arXiv:2303.00865v2 [cs.CV] 5 Jul 2023

AMIGO：用于十亿像素图像表征学习的共享上下文处理稀疏多模态图 Transformer

Ramin Nakhli 不列颠哥伦比亚大学ramin.nakhli@ubc.ca

Puria Azadi Moghadam HaoyangMi 不列颠哥伦比亚大学 约翰斯·霍普金斯大学

Hossein Farahani 不列颠哥伦比亚大学

Alexander Baras 约翰斯·霍普金斯大学

Blake Gilks 不列颠哥伦比亚大学

Ali Bashashati 不列颠哥伦比亚大学ali.bashashati@ubc.ca

摘要

处理十亿像素级全切片组织病理图像（*WSI*）是一项计算开销巨大的任务。多实例学习（*MIL*）已成为处理全切片图像的常规方法，即将这些图像拆分成较小的图像块以进行后续处理。然而，基于 *MIL* 的技术忽略了图像块内单个细胞的显式信息。本文通过定义共享上下文处理这一新概念，设计了一种多模态图变换器（*AMIGO*），利用组织中的细胞图为患者生成单一表征，同时发挥组织层次结构的优势，从而能够在细胞级和组织级信息之间动态聚焦。我们将模型在生存预测方面的表现与多种最先进方法进行了基准比较，结果表明，我们的方法显著优于所有这些方法，包括层次化视觉变换器（*ViT*）。更重要的是，我们表明，该模型对信息缺失具有很强的鲁棒性，即使仅使用 *20%* 的数据，也能取得相同的性能。最后，在两个不同的癌症数据集中，我们证明模型能够将患者分层为低风险组和高风险组，而其他最先进的方法未能实现这一目标。我们还公开发布了大型免疫组化图像数据集 *InUIT*，其中包含来自 *188* 名患者的 *1,600* 个组织芯片（*TMA*）芯片及其生存信息，使其成为这一领域中规模最大的公开数据集之一 申请表链接 [1]。

1. 引言

医学图像的数字化处理近年来在计算机视觉领域受到广泛关注，深度学习模型在该领域的应用涵盖多种图像类型（例如，组织病理图像、CT扫描和MRI扫描）以及众多任务（例如，分类、分割和生存期预测）[6, 11, 27, 29, 35, 37, 43]。这些模型能够直接从原始图像中学习预测性特征，这一范式转变为医学成像带来了令人振奋的机遇。对于数字化组织病理图像而言，这一点尤为重要，因为其中每个数据点都是一张数十亿像素的图像（也称为全切片图像或WSI）。与自然图像不同，每张WSI在不同放大倍率下都具有很高的细粒度，图像尺寸可达100，000×100000像素，这给计算机视觉带来了令人振奋的挑战。

处理全切片图像（WSI）的计算复杂性，常见方法是采用多实例学习（MIL）技术 [30]。具体而言，该方法将每张切片划分为更小的图像块（例如，256×256像素），将这些图像块输入特征提取器，并通过聚合这些表示来表征切片。这项技术在多种任务中都取得了令人鼓舞的结果，包括癌症亚型分类和生存预测。然而，它存在若干主要问题。首先，鉴于全视野切片的分辨率很高，即使采用不重叠的 256×256 窗口，也会产生海量图像块。因此，后续的 MIL 聚合方法必须采用简单的池化操作 [3, 17] 或层次聚合，以提高灵活性 [6]。然而，前者会限制

[1]本文已被 CVPR 2023 接收

聚合器的表征能力，而后者则需要消耗大量计算能力。其次，该方法严重依赖数据集规模，因此在可用数据点较少（例如，数百个）的情况下，模型容易过拟合。最后，尽管细胞是组织的主要组成部分，多实例学习方法主要关注图像块，这使模型的分辨率仅限于观察细胞群体的快照，而非单个细胞。因此，切片的最终表示缺少单个细胞之间的相互作用。

多项临床研究已充分证实，组织的异质性对癌症结局有着至关重要的影响 [31, 45]。例如，研究表明，乳腺癌患者肿瘤基质中较高水平的免疫细胞浸润与更长的生存期和更积极的治疗反应相关 [45]。因此，组织病理学图像分析的机器学习方法需要考虑肿瘤异质性和细胞间相互作用。然而，该领域的大多数研究都针对突出显示细胞核（不区分细胞类型）和细胞外基质的单张图像。近年来，少数研究考察了利用不同蛋白质标记物识别多种细胞类型的病理图像 [25, 41]。但这些研究仍采用单模态方法（即一张图像中只有一种细胞类型），忽略了这些图像的多模态背景（即组织中存在多种细胞类型）。

本研究探讨了图神经网络（GNN）在处理由组织病理图像生成的细胞图（即通过连接相邻细胞构建的图）中的应用（图1）。我们尤其关注细胞图，因为它使我们能够聚焦于细胞级信息及其相互作用。通过在不同尺度上灵活聚焦，从细胞级别到组织级别，这类信息使模型能够以多尺度视角观察组织；相比之下，多实例学习模型仅关注具有预设分辨率和光学放大倍率的图像块。细胞类型及其空间位置的信息有助于模型找到组织中对其表示更为重要的区域（例如，肿瘤区域或浸润肿瘤细胞的免疫细胞）。与多实例学习方法中代价高昂的层次化池化相比 [6]，图神经网络的消息传递机制提供了一种高效的方法来处理规模庞大的全视野切片，这得益于所有图节点之间的权重共享。由于参数数量减少，这种方法也降低了训练时对大量全视野切片的需求。

在这项工作中，我们提出了稀疏多模态图变换器模型（AMIGO），通过以细胞作为主要构建单元，学习组织病理图像的表示。从细胞层面出发，我们的模型

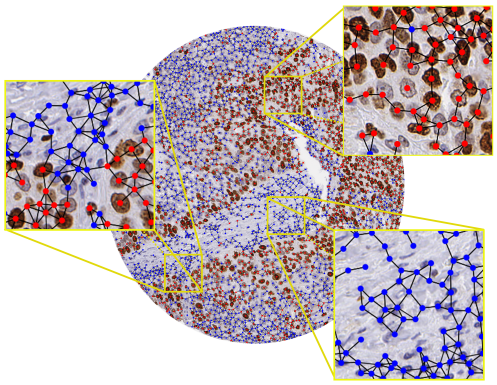


图1. 由经 Ki67 生物标志物染色的 $4,000 \times 4{,}000$ 像素 TMA组织芯片核心构建的细胞图。每个红点表示对 Ki67 呈阳性反应的细胞，而蓝点表示对该生物标志物呈阴性反应的细胞。高亮的图像块展示了组织中的代表性区域，这些区域的细胞空间分布和组织结构各不相同。典型的多实例学习方法无法捕捉这种异质性，因为它没有考虑图像块的位置，也缺少关于图像块内具体细胞的明确信息。

