半监督从 2D 自然图像预训练模型进行 3D 医学分割

Pak-Hei Yeung[1,2], Jayroop Ramesh[2], Pengfei Lyu[3], Ana Namburete[2], and Jagath Rajapakse[1]

[1] College of Computing and Data Science, Nanyang Technological University, Singapore
pakhei.yeung@ntu.edu.sg
[2] Oxford Machine Learning in NeuroImaging Lab, University of Oxford, United Kingdom
[3] Faculty of Robot Science and Engineering, Northeastern University, China

项目页面

摘要 本文探讨了将基于 2D 自然图像预训练的通用视觉模型的知识转移到改进 3D 医学图像分割的技术。我们专注于半监督设置，其中只有少量带有标签的 3D 医学图像可用，并且有一大批未标记的图像。为了应对这一挑战，我们提出了一种模型不可知框架，该框架逐步将 2D 预训练模型的知识蒸馏到从零开始训练的 3D 分割模型中。我们的方法 **M&N** 涉及通过伪掩模进行两模型迭代协同训练，这些伪掩模由彼此生成，并结合了我们提出的自适应调整学习率引导采样技术，以根据模型的预测准确性和稳定性在每个训练批次中动态调整有标签和无标签数据的比例，从而尽量减少不准确的伪掩码带来的负面影响。在多个公开可用的数据集上的广泛实验表明，M&N 实现了最先进的性能，在所有不同设置下均优于现有的十三种半监督分割方法。重要的是，消融研究表明 保持了模型不可知性，允许无缝集成不同的架构。这确保了其在更先进的模型出现时的适应能力。代码可在 https://github.com/pakheiyeung/M-N 处获取。

Keywords: 知识蒸馏 · 半监督分割 · 域适应。

1 介绍

深度学习的出现显著提升了三维医学图像分割的表现，这无疑是医学图像分析中最重要的任务之一。然而，从零开始训练一个深度学习模型通常需要大量的标注数据，在医学领域这是一个主要瓶颈 [28]。相比之下，其他领域的标注数据较为丰富，例如 2D 自然图像，在这些领域已经整理和利用了

大量数据集 [5,31] 来训练强大的视觉模型 [25]。这些预训练的模型在各种计算机视觉任务中展示了显著的能力。受到这些模型成功的启发，本文探讨了利用它们的知识来促进三维医学图像分割的可能性，特别是在只有少量手动标注的情况下。

具体来说，我们关注的是半监督的 3D 医学图像分割任务，其中只有少量标注的 3D 医学图像可用，同时还有一大批未标注的图像。该领域的最新进展通过各种利用未标注数据的策略实现了显著的性能，例如教师-学生框架 [2,30]、基于不确定性的方法 [16,20]、无监督领域自适应 [14] 以及基于原型和对比学习的框架 [19,23]。与本文相关的另一条研究路线探索了在 2D 和 3D 网络之间弥合差距以进行 3D 医学图像分析的方法，主要通过网络架构设计 [4,11,27] 实现。虽然已经实现了显著性能，但大多数这些方法都是为特定类型的网络量身定制的。相比之下，本工作提出了一种模型不可知框架，该框架能够使从任何 2D 预训练网络向任何 3D 分割网络的知识迁移，旨在提供一种更为灵活和通用的解决方案。

这项工作是由我们的初步发现所驱动的，这些发现表明使用预先在二维自然图像上训练的模型比从头开始为三维医学分割任务训练相同的网络表现更好。当标记的训练数据数量有限时，这种性能差距会进一步扩大。这些发现在 Tables 1 to 3 的前 3 行中有所展示。这表明，在自然图像上的预训练获取了可以转移到三维医学分割的知识，特别是在数据量较低的情况下。基于基于三维网络 [8,15] 在大型标记数据集上训练时实现了最先进的医疗分割性能的成功，我们提出了一个基本问题：我们能否利用二维预训练模型的知识来提高三维分割模型的性能，即使是在有限标记样本的训练中？

