结合注意力机制和邻域增强机制的牙颌模型分割网络

刘志华[1,2)], 薛久涛[1,2)], 唐浩[3)], 廖与禾[1,2)*]

[1)] (西安交通大学现代设计及转子轴承系统教育部重点实验室 西安 710049)
[2)] (西安交通大学陕西省机械产品质量保障与诊断重点实验室 西安 710049)
[3)] (重庆科技大学机械与动力工程学院 重庆 401331)
(yhliao@xjtu.edu.cn)

摘 要: 针对现有牙齿模型深度学习分割网络过度依赖坐标特征, 忽略了几何特征与坐标特征的差异性和互补作用的问题, 提出一种结合注意力机制和邻域增强机制的牙颌模型分割网络. 首先采用一个注意力增强模块, 通过不同的编码方式对输入特征进行邻域增强, 并结合几何特征与坐标特征的性质分别设计差值注意力机制和自注意力机制, 充分提取网格的上下文特征; 其次构建互补性池化函数, 融合最大池化、注意力池化以及平均池化促进不同特征间的互补作用, 提高分割网络的鲁棒性; 最后嵌入自适应注意力模块, 在平衡多尺度特征间的数值差异的同时获得综合特征来实现网格分割. 在 200 个真实患者的牙颌模型数据集上与其他网络进行对比的实验结果表明, 所提网络针对牙颌模型的分割准确率达 96.18%; 该网络优于对比的深度学习网络, 具有更优的牙颌模型分割性能.

关键词: 三维牙颌模型分割; 邻域增强; 语义分割网络; 注意力机制
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024-00037

Segmentation Network of Dental Mesh Model Based on Attention Mechanism and Neighbor Enhancement

Liu Zhihua[1,2)], Xue Jiutao[1,2)], Tang Hao[3)], and Liao Yuhe[1,2)*]

[1)] (*Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an* 710049)
[2)] (*Shaanxi Key Laboratory of Mechanical Product Quality Assurance and Diagnostics, Xi'an Jiaotong University, Xi'an* 710049)
[3)] (*School of Mechanical and Power Engineering, Chongqing University of Science and Technology, Chongqing* 401331)

Abstract: To address the issue of existing deep learning-based segmentation methods for dental models, which overly rely on coordinate features and overlook the differences and complementary nature of geometric and coordinate features, a segmentation network based on attention mechanism and neighbor enhancement is proposed. First, an attention-enhanced module is employed to enhance the neighborhood of input features through different encoding schemes. It designs a differential attention mechanism and a self-attention mechanism to extract comprehensive contextual features of the mesh by considering the properties of geometric and coordinate features separately. Next, a complementary pooling function is constructed to fuse maximum pooling, attention pooling, and average pooling, which promotes the complementary interaction among different features and enhances the robustness of the segmentation network. Finally, an adaptive attention module is embedded to balance the numerical differences among multi-scale features for effective mesh segmentation. Compared with other methods in 200 real patients dental model dataset, the segmenta-

收稿日期: 2024-01-17; 修回日期: 2024-05-15. **基金项目**: 国家重点研发计划(2019YFB1311903); 国家自然科学基金面上项目(51575424). **刘志华**(2000—), 男, 硕士研究生, 主要研究方向为计算机视觉; **薛久涛**(1995—), 男, 博士研究生, 主要研究方向为机械系统故障诊断; **唐浩**(1983—), 男, 博士, 正高级工程师, 硕士生导师, 主要研究方向为工业数据挖掘; **廖与禾**(1973—), 男, 博士, 副教授, 博士生导师, 论文通信作者, 主要研究方向为机械系统故障诊断.

tion accuracy of dental model is 96.18%. The experimental results show that the proposed method is superior to the compared deep learning networks and has better performance in the segmentation of dental models.

Key words: three-dimensional dental model segmentation; neighbor enhancement; semantic segmentation network; attention mechanism

牙齿畸形是常见的口腔疾病之一[1], 不仅影响患者的面容, 还可能引发胃肠道疾病, 对身体健康造成损害. 近年来, 随着数字图像技术和计算机技术的发展, CAD 系统在口腔医学领域得到了广泛的应用. CAD 系统基于口腔内扫描仪(intraoral optical scanners, IOS)获得患者的三维牙颌数字模型[2], 实现牙齿的自动分割、修复、矫正和重排等后续任务[3], 极大地提高了牙医的治疗效率. 本文研究由 IOS 获取的三维牙齿网格模型的分割问题.

与传统的 CT 图像分割相比, 三维牙颌模型分割过程中存在许多挑战: (1) 同一类型牙齿的几何特征因患者个体差异而异; (2) 牙颌模型中存在牙齿拥挤和排列不齐的问题, 导致分割边界难以确定; (3) 缺牙、断牙、智齿等问题存在不确定性, 严重破坏网格的几何特征.

为解决上述难题, 传统的分割方法通常是通过预先选择的几何特征(坐标[4]、法向量[5]、曲率[6]、谐波场[7])对模型进行分割, 或者将其投影到二维平面上[8-10], 再通过图像分割算法进行分割; 但这些方法不仅在模型投影时会降低分辨率, 还极度依赖网格特征的完整性, 当牙颌模型的几何特征因畸形而被破坏时, 将降低算法的鲁棒性.

近年来, 深度学习和加速图形处理技术的发展为牙齿模型的全自动分割奠定了基础. 一些方法将三维模型结构转化为二维图像[11-13], 使用卷积网络进行网格的特征提取和类型预测. 这些方法的网络结构较为简洁, 但忽略了网格之间的拓扑关系, 容易出现离散错误. 为此, 一些学者将特征提取模块应用到牙齿分割网络中[14-17], 将网格的邻域关系融入特征提取过程中; 但这些模块在特征融合过程中过度依赖坐标特征, 忽略了几何特征和语义特征之间的差异, 导致网络无法提取丰富的局部特征, 影响在牙齿边界处的分割精度.

