异常样本。
- ✧ **缺点:** 依赖于预任务设计的有效性; 需要额外的步骤来构建和训练单类分类器, 增加了模型的复杂性。

3.2.3 特征参考方法 (Feature Reference-based Methods)

特征参考方法通过比较当前输入样本的特征与参考特征之间的差异来进行异常检测。这类方法在工业缺陷检测中应用广泛，但在医学图像中的应用效果有限。

知识蒸馏 (Knowledge Distillation)

知识蒸馏方法利用教师网络提取的正常特征作为参考，训练学生网络以回归教师网络的正常特征图。期望学生网络在正常区域能够准确重构特征，而在异常区域由于缺乏对异常的知识，重构特征将出现偏差，从而用于异常检测。

代表性方法：

- ✧ **Bergmann 等人 (2020)**：训练学生网络以回归教师网络在正常样本上的特征图，通过学生与教师特征图的差异来评估异常程度。
- ✧ **Salehi 等人 (2021)**：改进了知识蒸馏方法，通过引入更多的特征层和更复杂的损失函数，提升异常检测的性能。

优缺点：

- ✧ **优点**：能够利用预训练教师网络的强大特征提取能力，提高异常检测的准确性。
- ✧ **缺点**：在医学图像中，正常与异常样本的特征分布可能难以有效分离，导致方法效果不佳；需要复杂的蒸馏过程和多层特征比较，增加了模型的计算成本。

特征建模 (Feature Modeling)

特征建模方法通过在预训练网络中提取正常图像的中间特征，并存储代表性的原型特征，通过计算测试图像特征与这些原型特征的距离来评估异常程度。

代表性方法：

- ✧ **Lee 等人 (2022)**：利用预训练的 ResNet 网络提取正常图像的中间特征，存储代表性的特征原型，并通过计算测试图像与这些原型的距离来进行异常检测。

- ✧ **Roth 等人 (2022)**：提出了一种基于特征建模的方法，通过聚类正常样本的特征，构建一个原型库，并在测试时计算样本特征与原型库的距离以检测异常。

优缺点：

- ✧ **优点**：能够高效地利用预训练网络的特征提取能力；适用于多种图像模态。
- ✧ **缺点**：在医学图像中，正常与异常样本的特征空间可能高度重叠，导致距离计算的有效性下降；需要维护和更新原型库，增加了存储和计算成本。

4. 讨论与分析

通过对比分析不同方法在多个医学图像数据集上的表现，可以得出以下几点结论：

重构方法的鲁棒性：在缺乏预训练的情况下，重构方法表现出较高的鲁棒性，特别是简单的 AE 方法在多个数据集上表现良好，尤其是在处理局部异常时。然而，复杂的重构方法如 MemAE 在医学图像中效果不佳，可能因特征空间中正常与异常样本难以分离所致。

自监督学习方法的潜力：双阶段自监督学习方法通常优于单阶段方法，尤其是在合成异常更具真实性的情况下。通过设计更逼真的合成异常，可以显著提升模型对真实异常的检测能力。然而，这些方法在处理不同数据集时可能表现不稳定，且对异常模式的敏感性较高。

特征参考方法的局限性：尽管在工业缺陷检测中表现出色，特征参考方法在医学图像中的应用效果有限。这可能与医学图像中异常样本的复杂性和多样性有关，导致特征空间中正常与异常样本难以有效分离。

不同方法之间存在一定的互补性。例如，重构方法和自监督学习方法都依赖于对正常样本的学习，但重构方法侧重于图像或特征的重构误差，而自监督学习方法则通过设计预任务来增强模型对异常的感知能力。此外，特征参考方法可以与其他方法结合，利用预训练模型的特征提取能力，进一步提升异常检测的效果。

5. 个人理解与见解

以下是个人的一些独特见解：

结合多种方法的混合策略：单一方法在处理复杂和多样的医学图像异常时可能存在局限性。未来的研究可以探索将重构方法、自监督学习方法与特征参考方法相结合，利用各自的优势，形成更为强大的混合模型。例如，可以在重构模型中引入自监督学习的预任务，或在特征参考方法中结合自监督学习的特征表示，以提升整体的异常检测性能。

优化重构误差度量：现有研究表明，重构误差的度量方式对异常检测性能有显著影响。通过引入学习型度量函数，能够更有效地捕捉异常相关的语义差异，可能进一步提升重构方法的性能。例如，可以利用深度学习方法自动学习一个适应不同数据集和模态的重构误差度量函数，而不是依赖于手工设计的距离函数（如 L2、SSIM、感知损失）。

提升模型的可解释性：医学图像分析中，模型的可解释性至关重要。未来研究应关注如何通过可视化和解释机制，使得异常检测结果更具临床参考价值。例如，可以结合类激活映射（CAM）等可视化技术，直观地展示模型认为异常的区域，帮助临床医生理解和验证检测结果。

6. 可能的未来研究方向以及改进建议

开发自适应的潜在空间配置方法：现有的潜在空间配置方法主要依赖于经验和试验，缺乏理论指导。未来可以结合信息论的方法，开发自适应的潜在空间配置策略，以动态调整模型参数，适应不同数据集的需求。例如，可以利用熵的概念量化正常数据的潜在空间信息量，并设计一个自适应机制，使潜在空间的熵与正常数据的熵相匹配。

