arXiv:2603.00985v1 [cs.CV] 1 Mar 2026

纹理-形状困境：面向三维医学变换器的边界安全合成生成

Jiaqi Tang[1⋆]、Weixuan Xu[2⋆]、Shu Zhang[3], Fandong Zhang[3], 和Qingchao Chen[1⋆⋆]

[1] 北京大学 jiaqi_tang@hsc.pku.edu.cn, qingchao.chen@pku.edu.cn[2] 中南大学 8208231610@csu.edu.cn[3] Deepwise有限公司 {zhangshu, zhangfandong}@deepwise.com

摘要。 视觉变换器（ViT）彻底革新了医学图像分析，但其对数据的巨大需求与临床档案的数据稀缺性及隐私限制相冲突。公式驱动的监督学习（FDSL）已成为解决这一瓶颈的有前景方案，它无需使用真实患者数据，即可根据数学公式合成无限量的标注样本。然而，现有的 FDSL 范式依赖强度均匀的简单几何形状，忽略了计算机断层扫描（CT）和 MRI 等成像模态固有的组织纹理与噪声模式，因此存在显著差距。本文指出了一种称为边界混叠的关键优化冲突：直接添加高频合成纹理会破坏学习结构边界所需的图像梯度信号，导致模型无法准确描绘真实的解剖边缘。为弥合这一差距，我们提出一种物理启发的空间解耦合成框架。该方法将合成过程正交化：首先基于边界距离构建梯度屏蔽缓冲区，以确保形状学习稳定；随后在对象核心区域注入物理驱动的频谱纹理。这一设计有效兼顾了稳健的形状表征学习与对采集噪声的不变性。在 BTCV 和 MSD数据集上的大量实验表明，我们的方法显著优于以往的 FDSL 方法以及使用真实世界医学数据集训练的自监督学习方法，在 BTCV 上提升 1.43%，在 MSD 任务上最高提升 1.51%，为医学视觉Transformer提供了一种可扩展、无需标注的基础。代码将在论文被接收后公开。

⋆ 这些作者对本研究做出了同等贡献。⋆⋆ 通讯作者。

1 引言

视觉变换器（ViT） [11] 改变了医学图像分析，但由于缺乏归纳偏置，它们对数据的需求极大且极易过拟合[4]。虽然自监督学习（SSL）利用无标注数据缓解了这一问题 [2, 19, 16], , 但仍受制于临床档案，延续了隐私限制和数据集特定偏差。因此，我们亟需无需依赖受限的真实世界数据采集流程、却能有效初始化 ViT 的预训练范式。

为应对这些局限，公式驱动监督学习（FDSL）已成为一种颇具吸引力且保护隐私的替代方案。FDSL 不依赖自然数据，而是直接根据数学公式合成海量图像及其稠密标注掩膜，使模型能够以完全监督的方式进行预训练。 [13, 15, 12, 8]。通过从合成模式中学习几何基元，FDSL 提供了理论上近乎无限的自动标注多样性来源，完全绕开了人工标注和患者隐私方面的瓶颈。此前，面向自然图像的 FDSL 工作（如 TileDB [8]和 RCDB [7], ）表明，仅在合成等高线上进行预训练，其表现便可媲美使用 ImageNet-21k [3]等大规模真实数据集所取得的效果。近期，PrimGeoSeg [14],将这一理念拓展至医学领域，利用三维几何基元（例如圆柱体、圆锥体）为体积网络注入空间感知能力。

