

基于可微拓扑搜索与可信动量优化的半监督3D牙齿分割方法

林泽峰

广东工业大学数学与统计学院, 广东 广州

收稿日期: 2026年3月24日; 录用日期: 2026年4月22日; 发布日期: 2026年4月30日

摘要

本文提出DiNTS-TMO, 一种结合可微拓扑搜索与可信动量优化的半监督3D牙齿分割框架。针对现有NAS方法在半监督场景下易受伪标签噪声干扰的问题, DiNTS-TMO采用DiNTS可微拓扑搜索自动发现网络结构, 并引入可信动量优化(TMO)机制, 通过对齐门控筛选无标签梯度, 在搜索阶段同时约束网络权重与架构参数的更新。基于STS-Tooth数据集两个子集的五折交叉验证结果, DiNTS-TMO在ROI子集上取得了最优的DSC和HD95, 在Integrity子集上取得了最优的DSC, 但在边界距离指标上仍存在改进空间。实验结果表明, 该方法在低标注3D牙齿分割任务中具有较强的竞争力, 并为结合NAS与半监督学习提供了一种可行思路。

关键词

3D牙齿分割, 深度学习, 神经架构搜索, 半监督学习, 可信动量优化

Semi-Supervised 3D Tooth Segmentation via Differentiable Topology Search and Trusted Momentum Optimization

Zefeng Lin

School of Mathematics and Statistics, Guangdong University of Technology, Guangzhou Guangdong

Received: March 24, 2026; accepted: April 22, 2026; published: April 30, 2026

Abstract

We propose DiNTS-TMO, a semi-supervised 3D tooth segmentation framework that combines differentiable topology search with Trusted Momentum Optimization (TMO). To address the sensitivity

of NAS to noisy pseudo labels, DiNTS-TMO uses DiNTS to automatically discover network topologies and introduces an alignment-gated TMO mechanism to filter unlabeled gradients, jointly constraining the updates of network weights and architecture parameters during search. Five-fold cross-validation on two STS-Tooth subsets shows that DiNTS-TMO achieves the best DSC and HD95 on the ROI subset and the best DSC on the Integrity subset, while still leaving room for improvement on boundary-distance metrics in the full-scan setting. These results indicate that DiNTS-TMO is a competitive approach for low-label 3D tooth segmentation and provide a practical route to combining NAS with semi-supervised learning.

Keywords

3D Tooth Segmentation, Deep Learning, Neural Architecture Search, Semi-Supervised Learning, Trusted Momentum Optimization

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

3D 牙齿分割是计算机辅助牙科诊断、正畸分析与治疗规划中的关键环节。该任务中牙齿结构紧密邻接，易受金属伪影、牙槽骨干扰和扫描质量波动影响，导致边界模糊且前景占比极低；同时，三维像素级标注依赖专业医生逐层修订，成本高、样本规模有限。因此，如何在低标注条件下获得稳定有效的 3D 分割模型，是一个具有现实意义的研究问题。

现有 3D 分割方法多依赖人工设计网络结构。U-Net [1]、3D U-Net [2]、V-Net [3]和 nnU-Net [4]推动了医学图像分割的发展，但其结构仍较依赖经验，难以针对特定数据自适应优化。神经架构搜索(NAS)为此提供了自动化方案，其中 DARTS [5]实现了高效可微搜索，DiNTS [6]则进一步面向 3D 医学图像分割构建了联合操作与拓扑的搜索空间。

在低标注场景下，半监督学习可通过未标注数据缓解监督不足。Temporal Ensembling [7]、Mean Teacher [8]以及 UA-MT [9]、CPS [10]、MetaSSL [11]、CCNet [12]等方法已取得一定成果。然而，伪标签往往会引入噪声梯度，并与有标签数据的优化方向发生冲突[13]。在可微 NAS 中，这类噪声不仅影响网络权重学习，还可能直接干扰架构参数更新。已有少量研究开始探索半监督学习与 NAS 的结合，但针对 3D 体积分割中“如何让无标签信号稳定参与架构搜索”这一问题，仍缺乏有效方法。

为此，本文提出 DiNTS-TMO，将 DiNTS 的可微拓扑搜索与可信动量优化(Trusted Momentum Optimization, TMO)相结合。TMO 以有标签梯度作为可信参考方向，通过对齐门控选择性利用无标签梯度，在搜索阶段同时约束网络权重和架构参数的更新。基于这一设计，DiNTS-TMO 能够在标注稀缺场景下协同完成架构搜索与模型训练，并在 STS-Tooth 数据集的验证中取得了具有竞争力的结果。

2. 相关工作

2.1. 神经架构搜索与半监督学习的协同

半监督学习与神经架构搜索的结合，为低标注条件下提升结构适应性提供了新的思路。现有代表性工作主要包括两类：Se2NAS [14]通过在搜索阶段引入自监督辅助任务，为架构优化提供额外监督；L-SSHNN [15]则利用师生一致性信号与单路径搜索策略，将未标注数据纳入结构选择过程。这些研究说明，

未标注数据不仅可以参与参数学习，也能够一定程度上参与架构寻优。

但在 3D 医学图像分割中，这一路径面临更高的稳定性挑战。由于 3D 牙齿 CBCT 存在背景复杂、边界模糊和标注稀缺等特点，未标注样本产生的伪标签更容易引入噪声梯度，并进一步干扰与权重耦合更新的架构参数。针对这一问题，本文提出 DiNTS-TMO，在可微拓扑搜索过程中引入梯度筛选机制，以降低不可靠无标签信号对搜索结果的干扰。

2.2. 面向冲突场景的梯度对齐与可信优化

当训练过程同时受到多源更新信号影响时，梯度冲突往往会削弱优化稳定性。围绕这一问题，现有研究主要提供了两类代表性思路：一类以梯度手术为代表，通过检测并修正任务间相互冲突的梯度分量，缓解负迁移现象[16]；另一类以“谨慎优化”为代表，根据当前梯度与历史动量方向的一致性筛选不可靠更新，从而减少震荡并提升优化效率[13]。

这些工作表明，基于方向一致性的梯度筛选是处理冲突更新的有效途径。本文据此提出可信动量优化(TMO)，在半监督场景中将有标签梯度视为更可靠的参考方向，并通过对齐门控 $\phi = \mathbb{I}(g_L \cdot g_U > 0)$ 选择性保留与其同向的无标签梯度。这样既能利用未标注数据提供补充学习信号，又能减弱伪标签噪声对网络权重和架构参数更新的干扰，为后续的可微拓扑搜索提供更稳定的优化基础。

