arXiv:2305.05006v2 [eess.IV] 4 Apr 2024

基于腺体布局合成带注释的结肠癌组织图像

Srijay Deshpande、Fayyaz Minhas 和 Nasir Rajpoot

组织图像 华威大学影像分析中心，英国考文垂 m

1. 摘要

生成带有标注的逼真组织图像是一项具有挑战性的任务，在许多计算病理组织学应用中都十分重要。合成生成的图像和标注对于训练和评估该领域的算法具有重要价值。为此，我们提出了一种交互式框架，可根据腺体结构布局生成成对的逼真结直肠癌组织学图像及其对应的腺体掩膜。该框架能够准确捕捉基质、杯状细胞和腺腔等关键特征。用户可以通过调整腺体数量、位置和大小等参数来控制腺体外观。与最先进的图像到图像转换模型相比，生成图像表现出良好的 Fréchet Inception 距离（FID）得分。此外，我们展示了合成标注在评估腺体分割算法方面的实用性。进一步地，我们提出了一种利用潜在扩散模型等先进深度生成模型构建腺体掩膜的方法。这些掩膜能够通过残差编码器-解码器网络生成组织图像。

关键词： Comp 计算病理学、生成对抗网络、扩散模型、深度学习 g

2. 引言

深度学习算法需要大量训练数据，这些数据可能难以获取且成本高昂，还可能需要高技能病理学家的参与。人们已提出许多生成高质量合成组织图像的技术。[1–4] 合成组织学图像有助于开发用于下游任务的预测模型，[3,5,6] 也可用于教育目的[7]和临床质量保证。[8]上述大多数合成组织学图像生成应用场景，都需要根据自定义腺体布局或用户定义的组织参数（例如腺体数量和疾病分级）合成图像。

在病理图像分析中，由于组织异质性以及实验室获取的组织图像存在差异，为了达到最佳性能，通常必须针对不同的组织类型（如腺体、脂肪组织和血管）重复进行数据标注。对于腺体分割等任务，手动生成突出显示腺体部分的组件掩膜，既费力又耗时。这给分割算法生成大量标注数据造成了重大障碍。一些研究人员探索了根据组织成分掩膜生成合成病理图像的方法。[5,9,10] 这些方法要么假定输入组件掩膜已存在，要么需要通过为细胞核等相应组织成分生成随机形状来显式构建组件掩膜，而这种做法可能导致错误且不切实际的结果。此外，为腺体等较大的多细胞结构制作组件掩膜也颇具挑战。因此，同时生成合成图像和组件掩膜更为理想，因为这样既有望降低标注成本，也能生成逼真的标注对。

在这项工作中，我们提出了一种用户交互式框架，能够利用输入的腺体布局，同时生成结直肠组织图像及其对应的组织成分掩膜。据我们所知，这是首个能够根据腺体布局控制生成带注释结直肠组织图像的框架。借助这一布局，用户可以指定结直肠组织图像中腺体的位置和大小。我们利用生成对抗网络（GANs）在对抗设置下的特性来创建逼真的组织图像。此外，我们提出了一种基于潜在扩散模型[11] 的方法，用于合成

作者补充信息：（通信请联系 Srijay Deshpande）Srijay Deshpande：电子邮箱：deshpandesrijay@gmail.com

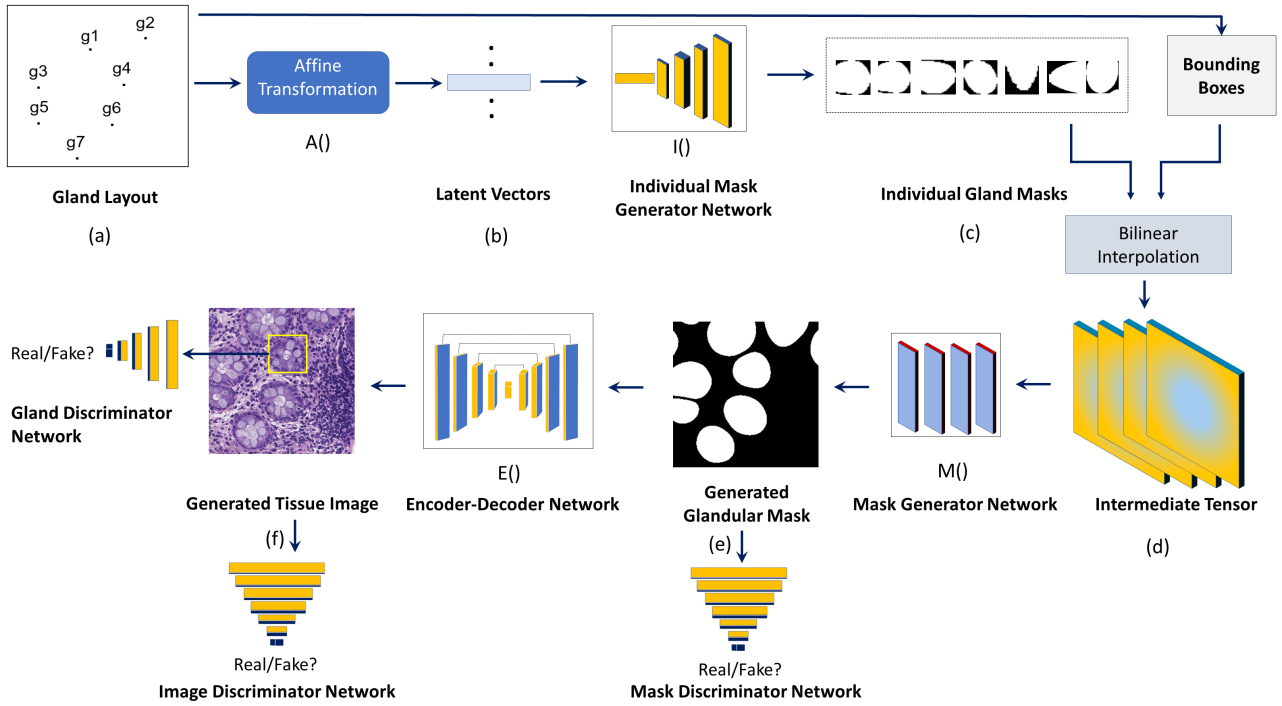


图1. 所提出框架的框图：(a) 显示输入的腺体布局，其中腺体排列在二维空间位置上。每个腺体由一个腺体专属向量表征，该向量经过仿射变换形成潜在向量 (b)。每个潜在向量输入单个掩膜生成器，生成二值单个腺体掩膜 (c)。生成的单个掩膜与输入边界框一起作为双线性插值算法的输入，该算法将生成的掩膜包裹在边界框内，从而创建中间张量 (d)。掩码生成网络使用中间张量并生成腺体掩膜 (e)，随后将其输入编码器-解码器生成网络，生成最终的组织图像 (f)。这里使用了三个判别器，分别用于生成的掩膜、图像以及图像中的腺体部分。