不同于上述网络模型, 本文提出基于注意力机制和邻域增强机制的网络(attention and neighborhood enhancement network, ATNE-NET)包括设计一个注意力增强模块(attention-enhanced module, AEM), 通过不同的编码方式对输入特征进行邻域进行邻域增强, 同时结合几何特征与坐标特征的性质, 分别设计差值注意力模块(differential attention-enhanced module, DAEM)和自注意力模块(self-attention- enhanced module, SAEM), 以充分提取网格的上下文特征; 构建互补性池化模块(complementary pooling module, CPM), 融合最大池化(max pooling, MP)、注意力池化以及平均池化(average pooling, AP)促进不同特征间的互补作用, 动态地调整网络对不同特征的关注程度, 以提高分割网络的鲁棒性; 在跳跃连接的特征融合过程中嵌入自适应注意力模块, 以平衡多尺度特征间的数值差异, 实现网格分割.

1 相关工作

1.1 三维形状分割

深度学习在三维形状分割中得到广泛应用, 并取得了一定的效果. 一些方法将三维模型转换为体素数据[18], 并使用卷积模块进行特征学习, 它们在初始分割中取得了较好的分割精度, 然而其在空间变换的过程中不可避免地会降低模型分辨率, 制约了精度的进一步提升. 为了解决上述问题, PointNet++[19]使用多层感知器(multilayer perceptron, MLP)和全局最大池化从几何特征中提取三维形状的局部和全局特征.

图卷积神经网络由于可以融合三维形状的邻域信息而取得了较好的分割效果. Zhang 等[20]提出动态图卷积神经网络(dynamic graph convolutional neural networks, DGCNN), 其中 EdgeConv 块在局部特征学习过程中有效地整合了邻域间的坐标特征差异. 基于 DGCNN 的思想, 局部特征聚合[21-23]、双边注意力[24-25]等模块均被用于学习更详细的局部特征, 并取得了较好的分割效果.

1.2 三维牙颌模型分割

在三维牙颌模型分割研究中, 传统的分割方法主要基于几何特征[4-7], 如边界线搜索、区域生长算法和分水岭算法等. 这些方法需要人工选择种子点, 然后使用预先选择的几何特征作为判别

依据对牙齿网格进行分割；但它们过度依赖几何特征的完整性，对模型特征变化的敏感性很高.

近年来，基于深度学习的三维分割网络被应用于牙齿分割任务. Zhang 等[26]将三维牙颌模型映射到二维谐波图像中提取特征; Tian 等[12]将点云数据转换为体素数据，并利用三维卷积模块提取特征；Lian 等[14]采用多尺度图约束模块和密集融合策略获取网格的邻域特征和不同尺度的语义特征; Wu 等[15]基于 MeshSegNet[14]，使用边缘卷积模块代替多尺度图约束模块，将坐标特征之间的差异信息整合到邻域特征中; Zheng 等[27]通过网络获取网格的偏移质心，然后基于密度实现黏附牙齿的分离. 为了进一步提取充分的局部特征，Ma 等[28]基于注意机制融合网格几何特征之间的差异信息; Cui 等[29]通过距离感知投票方案生成牙齿质心，并基于置信度感知的注意机制提高边界的光滑程度；Zhao 等[30]将邻域增强机制集成到特征提取模块中，并使用语义特征差异学习权值，使网络关注语义特征相似的网格；Lin 等[31]提出一种质心引导的分离注意方法学习模型的全局特征，并通过高斯邻居注意机制增强网格法向量特征间的差异性.

虽然上述方法通过在特征学习过程中整合邻域信息来捕获详细的局部特征，但仍然存在以下问题: (1) 在邻域信息融合过程中只利用坐标特征之间的差异，忽略了其与几何特征(如法向量、曲率)间的差异信息; (2) 在特征提取过程中，使用单一的池化方式对不同特征进行融合，会丢失大量有用信息. 与上述方法不同，本文网络所使用的特征提取方法，不仅考虑了网格的邻域信息，还根据坐标特征与几何特征二者在网格分割过程中的特性，分别设计不同的注意力机制和互补性融合机制提取并融合上下文的特征，以提高牙齿网格特征提取的效率和完整性.

2　ATNE-NET

2.1　概　述

ATNE-NET 中，牙齿上的每个网格由一个 27 维输入向量描述. 将网格的坐标特征与几何特征进行拼接，并对不同网格的特征向量沿每个维度进行 z-score 归一化. 其中，坐标特征指网格的顶点坐标(9 维)和中心点坐标(3 维)，几何特征包括网格的顶点法向量(9 维)、平面法向量(3 维)和顶点曲率特征(3 维).

ATNE-NET 的结构如图 1 所示，将模型的特征矩阵输入到网络中，通过 3 个彼此交叉的 AEM 模块与 CPM 实现网格的上下文特征提取和融合；然后使用自适应注意力模块实现不同维度特征间的跳跃连接，获取网格的综合特征；最后通过逐次 MLP 进行预测. ATNE-NET 的输出是一个 $N\times C$ 矩阵，每行向量表示相应网格单元属于 C 类的概率.

