

摘 要

随着近些年来对深度学习和智能医学的研究，数字化口腔诊治得到快速发展。获取完整的 3D 牙齿模型是数字化牙科诊治的关键步骤，而 CBCT（Cone beam CT，锥形束 CT）牙科数据，因其包含牙齿完整的形态学结构信息，逐渐成为牙齿临床诊治和临床实验中常用的数据之一，自动而准确的 CBCT 牙齿实例分割对后续诊疗任务特别是牙齿正畸工作具有重要意义。

近年来众多学者对 CBCT 牙齿分割进行了很多有意义的研究，并取得了很大的进展。传统的牙齿分割方法常常忽略了牙齿丰富的形态学特征，尤其是面对复杂的临床畸形牙齿案例时，多数牙齿分割方法依旧无法实现较好的分割结果。尽管目前深度学习方法可以大大提升牙齿实例分割的准确性，然而由于目前可用于网络训练用的 CBCT 数据非常少，大大影响了深度学习方法在这方面的应用。因此，准确的捕获单颗齿体完整的几何形状以及形态细节仍然是目前牙齿分割研究的挑战。针对上述问题，本文围绕从 CBCT 数据中精确的自动检测和实例分割牙齿进行研究，具体工作如下：

第一，针对 CBCT 数据牙齿复杂的几何结构与牙根等节特征难以捕获，从而导致牙齿分割精度不高的问题，提出了一种基于捕获依赖性和感受野调整的 CBCT 牙齿实例分割方法，新方法构建了一个新的两阶段深度学习框架，TSDNet（Tooth Segmentation Deeplearning Network），以实现牙齿 CBCT 数据自动而准确的实例分割。在第一阶段用以获取准确的牙齿质心，确保牙齿准确的空间定位，同时获取单颗牙齿的实例信息。在网络的第二阶段的编码过程引入了基于 3D self-attention 机制的牙齿几何信息引导模块，提升网络捕获齿体整体结构信息的能力。同时提出的基于多尺度空洞卷积的牙齿特征整合模块用以多尺度捕获牙齿表面和牙根等细节信息，很好的提升了牙齿分割精度。本文使用现实医疗机构获取到的 CBCT 牙齿数据集评估提出的方法，实验证明，TSDNet 具有较高的分割性能和精确的实例分割效果，相比现有先进方法具有一定的优势。

第二，当仅以牙齿质心作为牙齿定位关键点时，网络训练体素指向质心，在面对更复杂几何结构的临床畸形牙齿案例时，牙齿整体特征常常难以捕获，同时 CBCT 训练数据少，影响模型训练效果。针对以上问题，本文提出了一个 CBCT 中形变数据增强和形态学关键特征点引导的牙齿实例分割方法（Tooth Morphology Segmentation Deeplearning Network, TMSDNet）。该框架引入了层次化牙中轴预测网络，以 CBCT

牙齿数据逐层中心生成的关键点为层次化引导，网络训练体素逐层指向牙中轴。本文设计了一个多任务学习机制，输出实例级牙中轴作为辅助任务，最终得到精确的牙齿分割结果。同时，本文引入了三维薄板样条插值（Thin Plate Spline, TPS）对 CBCT 牙齿数据进行扭曲形变，模拟生成更多的复杂临床畸形牙齿案例，同时扩充了网络数据训练量。广泛的评估、消融研究以及与现有方法的比较表明，本文的方法达到了最先进的分割效果，特别是在具有挑战性的复杂临床畸形牙齿案例上也同样展现出了较好的分割性能，证明了本文的方法在临床医学上的具有一定的潜在适用性。

关键字：CBCT，牙齿实例分割，3D self-attention，关键点预测，层次化表示，数据增强。

Abstract

With the research on deep learning and smart medicine in recent years, digital dental consultation has been rapidly developed. Obtaining a complete 3D dental model is a key step in digital dentistry, and CBCT (Cone beam CT) dental data, because it contains the complete morphological structure information of teeth, is gradually becoming one of the common data in dental clinical diagnosis and treatment and clinical experiments, and the automatic and accurate CBCT dental example segmentation is important for the subsequent treatment tasks, especially orthodontic work. The automatic and accurate segmentation of CBCT dental examples is important for the subsequent treatment tasks, especially orthodontics.

