M.S. THESIS

End-To-End Deep Learning Network for 3D Tooth Segmentation.

딥러닝 기반 치아 분할 네트워크

2023 년 8 월

서울대학교 대학원

컴퓨터 공학부

임 호 연

M.S. THESIS

End-To-End Deep Learning Network for 3D Tooth Segmentation.

딥러닝 기반 치아 분할 네트워크

2023 년 8 월

서울대학교 대학원

컴퓨터 공학부

임 호 연

End-To-End Deep Learning Network for 3D Tooth Segmentation.

딥러닝 기반 치아 분할 네트워크

지도교수 신 영 길

이 논문을 공학석사학위논문으로 제출함

2023 년 6 월

서울대학교 대학원

컴퓨터 공학부

임 호 연

임호연의 공학석사 학위논문을 인준함

2023 년 6 월

위 원 장	이 상 구	(인)
부위원장	신 영 길	(인)
위 원	김 건 희	(인)

Abstract

The tooth area of the 3D dental mesh scanned by the 3D Intra Oral Scanner have to be identified for orthodontic treatment. It takes a long time to perform tooth segmentation manually. Therefore, many auto or semi-auto 3D tooth segmentation algorithms have been proposed. However, existing tooth segmentation algorithms exhibits unstable segmentation results for both normal and abnormal dental mesh cases. In addition, these algorithms also cannot predict fine-grained segmentation results near the boundary. To compensate for the shortcomings of existing tooth segmentation methods, we propose a 3d tooth segmentation method based on deep learning. Specifically, the clustering-based instance segmentation algorithm is adopted to make our tooth segmentation robust for both normal and abnormal dental mesh cases. Furthermore, we proposed Boundary Aware Point Sampling, which helps our tooth segmentation method predict fine-grained tooth segmentation results near the boundary. Lastly, Contrastive Boundary Learning can induce our backbone network to generate discernable features in the vicinity of the boundary. The experiment results indicate that our tooth segmentation method outperforms existing methods in terms of segmentation accuracy. Ablation studies demonstrate the effectiveness of the several components adopted for 3d tooth segmentation.

Keywords: deep learning, 3d point cloud analysis, 3d dental mesh segmentation, medical data analysis

Student Number: 2021-26211

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Dental treatment is undergoing significant changes due to advancements in electronic devices and software. All patient information has been digitized, and software programs that can automatically process this information have been developed. Dental mesh is a type of patient information that is used in orthodontic and prosthodontic treatments. In the past, the dental mesh was acquired through scanning a plaster cast that replicated the patient's dental anatomy. In recent times, acquiring dental mesh has become simpler with the advent of intraoral 3D scanners, and the use of these dental meshes in dental treatment has become more convenient. To utilize the dental mesh for treatment, the dentist has to identify the tooth area in the dental mesh, but manual tooth segmentation is a time-consuming process. Consequently, several automated or semi-automated 3D tooth segmentation methods have been proposed.

Traditional 3D tooth segmentation method used general mesh segmentation method([1], [2]). Wu et al. [3] proposed tooth domain-specific algorithms which utilize the inherited property of tooth. Liao et al. [4] suggested a semi-automatic

tooth segmentation algorithm, which is based on the theory of harmonic field to detect partition boundaries. These methods were found to be unreliable when applied to dental meshes that differed from the one considered in an algorithm design phase.

As deep learning has achieved significant success in computer vision, numerous deep learning-based 3d point cloud analysis methods have been proposed. PointNet[5] is a pioneer work that directly processes point cloud with deep neural networks. Subsequently, several deep neural networks have been proposed, including PointNet++[6], DGCNN[7], and Point Transformer[8]). Tooth segmentation methods can leverage point cloud segmentation methods by sampling the vertices of the dental mesh. For example, TSegNet[9] used the Pointnet++ backbone network for 3d tooth segmentation.

However, the existing deep learning-based tooth segmentation methods are not well-established. These methods are not robust in abnormal dental cases, which have missing teeth or braces. In addition, the variations in tooth shape among individuals make difficulties during the training process of deep neural networks. Moreover, the current approaches exhibit relatively low tooth segmentation accuracy in the vicinity of the boundary between teeth and gingiva or adjacent teeth.