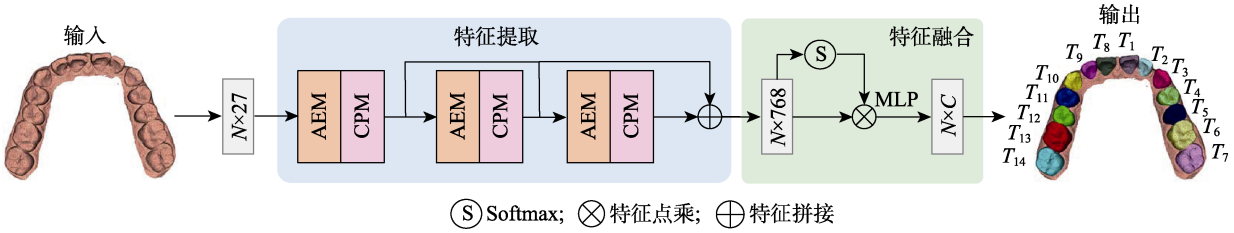


图 1　ATNE-NET 结构

2.2　AEM 的结构

2.2.1　DAEM 的结构

(1) KNN 图. 利用与中心网格 m_i 语义特征最接近的 K 个网格构建局部邻域图. 语义特征距离定义为

$$d_{i,f}=\sqrt{\sum_{q=1}^{n}\left(f_{i,q}-f_{j,q}\right)^2}\ .$$

其中，$f_{i,q}$ 表示 m_i 的第 q 个语义特征；$f_{j,q}$ 表示邻居网格 m_j 的第 q 个语义特征；n 表示一个网格的语义特征总数. 需要注意的是，在第 1 个 DAEM 中，由于只输入几何特征和坐标特征，直接基于网格中心点的坐标构造邻域图. KNN 图定义为 $\boldsymbol{G}(\boldsymbol{M},\boldsymbol{E})$，其中，$\boldsymbol{M}=\{m_1,m_2,\cdots,m_N\}$ 和 $\boldsymbol{E}\subseteq|\boldsymbol{M}|\times|\boldsymbol{M}|$ 分别表示 N 个网格集和的邻域关系集(由 KNN 定义). 对于网格单元 $m_i\in\boldsymbol{M}$，用 $N(i)$ 表示其邻域集 $\boldsymbol{G}(\boldsymbol{M},\boldsymbol{E})$.

(2) 邻域编码增强. 在牙齿边界处，相邻网格间的区分要依靠几何特征间的差异性. 为此，对于每个 m_i 的几何特征，AEM 通过邻域信息增强 m_i 的局部几何信息，如图 2 所示. 在特征增强时，该模块考虑 m_i 与邻域网格在几何特征之间的差异. 首

先从每个邻域网格 $m_k \in N(i)$ 获取几何特征，进行几何空间编码后得到增强特征矢量，公式为

$$\boldsymbol{m}_{i,n}^{k} = \boldsymbol{m}_{i,n} \oplus \boldsymbol{m}_{k,n} \oplus \left(\boldsymbol{m}_{i,n} - \boldsymbol{m}_{k,n}\right),$$

$$\boldsymbol{m}_{i,u}^{k} = \boldsymbol{m}_{i,u} \oplus \boldsymbol{m}_{k,u} \oplus \left(\boldsymbol{m}_{i,u} - \boldsymbol{m}_{k,u}\right).$$

其中，$\boldsymbol{m}_{i,n}$ 和 $\boldsymbol{m}_{k,n}$，$\boldsymbol{m}_{i,u}$ 和 $\boldsymbol{m}_{k,u}$ 分别表示 m_i 和 m_k 的法向量和曲率；$\oplus$ 表示特征拼接；$\|\cdot\|$ 表示欧几里得距离计算. 然后将 m_i 和邻域网格的曲率增强特征与法向量增强特征连接起来，并通过 MLP 对特征进行对齐，公式为

$$\boldsymbol{G}_{i}^{k} = \mathrm{MLP}^{\mathrm{dat}}\left(\boldsymbol{m}_{i,n}^{k} \oplus \boldsymbol{m}_{i,u}^{k}\right),$$

其中，$\boldsymbol{G}_{i}^{k}$ 表示对齐后的几何增强特征.

(3) 差值注意力权重. 采用基于差值注意力机制对几何增强特征赋予不同的重视程度，目的是引导网络关注与 m_i 更相似的 m_k 间的几何差异性. 差值注意力权重赋值的公式为

$$\boldsymbol{W}_{i}^{k} = \mathrm{Softmax}\left(\mathrm{MLP}^{\mathrm{dat}}\left(\boldsymbol{f}_i \oplus \boldsymbol{f}_d\right)\right) \otimes \boldsymbol{G}_{i}^{k}.$$

其中，$\otimes$ 表示执行特征向量的逐元素乘积；$\boldsymbol{f}_d = \boldsymbol{f}_i - \boldsymbol{f}_k$ 表示 m_i 与 m_k 的特征差异，使网络更加关注相似网格间的特征差异，以提高几何拓扑关系的学习效率. DAEM 中，首先通过特征增强机制显著地体现局部网格在几何空间的差异，然后使用差值注意力机制为语义邻近的几何特征赋予更大的关注权重. 这一系列操作可以在邻近却不同类的网格分类中提供显著的几何判别依据，在减缓网络对坐标特征依赖程度的同时提高边界分割精度.

