

DOI: 10.19650/j.cnki.cjsi.J2513717

基于多尺度卷积和选择性核双注意力机制的 半监督全景 X 射线图像龋齿分割*

薛钟毫^{1,2}, 姜金刚^{1,2}, 孙健鹏^{1,2}, 潘 洁³, 张嘉伟¹

(1. 哈尔滨理工大学先进制造智能化技术教育部重点实验室 哈尔滨 150080; 2. 哈尔滨理工大学机器人技术及
工程应用研究中心 哈尔滨 150080; 3. 北京大学口腔医学院 北京 100081)

摘 要:全景 X 射线图像的龋齿分割是进行早期龋齿检测以及后续治疗的重要前提,为实现全景 X 射线图像中龋齿的精确自动分割,提出一种具有多尺度卷积和选择性核双注意力机制的半监督学习框架,该方法旨在利用大量未标注数据增强模型泛化能力,并缓解龋齿病灶区域边界模糊、对比度低等问题。框架设计上,采用教师-学生双网络结构,通过多尺度卷积注意力机制对学生网络多层解码器进行深度监督,提升对边界细节和类间相似区域的判别能力。同时,引入选择性核注意力机制融合教师网络的多级预测结果,根据像素不确定性自适应选择不同卷积核,生成精确的不确定性掩模图,引导学生网络优化学习。实验在数据集 1 和 2 上进行,结果显示,在 265 切片数据上联合使用双注意力机制较基线模型在 Dice 系数、查准率和灵敏度分别提升 3.91%、2.14% 和 5.35%;在 530 切片数据上则提升 1.39%、5.69% 和 12.34%,验证了方法在大规模数据下的稳定性和适应性。与传统全监督模型相比,所提出的方法在 Dice 系数、查准率和灵敏度上最高分别提升 22.27%、17.64% 和 24.57%;相比最新半监督模型也分别提升最多 14.54%、14.81% 和 11.96%。本研究不仅有效提升了龋齿分割性能,同时也为全景 X 射线图像处理提供了一种精确的分割方案。

关键词: 龋齿分割;半监督学习;双注意力机制;全景 X 射线图像

中图分类号: TP391.41 TH7 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Semi-supervised caries segmentation of panoramic X-ray images based on multi-scale convolution and selective kernel dual-attention*

Xue Zhonghao^{1,2}, Jiang Jingang^{1,2}, Sun Jianpeng^{1,2}, Pan Jie³, Zhang Jiawei¹

(1. Key Laboratory of Advanced Manufacturing and Intelligent Technology, Ministry of Education, Harbin University of Science and Technology, Harbin 150080, China; 2. Robotics & its Engineering Research Center, Harbin University of Science and Technology, Harbin 150080, China; 3. Peking University School of Stomatology, Beijing 100081, China)

Abstract: Caries segmentation in panoramic X-ray images is an important prerequisite for early caries detection and subsequent treatment. In order to achieve accurate and automatic segmentation of caries in panoramic X-ray images, a semi-supervised learning framework with multi-scale convolution and selective kernel dual-attention mechanism is proposed. This framework aims to enhance the generalization capability of the model by leveraging a large amount of unlabeled data, while addressing challenges such as blurred lesion boundaries and low contrast in caries-affected regions. The framework adopts a teacher-student dual network architecture. It applies multi-scale convolutional attention to deeply supervise the multilayer decoder in the student network, thereby improving its ability to capture boundary details and distinguish between similar inter-class regions. Meanwhile, a selective kernel attention mechanism is introduced to fuse multi-level predictions from the teacher network, adaptively selecting convolution kernels based on pixel-level uncertainty to generate accurate uncertainty masks that guide the student's learning process. Experiments conducted on the dataset 1 and dataset 2 demonstrate that, on 265 slices, the dual attention mechanism achieves improvements over the baseline model of 3.91%, 2.14%, and 5.35% in Dice coefficient, precision, and sensitivity, respectively. And on 530 slices, the improvements reach 1.39%, 5.69%, and 12.34%, respectively. Compared with traditional fully supervised models, the proposed method achieves the highest improvements of 22.27%, 17.64%, and 24.57% in Dice coefficient, precision, and sensitivity, respectively. Compared with the latest semi-supervised model, it also improves by up to 14.54%, 14.81%, and 11.96%, respectively. This study not only effectively improves caries segmentation performance, but also provides a precise segmentation scheme for panoramic X-ray image processing.

收稿日期: 2025-01-24 Received Date: 2025-01-24

* 基金项目: 黑龙江省“优秀青年教师基础研究支持计划”项目(YQJH2024075)资助

5.69%, and 12.34%, verifying the method's stability and adaptability on larger-scale data. Compared with traditional fully supervised models, the proposed method achieves the highest improvements in Dice coefficient, precision, and sensitivity, with increases of 22.27%, 17.64%, and 24.57%, respectively. And compared with recent semi-supervised methods, it achieves improvements of up to 14.54%, 14.81%, and 11.96%, respectively. This study not only significantly enhances caries segmentation performance but also provides an accurate and robust solution for panoramic X-ray images.

Keywords: caries segmentation; semi-supervised learning; dual-attention mechanism; panoramic X-ray images

0 引言

龋齿是最常见的口腔疾病,世界卫生组织将龋齿、肿瘤和心血管疾病视为人类预防和治疗的三大关键疾病^[1]。口腔医师可根据全景X射线图像中龋齿病变的分割结果,获取病人病灶的数量、大小、形状、位置、病变程度等信息,对疾病进行诊断、制定适当的治疗方案。因此,龋齿的早期检测对于龋齿的治疗和预防至关重要。而准确、自动的分割龋齿病变是龋齿检测的重要步骤。目前,在龋齿的临床诊断领域,全景X射线图像与锥形束计算机断层扫描技术共同扮演着至关重要的角色^[2]。其中,全景X射线图像凭借其高效的成像速度及相对经济的成本优势,成为了广泛应用的首选工具。而全景X射线图像的正确解读很大程度上取决于临床牙医的经验,不仅费时费力,且大大增加了龋齿诊断的不确定性,导致漏诊和误诊率相对较高。因此研究高效准确的全景X射线图像的龋齿自动分割方法对龋齿的临床诊断与治疗具有重要意义。全景X射线图像中龋齿病变的准确分割是龋齿诊断的重要前提,也是龋齿治疗方案制定的重要基础。

