

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2024)03-0686-11

论文引用格式: Qin J, Lu T L, Ji B and Li Y Q. 2024. Tooth segmentation network for low-dose CT. Journal of Image and Graphics, 29(03):0686-0696(秦俊, 卢婷岚, 纪柏, 李雨晴. 2024. 面向低剂量CT的牙齿分割网络. 中国图象图形学报, 29(03):0686-0696)[DOI: 10.11834/jig.230540]

面向低剂量CT的牙齿分割网络

秦俊¹, 卢婷岚¹, 纪柏^{2*}, 李雨晴¹

1. 长春理工大学计算机科学技术学院, 长春 130000; 2. 吉林大学白求恩第一医院, 长春 130000

摘要: 目的 锥形束计算机断层扫描(cone beam computer tomography, CBCT)已成为口腔诊疗领域中最常用的一项医学影像技术。由于CBCT图像本身对比度低且牙齿形状复杂,在进行牙齿分割时容易导致分割边界模糊、牙齿根部错误分割的问题。现有方法往往无法达到预期效果,并且基于深度学习的分割网络在分割精度等性能提升到一定程度后存在生梯度爆炸、过拟合以及无法关注图像全局信息等限制。然而,牙齿分割在医生制定诊断和治疗计划方面至关重要。为了应对这一问题,提出了一种名为MF-CA Net的牙齿分割模型,以提高牙齿分割的准确性和鲁棒性。方法 MF-CA Net模型引入了多尺度特征提取模块(multi-scale feature extraction module, MFEM)和CA(coordinate attention)注意力机制,这些模块使网络能够准确地捕捉感兴趣的牙齿区域,并提取丰富而密集的多尺度特征信息,从而有效地指导分割任务。特别是在牙根分割方面,这些模块能够显著提高分割的精度。为了进一步提升分割算法的性能,还引入了联合损失函数,该损失函数综合考虑了像素级、局部级和全局级3个方向的牙齿边缘分割,以提高算法的准确性和稳健性。结果 实验在数据集上对MF-CA Net模型与6种主流方法进行了比较。实验结果表明,相较于其他分割方法,MF-CA Net模型在各项评价指标上都取得了显著的提升。尽管在精度(accuracy)指标上稍低于DeeplabV3+,但在Dice评价指标上达到了0.9495的高分数,相比PyConvU-Net提高了4%,相对于DeeplabV3+提高了约4%,对比U-Net提高了约16%。此外,平均交并比(mean intersection-over-union, mIoU)指标提升了3%~11%,F2值提升了5%。结论 本文提出的MF-CA Net网络模型可以实现对牙齿的精确分割。

关键词: 深度学习; 锥形束计算机断层扫描(CBCT); 牙齿分割; 注意力机制; 多尺度信息; 损失函数; 分割精度

Tooth segmentation network for low-dose CT

Qin Jun¹, Lu Tinglan¹, Ji Bai^{2*}, Li Yuqing¹

1. School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130000, China;

2. Baiqiu'en First Hospital of Jilin University, Changchun 130000, China

Abstract: Objective The field of dentistry can be developed and improved with the mutual penetration and integration of computer technology and modern dentistry. Cone beam computed tomography (CBCT) has become one of the most commonly used medical imaging techniques in dental diagnoses and treatments. CBCT has the advantages of low radiation dosage, simple operation, and low cost. However, the noise and artifacts of CBCT are more intense than those of conventional CT, and fuzzy tooth boundaries will affect doctors' diagnoses and subsequent treatments. In oral diagnoses and treatments, doctors usually need to manually segment the tooth model in CBCT to formulate subsequent treatment plans. However, this method is time-consuming and labor-intensive, and the segmentation results of the teeth are also significantly affected by

收稿日期: 2023-08-21; 修回日期: 2023-09-21; 预印本日期: 2023-10-04

* 通信作者: 纪柏 ji_bai@jlu.edu.cn

基金项目: 吉林省自然科学基金项目(20210204101YY)

Supported by: Natural Science Foundation of Jilin Province, China (20210204101YY)

doctors' subjective factors. In addition, the existing network method often fails to achieve the expected results. The segmentation network based on deep learning also relatively exists in the segmentation accuracy and other performance after network segmentation to improve gradient explosion, overfitting, and over-expression. Limitations include gradient explosion, overfitting, and the inability to focus on global image information. Therefore, people have been working to find a dental segmentation method with high automation and high accuracy. To address this problem, a dental segmentation model called multi scale feature extraction module and coordinate attention (CA) mechanism network (MF-CA Net) is proposed, which uses a series of innovative methods to improve the accuracy and robustness of dental segmentation. **Method** The MF-CA Net network uses the multi-scale feature extraction module (MFEM) to extract features at different scales of images and utilizes the CA attention mechanism that is currently excelling in improving network performance. MFEM uses four different convolution kernels for convolution, enabling the extraction of multi-scale features and facilitating the network to learn markedly robust representations. Meanwhile, dilation convolution uses four dilation rates to further increase the receptive field, enabling the network to obtain significantly detailed information and refine important features. The CA attention mechanism calculates the spatial and channel attention weights in the input feature maps. It adaptively weights the feature map, emphasizing more representative local structures and global contextual information. By embedding positional information into the channel attention, the CA mechanism assists the network in accurately localizing and identifying the objects of interest. These modules enable the network to accurately determine the tooth region of interest and extract extensive and dense multi-scale feature information to effectively guide the segmentation task. For tooth root segmentation, these modules can significantly improve the accuracy of segmentation. To further improve the performance of the segmentation algorithm, the MF-CA Net network model also uses structural similarity to construct the boundary loss function. Moreover, the algorithm uses a combination of the Dice, binary cross-entropy, and structural similarity (SSIM) loss functions as final loss function. The Dice loss function is used to compute the similarity between two sets of images, whereas the cross-entropy loss function is used to predict the segmentation result and pixels corresponding to the real segmentation result. This loss function integrates tooth edge segmentation in three directions, namely, pixel, local, and global levels, to improve the accuracy and robustness of the algorithm. **Result** To more accurately evaluate the performance of the proposed model in tooth segmentation, Dice similarity coefficient, mean intersection to merger ratio (mIoU), accuracy, recall, precision, and F2 score are used as evaluation metrics. This study compares the MF-CA Net model and six mainstream methods on the dataset. Experimental results show that the MF-CA Net model has significant improvement in most of the evaluation metrics compared with other segmentation methods. Although MF-CA Net is slightly lower than DeeplabV3+ in accuracy metrics, it achieves a high score of 0.949 5 in the Dice evaluation metrics, which is an improvement of 4% compared with PyConvU-Net, about 4% compared with DeeplabV3+, and about 16% compared with U-Net. In addition, the mIoU metric improves from 3% to nearly 11%. Precision value reaches 0.942 1, which is a 7% improvement compared with UNET++. The recall metric reaches 0.968 7, which is an 8% improvement compared with the UNET network. Lastly, the F2 metric reaches 0.954 3, which is a 5% improvement compared with the Res-UNet value. Results fully demonstrate the superiority of the MF-CA network model in tooth segmentation. **Conclusion** The proposed MF-CA network model successfully solves the difficult problem of tooth segmentation in CBCT images by introducing a multiscale feature extraction module, an attention mechanism, and a hybrid loss function. Many experimental results verify the proposed model's superiority in accurate tooth segmentation. Lastly, the proposed model is expected to be widely used in dental diagnoses and treatments, which is significant in oral diagnoses and treatments.