Many scholars have conducted many meaningful studies on CBCT tooth segmentation in recent years, and great progress has been made. Traditional tooth segmentation methods often ignore the rich morphological features of teeth, especially when facing complex clinical malocclusion cases, most tooth segmentation methods still fail to achieve better segmentation results. Although current deep learning methods can greatly improve the accuracy of tooth instance segmentation, the application of deep learning methods in this area is greatly hampered by the very small amount of CBCT data currently available for network training. Therefore, accurately capturing the complete geometry as well as morphological details of a single tooth remains a challenge for current tooth segmentation research. To address these issues, this paper focuses on the accurate automatic detection and instance segmentation of teeth from CBCT data, as follows.

First, to address the problem that the complex geometric structures of teeth with nodal features such as roots of CBCT data are difficult to capture, which leads to poor accuracy of tooth segmentation, a method of CBCT tooth instance segmentation based on capture dependency and perceptual field adjustment is proposed, and the new method constructs a new two-stage deep learning framework, TSDNet (Tooth Segmentation Deeplearning Network) to achieve automatic and accurate instance segmentation of dental CBCT data. The first stage is used to obtain the accurate tooth center of mass to ensure accurate spatial localization of the tooth, and to obtain the instance information of a single tooth. In the second stage of the network, the coding process introduces a guidance module for tooth

geometry information based on the 3D self-attention mechanism to enhance the network's ability to capture the overall structure information of the tooth. The proposed multi-scale cavity convolution-based tooth feature integration module is also used to capture tooth surface and root details at multiple scales, which improves the tooth segmentation accuracy. We evaluate the proposed method using CBCT dental datasets obtained from real medical institutions, and experimentally demonstrate that TSDNet has high segmentation performance and accurate instance segmentation, which has certain advantages over existing advanced methods.

Second, when only the center of mass of the tooth is used as the key point for tooth localization, the network training voxels point to the center of mass, and the overall tooth features are often difficult to capture when facing clinical malocclusion cases with more complex geometric structures, while the small amount of CBCT training data affects the model training effect. To address these problems, this paper proposes a tooth instance segmentation method (Tooth Morphology Segmentation Deeplearning Network (TMSDNet)) with morphological data enhancement and morphology key feature points guidance in CBCT. The framework introduces a hierarchical tooth mesiodistal prediction network with hierarchical guidance of key points generated from the center of CBCT tooth data layer by layer, and the network training voxels point to the tooth mesiodistal axis layer by layer. We designed a multi-task learning mechanism to output instance-level tooth mesial axes as a secondary task to finally obtain accurate tooth segmentation results. At the same time, we introduced the 3D Thin Plate Spline (TPS) to distort the CBCT tooth data to simulate the generation of more complex clinical malocclusion cases and expand the training volume of the network data. Extensive evaluation, ablation studies and comparisons with existing methods show that our method achieves state-of-the-art segmentation results, especially in challenging complex clinical malocclusion cases, and we also show good segmentation performance, demonstrating the potential applicability of our method in clinical medicine.

KEYWORDS: CBCT, Tooth instance segmentation, 3D self-attention, key point prediction, hierarchical representation, data augmentation.

目录

摘 要	1
第一章 绪论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	2
1.2.1 牙齿分割方法综述	2
1.2.2 医学图像数据增强方法	3
1.3 本文主要工作	5
1.3.1 主要研究内容	5
1.3.2 本文组织结构	6
第二章 图像分割的卷积神经网络理论基础	7
2.1 卷积神经网络	7
2.1.1 卷积层	7
2.1.2 池化层	8
2.1.3 激活函数	8
2.1.4 全连接层	10
2.1.5 空洞卷积层	10
2.2 用于医学图像分割的深度学习网络框架	11
2.2.1 FCN	11
2.2.2 U-Net	11
2.2.3 V-Net	12
2.2.4 R-CNN	12
2.2.5 Fast-R-CNN	13
2.2.6 Mask-RCNN	13
2.3 常用数据集和评价标准	14
2.3.1 数据集介绍	14
2.3.2 评价指标	15
2.4 本章小结	16
第三章 基于捕获依赖性和感受野调整的 CBCT 牙齿实例分割	17
3.1 引言	17
3.2 本章方法	18
3.2.1 牙齿质心预测	18
3.2.2 基于自注意力机制的牙齿几何结构信息引导模块	20