腺体掩膜，而无需依赖特定的腺体布局或类似输入。驯服变换器、[12] OpenAI 的 DALL-E、[13] 稳定扩散[11] 等前沿方法也采用了基于量化的架构来生成逼真的图像。我们工作的主要亮点如下：

1. 我们提出了一种框架，能够同时生成逼真的结直肠组织图像及其对应的组织成分/腺体掩膜。2. 所提出的框架允许用户通过调整腺体的位置和大小来改变腺体的外观。据我们所知，这是首个能够根据腺体布局生成受其控制的标注结直肠组织图像的框架。3. 我们展示了使用所提出的框架生成的标注对在评估腺体分割算法方面的有效性。4. 我们提供了一种基于向量量化变分自编码器（VQ-VAE）的方法，并结合扩散模型加以完善。这种组合能够生成逼真的腺体掩膜，随后利用残差编码器-解码器网络生成组织图像。

3. 方法

我们的目标是开发一个框架，根据输入的腺体布局构建结直肠癌组织图像及其对应的组织成分掩膜，以突出显示腺体区域。腺体布局可以描述为

作为第一象限的笛卡尔平面，用户可以在其二维空间位置上排列腺体。框架 f 使用腺体布局，首先构建组织成分掩膜，随后借此生成大小为 $N \times N$ 像素的完整组织图像。整体概览见图1。下面我们介绍所提出的框架的主要组成部分：

输入的腺体布局可视为由 n 个腺体组成的集合， $G_{layout} = \{g_k \equiv (\overrightarrow{l_k}, \overrightarrow{s_k}) \mid k = 1, 2...n\}$，其中 $\overrightarrow{l_k}$和 $\overrightarrow{s_k}$分别表示腺体 g_k的位置向量和大小向量。腺体 g_k由一个特定于该腺体的向量 $\overrightarrow{z_k}$表征，该向量从高斯噪声 $z \sim \mathcal{N}(0, 1)$中采样。高斯噪声用于确保最终图像中腺体对象的外观具有多样性。腺体特定向量 $\overrightarrow{z_k}$经过仿射变换 A，生成维度为 D的 n潜在嵌入 $\{a_k \mid k = 1, 2...n\}$，即 $a_k = A(\overrightarrow{z_k}; \theta_A)$，其中 θ_A 表示该函数的可训练权重。

3.1 腺体掩膜的生成

生成的潜在嵌入随后输入独立掩膜生成网络 I，以生成对应的单个腺体二值掩膜 $\{m_k \mid k = 1, 2...n\}$，每个掩膜的大小为 $B \times B$像素，即 $m_k = M(a_k; \theta_I)$，其中 θ_I 表示该网络的可训练参数。独立掩膜生成网络 由一系列模块组成，每个模块均包含一个转置卷积层，后接 ReLU 激活。

下一步是将生成的二值掩膜对齐到适当的位置，并构造与最终图像尺寸相同的组织成分掩膜，即 $N \times N$ 像素。为此，我们使用每个对象的输入边界框 $\{b_k \mid k = 1, 2...n\}$。这些边界框要么从数据集中获取（流程见第 4 节），要么根据输入的位置和尺寸参数生成（第 5.6 节）。将每个腺体特定向量 $\overrightarrow{z_k}$ 与相应的腺体掩膜 m_k逐元素相乘，并使用固定的双线性插值函数[14] F将其映射到边界框 b_k 的位置，从而得到维度为$D \times B \times B$的中间张量，即 $C = F(\{\overrightarrow{z_k}, m_k, b_k \mid k = 1, 2...n\})$。该中间张量有 D个通道，随后通过掩码生成网络 M将其缩减为 1 个通道，得到腺体掩膜或组织成分掩膜，即 $T = M(C, \theta_M)$，其中 θ_M表示其可训练参数。

3.2 组织图像生成

生成组织成分掩膜后，我们将其输入编码器-解码器生成网络 E。图像到图像转换编码器-解码器网络用作图像生成器，以构建最终的组织图像 $Z = E(T, \theta_E)$。

编码器由一系列（卷积 + ReLU）块组成，用于生成输入图像的固定尺寸编码。解码器通过一系列（转置卷积 + ReLU）块构建最终图像。受 Pix2Pix[15] , 启发，该网络 E 还在特征图尺寸相同的层之间采用跳跃连接，使第一个下采样层与最后一个上采样层相连，第二个下采样层与倒数第二个上采样层相连，依此类推。这些跳跃连接使图像生成器能够灵活地绕过编码部分，将信息传递至后续层，并使生成器能够利用早期编码块中的低级特征。

3.3 判别器

我们在对抗训练框架中使用 3 个判别器神经网络：掩码判别器 $D_M(M, \theta_{M_D})$用于生成的腺体掩膜（M），图像判别器 $D_Z(Z, \theta_{Z_D})$用于生成的组织图像 Z，以及腺体判别器 $D_G(Z_{g_i}, \theta_{G_D})$用于组织图像 $\{Z_{g_i} \mid i = 1, 2..n\}$ 内部的腺体部分，其中 n 是腺体数量； θ_{M_D}、 θ_{Z_D} 和 θ_{G_D} 分别表示这些判别器各自的可训练参数。前两个判别器采用 PatchGAN[15] 判别器，分别预测生成的组件掩膜和组织图像中不同部分的真实程度。基于这些判别器的对抗损失可确保组织成分掩膜和组织图像具有真实性。

腺体判别器由一系列卷积操作组成，并根据输入边界框，对从最终组织图像中裁剪出的生成腺体部分进行双线性插值，将其调整为固定大小，然后预测其真实性的单一分数。[14] 它确保生成的腺体逐一呈现真实的外观，包括杯状细胞和腔等微观组成部分。

