arXiv:2603.00979v1 [cs.CV] 1 Mar 2026

以假乱真：将解剖学逻辑注入医学分割的合成监督预训练

Jiaqi Tang[1]⋆、Mengyan Zheng[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] 北京航空航天大学 zmy050516@buaa.edu.cn[3] Deepwise Ltd. {zhangshu, zhangfandong}@deepwise.com

摘要。 视觉Transformer（ViT）在三维医学分割方面表现出色，但需要大量标注数据集。自监督学习（SSL）虽然利用无标签数据缓解了这一问题，但仍面临严格的隐私要求和实际操作障碍。公式驱动监督学习（FDSL）通过在合成数学基本图形上进行预训练，提供了一种保护隐私的替代方案。然而，一个关键的语义差距限制了其效能：通用形状缺乏真实解剖结构的形态保真度、固定空间布局和器官间关系，导致模型无法学习必要的全局结构先验。为弥合这一差距，我们提出一种**解剖信息引导的合成监督预训练** 框架，将FDSL的无限可扩展性与解剖学真实性相结合。我们用一个轻量级形状库取代基本图形，其中包含来自5名受试者的去标识化、仅标签分割掩膜。此外，我们还提出一种结构感知顺序放置策略，用于控制图块合成过程。该策略不采用随机放置，而是利用空间锚点确保正确定位，并通过拓扑图管理器官间交互（例如，防止不可能的重叠），以保证生理合理性。在BTCV和MSD数据集上的大量实验表明，我们的方法显著优于最先进的公式驱动监督学习基线方法和自监督学习方法，分别高出1.74%和最高1.66%；同时展现出稳健的缩放效应，即性能随合成数据量增加而提升。这为医学分割提供了一种数据高效且符合隐私要求的解决方案。代码将在论文被接收后公开。

1 引言

准确的 3D 医学图像分割对于临床诊断和治疗规划至关重要。尽管视觉 Transformer（ViT），例如 UNETR [4] 和

⋆ 这些作者对本研究贡献相同。⋆⋆ 通讯作者。

SwinUNETR [3],通过捕获长距离依赖展现出令人瞩目的性能，但它们缺乏卷积神经网络所固有的归纳偏置 [2]。因此，在有限标注数据上训练时，这些架构往往极度依赖数据，并容易过拟合。这造成了一个关键瓶颈：收集大规模、逐体素标注的医学数据集成本高得令人难以承受，而由于隐私法规和机构信息孤岛，即使是共享未标注数据也受到严重限制。

为减少对标注数据的依赖，自监督学习（*SSL*）已成为主流范式。采用掩码图像建模 [13, 5, 18] 或对比学习 [17, 19] 的方法，利用大量未标注的扫描图像来学习鲁棒表征。然而，自监督学习存在两个固有局限。首先，它无法消除数据整理在实际操作和法律层面的障碍，因此模型仍需要访问大规模领域内医学档案。其次，标准的自监督学习目标（如强度重建）主要关注局部特征，往往无法提供明确的监督。相比之下，公式驱动监督学习（FDSL） [7, 10, 12, 11, 9, 6], 提供了一种符合隐私要求且可无限扩展的解决方案。通过依据数学规则以程序化方式生成图像—标签对，FDSL 能够在不接触任何真实患者数据的情况下初始化分割模型。在医学领域，已有工作（如 PrimGeoSeg） [11]通过随机排列基本几何对象（例如多边形、圆柱体）来构建三维体积。然而，这些数学几何体与人体解剖结构之间仍存在显著的语义差距。现有 FDSL 方法依赖随机放置和简化几何形状，无法呈现医学扫描图像固有的**形态保真度和固定拓扑约束**。例如，在 *FDSL* 生成的体积中，圆柱体可能被随机放置在一个“类似肺部”的空间上方，导致其空间配置违背人体生理结构的拓扑规律。因此，在这类混乱的合成数据上预训练的模型或许能掌握低级边缘检测，却无法习得关键的**解剖先验**，从而难以区分对比度较低的软组织，而这类任务中空间上下文至关重要。