3.2.3 基于空洞卷积多尺度融合的牙齿特征整合模块.....	21
3.3 模型训练	22
3.3.1 损失函数	22
3.3.2 实验细节	23
3.4 实验结果与分析	23
3.4.1 性能比较	23
3.4.2 消融实验	25
3.4.3 局限性	27
3.5 本章小结	28
第四章 CBCT 中形变数据增强和形态学关键特征点引导的牙齿实例分割	29
4.1 引言	29
4.2 本章方法	30
4.2.1 TPS 数据增强方法	31
4.2.2 层次化牙齿中轴预测网络	32
4.3 模型训练	33
4.3.1 损失函数	33
4.3.2 实验细节	34
4.4 实验结果与分析	34
4.4.1 性能比较	34
4.4.2 消融实验	36
4.5 本章小结	37
第五章 总结与展望	39
5.1 本文总结	39
5.2 未来展望	40
参考文献	42
攻读学位期间取得的学术成果	48
致谢	49

第一章 绪论

1.1 研究背景和意义

数字化技术在个性化医疗领域，如数字化齿科有着广泛的应用。自动而准确的获取 3D 牙齿分割结果往往是数字化牙齿诊疗如数字正畸中重要的、也是基础的步骤之一，后续诊疗步骤一般都是基于独立的单颗牙齿。在现有的两种获取 3D 牙齿数据的方法中，口腔扫描数据^[1]仅能提供口腔内包含牙齿部分的表面信息，而不能提供牙齿的牙根等齿体结构，因此无法用它生成具有完整牙齿几何形状结构的 3D 牙齿模型；CBCT 数据^[2]则利用三维锥形束 X 射线扫描以获取包含完整牙齿齿体，牙床骨在内的全面口腔信息，其数据格式标准，非常适用于做牙齿数据的 3D 分割。因此，本文选取口腔 CBCT 数据进行牙齿的实例分割和识别工作的研究。

目前牙齿分割已经有了很多研究的进展。许多传统方法^[3-6]，在解决 CBCT 牙齿分割问题时，一般都需要提前对 CBCT 数据进行部分预处理操作，如通过人机交互在 CBCT 牙齿图形上选取种子点等。因此传统方法存在着交互性高，效率低等缺陷，非常依赖数据处理人员的操作经验，否则会直接影响到牙齿分割的结果，缺乏鲁棒性。随着深度学习方法引入到 CBCT 牙齿分割工作，实现了对牙齿的自动分割，使牙齿分割效率得到了很大的提升，如基于预训练 AlexNet 网络分割牙齿的方法^[7]；获取牙齿边缘信息并基于 Mask-RCNN 网络分割的 ToothNet 方法^[8]等。然而针对 CBCT 图像的实例分割仍然存在如下问题^[9,10]：（1）CBCT 数据中相邻牙之间存在粘连问题，很难对相邻牙齿进行区分识别进而实现准确分割；（2）牙齿齿体尤其是牙根部分形状复杂，不同牙齿的牙根有一到三个不等，且与周围牙槽骨对比度低，牙齿根部等细节很难捕获；（3）CBCT 图像中同时包含了背景软组织，容易对牙齿的分割形成噪声干扰。针对如上问题，即使一些最先进的深度学习方法，仍然无法较好的捕获和还原出复杂多样的几何齿体结构和牙根细节等特征，致使分割精度依旧不高；（4）CBCT 牙齿数据标注工作繁重，常常需要耗费大量的人力物力，训练数据少也是影响牙齿分割精度的关键问题。这些难题驱使我们设计一个新的网络框架，从 CBCT 中自动分割出单个牙齿并可以精确的还原出牙齿几何结构与细节信息。

1.2 国内外研究现状

目前已经提出的基于 CBCT 的牙齿实例分割方法通常可以分为依赖于对数据进行人机交互预处理等手工调节的传统方法和基于深度学习的方法。特别的,由于深度学习对数据深度依赖,而医学图像的获取和标注相对较为困难,因此,医学图像的数据增强是提高深度学习方法的手段之一。本节我们将对上述方法进行分类与综述。

