基于局部注意力机制的三维牙齿模型分割网络

张凌明[1,2],赵 悦[1,2],李鹏程[1,2],刘 洋[3],高陈强[1,2]
(1. 重庆邮电大学通信与信息工程学院,重庆 400065; 2. 信号与信息处理重庆市重点实验室,重庆,400065;
3. 重庆医科大学附属口腔医院,重庆,400060)

摘 要: 从三维牙齿模型中准确分割出牙齿部分是正畸计算机辅助诊疗的基础. 由于现有的三维模型分割网络对局部特征建模方式相对简单,这些方法无法有效提取牙齿边缘区域更细节的局部特征信息,进而导致这些区域出现牙齿多分、漏分等情况. 本文提出一种基于局部注意力机制的三维牙齿模型分割网络以提高牙齿边缘区域的分割性能. 首先,对原始牙齿模型中的三维网格数据进行多尺度的局部空间区域构建. 其次,根据每个局部区域内的网格空间分布和网格特征差异进行注意力权重的学习. 最后,基于学习到的网格权重进行局部特征聚合,以使得网络能自适应地去关注各个局部区域内更具有表达性网格特征. 在临床数据集上的实验结果表明,相对于现有方法,本文网络的分割结果在牙齿边界区域更加准确光滑.

关键词: 三维网格数据; 口腔扫描数据; 三维牙齿模型; 牙齿分割; 注意力机制
中图分类号: TP753 **文献标识码:** A **文章编号:** 0372-2112(2022)03-0681-10
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20201338

Dental Model Segmentation Network Based on Local Attention Mechanism

ZHANG Ling-ming[1,2], ZHAO Yue[1,2], LI Peng-cheng[1,2], LIU Yang[3], GAO Chen-qiang[1,2]
(1. *School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing*, 400065, *China*;
2. *Chongqing Key Laboratory of Signal and Information Processing, Chongqing*, 400065, *China*;
3. *The Stomatology Hospital, Chongqing Medical University, Chongqing*, 400060, *China*)

Abstract: Accurate tooth segmentation from 3D dental model is the basis of computer-aided-design (CAD) for orthodontic treatment. Due to the relatively coarse modeling of local feature, existing 3D shape segmentation networks cannot effectively extract more detailed local feature on teeth boundaries. This issue will further result in over-segmentation or under-segmentation on boundaries. In this paper, a 3D dental model segmentation network based on local attention mechanism is proposed to improve segmentation performance on teeth boundaries. Firstly, multi-scale local spaces are constructed for 3D mesh data of raw dental model. Secondly, attention weights are learned based on the spatial distribution and feature differences of meshes for each local space. Finally, a local feature aggregation is applied based on learned attention weights of meshes to make the network automatically focus on more representive mesh features in each local space. The proposed network is evaluated on a real-patient datasets, and the experimental results show that our network can more clearly and accurately segment teeth boundaries when compared with existing methods.

Key words: 3D mesh data; oral scanning data; 3D dental model; tooth segmentation; attention mechanism

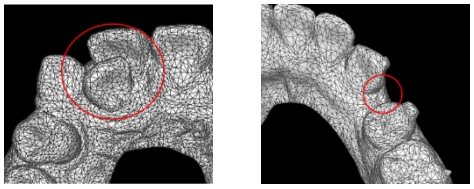
1 引言

三维牙齿模型是利用口内扫描仪对患者口腔内软硬组织进行实时扫描和重建而得到的一种数字化三维模型,其本质是由无序的三维点云或三维网格组成的非结构化数据. 三维牙齿模型的分割任务是指从模型中准确分割出牙齿区域,其分割结果可以高效地协助医生对患者牙齿进行移动、重排列等操作,以便可靠地

收稿日期:2020-11-26;修回日期:2021-06-02;责任编辑:梅志强
基金项目:国家自然科学基金(No.61571071,No.61906025);重庆市自然科学基金项目(No.cstc2018jcyjAX0227, No.cstc2020jcyj-msxmX0835);重庆市教委科学技术研究项目(No.KJQN201900607, No.KJQN202000647)

模拟正畸治疗后的效果. 同时,分割信息还可以为牙齿种植导板设计、3D生物打印种植体、以及患者后续治疗计划的制定提供重要参考信息. 因此,三维牙齿模型分割具有十分重要的意义.

然而,由于不同患者之间牙齿形状存在巨大差异,三维牙齿模型的分割任务也存在许多难点,如图1所示. 图1(a)所示的牙齿错位现象通常会导致不同类别牙齿在空间上距离更近甚至存在部分区域的重叠,而图1(b)所示的牙齿缺失现象也会造成整个模型的牙齿排列呈现更大的差异性,这都给网络对牙齿模型整体拓扑结构的学习造成较大困难.



(a) 牙齿错位 (b) 牙齿缺失

图1 三维牙齿模型中牙齿错位和缺牙示意图

到目前为止,已经提出了一些基于深度学习的三维牙齿模型分割方法[1,4]. 一种最直接的思路就是将三维牙齿模型当成三维点云数据或者三维网格数据进行处理,因此可将其分割任务进一步细化为点云或网格的分割任务. 例如,部分学者将三维牙齿模型中的点云或网格预处理为类似于图像领域的结构化数据形式,然后送入2D或3D卷积神经网络进行牙齿分割[1,2]. 这类方法较好地解决了卷积神经网络由于三维数据的无序性而无法直接应用于三维牙齿模型的问题. 其他学者则直接对现有的点云分割网络进行迁移或改进[3,4],例如,Lian等人[4]先通过构建邻接矩阵对PointNet[5]特征提取方法进行改进,然后用于三维牙齿模型的分割任务. 这类网络的结构往往可以很好地处理三维数据的无序性问题,因此可以避免额外的数据预处理操作.

虽然上述方法取得了一定的效果,但其中部分方法[1,2]对每个分割单位(点或网格)进行相对独立的特征提取,忽略了局部特征信息对牙齿分割的重要性. 其他方法[3,4]虽然对局部特征进行了建模,但其提取过程没有充分考虑局部空间内分割单位的真实分布情况,因此在牙齿边缘区域无法提取更细节的局部形状特征,进而导致这些区域出现较为严重的过分割或欠分割现象.

最近,学者们提出了一些基于局部空间分组的点云分割方法[6,7],并且在一定程度上优化了三维数据的局部特征提取问题. 然而,这些方法依旧采用相对简单的局部特征提取方法,例如文献[7]直接使用最大值池化(Max Pooling)对局部区域进行特征聚合. 由于最大值池化仅能选择性地保留部分局部特征信息,而忽略了各单位之间的深层关系(例如相对空间距离、特征相似度等). 因此,这种方式仍然会丢失部分细节信息.