我们对所有判别器采用对抗损失函数[16]。判别器 $D_t(X,\ \theta_{t_D})$ 尝试通过对生成器函数 $G(X,\ \theta_G)$ 生成的输入图像 X 进行分类来最大化损失，而生成器函数则试图最小化损失；其中 θ_{t_D} 和 θ_G 分别表示相应网络的可训练参数集合。 D_t 表示判别器类型（ D_M、 D_Z 或 D_G）。对抗式极小极大损失函数如下：

$$\min_{\theta_G}\max_{\theta_{t_D}} L^t_{GAN}(\theta_G,\theta_{t_D}) = E_{X\sim p_{data}(X)}[logD_t(X,\theta_{t_D})] + E_{X\sim p_X(X)}[log(1-D_t(G(X,\theta_G),\theta_{D_t})] \tag{1}$$

上述所有网络的具体架构见附录5。

3.4 损失分量

由于涉及多个网络，所提出的框架的训练损失由若干项组成。下面介绍该框架中使用的不同损失分量：

通过最小化由以下分量构成的损失函数，对包含可训练参数 $\{\theta_A,\theta_I,\theta_M,\theta_E,\theta_{M_D},\theta_{Z_D},\theta_{G_D}\}$ 的完整框架进行训练：

单个二值腺体掩膜重建损失： 该分量使用均方误差（MSE）来惩罚真实值 $\{\hat{}m_k \mid k=1,\ 2...n\}$ 与生成的单个二值腺体掩膜 $\{m_k \mid k=1,2...n\}$ 之间的差异，具体如下：

$$L_{GlandMaskRec}(\theta_A,\theta_I) = \sum_{k=1}^{n} MSE(\hat{m_k}, m_k) \tag{2}$$

其中，$\hat{}m_k$ 是真实值， m_k 是生成的单个腺体掩膜， n 是组织图像中的腺体数量。如第 3 节所述， m_k 取决于可训练参数 θ_A 和 θ_I。类似地，我们可以定义其他损失分量。

ˆ掩膜重建损失： 此损失用于惩罚真实值 T 与生成的组织成分掩膜 T 之间的差异，采用均方误差（MSE）计算如下，

$$L_{TissueMaskRec}(\theta_A,\theta_I,\theta_M) = MSE(\hat{T},T) \tag{3}$$

ˆ其中， T 为真实值， T 为生成的组织成分掩膜。

ˆ图像重建损失： 此分量通过使用 $L1$ 差异，衡量真实值 Z 与生成的组织图像 Z 之间的重建误差，

$$L_{ImageRec}(\theta_A,\theta_I,\theta_M,\theta_E) = \|\hat{Z}-Z\|_1 \tag{4}$$

ˆ其中， Z 为真实值， Z 为生成的组织图像。

对抗损失分量：如第 3.3 节所述，我们针对组织成分掩膜、组织图像以及从组织图像中裁剪出的腺体部分，分别采用了 3 个对抗损失分量：L^M_{GAN}、 L^Z_{GAN} 和 L^G_{GAN}。将 $t=M$、 $t=Z$ 和 $t=G$ 代入式（1）即可得到它们的表达式。

因此，基于对抗损失和重建损失的线性组合，整体学习问题可表述为以下对抗优化问题，

$$\begin{aligned}\min_{\theta_A,\theta_I,\theta_M,\theta_E}\ \max_{\theta_{M_D},\theta_{Z_D},\theta_{G_D}}\ & \lambda_1 L_{ImageRec}(\theta_A,\theta_I,\theta_M,\theta_E) + \lambda_2 L_{TissueMaskRec}(\theta_A,\theta_I,\theta_M)\\ &+ \lambda_3 L_{GlandMaskRec}(\theta_A,\theta_I) + \lambda_4 L^M_{GAN}(\theta_G=\{\theta_A,\theta_I,\theta_M\},\theta_{t_D}=\theta_{M_D})\\ &+ \lambda_5 L^Z_{GAN}(\theta_G=\{\theta_A,\theta_I,\theta_M,\theta_E\},\theta_{t_D}=\theta_{Z_D}) + \lambda_6 L^G_{GAN}(\theta_G=\{\theta_A,\theta_I,\theta_M,\theta_E\},\theta_{t_D}=\theta_{G_D})\end{aligned} \tag{5}$$

其中， λ_1， $\lambda_2..\lambda_6$ 表示相应损失分量的权重。

3.5 使用潜在扩散模型合成腺体掩膜

本节将讨论用于合成腺体掩膜的潜在扩散模型[11] 基础框架。该过程分为两个步骤：(i) 在腺体掩膜上训练 VQ-VAE[17]，以学习离散潜在表示；(ii) 使用以癌症类型（良性或恶性）为条件的扩散模型[18] 对新的潜在表示进行采样。下面将详细介绍这两个步骤。

3.5.1 使用 VQ-VAE 生成离散潜在表示

VQ-VAE[17] 模型首先将输入的腺体掩膜传入编码器神经网络，生成该掩膜的编码潜在表示。随后，编码后的潜在表示经过量化层。该层使用预定义的向量码本，将潜在表示的连续值映射为离散值。码本中的每个向量代表潜在表示的一种可能值。量化层根据每个实体与码本中各向量之间的欧几里得距离，为每个实体分配一个码本中的向量。经过量化层后获得的向量称为量化嵌入。

最后，将量化嵌入输入解码器网络，由其重建腺体掩膜。为了采样数据集之外的掩膜，我们获取有关潜变量的知识。通过学习这些潜变量上的先验，我们可以显著减少所需的内存和计算资源。这样，只需调整潜在空间的大小，即可生成高分辨率掩膜。

3.5.2 使用扩散模型采样腺体掩膜

为了根据潜变量或压缩掩膜的学习分布生成掩膜，我们为其建立先验分布。为此，我们采用扩散模型。[18] 扩散模型分两个阶段运行。第一阶段称为前向扩散，利用扩散过程将真实图像转换为含噪图像。这些含噪图像通常服从高斯分布。第二阶段是反向扩散过程，其中通过迭代地对从高斯分布中采样得到的初始噪声应用去噪函数或深度神经网络，以去除噪声并更新图像。重复这一过程多次，即可从目标分布中生成样本。扩散模型使用经过学习的去噪函数，通常由 U-Net 实现，[19] 以便在反向扩散过程的每一步对图像进行去噪。

