病理学告知的潜扩散模型在淋巴结转移异常检测中的应用

Jiamu Wang[1], Keunho Byeon[1], Jinsol Song[1], Anh Nguyen[1], Sangjeong Ahn[2], Sung Hak Lee[3*], and Jin Tae Kwak[1*]

[1] School of Electrical Engineering, Korea University, Seoul 02841, Korea
{taurusmumu, bkh5922, truetg, ngtienanh, jkwak}@korea.ac.kr
[2] Department of Pathology, Korea University Anam Hospital and Department of Biomedical Informatics, Korea University College of Medicine, Seoul 02841, Korea
vanitas80@korea.ac.kr
[3] Department of Hospital Pathology, Seoul St. Mary' s Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea
hakjjang@catholic.ac.kr

摘要 异常检测是数字病理学中的一种新兴方法，因为它能够高效地利用数据进行疾病诊断。虽然监督学习方法可以提供高精度，但它们依赖于大量注释的数据集，并且在数字病理学中面临数据稀缺的问题。然而，无监督异常检测通过识别与正常组织分布的偏差而不需要详尽的标注提供了可行的替代方案。最近，去噪扩散概率模型在无监督异常检测领域越来越受欢迎，在自然和医学图像数据集中取得了令人鼓舞的表现。在此基础上，我们将视觉语言模型与扩散模型结合起来用于数字病理学中的无监督异常检测，并在重建过程中使用组织病理学提示。我们的方法采用一组与正常组织相关的关键词来引导重建过程，有助于区分正常组织和异常组织。为了评估所提方法的有效性，我们在当地医院的一个胃淋巴结数据集上进行了实验，并通过一个公开的乳腺淋巴结数据集测试了其在领域转换下的泛化能力。实验结果突显了该方法在数字病理学中各种器官上的无监督异常检测潜力。代码：https://github.com/QuIIL/AnoPILaD.

Keywords: 无监督异常检测 · 扩散模型 · 视觉-语言模型 · 淋巴结转移。

1 介绍

淋巴结转移是癌症进展和治疗决策中的一个重要预后因素 [8] 。随着数字病理学的出现，已经提出了一些人工智能方法来自动化组织内的淋巴结转

[*] 这些作者对这项工作做出了同等贡献

移检测。其中许多方法基于监督学习方法，利用卷积神经网络（CNN）和基于变换器的模型以高精度识别转移 [9] 。虽然这些方法有效，但它们严重依赖于详尽的专业注释，这既耗时又耗费资源。为了解决这一限制，无监督学习作为一种可行的替代方案出现了，因为它不需要手动标注。在无监督学习中，模型仅通过正常（符合分布）样本进行训练，以学习符合分布模式的表示。然后，该训练好的模型通过识别偏离所学分布模式的情况来检测异常或不符合分布（OOD）样本，使其特别适合用于数字病理学中的大规模应用，在这种情况下，标注过的异常样本很少 [10] 。

最近，生成模型已被广泛用于无监督异常检测，主要有两种流行的方法：基于密度的方法和基于重构的方法。基于密度的方法，如变分自编码器（VAEs）[1, 18–20]，学习分布内数据的表示形式，对分布内的样本赋予更高的可能性，而对外部数据分布（OOD）的样本赋予更低的可能性。相比之下，基于重构的方法仅在正常数据上进行训练，以确保异常样本的重构质量较差，而正常样本的重构质量较高。这些方法包括基于自编码器（AE）[5–7]、生成对抗网络（GAN）[21–23] 和去噪扩散概率模型（DDPM）[2–4, 13] 的方法。

在本文中，我们介绍了 AnoPILaD，这是一种用于淋巴结病理图像异常检测的病理学信息潜扩散模型。该框架结合了潜扩散模型（LDM）[11] 和视觉语言模型（VLM）[12]，以提高对病理图像中异常的识别能力。AnoPILaD 利用一个 LDM 通过迭代扩散和去噪过程在潜在空间中学习近正常图像的一个紧凑表示，同时保留关键的组织学特征。AnoPILaD 还采用了一个 VLM 来选择特定于病理学的正常关键词，语义上引导重构过程朝着特定方向进行。以这种方式，AnoPILaD 实现了对正常样本的小偏差和对异常样本的大偏差，从而提高了异常检测的准确性和鲁棒性。