尽管取得了这些进展，当前 FDSL 范式与医学影像的内在特征之间仍存在差距。如图 1.A 所示，现有方法（包括 PrimGeoSeg [14],）主要依赖几何轮廓，并采用均匀的恒定强度区域，完全缺乏内部纹理。尽管这种表示足以学习结构边界，却未能考虑 CT 和 MRI 等成像模态中固有的异质组织模式与采集噪声。然而，弥合这一差距并非易事。如图 1.B 所示，在先前的实验中，我们观察到，直接添加高频纹理可能会破坏边界学习所必需的图像梯度信号，导致 FDSL 预训练性能显著下降（从 56% 降至 40%）。我们在理论分析（第 2.1 节）中将这一现象正式定义为边界混叠。因此，在这类相互冲突的信号上预训练的模型难以适应真实病理中细微的强度分布，从而限制了 FDSL 向下游任务的迁移能力（下游性能从 83% 降至 80%）。

为应对这些挑战，我们提出了一种新颖的**受物理启发的空间解耦合成框架**，旨在弥合数学公式与真实医学采集之间的域间差距。我们的核心思路是将合成过程正交分解为两个互补模块：**屏蔽纹理模型**和**空间解耦纹理合成**。不同于以往简单叠加噪声的方法，我们首先引入屏蔽纹理模型，通过距离变换构建梯度屏蔽缓冲区。该拓扑约束确保边界学习所必需的梯度信号保持纯净，不受内部高频伪影影响。

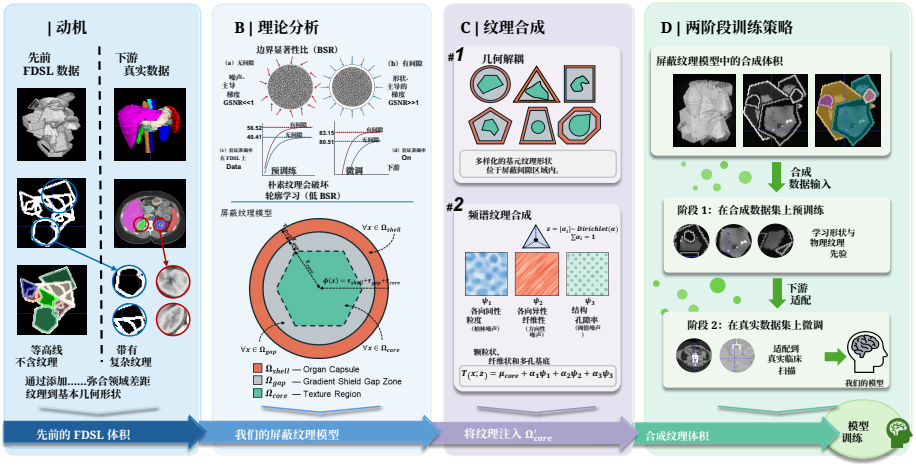


图1：所提出的物理启发的空间解耦合成框架概览。从左到右：（A）现有的 FDSL 依赖同质几何基元，导致其与真实临床数据之间存在纹理差距。（B）朴素的纹理注入会造成边界混叠，使随机梯度破坏结构线索；为此构建了梯度屏蔽缓冲区，以保留边界信号。（C）通过狄利克雷加权谱混合，将物理驱动纹理注入几何解耦核心。（D）合成后的体数据可用于合成预训练，随后在真实数据集上进行下游微调。

随后，我们将几何解耦与频谱纹理合成相结合。通过将物理驱动纹理注入几何上独立的核心区域，我们模拟真实解剖结构的生物物理外观，同时从数学上保证纹理模式不会干扰边界优化。

我们在 BTCV [10]和 MSD [1] 数据集上，使用标准的 UNETR [6] 和 SwinUNETR [5] 架构验证了我们的方法。大量实验表明，我们的方法显著优于从头训练和当前最先进的 FDSL基线方法：在 BTCV 上提升了 1.43%，在具有挑战性的 MSD 任务上最多提升了 1.51%。结果表明，将基于物理的纹理融入合成预训练能够有效降低对大规模标注医学数据集的依赖，为医学视觉 Transformer的训练提供了一种可扩展且保护隐私的途径。

2 方法论

以往的 FDSL 方法（如 PrimGeoSeg [14], ）完全依赖边界线索，无法捕捉真实医学扫描中固有的复杂纹理交互。受第 2.1 节理论分析的启发，我们提出边界显著性比（BSR），用于量化随机纹理梯度如何