为提取三维牙齿模型中更细节的局部形状特征,提高牙齿边缘区域的特征辨别能力,本文提出了一种基于局部注意力机制的分割网络. 网络首先以三维网格数据的形式对牙齿模型进行多尺度局部空间区域构建. 对于每一个局部区域,网络先进行空间信息增强以丰富网格数据的空间特征. 在此基础上,再根据网格的空间分布和相对特征差异自动学习注意力权重,并基于该权重进行局部特征聚合. 这种基于注意力机制的局部特征提取方式能帮助网络自适应地去关注不同局部区域内更具有表达性的网格特征,因此能在牙齿边缘区域提取更细节的局部形状特征,有效解决了现有方法无法准确分割牙齿边缘区域的问题. 本文网络在临床三维牙齿模型数据集上的实验结果表明:相对于现有的分割网络,本文方法能更准确地分割出牙齿边缘等低特征识别度区域,且能较好克服数据中存在的缺牙、牙齿错位等分割难点.

2 相关工作

2.1 传统三维牙齿模型分割方法

传统的三维牙齿模型分割方法通常利用预定义的空间几何特征如曲率、法向量等作为牙齿分割的参考信息. 这些方法大致可分为:(1)基于曲率的方法(curvature-based method)[8,12];(2)基于轮廓线的方法(contour-line-based method)[13,14];(3)基于谐波场的方法(harmonic-field-based method)[15]. 部分学者还将三维牙齿模型先映射为欧氏空间的2D图像进行图像分割[16,17]. 上述传统方法虽然比较直观,但由于不同人牙齿的形状变化较大,导致这些基于几何特征的方法鲁棒性差,易出现分割结果不稳定的情况. 同时部分传统分割方法还需要一定的人工交互,无法实现全自动的分割.

2.2 基于深度学习的方法

随着深度学习技术在自然图像和医学图像分割领域取得的巨大成功[18,19],许多学者也将其应用于三维牙齿模型的分割任务中. Xu等人[1]先手动提取三维牙齿模型中网格的几何特征,然后将每个网格的特征向量以矩阵形式送入卷积神经网络进行单个网格分类的分类任务. Farhad等人[3]则将牙齿模型以点云数据的形式送入PointCNN[6]网络进行点云分割任务,并配合鉴别器网络对分割结果进行优化. 上述方法虽然在总体上能取得较好的分割效果,但由于没有充分考虑局部区域内各单位之间的特征关联性,因此无法有效提取牙齿的局部形状信息. 而本文网络能充分参考局部区域内各单位的空间位置和特征信息进行局部特征提

取,以这种方式得到的特征信息能更好地反应牙齿的真实形状.

此外,为能将3D卷积神经网络迁移至三维牙齿模型的分割任务中,Tian等人[2]先利用稀疏八叉树分区[20]的方式(sparse octree partitioning)将三维牙齿模型进行体素化(voxelization),然后将其送入3D卷积神经网络对牙齿进行分割. 但该方法同时也增加了额外的计算开销,且在体素化阶段可能会引入量化误差.

近年来,学者们针对点云分割任务做了大量研究,如基于多视角的点云分割方法[21,24]、基于体素化的点云分割方法[25,30]等. 这些方法也为三维牙齿模型的分割任务提供了重要参考. 其中Qi等人提出的PointNet[5]是第一个可以直接将无序的三维空间数据作为输入的网络,有效地解决了非结构化数据的无序性问题. 鉴于PointNet[5]在点云处理上表现出的良好性能,Lian等人[4]在此基础上利用邻接矩阵对PointNet[5]的特征提取过程进行改进,并提出了一种新的三维牙齿模型分割网络MeshSegNet. 由于PointNet[5]无法提取三维数据的局部特征信息,Qi等人又提出了PointNet++[7]. PointNet++通过对三维空间数据进行局部区域构建的方式来提取局部特征信息,并在一定程度上提升了网络的学习能力,但由于其仅简单地使用最大值池化进行局部特征聚合,依旧会丢失了其他细节信息. 近年来,随着注意力机制在各领域取得的巨大成功[31,33],本文也设计了一种基于注意力机制的局部特征聚合,相比于对使用最大值池化,这种方式能更好地保留局部区域内细节形状信息.

3　本文方法

本文网络框架如图2所示,整个网络可分为局部特征提取阶段和特征逆向传播阶段. 在局部特征提取阶段,网络首先对输入的三维牙齿模型进行网格下采样以构建局部空间区域,然后对每个局部空间区域进行空间信息增强和基于局部空间注意力机制的特征聚合以提取局部特征信息. 提取的局部特征信息将作为下一层网络的输入进行同样的操作直至局部特征提取阶段结束. 在特征逆向传播阶段,网络以最后一次局部特征聚合模块的输出为起点,通过上采样和特征逆向传播将网格数量逐步恢复至原始牙齿模型具有的网格数量,并同时将两个阶段对应的网格特征信息进行融合学习. 最后,网络利用多层感知器(Multi-layer Perceptron, MLP)进行网格级别的牙齿分割预测.

