

# 基于 Medical Diffusion 的三维膀胱壁肿瘤扩充生成与半监督学习

## 摘要:

本文档总计从六个部分展示本项目的具体细节，分别是数据集概述与处理，数据模型的扩充，生成数据的评价，半监督训练，结果分析和图像展示

## 一 数据集概述与处理

本研究的数据集由 24 张  $512 \times 512 \times 24$  的 nii 格式 CT 扫描数据构成，每个数据包含 24 个连续的扫描切片，空间分辨率为  $512 \times 512$  像素。每张影像的像素标签分为 0、128 和 255 三个值，其中 0 代表背景，128 代表肿瘤，255 代表膀胱壁。

在处理这些数据之前，我们首先进行了数据清理和预处理。对于图像数量不足 24 张的扫描数据，我们选择进行填充处理，确保每个数据集都包含 24 张扫描影像。填充策略是使用该数据集中的最后一张影像进行复制，以维持原有影像的连续性和一致性。对于超过 24 张的扫描数据，我们进行了切割去除，确保所有的数据集都保持在 24 张影像的范围内。切割策略是从中间部分开始，等距离地去除超出部分，以尽可能保留原有图像中的有效信息。

对于图像的像素标签，我们没有进行额外的处理，因为这三个标签已足够清晰地表示背景、肿瘤和膀胱壁，并为后续的二半监督学习提供了必要的标记信息。此外，这些标签也将用于评估我们的模型在肿瘤识别和膀胱壁识别上的性能。

经过上述处理，我们得到了一份标准化的医学影像数据集，满足我们模型输入的需求。我们将利用这份数据集进行模型的训练和验证，探索 Medical Diffusion 模型和 DeepAtlas 半监督学习方法在膀胱壁肿瘤影像分析上的应用效果。

## 二 数据模型的扩充

### 2.1 VQ-GAN 编码器

为了将影像数据编码到低维潜在空间，我们首先采用向量量化自编码器（Vector Quantized Autoencoders, VQ-GAN）。这类自编码器已被证明是一种有效的选择，因为它们缓解了变分自编码器在输出模糊方面的问题。其工作原理是将自编码器的潜在特征向量映射到一个量化表示，该表示从学习到的代码本中获取。我们采用了 Esser 等人提出的 VQ-GAN 体系结构，它可以看作是一类向量量化自编码器，通过在输出处施加鉴别器损失来进一步提高图像重建质量。

具体而言，图像被输入编码器以构建潜在代码  $ze \in \mathbb{R}^{\frac{H}{s} \times \frac{W}{s} \times k}$ ，其中 H 表示高度，W 表示宽度，C 表示通道数，k 表示潜在特征图的数量，s 表示压缩因子。在向量量化步骤中，潜在特征向量通过用最接近的代码本向量 en 替换每个向量

来进行量化，这些代码本向量包含在学习到的代码本  $Z$  中。然后将量化特征向量输入解码器  $G$  中来重建图像。学习目标定义为最小化重建损失  $L_{REC}$ ，代码本损失  $L_{CodeBook}$  和承诺损失  $L_{Commit}$ 。我们使用感知损失作为重建损失，使用直通估计器来克服不可微的量化步骤。承诺损失定义为未量化的潜在特征向量和对应代码本向量之间的均方误差。

需要注意的是，梯度只计算连续的潜在特征向量，以迫使它们更接近量化的代码本向量。可学习的代码本向量通过在所有映射到它的潜在向量上保持指数移动平均进行优化。此外，在输出处使用基于片的鉴别器以获得更好的重建质量。为了支持 3D 输入，我们采用 Ge 等人的方法，用 3D 卷积替换 2D 卷积。此外，我们还在原始 VQ-GAN 模型中替换了鉴别器，采用一个基于切片的鉴别器（以随机切片作为输入）和一个 3D 鉴别器（以整个重建体积作为输入）。按照他们的方法，我们还添加了特征匹配损失，以稳定 GAN 的训练。

## 2.2 扩散模型

扩散模型是一类由潜在变量  $\mathbf{x}_1 \dots \mathbf{x}_t$  的马尔可夫链定义的生成模型。其主要思想是从图像  $x_0$  开始，通过连续添加高斯噪声（随着  $T$  步骤的增加，噪声的方差增大）扰动图像。神经网络根据在时间步  $t$  的图像的噪声版本和时间步本身进行条件化，并被训练来学习扰动图像的噪声分布，以便推断出时间  $t-1$  的数据分布  $p(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ 。当  $T$  足够大时，可以从先验分布  $N(0, I)$  中近似抽样  $\mathbf{x}_T$ ，然后