为了解决这个问题，我们提出了 M&N ，这是一个模型不可知的框架，该框架从预先在二维自然图像上训练好的视觉模型中提取知识，并将其应用到从头开始训练的三维模型中用于半监督医疗分割。我们的工作做出了以下贡献：首先，我们提出了一种迭代协同训练策略，其中 2D 和 3D 模型使用彼此生成的伪掩码进行训练。为了减轻不准确的伪掩码的影响，我们进一步提出了学习率引导采样，该方法自适应地调整一个批次中标注数据和未标注数据的比例以符合模型的预测准确性和稳定性。作为我们的第二贡献，我们在不同有限数据设置的各种公开可用的数据集上对 M&N 进行了基准测试。M&N 在所有实验中超越了现有的 13 种半监督分割方法，达到了最先进的性能。第三位，我们的消融研究表明 M&N 对不同的模型和架构是不可知的。这表明其通用性和与先进模型无缝集成以在未来实现更出色结果的潜力。

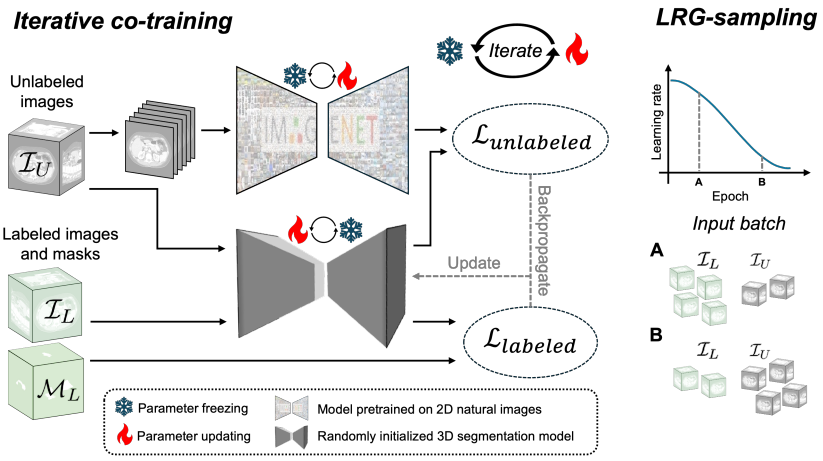

图 1. 我们提出的 M&N 框架的流程。二维和三维模型通过彼此生成的伪掩码进行迭代协同训练，并使用未标记损失（$\mathcal{L}_{unlabled}$），以及带有标签图像和掩码的标注损失（$\mathcal{L}_{labled}$）。此迭代过程在奇数和偶数周期之间交替进行。LRG-采样根据当前学习率动态调整每个批次中标记数据与未标记数据的比例，优化可用训练数据的利用。

2 方法

我们提出 M&N 用于半监督的 3D 医学图像分割。给定一个三维医学图像数据集，$\mathcal{I} = \{\boldsymbol{I}_i\}_{i=1}^{m}$，其中每个图像 $\boldsymbol{I}_i \in \mathbb{R}^{C_i \times H \times W \times D}$ 有 C_i 个通道，高度 H，宽度 W 和深度 D。我们假设有一组图像子集，$\mathcal{I}_L$，具有相应的标记掩码，$\mathcal{M}_L = \{\boldsymbol{M}_i\}_{i=1}^{n}, \boldsymbol{M}_i \in \mathbb{R}^{C_c \times H \times W \times D}$，包含 C_c 类，其中 $n \ll m$。剩余的图像，$\mathcal{I}_U$，是未标记的。

使用 $\mathcal{I}_L$、$\mathcal{M}_L$ 和 $\mathcal{I}_U$，我们的目标是从一个预训练的视觉模型 $f(\cdot;\theta_{nat})$ 中提取知识，该模型由 θ_{nat} 参数化并在 2D 自然图像上进行预训练，将其知识转移到一个 3D 分割模型 $g(\cdot;\theta_{med})$ 上，该模型由 θ_{med} 参数化。

2.1 在标记图像上的微调

我们首先对预训练的 2D 模型 $f(\cdot;\theta_{nat})$ 进行微调，通过沿深度维度提取 2D 切片 D，在标记的数据集 $\{\boldsymbol{I}_i, \boldsymbol{M}_i\}_{i=1}^{n}$ 上进行。同时，我们从头开始在 $\mathcal{I}_L$ 和 $\mathcal{M}_L$ 上训练三维分割模型 $g(\cdot;\theta_{med})$。在此阶段，两个模型使用定义为：

$$\mathcal{L}_l = w_{ce} \cdot \mathcal{L}_{ce}\left(\hat{\boldsymbol{M}}, \boldsymbol{M}\right) + w_{dice} \cdot \mathcal{L}_{dice}\left(\hat{\boldsymbol{M}}, \boldsymbol{M}\right), \tag{1}$$