Key words: deep learning; cone beam computed tomography (CBCT); tooth segmentation; attention mechanism; multi-scale information; loss function; segmentation accuracy

0 引言

随着计算机技术和现代口腔医学相互渗透融

合,口腔数字化技术的高速发展正推动着口腔医学的进步。由计算机断层扫描(computed tomography, CT)演化而来的锥形束CT(cone beam computed tomography, CBCT)实现了人体头部及口腔的三维成

像,使口腔医学得以进一步精准化和数字化。相比常用于其他人体部位的CT照射方式,CBCT照射具有成像速度快、辐射剂量低等优点,可以对不耐辐射人群使用,并且针对口腔这一靠近人体大脑部位进行低剂量CT照射,可以降低对人体辐射伤害。并且针对阻生牙、畸形牙、龋齿、缺牙以及牙列不齐等常见口腔疾病,需要以患者牙齿结构为参考,根据CBCT的结果,医生可以对患者的不同问题采用不同的治疗办法。

但是由于CBCT放射量低,参与成像的放射物质较少,所以成像清晰度相对于常剂量CT略低。并且牙齿本身形状复杂,在进行牙齿分割时容易导致分割边界模糊、牙齿根部错误分割的问题,故而通过图像处理算法对低剂量CT进一步处理,方便医生更快速地获取清晰的口腔CT。临床上,牙齿分割一般由有经验的医生对患者口腔CBCT数据的牙冠和牙根区域进行手动定位和勾画,这一过程不仅耗时,而且非常依赖医生的解剖学知识和临床经验,分割结果受主观因素影响较大。如今,人工处理图像已不能满足临床需求,因此建立一种自动化程度相对较高且准确的牙齿分割方法已成当务之急,牙齿自动分割既可以减轻医生的工作量,又可以提高诊断率和手术的治愈率。然而,从CBCT图像中自动分割牙齿仍然是一个具有挑战性的问题,因为牙齿在拓扑结构上表现出很大的变化。例如,磨牙可能在牙根处分裂成2个或3个结构。除了一般规则外,还存在磨牙只有1个牙根和3个以上牙根的特殊情况,这些情况在现实生活中相当常见(Xu等,2022)。

目前已有更多关于牙齿分割的算法相继提出,但其中大多数是基于水平集(Zhang, 2011; Gao和Chae, 2010; Gan等, 2015)或区域生长算法(Ji等, 2014; Gan等, 2018; Xia等, 2017)。这些方法通常需要强大的先验知识和优秀的初始化,但即使这样也不能从这些区域的背景组织中正确分割牙齿。

随着人工智能的发展和应用,卷积神经网络在图像处理领域崭露头角,出现了许多成功案例,这为牙齿分割任务带来了新的研究方向。基于卷积神经网络的图像分割算法大量涌现并取得了优秀的性能,其中全卷积网络(fully convolutional network, FCN)(Long等, 2015)和U-Net及其变体网络(Li等, 2021; 周涛等, 2021)在医学图像分割中表现尤为突出。率先使用深度学习对CBCT图像中的牙齿

进行分割的团队(Cui等, 2019)使用3D Mask RCNN作为基本网络,实现了对口腔CBCT图像中牙齿的自动分割和分类。Wirtz等人(2018)结合梯度图像特征与关于牙齿形状变化的空间关系的统计知识进行牙齿分割,该方法取得了一定的效果,但该模型无法分割智齿。Chung等人(2020)提出了一个姿势感知实例分割框架,用于分割口腔CBCT图像中的单颗牙齿。Zhao等人(2020)提出使用长短时记忆网络来分割口腔CBCT图像中的牙齿。Koch等人(2019)将基于U-Net架构的全卷积神经网络应用于牙齿分割任务,利用一般性的分割技巧获得了不错的分割效果。尽管目前已经提出了一些自动方法来提高牙齿分割的准确性,但由于CBCT图像本身的低对比度和牙齿的形状轮廓复杂,这些方法对智齿和牙根区域的分割效果并不尽如人意,都存在信息丢失的问题。

在本文中提出了一种基于多尺度特征提取模块(multi scale feature extraction module, MFEB)和CA(coordinate attention)的自动牙齿分割方法。具体来说, MFEB通过扩大感受野来使网络捕获全局和局部特征,并且采用了目前在提升网络性能方面表现优秀的CA注意力机制。该注意力机制可以更好地捕捉局部和全局上下文信息,更准确地定位和识别感兴趣的对象,这有助于提高分割的准确性。在损失函数方面,本文从像素、局部和整体3个方向优化牙齿边缘的分割效果,提高了分割算法的准确性和鲁棒性。实验结果表明,该网络模型明显优于目前多数主流分割网络的效果,这表明了设计的方法对牙齿分割的有效性,对口腔医学的诊断和治疗有重要意义。

1 相关工作

1.1 数据集预处理

数据集采用Dicom格式存储,为了后续方便传入网络等操作,本文将CBCT数据由16 bit映射到8 bit。由于CBCT采用的是锥形束投影,扫描到的视野范围比较广,所以得到的数据通常包括颌骨等多种信息,牙齿在其中仅占很小部分。给分割工作带来很大的困难。基于此,对CBCT图像进行了裁剪,去除图像中的无用区域,使牙齿的特征更加明显,便于输入网络进行特征提取。裁剪后图像尺寸