反向遍历马尔可夫链，从学习到的分布  $p_\theta(\mathbf{x}_0) := \int p_\theta(\mathbf{x}_0 : T) d\mathbf{x}_{1:T}$  中抽样新的图像。

用于模拟噪声的神经网络通常选择为 U-Net。为了支持 3D 数据，我们用 3D 卷积替换 U-Net 中的 2D 卷积。此外，按照 Ho 等人的方法，我们只在高分辨率图像平面上使用卷积（即核大小为  $3 \times 3 \times 1$ ），然后在这个高分辨率平面上跟随一个空间注意力块（将深度维度当作批处理大小的扩展），以提高计算效率。空间注意力块之后是深度注意力块，其中高分辨率图像平面轴被当作批处理轴。

## 2.3 整体模型

首先，在整个数据集上训练 VQ-GAN 模型，以学习数据的有意义的低维潜在表示。由于输入到扩散模型的图像应标准化到 -1 和 1 的范围内，我们必须确保图像的潜在表示也在这个范围内。假设在 VQ-GAN 模型的向量量化步骤中，学习到的代码本向量被强制靠近量化前的潜在特征向量，我们可以通过学习到的代码本中的最大值近似未量化特征表示的最大值。同样地，我们可以通过学习到的代码本中的最小值近似未量化特征表示的最小值。因此，通过对未量化特征向量进行简单的最小-最大标准化，我们得到了接近 -1 和 1 范围的潜在表示。这些潜在表示可以用于训练 3D 扩散模型。

然后，我们可以通过反向运行扩散过程，从一个从标准高斯分布中抽样的噪声开始，生成新的图像。该过程的输出经过 VQ-GAN 的学习代码本进行量化，然后输入解码器生成相应的图像。

### 3 生成数据的评价

我们使用 3 中的模型生成了 200 份膀胱壁肿瘤的三维扫描，并对其进行了综合评价。评价主要集中在以下三个方面：真实图像的外观、切片之间的一致性和解剖学的准确性。

#### 3.1 真实图像的外观

首先，我们评估了生成的 CT 扫描图像在视觉上的真实感。视觉质量是评估图像生成任务的重要标准。在这方面，我们的模型展现出了很高的能力。生成的膀胱壁肿瘤图像在视觉上几乎无法与真实的 CT 扫描图像区分开。肿瘤和膀胱壁的细节、形状和边缘被真实地再现，达到了几乎无法被肉眼区分的程度。这对于医学影像的训练和分析非常重要，因为细节的准确性直接影响后续的图像识别和诊断准确性。

#### 3.2 切片之间的一致性

其次，我们评估了生成的三维 CT 扫描在连续切片之间的一致性。在真实的 CT 扫描中，连续切片应该呈现出清晰、连贯的组织结构和病变区域的变化。在这方面，我们的模型也展现了良好的性能。生成的三维 CT 扫描在连续切片中展示出了自然和连贯的变化，不仅在大的组织结构上，如膀胱壁的整体形状和位置，还在小的细节上，如肿瘤的大小和形状。这种连贯性对于三维医学图像的识别和分析至关重要。

#### 3.3 解剖学的准确性

最后，我们评估了生成的 CT 扫描在解剖学上的准确性。在医学影像中，解剖学准确性是一个非常重要的评估指标。在这方面，我们的模型也表现出了出色的性能。生成的膀胱壁肿瘤图像精确地再现了膀胱壁和肿瘤的解剖结构，包括肿瘤的形状、位置和大小，以及膀胱壁的厚度和形状。这些都是在真实的 CT 扫描中可以观察到的重要解剖学特征。这种准确性对于训练医学影像识别和分析模型非常关键，因为它可以提供更准确和丰富的解剖学信息，从而提高模型的识别和分析能力。

我们的模型在生成的三维膀胱壁肿瘤 CT 扫描评估中展现了高质量和准确性。无论是在视觉真实感、切片一致性还是解剖学准确性方面，我们的模型都表现出了高水平的性能。这证明了我们的模型在医学影像生成和分析任务中具有很高的潜力。