全景X射线图像中的龋齿病变部分通常具有对比度低,边界模糊等特点。传统分割方法,如聚类、区域生长、水平集、图割、活动轮廓模型^[3-4]等,难以有效适应龋齿病变分割的复杂性和多样性。此外,该类方法通常需要人工干预,无法实现龋齿病变区域的自动分割。近年来,随着人工智能技术的飞速发展,尤其是深度学习方法的广泛应用,全监督学习在医学影像分割任务中展现出了巨大的潜力,为龋齿的精准识别与分割提供了新的可能。Zhao等^[5]在U-Net网络中嵌入全局和局部注意力模块并加入残差连接,使网络专注于自动捕捉真实的牙齿区域,细化了牙齿区域和边界。Ying等^[6]借鉴U-Net和Trans-U-Net的实现,在跳跃连接处加入空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP),并采用扩张卷积和特征金字塔融合方法,增强了网络多尺度和全局特征提取能力。Haghanifar等^[7]提出PaXNet,利用遗传算法和Capsule Network作为分类器,在全景X射线图像中分割牙齿,在具有多颗缺失牙齿的具有挑战性的颌骨上也表现出强大的性能。Zhu等^[8]提出CariesNet,使用结合编码器模块和全尺寸轴向注意力模块的U形网络,从通

道域和空间域提取注意力特征,更准确地分割龋齿病变区域。Mârginean等^[9]提出CariSeg,集成了DeeplabV3、U-net和特征金字塔网络(feature pyramid network, FPN)这3种网络架构,实现了在X射线图像上的精确龋齿分割。Son等^[10]引入了一种新颖的基于图的聚类算法,并结合片段权重与隶属度向量分析,综合判定疾病状态,确保了诊断结果的科学性与准确性。Qayyum等^[11]提出了一种基于学生-教师方法的自监督学习方法,利用龋齿区域的质心裁剪图像和不同的增强技术来训练自监督模型,提供了计算和性能增益。Wang等^[12]提出了一种结合半监督学习、多级扰动与一致正则化的方法,通过优化决策边界、构建稳健不确定性掩模图,并引入新数据集,有效提升了龋齿分割性能。

上述方法中,全监督学习方法虽然采用了优越的算法来提高对病变的检测,但全监督训练需要大量的像素级标注数据。对于自然图像语义分割任务而言,其像素级标签可通过非专业人士进行标注,但在医学图像分割中,采集精确的像素级标签需要医学专家花费大量的时间进行标注,因此医学图像标注数据的获取往往成本更加昂贵且耗时。半监督学习方法虽然可以利用少量的标记数据和大量的未标记数据学习分割龋齿,获得高质量的分割结果,但仍无法解决龋齿边界模糊引起的分割精度低,以及数据类别不平衡造成的龋齿病变分割困难等问题。

为了解决上述问题从而实现龋齿病变的准确分割,受注意力机制^[13-16]的启发,提出了一种多尺度卷积和选择性核双注意机制的龋齿分割网络(multi-scale convolutional and selective kernel dual attention caries segmentation network, MSC-SK DACariesNet)。该模型基于半监督学习框架^[17-20](semi-supervised learning, SSL),利用多尺度卷积注意力机制对多层解码器进行深度监督,并通过选择性核注意力机制融合多级预测结果,提高对边界模糊的龋齿病变区域的分割精度。

1 MSC-SK DACariesNet 网络构建

1.1 MSC-SK DACariesNet 网络结构

MSC-SK DACariesNet采用双网络结构,包括教师网络和学生网络。首先,学生网络从每个解码器层提取特征图,作为后续不确定区域决策的候选图。然后,为了增强模型在龋齿区域边界细节捕捉能力和在类间相似区域

的判别能力, MSC-SK DACariesNet 引入多尺度卷积注意力 (multi-scale convolutional attention, MSCA) 提取的特征对学生网络进行直接和深度监督。对于教师网络, MSC-SK DACariesNet 使用指数移动平均 (exponential moving average, EMA) 更新它的权重。此外, 为了生成鲁棒的不确定性图, MSC-SK DACariesNet 结合选择性核注意力机

制 (selective kernel attention, SKA) 根据当前像素的不确定性自适应地选择不同尺度的卷积核, 融合教师网络的直接预测和多级预测, 生成融合预测特征图。最后用融合预测特征图生成掩码图, 得到的结果对学生网络生成的带噪声的直接预测进行一致性约束, MSC-SK DACariesNet 的总体架构如图 1 所示。