2 方法

2.1 病理学告知的潜扩散模型

近期的研究经常使用 DDPMs 进行基于重建的异常检测。训练好的 DDPMp_θ 通过前向过程（扩散过程）添加噪声生成匹配分布模式 $\boldsymbol{z_0} \sim q(\boldsymbol{z_0})$ 的样本，在时间 $t-1$ 处具有可计算的后验概率：

$$q(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t, \boldsymbol{z}_0) = \mathcal{N}\left(\boldsymbol{z}_{t-1}\middle|\tilde{\mu}(\boldsymbol{z}_t(\boldsymbol{z_0}, \boldsymbol{\epsilon}), t), \tilde{\beta}_t \boldsymbol{I}\right) \tag{1}$$

其中 $t \sim [1, 1000]$,$\epsilon \sim \mathcal{N}(0, \boldsymbol{I})$,$\boldsymbol{z_t}$ 是一个带噪声的样本，而 $\tilde{\beta}_t$ 是一个预定义常数。该模型通过逆过程逐步去除噪声：

$$p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t) = \mathcal{N}\left(\boldsymbol{z}_{t-1}\middle|\mu_\theta(\boldsymbol{z}_t(\boldsymbol{z_0}, \boldsymbol{\epsilon_\theta}), t), \tilde{\beta}_t \boldsymbol{I}\right) \tag{2}$$

其中 $\boldsymbol{\epsilon_\theta}$ 是一个用于从 $\boldsymbol{z_t}$ 预测 ϵ 的近似器。该模型通过减少两个高斯分布之间的 KL 散度进行训练，一个简单的目标函数表示为：

$$\mathcal{L} = E_{\boldsymbol{z_t}, t, \epsilon \sim \mathcal{N}(0,1)}\left[\|\epsilon - \epsilon_\theta(\boldsymbol{z_t}, t)\|_2^2\right]. \tag{3}$$

一项先前的工作 [2] 采用了 DDPM 来检测乳腺淋巴结转移，称为 AnoDDPM，在此工作中正常样本是分布内数据而转移样本是 OOD 数据。预训练的 DDPM 对部分扩散的输入 $\boldsymbol{z_t} \sim q(\boldsymbol{z_t}|\boldsymbol{z_0})$ 进行了去噪，同时引导反向过程趋向于分布内的模式，并得到重构 $\hat{\boldsymbol{z_0}} \sim p_\theta(\boldsymbol{z_0}|\boldsymbol{z_{1:t}})$。然后，输入 $\boldsymbol{z_0}$ 与其重构 $\hat{\boldsymbol{z_0}}$ 之间的差异作为异常分数，预期正常样本的该分数较小而转移样本的较大。尽管成功，AnoDDPM 展示了大量误报 [2] ，表明其区分正常样本和 OOD 分布的能力不足（图 1 ）。为了解决这个问题并提高异常检测性能，AnoPILaD 集成了一种带有病理学特定文本提示的 LDM，基于这些提示可以提升重构质量并进而放大正常样本与异常样本之间差异的假设。我们使用以下目标训练 LDM：

$$\mathcal{L} = E_{\boldsymbol{z_t}, t, c, \epsilon \sim \mathcal{N}(0,1)}\left[\|\epsilon - \epsilon_\theta(\boldsymbol{z_t}, t, c)\|_2^2\right] \tag{4}$$

其中 c 表示一个文本提示。我们通过使用相同的重构过程来对淋巴结病理图像进行基于病理信息的重构，与 AnoDDPM 相同。唯一的例外是反转过程由给定的 $\hat{\boldsymbol{z_0}} \sim p_\theta(\boldsymbol{z_0}|\boldsymbol{z_{1:t}}, c)$ 文本条件引导。