为应对这些局限，我们提出一种混合范式，将FDSL的无限可扩展性与真实数据的生物学有效性相结合。我们提出**解剖学引导的合成监督预训练框架**。与依赖纹理重建的自监督学习方法或局限于几何基元的通用FDSL方法不同，我们的范式利用由生成式组合过程得到的密集解剖监督。具体而言，我们通过从仅包含几何掩码的最小受试者集合（$K = 5$）构建轻量级形状库，并丢弃全部患者特异性纹理信息，从而引入**结构感知合成策略** 。我们并非随机放置，而是注入显式解剖先验，包括按类别的空间锚点与拓扑关系图，以约束生成过程。这促使Transformer从头学习有效的全局解剖布局与器官间依赖关系，从而提供完全保护隐私的强像素级监督。我们系统地评估了本框架在BTCV [8] 和

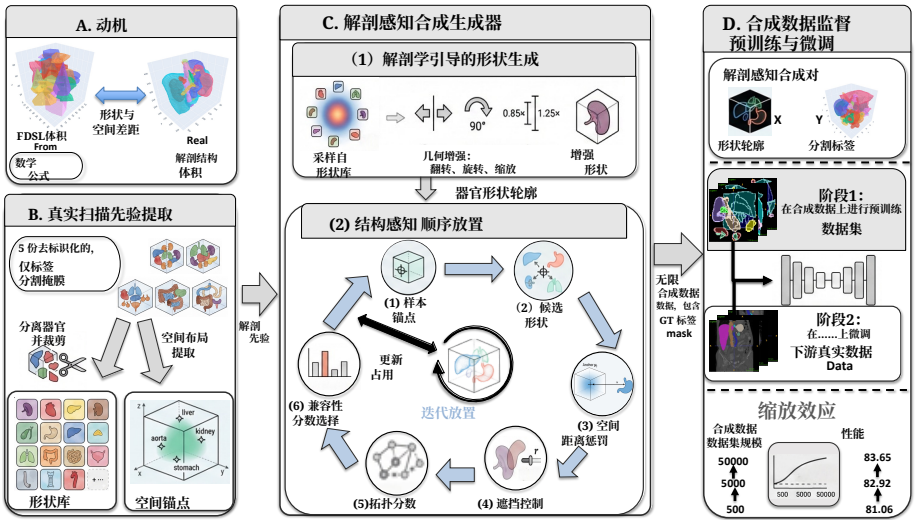

图1：总体框架。通过从极少的真实掩膜中提取形态与空间先验，我们的生成器迭代合成无限的生物学合理数据，以弥合FDSL的语义差距，从而实现稳健的模型预训练与下游微调。

MSD [1] 数据集上，基于UNETR [4]和SwinUNETR [3]骨干网络，性能超过最先进的公式驱动监督学习基线方法和自监督学习方法1.45%以上，最高达1.67%。

总之，我们提出一种新颖的预训练范式，可在无需真实数据的情况下提供密集的任务对齐监督。通过捕捉标准FDSL无法获得的全局结构依赖，并展现持续的缩放效应，我们的方法为数据高效医学人工智能建立了一条稳健且符合隐私要求的路径。

2 方法

图 1 展示了解剖信息引导的合成监督预训练框架，该框架无需依赖真实患者纹理，即可生成符合生物学规律的三维图像-标签对。该流程包含两个核心组件：提供逼真形态基元的解剖信息引导形状库（第 2.2 节），以及受空间与拓扑约束支配的结构感知顺序放置机制（第 2.3 节）。通过将这一生成过程以约束空间点过程进行数学建模（第 2.1 节），我们的方法能够合成多样且密集标注的体积数据，为分割模型预训练提供稳健的结构先验。

2.1 理论 formulation

合成预训练的根本目标是学习一个分割函数 f_θ，使其在无法访问的真实医学分布 $\mathcal{D}_{real}$ 上的期望风险最小化。公式驱动监督学习（FDSL）通过合成分布 $\mathcal{D}_{syn}$ 对此进行近似：从 $P(\mathbf{x},\mathbf{y})$ 分布中采样，生成带标签体积对 $(\mathbf{x},\mathbf{y})$，其中 $P(\mathbf{x},\mathbf{y}) = P(\mathbf{x}|\mathbf{y})P(\mathbf{y})$。此处，标签 $\mathbf{y}$ 由形状 $\mathcal{S}$和空间变换 $\mathcal{T}$构成，并严格表示为体素坐标集合。以往方法假设各形状结构在统计上相互独立，将先验分解为 $P(\mathbf{y}) = \prod P(S_i)P(T_i)$，并采用均匀的 T_i。这种不受约束的随机性会导致 $\mathcal{D}_{syn}$ 与 $\mathcal{D}_{real}$之间出现分布失配。