1.2.1 牙齿分割方法综述

(1) 依赖于人机交互的传统方法

依赖于人机交互的传统方法大致可以分为三类:基于图割的方法、基于先验引导的模型拟合方法和水平集方法。Evain 等人^[3]从牙科 CBCT 图像中分割单个牙齿时用了使用图形切割方法,其工作在数值标准上取得了较高的结果。但是这种图割的方法非常容易受到数据中前景目标定义的影响,在迭代计算的过程中,前景的变化对最终的分割结果影响很大。此外,该方法需要人工处理来构建统计形状模型,不能对牙齿实现自动的分割;基于先验引导的模型拟合方法,如 Wang 等^[4]首先通过对 CBCT 进行人机交互处理获取上下颌的初始分割概率图作为 CBCT 图像分割的先验指导,然后使用随机森林自动分割牙科 CBCT 图像,实现了牙齿的自动分割,但其需要手工获取牙齿分割的先验知识,费时费力且方法鲁棒性不足;基于水平集的方法^{[5][6][7][11]}可以较好的处理 CBCT 数据中复杂的牙齿形状变化。如 Gan^[6]利用两个步骤的水平集方法实现了 CBCT 数据中牙齿和牙槽骨的分割。其首先通过全局水平集提取高密度骨组织信息,然后利用局部水平集方法从牙槽骨中对单个牙齿进行分割。尽管水平集方法取得了不错的分割效果,但其方法对于相邻牙齿或牙槽骨之间的咬合与粘连问题依旧没有很好的解决;随着进一步对传统方法的研究, Jiang 等人^[12]通过手动拾取每相邻 2 颗牙齿间的一个特征点,利用每颗牙齿两侧的特征点计算模型局部区域的标量场与等值线,最后利用一种启发式的特征线提取算法对分割线进行优化得到精确的龈缘线,实现牙齿的精确分割的半自动分割方法,其方法虽然一定程度上提升了牙齿的分割效率,但其分割鲁棒性与自动化程度依旧较低。

(2) 基于深度学习的牙齿分割方法

随着深度学习在医学图像的应用与发展^[13-18],许多研究人员将深度学习方法应用到了 CBCT 牙齿图像分割处理中。Miki 等人^[7]使用预训练的 AlexNet 在 CBCT 图像

的 2D 切片上实现了牙齿类型分类;Cui.等人^[8]在 3D 区域提议引导的 Mask-R-CNN 网络上从 CBCT 图像中实现牙齿实例分割和识别,但其方法只关注排除智齿的牙齿数据集,当应用于错综复杂的牙齿数据时,ToothNet 中基于边界检测器可能会导致牙齿的错误分割;Ezho 等人^[19]提出通过由粗到细的框架在 3D CBCT 图像中分割单个牙齿,他们在小样本数据集上的训练结果,同样不具备鲁棒性。Wang^[20]等人提出了一种基于对称式全卷积残差网络分割算法,其通过增加网络层数,利用跳跃链接结构促进多层次特征融合的方法对牙齿图像进行提取与分割,取得了不错的效果。但依旧没有很好的解决相邻牙齿的粘连问题;Cui 等人^[21]提出一个牙骨架引导的多任务学习框架,通过多任务学习牙齿的边界和牙根,引导框架学习牙齿细节信息,取得了较好的结果。但其依靠牙齿的骨架几何结构引导,面对复杂的牙齿几何结构情况时,该方法也不足以保证分割结果鲁棒性。其他许多性能较好的深度学习框架已经广泛应用于医学 CBCT 图像处理工作。如 VNet^[22]是 UNet^[23]的变体,主要用于处理三维医学数据。其借用了 UNet 从压缩路径叠加特征图,从而补充信息的方法,与 UNet 相比,它在每个卷积阶段都使用了残差架构,这样可以更有效地利用特征图中的信息,在处理三维医学数据分割领域有着良好的性能,但是传统的 VNet 模型也很难应对 CBCT 数据的复杂性,不能很好的处理牙根部分与牙槽骨的粘连性。目前已经涌现出的很多 CBCT 牙齿分割方法,虽然性能有了很大提升,但依然存在着牙齿几何特征丢失,分割精度不高等问题。