2.2 加权提示生成

为了在重建过程中引入更强的归纳偏差，我们建议利用正常淋巴结的先验病理知识。具体来说，我们从文献中收集了 74 个描述正常淋巴结和组织细胞及微环境特征的病理关键词。这些关键词由一位经验丰富的病理科医生审阅并验证，以确保其临床相关性和准确性。

给定一张病理图像，我们将其与病理关键词对齐，识别出最相关的关键词，并使用它们生成一个文本提示，引导重建过程朝向基于病理信息的方向（图 2）。为了实现病理图像和关键词之间的对齐，我们采用了 CONCH [14]，这是一种在超过 117 万张病理图像-描述配对上预训练的视觉语言基础模型。

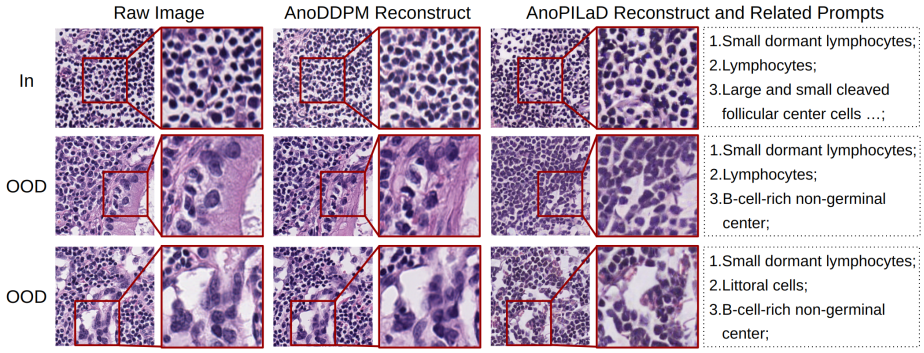

图 1. 淋巴结图像重建使用基于扩散的方法。使用文本提示，AnoPILaD 生成类似正常的病理图像，适用于分布内（in）和分布外（OOD）样本。

CONCH 的图像编码器和文本编码器分别用于生成每一对输入图像和关键词的图像嵌入和文本嵌入。然后，我们计算图像嵌入和文本嵌入之间的余弦相似度得分，并选择前五个最相似的关键词。所选关键词的相似性得分随后通过除以中位数得分进行归一化。使用选定的关键词及其归一化的得分，我们生成一个加权提示，如图 2所示。该加权提示被转换为嵌入向量，并输入到 LDM 中，遵循 Compel 库中所展示的过程 *。

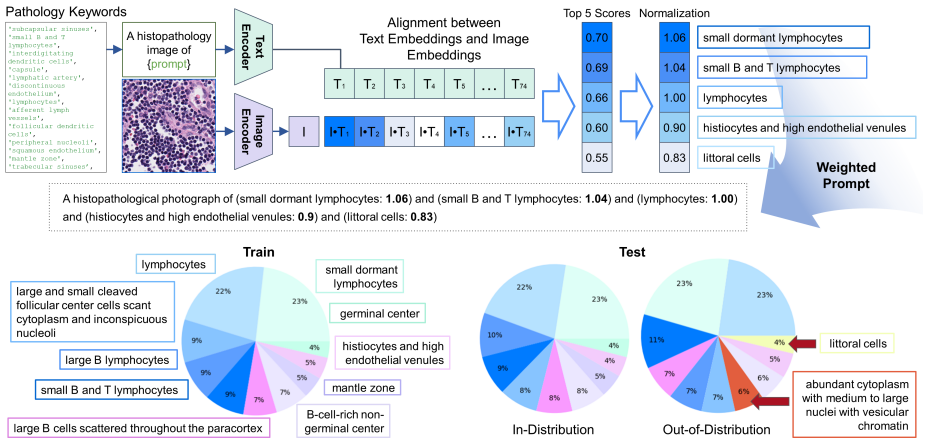

图 2.（上）生成加权文本提示的示例。（下）当地医院数据集中训练集和测试集前 10 个最频繁病理关键词的分布。

* https://github.com/damian0815/compel

表 1. 本研究中使用的两个数据集

数据集	切片级数据集				补丁级数据集			
	D^w_{tr}	D^w_{val}	D^w_{in}	D^w_{out} ($D^w_{\text{out,m}}$)	D^p_{tr}	D^p_{val}	D^p_{in}	D^p_{out} ($D^p_{\text{out,m}}$)
LH	643	50	58	57	1,373,475	102,240	174,703	115,330
C16	-	30	88	44 (20)	-	37,056	240,139	55,659 (55,536)