破坏边界定位信号。为此，我们提出一种受物理启发的空间解耦合成框架。如图 1 所示，该框架包含两个针对上述挑战提出的新组件：**（i）屏蔽纹理模型**（第 2.2 节），通过构建梯度屏蔽缓冲区，确保边界鲁棒性并保持较高的 BSR；以及**（ii）空间解耦纹理合成**（第 2.3 节），采用几何解耦和物理驱动的频谱混合，在对象核心区域模拟组织特异性外观。这种设计有效协调了纹理异质性与几何优化的稳定性，使网络能够学习鲁棒的形状先验，同时避免受到高频噪声伪影的干扰，并弥合纹理差距。

2.1 边界混叠：纹理-边界冲突

我们分析随机纹理场如何影响输入图像的空间强度梯度，这些梯度是神经网络检测解剖边界的基本线索。这项分析可作为一种诊断手段，揭示朴素纹理合成为何会导致特征损坏。

设置。设 $\Omega \subset \mathbb{R}^3$ 表示三维图像域。几何基元由二值占据图 M 表示：$\Omega \to \{0,1\}$。我们分析边界 ∂M 处的空间导数 $\nabla_{\mathbf{x}}$。在离散体素坐标系中（单位长度 $\Delta x = 1$），边界两侧的强度对比度 $\Delta\mu$ 意味着空间梯度幅值为 $\Delta\mu$。遵循先前的 FDSL 方法 [14], ，我们假设背景均匀 $B(\mathbf{x}) \equiv \mu_{bg}$，从而得到 $\nabla_{\mathbf{x}} B = \mathbf{0}$。

对于干净的几何输入 $\mathbf{X}^{geo}$，边界体素$\mathbf{x}_b$处的空间梯度为边缘检测提供了纯净信号，可由表面法向量很好地近似：

$$\nabla_{\mathbf{x}} \mathbf{X}_{geo}(\mathbf{x}_b) \approx \Delta\mu\, \hat{\mathbf{n}}(\mathbf{x}_b) + \mathcal{O}(\delta), \tag{1}$$

其中 $\hat{\mathbf{n}}$ 是朝外的单位法向量。该向量 $\nabla_{\mathbf{x}} \mathbf{X}^{geo}$ 表示网络应学习以编码形状先验的理想几何线索 。

朴素纹理合成下的信号损坏。 一种朴素方法是将随机纹理 $T(\mathbf{x})$ 直接叠加到形状上：$\mathbf{X}_{naive}(\mathbf{x}) = M(\mathbf{x})T(\mathbf{x}) + (1 - M(\mathbf{x}))B(\mathbf{x})$。根据离散乘积法则展开空间梯度，可揭示出一种冲突：

$$\nabla_{\mathbf{x}} \mathbf{X}_{naive}(\mathbf{x}_b) \approx \underbrace{[T(\mathbf{x}_b) - B(\mathbf{x}_b)]\, \hat{\mathbf{n}}(\mathbf{x}_b)}_{\text{Geometric Signal}} + \underbrace{M(\mathbf{x}_b) \nabla_{\mathbf{x}} T(\mathbf{x}_b)}_{\text{Texture Interference}}. \tag{2}$$

由于医学纹理常呈现高频变化（例如骨小梁），干扰项 $\|\nabla_{\mathbf{x}} T\|$ 的幅度通常与边界对比度相当，导致纹理梯度压过几何法向量，并干扰网络的特征提取器。为量化这种冲突，我们定义边界体素处的边界显著性比（*BSR*）：

$$\mathrm{BSR}(\mathbf{x}_b) = \frac{|T(\mathbf{x}_b) - B(\mathbf{x}_b)|^2}{\mathbb{E}_{\mathbf{z}}[\|\nabla_{\mathbf{x}} T(\mathbf{x}_b; \mathbf{z})\|^2] + \epsilon}. \tag{3}$$