逐步将信息传播至更大范围的细胞邻域，从而自然地编码组织的层次结构。更重要的是，与其他研究不同，我们以多模态方式处理这一问题，以便更全面地理解组织结构；这对于生存预测等基于上下文的任务至关重要。具体而言，对于单个患者，可能有多张组织病理图像可供使用，每张图像突出显示特定类型的细胞（通过使用特定蛋白质标志物对细胞进行染色），并由此形成一个独立的细胞图（图2）。因此，我们采用多模态方法，将不同模态的细胞图结合起来，为单个患者获得统一表示。这也进一步证实了我们关于细胞类型重要性以及不同细胞连接类型之间区别的观点。除了取得最先进的结果外，我们还惊讶地发现，我们的多模态方法对信息缺失具有很强的鲁棒性，这使我们能够依靠网络的这种重建能力，更高效地进行训练。我们的工作从多个方面推动了多实例学习、视觉变换器（ViT）和图神经网络的发展：

- 我们提出首个多模态细胞图处理模型，该模型基于具有共享上下文信息的多模态组织病理图像进行生存预测。

- 我们的模型消除了多实例学习模型的关键障碍，使得能够在单个图形处理器上高效训练多千兆像素图像，

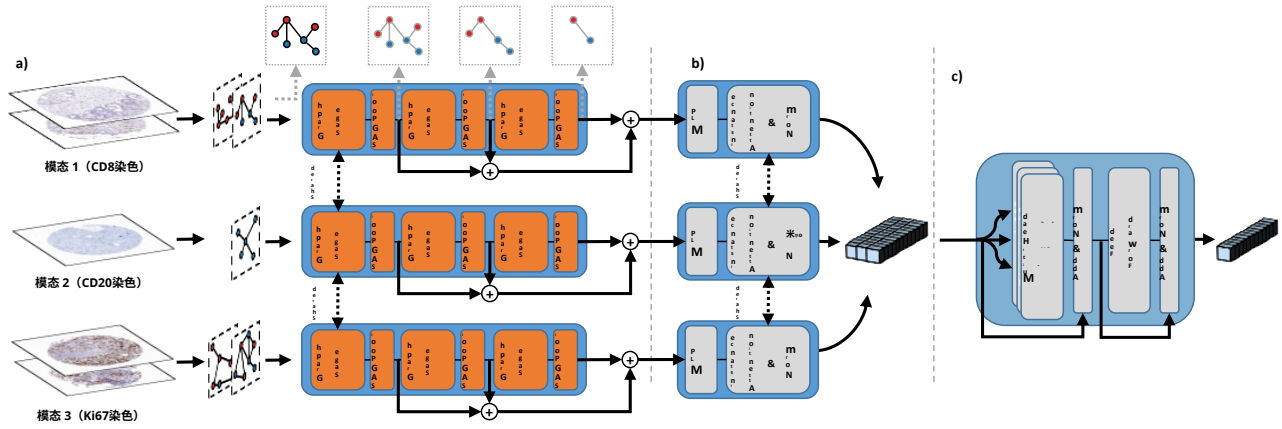


图 2. 我们提出的方法概览。a）首先从经不同生物标志物（例如，CD8、CD20 和 Ki67）染色的组织病理图像中提取细胞图，并将其输入与各自模态对应的编码器。编码器的初始层为共享层，从而有助于进一步泛化；后续层则提取各模态特有的功能。上方的图展示了模型的层次池化机制。b）通过共享的实例注意力层（共享上下文处理），将每种模态下多个图实例得到的表征进行整合，从而生成单个表示向量。c）使用 Transformer 合并所得向量，生成患者级嵌入，用于生存预测等下游任务。

并且性能优于包括 ViT 在内的所有基线模型。该模型还实现了视觉变换器的层次化结构，同时在端到端训练过程中显著减少了参数数量。

- 我们还发布了一个大型免疫组织化学图像数据集，其中包含来自 188 位患者的 1,600 个组织芯片（TMA）芯样及其生存信息，使其成为该领域规模最大的数据集之一。

2. 相关工作

2.1. 组织病理学中的多实例学习

受词袋思想启发，Zaheer 等人 [46]和 Brendel 等人 [4]是多实例学习模型领域的两位先驱，他们提出了用于图像表示学习的置换不变特征袋。类似地，数字病理学中早期的多实例学习研究也采用相同的方法，通过简单的图块级聚合算法来学习全视野切片的表征 [16, 19]。然而，后续研究在此方面采用了更灵活的设计。例如，IIse等人 [17] 使用基于注意力的操作，对所有图像块的表征进行池化；Cam- panella 等人 [5] 则使用循环神经网络（RNN）聚合排名靠前的图像块的表征。Li 等人。 [23]通过提出两阶段模型，引入了多分辨率多实例学习的思想：在某一 5×放大倍率下对图像块进行聚类，并在 10× 放大倍率下对信息量最大的图像块进行注意力池化。类似地，另一项同期研究通过自监督学习实现了多分辨率思想，作者还考虑了图像块的空间位置 [22]。在更最近的一项研究中，Zhang 等人. [47]在双层设置中引入了伪装袋，而 Chen 等人. [6]提出使用 Vision Transformer 对全视野切片进行层次化池化。然而，以上所有研究都忽略了图像中存在的细胞级细节，并且模型训练需要大量数据（数千名患者）。本文旨在通过采用以细胞为中心的方法，在图像不同区域间进行信息的层次化池化，以解决这些不足。