In this paper, we propose a segmentation method suitable for automated 3d tooth segmentation with a sampling and learning approach to improve the segmentation accuracy of the boundary part. First, we applied a point cloud segmentation method based on PointGroup[10] to segment the teeth, based on the fact that the teeth themselves have a cylinder shape that is easy to group. Specifically, the offset of each sampled point from the center point of the corresponding tooth is learned. Tooth segmentation is performed by clustering the points that is moved according to their offset. In addition, to improve

the segmentation accuracy in the vicinity of the boundary part, we proposed a sampling technique to sample more points in the boundary part. Moreover, Contrastive Boundary Learning[11] is used to learn discernable features in the vicinity of the boundary. Experiments show that our proposed method outperforms other tooth segmentation methods. Furthermore, ablation studies show the effectiveness of our proposed method.

1.1 Related work

1.1.1 Deep learning based 3d model analysis

Hanocka et al. [12] proposed MeshCNN that directly processes mesh with a convolutional neural network which is designed specifically for edges. MeshCNN was tested only with synthetic single objects rather than multiple real-world objects. Therefore, this method is not suitable for dental mesh.

Several methods have been proposed to analyze the 3d point cloud. Su et al. [13] proposed the MVCNN which is a pioneering work of multi-view based point cloud analysis method. It makes a lot of synthetic projection views and uses a 2d convolutional neural network for point cloud analysis. Maturana et al. [14] converts point cloud to volumetric occupancy grid and utilized a 3d convolutional neural network for dealing with 3d volumetric occupancy grid. These two methods can exploit image processing methods that have been actively studied. However, information is lost in conversion from point cloud to other representations.

Subsequently, techniques for directly processing point clouds were developed. Qi et al. [5] proposed PointNet which processes point cloud directly with rotation and order invariant deep neural network. After that, Qi et al. [6] proposed PointNet++ which is able to learn local features efficiently.

Recently, there have been a number of studies to adopt the transformer proposed in NLP for point cloud analysis. Guo et al. [15] and Engel et al .[16] proposed the networks that leverage the Point Transformer encoder layer for point cloud classification and segmentation tasks. These networks are difficult to predict fine-grained segmentation mask due to their architecture. Zhao et al. [8] introduced Point Transformer which is composed of vector attention layers. It is state-of-the-art on open point cloud segmentation benchmark dataset[17] because it can learn local information efficiently due to the inherent inductive bias of its architecture.

For exploiting these point cloud processing methods, we sample vertices of the original dental mesh to obtain a point cloud. The Features of the point cloud are extracted with the Point Transformer network in our tooth segmentation method.

1.1.2 3D point cloud instance segmentation

Instance segmentation is a task that detects the area of each known object. Inspired by 2d image object detection methods, Hou et al. [18], Yang et al. [19], and Shi et al. [20] proposed instance segmentation frameworks based on region proposal methods. These methods detect the region in which each object exists and then predict the fine-grained instance mask in the next step.

There are other methods that assume that the coordinates and features of the points belong to the same object are similar. Wang et al. [21] proposed SGPN network that groups each point which has similar features. Jiang et al. [10], Zhong et al. [22], Vu et al. [23], and Chen et al. [24] designed a network that predicts the offset between each point and the instance each point belongs to. These methods are state-of-the-art on open point cloud instance segmentation benchmark dataset[17]. Therefore, we adopted a similar approach to Point

Group[10] for our tooth segmentation method.

1.1.3 3D dental model segmentation

Recent developments in deep learning for computer vision tasks have motivated researchers to adopt deep learning for 3d dental model segmentation. Xu et al. [25] feed traditional geometric features of each triangle to the deep learning model. Although there is a limit to performance in that it relied heavily on past feature extraction methods, this method is meaningful because it is a pioneer work to leverage deep learning. Zanjani et al. [26] proposed a tooth instance segmentation framework named Mask-MCNet which is inspired by MaskRCNN which is a 2d image instance segmentation method. Lia et al. [27] proposed the graph-constrained learning module which is one of the graph convolutional networks. Cui et al. [9] introduced TSegNet which first predicts centroids and crops the point cloud around these centroids. Then, the segmentation module of TSegNet takes these cropped point clouds and output instance masks of each tooth instance. In addition to segmentation, Wu et al. [28] performed landmark localization tasks. We introduce a tooth instance segmentation method based on our proposed Tooth Group Network that is an end-to-end deep learning method.