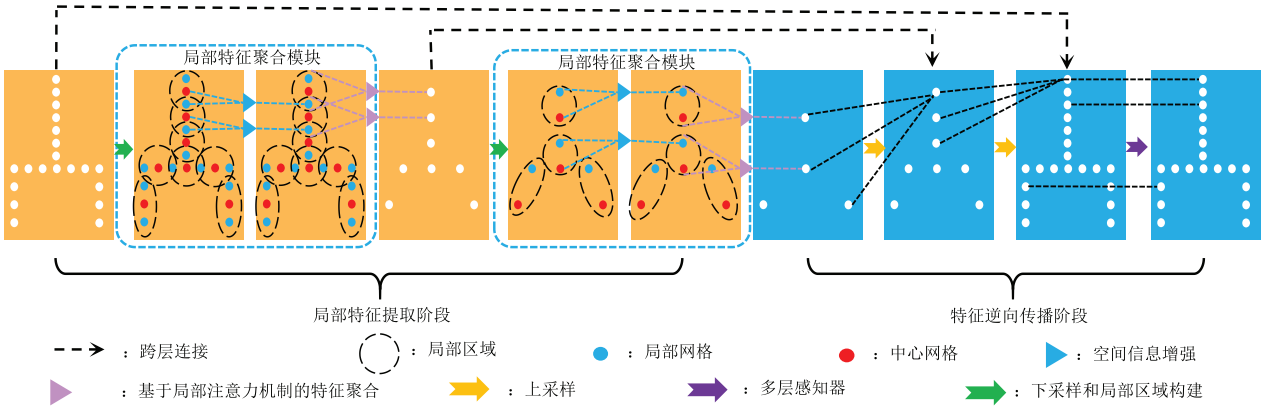


图2　本文网络的整体结构示意图

3.1　数据预处理

本文网络的初始输入包含三维牙齿模型中N个网格的初始特征信息和空间位置信息两部分. 由于每个网格初始特征信息仅含有网格三个顶点坐标信息组成的向量$\boldsymbol{V}\in R^9$. 若仅使用该信息作为输入,并不利于网络学习更细节的语义特征,且没有充分利用三维网格具有局部拓扑连接这一优势. 因此本文在数据预处理部分对每个网格进行额外的空间特征提取以丰富网格的初始特征信息. 定义三维牙齿模型中所有网格的集合为$M=\{m_1, m_2, \cdots, m_N\}$,$|M|=N$. 对于网格$m_i\in M$,先获取其网格的法向量信息$\boldsymbol{N}_{\text{mesh}}\in R^3$和其三个顶点法向量信息$\boldsymbol{N}_{\text{vertex}}\in R^9$,然后再将$\boldsymbol{N}_{\text{mesh}}$、$\boldsymbol{N}_{\text{vertex}}$和$\boldsymbol{V}$共同拼接成网格$m_i$的初始特征向量$\boldsymbol{f}_i\in R^{21}$. 而在定义每个网格的空间位置信息时,考虑到仅使用单个顶点坐标作为网格的空间位置信息并不准确,因此本文选择网格的中心点坐标作为每个网格的空间位置信息. 对于网格m_i,其中心点$\boldsymbol{p}_i\in R^3$的坐标为:

$$\boldsymbol{p}_i=\left(\frac{x_1^i+x_2^i+x_3^i}{3},\frac{y_1^i+y_2^i+y_3^i}{3},\frac{z_1^i+z_2^i+z_3^i}{3}\right)\tag{1}$$

其中,x^i,y^i,z^i分别表示网格m_i的顶点坐标值. 在完成N个网格的数据预处理操作后,将它们的初始特征信息和空间位置信息分别拼接成矩阵$\boldsymbol{F}=(\boldsymbol{f}_1,\boldsymbol{f}_2,\cdots,\boldsymbol{f}_N)$和$\boldsymbol{P}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_N)$,并作为网络的初始输入,其中$\boldsymbol{F}\in R^{N\times 21}$,$\boldsymbol{P}\in R^{N\times 3}$.

3.2 局部区域构建

本文先对牙齿模型中的网格数据进行局部区域的构建，然后再进行后续的特征学习．以第一次局部区域构建为例，网络先利用最远距离采样（Farthest Point Sampling，FPS）从输入的网格集合 M 中下采样出 $N_1(N_1<N)$ 个网格，并将这些被采样的网格定义为中心网格（如图2中红色圆点示意）．定义中心网格集合 $M_{\text{centre}}=\{m_1^{\text{c}},m_2^{\text{c}},\cdots,m_{N_1}^{\text{c}}\}$，对于任意中心网格 $m_i^{\text{c}}\in M_{\text{centre}}$，网络以其中心点坐标作为参考，在整个数据空间内选择距离该点最近的 k 个其他网格组成 m_i^{c} 的局部网格集合（如图2中蓝色圆点示意）．中心网格 m_i^{c} 和其局部网格集合便组成了一个局部空间区域（如图2中圆形虚线框示意）．当完成所有中心网格的局部区域构建后，原始输入的 N 个网格便被划分为 N_1 个局部空间区域（即 N_1 个网格分组），这些局部空间区域将会被送入后续的局部特征聚合模块进行局部特征提取．

3.3 局部特征聚合模块

局部信息聚合模块整体框图如图3所示．以网络第一个局部特征聚合模块为例，其输入为 N_1 个网格分组，输出为 N_1 个局部特征信息．模块首先对所有局部网格集合进行空间信息增强，并将增强结果与网格原始特征进行融合学习以得到更丰富特征信息．在此基础上，模块再根据中心网格和局部网格的真实空间分布和特征差异自动学习出局部网格的注意力权重，并基于该权重进行局部特征聚合．

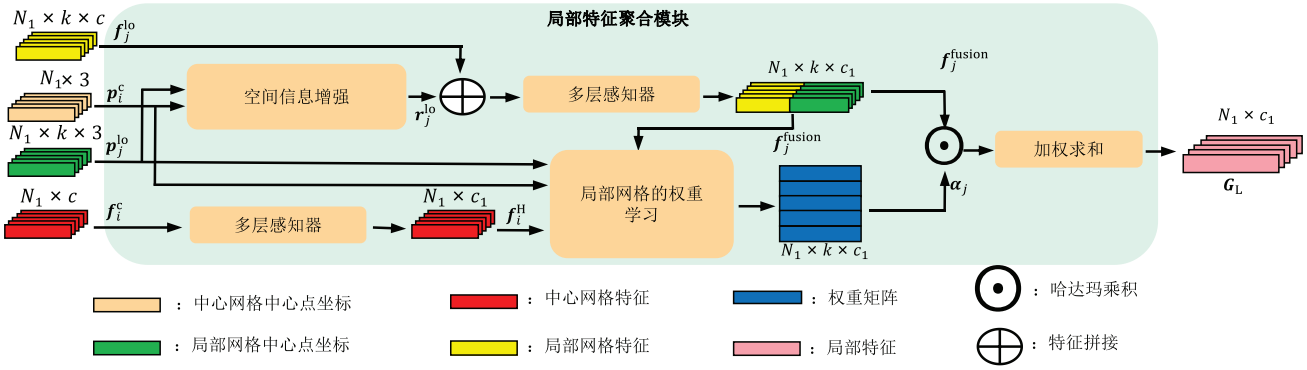


图3 局部特征聚合模块框图

3.3.1 局部特征聚合模块

对于任意一个局部空间区域L，定义中心网格 m_i^{c} 的中心点坐标为 $\boldsymbol{p}_i^{\text{c}}$，原始特征为 $\boldsymbol{f}_i^{\text{c}}$．定义区域L中局部网格集合 $M_{\text{local}}=\{m_1^{\text{lo}},m_2^{\text{lo}},\cdots,m_k^{\text{lo}}\}$，局部网格的中心点坐标矩阵 $\boldsymbol{P}_{\text{local}}=(\boldsymbol{p}_1^{\text{lo}},\boldsymbol{p}_2^{\text{lo}},\cdots,\boldsymbol{p}_k^{\text{lo}})$ 和局部网格的原始特征矩阵 $\boldsymbol{F}_{\text{local}}=(\boldsymbol{f}_1^{\text{lo}},\boldsymbol{f}_2^{\text{lo}},\cdots,\boldsymbol{f}_k^{\text{lo}})$．对于任意局部网格 $m_j^{\text{lo}}\in M_{\text{local}}$，其空间信息增强方式为：

$$\boldsymbol{r}_j^{\text{lo}}=\sigma(\boldsymbol{p}_i^{\text{c}}\oplus\boldsymbol{p}_j^{\text{lo}}\oplus(\boldsymbol{p}_j^{\text{lo}}-\boldsymbol{p}_i^{\text{c}})) \tag{2}$$