我们将扩散模型应用于使用 VQ-VAE[17] 模型学习得到的潜在向量，并以癌症类型（良性或恶性）为条件。我们通过交叉注意力[11] 机制实现这种条件控制。随后，将采样得到的潜在向量输入在上一步训练的 VQ-VAE 解码器模型，以生成逼真的腺体掩膜。在本实验中，我们使用的不是二值腺体掩膜，而是突出显示腺体（绿色）、基质区域（红色）和背景（蓝色）的组织成分掩膜，如图5所示。这样做的目的是对组织图像中尽可能多的成分之间的依赖关系进行建模，从而生成逼真的腺体掩膜。之后，可以从生成的组织成分掩膜中提取腺体，并将其他部分设为背景，从而得到二值腺体掩膜。为生成合成组织图像，我们采用残差编码器-解码器网络。[20]

4. 实验结果

4.1 数据采集

我们使用 DigestPath[21] 结肠镜组织分割数据集评估算法性能。该数据集收集自 DigestPath2019 挑战赛*。其中包含 660 张超大组织图像，平均尺寸为 5000×5000 像素。每张图像均带有腺体区域的像素级标注。该数据集最初仅包含恶性病变的标注（250 张图像）。为了获得良性腺体的组织分割掩膜，我们采用了半自动方法。为此，我们首先在 GlaS 数据集[23,24] 上训练了名为 Mild-Net[22] 的腺体分割模型，并获得了 DigestPath 数据集中正常等级图像的腺体分割掩膜，随后对这些掩膜进行了人工修正。我们从这些图像中提取了 1733 个尺寸为 512×512 的图像块，之后将其调整为 256×256。其中，我们保留了约 1300 个图像块用于训练（训练集），其余用于测试（测试集）。

*https://digestpath2019.grand-challenge.org/

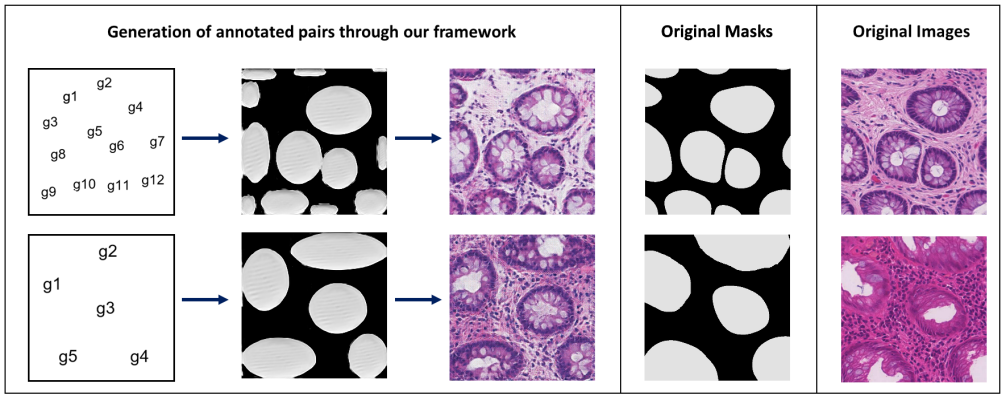


图 2. 根据输入腺体布局 (a) 生成的结直肠组织图像及其腺体分割掩膜的视觉结果。(b) 显示原始腺体分割掩膜，(c) 显示真实组织图像。

所获得的掩膜可以是二值的（区分腺体区域与其余部分），用于所提出的框架；也可以是三值的（腺体区域、基质区域和背景），用于生成第 3.5 节所述的合成掩膜。三值掩膜中的基质区域通过对苏木精-伊红组织图像进行像素阈值处理获得。附录 5.6 中介绍了如何从 DigestPath 数据集中收集的腺体掩膜中获取边界框，以及如何根据输入的位置 $\vec{l}$ 和大小 $\vec{s}$ 参数（腺体布局）构建边界框。

4.2 模型训练

为了训练整个框架，我们将目标组织图像尺寸设为 $N = 256$，腺体特定嵌入的输入噪声维度设为 $dim(z) = 6$，潜向量大小设为 $D = 32$，每个腺体生成的尺寸设为 $B = 64$。对于损失函数（见公式 5），经过交叉验证调优后，我们将 $\lambda_1 = \lambda_2 = \lambda_3 = 100.0$ 和 $\lambda_4 = \lambda_5 = \lambda_6 = 1.0$ 设为相应值。

我们使用 Adam 优化器[25] 、学习率 10^{-4} 和批量大小 1 训练所有模型，约进行 300K 次迭代。每次迭代中，我们先更新生成器权重 $\{\theta_A, \theta_I, \theta_M, \theta_E\}$，然后更新判别器$\{D_T, D_Z, D_G\}$。该框架使用 PyTorch 在 Nvidia Titan X 上实现，训练耗时近 2 天。

4.3 视觉结果

生成组织图像（来自测试集）的视觉结果如图2所示。我们可以观察到，腺体形状得到了保留，杯状细胞和基质区域等组织成分也构建得较为逼真，但腺腔存在一定程度的变形。生成的组织成分掩膜看起来也与实际掩膜相近。这些掩膜形状的轻微变形和变化，是在腺体的表征嵌入中使用高斯噪声所致，也使其显得更加真实。我们还研究了改变大小 $\vec{s}$和位置 $\vec{l}$后腺体外观的变化。图3展示了改变位置和大小后的组织图像结果。改变大小和位置后，边界框也会相应变化，从而有效改变腺体的大小和方向。

4.4 定量分析

我们使用 Fréchet Inception 距离（FID）[26] 评估生成图像（来自测试集）的质量。FID 是一种用于评估生成模型所生成图像质量的标准指标。它计算真实图像与生成图像的卷积特征图之间的距离。在我们的实验中，为了提取卷积特征，我们使用了在 ImageNet 数据集上训练的预训练 InceptionV3[27] 网络。[28] FID 分数越低，图像质量越好。由于该指标取决于图像大小，为了了解其数值尺度，我们还计算了真实值图像与同等大小随机噪声之间的 FID 分数。作为基线，我们

