arXiv:2304.04106v2 [eess.IV] 4 Jul 2023

MedGen3D：用于配对三维图像和掩膜生成的深度生成框架

韩坤[1], 熊一峰[1],游晨宇[2],普亚·霍斯拉维[1],孙善林[1],严向毅[1], 詹姆斯·邓肯[2], 和谢晓辉[1]

[1] 加利福尼亚大学欧文分校[2] 耶鲁大学

摘要。获取并标注足够的带标签数据对于开发准确且稳健的基于学习的模型至关重要，但在许多医学图像分割任务中，获取此类数据可能颇具挑战。一种有前景的解决方案是合成带有真实掩膜标注的逼真数据。然而，此前尚无研究探索生成带有掩膜的完整三维体积图像。在本文中，我们提出了MedGen3D，这是一种能够生成成对三维医学图像和掩膜的深度生成框架。首先，我们将三维医学数据表示为二维序列，并提出多条件扩散概率模型（MC-DPM），用于生成遵循解剖结构几何特征的多标签掩膜序列。随后，我们使用以生成的掩膜序列为条件的图像序列生成器和语义扩散精炼器，生成与这些掩膜相匹配的逼真三维医学图像。我们提出的框架确保了合成图像与分割图之间的准确对齐。在三维胸部CT和脑部MRI数据集上的实验表明，我们的合成数据既具有多样性，又忠实于原始数据，并展示了其对下游分割任务的益处。我们预计，MedGen3D合成成对三维医学图像和掩膜的能力将被证明对于训练用于医学成像任务的深度学习模型具有重要价值。

关键词： 深度生成框架 · 带掩膜的三维体积图像· 保真度与多样性 · 分割

1 引言

在医学图像分析中，获得大量准确标注的三维数据是实现分割和检测等任务高性能的先决条件 [26,17,29,9,31,34,35,32,33]。这反过来又能带来更精准的诊断和治疗方案。然而，获取并标注此类数据面临诸多挑战，包括医学图像的复杂性、对专业知识的要求以及隐私问题。

生成逼真的合成数据为解决上述挑战提供了一种有前景的方案，因为它消除了人工标注的需求，并降低了隐私风险。然而，大多数既往研究 [16,6,7,4,3,20,28,34,32,33]都侧重于二维图像合成，只有少数研究生成了对应的分割掩膜。例如， [15]使用双生成对抗网络（GAN） [14,35] 来

合成二维带标签的视网膜眼底图像，而 [12]结合标签生成器 [25] 与图像生成器 [24] 来生成二维脑部 MRI 数据。最近， [27] 使用 WGAN [5]生成小型三维图像块及对应的血管分割结果。

然而，此前尚无研究能够生成完整的三维体积图像及其对应的分割掩膜。生成带有对应分割掩膜的三维体积图像面临两大障碍。首先，由于 GPU 内存限制，直接将完整的三维体积输入神经网络并不可行，而降低分辨率可能会损害合成数据的质量。其次，在训练过程中将整个三维体积视为单个数据点并非最优方案，因为带标注的三维数据数量有限。因此，需要创新方法来克服这些挑战，并生成带有对应分割掩膜的高质量合成三维体积数据。

我们提出了 MedGen3D，这是一种新颖的基于扩散的深度生成框架，用于生成成对的三维体积医学图像和多标签掩膜。我们的方法将三维医学数据视为切片序列，并采用自回归过程依次生成三维掩膜和图像。在第一阶段，多条件扩散概率模型（MC-DPM）结合条件生成过程与无条件生成过程，生成掩膜序列。具体而言，MC-DPM 可以从随机噪声中直接生成任意位置的掩膜子序列（即若干个连续切片），也可以以已有切片为条件，向前或向后生成子序列。鉴于医学图像具有相似的解剖结构，切片索引作为额外条件，有助于掩膜子序列的生成。在第二阶段，我们引入了一个条件图像生成器，该生成器结合了来自 [30] 的序列到序列模型和语义扩散精炼器。通过以第一阶段生成的掩膜序列为条件，我们的图像生成器能够合成与掩膜对齐的逼真医学图像，同时保持相邻切片之间的空间一致性。

我们工作的主要贡献如下：1）我们提出的框架首次 应对了合成完整三维体积医学图像及其对应掩膜这一挑战；2）我们引入了一种多条件扩散概率模型，用于生成高保真度且多样性丰富的三维解剖掩膜；3）我们利用生成的掩膜作为条件，输入图像序列生成器和语义扩散精炼器，从而生成逼真且与生成掩膜准确对应的医学图像；4）我们展示了实验结果，证明所生成的三维多标签医学图像具有良好的保真度和多样性，并凸显了其在下游分割任务中的潜在益处。