### 4 半监督训练

在医学影像领域，手动标记数据集相对于图像数据来说非常宝贵，因为标记一个三维图像需要专家投入大量的时间和精力。在这种情况下，我们选择使用 DeepAtlas 作为我们的半监督学习框架，该框架通过联合学习图像配准和图像分割来实现在有限标记数据下训练高质量模型的目标。

## 4.1 概述

DeepAtlas 是一个深度学习框架，结合了图像配准和图像分割的半监督学习。与传统的深度无监督图像配准方法相比，DeepAtlas 可以在有分割标签的情况下使用现有的分割，否则通过分割网络进行计算，从而达到相同的配准效果。同时，分割网络的训练受益于配准，提供了一种真实的数据增强形式。通过对其他类似膝盖和脑部 3D 磁共振图像的实验，DeepAtlas 展示了同时提高分割和配准准确度的能力，允许我们使用少量标记数据训练高质量模型。

## 4.2 深度图谱训练流程

首先，我们构建了分割和配准模型，并使用少量分割数据（甚至只有一张）预训练分割模型，即使其性能不够强大也是可以接受的。然后，我们使用分割模型作为弱监督来源来辅助训练配准模型，反之亦然。最后，我们反复进行这个过程，或者在联合训练循环中同时进行分割和配准的训练。

### 4.2.1 分割辅助配准

如果我们有一个强大的分割算法可以在一类图像中标记解剖特征，我们可以使用它来评估配准模型的性能。具体步骤如下：

- 1 注册一对图像，将“移动”图像变形到“目标”图像。
- 2 获取这两张图像的分割结果。
- 3 使用配准的变形将移动图像的分割结果扭曲，得到目标图像的候选分割。
- 4 如果扭曲后的移动图像分割与目标图像分割匹配，说明配准算法表现良好。
- 5 如果我们希望使用深度神经网络进行配准，可以使用分割算法创建训练目标，产生的损失函数称为解剖损失。

### 4.2.2 配准辅助分割

假设我们有一个强大的配准算法，可以将一张图像变形到另一张图像，并以解剖学上真实的方式进行变形。只需一张手动标记的分割图像，我们就可以拥有同样强大的分割算法：

- 1 将需要分割的图像注册到已经有分割的图谱上。
- 2 使用配准的变形将图谱的分割结果扭曲回图像。
- 3 如果我们希望使用深度神经网络进行分割，可以将配准算法作为数据增强的形式。

## 5 结果分析

通过应用半监督训练的 DeepAtlas，我们观察到膀胱壁肿瘤分割模型性能的

显著提升，具体体现在 Dice 系数的提高。Dice 系数是衡量分割模型性能的重要指标，值越接近 1 表示模型的分割效果越好。在我们的训练中，Dice 系数从 0.81 提升到了 0.85。

### 5.1 Dice 系数的提升

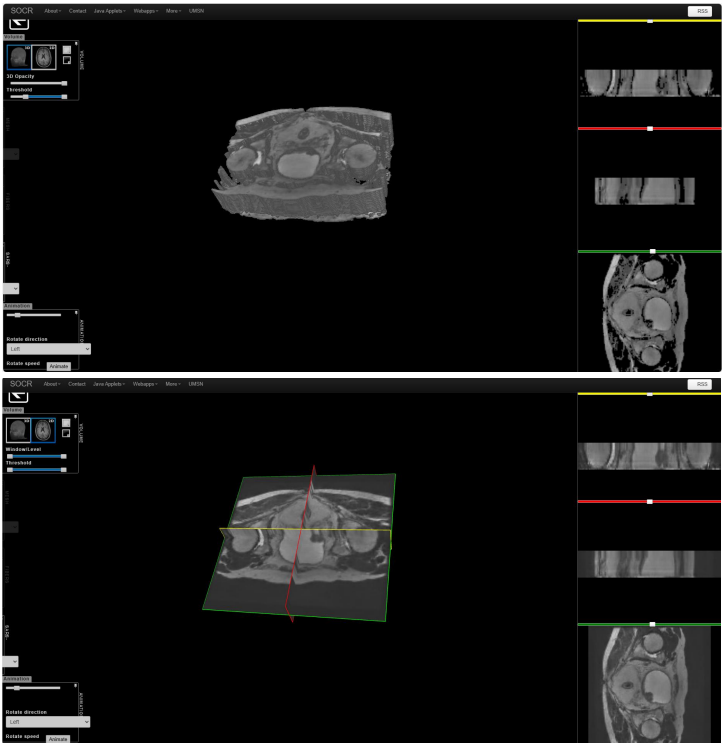
Dice 系数（也被称为 Dice Similarity Coefficient, DSC）是一种用于衡量两个样本相似度的指标，特别常用于医学影像分割的评估。在我们的情况下，Dice 系数用于衡量模型预测的肿瘤分割与真实分割之间的相似度。