当边界体素$\mathbf{x}_b$处的 BSR $\ll$ 1时，输入特征会被随机纹理伪影主导，而非结构边界，这种情况称为边界混叠。这意味着，要学习稳健的形状先验，我们必须在 ∂M的局部邻域内强制满足 $\|\nabla_{\mathbf{x}} T(\mathbf{x})\| \to 0$。这一理论洞见直接启发了我们的屏蔽纹理模型：该模型构建梯度屏蔽缓冲区，以保证较高的 BSR，同时保留内部纹理多样性。

2.2 纹理模型的屏蔽

为解决上述冲突，我们引入一种空间解耦机制。我们不再进行全局纹理化，而是基于欧几里得距离变换（EDT）将体积正交划分为不同功能层，$\mathcal{D}_M(\mathbf{x}) = \min_{\mathbf{y}\in\partial M} \|\mathbf{x}-\mathbf{y}\|_2$。我们将前景 Ω_{obj} 划分为三个不同的区域：

$$\Omega_k = \begin{cases} \Omega_{shell} & \text{if } 0 < \mathcal{D}_M(\mathbf{x}) \le \tau_{shell} \\ \Omega_{gap} & \text{if } \tau_{shell} < \mathcal{D}_M(\mathbf{x}) \le \tau_{shell} + \tau_{gap} \\ \Omega_{core} & \text{if } \mathcal{D}_M(\mathbf{x}) > \tau_{shell} + \tau_{gap} \end{cases}, \tag{4}$$

其中，Ω_{shell} 模拟器官包膜，强度恒定为 μ_{shell}。

在此模式下，区域 Ω_{gap} 充当关键的梯度屏蔽。通过在该缓冲区 μ_{gap} 内强制采用恒定强度过渡，我们严格施加 $\nabla \mathbf{X}(\mathbf{x}) = \mathbf{0}, \forall \mathbf{x} \in \Omega_{gap}$。关键的是，缓冲区宽度 τ_{gap} 设置为大于网络首层梯度算子核大小的值（例如，针对 3×3核的 $\tau_{gap} \ge 2$）。这一拓扑约束从理论上确保 $\text{BSR}(\mathbf{x}_b) \to \infty$，使网络能够学习仅由确定性对比度驱动的形状先验，避免受到纹理核心区随机性破坏的影响。

2.3 空间解耦的物理启发式纹理合成

尽管空间解耦保证了边界稳定性，但在 Ω_{core} 内使用简单的随机纹理可能导致平凡解，例如过拟合于同心图案。我们进一步提出一种物理驱动建模方法，通过几何解耦和频谱纹理合成提升真实感，同时避免这类捷径。

几何解耦。直接对外部形状 M 进行腐蚀以定义纹理区域，会产生很强的空间相关性，网络可能会利用这种相关性。为打破这种依赖关系，我们使用独立的几何基元 M_{inner}（例如，圆柱体内的棱柱）生成纹理承载区域 Ω'_{core}。最终的核心区域定义为经过变换的几何基元与距离变换安全区的交集：

$$\Omega'_{core} = \mathcal{A}(M_{inner}) \cap \{\mathbf{x} : \phi(\mathbf{x}) > \tau_{shell} + \tau_{gap}\}, \tag{5}$$