为 960×720 像素。

由于辐射剂量的限制, CBCT图像会存在图像质量差、噪声增加的问题。为了不影响牙齿分割的准确性, 对裁剪后的图像进行去噪处理, 为牙齿分割工作打下良好的基础。

1.2 数据标记

深度学习算法通常需要依赖于精确标注的数据集。然而, 目前并没有可供研究和应用的口腔CBCT标注数据集。因此, 需要对收集到的口腔数据进行手动标记, 创建带有标签的图像数据集, 以便进行后续的网络训练和测试。

牙齿分割的数据集包括预处理后的CBCT图像和与其对应的标签图像。其中标签图像是手工标注后生成的一种与原始图像尺寸相同的二值图像。标签图像中, 牙齿区域被标注为前景, 而其他组织区域被标注为背景。如图1所示为分割数据集示例。

2 牙齿分割框架

2.1 网络模型

牙齿分割模型MF-CA Net的整体架构如图2所示。首先使用预训练的ResNet50(residual network

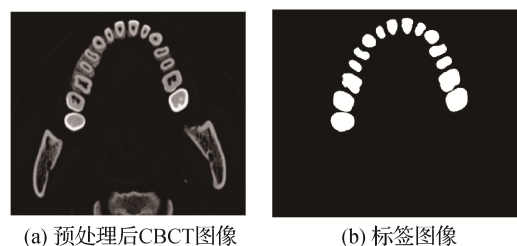


图1 分割数据集示例

Fig. 1 Examples of segmented dataset

((a) pre-processed CBCT image; (b) labeled image)

50)(He等, 2016)作为编码器, 从输入图像中提取4个不同的特征图。然后, 特征图中的每一个都需要通过一系列 3×3 卷积层、批量归一化和ReLU (linear rectification function) 激活函数后送入MFEB模块, MFEB的详细结构如图3所示。使用该模块旨在尝试减少当牙齿拓扑高度变化且牙齿边缘模糊时出现的错误分割。MFEB使用不同尺寸的卷积核分别对输入特征图进行处理, 从而提取牙齿CBCT图像的多种特征, 增强网络在提取图像特征信息方面的鲁棒性, 从而使网络能够捕捉更多细节并优化突出牙齿图像的重要特征, 最终优化分割效果。