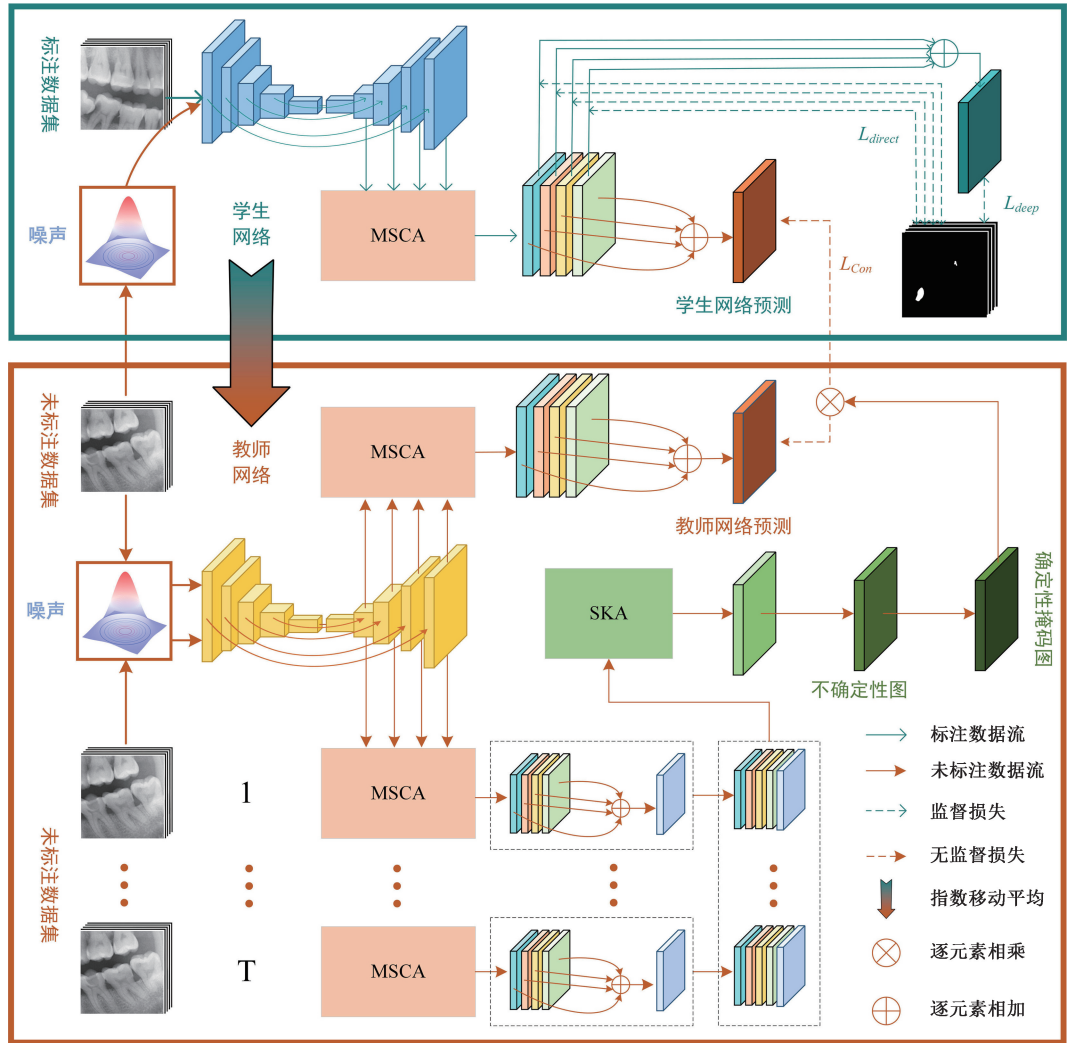


图1 MSC-SK DACariesNet 模型的工作流程

Fig. 1 Workflow diagram of the MSC-SK DACariesNet model

MSC-SK DACariesNet 模型采用了 ResNet34^[21] 作为编码器, 并且使用 Unet^[22] 架构作为分割模型的骨干网络。此外, 该网络还采用了一个双分支架构, 包括教师分支和学生分支。两者虽然结构完全相同, 但具有不同的权重。对于学生网络, 提供了标注数据集, 然后取交叉熵损失和 Dice 损失的平均值为 L_{direct} , 进行全监督训练。损失函数 L_{direct} 的定义如式(1)所示。

$$L_{direct} = \frac{1}{2} \frac{1}{M} \sum_{i=1}^M (l_{bce}(f(x_i), y_i) + l_{dice}(f(x_i), y_i)) \quad (1)$$

式中: M 是全监督学习数据的个数; $f(\cdot)$ 表示学生网络;

$l_{bce}(\cdot)$ 和 $l_{dice}(\cdot)$ 分别表示监督学习二元交叉熵损失和 Dice 损失; L_{direct} 表示学生网络 l_{bce} 和 l_{dice} 的均值。

对于教师网络, 使用 EMA^[23] 更新其参数权重, 如式(2)所示。

$$\theta_t = \alpha \theta_t + (1 - \alpha) \theta_s \quad (2)$$

式中: θ_t 是教师网络的参数; θ_s 是学生网络的参数; α 是控制更新速率的 EMA 衰减系数。

1.2 多尺度卷积注意力机制

为了生成更加稳定的中间预测, MSC-SK

DACarriesNet 使用 MSCA 提取的特征对学生网络进行直接和深度监督,如图 2 所示,这里用 k_1 表示卷积核的

第 1 个维度, k_2 表示卷积核的第 2 个维度, $k_1 \times k_2$ 表示使用 $k_1 \times k_2$ 大小的深度可分离卷积核。

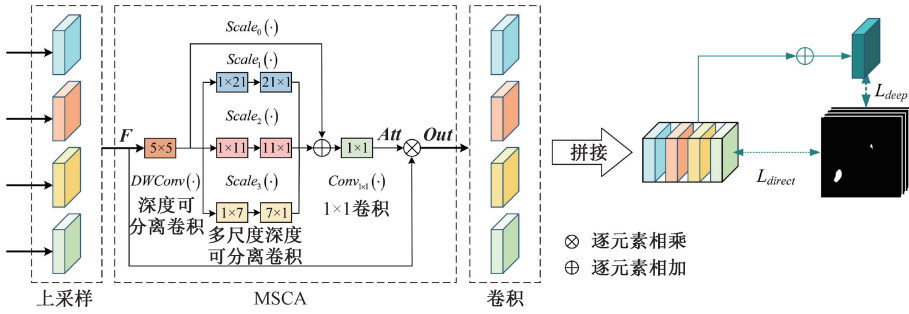


图 2 基于 MSCA 的直接和深度监督示意图

Fig. 2 Diagram of direct and deep supervision based on MSCA

MSCA 主要由 3 部分组成:用于聚合局部信息的深度可分离卷积、捕捉多尺度上下文的多分支结构,以及用于信息融合的 1×1 卷积。为获得不同尺度的特征表示, MSCA 采用了 3 种不同尺寸的深度可分离卷积核,分别为 7×7 、 11×11 和 21×21 。每个尺度的卷积核专门负责捕捉该尺度的特征信息。为了减少计算量,每个卷积核都被分解成两个更小的卷积核,例如将 7×7 卷积核分解为 1×7 和 7×1 的卷积核。这样的分解在保持原始卷积核效果的同时,大大降低了计算复杂度。通过这些多尺度卷积操作, MSCA 能够有效地捕捉图像中的多尺度信息,其具体操作如式(3)所示。

$$Att = Conv_{1 \times 1} \left(\sum_{i=0}^3 Scale_i(DWConv(F)) \right) \quad (3)$$

式中: F 表示输入特征; Att 是注意力图; $DWConv(\cdot)$ 表示深度可分离卷积; $Scale_i(\cdot)$ 中 $i \in \{0, 1, 2, 3\}$, 表示不同尺度的深度可分离卷积; $Scale_0$ 是恒等映射; $Scale_1$, $Scale_2$, $Scale_3$ 分别是 7×7 、 11×11 和 21×21 的卷积操作。随后, 1×1 卷积用于聚合信息和重新加权。它的输出直接作为注意力权重,重新加权 MSCA 的输入特征。其具体操作如式(4)所示。

$$Out = Att \otimes F \quad (4)$$

式中: Out 表示输出; $\otimes$ 是逐元素相乘操作。该加权过程使 MSCA 通过加权增强对关键特征和通道的关注,从而提升特征表达能力。其将不同尺度的特征图逐元素相加融合,经 1×1 卷积聚合信息后反馈至原始特征图,丰富特征多样性与语义完整性。具体而言,解码器输出的多尺度特征输入 MSCA 后,经卷积提取注意力权重以加权输入特征,生成输出特征图。该输出不仅为学生网络提供直接监督,也通过逐元素相加实现深度监督,从而多层次引导学生网络学习。