的标记损失 $\mathcal{L}_l$ 独立进行优化，其中 $\hat{\boldsymbol{M}}$ 是预测掩码，$\mathcal{L}_{ce}$ 是交叉熵损失，$\mathcal{L}_{dice}$ 是软 Dice 损失，而 w_{ce} 和 w_{dice} 分别是它们的权重。

预训练模型。M&N 是模型不可知的，允许使用具有不同学习目标预训练的各种二维视觉模型，$f(\cdot;\theta_{nat})$。对于采用编码器-解码器架构的 $f(\cdot;\theta_{nat})$，如 [25]，可以通过简单地替换最后一层以匹配类别数量 C_c 来进行微调。或者，需要附加一个解码器并对模型进行微调，针对只有预训练编码器的模型，例如 [9]。这两种情况都在 Section 3.3 中进行了评估。

微调策略。预训练的 $f(\cdot;\theta_{nat})$ 可以采用不同的策略进行微调，从更新所有权重，θ_{nat}，到只对部分层进行微调而保持其他层不变。我们采用低秩适应（LoRA）[10] 作为 M&N 的默认微调策略，但也调查了 Section 3.3 中的其他选项以提供全面比较。

2.2 迭代协同训练

两个模型，$f(\cdot;\theta_{nat})$ 和 $g(\cdot;\theta_{med})$，都在标记子集 $\mathcal{I}_L$ 和 $\mathcal{M}_L$ 以及未标记子集 $\mathcal{I}_U$ 上进行了训练，如 Fig. 1 所示。

奇数-数 epochs。二维切片，$\{\boldsymbol{S}_i^d\}_{i=1,d=1}^{n,D}$，沿着深度维度 D 从 $\mathcal{I}_U=\{\boldsymbol{I}_i\}_{i=1}^n$ 中提取，其中 $\boldsymbol{S}_i^d\in\mathbb{R}^{C_i\times H\times W}$。这些切片被输入到 $f(\cdot;\theta_{nat})$ 中生成伪掩码，$\{\boldsymbol{P}_i\}_{i=1}^n, \boldsymbol{P}_i\in\mathbb{R}^{C_c\times H\times W\times D}$:

$$\boldsymbol{P}_i = concat\left(f(\boldsymbol{S}_i^1;\theta_{nat}),\ f(\boldsymbol{S}_i^2;\theta_{nat}),\ ...,\ f(\boldsymbol{S}_i^D;\theta_{nat})\right), \tag{2}$$

其中 $concat(\cdot)$ 沿深度维度 D 连接 2D 预测的掩码。与由 $g(\cdot;\theta_{med})$ 预测的掩码一起:

$$[\hat{\boldsymbol{M}}_1,\ \hat{\boldsymbol{M}}_2, ...,\ \hat{\boldsymbol{M}}_n] = [g(\boldsymbol{I}_1;\theta_{med}),\ g(\boldsymbol{I}_2;\theta_{med}),\ ...,\ g(\boldsymbol{I}_n;\theta_{med})], \tag{3}$$

无标签损失 $\mathcal{L}_u$ 可以计算为:

$$\mathcal{L}_u = w_{kl}\cdot\mathcal{L}_{kl}\left(\hat{\boldsymbol{M}},\boldsymbol{P}\right) + w_{dice}\cdot\mathcal{L}_{dice}\left(\hat{\boldsymbol{M}},\boldsymbol{P}\right), \tag{4}$$

其中 $\mathcal{L}_{kl}$ 是 Kullback-Leibler 散度损失，w_{kl} 是其权重。

仅用 $\mathcal{L}_u$ 进行监督可能导致解决方案崩溃，例如，无论输入是什么都输出相同的预测。因此，计算有标签的损失，$\mathcal{L}_l$(Eq. (1))，这导致最终的协同