随着许多经典模块的提出大大改进了网络对医学图像的处理性能,如, self-attention^[24]模型可以通过建立局部或数据层内所有像素之间的关系来捕获像素之间的相关性,对于医学图像中不同器官语义信息的识别分割等展示出很好的改进,如 Li^[25]等人通过在卷积网络中使用 self-attention 机制,获取到了 CT 图像中不同器官间的像素依赖关系,利用像素相关性达到了去除 CT 图像中噪声的目的,获得了更清晰的器官图像;空洞卷积^[26]在一些图像分割工作中同样发挥了重要作用,不同的空洞率大小的卷积作用于图像时,网络感受野大小不同,对图像信息的捕获也不同,如在 CBCT 数据中,不同尺度的空洞卷积可以捕获从牙齿表面到牙根部位等不同的牙齿细节,多种空洞卷积融合可以实现对图像信息的多尺度捕获,以获取更多的图像细节。

1.2.2 医学图像数据增强方法

为了克服医学图像处理中的数据不足问题,研究人员提出了许多方法,大致可分

为以下几类。一类是基于改进深度学习网络结构的方法。Alom 等人提出了一个用于医学图像分割的 R2U-Net^[27], 该网络结合利用了 U-Net、Residual Network, 以及 RCN。Wang 等人提出了 ChestNet^[28], 将注意力分支纳入深度卷积网络中。注意力分支允许网络自适应地集中在病理异常区域, 这样网络就能从小范围的病理区域内学习更多的信息, 从而使网络可以从一个小数据集集中学习更多的信息。Wang 等人提出了一个深度交互式框架 DeepIGeoS^[29], 用于医学图像的分析。该方法可以通过 3D P-Net 与 CRF-Net 自动分割图像, 然后通过 CRF-Net 的 3D R-Net 交互式地完善结果。然而, 调整深度神经网络是一项艰巨的任务, 而且结果具有非常严重的不确定性。此外, 许多文章采用了基于图像补丁的技术^[30-33]。通过图像基于补丁的表示方法, 图像分割问题可以作为一个图像块的分类问题来解决, 但这种方法往往会丢失一些图像的全局信息。

第二类是通过传统的图像变换增加训练数据量。最简单和最常见的方法是刚性变换, 如旋转、翻转、缩放、剪切和裁剪^[34-36]。因为传统的变换是很容易实现的, 因此它在自然和医学图像处理中都有广泛的应用。然而, 这些变换并没有考虑到数据内器官的形变。而且, 翻转和剪切可能会改变目标结构的形状, 但人体器官的形状和方向是相对固定的, 所以增强后的图像的分布可能会远远偏离医学图像的真实分布, 因此, 传统图像变换的效果也是有限的。

第三类是利用生成对抗网 (GAN)^[37]来学习样本的内在关系, 并生成一些新的样本。近年来, 已经有一些使用 GAN 或其变体的相关研究来解决医学数据不足的问题。Calimeri 等人利用 GAN 进行脑部 MR 图像的增强^[38]。Bermudez 等人使用 DCGAN 来实现二维脑图像合成^[39]。GAN 在大多数自然图像的研究上取得了良好的效果^[40-44], 但是据我们所知, 这些网络仍然有一些如不稳定的训练等缺点。因此, GAN 通常被用于生成二维的医学图像, 而且使用比例很小。

第四类是利用基于模型的数据增强算法。它可以利用数据的统计信息, 建立一个统一的模型。如先前对 MNIST 进行的增强的工作^[45]。Wilms 等人使用统计学形状模型对手部图像进行建模^[46]。该算法可以生成大量不同姿势和大小的二维手部图像。Uzunova 等人采用统计学外观模型来建立了一个 MRI 心脏图像模型^[47], 并创建了一些新的二维心脏图像。结果表明该方法可以提高卷积网络的分割精度。然而, 目前基于模型的增强算法大多适用于二维图像, 而且需要标记大量的特征。由于许多人体器官内缺乏特征点, 如前列腺和心脏等, 传统的基于模型的数据增强算法通常难以在三

维医学图像上使用。

1.3 本文主要工作

1.3.1 主要研究内容

CBCT 图像牙齿分割是计算机视觉和数字化正畸领域的热点问题，在牙科临床诊断中具有极大的研究意义和商业价值。但如前所述，已经涌现出的很多 CBCT 牙齿分割方法，虽然性能有了很大提升，但依然存在着牙齿几何特征丢失，分割精度不高等问题。因此，新的研究希望增强网络对牙齿几何结构语义信息的捕获，以及对牙齿细节特征的保持能力。具体研究内容如下：