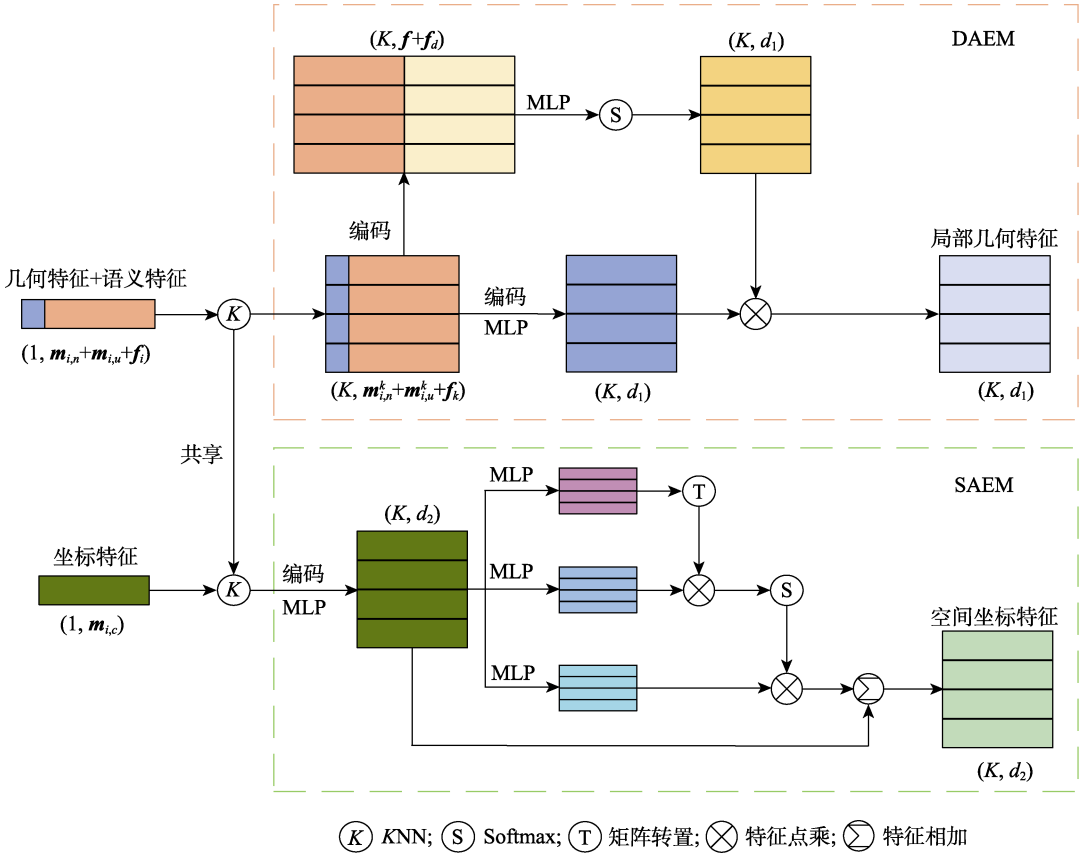


图 2 AEM 结构

2.2.2 SAEM

(1) 共享 *K*NN 图. SAEM 是对网格的坐标特征进行增强学习的，考虑直接基于坐标空间构图进行邻域增强可能会使网络对坐标特征过度依赖，产生拥挤牙齿错分割的情况，因此 SAEM 本身不构建图，直接共享 DAEM 通过语义特征构建的邻域图，以保证网络更加关注语义相似的网格.

(2) 邻域编码增强. 为了使 m_i 更好地感知其所在的局部区域，以及与 m_k 在坐标空间建立内在联系，基于邻域图对坐标特征进行增强和对齐，公式为

$$\boldsymbol{m}_{i,c}^{k} = \boldsymbol{m}_{i,c} \oplus \boldsymbol{m}_{k,c} \oplus \left(\boldsymbol{m}_{i,c} - \boldsymbol{m}_{k,c}\right) \oplus \left\|\boldsymbol{m}_{i,c} - \boldsymbol{m}_{k,c}\right\|.$$

其中，$\boldsymbol{m}_{i,c}$ 和 $\boldsymbol{m}_{k,c}$ 分别表示 m_i 和 m_k 的中心点坐标. 然后通过 MLP 对齐，公式为

$$\boldsymbol{C}_{i}^{k} = \mathrm{MLP}^{\mathrm{sat}}\left(\boldsymbol{m}_{i,c}^{k}\right).$$

(3) 自注意力权重. 为了使网络充分地理解模型的全局结构并抑制噪声，计算增强特征 $\boldsymbol{C}_i^k$ 的自注意力得分并进行聚合，使用残差连接得到体现网格在坐标空间中的全局分布和局部分布的空间坐标特征

$$\boldsymbol{V}_i^k = \mathrm{MLP}^{\mathrm{sat}}\left(\boldsymbol{S}\otimes\mathrm{MLP}^{\mathrm{sat}}\left(\boldsymbol{C}_i^k\right)+\boldsymbol{C}_i^k\right).$$

其中，$\boldsymbol{S}=\mathrm{Softmax}\left(\mathrm{MLP}^{\mathrm{sat}}\left(\boldsymbol{C}_k^i\right)^{\mathrm{T}}\otimes\mathrm{MLP}^{\mathrm{sat}}\left(\boldsymbol{C}_k^i\right)\right)$ 表示特征的自注意力得分.

SAEM 中，首先通过特征增强机制体现局部网格在坐标空间的差异，然后使用自注意力机制学习坐标间的上下文特征. 该模块不仅使网络感知到中心网络邻域分布，还可捕捉到邻域内网格的全局联系，使得同一语义空间内的相似网格紧密地连接在一起，避免出现离散的分类结果.

2.3　CPM 的结构

现有的网络通常使用注意力池化操作集成邻域特征，以获取含有最多有用信息的融合特征向量. 如图 3 所示，本文将 DAEM 输出的 $\boldsymbol{W}_i^k$ 与 SAEM 输出的 $\boldsymbol{V}_i^k$ 拼接，即

$$\boldsymbol{F}_i^k = \mathrm{MLP}^{\mathrm{cpm}}\left(\boldsymbol{W}_i^k\right)\oplus\mathrm{MLP}^{\mathrm{cpm}}\left(\boldsymbol{V}_i^k\right),$$

并使用注意力池化集成邻域信息，即

$$\boldsymbol{A}(\boldsymbol{f}_i)=\sum_{j=1}^{k}\mathrm{Softmax}\left(\mathrm{MLP}^{\mathrm{cpm}}\left(\boldsymbol{F}_i^k\right)\right)\otimes\boldsymbol{F}_i^k.$$

其中，k 表示 m_i 的邻居数，$\boldsymbol{A}(\boldsymbol{f}_i)$ 表示注意力池化特征. 不难理解，注意力池化虽然可以确保空间特征的重要信息尽可能实现传递，却会削弱几何特征和坐标特征间的差异显著程度. 本文认为，在池化过程中，可以利用 MP 尽可能地保留局部几何空间的显著差异信息，使网络的输出具有平移不变性；同时，利用 AP 减少噪声和细微坐标变化对网络特征提取的影响，提高模型的鲁棒性. 公式分别为