训练损失，$\mathcal{L}_c$：

$$\mathcal{L}_c = \frac{b_l}{b_l + b_u} \cdot \mathcal{L}_l + \frac{b_u}{b_l + b_u} \cdot \mathcal{L}_u, \tag{5}$$

其中 b_l 和 b_u 是一批中有标签和无标签数据的数量。

在每次反向传播步骤中，计算一个随机优化以最小化 $\mathcal{L}_c$，相对于 θ_{med} 来训练 $g(\cdot;\theta_{med})$：

$$\theta_{med} \leftarrow optim(\theta_{med}, \nabla_{\theta_{med}} \mathcal{L}_c, \eta_{med}), \tag{6}$$

其中 $optim(\cdot)$ 表示优化器，η_{med} 是 $g(\cdot;\theta_{med})$ 的学习率。

即使-数量周期。伪掩码，$\{\boldsymbol{P}_i\}_{i=1}^n$，由 $g(\cdot;\theta_{med})$ 生成：

$$[\boldsymbol{P}_1,\ \boldsymbol{P}_2, ...,\ \boldsymbol{P}_n] = [g(\boldsymbol{I}_1;\theta_{med}),\ g(\boldsymbol{I}_2;\theta_{med}),\ ...,\ g(\boldsymbol{I}_n;\theta_{med})] \tag{7}$$

同时，$\{\hat{\boldsymbol{M}}_i\}_{i=1}^n$ 由 $f(\cdot;\theta_{nat})$ 从输入的 2D 切片，$\{\boldsymbol{S}_i^d\}_{i=1,d=1}^{n,D}$ 输出：

$$\hat{\boldsymbol{M}}_i = concat\left(f(\boldsymbol{S}_i^1;\theta_{nat}),\ f(\boldsymbol{S}_i^2;\theta_{nat}),\ ...,\ f(\boldsymbol{S}_i^D;\theta_{nat})\right) \tag{8}$$

通过最小化共训练损失，$\mathcal{L}_c$（Eq. (5)），相对于 θ_{nat}（或根据在 Section 2.1 中描述的微调策略依赖于 θ_{nat} 的一个子集），$f(\cdot;\theta_{nat})$ 的训练可以总结为：

$$\theta_{nat} \leftarrow optim(\theta_{nat}, \nabla_{\theta_{nat}} \mathcal{L}_c, \eta_{nat}) \tag{9}$$

通过在 Eqs. (6) and (9) 之间迭代，两个模型，$f(\cdot;\theta_{nat})$ 和 $g(\cdot;\theta_{med})$，通过持续地从彼此学习来提高它们的表现。

2.3 学习率引导采样（LRG-采样）

由于标记集 $\mathcal{I}_L$ 和未标记集 $\mathcal{I}_U$ 之间的规模存在显著不平衡，在协同训练过程中从 $\mathcal{I}$ 中进行均匀数据采样可能会损害训练稳定性和最终性能。即使采用过采样的策略，固定的数据采样方法可能仍然不是最优的。

理想情况下，批处理中标签图像和无标签图像的比例应适应模型的当前状态。在早期阶段，当模型的预测仍然不稳定且不准确时，过分依赖其生成的伪掩码可能是次优的。随着训练的进行，当预测变得更加稳定和准确时，批次中的无标签图像用作伪掩码的数量应该相应增加以最大化未标记数据的利用。

此模式与学习率衰减一致，这是一种广泛用于训练深度模型的技术 [3]。因此，我们提出了由学习率引导的采样（LRG-采样）方法，该方法根据当前

每个时期的 learning rate,$\eta_{current}$，自适应地调整 b_l 和 b_u（Eq. (5)）。这可以表示为：

$$b_u = \left\lfloor \frac{\eta_{initial} - \eta_{current}}{\eta_{initial} - \eta_{final}} \cdot B \right\rfloor \tag{10}$$

$$b_l = B - b_u \tag{11}$$

其中 $\lfloor \rfloor$ 表示取整函数，B 是批量大小，以及 $\eta_{initial}$ 和 η_{final} 是调度的初始和最终学习率。

3 实验与结果

3.1 实验设置

我们对 M&N 进行了全面评估，将其与 13 种最先进的方法在不同设置下进行了比较（即。使用不同数量的标注数据进行训练），并通过消融研究调查了 M&N 中不同组件的影响。