1.3 多级蒙特卡洛选择性核注意力机制

为了提升不确定性计算, MSC-SK DACarriesNet 使用 SKA 所示的方式融合了上述获得的所有预测。SKA 是一种由多个选择性核单元组成的视觉注意力机制。由于选择性核单元可以融合不同大小卷积核适应不同尺度目标得到的特征图,所以使得 SKA 获得能够自适应调节感受野的能力,进而使得目标特征信息提取更加精确。SKA 对图像中不同目标相异特征信息的提取主要包括 3 个步骤,如图 3 所示。

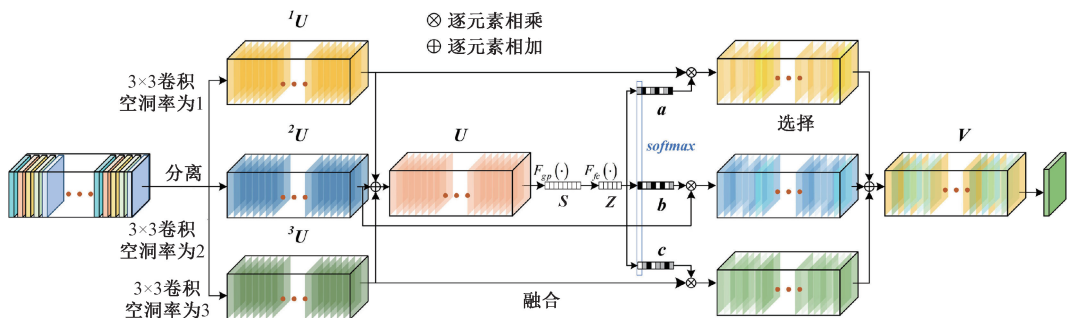


图 3 SKA 示意图

Fig. 3 SKA architectural diagram

分离:将输入的特征图像 $\mathbf{X} \in \mathbf{R}^{H \times W \times C}$ 分离,其中 $\mathbf{R}$ 、 $\mathbf{H}$ 、 $\mathbf{W}$ 、 $\mathbf{C}$ 分别为特征图像集、特征图高、宽和通道数,并行采用空洞率为1、2和3,卷积核大小为 3×3 的空洞卷积进行特征提取,得到感受野分别为 3×3 、 5×5 和 7×7 的特征图 ${}^1\mathbf{U}$ 、 ${}^2\mathbf{U}$ 和 ${}^3\mathbf{U}$ 。融合:将分离步骤中得到的 ${}^1\mathbf{U}$ 、 ${}^2\mathbf{U}$ 和 ${}^3\mathbf{U}$ 融合,即 $\mathbf{U} = {}^1\mathbf{U} + {}^2\mathbf{U} + {}^3\mathbf{U}$,进而将 $\mathbf{U}$ 全局平均池化得到统计各通道特征信息的特征向量 $\mathbf{S}$,如式(5)所示。随后特征向量 $\mathbf{S}$ 被输入至全连接层进行降维后得到向量 $\mathbf{Z}$,如式(6)所示。

$$\mathbf{S} = F_{gp}(\mathbf{U}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{U}(i, j) \quad (5)$$

$$\mathbf{Z} = F_{fc}(\mathbf{S}) = \delta(BN(\mathbf{W}_{fc} \cdot \mathbf{S})) \quad (6)$$

式中: $F_{gp}(\cdot)$ 表示全局池化; $F_{fc}(\cdot)$ 表示全连接层; $BN(\cdot)$ 表示批量归一化; $\delta(\cdot)$ 表示 ReLU 激活函数。选择:向量 $\mathbf{Z}$ 经过全连接拓展后得到3个向量,在其通道维度上由 softmax 函数求得两者对应的注意力机制权重向量 $\mathbf{a}$ 、 $\mathbf{b}$ 和 $\mathbf{c}$ 。将 $\mathbf{a}$ 、 $\mathbf{b}$ 和 $\mathbf{c}$ 分别与 ${}^1\mathbf{U}$ 、 ${}^2\mathbf{U}$ 和 ${}^3\mathbf{U}$ 相乘加权,对输出的特征图结果融合即得到特征信息加强后的向量 $\mathbf{V}$,如式(7)所示。

$$\mathbf{V} = \mathbf{a} \cdot {}^1\mathbf{U} + \mathbf{b} \cdot {}^2\mathbf{U} + \mathbf{c} \cdot {}^3\mathbf{U} \quad (7)$$

然后,收集 T 次蒙特卡洛法采样教师网络的输出,顺序拼接成一个 $T \times (L + 1) \times H \times W \times C$ 的张量。使用 SKA 融合所有输出,表示为 $\mathbf{y}_{ska}$,如式(8)所示。

$$\mathbf{y}_{ska} = f_{ska}(\hat{\mathbf{y}}_g, \hat{\mathbf{Y}}_g) \quad (8)$$

式中: $\hat{\mathbf{y}}_g$ 和 $\hat{\mathbf{Y}}_g$ 表示教师网络的直接和多层预测; $f_{ska}(\cdot)$ 表示选择性核注意力机制。随后,用熵来表示图像的不确定性,用 $\mathbf{M}_{uncertain}$ 表示 $\mathbf{y}_{ska}$ 的不确定性特征图,如式(9)所示。考虑到模型的预测随着训练的进行逐渐可靠,通常使用动态阈值自适应地增加确定性区域的范围,它由式(10)确定,即:

$$\mathbf{M}_{uncertain} = -2 \cdot \mathbf{y}_{ska} \cdot \log(\mathbf{y}_{ska}) \quad (9)$$

$$Thresh = \gamma \cdot (\beta + (1 - \beta) \cdot e^{-5 \cdot (1 - \frac{\varepsilon}{E})^2}) \quad (10)$$

式中: γ 和 β 为调整阈值的两个超参数, ε 和 E 分别为当前和最大训练轮数。选择不确定性特征图中小于当前训练阈值的部分作为确定性掩码特征图。从而确定了一致性损失,如式(11)所示。

$$L_{con} = \frac{\sum f_T(p < Thresh) \cdot \|\hat{\mathbf{y}}'_f - \hat{\mathbf{y}}'_g\|^2}{\sum f_T(p < Thresh)} \quad (11)$$