2 预备知识

2.1 扩散概率模型

扩散概率模型（DPM） [18]是一个长度为 T 的参数化马尔可夫链，旨在学习数据分布 $p(X)$。DPM 构建

前向扩散过程（FDP），以通过 $q(X_t \mid X_{t-1}) = \mathcal{N}(X_t$ 在任意时间步 t 获取扩散后的数据点 X_t；$\sqrt{1-\beta_t}X_{t-1}$，$\beta_t I)$，其中 $X_0 \sim q(X_0)$ 且 $p(X_T) = \mathcal{N}(X_T; 0, I)$。令 $\alpha_t = 1-\beta_t$ 和 $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$，训练反向扩散过程（RDP），使其通过最小化以下目标来预测前向扩散过程中添加的噪声：

$$Loss(\theta) = \mathbb{E}_{X_0 \sim q(X_0), \epsilon \sim \mathcal{N}(0,I), t}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}X_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\right\|^2\right], \quad (1)$$

其中，ϵ_θ 是预测噪声，θ 是模型参数。

2.2 无分类器引导

通过以 p的概率将条件 c 设为 $\emptyset$，可以借助无分类器引导 [19] 改进条件扩散模型的采样结果。在采样过程中，模型的输出会进一步朝着 $\epsilon_\theta(X_t \mid c)$ 的方向外推，并远离 $\epsilon_\theta(X_t \mid \emptyset)$，如下所示：

$$\hat{\epsilon}_\theta(X_t \mid c) = \epsilon_\theta(X_t \mid \emptyset) + s \cdot (\epsilon_\theta(X_t \mid c) - \epsilon_\theta(X_t \mid \emptyset)), \quad (2)$$

其中 $\emptyset$ 表示空条件，$s \geq 1$ 是引导尺度。

3 方法论

我们提出了一种生成带掩膜的复杂三维体积图像的序列过程，如图 1 所示。第一阶段生成多标签分割，第二阶段执行条件医学图像生成。具体细节将在以下各节中介绍。



图 1. 所提出的**MedGen3D**概览，包括一个三维掩膜生成器，用于从随机位置 z 开始自回归地生成掩膜序列；以及一个条件图像生成器，用于生成以所生成掩膜为条件的三维图像。

3.1 三维掩膜生成器

由于真实标注数据有限且受 GPU 内存限制，直接将整个三维体积输入网络并不切实际。因此，我们将三维医学数据视为一系列子序列。为了生成完整的掩码序列，首先从随机噪声中无条件生成一个由 m 个连续切片组成的初始子序列。随后，以现有切片为条件，按照自回归方式向前和向后扩展该子序列。

受第 2.2节分类器无关引导的启发，我们提出了一种通用的多条件扩散概率模型（MC-DPM），以统一所有三种条件生成（无条件、正向和反向）。如图 2所示，MC-DPM 能够直接从随机噪声生成掩膜序列，或以现有切片为条件进行生成。

此外，由于三维医学数据通常具有相似的解剖结构，相对位置相同的切片大致对应相同的解剖区域。因此，我们可以利用切片的相对位置作为

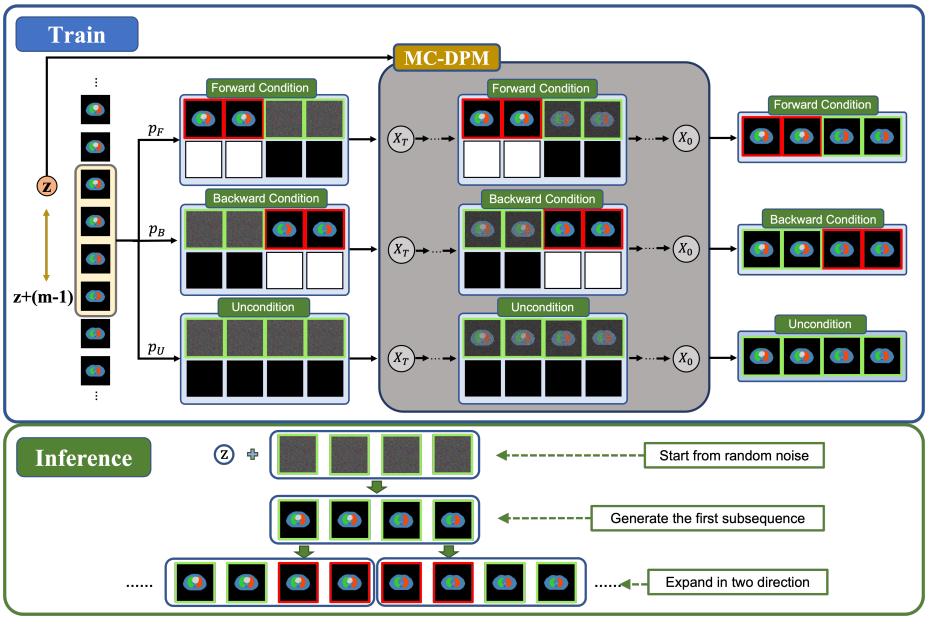


图2。 提出的三维掩膜生成器。给定目标位置 z，MC-DPM旨在根据预定义的概率，无条件地生成特定区域的掩膜子序列（长度为 m），或以第一个或最后一个 n 切片为条件进行生成。为表示条件切片，我们为切片分配二进制指示符。为实现清晰可视化，我们在推理过程中忽略二进制指示符，其中红色轮廓表示条件切片，绿色轮廓表示生成的切片。