数据集。我们在左心房（LA）腔数据集 [26] 和胰腺-CT 数据集 [18] 上进行了基准测试。**LA 数据集**包含 3D 钆增强心脏 MR 图像，带有 LA 腔的真实掩码。分辨率为 $0.625 \times 0.625 \times 0.625mm^3$。我们将每个图像的体素值归一化到范围 $[0, 1]$，然后进行标准化。训练集包含 100 个图像，测试集包含 56 个。**胰腺-CT 数据集**包含 82 张（62 张训练和 20 张测试）腹部对比增强 3D CT 图像，带有胰腺的地面真实掩模。我们将图像重采样到统一的分辨率 $0.85 \times 0.85 \times 0.75mm^3$。体素值被裁剪到范围 $[-175, 250]$Hounsfield Units，然后归一化为 $[0, 1]$。

实现细节。M&N 使用了 SegFormer-B2[25]，在 Imagenet-1K[5] 和 ADE20K[31] 上预训练，作为二维模型 $f(\cdot; \theta_{nat})$，以及一个随机初始化的 3D UNet[17] 作为三维分割网络 $g(\cdot; \theta_{med})$。其他 2D 和 3D 模型，包括预训练的 ResNet-50 编码器 UNet[9,17] 和 SwinUNETR[8]，也在消融研究中进行了评估。超参数设置为：$w_{ce} = w_{kl} = w_{dice} = 1$ 和 $B = 5$。在训练过程中，我们应用了一系列数据增强技术，包括随机缩放，比例为 $0.9 - 1.1$，gamma 对比度调整值为 $\gamma \in [0.8, 1.2]$，以及以 50% 的概率随机裁剪至 $160 \times 160 \times 64$，并包含前景。由 3D 模型预测的伪掩码用于确定未标记图像的前景区域。两个模型首先使用带标签的图像训练和微调了 500 个周期（包括 50 个预热周期），然后

共同训练了另外 3500 个周期，使用带有 $\eta_{initial} = 10^{-3}$ 的权重衰减余弦调度器对 $f(\cdot;\theta_{nat})$ 以及 $\eta_{initial} = 10^{-4}$ 对 $g(\cdot;\theta_{med})$ 和 $\eta_{final} = 0$。优化器采用了 AdamW[13]。

推理。对于每个体积，我们采样了尺寸为 $160 \times 160 \times 64$ 的块，与训练期间使用的维度一致，以确保对整个体积进行完整覆盖。相邻块以 50%的重叠进行采样，并对重叠区域内的预测结果取平均以获得最终输出。推理过程中未应用任何数据增强。

评估指标。四个度量标准，即 Dice 系数、Jaccard 指数、95% 豪斯多夫距离（HD_{95}）和平均表面距离（ASD），被用于评估。对于 Dice 系数和 Jaccard 指数，较高的值表示性能更好，而对于 HD_{95} 和 ASD 则希望数值较低。

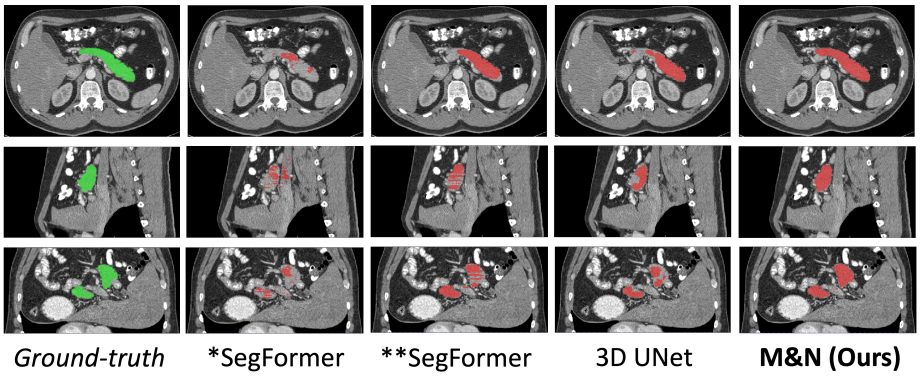

图 2. 胰腺-CT 数据集的定性结果 [18]。所有方法都使用相同数量（即 6）的标记图像进行训练。*SegFormer 是从零开始训练的，且**SegFormer 首先在 ADE20K[31] 上进行了预训练。