本文主要针对 CBCT 图像的牙齿自动检测与实例分割问题展开研究，提出了两种解决方法，一是基于捕获依赖性和感受野调整的 CBCT 牙齿实例分割方法，二是基于 CBCT 形变数据增强和形态学关键特征点引导的牙齿实例分割方法。特别的，为解决数据缺少问题，在第二种方法中还提出基于 TPS 数据形变的 CBCT 数据增强方法，在扩充训练数据的同时尽可能多的生成复杂临床畸形病例牙齿样本，提升网络对复杂牙齿样本的特征捕获能力。

本文的主要贡献总结如下：

1. 针对传统方法效率低下，泛化能力差，鲁棒性差等问题，提出了一个从牙齿质心定位到牙齿实例分割的两阶段深度学习框架。第一阶段利用基于 VNet 的质心预测网络获取每个形状复杂牙齿的牙齿质心，实现对单个牙齿的准确空间定位的同时获取单个牙齿的实例级牙齿信息；第二阶段针对 CBCT 图像中相似噪声繁多，对比度低导致牙齿几何结构难以捕捉的问题，提出了一种基于 3D 自注意力机制的牙齿几何结构信息引导模块，用于同时获取 CBCT 图像中切片内依赖关系和切片间依赖关系，提升网络对 CBCT 牙齿数据中不同组织像素点的分类与获取能力，使得构造的网络对牙齿完整几何结构拥有了更好的捕获能力；同时针对牙齿分割过程中牙根与牙表面等细节难以保持的问题，提出了一个基于空洞卷积多尺度融合的牙齿特征整合模块，以 3 个不同空洞率的空洞卷积并行捕获牙齿数据的上下文语义信息，多尺度的捕获牙齿从牙表面到牙根等细节信息。进而通过融合不同空洞率的空洞卷积层，使网络在保持分辨率不变的同时以更大感受野对牙齿细节信息进行多尺度捕获与还原，引导网络的逐级解码过程。

2. 针对由于复杂牙齿案例形态学特征捕获困难导致分割精度不高的问题, 提出一种 CBCT 形变数据增强和形态学关键特征点引导的牙齿实例分割方法。首先根据 3DTPS 算法对原始 CBCT 数据进行形变, 增加训练数据量的同时模拟生成更多的临床畸形齿案例; 为了提升网络对更复杂牙齿的特征捕获能力, 本文对 CBCT 牙齿数据进行逐层形态学关键特征点检测, 通过预测牙齿实例级牙中轴, 提升网络对更多复杂状态牙齿的特征捕获能力, 通过多任务训练网络得到优秀的牙齿分割结果。

1.3.2 本文组织结构

第一章为绪论部分, 主要介绍了牙齿自动检测与分割的研究背景及意义、国内外的研究现状以及本文的主要工作。

第二章主要介绍牙齿分割相关研究的理论基础。简要介绍了深度卷积神经网络的基本结构和用于分割的常用网络。最后, 介绍了几种常用的牙齿分割评估方法和本文所用数据集。

第三章介绍了提出的 CBCT 中基于捕获依赖性和感受野调整的牙齿实例分割方法。首先介绍了提出的两阶段网络中的质心定位网络, 详细介绍了牙齿质心预测方法。随后介绍了实例分割阶段的基于 3D 自注意力机制的牙齿几何结构信息引导模块和基于空洞卷积多尺度融合的牙齿特征整合模块。通过在本文数据集上与最新方法进行定性和定量实验及消融实验来证明该方法的有效性。

第四章介绍了提出的 CBCT 中形变数据增强和形态学关键特征点引导的牙齿实例分割方法。首先介绍了基于 3D TPS 的数据增强方法, 增加训练数据量的同时模拟生成更多的复杂临床畸形齿案例。其次介绍了层次化牙中轴预测方法和多任务训练的牙齿实例分割网络。最后给出实验结果。

第五章是本文的总结与展望。对本文提出的算法进行了总结, 分析方法的优点和不足之处, 并探讨未来可以改进的方向。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/687162004060006134>