$$\boldsymbol{M}(\boldsymbol{f}_i)=\mathrm{MP}\left(\boldsymbol{W}_i^k\right),$$

$$\boldsymbol{E}(\boldsymbol{f}_i)=\mathrm{AP}\left(\boldsymbol{V}_i^k\right).$$

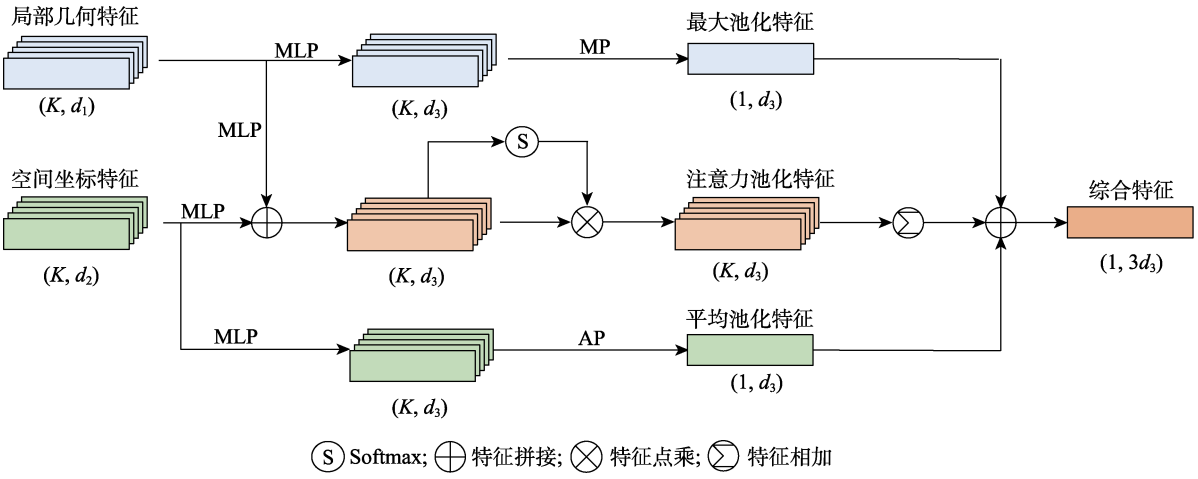


图 3　CPM 结构

综上所述，为了促进不同特征间的互补性能，ATNE-NET 根据各特征的分割特性构建 CPM 整合池化所得的 3 种特征，以获得更丰富信息的融合特征，公式为

$$\boldsymbol{T}(\boldsymbol{f}_i)=\mathrm{MLP}^{\mathrm{cpm}}\left(\boldsymbol{A}(\boldsymbol{f}_i)\oplus\boldsymbol{M}(\boldsymbol{f}_i)\oplus\boldsymbol{E}(\boldsymbol{f}_i)\right).$$

2.4　多尺度特征融合

经过一系列特征提取模块得到融合特征后，ATNE-NET 通过跳跃连接将这些不同维度的输出特征连接起来，公式为

$$\boldsymbol{T}_{\mathrm{mul}}=\boldsymbol{T}^1\oplus\boldsymbol{T}^2\oplus\boldsymbol{T}^3.$$

其中，$\boldsymbol{T}^l$ 表示第 l 个 CPM 的输出特征. 考虑不同尺度特征的重要程度不同，引入自适应注意模块，使网络动态地区分哪些特征更重要. 公式为

$$\boldsymbol{F}_{\mathrm{con}}=\mathrm{Softmax}\left(\mathrm{MLP}^{\mathrm{con}}\left(\boldsymbol{T}_{\mathrm{mul}}\right)\right)\otimes\boldsymbol{T}_{\mathrm{mul}}.$$

最后将 $\boldsymbol{F}_{\mathrm{con}}$ 输入 MLP，实现对网格的分类预测. 每行预测向量的维数为 C，对应每个网格可能的类别概率为 $\boldsymbol{P}=\mathrm{MLP}^{\mathrm{pre}}\left(\boldsymbol{F}_{\mathrm{con}}\right)$.

3　实验与结果分析

3.1　实验细节

3.1.1　数据集

本文实验的数据集由 Teeth3DS 数据库[32]提供的来自不同患者的 200 个三维牙颌网格口扫模型组

成(上、下牙颌一起训练), 并以 4∶1 的比例随机划分为 160 个训练数据和 40 个测试数据. 原始三维牙颌模型包含大约 50 000 个网格单元, 保留拓扑结构将其简化为 15 000 个网格单元; 数据中包含多种牙颌畸形类型(拥挤、断齿、错位、缺齿等), 并在相关专家的指导下使用 MeshLabeler 软件[33]实现标注. 为了提高网络的泛化能力, 对训练集进行增强, 每个牙颌模型有 50%的概率分别在[–10,10]距离之间平移, 沿 y 轴在 $[-\pi/6, \pi/6]$ 角度之间旋转, 以及在[0.8,1.2]倍率之间缩放, 从每个三维模型中生成 20 个新样本.

3.1.2 评估指标与损失函数

为了定量地验证 ATNE-NET 的性能, 本文使用总体精度(overall accuracy, OA), 交并比(intersection over union, IoU)和平均交并比(mean intersection over union, mIoU)作为评价指标, Dice 损失 L_{Dice} 作为损失函数, 它们的计算公式分别为

$$\text{OA} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}},$$

$$\text{IoU} = \frac{\text{TP}}{\text{FN}+\text{FP}+\text{TP}},$$

$$\text{mIoU} = \frac{1}{C}\sum_{1}^{C}\frac{\text{TP}}{\text{FN}+\text{FP}+\text{TP}},$$

$$L_{\text{Dice}} = \frac{2\text{TP}}{2\text{TP}+\text{FN}+\text{FP}}.$$

其中, TP, TN, FP 和 FN 分别表示真阳性(true positive, TP), 真阴性(true negative, TN), 假阳性(false positive, FP)和假阴性(false negative, FN)样本; C 表示网格类数.

3.1.3 网络细节

如图 1 所示, ATNE-NET 包含 3 个连续的特征提取模块和 1 个特征融合模块. 从第 1 个模块到第 3 个模块的 MLP^{dat}, MLP^{sat} 和 MLP^{cpm} 包含通道数分别为 64, 128 和 256 的 2D Conv; KNN 图的邻居个数 k =30; 特征融合模块的 MLP^{con} 包含 1 个 1 024 通道的 1D Conv; MLP^{pre} 包含 4 个连续的 1D Conv, 通道数分别为 1 024, 512, 256 和 C. 除了最后 1 个 Conv, 其余 Conv 后面都有 1 个 BN 和 Leaky-ReLU.

本文在 PyTorch 框架下进行所有训练与实验. ATNE-NET 在一台 NVIDIA GeForce RTX 3090 GPU 上进行 200 轮的训练, 批量大小设置为 2; 使用初始学习率为 10^{-3} 的 Adam 优化器, 每 20 轮减少 50%.