其中，$\mathcal{A}(\cdot)$ 表示随机仿射变换。这确保内部纹理边界 $\partial\Omega'_{core}$ 与器官边界在空间上不相关

表 1：BTCV 数据集上的分割性能比较。

预训练 PT	数量	类型	平均值	脾脏	右肾	左肾	胆囊	食管	Liv	Sto	主动脉或下腔静脉、胰腺、右前降支			
UNETR														
从头训练	0	–	75.86	93.02	89.79	90.60	51.34	71.12	95.02	72.90	82.95	76.71	69.05	76.29	62.02	55.33
PrimGeoSeg	5K	FDSL	78.90	93.19	93.47	93.35	54.49	68.91	95.88	78.83	87.82	82.47	72.73	80.08	65.06	59.35
Ours	5K	FDSL														
SwinUNETR														
从头训练	0	–	80.12	95.32	94.01	93.51	61.17	73.83	94.40	71.19	88.13	85.20	73.12	80.60	67.15	63.88
PrimGeoSeg	5K	FDSL	80.08	95.25	93.76	93.79	61.23	74.47	96.04	80.64	89.39	83.91	74.27	77.83	64.13	56.36
Ours	5K	FDSL														

∂M，迫使网络学习全局形状语义，而非局部强度变化。

频谱纹理合成。为弥合与医学扫描图像之间的纹理差距，我们将纹理场 $\mathcal{T}(\mathbf{x})$建模为生物物理原型的凸组合，而非简单的高斯噪声。我们引入一个频谱混合函数：

$$\mathcal{T}(\mathbf{x};\mathbf{z}) = \mu_{core} + \sum_{i=1}^{3} \alpha_i \Psi_i(\mathbf{x}), \quad \text{s.t. } \sum \alpha_i = 1, \tag{6}$$

其中$\mathbf{z} = \{\alpha_i\}$ 是从狄利克雷分布中抽取的混合权重。基函数 Ψ_i 捕捉不同的组织特性：（1）通过多尺度柏林噪声生成实质组织的各向同性颗粒度；（2）通过方向缩放的噪声场 $\Psi(\mathbf{S_{VX}})$生成各向异性纤维化；以及（3）通过阈值化噪声场模拟骨小梁，从而生成结构孔隙度。将这种复杂的高频信号严格限制在 Ω'_{core} 内，我们既能合成纹理逼真的体数据，又能确保边界处梯度为零，这对几何学习至关重要。

表2：多个MSD数据集上的分割性能比较。

方法	MSD数据集					
	任务02		任务06		任务09	
	UNETR	SwinUNETR	UNETR	SwinUNETR	UNETR	SwinUNETR
从头训练	95.14	95.34	70.68	76.14	92.52	95.92
PrimGeoSeg	95.61	95.77	78.81	80.39	96.30	96.96
Ours	**95.91**	**96.29**	**79.20**	**81.47**	**96.63**	**98.47**

3 实验与结果

3.1 实验设置

任务和数据集： 在本部分中，我们在 BTCV [10]和 MSD [1] 数据集上开展实验，以评估所提出的方法。BTCV 数据集包含

30 个带有多器官标注的计算机断层扫描体数据。对于 MSD，我们在任务02（心脏）、任务06（肺）和任务09（脾脏）上进行评估。遵循以往方法 [14], ，这些数据集按照 80/20 的比例划分为训练子集和验证子集，所有实验均在离线设置下进行。

实现细节：在预训练期间，我们使用大小为 $96\times96\times96$ 的感兴趣区域、批量大小为 4、学习率为 0.0002、权重衰减为 0.00001。模型采用 AdamW 优化，并使用预热余弦学习率调度器，训练迭代次数设为 25000。微调时，我们遵循先前研究 [14] 使用的标准超参数设置，批量大小为 4。具体而言，SwinUNETR [5] 在 BTCV 上的设置遵循 [17], 报告的配置，而 MSD 的超参数设置则遵循 [6] 所述的既定流程。

3.2 结果

完整数据集上的性能如表 1 和表 2 所示，在 BTCV 和 MSD 任务中，无论使用 UNETR 还是 SwinUNETR，我们的方法始终优于从头训练和仅几何预训练基线。在 BTCV 上，使用 SwinUNETR 时，平均 Dice系数较从头训练提高 1.39，较基线方法提高 1.43。大多数器官均有提升，包括胰腺和肾上腺等细小结构。在 MSD 上，Task06 的提升更为明显；该任务具有更显著的外观变化和边界模糊性，性能较从头训练提高 5.33，较 PrimGeoSeg 基线提高 1.08。