式中: p 表示 $\mathbf{M}_{uncertain}$ 中的像素值; $f_T(\cdot)$ 表示阈值激活函数,当 $p < Thresh$ 时为0, $p \geq Thresh$ 时为1。添加噪声样本的学生网络和教师网络的直接预测分别用 $\hat{\mathbf{y}}'_f$ 和 $\hat{\mathbf{y}}'_g$ 表示。最终得到总的损失如式(12)所示。

$$L = L_{deep} + L_{direct} + \lambda \cdot L_{con} \quad (12)$$

式中: λ 是标记和未标记训练过程之间权衡损失函数的平衡因子。这里, λ 的值定义如式(13)所示。

$$\lambda = \omega \cdot e^{-5 \cdot (1 - \frac{\varepsilon}{E})^2} \quad (13)$$

式中: ω 取一个相对较小的固定值,以确保在早期训练不稳定时,损失不会给学生网络带来麻烦。最后,将所有数据流放在了算法1中。

算法1: MSC-SK DACariesNet

输入:标注数据集 $\mathbf{D}_l$,未标注数据集 $\mathbf{D}_u$

输出:学生网络 $f(\cdot)$ 的参数 θ_s

1:初始化学生和教师网络 $f(\cdot)$, $g(\cdot)$ 的参数 θ_s, θ_t

2:从 $\mathbf{D}_l$ 采样 $(\mathbf{x}_i, \mathbf{y}_i)$, $\mathbf{D}_u$ 采样 $\mathbf{x}_j$

3:while 不收敛 do

4: $t \leftarrow 0$;

5: $\hat{\mathbf{y}}_f, \hat{\mathbf{Y}}_f \leftarrow f(\mathbf{x}_i)$;

6: $L_{direct} \leftarrow (\hat{\mathbf{y}}_f - \mathbf{y}_i)^2$;

7: $L_{deep} \leftarrow (\hat{\mathbf{Y}}_f - \mathbf{y}_i)^2$;

8: $\hat{\mathbf{y}}'_f \leftarrow f(\mathbf{x}_j + \text{noise}_l)$;

9: $\hat{\mathbf{y}}'_g \leftarrow g(\mathbf{x}_j + \text{noise}_u)$;

10: $\text{volume} \leftarrow \text{zeros_like}(T \cdot L, C, H, W)$;

11: for $t < T$ do

12: $\hat{\mathbf{y}}'_g, \hat{\mathbf{Y}}'_g \leftarrow g(\mathbf{x}_j + \text{noise}_t)$;

13: volume 拼接 $\hat{\mathbf{y}}'_g, \hat{\mathbf{Y}}'_g$;

14: $t \leftarrow t + 1$;

15: end

16: $\hat{\mathbf{y}}_g^{fusion} \leftarrow f_{ska}(\text{volume})$;

17: 通过式(9)求得 $\mathbf{M}_{uncertain}$;

18: $\mathbf{M}_{certain} \leftarrow \mathbf{M}_{uncertain} < Thresh$;

19: $\text{dist} \leftarrow \|\hat{\mathbf{y}}'_f - \hat{\mathbf{y}}'_g\|^2$;

20: $L_{con} \leftarrow \text{mean}(\mathbf{M}_{certain} \cdot \text{dist})$;

21: 通过 $L_{direct}, L_{deep}, L_{con}$ 更新学生网络 θ_s ;

22: 通过式(2)更新教师网络 θ_t ;

23: end

2 实验结果和分析

2.1 数据集

数据集1:采用的 DC1000^[12]是唯一公开的包含精确的像素级分割标签的龋齿分割数据集,该数据集包含1000张全景X射线图像,分为593张详细注释图像和407张粗略注释图像,宽度为2943 pixels,高度为1435 pixels。选取497个全景X射线图像作为训练集,为了提高的注释质量,将训练数据集裁剪成384×384的正方形区域,生成1460个有监督切片,929个无监督切片。此外,测试集选取100个全景X射线图像作为评估数据。

数据集2:该数据集基于公开牙科全景X射线图像数据集 Dental OPG XRAY Dataset^[24](具体网址见

<https://data.mendeley.com/datasets/c4hhrkxytw/4>) 构建, 原始图像来源于 3 家不同的牙科诊所, 共包含 119 张龋齿相关图像。由于图像尺寸不统一, 首先对图像进行了预处理, 通过中心裁剪或上采样的方式将其标准化为统一分辨率大小 $2\,943\text{ pixels}\times1\,435\text{ pixels}$ 。原始数据集中并未提供龋齿的像素级分割标注, 因此在专业牙科医师的指导下, 重新对龋齿区域进行了像素级分割标注。其中选取 100 张图像用于训练, 19 张图像用于测试。为了构建监督学习所需的训练样本, 本研究进一步将训练图像裁剪为 384×384 的正方形区域, 最终生成 290 个有监督切片和 180 个无监督切片。

2.2 实验环境和参数设置

实验训练环境为装有 Win 11 操作系统、Intel Core i7-13700K 处理器, 32 GB 内存和 NVIDIA GeForce RTX2060 Super GPU 的计算平台。编程语言为 Python, 深度学习框架为 PyTorch Lightning。MSC-SK DACariesNet 模型训练使用学习率为 0.001 的 AdamW 优化器, 动量和权重衰减系数分别设置为 0.9 和 0.001。全监督学习的批量大小设置为 4, 半监督的批量大小设置 8。在批量大小为 8 的半监督样本中, 监督样本和不监督样本各占 4 个批量大小。此外, 将最大训练轮数设置为 200, 将多层解码器的层数 L 设置为 4。此外, α 、 β 、 ω 这 3 个超参数分别设置为 0.99、0.75 和 0.1。

2.3 评估指标

为了定量评价网络分割性能, 使用了 3 种常用的评价指标, 包括 Dice 系数、灵敏度和查准率^[12,25]。在数据集 1 中, 设计了 2 组实验, 分别在 265 个切片和 530 个切片下对全监督与半监督方法进行了性能评估, 评价指标分别记为 $DICE_{265}$ 、 SEN_{265} 、 PRE_{265} 以及 $DICE_{530}$ 、 SEN_{530} 、 PRE_{530} 。在数据集 2 中, 由于样本规模较小, 仅在 265 个切片下进行了全监督与半监督的性能评估。

2.4 对比试验及结果分析

为了验证 MSC-SK DACariesNet 的综合性能, 在数据集 1 和 2 上将全监督和半监督分割算法在牙齿分割性能方面进行了比较, 实验结果如表 1 和 2 所示。对比算法包括全监督分割算法 Unet^[22]和 Unet++^[26]以及半监督分割算法不确定性校正的金字塔一致性^[27] (uncertainty rectified pyramid consistency, URPC) (生成金字塔预测, 用未标记的图像引导多尺度一致性损失)、不确定性感知的均值教师^[23] (uncertainty aware mean teacher, UAMT) (对学生-教师双网络架构进行一致性损失和分割损失训练)、跨层级对比学习与一致性约束^[28] (cross-level contrastive learning and consistency constraint, CLCC) (对切片前后图像进行一致性约束) 和多级不确定性感知^[12] (multi-level uncertainty aware, MLUA) (多级不确定性感