3 实验

3.1 数据集

我们从两个本地医院 (LH) 获取了一个胃淋巴结数据集，包含 808 个全幻灯片图像 (WSIs)，其中 751 个是正常的，57 个是有部分标注的转移性 WSIs。所有的转移性 WSIs 都被用作 OOD 测试集 ($D_{out}^{LH,w}$)。正常 WSIs 被分为一个包含 643 个 WSIs 的训练集 ($D_{tr}^{LH,w}$)，一个包含 50 个 WSIs 的验证集 ($D_{val}^{LH,w}$)，以及一个包含 58 个 WSIs 的分布内测试集 ($D_{in}^{LH,w}$)。我们将 WSIs 进一步划分为小块，得到 $D_{tr}^{LH,p}$（1,373,475），$D_{val}^{LH,p}$（102,224），$D_{in}^{LH,p}$（174,703 个小块，其中 138054 个来自 $D_{in}^{LH,w}$ 和 36649 个来自正常标注的 $D_{out}^{LH,w}$），以及 $D_{out}^{LH,p}$（115,330 个带有完全转移注释的小块）。此外，我们使用了 Camelyon16 挑战数据集（C16）[15]，一个乳腺淋巴结数据集，进行独立测试。我们利用 32 个正常 WSIs 作为验证集（$D_{val}^{C16,w}$），88 个正常 WSIs 作为分布内测试集（$D_{in}^{C16,w}$），以及 40 个转移性 WSIs 作为分布外测试集（$D_{out}^{C16,w}$）。在 40 个中，含有直径大于 2 毫米的肿瘤细胞簇的有 22 个 WSIs，被指定为 C16 Macro（$D_{\text{out,m}}^{C16,w}$）。LH 中的所有肿瘤区域直径均超过 2 毫米。我们将 WSIs 进一步分割成小块，结果分别为 $D_{val}^{C16,p}$（37,056），$D_{in}^{C16,p}$（240,139）和 $D_{out}^{C16,p}$（带有完全转移注释的 55,659 个切片）。对于补丁级别的评估，来自 LH 和 C16 的 WSI 使用像素级别注释被划分为 256×256 像素的补丁。所有 WSI 均在 20× 倍放大下进行处理。数据集概述见表 1。

3.2 实验详情

为了实现 AnoPILaD，我们使用了稳定扩散模型 v1.5。扩散模型经过了 400,000 步的微调，采用了 Adam 优化器、学习率为 1e-5 以及批量大小为 64，并利用低秩适应（LORA）进行更新矩阵维度为 4。输入图像尺寸为 256×256

像素。我们将 AnoPILaD 与基于密度和重构的 OOD 方法进行了比较。对于基于密度的方法，我们使用了 VAE 骨干网络 [18] 的负对数似然（NLL）及其三个变体：Regret [1]、LLR [20] 和 complexity [19]。潜向量的大小设置为 100，输入图像尺寸随机裁剪为 64×64 像素。对于基于重构的方法，我们采用了 f-AnoGAN [21]、AE 和 MemAE [5]。f-AnoGAN 使用随机裁剪的 64×64 输入图像块。对于 AE 和 MemAE，它们遵循了 [5] 的架构设计，并且输入图像尺寸为 256×256 像素。AnoDDPM 和 AnoPILaD 都是使用 Diffusers 库 [16] 实现的，并在推理过程中使用了具有 100 个时间步长的 PLMS 采样器 [17]。除非另有说明，训练过程遵循原始工作的描述。