2.2. 组织病理学中的图神经网络

图神经网络近年来受到广泛关注，因为它们凭借保留结构的能力，在各种任务中取得了出色的表现 [32, 36]。这类模型基于局部消息传递，因此适合捕捉组织病理图像中的空间信息 [13]。Adnan 等. [1] 从全切片图像中选取最重要的图像块，以此构建全连接图，并使用图神经网络对其进行处理，从而获得整个图的表示。另一方面，Lu 等. [24] 将相邻且相似的图像块合并为图中的一个节点，然后应用图神经网络。在另一项应用中，Zheng等. [48]使用图神经网络提供一种哈希机制，以检索与查询图像在上下文上相似的感兴趣区域。与我们的工作类似，Chen 等. [8] 和 Wang 等. [41]使用细胞图进行生存预测。然而，与我们提出的工作不同，这些研究忽略了细胞类型，并采用单模态设置进行预测。

2.3. 组织病理学中的多模态图像分析

在组织病理学中，结合组织病理图像与组学数据（例如，转录组学和突变）的研究已十分成熟。例如，Vale-Silva 等 [39] 结合组织病理图像、临床信息和 RNA 数据进行生存预测。Chen等 [8]利用克罗内克积，将处理后的组织病理图像数据与基因组学信息融合，以进行生存预测；同一作者 [9] 进一步分析其工作，将结果与胰腺癌中可解释的特征联系起来。尽管组织病理图像与基因组数据的多模态学习已得到广泛研究，但针对不同染色图像的多模态学习应用却在很大程度上被忽视。据我们所知，Dwivedi等的工作 [10] 是目前唯一一项融合不同染色图像进行分级预测的研究。然而，与我们提出的设计不同，他们采用多实例学习设计来处理这一问题。

3. 方法

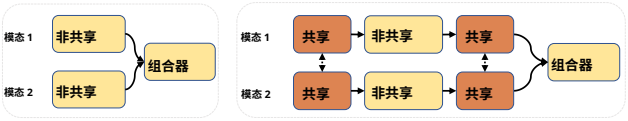
3.1. 问题表述

在本部分中，我们介绍本文其余部分将使用的符号。设 $\{x^m_{n,i}|n = 0, ..., N;\ m = 0, ..., M;\ i = 0, ..., C(n, m)\}$ 为数据集中的图像集合，其中 n 为患者数量，m 为模态数量，i 为图像标识符。在此设定下，N 表示患者总数，M 表示模态总数，$C(n, m)$ 表示患者 n 的模态 m 中可用图像的数量。我们的目标是预测每位患者的估计生存时间，也称为结局。具体而言，我们使用患者的所有可用图像（涵盖不同模态），以获得该患者的统一表示 $R_n \in \mathbb{R}^{1\times d}$，并据此预测其生存时间。为避免重复，本文其余部分假设 $x^m_n,_i$ 同时指图像及由其生成的细胞图。我们将在第 4.1 节说明这一预处理步骤。

3.2. 多模态共享上下文处理

在深入探讨我们的模型细节之前，我们必须先介绍一个新概念，称为共享上下文处理。处理多模态数据的常见策略是使用独立的编码器对每种模态进行编码（图 3a）。然而，我们认为，在处理包含可比上下文的不同模态时，结合共享处理与非共享处理步骤可能会带来益处（例如

。，来自不同染色的细胞图）。具体而言，我们认为，这类场景需要采用三步流程（图 3b）：1）使用共享模型从所有模态中提取基础特征，以提升泛化能力；2）使用各模态的独立模型进行模态特定分析；3）统一高层



(a) 非共享上下文　　(b) 共享上下文

图 3. 共享上下文与非共享上下文多模态处理架构。共享上下文架构能够在低层和高层特征上实现泛化，同时允许在中层进行模态特定的处理。

表征，并由一个适用于所有模态的共享模型处理。这种共享上下文处理方法能够统一低层和高层特征，同时为中层特征处理保留灵活性。在下一节中，我们将说明模型如何受益于这一设计。

3.3. 稀疏多模态细胞图神经网络

3.3.1 模态编码

一般而言，每位患者存在两种变异性：1）来自不同模态的多种图像；2）每种模态内的多种图像。我们的方法的第一阶段处理第二种变异性，并为每种模态设置一个专属处理分支。每个分支包含一个单模态编码器，后接一个实例注意力聚合器；给定 $\{x^m_{n,i}|i = 0, ..., C(n, m)\}$ 作为输入，该分支会生成一个 $R^m_n \in \mathbb{R}^{1\times d}$ 的单一表示向量。

该分支的编码器被设计为一个由三个 GraphSAGE 层构成的图神经网络模型 [14], ，每层后接一个 SAGPool [21]。SAGPooling层使模型能够通过选择图中最重要的节点来执行层次化池化。随后，将每次 SAGPool 后图节点嵌入的平均池化结果和最大池化结果拼接起来，对不同池化层的结果求和，再输入两层多层感知机（多层感知机）。

考虑到来自不同模态的输入图具有相近的上下文，因此可以受益于第3.2节所述的共享上下文处理。为实现低级特征统一，我们通过矩阵分解耦合各分支的第一层。具体而言，我们的 GraphSAGE层遵循公式（1）

$$\hat{h}^m_k = W_s W_m [h^m_k, \frac{1}{K} \sum_{j\in\mathcal{N}_k} h^m_j]. \quad (1)$$

ˆ 在此方程中，h^m_k和 h^m_k分别表示节点 k 在该层前后的嵌入，W_s 是各模态分支共享的权重，W_m 是 m的模态分支专属权重，$[.,.]$ 是拼接操作，$\mathcal{N}_k$ 是与节点 k相连的节点集合，而 K

表示 $\mathcal{N}_k$的大小。对于每个模态分支的第一个 GraphSAGE 层，我们将 W_s 设为可学习矩阵，而其他层则将其设为全 1 矩阵。

为了结合患者在特定模态下多个图的嵌入，我们采用实例注意力，随后进行实例归一化 [38], ，其形式与公式（2）类似。

$$R_n^m = InstanceNorm(\sum_{i=0}^{C(n,m)} \sigma(W R_{n,i}^m) R_{n,i}^m), \quad (2)$$

其中 $R_{n,\ i}^m$ 是 x_n^m对应的表示，i，W 是一个可学习矩阵，σ 是 Sigmoid函数。实例注意力层通过在所有模态间共享W，完成共享上下文处理的高层聚合部分（第 3.2 节）。