其中，$\boldsymbol{r}_j^{\text{lo}}$ 表示网格 m_j^{lo} 增强后的空间位置信息，$(\boldsymbol{p}_j^{\text{lo}}-\boldsymbol{p}_i^{\text{c}})$ 表示两点坐标差值，$\oplus$ 表示向量拼接操作（Concatenate），σ 表示增强函数，本文采用全连接网络作为 σ．从式(2)中可知，$\boldsymbol{r}_j^{\text{lo}}$ 不仅包含网格 m_j^{lo} 本身的绝对位置信息，还包含其所在局部空间区域的相对位置信息．随后 $\boldsymbol{r}_j^{\text{lo}}$ 与网格原始特征信息 $\boldsymbol{f}_j^{\text{lo}}$ 进行的融合学习，如式(3)所示：

$$\boldsymbol{f}_j^{\text{fusion}}=\text{MLP}\left(\boldsymbol{r}_j^{\text{lo}}\oplus\boldsymbol{f}_j^{\text{lo}}\right) \tag{3}$$

其中 MLP(•) 为多层感知器网络．融合学习的输出 $\boldsymbol{f}_j^{\text{fusion}}$ 将作为 m_j^{lo} 的新特征向量去参与后续的权重学习，同时，中心网格 m_i^{c} 的原始特征 $\boldsymbol{f}_i^{\text{c}}$ 也会经过多层感知器将特征空间映射至与 $\boldsymbol{f}_j^{\text{fusion}}$ 相同的维度，输出为 $\boldsymbol{f}_i^{\text{H}}$．

3.3.2 局部注意力机制

本文基于局部注意力机制对局部空间区域内的网格特征信息进行聚合．对于局部网格 m_j^{lo}，其权重向量 $\boldsymbol{\alpha}_j$ 的学习方式为：

$$\boldsymbol{\alpha}_j=\text{MLP}(\boldsymbol{p}_i^{\text{c}}\oplus\boldsymbol{p}_j^{\text{lo}}\oplus\left\|\boldsymbol{p}_j^{\text{lo}}-\boldsymbol{p}_i^{\text{c}}\right\|\oplus(\boldsymbol{f}_j^{\text{fusion}}-\boldsymbol{f}_i^{\text{H}})) \tag{4}$$

其中，$\|\bullet\|$ 表示求两点间的欧式距离，$\boldsymbol{f}_j^{\text{fusion}}-\boldsymbol{f}_i^{\text{H}}$ 表示局部网格 m_j^{lo} 相对于中心网格 m_i^{c} 的特征差异，$\boldsymbol{\alpha}_j$ 与 $\boldsymbol{f}_j^{\text{fusion}}$ 具有相同的向量维度．由式(4)可知，注意力权重是同时参考局部网格和中心网格的空间位置信息和与特征信息学习得到，这样得到的权重分布更符合牙齿形状的潜在几何特征．在学习到局部网格的权重分布后，局部空间区域L的特征聚合方式为：

$$\boldsymbol{G}_{\text{L}}=\sum_{j=1}^{k}\boldsymbol{\alpha}_j\odot\boldsymbol{f}_j^{\text{local}} \tag{5}$$

其中，$\boldsymbol{G}_{\text{L}}$ 表示局部空间区域L聚合后的局部特征信息，$\odot$ 为哈达玛乘积．

在完成 N_1 个局部空间区域的特征聚合后，模块会输出 N_1 个局部特征信息 $\boldsymbol{G}_1,\boldsymbol{G}_2,\cdots,\boldsymbol{G}_{N_1}$，如图3所示．这些局部特征信息将重新作为 N_1 个中心网格的原始特征并进行下一次的局部特征提取的，直到局部特征提取

阶段结束.

3.4 特征逆向传播

特征逆向传播阶段是局部特征提取阶段的逆过程,其通过上采样和特征逆向传播将下采样后的网格数量逐步恢复至原始网格数量以做分割预测. 与PointNet++[7]类似,特征逆向传播阶段每一次上采样会使网格集合从 M_l 恢复至 M_{l-1},其中 $\left|M_l\right|=N_l$,$\left|M_{l-1}\right|=N_{l-1}$($N_l<N_{l-1}$),如图2所示. 对于网格 $m_i^{l-1}\in M_{l-1}$,其特征 $\boldsymbol{f}_i^{l-1}$ 是由集合 M_l 中距离 m_i^{l-1} 最近的3个其他网格进行特征加权而得,如式(6)所示:

$$\boldsymbol{f}_i^{l-1}=\sum_{j=1}^{3}\frac{1/\left\|\boldsymbol{p}_i^{l-1}-\boldsymbol{p}_j^{l}\right\|}{\sum_{j}^{3}\left(1/\left\|\boldsymbol{p}_i^{l-1}-\boldsymbol{p}_j^{l}\right\|\right)}\boldsymbol{f}_j^{l} \tag{6}$$

其中 $\boldsymbol{p}_i^{l-1}$ 和 $\boldsymbol{p}_j^{l}$ 分别是 m_i^{l-1} 和 m_j^{l} 的中心点坐标. 这种由少量网格恢复出更多网格特征信息的逆向传播方式与局部特征提取阶段相反. 然后,$\boldsymbol{f}_i^{l-1}$ 再通过跨层连接(skip-connection)和多层感知器与 m_i^{l-1} 在局部特征提取阶段的所对应的特征信息进行融合学习. 在经过与局部特征提取操作相同次数的上采样操作后,网格数量将重新恢复至 N. 最后,网络利用多层感知器和softmax函数输出 $N\times C$ 的分割预测分数矩阵,其中矩阵第 i 行表示网格 m_i 属于每个类别的概率,本文选择最大预测概率所对应类别作为网格 m_i 的最终分割类别.