条件，引导 MC-DPM 生成目标区域的子序列，并控制生成序列的长度。

训练：对于给定的三维多标签掩膜 $M \in \mathbb{R}^{D\times H\times W}$，选择由 m个连续切片组成的子序列作为 $\{M_z, M_{z+1}, \ldots, M_{z+(m-1)}\}$，其中 z为随机选择的起始索引。对于每个子序列，我们根据概率 $p^C \in \{p_{Forward}, p_{Backward}, p_{Uncondition}\}$，通过选择前 n 个切片、后 n 个切片或不选择切片来确定条件切片 $X^C \in \{\mathbb{R}^{n\times H\times W}, \emptyset\}$。MC-DPM 的目标是生成其余切片，记为$X^P \in \mathbb{R}^{(m-\text{len}(X^C))\times H\times W}$。

为纳入位置条件，我们使用子序列 $\tilde{z} = z/D$ 的相对位置，其中 z 是该子序列起始切片的索引。随后，我们对位置条件进行嵌入，并将其与时间嵌入拼接，以辅助生成过程。我们还为子序列中的每个切片使用一个二值指示符，以表示条件切片是否存在。

带有条件切片 X^C 的反向扩散过程（RDP）的联合分布可写为:

$$p_\theta(X^P_{0:T}|X^C, \tilde{z}) = p(X^P_T)\prod_{t=1}^{T} p_\theta(X^P_{t-1} \mid X^P_t, X^C, \tilde{z}). \tag{3}$$

其中 $p(X^P_T) = \mathcal{N}\left(X^P_T; 0, I\right)$、 $\tilde{z} = z/D$ 和 p_θ 是由模型参数化的分布。

总体而言，模型将通过最小化以下损失函数进行训练，其中 $X_t^P = \sqrt{\bar{\alpha}_t} X_0^P + \sqrt{1-\bar{\alpha}_t}\epsilon$：

$$\text{Loss}(\theta) = \mathbb{E}_{X_0 \sim q(X_0), \epsilon \sim \mathcal{N}(0,I), p^C, z, t} \left[\left\| \epsilon - \epsilon_\theta \left(X_t^P, X^C, z, t \right) \right\|^2 \right]. \quad (4)$$

推理： 在推理过程中，首先根据随机位置 z，从随机噪声生成一个由 n 个切片组成的序列。随后，如图2所示，可基于已有切片，向两个方向扩展，以自回归方式生成完整的掩膜序列。详细的生成过程和网络结构请参阅**补充材料**。

3.2 条件图像生成器

在第二步中，我们采用序列到序列的方法，根据掩膜生成医学图像，如图3所示。

图像序列生成器： 在序列到序列生成任务中，新的切片由变形后的前一切片和新生成的纹理组合而成，并由连续掩膜 [30]进行加权。我们使用Vid2Vid [30]作为我们的图像序列生成器。我们使用其原始损失训练Vid2Vid，其中包括多尺度图像和视频判别器上的GAN损失、光流估计损失以及特征匹配损失。



图 3。 图像序列生成器。给定生成的三维掩膜，Vid2Vid 模型会依次生成初始图像。为了利用语义扩散模型（SDM）优化初始结果，我们首先施加少量噪声步数（10 步），然后使用三个语义扩散模型进行优化。最终结果是来自 3 个不同视图（轴位、冠状位和矢状位）的三维图像均值，相较于初始生成图像有显著提升。

语义扩散精炼器： 尽管 Vid2Vid 实现了较高的跨切片一致性和空间连续性，但块状伪影、模糊以及纹理生成欠佳等问题仍然存在。鉴于扩散模型已被证明能够生成更优质的图像 [11]，我们提出了一种语义扩散精炼器，利用扩散概率模型对先前生成的图像进行精炼。

对于 3 种不同视角中的每一种，我们都训练一个语义扩散模型（SDM），该模型以二维掩膜和噪声图像作为输入，生成与输入掩膜对齐的图像。在推理阶段，我们仅对生成的图像执行少量加噪步骤（10 步），以保留整体解剖结构和空间连续性。之后，我们使用预训练语义扩散模型对图像进行精炼。最终精炼后的三维图像是 3 种视角结果的均值。实验结果表明，在语义扩散精炼器的帮助下，生成图像的质量有了显著提升。

4 实验与结果

4.1 数据集与设置

数据集：我们在胸部部位使用三个胸腔CT数据集进行了实验，并在脑部部位使用两个脑部MRI数据集进行了实验。对于生成模型和下游分割任务，我们使用了以下数据集：

– SegTHOR [22]: 三维胸腔CT扫描（25个训练集、5个验证集、10个测试集）；
– OASIS [23]: 三维脑部MRI T1扫描（40个训练集 ，10个验证集 ，10 个测试）；
仅针对下游分割任务和迁移学习，我们从 StructSeg-Thorax [2] 和 Public-Thor [9],的三个三维胸腔 CT 数据集中，以及 ADNI [1]的三维脑部 MRI T1 数据集中，各选取 10 个用于微调的扫描、5 个验证扫描和 10 个测试扫描。