探索基于图像语义的重构方法：通过结合图像的高层语义信息，设计更为智能的重构模型，能够更好地捕捉和识别异常区域，提升检测准确性。例如，可以引入注意力机制，使得重构模型能够更关注图像中潜在的异常区域，从而提高异常检测的敏感性。

利用最新的视觉语言模型（VLMs）：随着视觉语言模型（如 CLIP、BLIP 等）的迅速发展，这些模型在跨模态学习和语义理解方面展现出强大的能力。未来研究可以探索如何将 VLMs 应用于医学图像异常检测，进一步提升模型的泛化能力和鲁棒性。例如，可以利用 VLMs 的多模态特征表示，将文本描述与图像特征结合，提升对复杂异常的识别能力。

7. 结论

本综述系统性地总结了医学图像异常检测领域中的主要方法，涵盖了重构方法、自监督学习方法和特征参考方法。通过在多个医学图像数据集上的对比分析，发现：

- ✧ 重构方法在无预训练情况下表现出较高的鲁棒性，尤其是简单的 AE 方法在处理局部异常时效果显著。
- ✧ 自监督学习方法，特别是双阶段方法，在合成异常更具真实性的情况下，能够显著提升对真实异常的检测能力。
- ✧ 特征参考方法在医学图像中的应用效果有限，可能因医学图像异常样本的复杂性和多样性导致特征空间中正常与异常样本难以分离。

参考文献

- [1] **Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., Langs, G.** (2017). *Unsupervised anomaly detection with generative adversarial networks to guide marker discovery*. In **International Conference on Information Processing in Medical Imaging (IPMI)** (pp. 146–157). Springer.
- [2] **Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., Schmidt-Erfurth, U.** (2019). *f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks*. **Medical Image Analysis**, 54, 30–44.
- [3] **Baur, C., Wiestler, B., Albarqouni, S., Navab, N.** (2019). *Deep autoencoding models for unsupervised anomaly segmentation in brain MR images*. In **Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries** (4th International Workshop, BrainLes 2018), pp. 161–169. Springer.
- [4] **Akcay, S., Atapour-Abarghouei, A., Breckon, T. P.** (2019). *Ganomaly: Semi-supervised anomaly detection via adversarial training*. In **Asian Conference on Computer Vision (ACCV)**, pp. 622–637. Springer.
- [5] **Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.** (2018). *Improving unsupervised defect segmentation by applying structural similarity to autoencoders*. arXiv preprint arXiv:1807.02011.

- [6] **Shvetsova, N., Bakker, B., Fedulova, I., Schulz, H., Dylov, D. V.** (2021). *Anomaly detection in medical imaging with deep perceptual autoencoders*. **IEEE Access**, 9, 118571–118583.
- [7] **Meissen, F., Paetzold, J., Kaissis, G., Rueckert, D.** (2022a). *Unsupervised anomaly localization with structural feature-autoencoders*. arXiv preprint arXiv:2208.10992.
- [8] **Guo, J., Lu, S., Jia, L., Zhang, W., Li, H.** (2023a). *ReContrast: Domain-specific anomaly detection via contrastive reconstruction*. In **Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS)**.
- [9] **Tan, J., Hou, B., Batten, J., Qiu, H., Kainz, B.** (2020). *Detecting outliers with foreign patch interpolation*. arXiv preprint arXiv:2011.04197.
- [10] **Tan, J., Hou, B., Day, T., Simpson, J., Rueckert, D., Kainz, B.** (2021). *Detecting outliers with Poisson image interpolation*. In **International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)** (pp. 581–591). Springer.
- [11] **Sato, J., Suzuki, Y., Wataya, T., Nishigaki, D., Kita, K., Yamagata, K., Tomiyama, N., Kido, S.** (2023). *Anatomy-aware self-supervised learning for anomaly detection in chest radiographs*. **iScience**, 26(4), 106483.
- [12] **Reiss, T., Cohen, N., Bergman, L., Hoshen, Y.** (2021). *PANDA: Adapting pre-trained features for anomaly detection and segmentation*. In **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 2806–2814).
- [13] **Reiss, T., Hoshen, Y.** (2023). *Mean-shifted contrastive loss for anomaly detection*. In **AAAI Conference on Artificial Intelligence (AAAI)** (pp. 2155–2162).
- [14] **Li, C. L., Sohn, K., Yoon, J., Pfister, T.** (2021). *CutPaste: Self-supervised learning for anomaly detection and localization*. In **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 9664–9674).
- [15] **Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.** (2020). *Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings*. In **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 4183–4192).
- [16] **Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., Rabiee, H. R.** (2021). *Multiresolution knowledge distillation for anomaly detection*. In

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 14902–14912).

- [17] **Lee, S., Lee, S., Song, B. C.** (2022). *CFA: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization*. **IEEE Access**, 10, 78446–78454.
- [18] **Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.** (2022). *Towards total recall in industrial anomaly detection*. In **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 14318–14328).
- [19] **Mao, Y., Xue, F. F., Wang, R., Zhang, J., Zheng, W. S., Liu, H.** (2020). *Abnormality detection in chest X-ray images using uncertainty prediction autoencoders*. In **International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)** (pp. 529–538). Springer.