知对解码器进行深度监督)。在两个数据集上的所有算法均采用与 MSC-SK DACariesNet 相同的训练集和测试集设置, 确保对比结果的公平性与一致性。

表 1 数据集 1 上的经典分割网络和半监督分割网络性能对比
Table 1 Performance comparison between classical segmentation networks and semi-supervised segmentation networks on dataset 1

模型	$DICE_{265}$	SEN_{265}	PRE_{265}	$DICE_{530}$	SEN_{530}	PRE_{530}
Unet ^[22]	42.78	44.18	49.94	53.81	58.34	57.20
Unet++ ^[26]	42.89	43.54	52.76	52.97	53.68	60.62
URPC ^[27]	50.51	46.37	67.67	59.92	52.07	81.36
UAMT ^[23]	57.91	55.55	67.91	67.43	65.83	74.19
CLCC ^[28]	57.93	57.53	62.55	68.24	66.31	73.87
MLUA ^[12]	61.14	59.04	69.16	69.54	68.43	73.71
本文	65.05	61.18	74.51	70.93	69.20	76.03

表 2 数据集 2 上的经典分割网络和半监督分割网络性能对比
Table 2 Performance comparison between classical segmentation networks and semi-supervised segmentation networks on dataset 2

模型	$DICE_{265}$	SEN_{265}	PRE_{265}
Unet ^[22]	42.08	43.86	49.72
Unet++ ^[26]	43.75	44.71	53.59
URPC ^[27]	49.58	42.77	60.72
UAMT ^[23]	56.34	54.42	67.64
CLCC ^[28]	56.91	57.41	62.39
MLUA ^[12]	61.06	58.27	69.71
本文	64.95	61.13	74.68

由表 1 和 2 可见, 在数据集 1 和 2 的训练效果中, U-Net 和 U-Net++ 两种全监督分割算法的性能整体最差, URPC 略优于这两者。UAMT 和 CLCC 表现较为接近, 均优于 URPC, 说明引入一致性损失机制对分割性能有所提升。MLUA 的性能虽进一步提升, 优于前述几种方法, 但仍明显低于本研究提出的 MSC-SK DACariesNet。这一趋势在两个数据集上均表现一致, 充分验证了 MSC-SK DACariesNet 在龋齿分割任务中具备出色的鲁棒性与准确性, 能够有效应对边界模糊、组织特征相似等复杂情况, 在区分病变与正常区域方面表现显著优于现有方法。

数据集 1 和 2 上龋齿病变区域分割结果如图 4 和 5 所示。选择了一些具有挑战性的全景 X 射线图像切片样本进行可视化对比, 这些图像中龋齿病变区域与正常牙齿组织之间的边界模糊、像素特征相似, 极易导致过分割或欠分割等问题。在这些具有代表性的复杂样例中,

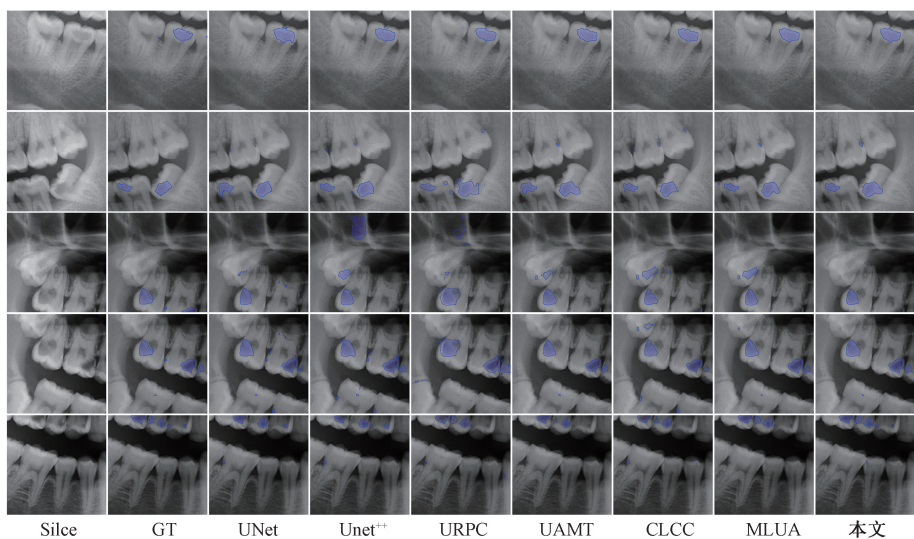


图 4 数据集 1 上的网络分割效果对比

Fig. 4 Comparison diagram of network segmentation result on dataset 1

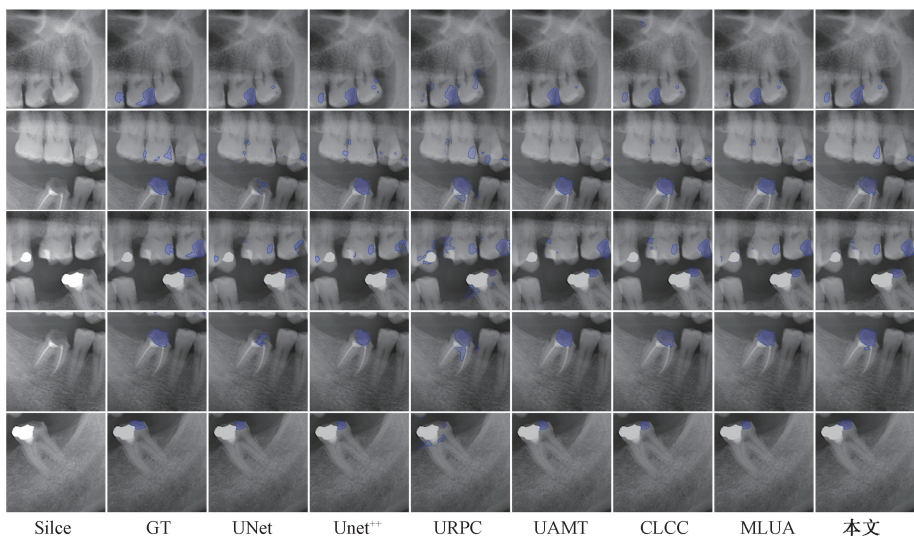


图 5 数据集 2 上的网络分割效果对比

Fig. 5 Comparison diagram of network segmentation result on dataset 2

对比测试显示:尽管所有方法在牙齿紧密交界区域普遍存在一定程度的过分割现象,且在与软组织过渡区域可能出现欠分割,但 MSC-SK DACariesNet 在两个数据集上的表现均更为稳健,能够更准确地定位并分割龋齿病变区域,显示出更强的鲁棒性和泛化能力。