通过半监督训练，我们观察到 Dice 系数从 0.81 提升到了 0.85。这意味着我们的模型在预测肿瘤分割方面的准确性有了显著提升。提高的 Dice 系数表明模型预测的肿瘤分割与实际分割结果更加接近，验证了半监督训练方法的有效性。

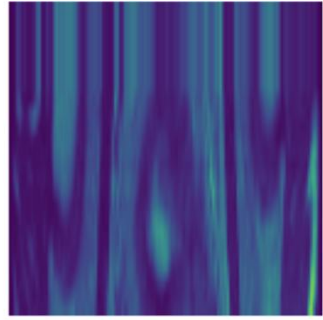
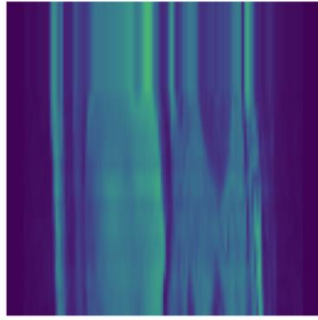
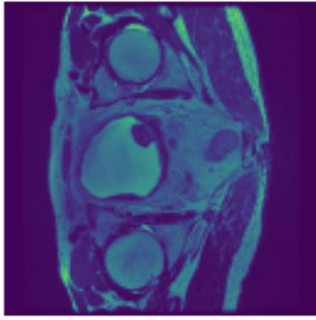
### 5.2 半监督训练的影响

这次 Dice 系数的提升验证了半监督训练在模型性能提升中的关键作用。在医学图像分割任务中，数据标注成本高昂，标注样本数量有限，半监督训练有效地利用了未标注数据，为模型提供更多学习信息。通过 DeepAtlas 框架的配准和分割相互配合，使模型能够从大量未标注图像中学习有用的信息，从而提高了分割效果。

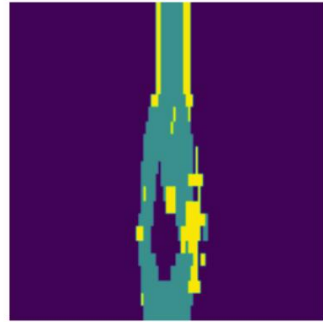
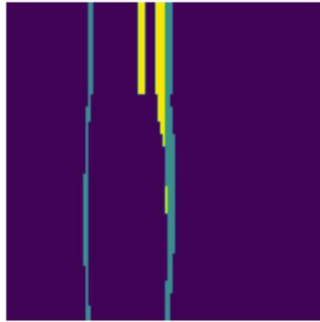
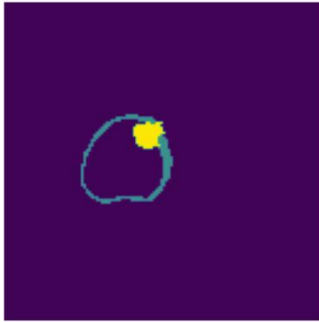
## 6 图像展示



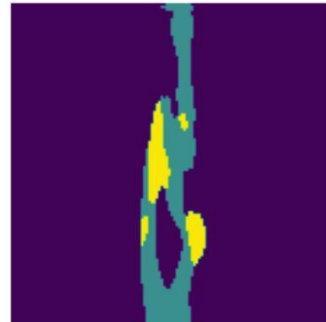
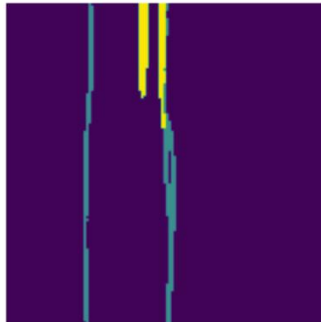
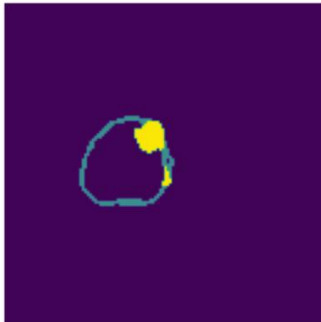
扩散生成的 3D 影像（共 200 份）



验证集的原始图像



真实的分割



预测的分割