实现： 对于胸腔数据集，我们将 CT 扫描裁剪并填充至 $(96 \times 320 \times 320)$。由一位经验丰富的放射肿瘤科医生检查六个器官（左肺、右肺、脊髓、食管、心脏和气管）的标注。我们还纳入身体掩膜，以辅助生成身体区域的图像。对于脑部 MRI 数据集，我们使用 Freesurfer [13] 获取四个区域（皮层、皮层下灰质、白质和脑脊液）的分割结果，然后将体积裁剪至 $(192 \times 160 \times 160)$。我们为胸腔和脑部数据集中不同区域或器官的掩膜赋予离散值，再将它们合并为一个三维体积。合成掩膜序列时，我们将掩膜的宽度和高度调整为 128×128，并将子序列m 的长度设为 6。我们使用 MONAI[8]提供的官方分割模型，并采用标准数据增强，包括空间变换和颜色变换。

设置： 我们将合成图像质量与 DDPM [18], 3D-α-WGAN[21] 和 Vid2Vid [30], 进行比较，并采用四种具有不同训练策略的分割模型，以展示其对下游任务的益处。

4.2 评估合成图像的质量。

合成数据集： 为解决标注三维医学数据有限的问题，我们仅使用了 SegTHOR 中的 30 例 CT扫描（25 例用于训练，5 例用于验证）和 OASIS 中的 50 例 MRI扫描（40 例用于训练，10 例用于验证），分别生成 110 例三维胸部CT扫描和 110 例三维脑部磁共振成像。

	胸部CT		脑部磁共振成像	
	FID ↓	学习感知图像块相似度 ↑	Fréchet起始距离 ↓	学习感知图像块相似度 ↑
去噪扩散概率模型 [18]	**35.2**	**0.316**	**34.9**	0.298
三维-α-WGAN [21]	136.2	0.286	136.4	0.289
Vid2Vid [30]	47.3	0.300	48.2	0.324
Ours	39.6	0.305	40.3	**0.326**

表1. 合成图像质量比较：基线方法与我们的方法。

我们通过从 3 个不同视角计算平均 Frèchet Inception Distance（Fréchet起始距离，FID）和 Learned PerceptualImage Patch Similarity（学习感知图像块相似度，LPIPS），比较我们的合成数据与DDPM [18]（为不同视角训练 3 个）、三维-α-WGAN [21],以及 Vid2Vid [30]的保真度和多样性。

根据表1，与去噪扩散概率模型相比，我们提出的方法具有略低的FID分数，但LPIPS分数相近。我们推测，这是因为去噪扩散概率模型是在没有明确解剖约束的二维图像上进行训练的，并且只能生成二维图像。另一方面，使用规模大得多的三维训练数据（三维胸腔数据146个、脑数据414个）的3D-α-WGAN [18], ，其FID和LPIPS分数明显差于我们的方法。此外，我们提出的方法优于Vid2Vid，显示了语义扩散精炼器的有效性。

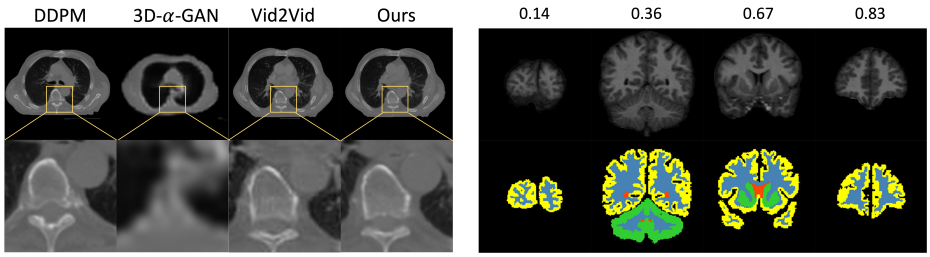


图4。 与 3D-α-WGAN 和 Vid2Vid 相比，我们提出的方法生成的图像在解剖学上更加准确，这体现在更清晰的器官边界和更逼真的纹理上。左：不同生成模型之间的定性比较。右：不同相对位置处的合成三维脑部磁共振成像切片可视化结果。

4.3 评估合成数据对分割任务的益处。

我们通过比较 4 种分割模型的 Sørensen–Dice系数（DSC），包括 Unet2D [26], UNet3D [10], UNETR [17],和 Swin-UNETR [29], 来探究合成数据对下游分割任务的益处。在表2和表3中，我们利用真实训练数据（来自 SegTHOR 和 OASIS）以及合成数据，采用 5 种不同策略训练分割模型，并在全部 3 个胸腔CT数据集和 2 个脑部MRI数据集上进行测试。在表4中，我们旨在展示合成数据能否在每个测试数据集（StructSeg-Thorax、Public-Thor 和 ADNI）提供的有限真实微调数据下，通过 4 种训练策略促进迁移学习。