3.2 对比实验

将 ATNE-NET 依次与 2 种三维形状分割网络(PointNet++[19], DGCNN[20]), 4 种三维牙齿模型分割网络(MeshSegNet[14], TeethGNN[27], LSAM[30]和 DBGAN[31])进行实验, 表 1 所示为 OA 和 mIoU 指标的定量比较. 其中, IoU 仅显示右侧牙齿和牙龈的分割结果, $T_1 \sim T_6$ 表示右侧不同牙齿的标签, T_0 表示牙龈网格的标签(如图 1 所示). 可以看出: (1) 与同样使用特征增强机制的 LSAM[30]和 TeethGNN[27]相比, ATNE-NET 的 OA 分别提高 2.93 个百分点和 1.15 个百分点, 表明该网络在获得更丰富的局部特征方面是有效的. 因为 LSAM[30]和 TeethGNN[27]这 2 种网络在特征增强过程中只考虑了坐标信息, 而 ATNE-NET 不仅考虑坐标特征, 还通过差值注意力机制增强了邻近网格几何特征间的差异性. (2) ATNE-NET 的分割精度优于 MeshSegNet[14]和 DBGAN[31], 因为这 2 种对比网络具有粗糙的特征融合方式, 证明了 ATNE-NET 在促进特征互补方面的有效性. (3) 与所有对比网络相比, ATNE-NET 在每颗牙齿分割中获得了更高的 IoU, 证明了其在牙齿精细分割过程中的泛化性能.

图 4 所示为表 1 中 7 种网络分割结果的可视化结果. 可以看出: (1) 第 1~2 行的红色圆圈中, PointNet++[19]和 DGCNN[20]对错位牙齿的分割精度

表 1 7 种网络的分割结果比较 %

主干网络	OA	mIoU	IoU							
			T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_0
PointNet++[19]	88.01	76.15	65.66	72.31	75.98	80.66	78.81	73.15	77.12	85.64
DGCNN[20]	89.93	83.15	80.16	77.34	81.45	85.13	84.24	78.37	82.39	88.17
MeshSegNet[14]	91.26	84.07	80.59	81.49	81.36	83.41	82.73	79.69	85.58	90.83
TeethGNN[27]	93.38	84.52	81.63	82.56	81.37	82.73	83.48	82.13	82.55	93.43
LSAM[30]	93.25	85.61	82.81	84.76	84.05	85.34	85.46	82.49	84.55	92.67
DBGAN[31]	95.03	88.27	85.76	86.28	88.51	87.72	87.53	88.62	87.67	93.77
ATNE-NET	**96.18**	**90.13**	**88.78**	**88.67**	**89.02**	**88.15**	**89.22**	**88.51**	**87.65**	**94.97**

注. 粗体表示最优值.



a. 原始模型　b. PointNet++[19]　c. DGCNN[20]　d. MeshSegNet[14]　e. TeethGNN[27]　f. LSAM[30]　g. DBGAN[31]　h. ATNE-NET　i. 标注

图 4　ATNE-NET 与 6 种对比网络的分割结果可视化对比

低，因为这些网络在特征提取时忽略了邻域网格间的特征差异，依赖于网格的坐标特征进行分割；基于特征图的网络 MeshSegNet[14]和 LSAM[30]在错位牙齿的分割上精度有一定提升，而使用网络进行质心定位的 TeethGNN[27]和 DBGAN[31]同样准确地识别了错位牙齿的区域，但是这些网络在牙齿边界处均存在过分割现象；ATNE-NET 在牙齿边界处体现出了优异的分割性能，这是因为差值注意力机制使得网络对边界处网格的细微几何特征变化更加敏感. (2) 第 3~4 行中，即使边界结构相对平坦(如棕色虚线圈所示)，ATNE-NET 在牙齿和牙龈的边界处依然实现了很好的分割效果；将特征简单拼接在一起的 DBGAN[31]和使用聚类算法的 TeethGNN[27]在牙齿边界处依然存在过分割现象. 实验结果进一步证明了 ATNE-NET 在促进特征互补方面的有效性.

3.3　消融实验

3.3.1　AEM 中邻域增强机制的有效性

为了证明 AEM 中的邻域增强机制的有效性，通过不同的增强类型获得 3 种变体：(1) 所有特征都不进行邻域增强，通过中心网格特征重复获得与增强矩阵相同维度的输入(基准网络)；(2) 只对几何特征增强(基准网络+几何增强)；(3) 只对坐标特征进行增强(基准网络+坐标增强).

将这些变体与 ATNE-NET 进行比较，定量比较结果如表 2 所示. 可以看出，随着几何特征和坐标特征的增强，网络的整体性能逐渐提高；基准网络无法有效地学习网格特征，因为 MLP 无法提取网格间的拓扑关系；坐标增强变体比几何增强变体的分割性能的 OA 和 mIoU 提高 18.26 个百分点和 18.37 个百分点，这表明坐标特征包含的拓扑结构是网络局部特征学习的关键；ATNE-NET 取得了更高的精度，几何特征的增强使得网络不过度依赖坐标特征，有效地感知相邻网格间几何特征的微小变化.

表 2　AEM 中不同邻域增强机制的消融结果比较　%

变体方法	OA	mIoU
基准网络	56.25	48.19
基准网络+几何增强	71.35	65.37
基准网络+坐标增强	89.61	83.74
ATNE-NET	**96.18**	**90.13**

注. 粗体表示最优值.

3.3.2　AEM 中注意力机制的有效性

为了突出各特征的分割特性，AEM 中设计了 2 种注意力机制对增强后的特征进行权重优化. 为了证明不同注意力机制的有效性，使用不同的注意力方法获得 3 种变体：(1) 所有增强特征都不进行注意力权重优化(基准网络)；(2) 只对几何特征使用差值注意机制(基准网络+差值注意)；(3) 只对坐标特征使用自注意力机制(基准网络+自注意力).