2.5 消融实验结果及分析

为了验证所提出方法的有效性,以 MLUA 为基础模型,设计了 4 组消融实验,系统分析 2 种注意力机制(MSCA 与 SKA)对模型性能的影响。本研究在数据集 1 和 2 上均进行了实验验证,以确保评估结果的全面性与泛化能力的体现。在实验设置中,当仅引入 MSCA 模块时, MSC-SK DACariesNet 采用平均池化融合多尺度采样

的预测结果;当仅引入 SKA 模块时,采用卷积操作提取解码器生成的特征。消融实验结果如表 3 和 4 所示,清晰揭示了 MSCA 与 SKA 模块对模型性能的提升作用。在数据集 1 和 2 的 265 个切片上,双注意力机制展现出显著增益。在数据集 1 上单独引入 MSCA 时, Dice 系数由 61.14% 提升至 63.77%, 提升了 2.63%; 在数据集 2 上 Dice 系数由 61.06% 提升至 62.89%, 提升了 1.83%, 表明 MSCA 模块在增强边界细节感知和提高判别能力方面的有效性。此外,在数据集 1 上单独引入 SKA 时, Dice 系数由 61.14% 提升至 63.50%, 提升了 2.36%。在数据集 2 上 Dice 系数由 61.06% 提升至 63.07%, 提升了 2.01%, 显示其对不确定区域的自适应增强能力。进一

表 3 数据集 1 上的两种注意力机制下的消融实验
Table 3 Ablation study under two attention mechanisms on dataset 1

模型	$DICE_{265}$	SEN_{265}	PRE_{265}	$DICE_{530}$	SEN_{530}	PRE_{530}
MLUA	61. 14	59. 04	69. 16	69. 54	68. 43	73. 71
MLUA+MSCA	63. 77	60. 99	72. 26	68. 60	67. 35	74. 67
MLUA+SKA	63. 50	62. 66	69. 18	70. 22	70. 86	72. 44
MLUA+MSCA+SKA	65. 05 ↑	61. 18	74. 51 ↑	70. 93 ↑	69. 20	76. 03 ↑

表 4 数据集 2 上的两种注意力机制下的消融实验

Table 4 Ablation study under two attention mechanisms on dataset 2

模型	$DICE_{265}$	SEN_{265}	PRE_{265}
MLUA	61. 06	58. 27	69. 71
MLUA+MSCA	62. 89	61. 14	71. 34
MLUA+SKA	63. 07	61. 83	69. 77
MLUA+MSCA+SKA	64. 95 ↑	61. 13	74. 68 ↑

步地,当同时引入 MSCA 与 SKA 两个注意力机制时,模型性能进一步提升。在数据集 1 中 Dice 系数由 61. 14% 提升至 65. 05%,提升了 3. 91%;在数据集 2 中 Dice 系数由 61. 06% 提升至 64. 95%,提升了 3. 89%,进一步验证了双注意力融合策略在区分龋齿病变区域与背景干扰区域方面的有效性与稳健性。

在此基础上,在数据集 1 的 530 个切片上,双注意力机制协同效应更为突出。单独引入 MSCA 时 Dice 系数由 69. 54% 下降至 68. 60%,下降了 0. 94%。单独引入 SKA 时 Dice 系数由 69. 54% 提升至 70. 22%,提升了 0. 68%。而同时引入 MSCA 与 SKA 两个注意力机制时 Dice 系数由 69. 54% 提升至 70. 93%,提升了 1. 39%。MSCA 提升了网络的 Dice 系数和查准率,SKA 改善了网络的灵敏度,二者同时引入时 SKA 有效补偿了 MSCA 的灵敏度损失,使得最终 Dice 系数,灵敏度和查准率较基线均有显著提升,说明双注意力融合策略在大规模数据下具有更强的稳定性和泛化能力。

此外,MSC-SK DACariesNet 展现出出色的大数据适应性。在数据集 1 中,当训练数据规模从 265 切片扩展至 530 切片时,Dice 系数由 61. 14% 提升至 70. 93%,提升了 9. 84%;灵敏度由 58. 21% 提升至 70. 55%,提升了 12. 34%;查准率由 63. 45% 提升至 74. 72%,提升了 11. 27%,验证了算法在更大规模数据集上的稳定优化能力。类似的趋势也在数据集 2 中得到验证,进一步印证了 MSC-SK DACariesNet 在不同数据来源与样本分布条件下,均具备良好的可扩展性和鲁棒性。

3 结 论

针对全景 X 射线图像龋齿病变对比度低、边界模糊、尺度多变等特点,提出了一种基于 MSC-SK DACariesNet 的全景 X 射线图像龋齿分割方法。MSC-SK DACariesNet 利用小部分标记图像来训练教师模型,利用大量未标记图像来训练学生模型,提供了计算和性能增益。此外,所提出的模型增加了多尺度卷积和选择性核双重注意力机制来提高其对各种尺度龋齿的分割性能。为验证方法的有效性,在数据集 1 和 2 上进行了大量实验。结果表明,Dice 系数、查准率和灵敏度相较于经典的全监督学习算法分别提高了 22. 16% ~ 22. 27%、17% ~ 17. 64% 和 21. 75% ~ 24. 57%,相较于最近的半监督学习算法分别提高了 3. 91% ~ 14. 54%、2. 14% ~ 14. 81% 和 5. 35% ~ 11. 96%,相较于添加注意力机制前的算法,分别提高了 3. 91%、2. 14% 和 5. 35%。综上所述,所提出的方法在全监督和半监督框架中达到了较好的性能,并在不同数据来源、样本数量及标注精度条件下,始终保持稳定的性能优势,展现出良好的鲁棒性和广泛适用性。此外,提出的分割模型可以集成部署在相关医学影像勾画平台,以提升计算机辅助口腔诊疗的精度和效率。

参考文献

[1] XU Y, YOU Y, YI L Y, et al. Dental plaque-inspired versatile nanosystem for caries prevention and tooth restoration[J]. Bioactive Materials, 2023, 20: 418-433.

[2] MAJANGA V, VIRIRI S. A survey of dental caries segmentation and detection techniques [J]. Scientific World Journal, 2022, 2022: 1-19.

[3] 蔡玉芳, 王涵, 李琦, 等. 联合自然梯度和 AdamW 算法的 RSF 图像分割模型[J]. 仪器仪表学报, 2023, 44(3): 261-270.

CAI Y F, WANG H, LI Q, et al. RSF image segmentation model joint natural gradient and AdamW algorithm[J]. Chinese Journal of Scientific Instrument, 2023, 44(3): 261-270.