3.3.2 跨模态聚合

为了结合不同模态的表征（R_n^m），我们采用了一个 Transformer 模型 [40]。每个注意力头的公式如下，见公式（3）。

$$H_n = softmax(\frac{Q_n K_n^T}{\sqrt{d}}) V_n, \quad (3)$$

其中 Q_n，K_n，V_n 是 $M \times d$ 矩阵，由患者的表征矩阵 n（即 $concat\{R_n^1, ..., R_n^M\}$）经过线性变换得到。最后，将所有注意力头拼接后输入多层感知机，并在各模态间取平均，以生成患者的嵌入。

通过对表征形成整体理解，Transformer 使我们的模型能够对所有染色进行交叉注意力，从而突出最具信息量的染色。

3.3.3 稀疏处理

尽管以往研究强调为各种图相关应用学习输入图的精确拓扑结构 [44], ，但我们发现，我们的多模态方法对信息缺失具有很强的鲁棒性。与计算机视觉领域近期的研究类似（例如，MAE [15]），我们利用这一发现进一步降低模型的计算复杂度。具体而言，在每种模态中，我们对输入图的特征矩阵和邻接矩阵执行掩码操作，如公式（4）所示

$$\hat{X} = MX, \hat{A} = MAM^T, M = P\mathcal{I}, \quad (4)$$

其中，$X \in \mathbb{R}^{c \times d}$ 是节点的特征矩阵，$A \in \mathbb{R}^{c \times c}$ 是图的邻接矩阵，c 是节点数图中的节点，$\mathcal{I} \in \mathbb{R}^{1 \times c}$ 是全 1 矩阵，$P \in \mathbb{R}^{c \times 1}$ 是掩码矩阵，其中每个元素均来自参数为 $1-s$ 的伯努利分布，s 是稀疏率。随着 s 的增加，X 和 A 中的非零元素数量都会减少，从而减少后续计算操作。我们将此称为稀疏处理，并将证明即使稀疏率较高，我们的模型仍能保持性能。

3.3.4 损失函数和 BCP 技术

生存预测是一项具有挑战性的任务，其中包括将失效时间（死亡）估计为连续变量 [20]。从最大似然估计的角度来说，这意味着对于在特定时间失效的受试者，我们必须最大化该受试者相对于其他尚未失效受试者的失效概率。设 t_j 和 $R(t_j)$分别表示受试者 j 的失效时间，以及存活至时间 t_j的受试者集合。受试者 j 的失效概率根据公式（5）计算

$$P_j(T = t_j | R(t_j)) = \frac{P_j(T = t_j | T \geq t_j)}{\sum_{i:t_i \geq t_j} P_i(T = t_j | T \geq t_j)}. \quad (5)$$

我们的训练目标是最大化每个j对应的这一概率。具体而言，一个包含 $\mathcal{B}$ 个样本的小批次上的总损失期望将按公式（6）计算，其中 $\mathcal{U}(.)$ 是受试者上的均匀分布

$$L_{batch} = -E_{i \sim \mathcal{U}(.)}[\log P_i(T = t_i | R(t_i))]. \quad (6)$$

然而，上述损失在实践中存在一个问题。问题源于公式（5）中的损失仅对具有确定失效时间的受试者有定义；对于最近一次随访时仍处于生存状态的受试者（我们通常将这类受试者称为删失数据），该损失则未定义。因此，由于损失未定义（显式梯度），删失受试者在公式（6）的反向传播步骤中，无法作为独立数据点提供任何梯度，从而妨碍模型的正常训练。尽管如此，需要注意的是，这些受试者仍会通过未删失受试者损失的公式（5）分母参与反向传播（隐式梯度）。

为缓解这一问题，我们将损失函数重新表述为式（7），其中 $\mathcal{U}_C(.)$ 和 $\mathcal{U}_N(.)$ 分别是在删失受试者和未删失受试者上的均匀分布，而 k 服从参数为 α的伯努利分布

$$L_{batch} = -E_{i \sim k\mathcal{U}_C(.) + (1-k)\mathcal{U}_N(.)}[\log P_i(T = t_i | R(t_i))]. \quad (7)$$

可以注意到，当 α 等于删失病例所占的百分比时，这两个方程相等。然而，我们将证明，为该参数选择一个合适的值，可以在删失数据的隐式梯度和显式梯度之间实现平衡。我们将 α 称为批次删失比例（BCP），并说明它会对结果产生显著影响。

4. 实验

4.1. 数据集和预处理步骤

本研究使用了两个免疫组织化学（IHC）数据集：（1）InUIT：内部高级别浆液性卵巢癌队列，包含来自188例患者的1,600个经Ki67、CD8和CD20生物标志物染色的组织芯片核心；（2）MIBC：肌层浸润性膀胱癌队列，包含来自58例患者的585个经Ki67、CK20和P16染色的组织芯片核心[26]。每位患者至少有一个经每种生物标志物染色的TMA组织芯片核心，且所有患者均有最新的生存状态（存活或死亡）以及总生存时间（自诊断起计算）记录。

TMA组织芯片核心中的每个细胞会呈现红色或蓝色，这表示该细胞对于相应生物标志物呈阳性或阴性（图1）。我们对每个TMA组织芯片核心应用细胞分割算法，以定位细胞并识别其类型（阳性/阴性）[12]。随后，将每个细胞视为图中的一个节点，并使用K近邻算法将这些节点彼此连接（K=5）。类似于 [18], 我们假设细胞间通信距离受到生物学限制。因此，我们移除了长度超过60像素的边。随后，在图像中以相应细胞为中心的 72×72 裁剪图像上应用预训练的ResNet34，获得每个节点的嵌入表示。此外，节点类型（阳性/阴性）及其位置（相对于图像尺寸）也被添加到嵌入中。

4.2. 实现细节

所有实验均在单个 GeForce RTX 3090 上使用 PyTorch 和 DGL 软件包完成。每个分支中的GraphSAGE层和多层感知机层的输出特征大小分别设置为 128 和 32。模型训练采用 Adam优化器，学习率为 0.002、余弦调度器、权重衰减为 0.0001，批次大小为 128。跨模态聚合器的 Transformer模块包含 4 个多头自注意力模块，并将 BCP 设置为 0.1。我们还使用了 0.8 的稀疏率，该设置仅应用于训练阶段。