解码器部分首先对特征图进行双线性上采样, 将输入特征图的高度和宽度增加两倍(朱翌和李秀,

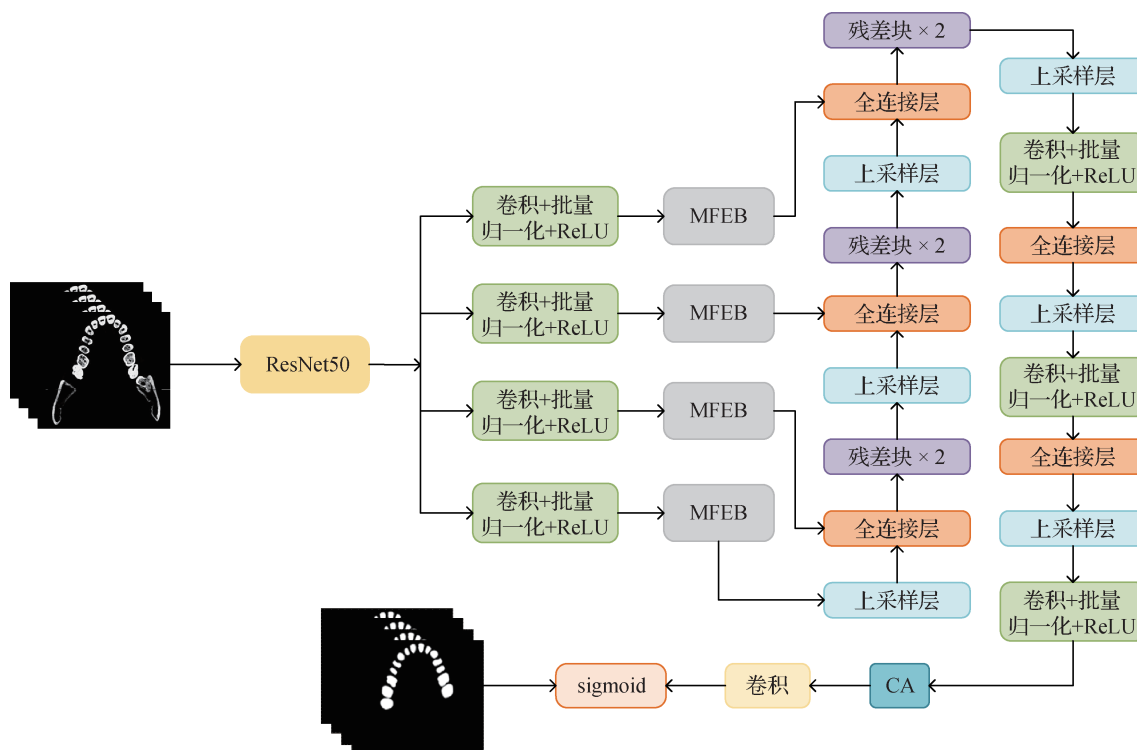


图2 MF-CA Net模型结构图

Fig. 2 Structure of MF-CA Net model

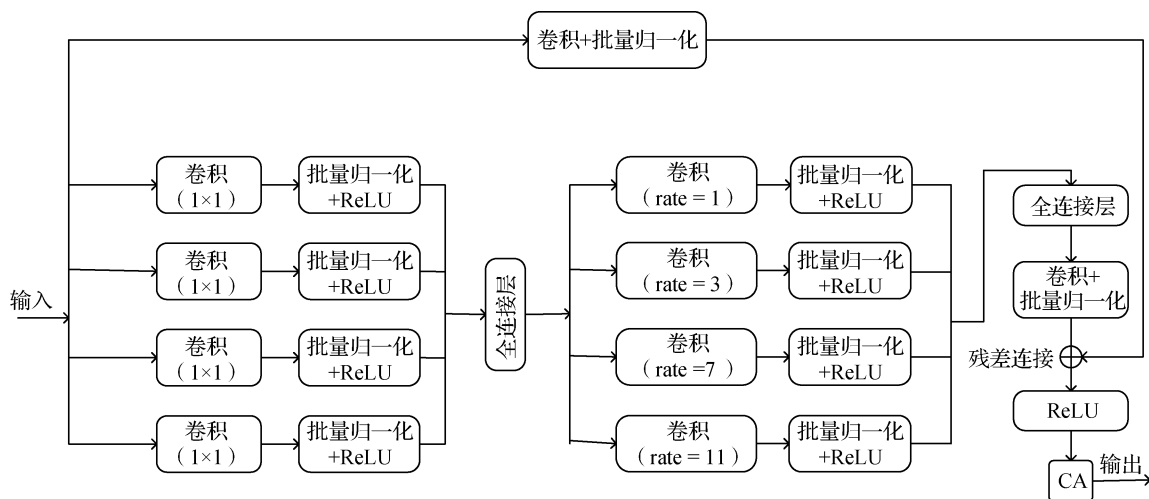


图3 多尺度特征提取模块结构图

Fig. 3 Structure of multi-scale feature extraction module

2023)。将上采样的特征图与另一个MFEB的输出连接起来,这样可以使解码器获取到更多的语义信息,增加了它的特征表示。通过两个残差块,其中每个残差块由一个卷积块和一个连接卷积块的输入和输出的恒等映射组成。卷积块以两个 3×3 卷积层开始,每个卷积层后都有一个批量归一化和一个ReLU激活函数。将第1个残差块的输出传递给双线性上采样层,以将其空间维度增加两倍。之后是 3×3 卷积层、批量归一化和ReLU激活函数。然后将ReLU激活函数的输出与第2个残差块的输出连接起来。再通过一个双线性上采样层,其中连接的特征图被上采样两倍,然后是一个 3×3 卷积层、批量归一化和ReLU激活函数。然后将ReLU激活函数的输出与第3个残差块的输出连接起来。最后,特征图再次进行上采样并通过 3×3 卷积层、批量归一化和ReLU激活函数。然后将特征图通过CA注意力模块,它可以增强网络的特征的表达能力,帮助网络选择有效的特征并抑制不相关特征的影响,从而提高特征表示及其鲁棒性。

2.2 多尺度特征提取模块

提出的MFEB的详细结构如图3所示。对于输入的特征图首先使用4个卷积核大小不同的卷积($1 \times 1, 3 \times 3, 7 \times 7, 11 \times 11$),卷积核的大小不同,对输入图像数据做出的卷积运算也不同。换言之,所提取出来的图像特征信息也不同。多尺度卷积核可以有效避免在卷积核设置上对整体网络模型提取图像特征信息能力的限制,增强了网络提取图像特征信息的鲁棒性,使网络能够学习更强大的

表示。

然后再用4个扩张率分别为1、3、7、11的扩张卷积进一步增大感受野。不同的感受野会带来多尺度上下文信息,使网络能够捕捉到更多的细节,并提炼出重要的特征,这对于分割任务是非常重要的。每个卷积层之后是批量归一化和ReLU激活函数。之后,对这些特征进行连接,并将它们传入 1×1 卷积,然后是残差连接。最后,生成的特征图通过CA注意力机制进一步突出特征。

2.3 CA模块

已经证明注意力机制能够在各种计算机视觉任务中发挥重要作用,包括图像分类和图像分割。通道注意力是当前备受关注的技术之一,它可以显著提高模型的性能,但是通常情况下会忽略位置信息。而在计算机视觉任务中,位置信息通常至关重要。因此,研究人员引入了一种创新的注意力机制,称为CA。CA注意力机制通过计算输入特征图中的空间和通道注意力权重,自适应地调整特征图的权重分布,以更好地捕捉局部结构和全局上下文信息。更重要的是,CA注意力机制将位置信息与通道注意力相结合,从而有助于网络更准确地定位和识别感兴趣的对象。CA注意力模块的结构如图4所示。

通道注意力机制通常采用二维全局池化操作,将特征张量压缩为单个特征向量。然而,CA注意力采用了一种不同的方法,将通道注意力分解为两个一维特征编码过程,分别沿两个空间方向整合特征信息。这个设计允许网络更好地捕捉到视觉任务中

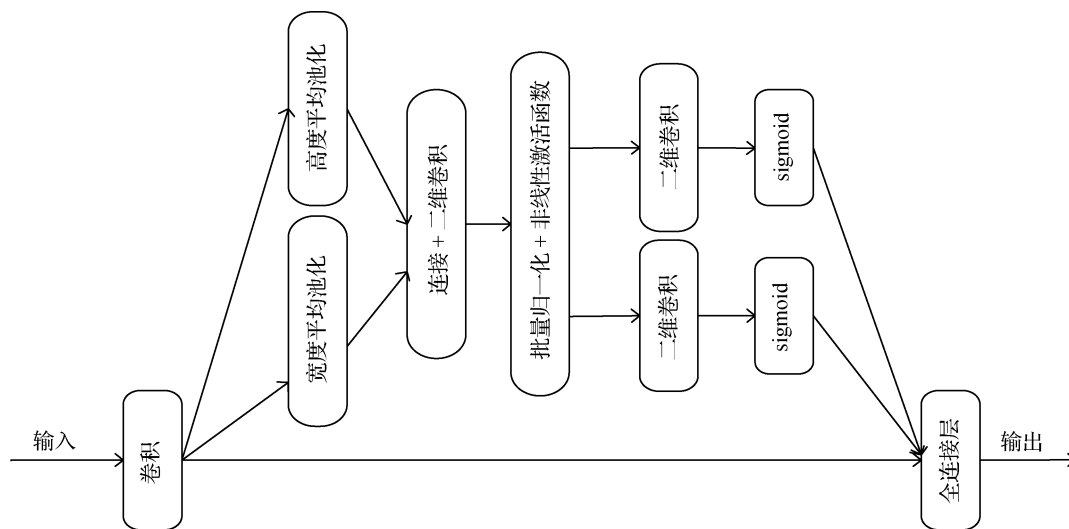


图4 CA注意力模块结构图

Fig. 4 Structure of CA attention module

关键的空间位置之间的远程依赖关系。CA注意力模块的目标是增强移动网络的特征表达能力,它可以对网络中的任何中间特征张量进行变换,而输出的张量尺寸保持不变。

CA注意力模块以一种独特的方式获取图像宽度和高度上的关注,并对精确的位置信息进行编码。首先将输入特征图分成两个方向:宽度和高度,并对它们分别进行全局平均池化。这样,就会得到在宽度和高度方向上的两个特征图,相应的计算为