4 实验结果与分析

4.1 实验数据

本文所使用的实验数据包括40例由人工精标注的数字化三维牙齿模型,数据来源均为口内扫描仪对不同的患者进行扫描而得. 由于每一例原始牙齿模型所包含的网格数量大约在10万到30万之间且互不相同. 为减少数据的冗余以及保持网络训练时的数据一致性,每一例牙齿模型在保证基本拓扑结构的基础上被统一下采样至16000个网格用于网络的训练和测试(即 $N=16000$). 本文定义的牙齿分割类别种数 $C=8$,包括由中切牙到第2磨牙的7种牙齿类别(左右对称)和1种牙龈类别. 本文还对训练数据进行如下数据增强操作:(1) 随机旋转角度 ε,$\varepsilon\in[-\pi/6,\ \pi/6]$;(2)随机坐标平移,平移量 $\gamma\in[-10,\ 10]$. 每一例训练数据都会进行上述两种数据增强操作以产生60例新的训练数据参与网络训练. 本文所有实验均采用3折交叉验证.

4.2 实验环境与网络参数设置

本文网络是利用Pytorch深度学习工具实现,GPU版本为NVIDIA GeForce GTX 1080,操作系统为Ubuntu 16.04 64bit. 训练时采用的优化器为Adma,损失函数为交叉熵损失函数(Cross-Entropy loss),初始学习率为0.001,每训练20轮进行0.5倍衰减,最低学习率为0.00001,训练过程中batch_size设置为4. 网络总训练轮数为200 epoch. 实际训练网络结构包含4次局部特征提取操作(局部区域构建+局部特征聚合模块)以及4次上采样操作. 进行局部区域构建时的下采样的中心网格个数和局部网格个数 k 随着次数的增加分别为[4000,2000,1000,500]和[32,32,16,16].

4.3 对比方法

在相同的实验环境和实验平台下,本文网络的分割性能与PointCNN[6],PointNet[5],PointNet++[7]以及Lian[4]等人提出三维牙齿模型分割网络MeshSegNet进行了对比. 所采用的评估指标包括分割准确率(Accuracy)、平均交并比(mean Intersection-over-Union, mIOU)以及单个类别的交并比. 四种对比方法的相关实验细节描述如下.

(1) PointNet:本文所采用的PointNet与文献[5]中的网络结构一致. 同时,为保证对比的公平性,本文采用是 $N\times 21$ 的网格初始特征信息矩阵作为PointNet的输入,与本文网络的输入保持一致. 训练过程中参数batch_size设置为4,总训练轮数为200 epoch,其他训练参数设置与3.2小节一致.

(2) PointNet++:本文所采用的PointNet++与文献[7]中的网络结构的基本一致. 在进行局部区域划分时采用与本文方法一样的 k 近邻算法和相关参数设置. PointNet++的输入包含 $N\times 21$ 的网格初始特征信息矩阵和 $N\times 3$ 的网格空间位置信息矩阵两部分,输入和本文网络保持一致. PointNet++训练时batch_size设置为4,总训练轮数为200 epoch,其他训练参数设置与3.2小节一致.

(3) PointCNN:本文所采用的PointCNN的网络结构和网络内部参数设置与文献[6]中一致,训练时batch_size设置为4,总训练轮数为200 epoch,其他训练参数与3.2小节一致.

(4) MeshSegNet:本文所采用的MeshSegNet的网络结构和网络内部参数设置与文献[4]中一致,训练时batch_size设置为4,总训练轮数为200 epoch,原文中MeshSegNet的输入为6000个网格,为保证对比的公平性,本文对将其输入扩充至16000个网格,以达到与本文方法的输入保持一致.

4.4 实验结果分析

本文方法与现有四种分割网络在3折交叉验证下的分割准确率指标和分割交并比指标对比分别如表1、表2所示. 表1中每一行数据表示不同分割网络在测试集上的分割准确率. 表2中每一行数据表示不同分割网络在单个分割类别上的分割交并比和平均交并比,其中T0表示牙龈类别,T1~T7分别表示的左右对称的

表1 本文网络与现有方法在3折交叉验证下的分割准确率对比(均值±标准差)

模型	Accuracy
PointNet[5]	0.807±0.011
PointNet++[7]	0.892±0.009
PointCNN[6]	0.879±0.008
MeshSegNet[4]	0.925±0.003
本文方法	**0.943±0.007**

中切牙至第2磨牙7种类别. 从分割结果可以看出,由于PointNet[5]在特征学习时缺失对局部特征提取,因此在准确率和mIoU两种指标上都明显低于其他方法,这说明局部形状信息对牙齿模型的分割具有十分重要的作用. PointNet++[7]和PointCNN[6]在两种分割指标上相对于PointNet[4]有明显提高,但由于它们的局部特征聚合方式相对简单,分割准确性无法进一步提高. MeshSegNet[4]在四种对比方法中取得了最好的分割性能,但由于其本质上是对局部区域内网格分配相同的权重(邻接矩阵中仅用0和1表示是否具有连接关系),忽略了真实的网格分布. 本文网络根据局部网格和中心网格的内在关系自动学习出注意力权重并进行特征聚合,这样的局部特征提取方式能更好地学习牙齿的局部形状信息,尤其是在牙齿边缘或相邻牙齿区域更具有优势. 所以从实验结果可以看出本文方法在准确率和IoU上都明显优于另外四种对比方法.

本文网络与其他方法的分割结果可视化对比如图4所示. 通过对比可知,本文方法的分割明显优于其他四种对比方法. 例如,在第一行所示牙齿模型在侧切牙区域(红色箭头所指处)存在较为严重的牙齿错位,以及第二行所示牙齿模型两侧的尖牙区域(蓝色箭头所指)也存在明显的缺牙现象. 其他四种分割网络在上述的两个区域都存在一定程度的过分割和欠分割现象. 而本文网络由于能在牙齿边缘区域学习出更细节的特征差异,因此即使牙齿模型中存在上述分割难点,

表2 本文网络与现有方法在3折交叉验证下的分割交并比对比

模型	T0	T1	T2	T3	T4	T5	T6	T7	mIoU
PointNet[5]	0.826±0.005	0.413±0.671	0.495±0.087	0.528±0.066	0.619±0.068	0.621±0.059	0.567±0.110	0.684±0.064	0.590±0.006
PointNet++[7]	0.863±0.006	0.719±0.038	0.741±0.044	0.734±0.033	0.787±0.009	0.757±0.196	0.705±0.137	0.776±0.003	0.760±0.198
PointCNN[6]	0.849±0.006	0.696±0.030	0.731±0.024	0.725±0.003	0.786±0.015	0.748±0.010	0.651±0.017	0.721±0.016	0.738±0.014
MeshSegNet[4]	0.911±0.009	0.811±0.051	0.821±0.015	0.814±0.005	0.838±0.008	0.840±0.009	0.811±0.005	0.847±0.023	0.837±0.009
本文方法	**0.915±0.008**	**0.848±0.015**	**0.877±0.019**	**0.856±0.032**	**0.880±0.020**	**0.868±0.018**	**0.831±0.028**	**0.848±0.026**	**0.865±0.019**