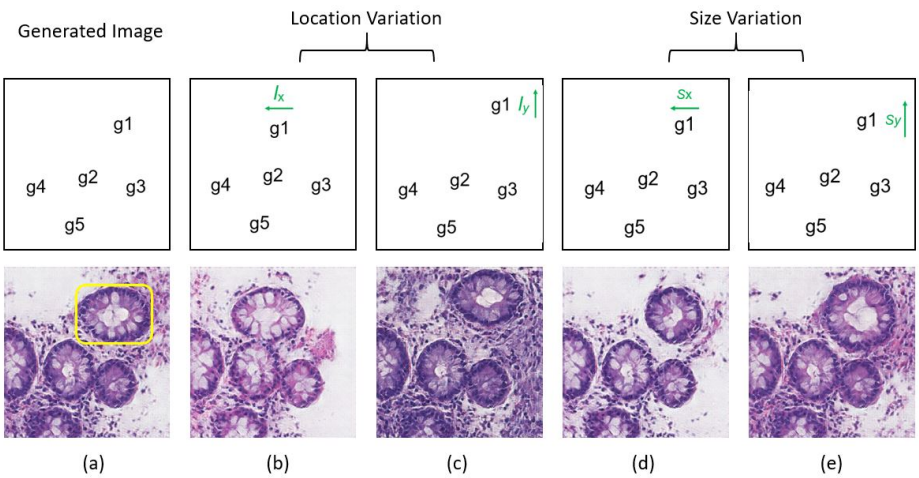


图3。最左侧的图像 (a) 展示了由所提出的框架生成的生成样本。其右侧的图像展示了改变黄色边框腺体的位置 $\vec{l} = (l_x, l_y)$ 和大小 $\vec{s} = (s_x, s_y)$ 后的外观变化。(b) 和 (c) 分别展示了该腺体向左侧移动（降低 l_x）和向上移动（增加 l_y）的情况。对于同一腺体，在分别减小 (s_x) 和增大 (s_y) 后，(d) 展示了其水平方向的收缩，(e) 展示了其垂直方向的扩展。

对图像到图像 Pix2Pix 网络[15] 进行调整，使其能够根据现有的组织成分掩膜生成组织图像，并计算其生成的组织图像的 FID。比较结果见表1。表1显示，与随机噪声相比，所提出的模型取得了更好的结果，但略逊于 Pix2Pix。原因可能是，我们的目标是在生成最终组织图像的同时构建腺体分割掩膜，而 Pix2Pix 假设真实值掩膜已存在，并据此构建组织图像。因此，生成掩膜时的构建误差可能会影响我们框架的性能，并略微降低生成图像的质量。不过，从 FID值的数值尺度来看，这两种框架之间的差异并不显著。

模型	FID
随机	485 ± 5.7
Pix2Pix	120 ± 2.7
所提出的框架	134 ± 2.4

表1. Fréchet Inception 距离（FID）分数比较

4.5 通过腺体分割进行评估

我们还使用基于 U-Net 的腺体分割算法评估所提出的框架生成的标注对的质量。[19] 我们使用训练集中的图像块训练 U-Net，图像块尺寸为 256×256，并计算测试集中真实图像和合成图像的分割掩膜。我们使用 Dice 分数[29]，分别计算真实成分掩膜与 U-Net 在真实图像上生成的掩膜之间，以及生成的成分掩膜与 U-Net 在合成图像上生成的掩膜之间的 Dice 分数。示例结果见图4。

在上述两种情况下，我们得到的平均 Dice 指数分别为 0.9022（标准差为 0.006）和 0.9001（标准差为 0.012）。所得分数高度相似，这验证了生成的组织图像及其组织成分掩膜均适用于腺体分割算法的评估。

4.6 使用潜在扩散模型生成的掩膜进行图像合成

通过基于潜在扩散模型的框架（第3.5节所述）生成的组织成分掩膜，以及使用残差编码器-解码器网络创建的相应组织图像，[20] 显示于图5中。真实图像与通过该方法生成的图像之间的FID为

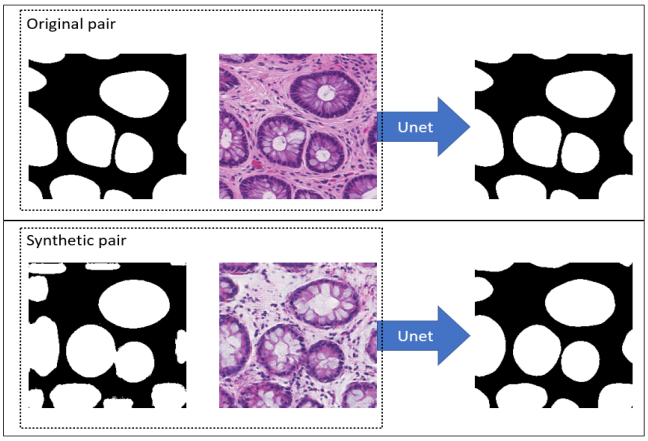


图4. 组织图像及其对应腺体分割掩膜的真实（上方）和构造（下方）标注对样例。右侧所示掩膜由基于 U-Net 的分割算法分别应用于原始图像（上方）和合成图像（下方）生成。

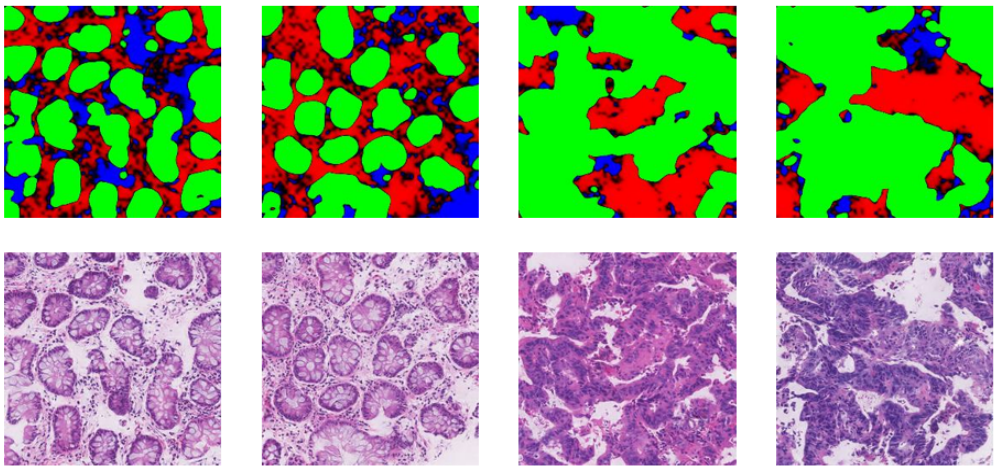


图5. 使用潜在扩散模型合成的标注对。从右侧数，前两对展示的是良性级别的结肠组织图像标注对，最后两对展示的是恶性类型的图像。

130.5（±2.8）。将这些FID值与此前获得的结果进行比较，可以明显看出，合成图像与真实图像非常相似。

5. 结论与未来方向

在本研究中，我们提出了一种交互式框架，用于生成结直肠组织图像及其组织成分掩膜的标注对。我们在 DigestPath 数据集上进行了实验，结果表明，该框架能够生成逼真的图像，并保留基质、杯状细胞和腺腔等形态学特征。与最先进的图像到图像转换模型相比，生成图像的 FID 分数表现良好。我们展示了通过改变腺体的大小和位置，可以呈现腺体外观的多样性。此外，我们还证明了合成标注对可用于评估腺体分割算法。