将这些变体与 ATNE-NET 进行比较，定量比较结果如表 3 所示. 可以看出，引入注意力机制后均有效地提高了模型的分割精度，其中，与自注意力机制相比，差值注意力机制取得了更高的提升精度，这是因为差值注意力可以解决边界处分割模糊的问题. 为了进一步体现 ATNE-NET 的有效性，对分割结果进行可视化，结果如图 5 所示. 可以看出，使用差值注意的网络在齿间的边界处精度较高，而使用自注意力的网络在齿间孤立错分结果较少，这是由于差值注意力使网络更加关注

同类网格间的几何差异，相邻却拥有较大几何差值的网格将被赋予其他标签，进而使边界光滑；而自注意力使网络更加关注同类网格间的连通性，避免孤立错误.

表 3 AEM 中不同注意力机制的消融结果比较 %

变体方法	OA	mIoU
基准网络	93.03	84.22
基准网络+差值注意	95.26	88.12
基准网络+自注意力	94.51	86.37
ATNE-NET	**96.18**	**90.13**

注. 粗体表示最优值.

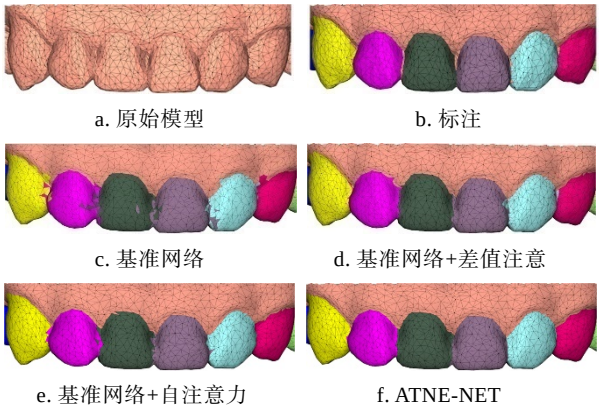


图 5 不同注意力机制的可视化分割结果

3.3.3 CPM 的有效性

经过特征 AEM 后，本文提出的 CPM 在最大限度地保留各自分割特性的条件下进行特征融合. 为了验证 CPM 的有效性，使用不同的池化组合方法进行了实验：(1) 对 2 类拼接后的特征通过注意力池化融合，并分别对 2 类特征使用 MP 融合(双 MP)；(2) 对 2 类拼接后的特征通过注意力池化融合，并分别对 2 类特征使用 AP 融合(MP)；(3) 对 2 类拼接后的特征通过注意力池化融合，并分别对坐标特征使用 MP 融合，对几何特征使用 AP 融合(平均池+最大池).

将不同池化方法与 ATNE-NET 进行比较，定量比较结果如表 4 所示. 可以看出 AP+MP 变体的精度提升最少，原因是这种组合方式与对应特征的特性不符，抑制了特征间的互补作用. 为了进一步验证上述结论，图 6 所示为这些变体的可视化结果. 与表 4 的定量分析结果一致，可以看出，ATNE-NET 的结果最好；双最大池中出现较多的离散错误，双平均池中齿间边界较为粗糙，这是由于 MP 作用于坐标特征，使网络更加关注距离较远的网格，进而产生了离散误判；而 AP 则抑制了几何特征间的差异显著程度.

表 4 不同池化方法的消融结果比较 %

池化方法	OA	mIoU
双最大池	95.87	89.82
双平均池	95.04	88.73
平均池+最大池	94.16	87.65
ATNE-NET	**96.18**	**90.13**

注. 粗体表示最优值.

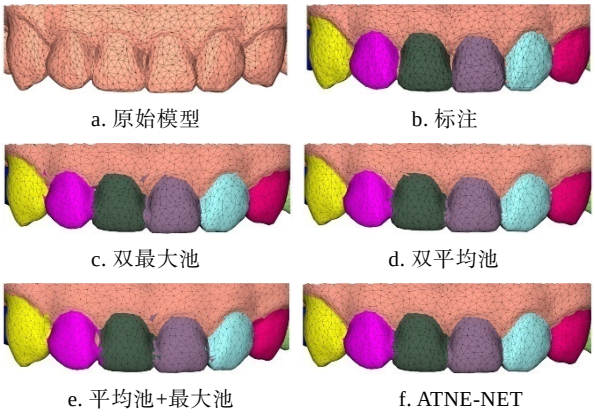


图 6 不同池化方法的可视化分割结果

3.3.4 自适应特征融合策略的有效性

在特征融合的跳跃连接的过程中，使用自适应注意力机制为不同维度的特征分配合理的权重. 为了评估这一措施的有效性，将 ATNE-NET 与拼接融合的变体进行实验，即网络直接将特征拼接在一起，定量比较结果如表 5 所示. 可以看出，使用自适应注意机制的网络分割效果更好，说明自适应特征融合策略可以对特征赋予有效的权重，平衡多尺度特征间的数值差异，避免某一特征数值过大而忽略其他特征.

表 5 特征融合对比分割结果 %

方法	OA	mIoU
拼接融合	95.06	88.91
ATNE-NET	**96.18**	**90.13**

注. 粗体表示最优值.

4 结 语

本文提出一种 ATNE-NET 用于牙颌模型分割. 首先通过增强机制提高特征的局部感知能力，并针对几何特征与坐标特征的分割性质分别设计了 DAEM 与 SAEM，促进 2 类特征的互补性；然后设计了 CPM，在保留最大有用信息的同时尽可能地确保 2 类特征间的差异性；最后自适应地融合不同

维度的特征，为其赋予不同权重．将 ATNE-NET 与 6 种先进的方法在 200 个真实患者的牙颌模型数据集上进行比较，并通过消融实验直观地分析了核心模块的有效性．实验结果证明了 ATNE-NET 的优越性，其适合具有挑战性的实际任务．

参考文献(References):

[1] Sharifovich O S, Axmadjanovna Y U, Amrulloevich G S. Comprehensive treatment and diagnosis of functional and metabolic disorders in children with dental deformities[J]. Journal of Pharmaceutical Negative Results, 2023, 14(3): 3708-3715

[2] Ural Ç, Kaleli N. Direct digitalization devices in today's dental practice: intra oral scanners[J]. Journal of Experimental and Clinical Medicine, 2021, 38(S2): 136-142