	SegTHOR*				StructSeg-Thorax				Public-Thor			
	Unet 2D	Unet 3D	UNETR	Swin UNETR	Unet 2D	Unet 3D	UNETR	Swin UNETR	Unet 2D	Unet 3D	UNETR	Swin UNETR
E2-1	0.817	0.873	0.867	0.878	0.722	0.793	0.789	0.810	0.822	0.837	0.836	0.847
E2-2	0.815	0.846	0.845	0.854	0.736	0.788	0.788	0.803	0.786	0.838	0.814	0.842
E2-3	0.845	0.881	0.886	0.886	0.772	0.827	0.824	0.827	0.812	0.856	0.853	0.856
E2-4	**0.855**	0.887	**0.894**	**0.899**	0.775	**0.833**	**0.825**	0.833	**0.824**	0.861	0.852	**0.867**
E2-5	0.847	**0.891**	0.890	0.897	**0.783**	**0.833**	0.823	**0.835**	0.818	**0.864**	**0.858**	**0.867**

表2。 实验2：不同胸部分割模型的 DSC。共有5种训练策略，分别为：**E2-1：** 使用真实 SegTHOR 训练数据进行训练；**E2-2：** 使用合成数据进行训练；**E2-3：** 使用合成数据和真实数据进行训练；**E2-4：** 使用真实训练数据对 E2-2 中的模型进行微调；以及**E2-5：** 使用真实训练数据对 E2-3 中的模型进行微调。（* 表示训练数据来源。）

根据表2和表3，二维与三维分割模型之间显著的DSC差异凸显了三维标注数据的关键作用。

虽然纯合成数据（**E2-2**）无法达到与真实训练数据（**E2-1**）相同的性能，但真实数据与合成数据的结合（**E2-3**）在大多数情况下都能提升模型性能，Public-Thor 数据集上的 Unet2D 除外。此外，使用真实数据对预训练模型进行微调（**E2-4** 和 **E2-5**），其性能始终优于仅使用真实数据训练的模型。有关 Swin-UNETR 模型的器官级 DSC 比较详情，请参阅**补充材料** 。

	OASIS*				ADNI			
	Unet 2D	Unet 3D	UNETR	Swin UNETR	Unet 2D	Unet 3D	UNETR	Swin UNETR
E2-1	0.930	0.951	0.952	0.954	0.815	0.826	0.880	0.894
E2-2	0.905	0.936	0.935	0.934	0.759	0.825	0.828	0.854
E2-3	0.938	0.953	0.953	0.955	0.818	0.888	0.898	**0.906**
E2-4	**0.940**	**0.955**	**0.954**	**0.956**	**0.819**	0.891	**0.903**	0.903
E2-5	**0.940**	0.954	**0.954**	**0.956**	**0.819**	**0.894**	0.902	**0.906**

表3。实验2：脑分割模型的DSC。详细训练策略请参阅表2。（*表示训练数据来源。）

根据表4，对于迁移学习，利用预训练模型（**E3-2**）相比从头训练（**E3-1**）能够取得更好的性能。此外，使用合成数据（**E3-3** 与**E3-4**）对模型进行预训练，有助于将迁移学习应用于标注数据有限的新数据集。

	胸部CT		脑部磁共振成像
	StructSeg-Thorax*	Public-Thor*	ADNI*
E3-1	0.845	0.897	0.946
E3-2	0.865	0.901	0.948
E3-3	0.878	0.913	**0.949**
E3-4	**0.882**	**0.914**	**0.949**

表4。 实验3：使用真实数据集对微调后的 Swin-UNETR 进行 DSC 评估。共有4种训练策略：**E3-1：** 使用有限的微调数据为每个数据集从头训练；**E3-2** 微调实验2中的模型 E2-1；**E3-3** 微调实验2中的模型 E2-4；以及**E3-4** 微调实验2中的模型 E2-5。（* 表示微调数据来源。）

我们在**补充材料**中提供了生成的三维体积图像的视频演示，这些演示能够更全面地呈现生成图像的质量。

5 结论

本文介绍了MedGen3D，这是一种用于合成三维医学掩膜-图像对的新型框架。我们的实验表明，在标注数据有限的情况下，该框架在逼真数据生成和下游分割任务中具有良好的潜力。未来的工作包括融合图像序列生成器和语义扩散精炼器，以实现端到端训练，并扩展该框架以合成跨成像模态的三维医学图像。总的来说，我们相信这项工作为生成与掩膜配对的高质量三维医学图像开辟了新的可能性，并期待该领域未来的发展。