3. 方法

本研究提出了一种基于可微网络拓扑搜索(DiNTS)和可信动量优化(TMO)的半监督学习框架，用于解决 3D 牙齿分割中数据标注稀缺和解剖结构复杂的问题。整体框架如图 1 所示。

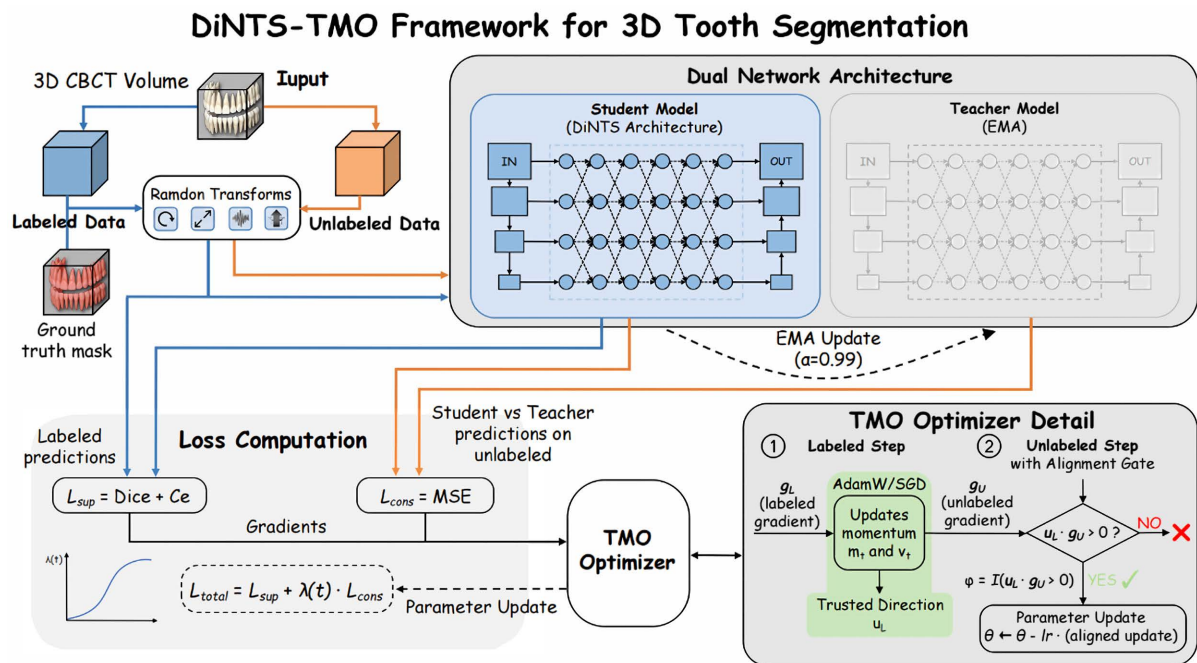


Figure 1. DiNTS-TMO semi-supervised 3D tooth segmentation framework

图 1. DiNTS-TMO 半监督 3D 牙齿分割框架

3.1. DiNTS 网络架构搜索

DiNTS (Differentiable Network Topology Search) [6]是一种专门针对 3D 医学图像分割设计的高效神经

架构搜索方法。其搜索空间采用分层设计，包含微观操作层和宏观拓扑层两个层次，实现了对网络结构的全面搜索。

微观操作(Cell-level Operations)与 DARTS [5]中所有计算节点共享相同操作不同，DiNTS 对每个计算单元(Cell)独立搜索最优操作。如表 1 所示，候选操作集包含 5 种基本操作。其中，跳跃连接实现恒等映射，允许特征直接传递；标准 $3 \times 3 \times 3$ 卷积提供完整的三维空间特征提取能力；三种伪 3D (P3D) 卷积变体[17]则将标准 3D 卷积分解为两个连续的低秩卷积，分别沿不同空间轴组合，在保持等效感受野的同时显著降低计算复杂度和参数量。每个 Cell 还包含 ReLU 激活函数和实例归一化(Instance Normalization)层，分别置于卷积操作的前后(跳跃连接除外)。值得注意的是，Cell 层操作不包含空洞卷积或池化等多尺度特征聚合操作，空间尺度的变换由宏观拓扑层的上/下采样边统一处理。

Table 1. DiNTS microscopic operation search space

表 1. DiNTS 微观操作搜索空间

索引	操作类型	卷积核配置
0	跳跃连接(Skip Connection)	-
1	3D 卷积	$3 \times 3 \times 3$
2	伪 3D 卷积(P3D-A)	$3 \times 3 \times 1 \rightarrow 1 \times 1 \times 3$
3	伪 3D 卷积(P3D-B)	$3 \times 1 \times 3 \rightarrow 1 \times 3 \times 1$
4	伪 3D 卷积(P3D-C)	$1 \times 3 \times 3 \rightarrow 3 \times 1 \times 1$