3.2 基线比较

我们在 LA 数据集上使用 4 张和 8 张标记图像训练了 M&N ，并与 13 种最先进的方法进行了比较，并在用 6 张标记图像训练的 Pancreas-CT 数据集上也进行了同样的比较。

如 Tables 1 to 3 所示，我们提出的 M&N 在不同的数据集和设置下始终优于所有现有方法（即。的结果见**粗体**）。在不同模式（即。MRI 和 CT）和

表 1. LA 结果（8 个标签）

方法	指标			
	骰子 (%)↑	Jaccard（百分比）↑	HD_{95} (体素)↓	ASD (体积)↓
3D U 网	88.26	79.89	11.19	3.26
*SegFormer	83.05	73.25	15.79	5.10
分段式转换器	88.61	79.87	10.49	3.35
SASSNet[12]	87.32	77.72	9.62	2.55
MC-Net[24]	87.71	78.31	9.36	2.18
SS-网络 [23]	88.55	79.62	7.49	1.9
MC-Net+[22]	88.96	80.25	7.93	1.86
相机 [6]	89.62	81.28	8.76	2.02
DK-UXNet[21]	90.41	82.69	7.32	1.71
使用-修改 [29]	84.25	73.48	13.84	3.36
LG-ER-MT[7]	85.54	75.12	13.29	3.77
杜姆 [20]	85.91	75.75	12.67	3.31
AD-MT[30]	90.55	82.79	5.81	1.7
BCP[2]	89.62	81.31	6.81	1.76
共生物联网 [16]	89.2	80.68	6.44	1.9
图 CL[19]	90.24	82.31	6.42	1.71
M&N 的	**91.56**	**84.47**	**4.59**	**1.40**

粗体表示顶级性能
↑ 意味着更高的值更准确
* 从头开始训练的 SegFormer
** 在 ADE20K 上预训练的 SegFormer [31]

表 2. LA 结果（4 个标签）

方法	指标			
	骰子	Jaccard 系数	HD_{95}	ASD
3D UNet	82.01	72.42	29.95	9.91
*SegFormer	71.01	58.4	14.34	7.07
**SegFormer	84.23	74.49	8.24	3.57
SS-Net[23]	86.33	76.15	9.97	2.31
CAML[6]	87.34	77.65	9.76	2.49
DK-UXNet[21]	85.96	75.91	11.72	2.64
AD-MT[30]	89.63	81.28	6.56	1.85
BCP[2]	88.02	78.72	7.9	2.15
GraphCL[19]	88.8	80	7.16	2.1
M&N	**90.47**	**82.66**	**4.72**	**1.59**

表 3. 胰腺-CT 结果（6 个标签）

方法	指标			
	骰子	Jaccard	HD_{95}	ASD
3D UNet	59.53	44.17	18.52	1.79
*SegFormer	43.76	28.32	19.67	6.86
**SegFormer	68.57	53.34	16.39	4.77
MC-Net[24]	68.94	54.74	16.28	3.16
MC-Net+[22]	74.01	60.02	12.59	3.34
AD-MT[30]	80.21	67.51	7.18	1.66
Co-BioNet[16]	77.89	64.79	8.81	**1.39**
M&N	**81.67**	**69.53**	**6.56**	1.67

目标结构（即。左心房和胰腺）上使用不同数量的标记训练数据进行测试，我们的结果展示了 M&N 的鲁棒性和通用性。

表格中前三行的结果，特别是 Table 3，和 Fig. 2，这些展示了更具挑战性的 CT 胰腺分割任务，验证了我们工作的动机。具体来说，基于 2D 自然图像的预训练显著帮助了医学分割（即 *SegFormer vs. **SegFormer），而仅用少量标注数据训练的 3D UNet 难以达到令人满意的结果。然而，它仍然超过了从零开始训练的 2D 网络（即。3D UNet vs. *SegFormer），证明了 M&N 的选择将预训练 2D 模型的知识蒸馏到 3D 模型中的合理性。