对于每种方法，我们计算了每个测试集中所有切片的异常得分的 z 分数，并通过计算接收者操作特征曲线下的面积（AUC）和精度-召回率曲线下的面积（AUPR）来评估切片级别的 OOD 检测。对于 WSI 级别的评估，我们生成了一个 z 分数热图，并应用了一次形态学腐蚀操作 [24]，使用一个 2×2 窗口，因为转移区域可以非常小。然后，我们评估了每个模型的分类和分割性能。对于分类，我们采用了两种方法：最大 z 分数（Z_{MAX}）和第 99 百分位数 z 分数的平均值（Z_{99}）。这些用于计算 WSI 级别的 AUC 和 AUPR。对于分割，我们计算了平均切片级别 DICE 和交并比（IoU）D^w_{out} 来评估预测与标注之间的重叠，并且由于没有正区域，还计算了平均切片级别真阴性率（TNR）D^w_{in}。分割预测阈值为零。为了确定两种基于扩散模型的方法的重建时间步长，我们测试了八个时间步长值如 [2] 所示，并根据 LH 测试集上的最佳 WSI 分类性能选择了时间步长，对于这两种方法都是 674。

4 结果

我们训练了 AnoPILaD 和所有竞争模型，仅使用 D_{tr}，并分别在源自不同器官的两个独立数据集上评估它们的切片级和 WSI 级性能。因此，此次评估提供了关于由于组织类型变化而导致领域偏移时，模型稳健性的见解。

表 2. 补丁级别异常检测结果

		NLL	Regret	LLR	complexity	f-AnoGAN	AE	MemAE	AnoDDPM	AnoPILaD
LH	AUC	0.4982	0.6720	0.6078	0.7931	0.2289	0.9254	0.9290	0.8555	**0.9587**
	AUPR	0.5552	0.6718	0.6260	0.7139	0.3377	0.8906	0.8886	0.7841	**0.9499**
C16	AUC	0.4983	0.6720	0.7065	0.7752	0.1735	0.6584	0.6611	0.6857	**0.8884**
	AUPR	0.5552	0.6718	0.4765	0.5140	0.1104	0.4759	0.4880	0.5741	**0.6987**

表 2展示了补丁级别异常检测的性能。AnoPILaD 远远超过了所有其他方法。在四种基于密度的方法（NLL、Regret、LLR 和复杂度）中，NLL 未能区分 OOD 补丁与正常补丁。其变体提高了性能，其中复杂度方法取得了最大的改进。尽管有这些改进，复杂度仍远逊于 AnoPILaD，在 AUC 上存在 ∼0.16 的显著差距，并且在 AUPR 上有 ∼0.23 的显著差距。在基于重构的方法中，f-AnoGAN 表现最差，而基于 AE 的方法（AE 和 MemAE）获得了最高分。在 LH 中，它们的 AUC 和 AUPR 得分分别比 AnoPILaD 低约 0.03 和 0.05。然而，在 C16 中，性能差距大幅增加到 AUC 上的 ∼0.22 以及 AUPR 上的 ∼0.22，表明了由于器官类型差异导致的领域变化时 AnoPILaD 具有优越的鲁棒性。

表 3. WSI 级别异常检测结果

		LH			C16			C16 宏观		
分类		AUC	AUPR	-	AUC	AUPR	-	AUC	AUPR	-
AE	Z_{max}	0.9622	0.9612	-	0.6612	0.5961	-	0.6398	0.3677	-
	Z_{99}	0.9395	0.9381	-	0.5798	0.5217	-	0.5523	0.2918	-
成员自动编码器	Z_{max}	0.9504	0.9440	-	0.6505	0.5689	-	0.6381	0.3657	-
	Z_{99}	0.9365	0.9382	-	0.5686	0.5313	-	0.5597	0.3141	-
阿诺 DDPM	Z_{max}	0.7840	0.6616	-	0.4992	0.3885	-	0.4926	0.2146	-
	Z_{99}	0.9383	0.8995	-	0.4551	0.3905	-	0.5119	0.2347	-
阿诺皮拉德	Z_{max}	0.9837	0.9740	-	**0.6745**	**0.6140**	-	**0.8062**	**0.5965**	-
	Z_{99}	**0.9943**	**0.9948**	-	0.6367	0.5902	-	0.8023	0.5886	-
		LH			C16			C16 宏观		
分割段落		TNR	DICE	IoU	TNR	DICE	IoU	TNR	DICE	IoU
AE		0.7981	0.3812	0.2902	0.4536	0.1745	0.1317	0.4536	0.3249	0.2549
MemAE		0.7957	0.3863	0.2932	0.5593	0.1377	0.1039	0.5594	0.2173	0.2124
AnoDDPM		0.7850	0.4319	0.3259	0.7842	0.1765	0.1142	0.7842	0.3131	0.2075
AnoPILaD		**0.8097**	**0.4322**	**0.3311**	**0.8312**	**0.3098**	**0.2326**	**0.8312**	**0.5420**	**0.4275**