所提出的框架的一个明显局限是，构建图像需要腺体布局，这仍需投入一些工作。不过，我们还提出了另一种构建腺体掩膜的方法，从而生成组织图像。

未来可以进一步拓展这一思路，为计算病理组织学中的各种任务生成带注释的组织图像对，例如细胞核分割、癌症分级等。生成的图像对有望取代

真实世界的训练数据，在使用过程中绕过法律和安全障碍，协助数字病理算法的训练与验证，并降低数据获取的成本和工作量。

参考文献

[1] Quiros, A. C., Murray-Smith, R. 和 Yuan, K.，《病理学生成对抗网络：学习癌组织的深度表征》，arXiv 预印本 arXiv:1907.02644（2019）。[2] Mahmood, F., Borders, D., Chen, R. J., Mckay, G. N., Salimian, K. J., Baras, A. 和 Durr, N. J.，《针对组织病理图像多器官细胞核分割的深度对抗训练》，IEEE医学影像汇刊 39(11)，3257–3267（2020）。[3] Deshpande, S., Minhas, F., Graham, S. 和 Rajpoot, N.，《Safron：用于生成结直肠癌组织学图像的跨前沿网络拼接》，医学图像分析 77，102337（2022）。[4] Deshpande, S., Dawood, M., Minhas, F. 和 Rajpoot, N.，《Synclay：从定制细胞布局交互式合成组织学图像》，医学图像分析 91，102995（2024年1月）。[5] Hou, L., Agarwal, A., Samaras, D., Kurc, T., Gupta, R. 和 Saltz, J.，《稳健的组织病理图像分析：标注还是合成？》，8525–8534（2019年6月）。[6] Levine, A. B., Peng, J., Farnell, D., Nursey, M., Wang, Y., Naso, J., Ren, H., Farahani, H., Chen, C., Chiu, D., Talhouk, A., Sheffield, B., Riazy, M., Ip, P., Parra-Herran, C., Mills, A., Singh, N., Tessier-Cloutier, B., Salisbury, T. D., Lee, J., Salcudean, T., Jones, S. J. M., Huntsman, D., Gilks, C., Yip, S. 和 Bashashati, A.，《利用生成对抗网络合成诊断级癌症病理图像》，《病理学杂志》252（2020）。[7] McGaghie, W. C.，《作为转化科学的医学教育研究》，《科学·转化医学》2(19)，19cm8–19cm8（2010）。[8] He, J., Baxter, S., Xu, J., Xu, J., Zhou, X. 和 Zhang, K.，《人工智能技术在医学中的实际应用》，《自然·医学》25（2019年1月）。[9] Senaras, C., Niazi, M. K. K., Sahiner, B., Pennell, M. P., Tozbikian, G., Lozanski, G. 和 Gurcan, M. N.，《使用条件生成对抗网络优化生成高分辨率仿体图像：在Ki67乳腺癌图像定量分析中的应用》，《PLOS ONE》13(5)，e0196846（2018）。[10] Deshpande, S., Minhas, F. 和 Rajpoot, N.，《小规模训练，大规模生成：结直肠癌组织学图像合成》，载于 [《医学影像中的模拟与合成》],Burgos, N., Svoboda, D., Wolterink, J. M. 和 Zhao, C. 编，164–173，施普林格国际出版公司，尚（2020）。[11] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. 和 Ommer, B.，《利用潜在扩散模型进行高分辨率图像合成》，载于 [IEEE计算机视觉与模式识别会议（CVPR）论文集],（2022）。[12] Esser, P., Rombach, R. 和 Ommer, B.，《驯服变换器以合成高分辨率图像》，载于[IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集], 12873–12883（2021年6月）。[13] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M. 和 Sutskever, I.，《零样本文本到图像生成》，载于 [第38届国际机器学习会议论文集], Meila, M. 和 Zhang, T. 编，《机器学习研究论文集》139，8821–8831，PMLR（2021年7月18–24日）。[14] Jaderberg, M., Simonyan, K., Zisserman, A. 和 Kavukcuoglu, K.，《空间变换网络》，载于[NIPS]，（2015）。[15] Isola, P., Zhu, J.-Y., Zhou, T. 和 Efros, A. A.，《利用条件对抗网络进行图像到图像转换》，载于 [*2017* IEEE计算机视觉与模式识别会议（CVPR）论文集], 5967–5976（2017）。[16] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. 和 Bengio, Y.，《生成对抗网络》，载于 [神经信息处理系统进展], 2672–2680（2014）。[17] van den Oord, A., Vinyals, O. 和 Kavukcuoglu, K.，《神经离散表征学习》，载于 [第31届神经信息处理系统国际会议论文集],NIPS'17，6309–6318，Curran Associates Inc.，雷德胡克，纽约州，美国（2017）。[18] Ho, J., Jain, A. 和 Abbeel, P.，《去噪扩散概率模型》，载于 [神经信息处理系统进展],Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. 和 Lin, H. 编，33，6840–6851，Curran Associates, Inc.（2020）。