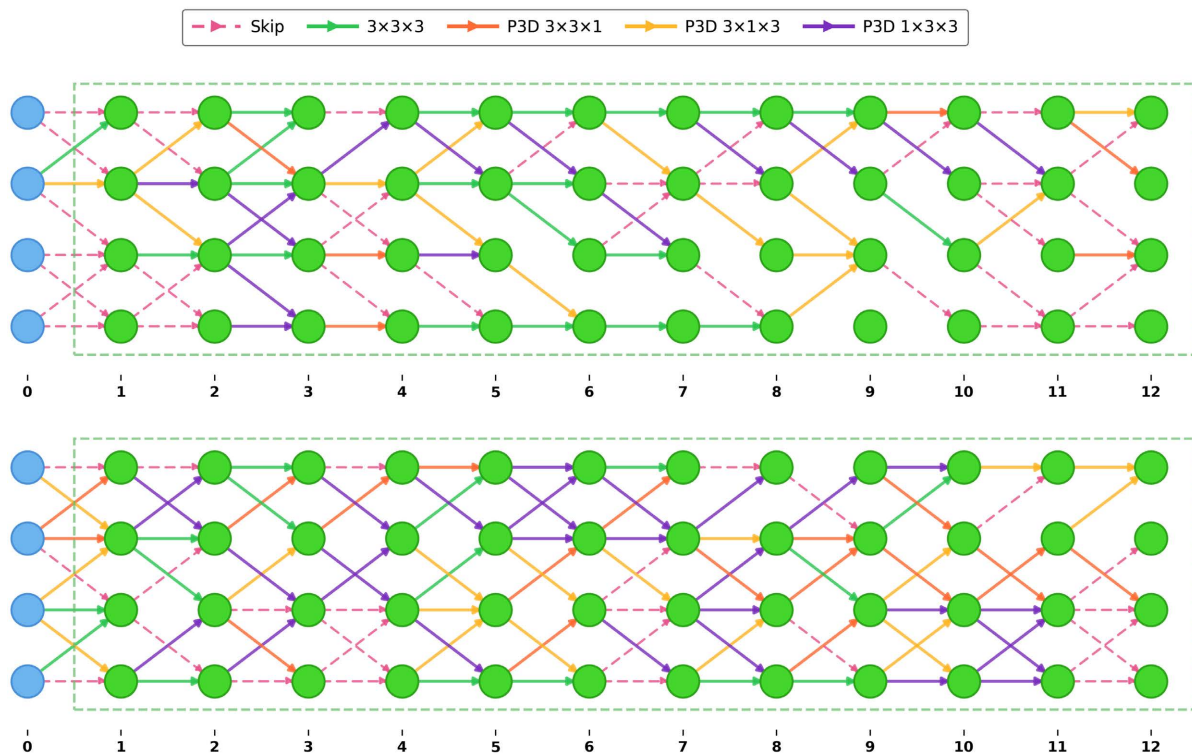


Figure 2. The architecture searched by DiNTS-TMO on the ROI (top) and Integrity (bottom) subsets

图 2. DiNTS-TMO 在 ROI (上)和 Integrity (下)子集上搜索得到的架构

宏观拓扑(Network Topology)宏观拓扑层定义了特征图在不同空间分辨率之间的流动路径。搜索空间由 L 个层组成, 每层包含 D 个不同分辨率的特征节点(对应原始分辨率的 1, 1/2, 1/4, 1/8 等尺度)。相邻层之间通过候选边全连接, 每条边可选择三种状态: 保持分辨率、2 倍下采样或 2 倍上采样。对于下采样/上采样边, 会在 Cell 操作之前额外执行相应的空间变换。每个特征节点的输出是其所有激活输入边特征的求和。

与 Auto-DeepLab [18] 仅支持单路径拓扑不同, DiNTS 支持多路径拓扑搜索, 能够发现类似 U-Net 的编码器-解码器结构以及更复杂的跨尺度连接模式。为了实现通过梯度下降对离散网络拓扑的搜索, DiNTS 采用连续松弛化方法。具体而言, 设特征节点 j 的输出为 x_j , 其由所有前驱节点 $i \in \text{Pre}(j)$ 经过各候选操作 $o \in \mathcal{O}$ 后的特征加权求和得到:

$$x_j = \sum_{i \in \text{Pre}(j)} \sum_{o \in \mathcal{O}} \frac{\exp(\alpha_{i,j}^o)}{\sum_{o' \in \mathcal{O}} \exp(\alpha_{i,j}^{o'})} o(x_i) \quad (1)$$

其中, 可学习的架构参数 $\alpha_{i,j}^o$ 控制着相邻节点之间各条候选边的选择概率。搜索结束后, 结合拓扑感知的离散化算法抽取最终架构, 有效缓解了连续搜索与离散架构之间的性能差距(discretization gap)。搜索得到的架构如图 2 所示, 展示了其能够构建复杂且高效的拓扑结构。

双层优化策略(Bi-level Optimization) DiNTS 采用双层优化框架[5]联合搜索网络架构与权重:

$$\min_{\alpha} \mathcal{L}_{val}(w^*(\alpha), \alpha) \quad \text{s.t.} \quad w^*(\alpha) = \arg \min_w \mathcal{L}_{train}(w, \alpha) \quad (2)$$

其中 α 为架构参数, w 为网络权重。与仅给出概念描述不同, 本文搜索阶段的架构目标在实现中写为:

$$\mathcal{L}_{arch} = \mathcal{L}_{sup}^A + \lambda(t) \mathcal{L}_{cons}^A + \omega(t) (\mathcal{R}_{ent} + \mathcal{R}_{ram} + 10^{-3} \mathcal{R}_{topo}) \quad (3)$$

其中 \mathcal{L}_{sup}^A 和 \mathcal{L}_{cons}^A 分别表示独立搜索批次上的监督损失与一致性损失, $\lambda(t)$ 为一致性损失的 ramp-up 权重, $\omega(t)$ 为 warmup 后线性增长的结构正则组合权重。两者分别定义为:

$$\lambda(t) = \lambda_{\max} \cdot \exp\left(-5 \left(1 - \frac{t}{T_{ramp}}\right)^2\right) \quad (4)$$

$$\omega(t) = \begin{cases} 0, & t < T_w, \\ \frac{t - T_w}{N_{iter} - T_w}, & t \geq T_w, \end{cases} \quad (5)$$

式中 t 为当前迭代数, λ_{\max} 为一致性损失权重上限, T_{ramp} 一致性权重爬坡长度, N_{iter} 为总搜索迭代数, T_w 为 warmup 迭代数。设 $p_a = \text{softmax}(\log \alpha_a)$ 、 $p_c = \text{softmax}(\log \alpha_c)$, 则熵正则项定义为:

$$\mathcal{R}_{ent} = -\text{mean}(p_a \odot \log(p_a + 10^{-5})) - \text{mean}(p_c \odot \log(p_c + 10^{-5})) \quad (6)$$

RAM 约束项用于约束搜索结构的目标显存占比, 在实现中写为:

$$\mathcal{R}_{ram} = \left| \rho_{ram} - \frac{C_{ram}}{C_{ram}^{full}} \right| \quad (7)$$

其中 ρ_{ram} 为可调的目标显存比例超参数, C_{ram} 为当前拓扑的显存开销估计, C_{ram}^{full} 为满连接拓扑下的显存开销。 ρ_{ram} 越小, 搜索过程越倾向于选择更轻量的拓扑; ρ_{ram} 越大, 则会放宽对显存开销的约束。本文实验统一设定 $\rho_{ram} = 0.8$ 。拓扑正则项用于约束相邻 block 之间的节点激活模式连续性, 在实现中写为:

$$\mathcal{R}_{topo} = \sum_{b=1}^{B-1} \sum_{n=1}^N \left[-p_{b,n}^{in} \log(p_{b+1,n}^{out} + 10^{-5}) - (1 - p_{b,n}^{in}) \log(1 - p_{b+1,n}^{out} + 10^{-5}) \right] \quad (8)$$