在 WSI 级别上，我们进一步证实了来自 patch 级别异常检测的结果。表 3显示了四个表现最佳的模型（AE、MemAE、AnoDDPM 和 AnoPILaD）在 patch 级别异常检测中的 WSI 级别异常检测结果。对于正常与异常切片分类，AnoPILaD 在两个数据集、评估指标以及评分策略（Z_{MAX} 和 Z_{99}）上都优于所有竞争模型。AnoPILaD 和其他三个模型的性能在 LH 和 C16 之间存在显著差异。在 LH 中，四个模型获得了 0.7840∼0.9943 AUC 和 0.6616∼0.9948 AUPR 的分数，而在 C16 中，它们的表现有所下降，AUC 为 0.4551∼0.6745，

AUPR 为 0.3885∼0.6140。然而，AnoPILaD 表现出最小的最佳性能下降，AUC 从 0.3092∼0.3576 到 AUPR 的 0.3600∼0.4046，突显其在数据集和器官类型上的鲁棒性。关于大型转移区域（C16 Macro），所有模型的性能普遍下降；然而，AnoPILaD 获得的 AUC 范围从 0.8023 到 0.8062，表明其在淋巴结跨器官异常检测中具有优越潜力。

评估转移区域的分割结果时，AnoPILaD 的优势非常明显，在所有评估场景中均获得了最高分。我们还观察到两种异常检测方法之间存在明显差异。尽管基于自编码器的方法（AE 和 MemAE）与基于扩散的方法（AnoPILaD 和 AnoDDPM）在分类性能上相当，但它们的分割性能显著较差。如图 3所示，MemAE 为幻灯片中的所有像素分配了相似的分数，表明其在检测转移区域时缺乏特异性。AE 的行为几乎与 MemAE 相同。这些结果表明，在评估异常检测时分割性能的重要性，因为仅靠分类可能无法准确评估模型定位异常区域的能力。

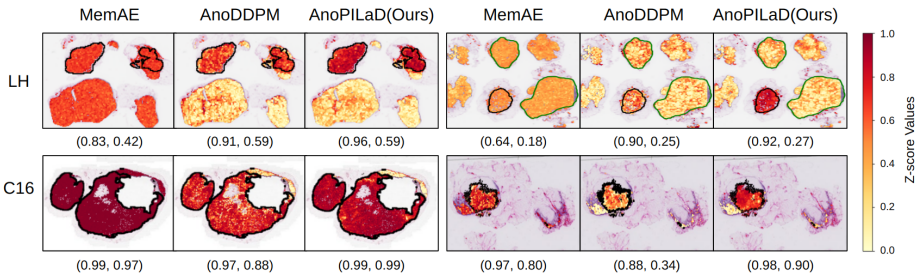

图 3. Z 分数热图来自两个数据集的四张转移滑片，黑色轮廓显示转移注释，绿色轮廓显示正常注释。括号中的数字表示每张幻灯片的（Z_{99}, Dice）。

我们通过 AnoPILaD 和 AnoDDPM 的可视化结果进行了比较（图 1）。从原始图像开始，AnoPILaD 和 AnoDDPM 分别在有无文本引导的情况下旨在重建类似正常的结构。对于一个分布内的样本（第一行），两个模型都能成功生成具有正常组织学特征的图像，保留了小而密集的淋巴细胞结构。然而，对于分布外样本（第二和第三行），这些样本包含被破坏的组织结构的转移区域，两种模型之间的差异变得更加明显。AnoDDPM 部分重建了转移区域，并未能完全抑制多形性核和纤维化组织，导致残留异常，模糊了正常与异常结构之间的界限。相比之下，在文本提示的引导下，AnoPILaD 生