[19] Ronneberger, O., Fischer, P. 和 Brox, T.，“U-Net：用于生物医学图像分割的卷积网络”，载于 [医学图像计算与计算机辅助干预国际会议], 234，241，Springer（2015）。[20] Ashual, O. 和 Wolf, L.，“在交互式场景生成中指定对象属性和关系”，*2019 IEEE/CVF*计算机视觉国际会议（*ICCV*），4560–4568（2019）。[21] Li, J., Yang, S., Huang, X., Da, Q., Yang, X., Hu, Z., Duan, Q., Wang, C. 和 Li, H.， [基于半监督学习框架的印戒细胞检测], 842，854（2019年5月）。[22] Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.-A., Snead, D., Tsang, Y. W. 和 Rajpoot, N.，“Mild-Net：用于结肠组织学图像腺体实例分割的最小信息损失扩张网络”，医学图像分析**52**，199–211（2019）。[23] Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P.-A., Guo, Y. B., Wang, L. Y., Matuszewski, B. J., Bruni, E., Sanchez, U., Böhm, A., Ronneberger, O., Cheikh, B. B., Racoceanu, D., Kainz, P., Pfeiffer, M., Urschler, M., Snead, D. R. 和 Rajpoot, N. M.，“结肠组织学图像中的腺体分割：GLAS挑战赛”，医学图像分析**35**，489–502（2017年1月）。[24] Sirinukunwattana, K., Snead, D. R. J. 和 Rajpoot, N. M.，“用于结肠组织学图像中腺体结构的随机多边形模型”，*IEEE*医学影像汇刊 **34**(11)，2366–2378（2015）。[25] Kingma, D. 和 Ba, J.，“Adam：一种随机优化方法”，学习表征国际会议（2014年12月）。[26] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. 和 Hochreiter, S.，“采用双时间尺度更新规则训练的生成对抗网络收敛至局部纳什均衡”，载于 [神经信息处理系统进展],6626–6637（2017）。[27] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. 和 Wojna, Z.，“重新思考用于计算机视觉的 Inception架构”，载于 [*IEEE*计算机视觉与模式识别会议论文集],2818–2826（2016）。[28] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. 和 Fei-Fei, L.，“ImageNet：一个大规模分层图像数据库”，载于 [*2009 IEEE*计算机视觉与模式识别会议], 248，255，IEEE（2009）。[29] Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells III, W. M., Jolesz, F. A. 和 Kikinis, R.，“基于空间重叠指数的图像分割质量统计验证1：科学报告”，《学术放射学》**11**(2)，178–189（2004）。[30] Bradski, G.，“OpenCV库”，《*Dr. Dobb*软件工具杂志》（2000）。

附录

此处描述　　所提出的框架中各个组件的神经网络架构

5.1 独立掩码生成网络

我们使用独立掩码生成网络为每个腺体生成独立掩码。输入是对原始腺体嵌入进行仿射变换后得到的各个腺体潜在向量，输出是 64×64 像素的腺体掩膜，所有元素的取值范围均为 0 到 1。掩码回归网络由一系列独立掩码生成器模块组成，每个模块包含插值 $+$ 卷积 $+$ 批归一化 $+$ ReLU激活操作。具体架构如表 2 所示，而独立掩码生成器模块 的架构如表 3 所示。

索引	输入	操作	输出形状
(1)	-	腺体潜向量	32
(2)	(1)	重塑	32 x 1 x 1
(3)	(2)	独立掩码生成器模块	32 x 2 x 2
(4)	(3)	独立掩码生成器模块	32 × 4 × 4
(5)	(4)	独立掩码生成器模块	32 x 8 x 8
(6)	(5)	独立掩码生成器模块	32 x 16 x 16
(7)	(6)	独立掩码生成器模块	32 x 32 x 32
(8)	(7)	独立掩码生成器模块	32 x 64 x 64
(9)	(8)	二维卷积 (K=1, $32 \rightarrow 1$)	1 × 64 × 64
(10)	(9)	Sigmoid 函数	1 x 64 x 64

表 2. 独立掩膜生成网络的架构。该函数实现了正文中的函数 I 。记号 Conv2d(K，$C_{in} \rightarrow C_{out}$) 表示一种卷积，具有 $K \times K$ 个卷积核、 C_{in} 个输入通道和 C_{out} 个输出通道；所有卷积的步幅均为 1，并采用零填充，以确保输入和输出的空间尺寸相同。

操作	输出形状
插值	32 x 2S x 2S
二维卷积 (K=3, $32 \rightarrow 32$)	32 × 2S × 2S
批归一化	32 x 2S x 2S
ReLU	32 x 2S x 2S

表 3. 独立掩码生成器模块的架构　输入为形状为 $C \times S \times S$ 的特征图，其中 C 为最后一层特征图的通道数，而 $S \times S$ 为高度和宽度的维度。

5.2 掩码生成网络

生成的中间张量（在 3.1 节中解释）有 32 个通道，经过掩码生成网络后减少至 1 个，从而形成腺体掩膜或组织成分掩膜。该网络由一系列卷积 $+$ReLU 操作组成。具体架构如表4所示。

5.3 编码器-解码器网络

最终组织图像借助编码器-解码器网络，由生成的腺体掩膜生成。编码器由一系列“编码”模块（见表6）组成，用于生成输入掩膜的低尺寸编码；解码器则由一系列“解码”模块（见表7）组成，根据该编码生成最终组织图像。编码器-解码器网络的具体架构如表5所示。

5.4 掩膜和图像判别器

我们采用的判别器对腺体掩膜（D_M）和组织图像（D_Z）使用相似的架构，以形状为 $C \times 256 \times 256$ 的真实或伪造图像作为输入（组件掩膜为 $C = 1$，组织图像为 $C = 3$），并将输入图像中大小为 7×7 的重叠网格图像块分类为真实或伪造。判别器的具体架构如表8所示

索引	输入	操作	输出形状
(1)	-	生成累积掩码	32 x 256 x 256
(2)	(1)	二维卷积 (K=3, $32 \rightarrow 16$)	16 x 256 x 256
(3)	(2)	带泄漏的修正线性单元	16 x 256 x 256
(4)	(3)	二维卷积（K=3, $16 \rightarrow 8$）	8 x 256 x 256
(5)	(4)	带泄漏的修正线性单元	8 × 256 × 256
(6)	(5)	二维卷积（K=3, $8 \rightarrow 4$）	4 × 256 × 256
(7)	(6)	带泄漏的修正线性单元	4 x 256 x 256
(8)	(7)	二维卷积 (K=3, $8 \rightarrow 4$)	1 x 256 x 256
(9)	(8)	带泄漏的修正线性单元	1 x 256 x 256