为弥合这一差距，我们提出一种**解剖学引导合成** 框架，将 $P(\mathbf{y})$ 重新定义为一种约束空间点过程，其服从以解剖关系图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 为条件的吉布斯分布：

$$P(\mathbf{y}|\mathcal{G}) = \frac{1}{Z} \exp\left(-\sum_{i \in \mathcal{V}} \psi_{unary}(S_i, T_i) - \sum_{(i,j) \in \mathcal{E}} \psi_{binary}(T_i, T_j) \right)$$

该先验以符合人体生理学为基础，同时保留足够的方差。两个核心能量分量实现了这一点：
一元势 $\psi_{unary}(S_i, T_i)$**：** 编码联合形状位置先验。形状项以非参数先验取代通用基元，该先验来自轻量级形状库 $\mathcal{B}$（第 2.2 节）。位置项则以基于锚点的惩罚项 $\psi_{unary}^{pos}(T_i) \propto \|\mathbf{t}_i - \mathbf{a}_i\|_2$ 取代均匀放置，其中锚点 $\mathbf{a}_i$ 从群体统计数据 $\mathcal{N}(\mu_i, \Sigma_i)$ 中采样。**二元势** $\psi_{binary}(T_i, T_j)$**：** 编码器官间拓扑依赖关系。它会惩罚物理上不可能的重叠，并奖励合理的边界接触，从而将随机放置转化为具有结构感知能力的组合。

由于在 $P(\mathbf{y}|\mathcal{G})$ 下的精确推断因未归一化的 Z 而难以处理，我们通过序贯贪心策略近似联合分布：$P(\mathbf{y}|\mathcal{G}) \approx \prod_{t=1}^{|\mathcal{V}|} P(T_t|Y_{t-1}, \mathcal{G})$，其中 Y_{t-1} 表示先前放置器官的体素集合。这明确地将我们的理论先验与第 2.3 节详述的可处理候选项排序生成过程衔接起来。最后，图像 $\mathbf{x} \sim P(\mathbf{x}|\mathbf{y})$ 被渲染为轮廓壳层而非实心体积，迫使编码器捕获对纹理不变的结构边界，而监督标签仍为稠密填充的体积掩膜。

2.2 解剖学引导的形状生成

标准的 FDSL 方法依赖于过于简化的几何基元，因而无法捕捉人体解剖结构复杂的边界曲率。为了注入真实的形态学先验，我们构建了一个 形状库 $\mathcal{B}$，它源自 TotalSegmentator [16], 中的少量真实受试者（K=5），包含跨越不同类别的 32 个解剖类别。

为确保形状干净且可复用，我们进行连通分量分析，以分离出各个器官实例，随后进行边界框

裁剪，并留出最小边距。至关重要的是，我们将原始离散类别索引重新映射到连续范围 $\{1,\dots,C\}$，确保为分割头提供紧凑的标签空间。在合成过程中，我们采用大幅度的几何增强，包括随机翻转、 90° 旋转和$[0.85,1.25]$范围内的各向同性缩放，以扩展 $\mathcal{B}$ 的有效多样性。这确保模型学到稳健的类别无关边界特征，而不是记住特定患者模板。

2.3 结构感知顺序放置

为高效解决放置问题，同时避免代价高昂的梯度优化，我们采用结合候选排序机制的蒙特卡洛采样方法。器官 c 的放置由 解剖锚点分布 $\mathcal{A}_c \sim \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ 引导，其中 $\boldsymbol{\mu}_c$ 和 $\boldsymbol{\Sigma}_c$ 分别表示从训练语料中得出的归一化均值质心和空间方差。

对于每个实例，我们首先采样一个特定的锚点 $\mathbf{a} \sim \mathcal{A}_c$。然后，我们通过围绕 $\mathbf{a}$ 采样随机扰动，生成一组 N 候选姿态 $\Pi = \{\pi_j\}_{j=1}^{N}$。通过最大化离散评分函数 $S(\pi)$，选出最优姿态 π^*；该函数综合了空间、物理和拓扑保真度：

$$\pi^* = \arg\max_{\pi_j \in \Pi} \left[S_{spatial}(\pi_j, \mathbf{a}) + S_{phys}(\pi_j, Y_{t-1}) + S_{topo}(\pi_j, Y_{t-1}, \mathcal{G}) \right]. \quad (1)$$