与现有的最佳方法 AD-MT[30] 相比，我们使用了 3D UNet，而 AD-MT 则使用 VNet[15] 作为分割模型。这两个模型具有非常相似的架构，进一步

证实 M&N 的更好性能主要归因于学习框架设计的不同，而非网络架构的差异。

我们关注的是标注数据较少的情况，其中可用的标注图像少于 10 张（即。4, 6 和 8）。这一阈值是根据我们之前与临床医生合作的经验故意选择的，当所需数量增加一个数量级时（例如 9 到 10 或 90 到 100），他们标注数据的积极性会显著下降。这种现象与心理学和市场营销概念一致，例如数字范畴感知和左数位效应 [1]。通过将我们的评估限制在少于 10 张标注图像上，我们的实验设置旨在反映注释资源稀缺的实际情况。

3.3 消融研究

我们研究了 M&N 中的不同组件对 LA 数据集的影响，使用 8 张标记图像进行训练，结果呈现在 Table 4 中。

模型不可知分析。虽然默认的预训练 SegFormer 具有编码器-解码器架构，我们将它替换为一个预训练的 ResNet-50[9] 编码器，并通过跳跃连接将其与随机初始化的解码器相连，称为 ResUNet。如 Table 4（第 1 行）所示，其性能略有下降，但仍优于现有最佳方法 AD-MT[30]（Table 1）。此外，我们将 3D UNet 替换为 SwinUNETR[8] 并观察到性能有类似的轻微下降（Table 4 第 2 行），但仍优于 AD-MT[30] 在大多数指标上。这些实验验证了 M&N 的模型无关特性，确认其能够适应各种架构同时保持优异性能。

微调策略。我们比较了预训练的 SegFormer 的三种微调策略，即 LoRA[10]、解码器微调和整个网络微调。虽然冻结编码器仅更新解码器导致性能下降（第 5 行），其他两种策略表现相当（第 6 行），其中 LoRA 在 4 个指标中的 3 个略胜一筹。由于 LoRA 参数效率更高，我们将其作为 M&N 的默认策略。

训练和数据采样。我们通过使用固定的伪掩码训练来消融迭代协同训练，其中预训练的 2D 模型首先在带有标签的图像（Section 2.1）上进行微调，然后用于生成伪掩码以训练 3D 模型而不做进一步更新。这导致了性能下降（第 3 行）。此外，用均匀采样数据替换 LRG-sampling 导致性能显著降低（第 4 行）。这些实验验证了我们在 M&N 中提出的组件的有效性。

表 4. M&N 在 LA 数据集（8 个标签）上的消融研究。

预训练的二维模型	微调战略*	3D 网络	迭代共同训练	大样本-采样	指标			
					骰子 (%)↑	Jaccard（%）↑	HD_{95} (体素)↓	ASD（体积）↓
ResUNet**	Whole	3D UNet	✓	✓	91.11	83.74	5.38	1.41
SegFormer	LoRA	SwinUNETR	✓	✓	90.67	83.03	5.24	2.16
SegFormer	LoRA	3D UNet	✗	✓	89.4	81.32	6.06	1.74
SegFormer	LoRA	3D UNet	✓	✗	87.14	77.35	8.43	2.36
SegFormer	Decoder	3D UNet	✓	✓	89.49	81.28	6.44	1.75
SegFormer	Whole	3D UNet	✓	✓	91.39	84.20	5.02	**1.38**
SegFormer	LoRA	3D UNet	✓	✓	**91.56**	**84.47**	**4.59**	1.40

* 微调策略指的是我们更新预训练的 2D 模型的哪一部分。

** 带有在 Imagenet-1K 上预训练的 ResNet-50 编码器的 UNet[5]

4 结论

总结，我们提出了 M&N ，一个无需特定模型的框架，用于将从 2D 自然图像预训练的一般视觉模型中的知识转移到增强半监督 3D 医学图像分割中。通过迭代地共同训练 2D 和 3D 模型，并在整个训练过程中自适应调整每批中标注和未标注图像的比例，M&N 在不同有限标注数据设置下的多个公开可用数据集上实现了最先进的性能，超越了 13 种现有方法。未来工作计划，我们将扩展 M&N 至医学图像分析中的其他任务，例如图像配准。我们的最终目标是利用丰富的跨域知识促进医学图像分析的发展。