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

式中, $z_c^h(h)$, $z_c^w(w)$ 表示在第 C 通道、高度 h 、宽度 w 处的输出, W 、 H 分别表示在水平、垂直方向上的窗口大小, $x_c(h, i)$ 表示输入张量在高度 h 、水平索引 i 处的元素值, $x_c(j, w)$ 表示输入张量在宽度 w 、垂直索引 j 上的元素值。

将获得的两个方向上的特征图拼接在一起,随后将它们送入共享的卷积核为 1×1 的卷积模块,将其维度降低为原来的 Clr ,接着将经过批量归一化处理的特征图 F 送入 sigmoid 激活函数中得到形如 $1 \times (W + H) \times Clr$ 的特征图 f , 具体为

$$f = \delta(F, ([z^h, z^w])) \quad (3)$$

式中, $[\cdot]$ 表示沿空间维度的级联操作, $\delta(\cdot)$ 是非线性激活函数。

接下来,将特征图 f 按照原来的高度和宽度送

入 1×1 的卷积。分别得到通道数与原来相同的特征图 F_h 和 F_w 。然后,将这两个新特征图分别送入 sigmoid 激活函数,以获得特征图在高度上的注意力权重 g^h 和在宽度方向的注意力权重 g^w , 具体为

$$g^h = \sigma(F_h(f^h)) \quad (4)$$

$$g^w = \sigma(F_w(f^w)) \quad (5)$$

最后,在原始特征图上通过乘法加权计算,得到最终在宽度和高度方向上带有注意力权重的特征图,具体为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

该模块在 MF-CA Net 模型和 MFEB 的结构中均有使用,目的是可以更好地捕捉和传递多尺度的牙齿特征。

2.4 联合损失函数

考虑到人类视觉系统对边缘信息的感知特性,网络利用结构相似性构造边界损失函数,以增强模型降低边界差异的能力。设计的联合损失函数从像素、局部和整体3个方向优化牙齿边缘的分割效果。使用 Dice 损失函数、二元交叉熵损失函数和 SSIM (structural similarity) 损失函数的联合来作为最终损失函数。Dice 损失是一种整体水平的损失函数,将真实分割结果与预测分割结果看做两个集合 A 和 B , 从而计算两个集合的相似度。具体为

$$L_{\text{Dice}} = 1 - \frac{2|A \cap B|}{|A| + |B|} \quad (7)$$

交叉熵损失是一种像素级别的损失,依次计算预测的分割结果和真实分割结果的相应像素位置。

具体为

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log y_i^* + (1 - y_i) \log (1 - y_i^*)] \quad (8)$$

式中, N 表示像素总数, y_i 表示像素 i 的预测, y_i^* 表示像素 i 的真实标签。

人类视觉系统可以很容易地从场景中提取结构特征, 而结构相似性 (structural similarity, SSIM) 往往意味着目标边缘的信息。SSIM (Wang 等, 2003) 最初是为图像质量评估而提出的, 它能在了一幅图像中捕捉到结构信息。因此, 将其作为损失函数的一部分, 以学习标记图像的结构信息, 计算方法为

$$L_{\text{SSIM}} = 1 - \frac{(2u_x u_y + C_1)(2\sigma_{xy} + C_2)}{(u_x^2 + u_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (9)$$

式中, u_x 和 u_y 分别是 X 和 Y 的均值, σ_x^2 和 σ_y^2 分别是 X 和 Y 的方差, σ_{xy} 表示 X 和 Y 的协方差 $C_1 = 0.001$, $C_2 = 0.001$, 这是为了避免分母为 0。

SSIM 损失是一个局部级别的损失函数, 与像素级别和整体级别的损失函数不同, 它能从不同的尺度 (分辨率) 看待像素的邻域差异, 可以让网络更注重图像的结构和边缘等信息, 从而得到更优的分割结果。

鉴于上述 3 种损失各自的不同关注点, 本文利用它们来联合优化网络性能, 可以综合考虑多个方面的性能指标, 从而更好地优化分割模型。计算方式为

$$L = \alpha_1 L_{\text{Dice}} + \alpha_2 L_{\text{BCE}} + \alpha_3 L_{\text{SSIM}} \quad (10)$$

通过多次实验结果对比, 将参数设置为 $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, $\alpha_3 = 0.2$ 。这样的参数设置具有最好的分割效果。

3 实验结果及分析

3.1 实验环境设置

实验在 Ubuntu64 位操作系统的 DGX Station 上运行, 具体硬件信息如表 1 所示。提出的 MF-CA Net 网络是基于 Python3 和 Pytorch 库进行构建的。模型更新部分使用初始学习率为 1×10^{-4} 的 Adam 优化器来训练网络。网络的训练数据批大小设置为 2, 训练迭代次数为 50 epoch。

3.2 数据集及评价指标

在医院口腔科收集的 20 例 CBCT 扫描中, 广泛评估了所提出的框架。首先在 CBCT 图像上手动裁

表 1 配置环境

环境配置	参数配置
操作系统	Ubuntu
内存	256 GB
显卡	Tesla V100
CPU	-Xeon-E5-2698 v4
硬盘容量	4 TB

剪牙齿区域, 将其尺寸调整为 960×720 像素。牙科医生对 CBCT 图像数据中的牙齿部分进行逐帧手动标记, 从而获得分割标签。实验使用训练数据 5 000 幅, 测试集 1 000 幅。

图像分割的评价指标能直接表明分割模型的性能。为了更准确地评估提出模型在牙齿分割任务上的性能, 本文使用了 Dice、平均交并比 (mean intersection-over-union, mIoU)、精度 (accuracy)、召回率 (recall)、准确性 (precision)、F2 分数作为评价指标。

Dice 相似系数计算的是真实标签和预测结果之间的相似度。若医学影像的真实标签用 A 表示, 分割模型的预测结果用 B 表示, 则 Dice 相似系数的表达式为

$$f_{\text{Dice}} = \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN} \quad (11)$$

式中, TP 为真正例, 表示在真实标签中为真, 在预测结果中也为真。 FP 为假正例, 表示在真实标签中为假, 在预测结果中为真。 FN 为假负例, 表示在真实标签中为假, 在预测结果中为假。 TN 为真负例, 表示在真实标签中为真, 在预测结果中为假。 Dice 相似系数的取值范围为 $[0, 1]$, 分割结果越接近真实标签, 则 Dice 相似系数的值越接近 1。