空间保真度 项 $S_{spatial}$ 确保生成的器官靠近采样得到的解剖锚点。其定义为负欧氏距离 $S_{spatial} = -\lambda_{anc}\|\mathbf{t}_\pi - \mathbf{a}\|_2$，其中 $\mathbf{t}_\pi$ 是候选掩膜的质心。该项起到随机拴绳的作用，在确保整体解剖正确性的同时，允许自然变化。

物理约束 项 S_{phys} 会对与当前已占用体积 Y_{t-1}之间不自然的重叠进行惩罚。我们将其定义为 $S_{phys} = -\lambda_{ovl} \cdot \mathrm{IoU}(\mathbf{m}^\pi, Y_{t-1})$，其中 IoU 表示交并比。此外，我们对排除集 $\mathcal{E}_{avoid}$中定义的生物学不相容对（例如骨骼与内脏）施加硬约束。如果对于任意 $k \in \mathcal{E}_{avoid}$，$\mathrm{IoU}(\mathbf{m}^\pi, Y_{(k)prev}) > \tau_{hard}$，则通过设置 $S(\pi) = -\infty$立即拒绝该候选项。

为模拟器官间交互，我们引入一个由关系图 $\mathcal{G}$控制的**拓扑分数** 项S_{topo}。该项通过指示函数鼓励特定的几何关系。对于包含关系（例如，气管位于肺内），当候选项的包含率超过阈值τ_{in}时，我们给予奖励。同样，对于邻接关系（例如，肝脏与主动脉接触），当候选项具有足够数量的边界接触体素时，我们给予奖励：

$$S_{topo} = \sum_{(c,k)\in\mathcal{E}_{in}} \lambda_{in}\mathbb{I}\left(\frac{|\mathbf{m}^\pi \cap Y^{(k)}|}{|\mathbf{m}^\pi|} > \tau_{in}\right) + \sum_{(c,k)\in\mathcal{E}_{adj}} \lambda_{adj}\mathbb{I}\left(|\partial\mathbf{m}^\pi \cap Y^{(k)}| > \nu_{contact}\right). \quad (2)$$

通过最大化这一综合分数，生成器会选出最能平衡随机变化与解剖合理性的候选结果。最终体积

表1：BTCV 数据集上的分割性能比较。最佳结果以粗体突出显示。

预训练	PT 数量	平均值	脾脏	右肾	左肾	胆囊	Eso	Liv	Sto	Aor	下腔静脉	静脉	胰腺	rad	lad
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 [11]	5K	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	**80.64**													
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 [11]	5K	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	**81.53**													

通过按体积从大到小依次叠加所选掩膜构建，从而模拟医学影像中固有的遮挡效应。

3 实验

3.1 实验设置

任务与数据集：我们在计算机断层扫描和磁共振成像两种成像模态上进行评估，以考察跨模态泛化。对于计算机断层扫描，我们使用 BTCV [8]（多器官）以及 MSD [1] Task06（肺）和 Task09（脾脏）。对于磁共振成像，我们在 MSD Task02（心脏）上进行评估。这一磁共振成像评估尤为重要，因为它检验了从合成 CT 预训练中学习到的解剖先验能否在没有显式 MRI 预训练数据的情况下迁移到不同的成像模态。所有数据集均按 80/20 划分为训练集/验证集，实验均离线进行。**数据合成设置：**我们构建形状库 $\mathcal{B}$，数据来自 $K{=}5$ TotalSegmentator 的 [16] 名受试者，并采用几何增强（翻转 $p{=}0.5$、90° 旋转、缩放$[0.85, 1.25]$）。结构感知放置对每个器官使用 $N_c{=}40$ 个候选，阈值分别为 $r{=}0.7$、$\sigma{=}0.12$、$\tau{=}0.35$，最小接触为 $v_{\min}{=}20$，包含比为 0.30。兼容性权重为 $w_o{=}-1.0$、$w_a{=}0.8$、$w_c{=}w_t{=}1.0$。

训练细节：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 4、学习率 0.0002 和权重衰减 0.00001。模型使用 AdamW 和预热余弦学习率调度器进行优化。预训练数据集包含 $5K$ 个样本，训练迭代次数设为 25000。在 BTCV 和 MSD 数据集上进行微调时，我们遵循先前工作 [11]中采用的标准超参数设置，批量大小为 4。所有下游实验均使用 Dice 相似系数进行评估。