与其他自监督学习方法的比较 我们在 BTCV 数据集上，将合成预训练策略与在 RadGenome 数据集 [20]上使用真实计算机断层扫描数据集预训练的代表性自监督学习 [18, 17] 方法进行比较。尽管 SwinUNETR 自监督学习方法略有提升，我们的方法取得了 81.51 的 Dice 系数，优于真实数据自监督学习方法和 FDSL基线方法。

表 3：与在 RadGenome数据集 [20]上预训练的自监督学习方法的比较

方法	预训练数据	数据量	平均 Dice
SwinMM [18]	真实数据（计算机断层扫描）	5,000	76.72
SwinUNETR [17]	真实（计算机断层扫描）	5,000	80.56
Ours	**合成**	**5,000**	**81.51**

3.3 分析

我们进一步分析了本方法中的四个关键因素：数据集规模、纹理来源、边界间隙宽度和骨干网络架构。对于前三个因素，实验在 BTCV 上使用 SwinUNETR 进行。结果汇总

表4：在不同设置下使用 SwinUNETR 对 BTCV 数据集进行的分析，包括预训练数据集规模、纹理类型和间隙宽度。

数据集规模		**纹理类型**		**间隙宽度**	
规模	平均 Dice	Type	平均 Dice	宽度	平均 Dice
500	80.16	果实	80.39	$w = 0$	80.02
5000	81.51	小梁状	80.56	$w = 1$	80.22
15000	82.26	纤维状	81.10	$w = 5$	80.55
50000	**82.40**	颗粒状	80.41	$w = 9$	**80.59**
–	–	**ours**	**81.51**	$w = 13$	80.15

于表4。在骨干网络分析中，我们进一步将 UNet 与基于 Transformer 的架构进行比较。

规模效应我们调整合成体积的数量。规模从 500 增加到 5,000，Dice系数从 80.16 提升至 81.51。进一步扩大到 15,000 和 50,000 时，Dice系数分别达到 82.26 和 82.40。随着合成数据集规模增大，结果稳步提升，但规模越大，增益越小。

纹理选择分析本节评估纹理选择的影响。除单一类型的物理模拟外，我们还引入真实水果纹理 [9] 进行比较。不同的纹理来源会导致明显的性能差异。与单一类型纹理和基于水果的纹理相比，我们的纹理设计取得了最高的 Dice系数，这表明有效的预训练更依赖于结构一致性，而非外观多样性。

间隙宽度分析我们调整边界区域与内部纹理之间的黑色间隙宽度 w。移除间隙后，Dice系数为 80.02。增大间隙宽度可持续提升性能。在 w =9 时取得最佳结果，Dice系数为 80.59，这表明明确分隔边界区域有助于模型学习更具判别性的结构感知表征。

骨干网络泛化分析 除了表1中报告的实验外，我们还在标准3D U-Net架构上进一步评估了我们的方法。从头训练获得了**78.87**Dice系数，而基线方法将性能提升至**80.64**。我们的方法进一步将Dice系数提升至**81.94**。这些结果表明，所提出的合成预训练策略不仅适用于基于Transformer的架构，也能有效增强卷积骨干网络。

4 结论

在本文中，我们通过一种受物理学启发的新型空间解耦合成框架，弥合了 FDSL 中的域间差距。通过识别并解决边界混叠冲突，我们的方法将稳健几何先验的学习与异质纹理建模正交化。大量实验表明，这种屏蔽与解耦范式不仅使我们的模型超越当前最先进的FDSL方法和SSL基线方法，更重要的是，使其达到了

与使用大规模真实医学数据集进行预训练相当的性能。因此，我们的方法为医学视觉Transformer的初始化提供了一种可扩展且保护隐私的基础方案，不受临床数据收集限制。