在分割任务中, 交并比计算真实标注与分割结果之间的交并比, 其表达式为

$$f_{\text{IoU}} = \frac{TP}{TP + FP + FN} \quad (12)$$

平均交并比是基于交并比的一种分割度量标准, 计算每个类别计算出的交并比求和之后取平均值, 其表达式为

$$f_{\text{mIoU}} = \frac{1}{k} \sum_{i=1}^k \frac{TP}{TP + FP + FN} \quad (13)$$

式中, k 为类别数, 对于医学影像中的二值分割来

说,只有属于分割标注和不属于分割标注两个区域,所以类别数 $k=2$ 。

F2分数是准确率和召回率的调和数,其表达式为

$$F_2 = \frac{5TP}{5TP + FP + 4FN} \quad (14)$$

3.3 对比实验

对提出的方法和几种主流的基于深度学习的图像分割方法进行了一系列的对比实验,并对实验结果做了详细的数据分析。对比组的模型分别是U-Net(Ronneberger等,2015)、Res-UNet(Zhang等,2018)、UNet++(Zhou等,2018)、DenseASPP、DeeplabV3+(Chen等,2018)和PyConvU-Net(Li等,

2021),这6个网络模型在医学图像分割领域都有出色表现。为了公平地比较这些模型的分割性能,所有模型都使用ResNet50作为主干特征提取网络,并在同一个测试数据集上统计了性能指标,其中定量结果如表2所示,从表中可以看出,与其他医学图像分割方法相比,本文方法取得了领先的分割性能,在Dice评价指标上达到了0.9495的高分数,对比PyConvU-Net和DeeplabV3+提高了约4%,对比U-Net提高了约16%。各评价指标都是在多次实验后取平均值,虽然本文模型在accuracy指标上与DeeplabV3+相比略低,但其余5个指标上都优于DeeplabV3+,因此依然可以证明所提方法的有效性。

表2 MF-CA Net与现有分割方法的性能比较

Table 2 Performance comparison between MF-CA Net and existing segmentation methods

模型	Dice	mIoU	precision	recall	accuracy	F2
U-Net	0.792 8	0.791 4	0.873 2	0.882 6	0.911 2	0.901 4
Res-UNet	0.832 5	0.846 5	0.897 4	0.925 0	0.982 5	0.884 6
UNet++	0.813 7	0.803 9	0.866 9	0.907 2	0.979 7	0.905 7
DenseASPP	0.844 4	0.839 2	0.902 8	0.931 5	0.979 9	0.921 1
DeeplabV3+	0.910 3	0.863 4	0.895 0	0.927 7	0.990 7	0.953 3
PyConvU-Net	0.908 7	0.869 0	0.907 2	0.914 9	0.984 9	0.941 9
MF-CA Net(本文)	0.949 5	0.899 7	0.942 1	0.968 7	0.990 4	0.954 3

注:加粗字体表示各列最优结果。

图5展示了本文方法和对比方法的一些典型分割结果示例。从图中可以看出,本文方法与其他医学分割方法相比更接近真实标签图像的分割结果,尤其是牙根部分的分割结果形状准确,边缘结构清晰。观察结果发现,只有本文方法可以将智齿准确地分割出来,其他算法对智齿的分割效果很差,甚至未能分割出智齿。其中DeeplabV3+方法在分割时容易在牙齿内部出现空洞和单独的黑点,推测可能是其为了提高分割速度而引入的深度可分离卷积,导致网络不能有效地利用不同通道在相同空间位置上的特征信息。在拓扑结构变化较大的牙根区域,其他对比算法出现了明显的“欠分割”现象,即网络错误地预测背景像素和牙齿像素的类别,导致牙根不能被正确分割。相比之下,本文方法在分割过程中轻松缓解了“欠分割”问题,表现出很高的鲁棒性。

3.4 消融实验

为了验证提出方法的有效性,设计了消融实验,并将消融实验分为两个部分:网络模块消融实验和损失函数消融实验。网络模块消融实验讨论了MFEB模块与CA注意力机制的有效性,其中“MFEB{1}”表示多尺度特征提取模块只有卷积核大小为1,并且扩张率也为1的分支,“MFEB{4}”表示多尺度特征提取模块有4个分支,其中卷积扩张率分别为1,3,7,11。“ L_{BCE} ”表示单独使用二元交叉熵损失,“ L_{BD} ”表示同时使用二元交叉熵损失和Dice损失作为网络的最终损失函数,“ L_{BDS} ”表示联合使用BCE损失、Dice损失和结构相似性损失作为最终损失函数,即上文提出的联合损失函数。此外,还将目前深度学习中表现优异的CBAM(convolutional block attention module)注意力机制加入到网络中与使用CA注意力机制进行了对比,探讨它们各自对网络性能的提升

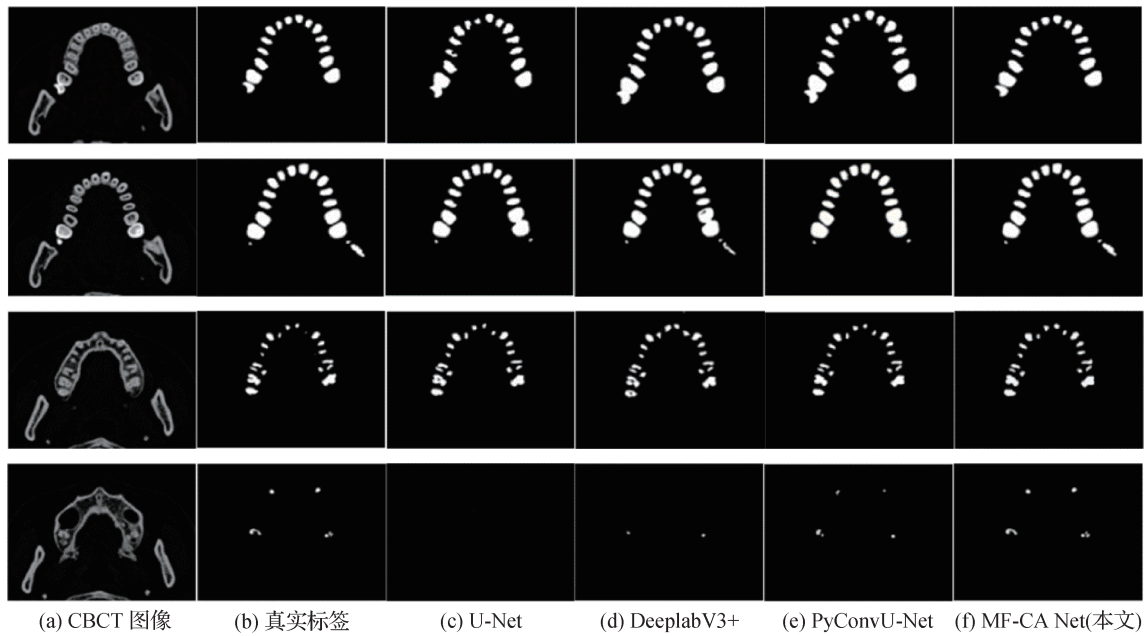


图5 不同方法在CBCT上的分割结果

Fig. 5 Segmentation results of different methods on CBCT

((a)CBCT images;(b)ground truth;(c)U-Net;(d)DeeplabV3+;(e)PyConvU-Net;(f)MF-CA Net)