Chapter 2

Proposed Method

2.1 Overview

We propose the tooth segmentation method that aims to segment the tooth instance in a dental mesh and classify the tooth type of each tooth instance. As shown in Fig. 2.1 (a), the tooth segmentation pipeline takes the dental mesh and outputs the tooth instance labels for each vertex of the dental mesh. First, the dental mesh is preprocessed(Chapter **??**) and the sampled point cloud is obtained by applying Farthest Point Sampling to the vertices of the dental mesh. Each point in this sampled point cloud has 3d coordinates and normals. The Tooth Group Network(Chapter 2.2, Fig. 2.1 (b)) accepts the sampled points and generates tooth instance labels and tooth type labels. After that, points near the boundary between the teeth and the gingiva is additionally sampled through Boundary Aware Point Sampling to obtain fine-grained tooth instance labels(Chapter 2.3). Finally, the tooth type of each tooth instance can be predicted by the tooth type labels generated from the Tooth Group Net-

work(Chapter 2.4).

2.2 Tooth Group Network

The Tooth Group Network consists of two modules: the Point Group Module and the Mask Refinement Module. Each module has a Point Transformer as a backbone Network. In this section, we will provide a concise overview of the backbone network, followed by detailed explanations for the Point Group Module and the Mask Refinement Module.

Backbone network. The Tooth Group Network has the Point Transformer[8] as a backbone network for extracting features. The Point Transformer is known as the state-of-the-art backbone network for point cloud semantic segmentation. It extracts local features more effectively compared to other semantic segmentation models like Pointnet++[6] or DGCNN[7], making it suitable for the tooth segmentation task that requires fine-grained tooth instance labels. The structure of the Point Transformer we used is shown in Fig. 2.2. A sequence of Point Transformer layers encodes the features of each point. Each layer of the Point Transformer consists of a point Transformer block and a transition down/up block. The Point Transformer block extracts features by applying vector self-attention and position encoding to the input feature. The transition-down block samples the input points of each layer to enable efficient operation. In order to obtain features of all points, points are upsampled through interpolation in the transition up layer. Multi-scale head[11] predicts offsets and tooth type labels in the Point Group module while it predicts tooth-gingiva mask labels for the Mask Refinement Module. The multi-scale head makes a more accurate prediction than the naive mlp layer by exploiting the features of each decoder layer.