参考文献

1. https://adni.loni.usc.edu/ 2. https://structseg2019.grand-challenge.org/dataset/ 3. Abbasi-Sureshjani, S., Amirrajab, S., Lorenz, C., Weese, J., Pluim, J., Breeuwer, M.: 基于 XCAT 解剖模型的四维语义心脏磁共振图像合成。载于：深度学习医学影像。第 6–18 页。PMLR（2020）4. Abhishek, K., Hamarneh, G.: Mask2lesion：基于掩膜约束的对抗性皮肤病变图像合成。载于：医学影像模拟与合成：第四届国际研讨会 SASHIMI 2019，与 MICCAI 2019 同期举行，中国深圳，2019 年 10 月 13 日，论文集。第 71–80 页。施普林格（2019）5. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein 生成对抗网络。arXiv 预印本 arXiv：Arxiv-1701.07875（2017）6. Baur, C., Albarqouni, S., Navab, N.: Melanogans：利用生成对抗网络合成高分辨率皮肤病变图像。arXiv 预印本 arXiv：1804.04338（2018）7. Bermudez, C., Plassard, A.J., Davis, L.T., Newton, A.T., Resnick, S.M., Land- man, B.A.: 利用深度学习构建脑部磁共振成像的隐式流形。载于：2018 年医学影像：图像处理。第 10574 卷，第 408–414 页。SPIE（2018）8. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., 等：MONAI：用于医疗保健深度学习的开源框架。arXiv 预印本 arXiv：2211.02701（2022）9. Chen, X., Sun, S., Bai, N., Han, K., Liu, Q., Yao, S., Tang, H., Zhang, C., Lu, Z., Huang, Q., 等：基于深度学习的全身计算机断层扫描图像危及器官自动分割系统，用于放射治疗。《放射治疗与肿瘤学》**160**，175–184（2021）10. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 三维 U-Net：从稀疏标注中学习密集体积图像分割。载于：医学图像计算与计算机辅助介入——MICCAI 2016：第 19 届国际会议，希腊雅典，2016 年 10 月 17–21 日，论文集，第 II 部分 19。第 424–432 页。施普林格（2016）11. Dhariwal, P., Nichol, A.: 扩散模型在图像合成方面优于生成对抗网络。《神经信息处理系统进展》**34**，8780–8794（2021）12. Fernandez, V., Pinaya, W.H.L., Borges, P., Tudosiu, P.D., Graham, M.S., Ver- cauteren, T., Cardoso, M.J.: 分割模型能否使用完全合成生成的数据进行训练？载于：医学影像模拟与合成：第七届国际研讨会 SASHIMI 2022，与 MICCAI 2022 同期举行，新加坡，2022 年 9 月 18 日，论文集。第 79–90 页。施普林格（2022）13. Fischl, B.: Freesurfer。《神经影像》**62**(2)，774–781（2012）14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: 生成对抗网络。《ACM 通讯》**63**(11)，139–144（2020）15. Guibas, J.T., Virdi, T.S., Li, P.S.: 由双生成对抗网络生成的合成医学图像。arXiv 预印本 arXiv：1709.01872（2017）16. Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., Fu- rukawa, Y., Mauri, G., Nakayama, H.: 基于生成对抗网络的合成脑部磁共振图像生成。载于：2018 年 IEEE 第 15 届生物医学成像国际研讨会（ISBI 2018）。第 734–738 页。IEEE（2018）17. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: UNETR：用于三维医学图像分割的 Transformer。载于：IEEE/CVF 冬季计算机视觉应用会议论文集。第 574–584 页（2022）

18. Ho, J., Jain, A., Abbeel, P.：去噪扩散概率模型。神经信息处理系统进展 **33**，6840–6851（2020） 19. Ho, J., Salimans, T.：无分类器扩散引导。arXiv预印本 arXiv: Arxiv- 2207.12598（2022） 20. Kim, S., Kim, B., Park, H.：用于改进数据增强的脑肿瘤多对比度磁共振图像合成。医学物理 **48**（5），2185–2198（2021） 21. Kwon, G., Han, C., Kim, D.s.：使用自动编码生成对抗网络生成三维脑磁共振成像。在：医学图像计算与计算机辅助干预–MICCAI 2019：第22届国际会议，中国深圳，2019年10月13–17日，论文集，第III部分 22。第118–126页。施普林格（2019） 22. Lambert, Z., Petitjean, C., Dubray, B., Kuan, S.：SegTHOR：CT图像中胸部危及器官的分割。在：2020年第十届图像处理理论、工具与应用国际会议（IPTA）。第1–6页。IEEE（2020） 23. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.：开放影像研究系列（OASIS）：年轻、中年、未患痴呆及患痴呆老年人的横断面磁共振成像数据。认知神经科学杂志 **19**（9），1498–1507（2007） 24. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.：采用空间自适应归一化的语义图像合成。在：IEEE/CVF计算机视觉与模式识别会议论文集。第2337–2346页（2019） 25. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.：采用潜在扩散模型的高分辨率图像合成。在：IEEE/CVF计算机视觉与模式识别会议论文集。第10684–10695页（2022） 26. Ronneberger, O., Fischer, P., Brox, T.：U-Net：用于生物医学图像分割的卷积网络。在：医学图像计算与计算机辅助干预–MICCAI 2015：第18届国际会议，德国慕尼黑，2015年10月5–9日，论文集，第III部分 18。第234–241页。施普林格（2015） 27. Subramaniam, P., Kossen, T., Ritter, K., Hennemuth, A., Hildebrand, K., Hilbert, A., Sobesky, J., Livne, M., Galinovic, I., Khalil, A.A., 等：使用生成对抗网络生成三维TOF-MRA体积和分割标签。医学图像分析 **78**，102396（2022） 28. Sun, L., Chen, J., Xu, Y., Gong, M., Yu, K., Batmanghelich, K.：用于三维高分辨率医学图像合成的分层摊销生成对抗网络。IEEE生物医学与健康信息学杂志 **26**（8），3966–3975（2022） 29. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.：用于三维医学图像分析的Swin Transformer自监督预训练。在：IEEE/CVF计算机视觉与模式识别会议论文集。第20730–20740页（2022） 30. Wang, T.C., Liu, M.Y., Zhu, J.Y., Liu, G., Tao, A., Kautz, J., Catanzaro, B.：视频到视频合成。arXiv预印本 arXiv:1808.06601（2018） 31. Yan, X., Tang, H., Sun, S., Ma, H., Kong, D., Xie, X.：After-UNet：用于医学图像分割的轴向融合变换器U-Net。在：IEEE/CVF冬季计算机视觉应用会议论文集。第3971–3981页（2022） 32. You, C., Dai, W., Liu, F., Su, H., Zhang, X., Staib, L., Duncan, J.S.：挖掘自身解剖结构：重新审视极少标签下的医学图像分割。arXiv预印本 arXiv:2209.13476（2022） 33. You, C., Dai, W., Min, Y., Liu, F., Zhang, X., Clifton, D.A., Zhou, S.K., Staib, L.H., Duncan, J.S.：重新思考半监督医学图像分割：方差缩减视角。arXiv预印本 arXiv:2302.01735（2023）