3.2 结果与分析

BTCV 多器官分割。如表 1 所示，在 BTCV [8]数据集上，使用 UNETR 时，我们的方法取得了 80.64% 的平均 Dice 系数，比从头训练基线高 4.78%，比 PrimGeoSeg 高 1.74%。在弱边界结构（例如胆囊 $+11.32$%、胃 $+7.61$%）上取得的显著提升表明，我们的解剖先验能够有效弥补模糊的

表2：不同预训练策略在MSD数据集上的分割性能比较。报告了 UNETR 和 Swi-nUNETR 两种架构的结果。

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	**96.02**	**95.93**	**80.47**	**81.45**	96.20	**97.38**

表3：与在真实医学数据集上进行预训练的自监督学习（SSL）方法的比较。所有模型均采用SwinUNETR骨干网络，并在BTCV上进行微调。（注：由于微调期间采用了前景裁剪，数值与表1略有不同，但相对提升保持一致。）

方法	预训练数据	数据量	平均 Dice 系数
从头训练	None	0	80.12
SwinMM [15]	真实（CT）	5,000	76.72
SwinUNETR [14]	真实（CT）	5,000	80.56
PrimGeoSeg	合成数据	5,000	80.08
Ours	**合成数据**	**5,000**	**81.51**

外观线索。对于 SwinUNETR，我们的方法取得了最佳整体性能（81.53%），在复杂的血管与腺体器官上准确率更高。值得注意的是，PrimGeoSeg 在此处未能超过从头训练基线（80.08% 对比 80.12%），这表明通用基元缺乏初始化复杂层次化Transformer所需的结构复杂性。

MSD基准泛化。 我们的方法还展现出稳健的跨数据集和跨模态可迁移性，并在 MSD [1]数据集上得到了验证。在 Task06 （肺）上，我们的方法始终优于从头训练基线和最先进的 FDSL 方法（PrimGeoSeg）。具体而言，对于 UNETR，我们的表现较从头训练提升了 9.79%，较 PrimGeoSeg 提升了 1.66 %。同样，SwinUNETR 的表现分别提升了 5.31% 和 1.06%。至关重要的是，尽管合成预训练数据完全源自计算机断层扫描，我们的方法仍能有效泛化至磁共振成像（Task02 心脏）。我们的方法取得了 96.02%（UNETR）和 95.93%（SwinUNETR）的成绩，创下新的最先进水平，验证了通过结构感知合成所学习到的空间关系和拓扑约束具有固有的模态不变性。

缩放效应。 我们评估了合成数据量（500、5,000 和 50,000）对 UNETR 下游 BTCV 性能的影响（图 2）。我们的方法始终优于从头训练基线（80.02%），分别达到 81.06%、

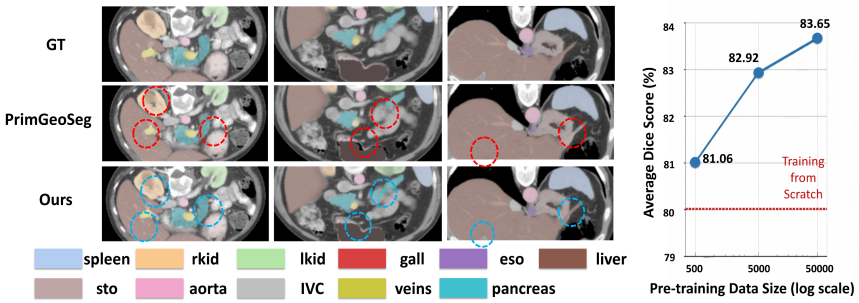

图2：**左：**定性比较结果。红色虚线表示误识别区域，蓝色虚线表示识别更准确的区域。**右：**将生成数据数量从500控制到50000时的缩放效应。

82.92% 和 83.65% 的平均 Dice 系数。随着数据规模扩大，边际收益逐渐降低（从 500 增至 5,000 时为 $+1.86\%$，而从 $5,000$ 增至 $50,000$ 时为 $+0.73\%$），这表明 5,000 个样本在计算成本与准确率之间实现了切实可行的权衡。