参考文献

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., 等：医学分割十项全能挑战赛。《自然·通讯》**13**(1), 4128 (2022) 2. Chen, Z., Agarwal, D., Aggarwal, K., Safta, W., Balan, M.M., Brown, K.：掩码图像建模推动三维医学图像分析的发展。载于：IEEE/CVF冬季计算机视觉应用会议论文集。第1970–1980页 (2023) 3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.：ImageNet：一个大型层次化图像数据库。载于：2009年IEEE计算机视觉与模式识别会议。第248–255页。IEEE (2009) 4. Dosovitskiy, A.：一张图像相当于16×16个词：用于大规模图像识别的Transformer。arXiv预印本 arXiv:2010.11929 (2020) 5. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H., Xu, D.：Swin UNETR：用于MRI图像中脑肿瘤语义分割的Swin Transformer (2022)，https://arxiv.org/abs/2201 01266 6 . . 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) 7. Kataoka, H., Hayamizu, R., Yamada, R., Nakashima, K., Takashima, S., Zhang, X., Martinez-Noriega, E.J., Inoue, N., Yokota, R.：以自动生成的轮廓取代带标签的真实图像数据集。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第21232–21241页 (2022) 8. Kataoka, H., Matsumoto, A., Yamada, R., Satoh, Y., Yamagata, E., Inoue, N.：基于递归铺砌图案的公式驱动监督学习。载于：IEEE/CVF国际计算机视觉会议论文集。第4098–4105页 (2021) 9. 鲁汶大学：植物器官三维X射线图像：X-Plant数据库，
h
t
t
ps://www.biw.kuleuven.be/biosyst/mebios/downloads/3d-x-ray-images-of-plant-organs 10. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.：颅腔之外的MICCAI多图谱标注——研讨会与挑战赛。载于：MICCAI颅腔之外的多图谱标注研讨会与挑战赛论文集。第5卷，第12页。德国慕尼黑 (2015) 11. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.：Swin Transformer：采用移位窗口的层次化视觉Transformer。载于：IEEE/CVF国际计算机视觉会议论文集。第10012–10022页 (2021) 12. Nakamura, R., Kataoka, H., Takashima, S., Noriega, E.J.M., Yokota, R., Inoue, N.：利用极少量合成图像预训练视觉变换器。载于：IEEE/CVF国际计算机视觉会议论文集。第20360–20369页 (2023)

13. Shinoda, R., Hayamizu, R., Nakashima, K., Inoue, N., Yokota, R., Kataoka, H.：Segrcdb：通过公式驱动的监督学习实现语义分割。载于：IEEE/CVF国际计算机视觉会议论文集。第20054–20063页 (2023) 14. Tadokoro, R., Yamada, R., Kataoka, H.：为三维医学图像分割预训练自动生成的体积形状。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第4740–4745页 (2023) 15. Takashima, S., Hayamizu, R., Inoue, N., Kataoka, H., Yokota, R.：视觉原子：利用正弦波预训练视觉变换器。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第18579–18588页 (2023) 16. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.：面向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（CVPR）论文集。第20730–20740页 (2022年6月) 17. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.：面向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) 18. Wang, Y., Li, Z., Mei, J., Wei, Z., Liu, L., Wang, C., Sang, S., Yuille, A.L., Xie, C., Zhou, Y.：SwinMM：结合Swin Transformer的掩码多视图方法，用于三维医学图像分割。载于：国际医学图像计算与计算机辅助干预会议。第486–496页。施普林格 (2023) 19. Wu, L., Zhuang, J., Chen, H.：VoCo：一种简单而有效的三维医学图像分析体积对比学习框架。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第22873–22882页 (2024年6月) 20. Zhang, X., Wu, C., Zhao, Z., Lei, J., Zhang, Y., Wang, Y., Xie, W.：Radgenome-Chest CT：一个用于胸部CT分析的基于视觉语言的对齐数据集。arXiv预印本 arXiv:2404.16754 (2024)