34. You, C., Dai, W., Staib, L., Duncan, J.S.：采用解剖结构感知对比蒸馏自举半监督医学图像分割。arXiv预印本 arXiv:2206.02307（2022） 35. You, C., Zhao, R., Liu, F., Dong, S., Chinchali, S., Topcu,U., Staib,L.,Duncan, J.：用于医学图像分割的类别感知对抗变换器。在：NeurIPS（2022）

arXiv:2304.04106v2 [eess.IV] 4 Jul 2023

算法 1 三维掩膜生成（推理）

要求： n：条件切片的数量，m：子序列的长度，L：
要生成的序列长度， P_θ：预测概率分布
$Z \leftarrow L -(m-1)$
$z \sim$ 均匀分布 $(\{0, 1, ..., Z\})$ ▷ 随机选择一个 z 作为起始位置
初始化一个空的掩码序列 $\mathcal{M}$
$X \leftarrow \{M_z, M_{z+1}, ..., M_{z+(m-1)}\} \sim P_\theta\ (X_0^P \mid X^C = \emptyset,\ z)$
$\mathcal{M} \leftarrow \mathcal{M} \cup X$
//前向采样
$z' \leftarrow z$
当 $z' <= Z$ **时进行#+#+#+#+**
$z' \leftarrow z' +(m-n)$
$X^C \leftarrow \mathcal{M}[-n:]$ ▷ 选择最后 n 个掩膜作为条件
$X^P \sim P_\theta(X_0^P \mid X^C,\ z')$ ▷ 采样接下来的 (m-n) 个掩膜
$\mathcal{M} \leftarrow \mathcal{M} \cup X^P$ ▷ 将生成的掩膜添加到序列末尾
结束 while 循环
//反向采样
$z' \leftarrow z$
while $z' >= 0$ **do**
$z' \leftarrow z' - (m-n)$
$X^C \leftarrow \mathcal{M}[: n]$ ▷ 选择前 n 个掩膜作为条件
$X^P \sim P_\theta(X_0^P \mid X^C, z')$ ▷ 采样之前的 (m-n) 个掩膜
$\mathcal{M} \leftarrow X^P \cup \mathcal{M}$ ▷ 将生成的掩膜添加到序列的开头
结束循环时
返回 $\mathcal{M}$ ▷ 返回生成的掩码序列

	SegTHOR*						StructSeg-Thorax						Public-Thor					
	ll	rl	ht	eso	tra	spin	ll	rl	ht	eso	tra	spin	ll	rl	ht	eso	tra	spin
E2-1	**0.98**	**0.99**	0.90	0.64	0.86	0.91	**0.95**	**0.960.91**	0.60	0.68	0.77	**0.97**	**0.98**	0.88	0.69	0.73	0.84	
E2-2	**0.98**	0.980.910.55		0.830.89		0.94	0.950.89		0.54	0.63	0.86	**0.970.98**		0.88	0.68	0.69	0.85	
E2-3	**0.98**	0.980.92		0.64	0.870.93		**0.95**	0.95	0.90	0.64	0.65	**0.87**	**0.970.98**	0.88	0.73	0.72	**0.86**	
E2-4	**0.98**	**0.990.93**		**0.67**	**0.90**	**0.94**	**0.95**	**0.96**	0.90	0.64	**0.69**	**0.87**	**0.97**	**0.98**	**0.90**	0.74	0.75	**0.86**
E2-5	**0.98**	**0.99**	0.92	**0.67**	0.88	**0.94**	**0.950.96**		0.90	**0.67**	**0.69**	0.86	**0.97**	**0.98**	0.89	**0.75**	**0.77**	0.85