4.3. 生存预测

生存期预测结果汇总见表1。为比较不同模型，与以往研究类似，我们 [8], 采用一致性指数（C-Index）来衡量患者生存排序的质量 [33]。此外，我们使用 LogRank 检验的 P 值来展示模型区分高风险和低风险患者的能力（更多信息见第4.7 节）。所有实验均采用按患者划分的 3 折交叉验证设置。与以往研究不同的是，我们使用 3 个不同的随机种子进行每项实验，以考虑初始化变异性。结果证实，我们的模型优于所有基线方法（包括 ViT），且在两个数据集上的两项指标中均表现稳定，而基线方法则不然。具体而言，我们的模型在 InUIT 和 MIBC 数据集上的一致性指数分别达到0。57 和0。61，而表现最接近的基线方法分别为0。55 和0。59。此外，我们的模型能够在两个数据集上显著区分低风险和高风险患者（P 值 < 0.01），且是在 InUIT 数据集上唯一做到这一点的方法。值得一提的是，少数基线模型的一致性指数为0。5 （相当于随机预测），这可能归因于前述基于多实例学习的技术所存在的问题。我们还注意到，与基线方法相比，我们的模型性能波动更小、参数数量更少。这一观察结果体现了模型的泛化能力和效率，这可能与其细胞基础有关。

基线模型的设置与 [7] 的设置类似。尽管原始设置使用了维度为 1024 的 ResNet50 进行特征提取，但我们也使用 ResNet34 开展了实验，以确保我们的模型与基线方法之间进行公平比较（更多细节和结果请参见补充材料）。

4.4. 消融研究

我们对模型的不同部分进行了消融研究，结果见表 2。这些实验包括：移除实例注意力后的实例归一化（无实例归一化）、解耦模态分支权重（无权重共享）、完全耦合模态分支权重（完全权重共享）、解耦实例注意力层的权重（非共享注意力）、不考虑批次删失比例（无 BCP）、使用 Transformer 替代实例注意力（Transformer注意力），以及在推理时应用稀疏性（推理时稀疏性）。结果表明，根据数据集的不同，每种消融特征都会导致模型性能显著下降，凸显了各项设计选择的重要性（更多结果见补充材料）。

如前所述，尽管我们数据中的不同模态代表不同的染色，但我们认为所有这些模态都包含共享的上下文信息。因此，我们可以利用这一点，采用前文介绍的共享上下文处理思路。

方法	特征提取器	参数	InUIT		MIBC	
			一致性指数 (↑)	P值 (↓)	一致性指数（↑）	P值（↓）
DeepSet	ResNet34	$395K$	0.50 ± 0.0	0.43	0.50 ± 0.001	–
	ResNet50	$657K$	0.53 ± 0.007	0.40	0.45 ± 0.004	0.28
注意力多实例学习	ResNet34	$657K$	0.51 ± 0.004	0.62	0.59 ± 0.007	0.04
	ResNet50	$920K$	0.55 ± 0.004	0.65	0.55 ± 0.004	0.57
DGC	ResNet34	$658K$	0.53 ± 0.007	0.46	0.58 ± 0.007	$< \mathbf{0.001}$
	ResNet50	$790K$	0.55 ± 0.005	0.31	0.54 ± 0.007	0.64
Patch-GCN	ResNet34	$1.3M$	0.53 ± 0.008	0.45	0.50 ± 0.004	0.005
	ResNet50	$1.4M$	0.50 ± 0.004	0.25	0.46 ± 0.009	0.33
Pathomic Fusion	CPC	$368K$	0.51 ± 0.001	0.43	0.52 ± 0.003	0.56
HIPT	层次化视觉 Transformer	$23.8M$	0.50 ± 0.002	0.18	0.53 ± 0.010	0.1
AMIGO（我们的方法）	ResNet34	$451K$	$\mathbf{0.57 \pm 0.002}$	**0.01**	$\mathbf{0.61 \pm 0.004}$	$< \mathbf{0.001}$

表1. 我们的模型与所有基线方法在两个数据集上的生存预测性能比较。

消融特征	一致性指数 (↑)	
	InUIT	MIBC
基线	$\mathbf{0.57 \pm 0.002}$	$\mathbf{0.61 \pm 0.004}$
无实例归一化	0.53 ± 0.001	0.54 ± 0.005
无权重共享	0.56 ± 0.001	0.54 ± 0.016
完全权重共享	0.53 ± 0.001	0.51 ± 0.005
无 BCP	0.54 ± 0.001	0.58 ± 0.013
Transformer注意力	0.53 ± 0.002	0.55 ± 0.011
推理时稀疏性	0.56 ± 0.001	0.55 ± 0.004
非共享注意力	0.54 ± 0.001	0.58 ± 0.003

表 2. 消融研究。

我们关于消除此步骤的消融实验（表2中的无权重共享和非共享注意力行）证实了这一假设。另一方面，有人可能会认为，对所有模态使用同一个网络也能达到这一目的。然而，相应的消融研究（表中的完全权重共享行）否定了这一观点。此外，由于所有模态都通过共享实例注意力进行处理，我们的模型可以在将嵌入传递给跨模态聚合器之前，先使用归一化层（无实例归一化），从而获益。

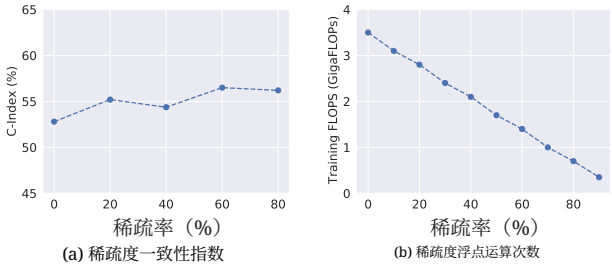
移除 BCP 后的结果也表明，删失数据与非删失数据的比例权衡十分重要，因为这可以改善反向传播中的梯度信号。最后，在推理时避免引入稀疏性可以提升性能，因为模型能够获取作出预测所需的全部信息。

4.5. 稀疏性鲁棒性与计算效率

我们研究中最重要的发现之一是，模型能够稳健地应对训练数据稀疏性。具体而言，我们发现，无论输入图的稀疏率如何，模型的最终性能都保持稳定。尤其是，尽管以往的数字组织病理学研究 [28]表明，学习细胞图的完整拓扑结构对于下游任务至关重要，但我们注意到，较低的稀疏率（20%）可以提升模型性能（图4a）。这一观察结果与以往的发现一致：数据增强能够防止过拟合，从而使深度学习模型受益 [34]。然而，令人意外的是，随着稀疏率提高，模型的性能几乎保持不变。另一方面，稀疏率与模型的计算成本呈线性反向关系（图4b），这表明较高的稀疏率会减少计算操作的次数。因此，在保持相同性能的同时，模型的计算成本可以显著降低（从 3.5 降至 0.7 十亿次浮点运算）。