成了淋巴细胞排列更加均匀的图像，同时抑制了组织结构中的畸变。这使得正常和转移区域之间的区分更加清晰。

5 结论

我们提出 AnoPILaD，一种病理信息引导的 LDM，用于淋巴结无监督异常检测。通过利用提示提供的组织学背景，AnoPILaD 在重建过程中引入了更强的归纳偏差，增强了对异常特征的敏感性，并提高了检测性能。在两种器官类型的斑块级和切片级性能评估中，AnoPILaD 明显优于其他异常检测方法，包括基于密度和重建的方法。未来的研究将包括扩展 AnoPILaD，以进一步提高其在跨器官数据集上的性能，提高其在更广泛的病理应用中的适应性和鲁棒性。

参考文献

1. Xiao, Zhisheng, Qing Yan, and Yali Amit. "Likelihood regret: An out-of-distribution detection score for variational auto-encoder." Advances in neural information processing systems 33 (2020): 20685-20696.
2. Linmans, Jasper, et al. "Diffusion models for out-of-distribution detection in digital pathology." *Medical Image Analysis* 93 (2024): 103088.
3. Graham, Mark S., et al. "Denoising diffusion models for out-of-distribution detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.
4. Wyatt, Julian, et al. "Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
5. Gong, Dong, et al. "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.
6. Chong, Yong Shean, and Yong Haur Tay. "Abnormal event detection in videos using spatiotemporal autoencoder." *Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21 – 26, 2017, Proceedings, Part II 14*. Springer International Publishing, 2017.

7. Hinton, Geoffrey E., and Richard Zemel. "Autoencoders, minimum description length and Helmholtz free energy." *Advances in neural information processing systems* 6 (1993)

8. Nathanson, S. David. "Insights into the mechanisms of lymph node metastasis." *Cancer* 98.2 (2003): 413-423.

9. Budginaite, Elzbieta, et al. "Computational methods for metastasis detection in lymph nodes and characterization of the metastasis-free lymph node microarchitecture: A systematic-narrative hybrid review." *Journal of Pathology Informatics* 15 (2024): 100367.

10. Salehi, Mohammadreza, et al. "A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges." arXiv preprint arXiv:2110.14051 (2021).

11. Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2022.

12. Zhang, Jingyi, et al. "Vision-language models for vision tasks: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).

13. Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

14. Lu, Ming Y., et al. "Towards a visual-language foundation model for computational pathology." *arXiv preprint arXiv:2307.12914* (2023).

15. Bejnordi, Babak Ehteshami, et al. "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer." *Jama* 318.22 (2017): 2199-2210.

16. Diffusers: State-of-the-art diffusion models" by Patrick von Platen, Suraj Patil, Anton Lozhkov, et al. (2022), available on GitHub as a repository named "huggingface/diffusers

17. Liu, Luping, et al. "Pseudo numerical methods for diffusion models on manifolds." arXiv preprint arXiv:2202.09778 (2022).

18. Kingma, Diederik P. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).

19. Serrà, Joan, et al. "Input complexity and out-of-distribution detection with likelihood-based generative models." *arXiv preprint arXiv:1909.11480* (2019).

20. Ren, Jie, et al. "Likelihood ratios for out-of-distribution detection." *Advances in neural information processing systems* 32 (2019).

21. Schlegl, Thomas, et al. "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks." *Medical image analysis* 54 (2019): 30-44.

22. Schlegl, Thomas, et al. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery." *International conference on information processing in medical imaging*. Cham: Springer International Publishing, 2017.

23. Goodfellow, Ian, et al. "Generative adversarial networks." *Communications of the ACM* 63.11 (2020): 139-144.

24. Virtanen, Pauli, et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." Nature methods 17.3 (2020): 261-272.