与自监督学习方法的比较。 我们将本方法与在真实医学数据上预训练的最先进自监督学习方法进行比较（表3）。在BTCV上使用SwinUNETR时，我们的解剖信息引导合成预训练（81.51%）同时超越了从头训练基线和PrimGeoSeg。值得注意的是，它还超过了在5,000个真实CT体积数据上预训练的SwinUNETR [14]（80.56%）。相比之下，基于重建的自监督学习（如SwinMM [15]）表现不及从头训练基线（76.72%），这表明掩码重建目标在细微域偏移下可能难以捕获精确的语义特征。这些结果表明，来自结构化合成数据的密集像素级解剖监督非常有效，提供了一种符合隐私要求的替代方案，其表现优于在敏感真实扫描数据上的无监督表示学习。

4 结论

本文提出了一种解剖学信息引导的合成监督预训练框架，将公式驱动监督学习（FDSL）的无限可扩展性与医学扫描图像的生物学真实性相结合。通过采用结构感知组合策略，确保生理合理性并满足拓扑约束，我们的方法在多个计算机断层扫描和磁共振成像基准测试中，显著优于当前最先进的自监督方法和通用合成基线方法。尤为重要的是，我们的工作验证了一条极具前景的技术路线：注入明确解剖学逻辑的稠密像素级合成监督，能够有效替代对海量且涉及隐私的真实世界数据集的依赖。这一范式表明，对于医学预训练而言，结构先验比纹理

重建更为重要，并为数据受限场景下训练稳健的3D医学Transformer奠定了可扩展、数据高效且严格遵守隐私保护要求的基础。

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. Dosovitskiy, A.：一张图像胜过 16×16 个词：用于大规模图像识别的 Transformer。arXiv 预印本 arXiv:2010.11929 (2020) 3. 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 4. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.：UNETR：用于 3D 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) 5. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.：掩码自编码器是可扩展的视觉学习器 (2021)，https://arxiv.org/abs/2111.06377 6. Kataoka, H., Hayamizu, R., Yamada, R., Nakashima, K., Takashima, S., Zhang, X., Martinez-Noriega, E.J., Inoue, N., Yokota, R.：用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) 7. Kataoka, H., Matsumoto, A., Yamada, R., Satoh, Y., Yamagata, E., Inoue, N.：基于递归平铺模式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) 8. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.：MICCAI 颅腔外多图谱标注——研讨会与挑战赛。载于：MICCAI 颅腔外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) 9. Nakamura, R., Kataoka, H., Takashima, S., Noriega, E.J.M., Yokota, R., Inoue, N.：使用极少量合成图像预训练视觉 Transformer。载于：IEEE/CVF 国际计算机视觉会议论文集。第 20360–20369 页 (2023) 10. Shinoda, R., Hayamizu, R., Nakashima, K., Inoue, N., Yokota, R., Kataoka, H.：SegRCDB：通过公式驱动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) 11. Tadokoro, R., Yamada, R., Kataoka, H.：为 3D 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) 12. Takashima, S., Hayamizu, R., Inoue, N., Kataoka, H., Yokota, R.：视觉原子：利用正弦波预训练视觉 Transformer。载于：IEEE/CVF 计算机视觉与模式识别会议论文集。第 18579–18588 页 (2023)

13. 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。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) 14. 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。载于：IEEE/CVF 计算机视觉与模式识别会议论文集。第 20730–20740 页 (2022) 15. Wang, Y., Li, Z., Mei, J., Wei, Z., Liu, L., Wang, C., Sang, S., Yuille, A.L., Xie, C., Zhou, Y.：SwinMM：用于 3D 医学图像分割的 Swin Transformer 掩码多视图方法。载于：医学图像计算与计算机辅助干预国际会议论文集。第 486–496 页。施普林格 (2023) 16. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.：TotalSegmentator：对 CT 图像中 104 个解剖结构进行稳健分割。《放射学：人工智能》5(5) (2023年9月)。https://doi.org/10.1148/ryai.230024, http://dx.doi.org/10 230024 17 1148. /ryai. . Wu, L., Zhuang, J., Chen, H.：VoCo：一种简洁而有效的三维医学图像分析体积对比学习框架 (2024)，https://arxiv.org/abs/2402 17300 18 . . Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., Hu, H.：SimMIM：一种用于掩码图像建模的简洁框架 (2022)，https://arxiv.org/abs/2111 09886 19 . . Zhou, Z., Sodha, V., Siddiquee, M.M.R., Feng, R., Tajbakhsh, N., Gotway, M.B., Liang, J.：Models Genesis：用于三维医学图像分析的通用自学习模型 (2019)，https://arxiv.org/abs/1908.06912