- [4] 于天河, 徐博超, 侯善冲, 等. 基于图像的道路裂缝分割及量化方法研究[J]. 仪器仪表学报, 2024, 45(9): 77-91.
YU T H, XU B CH, HUO SH CH, et al. Research on image-based segmentation and quantification of road cracks[J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 77-91.
- [5] ZHAO Y, LI P CH, GAO CH Q, et al. TSASNet: Tooth segmentation on dental panoramic X-ray images by two-stage attention segmentation network[J]. Knowledge-Based Systems, 2020, 206: 106338.
- [6] YING S, WANG B W, ZHU H H, et al. Caries segmentation on tooth X-ray images with a deep network[J]. Journal of Dentistry, 2022, 119: 104076.
- [7] HAGHANIFAR A, MAJDABADI M M, HAGHANIFAR S, et al. PaXNet: Tooth segmentation and dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier[J]. Multimedia Tools and Applications, 2023, 82(18): 27659-27679.
- [8] ZHU H H, CAO ZH, LIAN L Y, et al. CariesNet: A deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image[J]. Neural Computing and Applications, 2023, 35(22): 16051-16059.
- [9] MĂRGINEAN A C, MUREȘANU S, HEDEȘIU M, et al. Teeth segmentation and carious lesions segmentation in panoramic X-ray images using CariSeg, a networks' ensemble[J]. Heliyon, 2024, 10(10): e30836.
- [10] SON L H, TUAN T M, FUJITA H, et al. Dental diagnosis from X-Ray images: An expert system based on fuzzy computing[J]. Biomedical Signal Processing and Control, 2018, 39: 64-73.
- [11] QAYYUM A, TAHIR A, BUTT M A, et al. Dental caries detection using a semi-supervised learning approach[J]. Scientific Reports, 2023, 13(1): 749.
- [12] WANG X Y, GAO S ZH, JIANG K SH, et al. Multi-level uncertainty aware learning for semi-supervised dental panoramic caries segmentation[J]. Neurocomputing, 2023, 540: 126208.
- [13] LI X, FU CH, WANG Q, et al. DMSA-UNet: Dual multi-scale attention makes U-Net more strong for medical image segmentation[J]. Knowledge-Based Systems, 2024, 299: 112050.
- [14] 王琦, 张涛, 徐超伟, 等. 多尺度注意力融合与视觉Transformer方法优化的电阻抗层析成像深度学习方法[J]. 仪器仪表学报, 2024, 45(7): 52-63.
WANG Q, ZHANG T, XU CH W, et al. Optimized learning method for electrical impedance tomography with multi-scale attention fusion and vision Transformer[J]. Chinese Journal of Scientific Instrument, 2024, 45(7): 52-63.
- [15] GUO M H, LU CH Z, HOU Q B, et al. SegNeXt: Rethinking convolutional attention design for semantic segmentation[C]. 36th International Conference on Neural Information Processing Systems, 2022, 35: 1140-1156.
- [16] LI X, WANG W H, HU X L, et al. Selective Kernel Networks[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510-519.
- [17] 陈仁祥, 张晓, 张旭, 等. 改进掩码自编码器的滚动轴承半监督故障诊断[J]. 仪器仪表学报, 2024, 44(1): 26-33.
CHEN R X, ZHANG X, ZHANG X, et al. Labeled sample augmentation based on deep embedding relation space for semi-supervised fault diagnosis of gear box[J]. Chinese Journal of Scientific Instrument, 2024, 44(1): 26-33.
- [18] WANG X Y, YUAN Y W, GUO D Y, et al. SSA-Net: Spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning[J]. Medical Image Analysis, 2022, 79: 102459.
- [19] CHEN J M, YANG M, LING J. Attention-based label consistency for semi-supervised deep learning based image classification[J]. Neurocomputing, 2021, 453: 731-741.
- [20] MAI X Y, COUILLET R. Consistent semi-supervised graph regularization for high dimensional data[J]. Journal of Machine Learning Research, 2021, 22(94): 1-48.
- [21] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition[C]. 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [22] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]. Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [23] YU L Q, WANG SH J, LI X M, et al. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation[C]. Medical Image Computing and Computer Assisted Intervention, 2019: 605-613.
- [24] RAHMAN R B, TANIM S A, ALFAZ N, et al. Dental

OPG XRAY dataset [DS/OL]. (2024-08-27) [2025-05-10].

[25] KUMARI A , RAO S N , RAMANA P R . Design of hybrid dental caries segmentation and caries detection with meta-heuristic-based ResneXt-RNN [J] . Biomedical Signal Processing and Control , 2022 , 78 : 103961 .

[26] ZHOU Z W , SIDDIQUEE M R , TAJBAKHSN , et al . UNet++ : A nested U-Net architecture for medical image segmentation [C] . Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support , 2018 : 3-11 .

[27] LUO X D , WANG G T , LIAO W J , et al . Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency [J] . Medical Image Analysis , 2022 , 80 : 102517 .

[28] ZHAO X K , FANG CH W , FAN D J , et al . Cross-level contrastive learning and consistency constraint for semi-supervised medical image segmentation [C] . 2022 IEEE 19th International Symposium on Biomedical Imaging , 2022 : 1-5 .

作者简介



薛钟毫,2021 年于哈尔滨理工大学获得学士学位,现为哈尔滨理工大学博士研究生,主要研究方向为医疗机器人。
E-mail: 2220110128@ stu. hrbust. edu. cn
Xue Zhonghao received the B. Sc. degree

in the Mechatronic Engineering from Harbin University of Science and Technology in 2021, where he is currently a doctoral student. His main research interest includes medical robot.



姜金刚(通信作者),分别于 2005、2008 和 2013 年于哈尔滨理工大学机械电子工程专业获得学士、硕士和博士学位,现为哈尔滨理工大学教授、博士生导师,主要研究方向为医疗机器人和仿生机器人。
E-mail: jiangjingang@ hrbust. edu. cn

Jiang Jingang (Corresponding author) received the B. Sc. , M. Sc. and Ph. D. degrees in the Mechatronic Engineering from Harbin University of Science and Technology in 2005, 2008 and 2013, respectively. Now he is a professor at Harbin University of Science and Technology. His current research interests include medical robot and biomimetic robot.



潘洁,分别于 2002 年和 2008 年于北京大学口腔医学院获得医学博士和牙科博士,现为北京大学口腔医学院主任医师,主要研究方向为牙髓病学。
E-mail: panjie1972@ 163. com

Pan Jie received the M. D. and DDS. degrees in stomatology from the School of Stomatology, Peking University in 2002 and 2008, respectively. Now she is the chief physician with the School of Stomatology, Peking University. Her main research interests include endodontics.