表 4. 掩码生成网络的架构。该网络实现了正文中的函数 M 。带泄漏的修正线性单元使用的负斜率系数为 0.2

索引	输入	操作	输出形状
(1)	-	生成组件掩码	1 x 256 x 256
(2)	(1)	编码(1,64)	64 x 128 x 128
(3)	(2)	编码(64,128)	128 x 64 x 64
(4)	(3)	编码(128,256)	256 x 32 x 32
(5)	(4)	编码(256,512)	512 x 16 x 16
(6)	(5)	编码(512,512)	512 x 8 x 8
(7)	(6)	编码(512,512)	512 x 4 x 4
(8)	(7)	编码(512,512)	512 x 2 x 2
(9)	(8)	编码(512,512)	512 x 1 x 1
(10)	(9,8)	解码(512,512)	1024 x 2 x 2
(11)	(10,7)	解码(1024,512)	1024 x 4 x 4
(12)	(11,6)	解码(1024,512)	1024 x 8 x 8
(13)	(12,5)	解码(1024,512)	1024 x 16 x 16
(14)	(12,4)	解码(1024,256)	512 x 32 x 32
(15)	(14,3)	解码(512,128)	256 x 64 x 64
(16)	(15,2)	解码(256,64)	128 x 128 x 128
(17)	(16)	上采样	128 x 256 x 256
(18)	(17)	二维卷积 (K=4, $128 \rightarrow 3$)	3 x 256 x 256
(19)	(18)	Tanh	3 x 256 x 256

表5. 编码器-解码器网络的架构。该网络实现了正文中的函数 E 。

操作	输出形状
二维卷积 (K=4, $C_{in} \rightarrow C_{out}$)	C_{out} x S x S
实例归一化 (if normalize=True)	C_{out} x S x S
带泄漏的修正线性单元	C_{out} x S x S
Dropout（如果 dropout=True）	C_{out} x S x S

表6。“编码”模块的架构。带泄漏的修正线性单元使用的负斜率系数为 0.2。输入为大小为$C_{in} \times S \times S$的图像

操作	输出形状
二维转置卷积(K=4, $C_{in} \rightarrow C_{out}$)	C_{out} x S x S
实例归一化	C_{out} x S x S
ReLU	C_{out} x S x S
Dropout（如果 dropout=True）	C_{out} x S x S

表7。“解码”模块的架构。输入为大小为 $C_{in} \times S \times S$ 的图像

5.5 腺体判别器

我们的腺体判别器 D_G 会读取真实或生成的组织图像中对应腺体区域的图像像素，并将其分类为真实或生成。腺体区域通过其边界框裁剪

索引	输入	操作	输出形状
(1)	-	生成图像	C x 256 x 256
(2)	(1)	二维卷积 (K=4, C $\rightarrow$ 16, S=2)	16 x 128 x 128
(3)	(2)	带泄漏的修正线性单元	16 x 128 x 128
(4)	(3)	二维卷积 (K=4, 16 $\rightarrow$ 32, S=2)	32 x 64 x 64
(5)	(4)	带泄漏的修正线性单元	32 x 64 x 64
(6)	(5)	实例归一化	32 x 64 x 64
(7)	(6)	二维卷积 (K=4, 32 $\rightarrow$ 64, S=2)	64 x 32 x 32
(8)	(7)	带泄漏的修正线性单元	64 x 32 x 32
(9)	(8)	实例归一化	64 x 32 x 32
(10)	(9)	二维卷积 (K=4, 64 $\rightarrow$ 128, S=2)	128 x 16 x 16
(11)	(10)	带泄漏的修正线性单元	128 x 16 x 16
(12)	(11)	实例归一化	128 x 16 x 16
(13)	(12)	二维卷积（K=4, 128 $\rightarrow$ 256, S=2）	256 x 8 x 8
(14)	(13)	带泄漏的修正线性单元	256 x 8 x 8
(15)	(14)	实例归一化	256 x 8 x 8
(16)	(15)	二维卷积 (K=4, 256 $\rightarrow$ 1 , S=1)	1 x 7 x 7

表8。判别器网络的架构。当输入为组织成分掩膜时，C=1；当输入为组织图像时，C=3 。除最后一个二维卷积操作外，所有操作的步幅均为 2。带泄漏的修正线性单元使用 0.2 的负斜率系数。

索引	输入	操作	输出形状
(1)	-	从生成的图像中裁剪腺体部分	3 × 64 × 64
(2)	(1)	二维卷积 (K=5, 3 $\rightarrow$ 16, S=2)	16 x 30 x 30
(3)	(2)	批归一化	16 x 30 x 30
(4)	(3)	带泄漏的修正线性单元	16 x 30 x 30
(5)	(4)	二维卷积 (K=5, 16 $\rightarrow$ 32, S=2)	32 x 13 x 13
(6)	(5)	批归一化	32 x 13 x 13
(7)	(6)	带泄漏的修正线性单元	32 x 13 x 13
(8)	(7)	二维卷积（K=5, 32 $\rightarrow$ 64, S=2）	64 x 5 x 5
(9)	(8)	全局平均池化	64
(10)	(9)	仿射变换	1024
(11)	(10)	仿射变换	1

表9。腺体判别器的架构， D_G。带泄漏的修正线性单元使用 0.2 的负斜率系数

坐标裁剪出来，并使用双线性插值法调整至 64×64 像素。腺体判别器的具体架构如表 9 所示

5.6 边界框获取

按照第 4.1 节所述，我们从 DigestPath 数据集中收集组织图像及其标注的腺体掩膜后，使用 OpenCV（开源计算机视觉库）Python 库[30] 提取腺体的位置，即黑白组织成分掩膜中白色斑块对象的质心（如图1中的掩膜所示）。随后，我们还使用该库的内置函数 boundingRect() 获取这些已识别腺体对象的边界框。

在推理过程中，除了使用通过上述流程获得的真实标注边界框外，我们还根据腺体布局中的输入尺寸 $\overrightarrow{s_k}$ 和位置 $\overrightarrow{l_k}$ 属性，为腺体 g_k 构建边界框。给定输入尺寸 $\overrightarrow{l_k} = (s_{kx},\ s_{ky})$，其中 s_x 和 s_y 分别是腺体的水平和垂直跨度，以及质心位置 $\overrightarrow{l_k} = (l_{kx},\ l_{ky})$， g_k 的边界框坐标计算如下：

$$b_k = (l_{kx} - (s_{kx}/2), (l_{ky} - (s_{ky}/2), (l_{kx} + (s_{kx}/2), (l_{ky} + (s_{ky}/2)). \tag{6}$$

总体而言，所提出的框架的输入是腺体的位置、尺寸，以及从数据集中获取或通过上述流程构建的边界框；输出则是组织图像及其组织成分掩膜这一对结果。