其中 $p_{b,n}^{in}$ 和 $p_{b+1,n}^{out}$ 分别表示第 b 个与第 $b+1$ 个 block 上第 n 类节点激活模式的输入/输出概率。该项鼓励跨 block 的拓扑连接更平滑，从而降低搜索结果在结构层面出现剧烈震荡的风险。

本文在两个子集上均采用“两阶段”策略：前期仅优化网络权重，后期再交替更新网络权重与架构参数。当前实现中，搜索空间固定为 12 个拓扑块、4 个深度层级。为便于跨数据集复现，本文统一按有效迭代数报告调度超参数：两个子集的总搜索预算均约为 2×10^4 次迭代，warmup 迭代数固定为 $T_w = 5000$ ，一致性权重的爬坡周期 T_{ramp} 同样设为 5000 次迭代。这意味着前 5000 次迭代仅用于稳定网络权重与教师信号；随后再逐步引入架构参数更新以及结构正则项的组合权重，从而避免早期噪声过早进入架构搜索。

3.2. 可信动量优化(TMO)半监督策略

为稳定利用未标注数据，本文将 TMO 引入 DiNTS 的半监督训练流程。在搜索阶段，TMO 分别整合进权重优化器 TMOSGD 与架构优化器 TMOAdamW；在重训练阶段，当搜索得到的拓扑固定后，TMO 继续以 TMOSGD 作用于网络权重更新。两类参数共享同一门控思想：以有标签梯度建立可信方向，再仅吸收与其同向的无标签分量。统一门控写为：

$$\phi^\theta = \mathbb{I}(u_L^\theta \odot \tilde{g}_U^\theta > 0), \bar{\phi}^\theta = \frac{\phi^\theta}{\max(\text{mean}(\phi^\theta), 0.2)} \quad (9)$$

其中 $\theta \in \{w, \alpha_a, \alpha_c\}$ ， \tilde{g}_U^θ 为加入权重衰减后的无标签梯度；归一化门控的下界固定为 0.2，用于避免有效步长因可用梯度过少而过度缩小。

TMOSGD 权重更新对网络权重 w ，代码中使用带动量的 SGD 形式，动量系数 $\mu = 0.9$ 。在监督步中，先计算带 L_2 正则的可信梯度：

$$\tilde{g}_L^w = g_L^w + \lambda_{wd} w_t, m_t^w = \mu m_{t-1}^w + \tilde{g}_L^w, u_L^w = m_t^w \quad (10)$$

在无标签步中，仅对齐并累加同向分量：

$$\tilde{g}_U^w = g_U^w + \lambda_{wd} w_t, m_t^w \leftarrow m_t^w + (\tilde{g}_U^w \odot \phi^w) \quad (11)$$

$$w_{t+1} = w_t - \eta_w (m_t^w \odot \bar{\phi}^w) \quad (12)$$

因此，TMOSGD 的门控直接作用于带动量的更新方向，未对齐的无标签梯度不会进入最终参数步长。

TMOAdamW 架构更新对架构参数 $\alpha = \{\alpha_a, \alpha_c\}$ ，代码中为 $\log \alpha_a$ 与 $\log \alpha_c$ 分别维护独立的 TMOAdamW 优化器，其超参数为 $\beta_1 = 0.9$ 、 $\beta_2 = 0.999$ 、 $\epsilon = 10^{-8}$ ，且架构参数权重衰减设为 0。在监督步中：

$$m_t^\alpha = \beta_1 m_{t-1}^\alpha + (1 - \beta_1) g_L^\alpha \quad (13)$$

$$v_t^\alpha = \beta_2 v_{t-1}^\alpha + (1 - \beta_2) (g_L^\alpha)^2 \quad (14)$$

$$u_L^\alpha = \frac{\hat{m}_t^\alpha}{\sqrt{\hat{v}_t^\alpha + \epsilon}} \quad (15)$$

无标签步中仅保留与 u_L^α 同向的分量：

$$m_t^\alpha \leftarrow m_t^\alpha + (1 - \beta_1) (g_U^\alpha \odot \phi^\alpha) \quad (16)$$

$$v_t^\alpha \leftarrow v_t^\alpha + (1 - \beta_2) (g_U^\alpha \odot \phi^\alpha)^2 \quad (17)$$

$$u_t^\alpha = \frac{\hat{m}_t^\alpha}{\sqrt{\hat{v}_t^\alpha + \epsilon}}, \alpha_{t+1} = \alpha_t (1 - \eta_\alpha \lambda_{wd}) - \eta_\alpha (u_t^\alpha \odot \bar{\phi}^\alpha) \quad (18)$$

在本文实验中, $\lambda_{wd} = 0$ 对应架构参数不施加额外衰减。

该机制使未标注样本仅在与有标签梯度一致时参与更新。在搜索阶段, 权重分支与架构分支都按“有标签监督步 + 无标签一致性步”的顺序更新: 先在有标签批次上建立可信方向, 再在未标注批次上执行门控更新。对于架构分支, warmup 结束后才启用, 并且其有标签步在独立搜索批次上最小化监督分割损失与线性增长的结构正则项, 而无标签步仅最小化一致性损失。随后, 若启用半监督搜索, 则使用学生参数对教师模型执行 EMA 更新。进入重训练阶段后, 搜索得到的拓扑保持固定, 不再更新架构参数, 仅保留权重分支的监督步、无标签一致性同步和 EMA 更新。

在权重分支中, 本文采用 Mean Teacher [8] 范式, 教师参数由学生参数的指数移动平均(EMA)更新, 衰减率设为 0.99。对应的监督与一致性损失写为:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda(t) \cdot \mathcal{L}_{cons} \quad (19)$$

其中监督损失 \mathcal{L}_{sup} 由 Dice 损失与交叉熵构成:

$$\mathcal{L}_{sup} = 1 - \frac{2 \sum_i P_i G_i}{\sum_i P_i + \sum_i G_i} - \sum_i G_i \log(P_i) \quad (20)$$