ll：左肺 rl：右肺 ht：心脏 eso：食管 tra: 气管 spin：脊髓

表1。实验2：胸部部位的Swin-UNETR器官级DSC比较。有关详细策略，请参阅**表2** 中的**主文** 提交。

	OASIS*				ADNI			
	皮层	皮层下 Gray	白色物质	CSF	皮层	皮层下 Gray	白色物质	CSF
E2-1	0.941	0.965	0.970	0.941	0.884	0.856	0.929	0.908
E2-2	0.922	0.951	0.958	0.909	0.847	0.841	0.910	0.818
E2-3	0.942	0.964	0.970	0.944	**0.895**	**0.876**	**0.932**	**0.925**
E2-4	**0.943**	0.965	**0.971**	0.946	0.888	0.870	0.931	0.924
E2-5	**0.943**	**0.966**	**0.971**	**0.947**	**0.895**	0.875	0.931	0.919

表2. 实验2：Swin-UNETR 在脑部部位的器官级 DSC 对比。

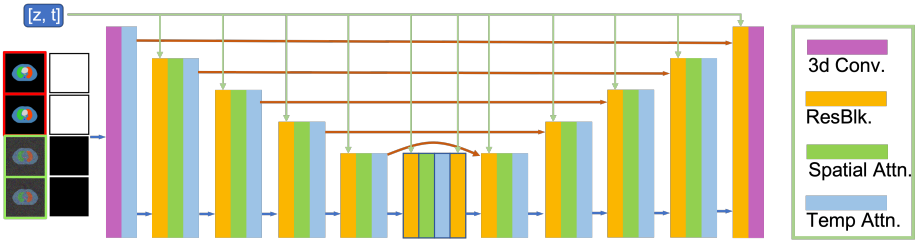


图1. MC-DPM的网络结构。

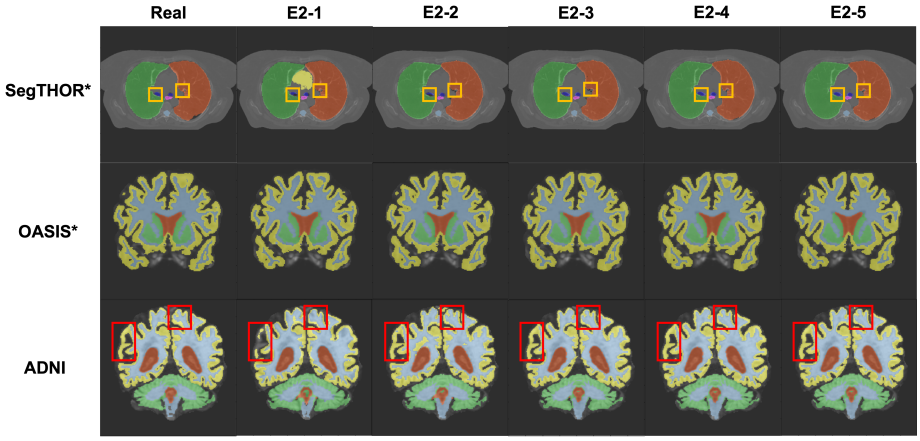


图2. 实验2：使用Swin-UNETR对不同训练策略进行定性比较。在脑部分割任务中，当训练数据和测试数据来自不同数据集（ADNI：在OASIS上训练，在ADNI上测试）时，合成数据带来的改进比训练数据和测试数据来自同一数据集（OASIS*）时更加明显。

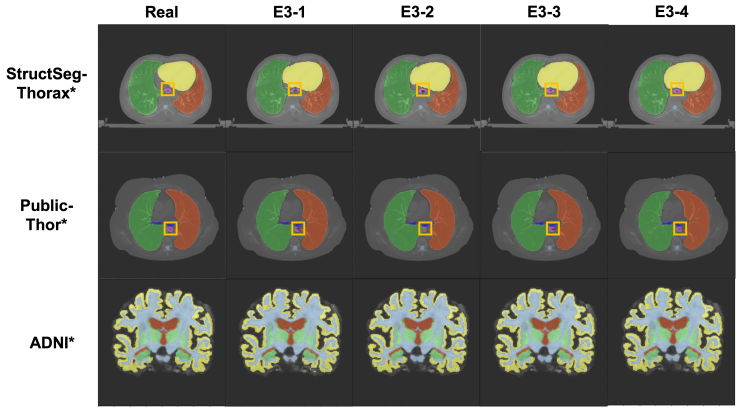


图3. 实验3：使用Swin-UNETR对不同微调策略进行定性比较。详细策略请参阅**表4** 中的主 投稿。