效果。损失函数消融实验中分别讨论了单独使用与联合使用各个损失函数对网络分割结果的影响,表3和表4为消融研究的结果。

从表3可以看出,本文方法在各项评价指标上均优于其他实验模块设置的相应指标,其中MFEB的加入使Dice得分从0.7393上升到0.9011。图6中对分割结果示例进行了展示,可以看出,

MFEB和CA注意力模块的加入,对网络的分割性能产生了巨大的提升,牙齿边缘分割结果准确而清晰。

从表4可以明显看出,提出的联合损失函数效果优于单独使用各个损失函数。相比单独使用二元交叉熵损失,联合损失函数使评价指标Dice得分提高了8.31%;结构相似性损失的加入使得网络的各

表3 网络模块消融实验结果

Table 3 Experimental results of network module ablation

模型	Dice	mIoU	precision	recall	accuracy	F2
MFEB{1}+ L_{BDS}	0.739 3	0.642 4	0.702 3	0.890 1	0.928 9	0.864 3
MFEB{4}+ L_{BDS}	0.901 1	0.807 6	0.816 5	0.921 3	0.959 9	0.884 8
MFEB{4}+CBAM+ L_{BDS}	0.907 6	0.809 9	0.837 1	0.944 3	0.972 1	0.903 9
MFEB{4}+CA+ L_{BDS} (本文)	0.949 5	0.899 7	0.942 1	0.968 7	0.990 4	0.954 3

注:加粗字体表示各列最优结果。

表4 损失函数消融实验结果

Table 4 Experimental results of loss function ablation

模型	Dice	mIoU	precision	recall	accuracy	F2
MFEB{4}+CA+ L_{BCE}	0.866 4	0.795 8	0.803 2	0.901 7	0.933 1	0.887 5
MFEB{4}+CA+ L_{BD}	0.901 3	0.830 9	0.881 6	0.922 3	0.988 3	0.903 6
MFEB{4}+CA+ L_{BDS} (本文)	0.949 5	0.899 7	0.942 1	0.968 7	0.990 4	0.954 3

注:加粗字体表示各列最优结果。

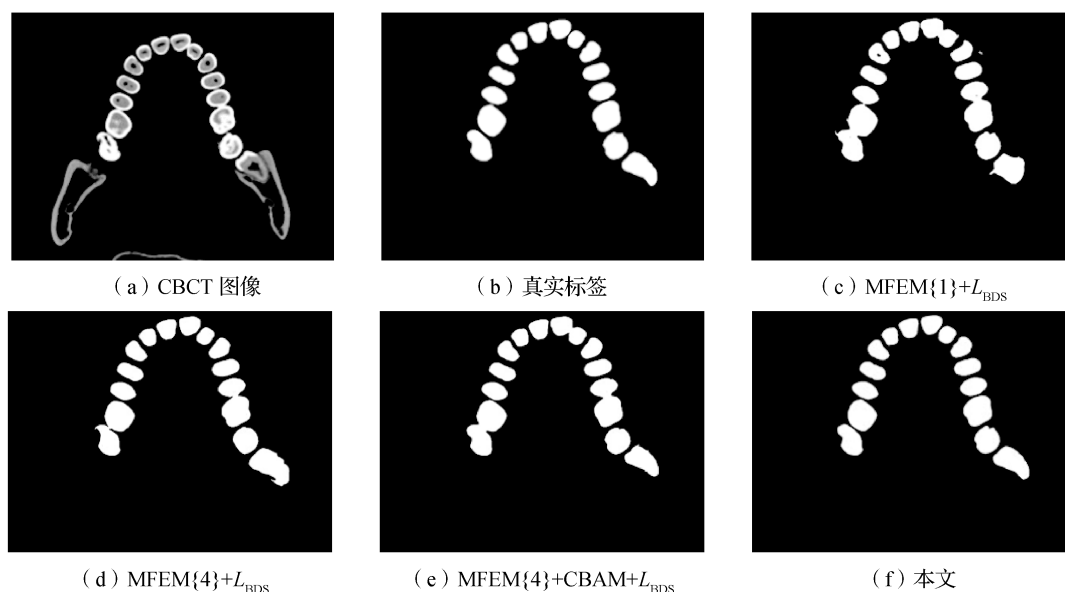


图6 不同模块设置下的网络模型的分割结果

Fig. 6 Segmentation results of the network model with different module settings

((a)CBCT image;(b)ground truth;(c)MFEM{1}+ L_{BDS} ;(d)MFEM{4}+ L_{BDS} ;(e)MFEM{4}+CBAM+ L_{BDS} ;(f)ours)

项性能指标均有所提高,其中 Dice 得分升高了近5%。

4 结论

本文提出了一种名为 MF-CA Net 的牙齿分割网络模型,该模型采用多尺度特征提取模块、CA 注意力机制和联合损失函数,以提高牙齿分割的鲁棒性和准确性,并且该模型有效解决了由于 CBCT 图像对比度低且牙齿轮廓复杂而导致的牙齿分割困难的问题,实现了对牙齿的精确分割,为医生的诊断和治疗计划提供了有力的支持。

通过多次实验发现,MFEB 和 CA 注意力机制的添加可以有效地提取口腔 CBCT 图像中的多尺度牙齿特征,对分割效果的提升起到了重要的作用。使用 ResNet50 作为模型的编码器,可以预训练权重,帮助模型更快地收敛到牙齿分割任务,实现高精度的牙齿分割结果。并且即使是扩大数据集,也可以保持高质量的分割性能。