[3] Wei G D, Cui Z M, Liu Y M, *et al*. TANet: towards fully automatic tooth arrangement[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 481-497

[4] Tong W H, Yang X K, Pan M D, *et al*. Spectral mesh segmentation via ℓ_0 gradient minimization[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(4): 1807-1820

[5] Kakehbaraei S, Arvanaghi R, Seyedarabi H, *et al*. 3D tooth segmentation in cone-beam computed tomography images using distance transform[J]. Biomedical Signal Processing and Control, 2023, 79: Article No.104122

[6] Zhou Z S, Wan H L, Zhang H Y, *et al*. Segmentation of spontaneous intracerebral hemorrhage on CT with a region growing method based on watershed preprocessing[J]. Frontiers in Neurology, 2022, 13: Article No.865023

[7] Jiang Xiaotong, Dai Ning, Zhang Changdong, *et al*. A semi-automatic precise tooth segmentation algorithm of dental model[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(5): 820-829(in Chinese)
(姜晓通，戴宁，张长东，等．牙齿半自动精确分割算法[J]. 计算机辅助设计与图形学学报, 2020, 32(5): 820-829)

[8] Kondo T, Ong S H, Foong K W C. Tooth segmentation of dental study models using range images[J]. IEEE Transactions on Medical Imaging, 2004, 23(3): 350-362

[9] Wongwaen N, Sinthanayothin C. Computerized algorithm for 3D teeth segmentation[C] //Proceedings of the International Conference on Electronics and Information Engineering. Los Alamitos: IEEE Computer Society Press, 2010: V1-277-V1-280

[10] Jiang B X, Zhang S Z, Shi M Y, *et al*. Alternate level set evolutions with controlled switch for tooth segmentation[J]. IEEE Access, 2022, 10: 76563-76572

[11] Xu X J, Liu C, Zheng Y Y. 3D tooth segmentation and labeling using deep convolutional neural networks[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 25(7): 2336-2348

[12] Tian S K, Dai N, Zhang B, *et al*. Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks[J]. IEEE Access, 2019, 7: 84817-84828

[13] Sun D Y, Pei Y R, Song G Y, *et al*. Tooth segmentation and labeling from digital dental casts[C] //Proceedings of the 17th IEEE International Symposium on Biomedical Imaging. Los Alamitos: IEEE Computer Society Press, 2020: 669-673

[14] Lian C F, Wang L, Wu T H, *et al*. Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3D intraoral scanners[J]. IEEE Transactions on Medical Imaging, 2020, 39(7): 2440-2450

[15] Wu T H, Lian C F, Lee S, *et al*. Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3D intraoral scans[J]. IEEE Transactions on Medical Imaging, 2022, 41(11): 3158-3166

[16] Hu H Q, Li Z X, Gao W C. MPCNet: improved MeshSegNet based on position encoding and channel attention[J]. IEEE Access, 2023, 11: 23326-23334

[17] Lian C F, Wang L, Wu T H, *et al*. MeshSNet: deep multi-scale mesh feature learning for end-to-end tooth labeling on 3D dental surfaces[C] //Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2019: 837-845

[18] Remmel T K, Extending morphological pattern segmentation to 3D voxels[J]. Landscape Ecology, 2022, 37(2): 373-380

[19] Qi C R, Yi L, Su H, *et al*. PointNet++: deep hierarchical feature learning on point sets in a metric space[C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2017: 5105-5114

[20] Zhang R T, Zhang G Y, Yin J H, *et al*. Mesh-based DGCNN: semantic segmentation of textured 3-D urban scenes[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: Article No.4402812

[21] Ren D Y, Wu Z Y, Li J W, *et al*. Point attention network for point cloud semantic segmentation[J]. Science China Information Sciences, 2022, 65(9): Article No.192104

[22] Wang Y, Tang X T, Yue C K. Enhancing the local graph semantic feature for 3D point cloud classification and segmentation[J]. IEEE Access, 2022, 10: 74620-74628

[23] Liu C, Zeng D D, Akbar A, *et al*. Context-aware network for semantic segmentation toward large-scale point clouds in urban environments[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: Article No.5703915

[24] Zeng Z Y, Xu Y Y, Xie Z, *et al*. LEARD-Net: semantic segmentation for large-scale point cloud scene[J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 112: Article No.102953

[25] Yin F K, Huang Z L, Chen T, *et al*. DCNet: large-scale point cloud semantic segmentation with discriminative and efficient feature aggregation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 4083-4095

[26] Zhang J D, Li C P, Song Q, *et al*. Automatic 3D tooth segmentation using convolutional neural networks in harmonic parameter space[J]. Graphical Models, 2020, 109: Article No.101071

[27] Zheng Y Y, Chen B J, Shen Y F, *et al*. TeethGNN: semantic 3D teeth segmentation with graph neural networks[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(7): 3158-3168

[28] Ma T, Yang Y Z, Zhai J C, *et al*. A tooth segmentation method

based on multiple geometric feature learning[J]. Healthcare, 2022, 10(10): Article No.2089

[29] Cui Z M, Li C J, Chen N L, *et al*. TSegNet: an efficient and accurate tooth segmentation network on 3D dental model[J]. Medical Image Analysis, 2021, 69: Article No.101949

[30] Zhao Y, Zhang L M, Yang C S, *et al*. 3D Dental model segmentation with graph attentional convolution network[J]. Pattern Recognition Letters, 2021, 152: 79-85

[31] Lin Z J, He Z S, Wang X, *et al*. DBGANet: dual-branch geometric attention network for accurate 3D tooth segmentation[OL]. [2024-01-17]. https://ieeexplore.ieee.org/document/10313281

[32] Ben-Hamadou A, Smaoui O, Chaabouni-Chouayakh H, *et al*. Teeth3DS: a benchmark for teeth segmentation and labeling from intra-oral 3D scans[OL]. [2024-01-17]. https://arxiv.org/abs/2210.06094

[33] Wu T H, Lian C F, Piers C, *et al*. Machine (deep) learning for orthodontic CAD/CAM technologies[M] //Ko C C, Shen D G, Wang L. Machine Learning in Dentistry. Cham: Springer, 2021: 117-129