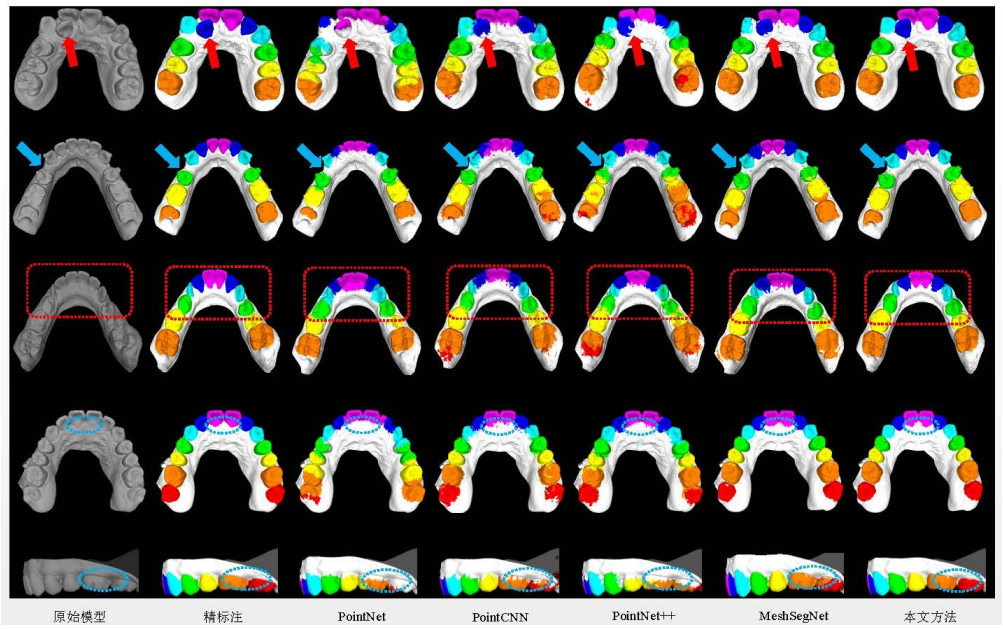


图4 本文网络与四种对比网络的分割结果可视化对比

本文网络仍然准确地分割出较为完整的牙齿结构. 第三行展示的数据的牙齿形状和其他模型相比具有较大的差异(红色虚线所示),而本文网络在此区域内的分割结果也明显优于其他网络,这也说明本文网络具有较强的泛化能力. 由第四行和第五行展示的数据在牙齿边界分割结果可知,PointNet[5]在边界部分存

在十分严重的欠分割问题,这进一步证明了局部特征信息对于牙齿边缘分割十分重要. PointCNN[6]和PointNet++[7]虽然效果要优于PointNet[5],但依旧存在较为严重的牙齿多分现象. MeshSegNet[4]在整体上取得比较准确的边缘分割准确率,但其在磨牙部分容易出现过分割现象(如第五行蓝色虚线框所示),这说明该方法在相邻牙齿区域的分割能力还存在不足. 本文利用基于局部注意力机制的特征聚合方式能帮助网络能提取更细节的局部形状信息,因此在牙齿边缘区域分割效果明显优于其他四种方法.

4.5　消融实验

4.5.1　初始特征信息组合

本文所使用的初始特征信息除了三维牙齿模型中各网格顶点的坐标V外,还增加了网格的法向量$\boldsymbol{N}_{\text{mesh}}$和网格顶点的法向量$\boldsymbol{N}_{\text{vertex}}$. 为验证这些额外的初始特征信息在提升网络分割性能上的有效性,本文在保持其它条件不变的情况下,使用不同的初始特征信息组合作为输入进行网络训练以及测试结果对比.

输入的特征信息组合包括:(1)仅使用顶点坐标;(2)顶点坐标+网格法向量;(3)顶点坐标+顶点法向量;(4)顶点坐标+网格法向量+顶点法向量. 分割指标对比在3折交叉验证下如表3所示. 由表3可知,随着初始特征信息逐渐丰富,网络的分割准确率和平均交并比也相应提高. 当输入包含所有初始特征信息时,网络达到最好的分割性能. 通过观察可知,在顶点坐标信息的基础上增加顶点法向量相对于增加网格法向量,网络分割性能提升得更为明显. 一个可能的原因是网格法向量信息仅是针对单个网格计算而得,而每个顶点的法向量信息是包含该顶点的周围所有网格的法向量信息,因此顶点法向量含有的空间特征更丰富.

四种输入组合所训练的网络的分割结果的可视化如图5所示. 由于空间相邻网格的坐标信息十分相似,所以当输入仅含有顶点坐标信息时,网络无法很好地学习出网格之间的特征差异,因此在牙齿边缘等区域存在一定程度的过分割现象. 然而,位于牙齿边缘区域的网格拓扑形状具有较为明显的变化,这使得网格在向量信息上具有很高的特征辨识度. 所以当输入的特征信息包含网格法向量和顶点法向量后,可以更好地辅助网络学习出边缘区域的网格特征差异,因此其分割结果在牙齿边缘区域更加准确光滑.

表3　本文网络在不同输入组合下的分割指标

输入组合	Accuracy	mIoU
顶点坐标	0.887±0.008	0.752±0.167
顶点坐标+网格法向量	0.919±0.060	0.820±0.013
顶点坐标+顶点法向量	0.934±0.012	0.851±0.005
顶点坐标+顶点法向量+网格法向量	**0.943±0.007**	**0.865±0.019**

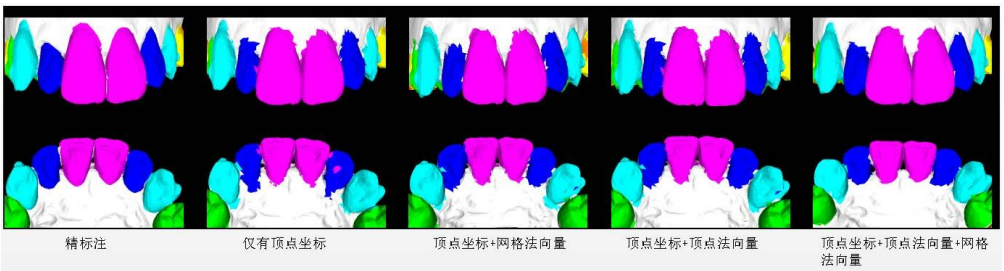


图5　不同输入组合的分割结果可视化对比

4.5.2　空间信息增强和局部注意力机制

为验证空间信息增强和局部注意力机制对提升网络分割性能的有效性,本文分别对这两个模型进行了消融实验. 实验设置如下:(1)仅使用空间信息增强,局部特征聚合方式采用最大值池化;(2)仅局部注意力机制;(3)本文完整网络结构. 上述三种网络结构的分割指标如表4所示. 实验结果表明,若只使用空间信息增强或只使用局部注意力机制,网络都无法达到最好的分割性能. 同时通过与完整网络结构的分割结果对比可知,在使用空间信息增强的基础上增加局部注意力机制,网络的分割准确率可提升3.3%,反之在使用局部注意力机制的基础上增加空间信息增强,网络的分割准确率可提升1.6%,实验结果验证了本文提出的空间信息增强和局部注意力机制都使得网络的分割性能得到进一步提升. 分割结果的可视化对比如图6所示,通过对比仅使用局部注意力机制和仅使用空间信息增强两种情况下网络