4.6. BCP 效果

我们还测量了 BCP 对模型性能的影响。如图所示5，BCP 为 0（批次中没有删失数据）时，由于训练过程中显式梯度信号增加，其性能优于典型的均匀批处理。另一方面，较高的 BCP 值会消除显式梯度，因此其性能低于均匀批处理。不过，图中结果证实了我们的假设：在隐式梯度与显式梯度之间进行权衡，选择合适的 BCP 值（0.1）可实现最高性能。



(a) 稀疏度一致性指数　(b) 稀疏度浮点运算次数

图 4. 图 4a 表明，我们的模型在 InUIT 数据集上的最终性能对训练数据的稀疏性具有鲁棒性。图 4b表明，我们模型的计算成本（浮点运算次数）与稀疏率呈反向线性关系。

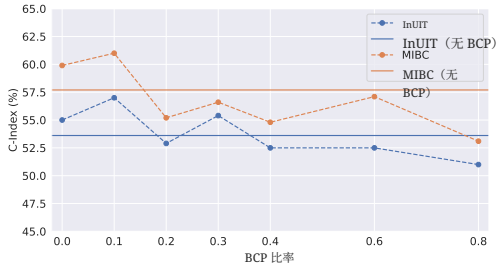
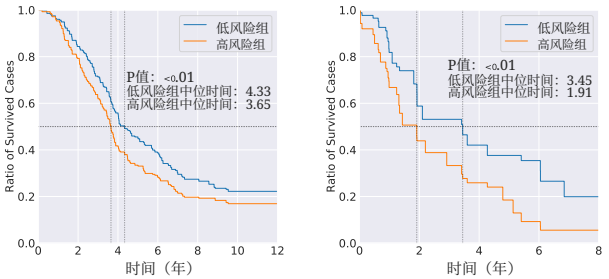


图 5. 基于 BCP 比率的模型性能。较低的 BCP 值会导致删失病例的显式梯度增加、隐式梯度减少。这两类梯度之间的权衡可以带来最高的性能。



(a) InUIT　(b) MIBC

图 6. 模型识别出的低风险（预测风险 ＜低于风险中位数）和高风险（预测风险 ＞高于风险中位数）患者队列的生存曲线。

4.7. 患者分层

虽然一致性指数可用于衡量各种生存预测模型的表现，但对于患者管理而言，它提供的信息并不充分。若要使基于机器学习的生存预测模型能够应用于临床，就需要证明它们能够将患者划分为不同风险组。因此，我们根据模型预测的风险评分，将每个队列中的患者分为两组（即低风险组和高风险组）。这些队列的 Kaplan-Meier（KM）生存曲线见图6。为了检验低风险与高风险患者类别的 KM 曲线之间差异的显著性，我们采用了对数秩检验 [2]。如图6所示，我们的模型成功地将卵巢癌和膀胱癌病例分别划分为低风险和高风险队列，凸显了模型从组织病理图像中提取有意义的上下文信息的能力。具体而言，InUIT数据集中高风险和低风险队列的中位生存时间分别为3.65 和4.33 年（对数秩检验 P 值 = 0.01），而 MIBC数据集中分别为 1.91 和3.45 年（对数秩检验 P 值 ＜ 0.001）。值得一提的是，我们的发现与此前的一项研究一致关于膀胱癌数据集的研究，其中 Mi 等人。 [26]针对肌层浸润性膀胱癌患者也实现了类似的分组。尽管他们使用了从细胞中人工设计的特征，但我们在考虑细胞间相互作用的同时，以端到端可训练方式处理这一问题。我们的卵巢癌数据集代表一种侵袭性极强的亚型（即高级别浆液性），过去几十年来，大多数研究（尽管主要未获成功）都致力于为这些患者寻找治疗反应的生物标志物。Wang 的一项研究等人 [42]表明，可以从全局基因组异常谱中找到此类标志物；而我们的研究首次基于常规组织病理学切片图像取得了令人鼓舞的结果。

5. 结论

在本研究中，我们首次开发了一种多模态图神经网络，通过聚焦细胞及其相互作用来处理组织病理图像。除了引入批次删失比例（BCP）和共享上下文处理等新技术外，我们还证明，所提出的模型在代表卵巢癌和肌层浸润性膀胱癌的两个数据集上，性能优于所有同类模型。更重要的是，我们证明了该模型对数据稀疏性具有很强的鲁棒性，即使训练时仅使用 20% 的数据，仍能取得相对相近的性能。利用这一发现，我们得以进一步降低模型的计算成本。我们还评估了该模型作为患者分层工具的适用性；它能够将患者划分为具有统计学显著差异的低风险组和高风险组。

我们认为，所提出的模型凸显了细胞的异质性、空间位置和相互作用对于不同癌症类型图像表征的重要性。我们希望这项工作能够为高效的细胞层面处理开辟新的有趣途径——

组织病理图像。鉴于我们的模型在对仅凭基因组信息才能区分的患者队列进行分层方面取得了成功，我们可以将其用作连接组织病理图像与基因表达、突变及基因组特征的引擎，从而开展更深入的分析和生物学探究。此外，我们方法的细胞中心性为识别可视化且可解释的生物学实体提供了机会；这些实体在预测结局方面发挥关键作用，也可应用于临床。

致谢

本研究获得Terry Fox Research Institute、加拿大卫生研究院、加拿大自然科学与工程研究理事会、Michael Smith Foundation for Health Research、OVCARE Carraresi、温哥华综合医院—不列颠哥伦比亚大学医院基金会以及美国国立卫生研究院的支持（资助编号：R01CA138264）。