最后本文在数据集上进行了大量的牙齿分割实验,最终在多个指标上都实现了超出目前多数主流分割网络的效果,表明了所提方法的有效性。

但是本文没有考虑牙齿之间的相互关系和约束,例如牙齿的位置、角度等,这可能会导致分割结

果不自然或不准确。在后续研究中,将引入牙齿之间的相互关系和约束,以提高分割结果的自然性和准确性,例如使用空间关系或图模型来约束牙齿的位置和角度。此外,提出的方法是基于全监督的图像分割,然而标注数据集十分烦琐且困难,并且标注结果十分依赖于标注者的经验。相对而言,弱监督学习不需要完整的训练标签就能够获得较为准确的分割结果,因此后续研究将尝试弱监督方法应用到牙齿分割工作中。使模型可以面对更加复杂的情况,达到更好的分割效果。

参考文献(References)

- Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]
- Chung M, Lee M, Hong J, Park S, Lee J, Lee J, Yang I H, Lee J and Shin Y G. 2020. Pose-aware instance segmentation framework from cone beam CT images for tooth segmentation. *Computers in Biology and Medicine*, 120: #103720 [DOI: 10.1016/j.combiomed.2020.103720]
- Cui Z M, Li C J and Wang W P. 2019. ToothNet: automatic tooth instance segmentation and identification from cone beam CT images//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, USA:

- IEEE: 6361-6370 [DOI: 10.1109/CVPR.2019.00653]
- Gan Y Z, Xia Z Y, Xiong J, Li G L and Zhao Q F. 2018. Tooth and alveolar bone segmentation from dental computed tomography images. *IEEE Journal of Biomedical and Health Informatics*, 22(1): 196-204 [DOI: 10.1109/JBHI.2017.2709406]
- Gan Y Z, Xia Z Y, Xiong J, Zhao Q F, Hu Y and Zhang J W. 2015. Toward accurate tooth segmentation from computed tomography images using a hybrid level set model. *Medical Physics*, 42(1): 14-27 [DOI: 10.1118/1.4901521]
- Gao H and Chae O. 2010. Individual tooth segmentation from CT images using level set method with shape and intensity prior. *Pattern Recognition*, 43(7): 2406-2417 [DOI: 10.1016/j.patcog.2010.01.010]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Ji D X, Ong S H and Foong K W C. 2014. A level-set based approach for anterior teeth segmentation in cone beam computed tomography images. *Computers in Biology and Medicine*, 50: 116-128 [DOI: 10.1016/j.combiomed.2014.04.006]
- Koch T L, Perslev M, Igel C and Brandt S S. 2019. Accurate segmentation of dental panoramic radiographs with U-Nets//*2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. Venice, Italy: IEEE: 15-19 [DOI: 10.1109/ISBI.2019.8759563]
- Li C Y, Fan Y X and Cai X D. 2021. PyConvU-Net: a lightweight and multiscale network for biomedical image segmentation. *BMC Bioinformatics*, 22(1): #14 [DOI: 10.1186/s12859-020-03943-2]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Wang Z, Simoncelli E P and Bovik A C. 2003. Multiscale structural similarity for image quality assessment//*Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, USA: IEEE: 1398-1402 [DOI: 10.1109/ACSSC.2003.1292216]
- Wirtz A, Mirashi S G and Wesarg S. 2018. Automatic teeth segmentation in panoramic X-ray images using a coupled shape model in combination with a neural network//*Proceedings of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention*. Granada, Spain: Springer: 712-719 [DOI: 10.1007/978-3-030-00937-3_81]
- Xia Z Y, Gan Y Z, Chang L C, Xiong J and Zhao Q F. 2017. Individual tooth segmentation from CT images scanned with contacts of maxillary and mandible teeth. *Computer Methods and Programs in Biomedicine*, 138: 1-12 [DOI: 10.1016/j.cmpb.2016.10.002]
- Xu L J, Zhang J Y, Huang Z H and Wang X Z. 2022. Successful individualized endodontic treatment of severely curved root canals in a mandibular second molar: a case report. *World Journal of Clinical Cases*, 10(14): 4632-4639 [DOI: 10.12998/wjcc.v10.i14.4632]
- Zhang W J. 2011. Algorithms of 3D segmentation and reconstruction based on teeth CBCT images//*Communication Systems and Information Technology: Selected Papers from the 2011 International Conference on Electric and Electronics*. Nanchang, China: Springer: 713-720 [DOI: 10.1007/978-3-642-21762-3_93]
- Zhang Z X, Liu Q J and Wang Y H. 2018. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5): 749-753 [DOI: 10.1109/LGRS.2018.2802944]
- Zhao Y, Li P C, Gao C Q, Liu Y, Chen Q Y, Yang F and Meng D Y. 2020. TSASNet: tooth segmentation on dental panoramic X-ray images by two-stage attention segmentation network. *Knowledge-Based Systems*, 206: #106338 [DOI: 10.1016/j.knsys.2020.106338]
- Zhou T, Dong Y L, Huo B Q, Liu S and Ma Z J. 2021. U-Net and its applications in medical image segmentation: a review. *Journal of Image and Graphics*, 26(9): 2058-2077 (周涛, 董雅丽, 霍兵强, 刘珊, 马宗军. 2021. U-Net网络医学图像分割应用综述. *中国图象图形学报*, 26(9): 2058-2077) [DOI: 10.11834/jig.200704]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//*4th International Workshop, DLMIA 2018, and 8th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]
- Zhu Y and Li X. 2023. A survey of medical image captioning technique: encoding, decoding and latest advance. *Journal of Image and Graphics*, 28(7): 1990-2010 (朱翌, 李秀. 2023. 医学图像描述综述: 编码、解码及最新进展. *中国图象图形学报*, 28(7): 1990-2010) [DOI: 10.11834/jig.211021]

作者简介

秦俊,男,副教授,主要研究方向为医学图像处理、手术引导辅助系统、医疗机器人、复杂环境下视频运动物体识别及追踪。E-mail: qinjun@cust.edu.cn

纪柏,通信作者,女,教授,主要研究方向为肝胆胰外科临床诊疗和肝消融手术临床及后续诊疗。

E-mail: ji_bai@jlu.edu.cn

卢婷岚,女,硕士研究生,主要研究方向为图像处理应用技术。E-mail: 1019115899@qq.com

李雨晴,女,硕士研究生,主要研究方向为图像处理应用技术。E-mail: 2501527622@qq.com