式中 P 和 G 分别代表预测概率和金标准(Ground Truth); 一致性损失 $\mathcal{L}_{cons} = \|P_{stu} - P_{tea}\|_2^2$ 。权重分支与架构分支共享同一 ramp-up 权重 $\lambda(t)$, 其定义与前述架构目标中的一致性调度保持一致; 这样做是为了避免训练初期教师模型带来过强约束。

4. 实验

4.1. 数据集与预处理

本研究使用 STS-Tooth 数据集进行实验验证[19]。该数据集是一个多中心、专用于半监督牙齿分割的 3D CBCT 医学影像数据集。为系统评估模型在不同数据场景下的性能, 我们按照该数据集的原始设计, 将其划分为两个子集进行实验:

STS3D-ROI-Subset: 该子集包含 ROI (Region of Interest) 裁剪后的牙齿 CBCT 数据, 由 130 个 3D 牙齿样本构成。其中 22 个样本为有标签集, 其余 108 个样本为无标签集, 标签为牙齿分割的二值掩模(背景和牙齿)。ROI 裁剪从完整 CBCT 扫描中提取包含牙齿的目标区域(如整排牙齿的中间部分或半侧区域), 聚焦于有牙齿分布的有效区域, 减少了大量无关背景信息。

STS3D-Integrity-Subset: 该子集包含完整牙齿扫描数据, 由 241 个离体牙齿样本的完整 3D CBCT 扫描构成。其中 10 个样本为有标签集, 其余 231 个样本为无标签集。该子集保留了完整的牙齿结构信息, 适用于评估模型在完整扫描场景下的性能。

为确保实验的统计可靠性和模型泛化能力的准确评估, 两个子集均采用 5 折交叉验证策略进行模型评估。对于半监督实验比较, 所有方法均采用相同的 1:1 标注/未标注数据协议, 即在每个 fold 中, 参与训练的未标注样本数量设为该 fold 中有标注训练样本数量的 1 倍, 以保证不同方法处于一致的监督强度下。数据预处理流程全部统一。

4.2. 实验设置

4.2.1. 实现细节

本研究基于 PyTorch 2.7 [20] 和 MONAI 1.5 [21] 框架实现。所有实验均在配备 NVIDIA RTX 4090 GPU (24 GB 显存) 的工作站上进行。对于全监督方法(UNet [1]、DynUNet [4]、Attention U-Net [22]、UNETR [23]、SwinUNETR [24]、SegResNet [25]、VNet [3]), 采用 SGD 优化器, 初始学习率为 1×10^{-4} , 训练 200

个 epoch。对于 DiNTS 方法[6]，架构搜索阶段设置搜索空间为 12 个拓扑块、3 个深度层级。架构参数使用 Adam 优化器(学习率 3×10^{-3})，网络权重使用 SGD 优化器(初始学习率 0.025，动量 0.9，权重衰减 4×10^{-5})，总计训练 21,000 次迭代。重训练阶段使用搜索得到的最优架构，训练 200 个 epoch。

4.2.2. 评价指标

本研究采用两个广泛使用的医学图像分割评价指标：

Dice 相似系数(Dice Similarity Coefficient, DSC)衡量预测分割与真实标签之间的重叠程度，定义为：

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|} \quad (21)$$

其中 P 表示预测区域， G 表示真实标签区域。DSC 取值范围为[0, 1]，值越大表示分割效果越好。

95% Hausdorff 距离(95% Hausdorff Distance, HD95)衡量预测边界与真实边界之间的距离，定义为两个方向表面距离的 95%分位数的最大值：

$$\text{HD95} = \max \{d_{95}(S_P, S_G), d_{95}(S_G, S_P)\} \quad (22)$$

其中 S_P 和 S_G 分别表示预测和真实标签的表面点集， d_{95} 表示 95%分位数距离。相比传统 Hausdorff 距离，HD95 对异常值更鲁棒，HD95 值越小表示边界预测越准确。

平均表面距离(Average Surface Distance, ASD)用于度量预测边界与真实边界之间的平均偏差，定义为：

$$\text{ASD} = \frac{\sum_{x \in S_P} d(x, S_G) + \sum_{y \in S_G} d(y, S_P)}{|S_P| + |S_G|} \quad (23)$$

ASD 对整体边界偏移更敏感，而不像 HD95 那样主要受少量极端离群点主导。本文所有表面距离指标均根据原始 NIfTI 元数据中的 spacing 计算物理距离(mm)，以保证完整扫描场景下的边界评估具有一致的物理意义。

4.3. 全监督方法对比

为验证本文搜索得到的骨干拓扑在全监督设置下的表现，本节对比了多种主流 3D 医学图像分割网络在仅使用有标签数据时的性能。所有方法均在相同监督设定下训练，确保公平对比。表 2 展示了各方法在 STS3D-ROI-Subset 和 STS3D-Integrity-Subset 两个子集上的 5 折交叉验证结果。

表 2 展示了多种主流 3D 医学分割网络在极少量标注条件下的全监督结果。结合 DSC 和 HD95，可以得到以下几点观察：

首先，具备局部归纳偏置的 CNN 结构提供了较强的基线性能。SegResNet 与 Attention U-Net 在两个子集上的结果相对稳定：在 ROI 子集中，其 DSC 分别为 87.63%和 87.62%；在 Integrity 子集中，SegResNet 的 HD95 为 6.84 mm，是该组比较方法中最好的边界距离结果之一。

其次，Transformer 架构在当前低标注设置下未表现出明显优势。UNETR 与 SwinUNETR 在 Integrity 子集上的 DSC 分别为 84.72%和 49.90%，均低于多数 CNN 基线，说明在当前数据规模和训练条件下，全局自注意力结构对该任务的适应性仍需进一步验证。

最后，本文搜索得到的骨干拓扑表现出较强竞争力。即使不引入半监督模块，仅将搜索拓扑作为 backbone 使用，也在 ROI 子集上取得最低 HD95 (3.85 mm)，并在 Integrity 子集上取得最高 DSC (91.91%)。这表明，数据驱动的拓扑搜索能够为当前任务提供有效的基础架构。

Table 2. Performance comparison of fully supervised segmentation based on different network architectures
表 2. 不同网络架构的全监督分割性能对比