致谢 叶勇由南洋理工大学的总统博士后奖学金资助。我们感谢 Madeleine Wyburd 博士和 Valentin Bacher 先生对这项工作提出的宝贵建议和评论。

参考文献

1. Anderson, E.T., Simester, D.I.: Effects of $9 price endings on retail sales: Evidence from field experiments. Quantitative Marketing and Economics **1**, 93–110 (2003)
2. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11514–11524 (2023)
3. Bengio, Y.: Practical recommendations for gradient-based training of deep architectures. In: Neural networks: Tricks of the trade: Second edition, pp. 437–478. Springer (2012)

4. Delchevalerie, V., Frénay, B., Mayer, A.: From three to two dimensions: 2d quaternion convolutions for 3d images. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) (2024)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 248–255. IEEE (2009)
6. Gao, S., Zhang, Z., Ma, J., Li, Z., Zhang, S.: Correlation-aware mutual learning for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 98–108. Springer (2023)
7. Hang, W., Feng, W., Liang, S., Yu, L., Wang, Q., Choi, K.S., Qin, J.: Local and global structure-aware entropy regularized mean teacher model for 3d left atrium segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 562–571. Springer (2020)
8. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop. pp. 272–284. Springer (2021)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
10. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (ICLR) (2022), https://openreview.net/forum?id=nZeVKeeFYf9
11. Jang, J., Hwang, D.: M3t: Three-dimensional medical image classifier using multi-plane and multi-slice transformer. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20718–20729 (2022)
12. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 552–561. Springer (2020)
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR) (2017), https://api.semanticscholar.org/CorpusID:53592270
14. Lyu, P., Yeung, P.H., Yu, X., Xia, J., Chi, J., Wu, C., Rajapakse, J.C.: Bridging the inter-domain gap through low-level features for cross-modal medical image segmentation. arXiv preprint arXiv:2505.11909 (2025)

15. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)

16. Peiris, H., Hayat, M., Chen, Z., Egan, G., Harandi, M.: Uncertainty-guided dual-views for semi-supervised volumetric medical image segmentation. Nature Machine Intelligence **5**(7), 724–738 (2023)

17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 234–241. Springer (2015)

18. Roth, H., Farag, A., Turkbey, E., Lu, L., Liu, J., Summers, R.: Data from pancreas-ct (version 2)[data set]. the cancer imaging archive (2016)

19. Wang, M., houcheng su, Li, J., Li, C., Yin, N., Shen, L., Guo, J.: GraphCL: Graph-based clustering for semi-supervised medical image segmentation. In: International Conference on Machine Learning (ICML) (2025), https://openreview.net/forum?id=Q2av1PZmfT

20. Wang, Y., Zhang, Y., Tian, J., Zhong, C., Shi, Z., Zhang, Y., He, Z.: Double-uncertainty weighted method for semi-supervised learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 542–551. Springer (2020)

21. Wu, R., Li, D., Zhang, C.: Semi-supervised medical image segmentation via query distribution consistency. In: IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)

22. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. Medical Image Analysis **81**, 102530 (2022)

23. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 34–43. Springer (2022)

24. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 297–306. Springer (2021)

25. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems (NeurIPS) **34**, 12077–12090 (2021)

26. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical Image Analysis **67**, 101832 (2021)

27. Yang, J., Huang, X., He, Y., Xu, J., Yang, C., Xu, G., Ni, B.: Reinventing 2d convolutions for 3d images. IEEE Journal of Biomedical and Health Informatics **25**(8), 3009–3018 (2021)

28. Yeung, P.H., Namburete, A.I., Xie, W.: Sli2vol: Annotate a 3d volume from a single slice with self-supervised learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 69–79. Springer (2021)

29. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 605–613. Springer (2019)

30. Zhao, Z., Wang, Z., Wang, L., Yu, D., Yuan, Y., Zhou, L.: Alternate diverse teaching for semi-supervised medical image segmentation. In: European Conference on Computer Vision (ECCV). pp. 227–243. Springer (2024)

31. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A.: Scene parsing through ade20k dataset. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 633–641 (2017)