参考文献

[1] Mohammed Adnan、Shivam Kalra和Hamid R Tizhoosh。使用图神经网络进行组织病理图像的表征学习。载于*IEEE/CVF*计算机视觉与模式识别会议研讨会论文集，第988–989页，2020年。3[2] J Martin Bland和Douglas G Altman。LogRank检验。*Bmj*，328(7447):1073，2004年。8[3] Joseph Boyd、Mykola Liashuha、Eric Deutsch、Nikos Paragios、Stergios Christodoulidis和Maria Vakalopoulou。利用视觉视野扩展进行数字病理学的自监督表征学习。载于*IEEE/CVF*国际计算机视觉大会论文集，第639–647页，2021。1[4] Wieland Brendel和Matthias Bethge。使用局部特征袋模型近似卷积神经网络，在ImageNet上取得了出人意料的良好效果。*arXiv*预印本*arXiv:1904.00760*，2019年。3[5] Gabriele Campanella、Matthew G Hanna、Luke Geneslaw、Allen Miraflor、Vitor Werneck Krauss Silva、Klaus J Busam、Edi Brogi、Victor E Reuter、David S Klimstra和Thomas J Fuchs。利用弱监督深度学习对全视野切片图像进行临床级计算病理学分析。《自然·医学》，25(8):1301–1309，2019年。3[6] Richard J Chen、Chengkuan Chen、Yicong Li、Tiffany Y Chen、Andrew D Trister、Rahul G Krishnan和Faisal Mahmood。通过分层自监督学习将视觉变换器扩展至十亿像素图像。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第16144–16155页，2022。1，2，3[7]Richard J Chen、Ming Y Lu、Muhammad Shaban、Chengkuan Chen、Tiffany Y Chen、Drew FK Williamson和Faisal Mahmood。全视野切片图像是二维点云：使用基于图块的图卷积网络进行上下文感知的生存预测。载于医学图像计算与计算机辅助干预国际会议，第339–349页。施普林格，2021。6

[8] Richard J Chen、Ming Y Lu、Jingwen Wang、Drew FK Williamson、Scott J Rodig、Neal I Lindeman和 Faisal Mahmood。《Pathomic fusion：一种融合组织病理学与基因组特征以进行癌症诊断和预后的集成框架》。IEEE医学影像汇刊，2020年。3、4、6[9] Richard J Chen、Ming Y Lu、Drew FK Williamson、Tiffany Y Chen、Jana Lipkova、Zahra Noor、Muhammad Shaban、Maha Shady、Mane Williams、Bumjin Joo 等。《通过多模态深度学习进行泛癌整合组织学—基因组分析》。《癌细胞》，40(8):865–878，2022。4[10] Chaitanya Dwivedi、Shima Nofallah、Maryam Pouryahya、Janani Iyer、Kenneth Leidal、Chuhan Chung、Timothy Watkins、Andrew Billin、Robert Myers、John Abel 等。《用于病理学多模态整合的多染色图融合》。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第1835 1845页，2022年 4 – 。[11] Jevgenij Gamper 和 Nasir Rajpoot。《多实例图注生成：从组织病理学教科书和文章中学习表征》。发表于IEEE/CVF计算机视觉与模式识别会议论文集，第16549–16559页，2021年。1[12] Parmida Ghahremani、Yanyun Li、Arie Kaufman、Rami Vanguri、Noah Greenwald、Michael Angelo、Travis J Hollmann 和 Saad Nadeem。《DeepLIIF：用于免疫组化图像定量的深度学习推断多重免疫荧光》。bioRxiv，2021年。6[13] Yonghang Guan、Jun Zhang、Kuan Tian、Sen Yang、Pei Dong、Jinxi Xiang、Wei Yang、Junzhou Huang、Yuyao Zhang 和 Xiao Han。《用于全切片图像表征与分类的节点对齐图卷积网络》。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第18813 18823页，2022年 3 – 。[14] Will Hamilton、Zhitao Ying 和 Jure Leskovec。《大型图上的归纳式表征学习》。神经信息处理系统进展，30，2017。4[15] Kaiming He、Xinlei Chen、Saining Xie、Yanghao Li、Piotr Doll´ar 和 Ross Girshick。《掩码自编码器是可扩展的视觉学习器》。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第16000–16009页，2022年。5[16] Le Hou、Dimitris Samaras、Tahsin M Kurc、Yi Gao、James E Davis 和 Joel H Saltz。《用于十亿像素分辨率图像分类的高效多实例卷积神经网络》。arXiv预印本arXiv:1504.07947，7:174–182，2015年。3[17] Maximilian Ilse、Jakub Tomczak 和 Max Welling。《基于注意力的深度多实例学习》。发表于国际机器学习会议，第2127–2136页。PMLR，2018年。1、3[18] Sajid Javed、Arif Mahmood、Muhammad Moazam Fraz、Navid Alemi Koohbanani、Ksenija Benes、Yee-Wah Tsang、Katherine Hewitt、David Epstein、David Snead 和 Nasir Rajpoot。《用于结直肠癌组织学图像组织表型分析的细胞群落检测》。Medical Image Analysis，63:101696，2020年。6