方法	STS3D-ROI-Subset		STS3D-Integrity-Subset	
	DSC (%)	HD95 (mm)	DSC (%)	HD95 (mm)
UNet	87.43 ± 1.56	10.82 ± 4.80	89.45 ± 0.92	22.66 ± 20.41
Attention U-Net	87.62 ± 2.54	7.68 ± 4.78	91.65 ± 0.97	11.61 ± 12.00
VNet	87.32 ± 2.15	10.92 ± 3.66	53.54 ± 6.20	180.71 ± 24.55
nnUNet	87.73 ± 2.98	16.61 ± 11.21	90.46 ± 1.37	16.54 ± 19.67
SegResNet	87.63 ± 1.75	7.94 ± 5.39	91.53 ± 1.10	6.84 ± 2.62
UNETR	85.92 ± 2.18	16.32 ± 5.02	84.72 ± 1.97	75.08 ± 46.41
SwinUNETR	86.69 ± 2.18	27.85 ± 9.78	49.9 ± 1.60	200.07 ± 30.55
DiNTS	87.44 ± 3.17	7.32 ± 7.33	91.31 ± 0.90	20.23 ± 13.98
DiNTS-TMO	87.95 ± 1.98	3.85 ± 7.74	91.91 ± 1.84	11.98 ± 26.30

图 3 的可视化结果与定量结论一致。V-Net 与 SwinUNETR 在部分样例中出现了较明显的骨骼误分割，而本文搜索拓扑与原始 DiNTS 的预测轮廓更接近真实标签，对复杂背景下离群骨骼结构的误分割更少。

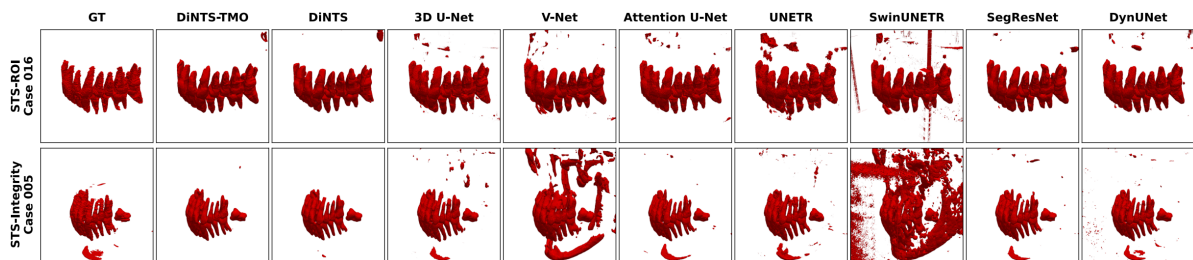


Figure 3. Visual comparison of 3D tooth segmentation results using fully supervised methods

图 3. 全监督方法的 3D 牙齿分割结果可视化对比

4.4. 半监督方法对比

为验证本文提出的 DiNTS-TMO 的有效性，本节将 DiNTS-TMO 与现有半监督医学图像分割方法进行对比。表 3 展示了 UA-MT [9]、CPS [10]、MetaSSL [11]、CCNet [12] 和 DiNTS-TMO 在 STS3D-ROI-Subset 和 STS3D-Integrity-Subset 两个数据集上的 5 折交叉验证结果。

表 3 显示，DiNTS-TMO 在两个数据子集上呈现出不同形式的性能特点。在 STS3D-ROI-Subset 上，DiNTS-TMO 同时取得了最高的 DSC (87.95%) 和最低的 HD95 (3.85 mm)，说明在经过 ROI 裁剪、背景干扰相对受控的场景中，DiNTS 搜索得到的拓扑与 TMO 的双路门控更新能够同时改善区域重叠和边界定位精度。与 CPS、CCNet 等比较方法相比，DiNTS-TMO 在该子集上的整体结果更均衡。

在 STS3D-Integrity-Subset 上，DiNTS-TMO 获得了最高的 DSC (91.88%)，表明该方法在完整牙列场景下对目标区域的整体重建建模较为充分；但其 HD95 为 11.98 mm，稍落后于 CCNet (11.76 mm)。这说

明 DiNTS-TMO 在复杂全景扫描中的主体分割表现较好,但在局部边界距离指标上仍有进一步优化空间。总体来看, DiNTS-TMO 并非在所有指标上都占优,但其在 ROI 子集上的综合表现和在 Integrity 子集上的最优 DSC 表明, TMO 参与权重参数与架构参数更新的搜索机制在不同复杂度的数据场景下具有一定的有效性。

Table 3. Comparison of segmentation performance of different semi-supervised methods
表 3. 不同半监督方法的分割性能对比

方法	STS3D-ROI-Subset		STS3D-Integrity-Subset	
	DSC (%)	HD95 (mm)	DSC (%)	HD95 (mm)
UA-MT	73.99 ± 5.64	20.80 ± 5.78	71.28 ± 3.36	165.13 ± 51.09
CPS	87.05 ± 4.46	7.34 ± 8.26	90.36 ± 2.24	19.27 ± 34.22
MetaSSL	81.27 ± 5.30	14.50 ± 7.95	86.10 ± 2.97	42.35 ± 39.23
CCNet	85.59 ± 4.29	5.51 ± 7.87	88.56 ± 2.25	11.76 ± 11.58
DiNTS-TMO	87.95 ± 1.98	3.85 ± 7.74	91.91 ± 1.84	11.98 ± 26.30

图 4 展示了两个代表性样例的三维重建结果。可以看出, UA-MT 与 MetaSSL 更容易出现大范围背景误检, CPS 与 CCNet 能够较好抑制噪声,但仍存在少量离散碎片。相比之下, DiNTS-TMO 的预测轮廓与 GT 更接近,在两个样例中都保持了更好的主体完整性。