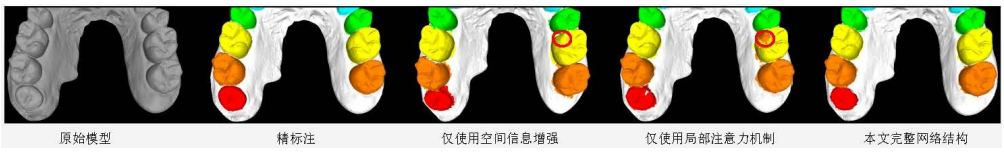


图6　本文网络使用不同模块的分割结果可视化对比

的分割结果可知,前者在牙齿边缘区域的分割准确性明显优于后者,这也进一步说明基于局部注意力机制的特征聚合相对于最大值池化能学习牙齿更多的细节形状信息. 但仅使用局部注意力机制的网络结构却在牙齿区域出现了部分错误分割的现象(红色实线区域所示),其主要原因是该例牙齿模型右侧缺少第二磨牙,从而导致左右部分牙齿分布不对称. 而仅使用了空间信息增强的网络结构虽然在牙齿边缘区域分割性能欠佳,但其同时参考了网格的绝对位置信息和相对位置信息,因此并没有受到因牙齿分布不均匀带来的影响. 本文完整网络结构同时具有上述两个模块的优点,且从分割结果可知,空间信息增强对局部注意力机制具有一定的促进作用.

表4 本文网络使用不同模块的分割指标

模型	Accuracy	mIoU
仅空间信息增强	0.910±0.004	0.802±0.009
仅局部注意机制	0.926±0.004	0.835±0.006
完整网络结构	**0.943±0.007**	**0.865±0.019**

4.5.3 不同网格分辨率对网络分割性能的影响

为讨论网络在输入不同网格分辨率(即不同的输入网格个数N)的牙齿模型时分割性能的变化,本文在网络训练阶段对牙齿模型中的网格数量进一步随机下采样至$N=12000$、$N=8000$和$N=4000$进行网络训练,在网络测试阶段依旧保持$N=16000$进行网络分割性能测试. 分割指标对比如表5所示. 实验结果表明,随着牙齿模型在局部区域的网格分辨率降低,网络的分割指标也相应降低. 其中mIoU的下降程度最大,其原因是当三维牙齿模型所具有的网格数量越少,其局部区域所能提供的空间信息将更加粗糙,网络难以学习出识别度高的特征用于分割预测. 然而,即使在$N=12000$的网格分辨率下,本文网络分割指标依旧优于其他对比方法在$N=16000$的网格分辨率下所取得的分割指标,这也进一步证明了本文网络的鲁棒性.

表5 本文网络在不同网格分辨率下的分割指标

分辨率	Accuracy	mIoU
$N=4000$	0.884±0.006	0.768±0.016
$N=8000$	0.916±0.032	0.815±0.136
$N=12000$	0.927±0.005	0.841±0.012
$N=16000$	**0.943±0.007**	**0.865±0.019**

5 结论

针对三维牙齿模型的分割任务,本文提出一种基于局部注意力机制的端到端分割网络. 网络先通过对三维牙齿模型进行多尺度的局部区域构建,并利用空间信息增强模块对三维网格进行特征丰富. 在此基础上,网络再根据区域内网格的真实空间分布和网格特征差异自动学习注意力权重,并基于该权重进行局部特征聚合以帮助网络自适应地去关注不同局部区域内更具有表达性的网格特征,有效地解决了现有方法存在的局部特征提取问题. 通过在临床数据集上的实验表明,本文网络相对于现有的部分方法在牙齿边缘区域能取得更好的分割性能.

参考文献

[1] XU Xiao-jie, LIU Chang, ZHENG You-yi. 3D tooth segmentation and labeling using deep convolutional neural networks[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 25(7): 2336-2348.

[2] TIAN Su-kun, DAI Ning, ZHANG Bei, et al. Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks[J]. IEEE Access, 2019, 7(1): 84817-84828.

[3] ZANJANI F G, MOIN D A, VERHEIJ B, et al. Deep learning approach to semantic segmentation in 3d point cloud intra-oral scans of teeth[C]//International Conference on Medical Imaging with Deep Learning. London: PMLR, 2019: 557-571.

[4] 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]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2019: 837-845.

[5] QI C R, SU Hao, MO Kai-chun, et al. Pointnet: deep learning on point sets for 3d classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Press, 2017: 652-660.

[6] LI Yang-yan, BU Rui, SUN Ming-chao, et al. Pointcnn: convolution on x-transformed points[C]//Advances in Neural Information Processing Systems. New York: ACM Press, 2018: 820-830.

[7] QI C R, YI Li, SU H, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space[C]//Advances in Neural Information Processing Systems. New York: ACM Press, 2017: 5099-5108.

[8] YUAN Tian-ran, LIAO Wen-he, DAI Ning, et al. Single-tooth modeling for 3D dental model[J]. International Journal of Biomedical Imaging, 2010, 2010(1): 1-14.

[9] ZHAO Ming-xi, MA Li-zhuang, TAN Wu-zheng, et al. Interactive tooth segmentation of dental models[C]//2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. Shanghai: IEEE Press, 2006: 654-657.

[10] KUMAR Y, JANARDAN R, LARSON B, et al. Improved segmentation of teeth in dental models[J]. Computer-Aided Design and Applications, 2011, 8(2): 211-224.

[11] KRONFELD T, BRUNNER D, BRUNNETT G. Snake-based segmentation of teeth from virtual dental casts[J]. Computer-Aided Design and Applications, 2010, 7(2): 221-233.

[12] WU Kan, CHEN Li, LI Jing, et al. Tooth segmentation on dental meshes using morphologic skeleton[J]. Computers & Graphics, 2014, 38(1): 199-211.