[19] Zhipeng Jia、Xingyi Huang、I Eric、Chao Chang 和 Yan Xu。《用于组织病理学图像分割的受约束的深度弱监督》。IEEE医学影像汇刊，36(11):2376–2388，2017。3[20] Jared L Katzman、Uri Shaham、Alexander Cloninger、Jonathan Bates、Tingting Jiang 和 Yuval Kluger。《Deep-surv：使用 Cox比例风险深度神经网络的个性化治疗推荐系统》。BMC医学研究方法学，18(1):1–12，2018年。5[21] Junhyun Lee、Inyeop Lee 和 Jaewoo Kang。《自注意力图池化》。发表于国际机器学习会议，第3734–3743页。PMLR，2019。4[22] Bin Li、Yin Li 和 Kevin W Eliceiri。《用于全切片图像分类并结合自监督对比学习的双流多实例学习网络》。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第14318–14328页，2021。3[23] Jiayun Li、Wenyuan Li、Anthony Sisk、Huihui Ye、W Dean Wallace、William Speier 和 Corey W Arnold。《一种结合多实例学习的多分辨率组织病理学图像分类与定位模型》。Computers in Biology and Medicine，131:104253，2021。3[24] Wenqi Lu、Simon Graham、Mohsin Bilal、Nasir Rajpoot 和 Fayyaz Minhas。《捕捉多十亿像素病理图像中的细胞拓扑结构》。发表于IEEE/CVF计算机视觉与模式识别会议研讨会论文集，第260–261页，2020年。3[25] Armin Meier、Katharina Nekolla、Lindsay C Hewitt、Sophie Earle、Takaki Yoshikawa、Takashi Oshima、Yohei Miyagi、Ralf Huss、G¨unter Schmidt 和 Heike I Grabsch。《将无假设深度生存学习应用于胃癌肿瘤微环境》。《病理学杂志：临床研究》，6(4):273–282，2020年。2[26] Haoyang Mi、Trinity J Bivalacqua、Max Kates、Roland Seiler、Peter C Black、Aleksander S Popel 和 Alexander S Baras。《利用细胞核形态和组织结构预测肌层浸润性膀胱癌新辅助化疗反应的模型》。Cell Reports Medicine，2(9):100382，2021。6、8[27] Ramin Nakhli、Amirali Darbandsari、Hossein Farahani 和 Ali Bashashati。《CCRL：对比式细胞表征学习》。发表于《计算机视觉——ECCV 2022研讨会：以色列特拉维夫，2022年10月23—27日，论文集，第VII部分》，第397–407页。施普林格，2023年。1[28]Pushpak Pati、Guillaume Jaume、Antonio Foncubierta-Rodr´ıguez、Florinda Feroce、Anna Maria Anniciello、Giosue Scognamiglio、Nadia Brancati、Maryse Fiche、Estelle Dubruc、Daniel Riccio 等。《数字病理学中的层次化图表征》。Medical Image Analysis，75:102264，2022年。7[29] Simon Reiß、Constantin Seibold、Alexander Freytag、Erik Rodner 和 Rainer Stiefelhagen。《每项标注都至关重要：用于医学图像分割的多标签深度监督》。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，第9532 9542页，2021年 1 – 。[30] Yash Sharma、Aman Shrivastava、Lubaina Ehsan、Christo- pher A Moskaluk、Sana Syed 和 Donald Brown。《聚类-to-conquer：一种用于全切片图像分类的端到端多实例学习框架。载于《Medical Imaging with Deep Learning》，第682–698页。PMLR，2021年。1[31] Beomseok Son、Sungmin Lee、HyeSook Youn、EunGi Kim、Wanyeon Kim和BuHyun Youn。肿瘤微环境在治疗耐药性中的作用。《Oncotarget》，8(3):3933，2017年。2[32] Maximilian Stadler、Bertrand Charpentier、Simon Geisler、Daniel Z¨ugner和Stephan G¨unnemann。图后验网络：用于节点分类的贝叶斯预测不确定性。《神经信息处理系统进展》，34:18033–18048，2021。3[33] Harald Steck、Balaji Krishnapuram、Cary Dehing-Oberije、Philippe Lambin和Vikas C Raykar。论生存分析中的排序：一致性指数的界限。《神经信息处理系统进展》，20，2007年。6[34] Susheel Suresh、Pan Li、Cong Hao和Jennifer Neville。通过对抗式图增强改进图对比学习。《神经信息处理系统进展》，34:15920–15933，2021。7[35] Aiham Taleb、Matthias Kirchler、Remo Monti和Christoph Lippert。Contig：结合遗传学的医学影像自监督多模态对比学习。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第20908–20921页，2022年。1[36] Shixiang Tang、Dapeng Chen、Lei Bai、Kaijian Liu、Yixiao Ge和Wanli Ouyang。面向少样本学习的互助CRF-GNN。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第2329–2339页，2021年。3 – .[37] Yucheng Tang、Dong Yang、Wenqi Li、Holger R Roth、Bennett Landman、Daguang Xu、Vishwesh Nath和Ali Hatamizadeh。用于三维医学图像分析的Swin Transformer自监督预训练。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第20730–20740页，2022年。1[38] Dmitry Ulyanov、Andrea Vedaldi 和Victor Lempitsky。实例归一化：快速风格化所缺少的关键要素。arXiv预印本 arXiv:1607.08022，2016年。5[39] Lu´ıs A Vale-Silva和Karl Rohr。使用多模态深度学习进行长期癌症生存预测。《Scientific Reports》，11(1):1–12，2021年。4[40] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser和Illia Polosukhin。Attention Is All You Need。《神经信息处理系统进展》，30，2017年。5[41] Yanan Wang、Yu Guang Wang、Changyuan Hu、Ming Li、Yanan Fan、Nina Otter、Ikuan Sam、Hongquan Gou、Yiqun Hu、Terry Kwok等。细胞图神经网络可精准预测胃癌患者生存。《NPJ Precision Oncology》，6(1):1–12，2022年。2, 3[42] Yi Kan Wang、Ali Bashashati、Michael S Anglesio、Dawn R Cochrane、Diljot S Grewal、Gavin Ha、Andrew McPherson、Hugo M Horlings、Janine Senz、Leah M Prentice等。异常DNA修复机制的基因组后果可对卵巢癌组织学类型进行分层。《Nature Genetics》，49(6):856–865，2017年。8

[43] An Xu、Wenqi Li、Pengfei Guo、Dong Yang、Holger R Roth、Ali Hatamizadeh、Can Zhao、Daguang Xu、Heng Huang和Ziyue Xu。弥合跨机构联邦医学图像分割的泛化差距。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第20866–20875页，2022年。1[44] Keyulu Xu、Weihua Hu、Jure Leskovec和Stefanie Jegelka。图神经网络的能力有多强？*arXiv*预印本*arXiv:1810.00826*，2018年。5[45] Yinyin Yuan。肿瘤微环境中的空间异质性。《冷泉港医学展望》，6(8):a026583，2016年。2[46] Manzil Zaheer、Satwik Kottur、Siamak Ravanbakhsh、Barnabas Poczos、Russ R Salakhutdinov和Alexander J Smola。深度集合。《神经信息处理系统进展》，30，2017年。3[47] Hongrun Zhang、Yanda Meng、Yitian Zhao、Yihong Qiao、Xiaoyun Yang、Sarah E Coupland和Yalin Zheng。DTFD-MIL：用于组织病理学全切片图像分类的双层特征蒸馏多实例学习。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第18802–18812页，2022年。3[48] Yushan Zheng、Bonan Jiang、Jun Shi、Haopeng Zhang和Fengying Xie。使用图神经网络编码组织病理学全切片图像，实现可扩展的诊断相关区域检索。载于医学图像计算与计算机辅助干预国际会议论文集，第550–558页。施普林格，2019年。3