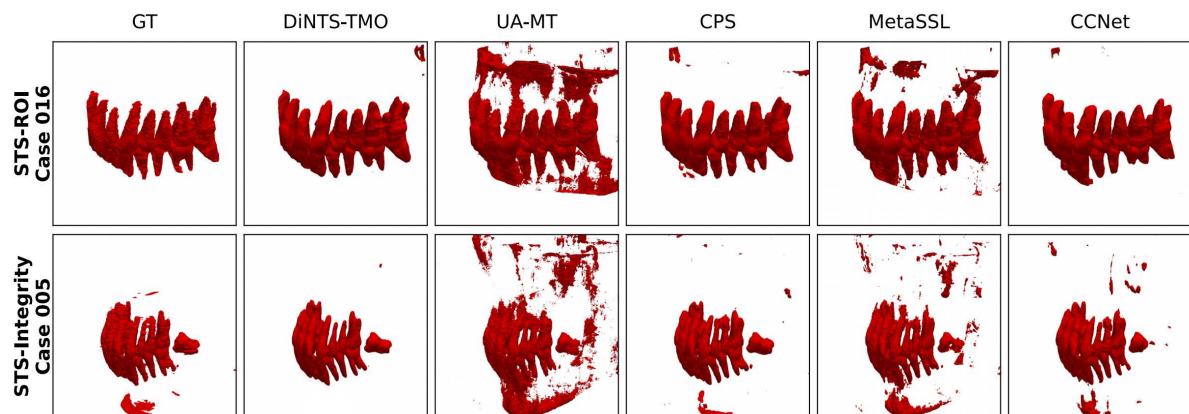


Figure 4. Visual comparison of 3D tooth segmentation results using semi-supervised methods
图 4. 半监督方法的 3D 牙齿分割结果可视化对比

4.5. Integrity 子集边界误差分析

为进一步解释 Integrity 子集上“DSC 最优但 HD95 不占优”的现象,表 4 补充报告了 Average Surface Distance (ASD)。ASD 更关注整体平均表面偏差,而 HD95 对少量极端离群点更敏感。真实评估结果显示, DiNTS-TMO 的 ASD 为 2.82 ± 2.07 mm,略优于 CPS 的 2.84 ± 1.65 mm,并明显优于 CCNet 的 3.83 ± 1.81 mm。这说明从整体边界质量看, DiNTS-TMO 在大多数病例上的平均表面偏差并未恶化。进一步的逐例结果表明,在 10 个 Integrity 病例中, DiNTS-TMO 有 7 例取得了最低 ASD;但 Case 004 与 Case 007 的极端失败会显著拉高边界距离分布尾部,从而解释其 HD95 仍不占优的现象。

Table 4. Boundary distance evaluation on the STS3D-Integrity-Subset
表 4. STS3D-Integrity-Subset 上的边界距离评估

方法	ASD (mm)
CPS	2.84 ± 1.65
CCNet	3.83 ± 1.81
DiNTS-TMO (Ours)	2.82 ± 2.07

图 5 给出了 Integrity 子集上两类具有代表性的失败模式。Case 004 中，模型在主体牙弓上的轮廓基本正确，但在远离主体的位置出现了两处明显的离群假阳性碎片。这类错误虽然只占少量体素，却会同时抬高 ASD 与 HD95，并对后者产生更强的尾部放大效应。Case 009 则主要表现为局部漏分与轮廓偏移并存，说明即使没有远端大块离群碎片，局部边界的系统性外扩或收缩也会累积成可观的平均表面偏差。值得注意的是，从局部放大图可以观察到，DiNTS-TMO 的预测轮廓在个别牙位与图像中的实际强度边界更为一致，这提示参考标注在少数局部边界处可能存在一定不确定性。因此，Integrity 子集上的边界距离偏差除了反映模型误差外，也可能受到单一参考标注误差的额外扰动。总体而言，DiNTS-TMO 在大多数病例中具有较好的整体边界质量，但其当前局限在于对少数困难区域的边界控制仍不够稳健。

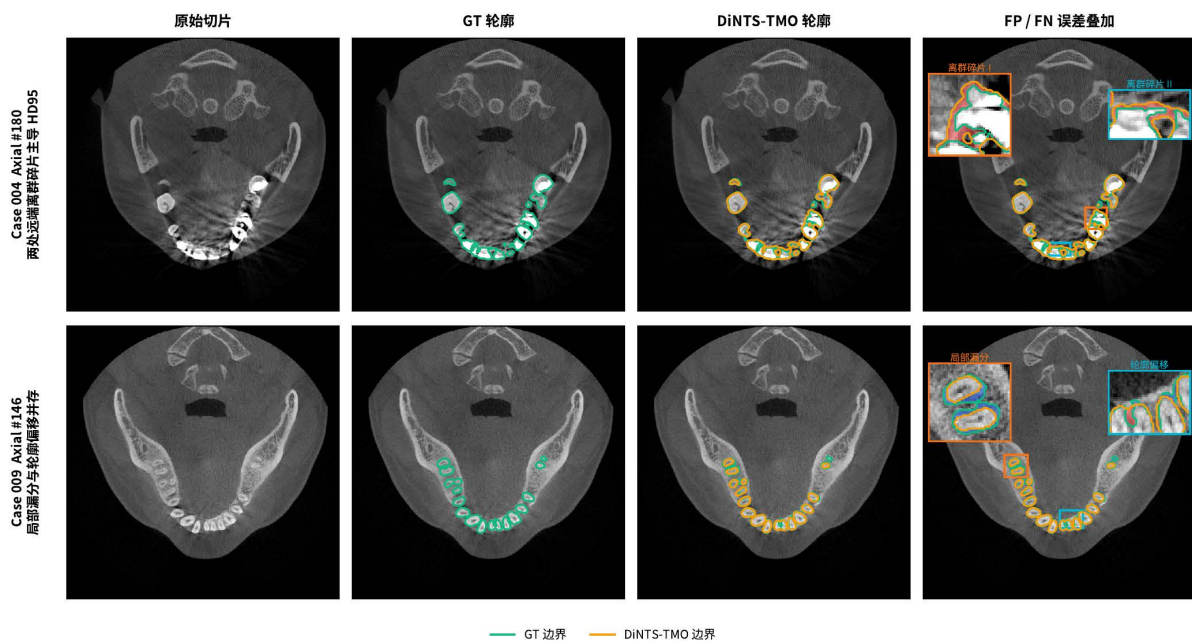


Figure 5. Two-dimensional boundary error visualization on the Integrity subset

图 5. Integrity 子集上的二维边界误差可视化

进一步结合图 2 可见，Integrity 子集搜索到的拓扑并未像 ROI 子集那样较早收缩为单条深层主干，而是在多数 block 中保留了更多并行分辨率节点和更持续的双向跨尺度连边。这种结构有助于在完整扫描场景中同时维持牙弓主体定位与局部边界恢复，因此 DiNTS-TMO 在该子集上取得更高 DSC，并在 ASD 上表现为更小的平均表面偏差；但其副作用是局部异常响应也更容易沿多条路径被保留和传播，从而在少数困难病例中形成远端离群碎片或局部轮廓偏移，最终拉高边界距离指标。