[13] SINTHANAYOTHIN C, THARANONT W. Orthodontics treatment simulation by teeth segmentation and setup [C]//2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. Los Alamitos: IEEE Press, 2008: 81-84

[14] MA Ya-qi, LI Zhong-ke. Computer aided orthodontics treatment by virtual segmentation and adjustment[C]// 2010 International Conference on Image Analysis and Signal Processing. Xiameng: IEEE Press, 2010: 336-339.

[15] ZOU Bei-ji, LIU Shi-jian, LIAO Sheng-hui, et al. Interactive tooth partition of dental mesh base on tooth-target harmonic field[J]. Computers in Biology and Medicine, 2015, 56(1): 132-144.

[16] 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.

[17] WONGWAEN N, SINTHANYOTHIN C. Computerized algorithm for 3D teeth segmentation[C]//2010 International Conference on Electronics and Information Engineering. Kyoto: IEEE Press, 2010: 277-280.

[18] 罗会兰, 张云. 基于深度网络的图像语义分割综述[J]. 电子学报, 2019, 47(10): 2211-2220.
LUO Hui-lan, ZHANG Yun. A survey of image semantic segmentation based on deep network[J]. Acta Electronica Sinica, 2019, 47(10): 2211-2220. (in Chinese)

[19] 梁新宇, 林洗坤, 权冀川, 肖铠鸿. 基于深度学习的图像实例分割技术研究进展[J]. 电子学报, 2020, 48(12): 2476-2486.
LIANG Xin-yu, LIN Xi-kun, QUAN Ji-chuan, XIAO Kai-hong. Research on the progress of image instance segmentation based on deep learning[J]. Acta Electronica Sinica, 2020, 48(12): 2476-2486. (in Chinese)

[20] WANG Peng-Shuai, LIU Yang, GUO Yu-xiao, et al. O-cnn: Octree-based convolutional neural networks for 3d shape analysis[J]. ACM Transactions on Graphics , 2017, 36(4): 1-11.

[21] LE T, BUI G, Duan Ye. A multi-view recurrent neural network for 3D mesh segmentation[J]. Computers & Graphics, 2017, 66(1): 103-112.

[22] KALOGERAKIS E, AVERKIOU M, MAJI S, et al. 3D shape segmentation with projective convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Press 2017: 3779-3788.

[23] DAI A, NIESSNER M. 3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation[C]// Proceedings of the European Conference on Computer Vision. Munich: IEEE Press, 2018: 452-468.

[24] YOU Hao-xuan, FENG Yi-fan, JI Rong-rong, et al. Pvnet: A joint convolutional network of point cloud and multi-view for 3d shape recognition[C]//Proceedings of the 26th ACM International Conference on Multimedia. New York: ACM Press, 2018: 1310-1318.

[25] KLOKOV R, LEMPITSKY V. Escape from cells: Deep kd-networks for the recognition of 3d point cloud models [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE Press, 2017: 863-872.

[26] RIEGLER G, OSMAN U A, GEIGER A. Octnet: Learning deep 3d representations at high resolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Press, 2017: 3577-3586.

[27] GRAHAM B, ENGELCKE M, MAATEN L VAN DER. 3d semantic segmentation with submanifold sparse convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Press, 2018: 9224-9232.

[28] WANG Zong-ji, LU Feng. VoxSegNet: Volumetric CNNs for semantic part segmentation of 3D shapes[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 26(9): 2919-2930.

[29] MENG H Y, GAO Lin, LAI Yu-kun, et al. VV-Net: Voxel vae net with group convolutions for point cloud segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul: IEEE Press, 2019: 8500-8508.

[30] 彭秀平, 仝其胜, 林洪彬, 等. 一种面向散乱点云语义分割的深度残差-特征金字塔网络框架[J]. 自动化学报, 2021, 47(12): 2831-2840.
PENG Xiu-Ping, TONG Qi-Sheng, LIN Hong-Bin, et al.

A deep residual — feature pyramid network framework for scattered point cloud semantic segmentation[J]. Acta Automatica Sinica, 2021, 47(12): 2831-2840.(in Chinese)

[31] 盖杉, 王俊生. 基于深度学习的非局部注意力增强网络图像去雨算法研究[J]. 电子学报, 2020, 48(10): 1899-1908.
GAI Shan, WANG Jun-sheng. Image Raindrop Algorithm Research Using Nonlocal Attention Enhanced Network Based on Deep Learning[J]. Acta Electronica Sinica, 2020, 48(10): 1899−1908. (in Chinese)

[32] 唐海桃, 薛嘉宾, 韩纪庆. 一种多尺度前向注意力模型的语音识别方法[J]. 电子学报, 2020, 48(07): 1255-1260.
TANG Hai-tao, XUE Jia-bin, HAN Ji-qing. A method of multi-scale forward attention model for speech recognition[J]. Acta Electronica. Sinica, 2020, 48(07): 1255-1260. (in Chinese)

[33] 张志昌, 曾扬扬, 庞雅丽. 融合语义角色和自注意力机制的中文文本蕴含识别[J]. 电子学报, 2020, 48(11): 2162-2169.
ZHANG Zhi-chang, ZENG Yang-yang, PANG Ya-li. A chinese textual entailment recognition method incorporating se mantic role and self-attention[J]. Acta Electronica. Sinica, 2020, 48(11): 2162-2169. (in Chinese)

作者简介

张凌明 男, 1997 年 9 月生于重庆, 现为重庆邮电大学通信与信息工程学院硕士研究生, 主要研究方向为计算机视觉、医学图像处理.
E-mail: zhanglingming1997@qq.com

赵 悦 女, 1988 年 10 月生于吉林. 2017 年于长春理工大学获得学士学位, 2017 年获得吉林大学博士学位. 现任重庆邮电大学通信与信息工程学院讲师. 主要研究方向包括图像处理和分析、模式识别、医学图像处理.
E-mail: zhaoyue@cqupt.edu.cn

李鹏程 男, 1995 年 12 月生于重庆. 现为重庆邮电大学通信与信息工程学院博士研究生. 主要研究方向为医学图像处理与分析, 计算机视觉和机器学习.
E-mail: lipengchengme@163.com

刘 洋 男, 1987 年 6 月生于重庆, 2014 年于北京大学口腔医学院获得口腔医学博士, 2017 年获得北京大学口腔医学院正畸学博士. 现任重庆医科大学附属口腔医院正畸科医师. 主要研究方向, 正畸牙移动的生物学机制, 牙齿图像处理及应用.
E-mail: yangliu@cqmu.ecu.cn

高陈强(通讯作者) 男, 1981 年 8 月生于重庆. 于华中科技大学获得博士学位. 现任重庆邮电大学通信与信息工程学院一名教授和博士生导师. 主要研究方向包括红外图像分析、目标检测与识别、行为识别.
E-mail: gaocq@cqupt.edu.cn