5. 结论与展望

本文针对低标注条件下的 3D 牙齿分割问题, 提出了结合 DiNTS 可微拓扑搜索与可信动量优化的 DiNTS-TMO。方法上, TMO 在搜索阶段同时作用于网络权重和架构参数的更新, 以有标签梯度作为可信方向, 并通过对齐门控筛选无标签梯度, 从而降低噪声伪标签对半监督架构搜索的干扰。

在 STS-Tooth 数据集的两个子集上的五折交叉验证结果表明, DiNTS-TMO 在 ROI 子集上取得了最优的 DSC 和 HD95, 在 Integrity 子集上取得了最优的 DSC, 说明该方法在低标注 3D 牙齿分割任务中具有较强的竞争力。同时, 结果也表明该方法在完整扫描场景下的边界距离指标仍有改进空间。未来的工作将考虑引入更直接的边界约束或不确定性建模策略, 以进一步改善复杂场景中的边界质量, 并在更多中心临床数据上验证方法的稳定性与泛化能力。

参考文献

- [1] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [2] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O. (2016) 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: *Lecture Notes in Computer Science*, Springer, 424-432. https://doi.org/10.1007/978-3-319-46723-8_49
- [3] Milletari, F., Navab, N. and Ahmadi, S. (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 *Fourth International Conference on 3D Vision (3DV)*, Stanford, 25-28 October 2016, 565-571. <https://doi.org/10.1109/3dv.2016.79>
- [4] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J. and Maier-Hein, K.H. (2021) NNU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nature Methods*, **18**, 203-211. <https://doi.org/10.1038/s41592-020-01008-z>
- [5] Liu, H., Simonyan, K. and Yang, Y. (2019) DARTS: Differentiable Architecture Search. *Proceedings of the International Conference on Learning Representations*, New Orleans, 6-9 May 2019, 1-13.
- [6] He, Y., Yang, D., Roth, H., Zhao, C. and Xu, D. (2021) DiNTS: Differentiable Neural Network Topology Search for 3D Medical Image Segmentation. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 5837-5846. <https://doi.org/10.1109/cvpr46437.2021.00578>
- [7] Laine, S. and Aila, T. (2017) Temporal Ensembling for Semi-Supervised Learning. *Proceedings of the International Conference on Learning Representations*, Toulon, 24-26 April 2017, 1-13.
- [8] Tarvainen, A. and Valpola, H. (2017) Mean Teachers are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning. *Advances in Neural Information Processing Systems*, **30**, 1195-1204.
- [9] Yu, L., Wang, S., Li, X., Fu, C.W. and Heng, P.A. (2019) Uncertainty-Aware Self-Ensembling Model for Semi-Supervised 3D Left Atrium Segmentation. In: *Lecture Notes in Computer Science*, Springer, 605-613. https://doi.org/10.1007/978-3-030-32245-8_67
- [10] Chen, X., Yuan, Y., Zeng, G. and Wang, J. (2021) Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 2613-2622. <https://doi.org/10.1109/cvpr46437.2021.00264>
- [11] Zhao, W., Zhong, L., Liao, X., Liao, W., Zhang, S., Zhang, S., *et al.* (2026) MetaSSL: A General Heterogeneous Loss for Semi-Supervised Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, **45**, 751-763. <https://doi.org/10.1109/tmi.2025.3605617>
- [12] Huang, H., Chen, Z., Chen, C., Lu, M. and Zou, Y. (2023) Complementary Consistency Semi-Supervised Learning for 3D Left Atrial Image Segmentation. *Computers in Biology and Medicine*, **165**, Article 107368. <https://doi.org/10.1016/j.compbiomed.2023.107368>
- [13] Liang, K., Chen, L., Liu, B. and Liu, Q. (2026) Cautious Optimizers: Improving Training with One Line of Code. *Proceedings of the International Conference on Learning Representations*, Rio de Janeiro, 23-27 April 2026, 1-26.
- [14] Pauletto, L., Amini, M. and Winckler, N. (2022) Se²NAS: Self-Semi-Supervised Architecture Optimization for Semantic Segmentation. 2022 *26th International Conference on Pattern Recognition (ICPR)*, Montreal, 21-25 August 2022, 54-60. <https://doi.org/10.1109/icpr56361.2022.9956294>
- [15] Chen, R., Nian, F., Cen, Y., Peng, Y., Wang, H., Yu, Z., *et al.* (2025) L-SSHNN: A Larger Search Space of Semi-Supervised Hybrid NAS Network for Echocardiography Segmentation. *Expert Systems with Applications*, **276**, Article

127084. <https://doi.org/10.1016/j.eswa.2025.127084>
- [16] Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K. and Finn, C. (2020) Gradient Surgery for Multi-Task Learning. *Advances in Neural Information Processing Systems*, **33**, 5824-5836.
- [17] Qiu, Z., Yao, T. and Mei, T. (2017). Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 5534-5542. <https://doi.org/10.1109/iccv.2017.590>
- [18] Liu, C., Chen, L., Schroff, F., Adam, H., Hua, W., Yuille, A.L., *et al.* (2019) Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 82-92. <https://doi.org/10.1109/cvpr.2019.00017>
- [19] Wang, Y., Ye, F., Chen, Y., Wang, C., Wu, C., Xu, F., *et al.* (2025) A Multi-Modal Dental Dataset for Semi-Supervised Deep Learning Image Segmentation. *Scientific Data*, **12**, Article No. 117. <https://doi.org/10.1038/s41597-024-04306-9>
- [20] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., *et al.* (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, **32**, 8024-8035.
- [21] MONAI Contributors (2026) Modules—MONAI 1.5.0 Documentation. <https://docs.monai.io/en/1.5.0/modules.html>
- [22] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas. *Proceedings of the International Conference on Medical Imaging with Deep Learning*, Amsterdam, 4-6 July 2018, 1-10.
- [23] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., *et al.* (2022) UNETR: Transformers for 3D Medical Image Segmentation. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2022, 1748-1758. <https://doi.org/10.1109/wacv51458.2022.00181>
- [24] Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., *et al.* (2022) Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 20698-20708. <https://doi.org/10.1109/cvpr52688.2022.02007>
- [25] Myronenko, A. (2019) 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M. and van Walsum, T., *Lecture Notes in Computer Science*, Springer, 311-320. https://doi.org/10.1007/978-3-030-11726-9_28