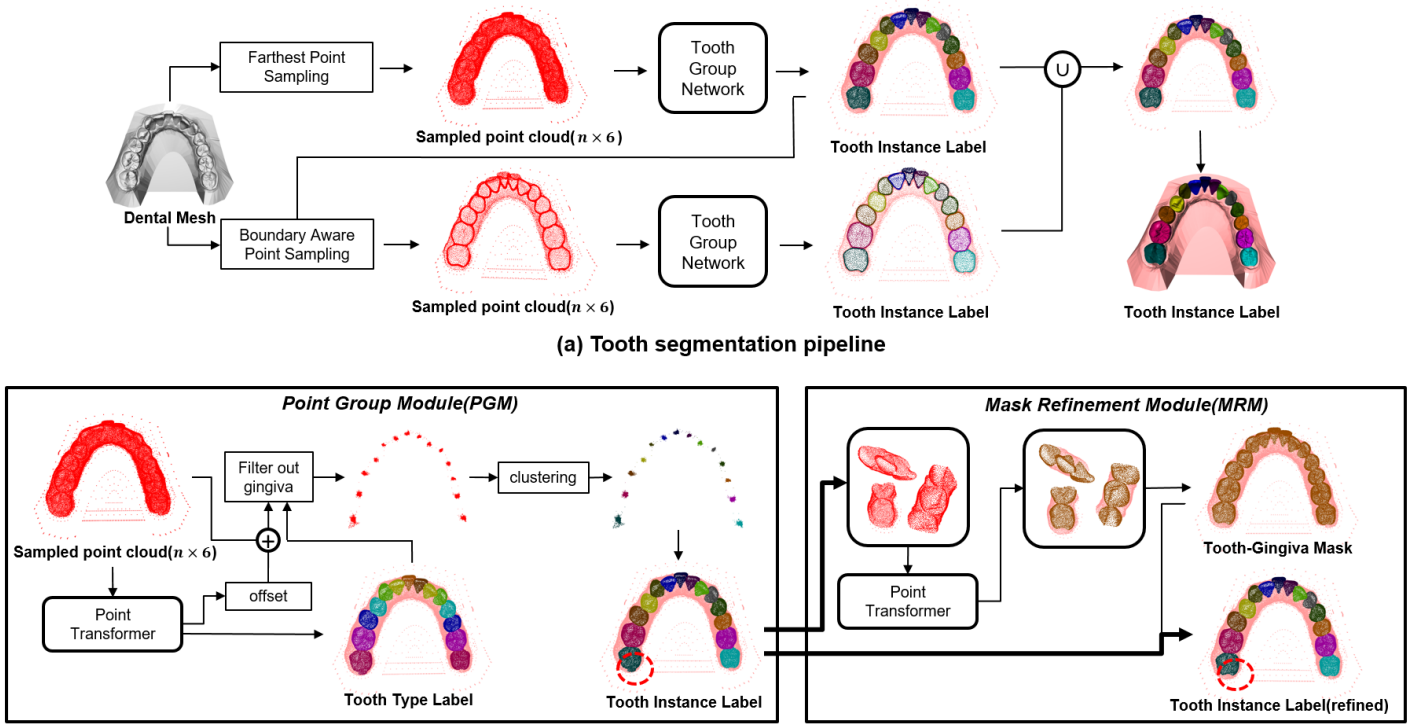


Figure 2.1: (a) Our tooth segmentation pipeline predicts tooth instance labels for each vertex of the dental mesh. (b) Point Group Module takes sampled point cloud and outputs tooth type labels and tooth instance labels. These tooth instance labels are then refined by Mask Refinement Module. The color of the tooth instance label means each tooth instance, and the color of the tooth type label means the type of each tooth instance. For example, green point in tooth type label means the canine

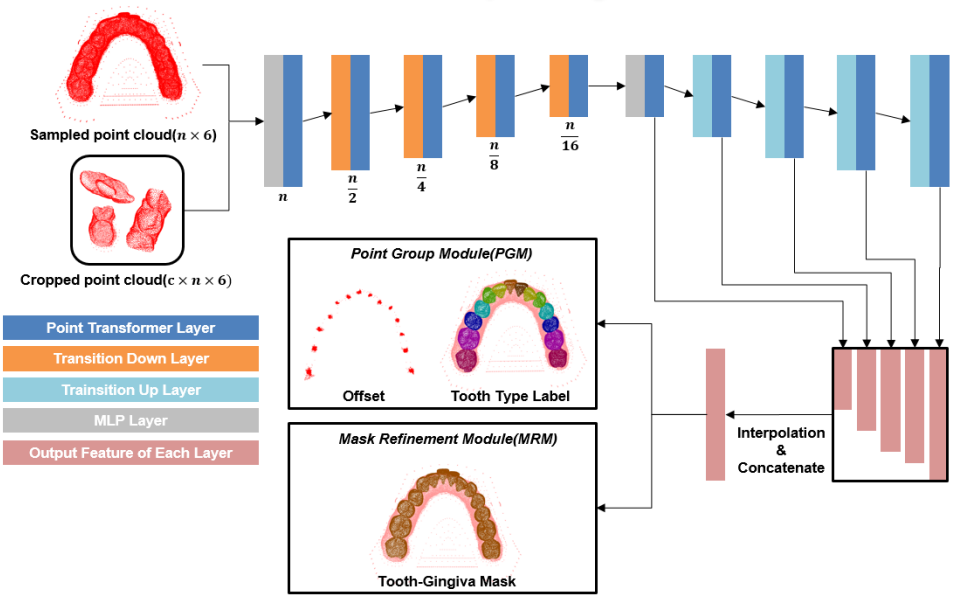


Figure 2.2: The entire architecture of the Point Transformer used in Point Group Network. Sampled point cloud($n \times 6$) is fed into the Point Transformer in the Point Group Module while cropped point cloud($c \times n \times 6$) is fed into the Mask Refinement Module. Here, n means sampled point number, and c means the number of tooth instances in the dental mesh.

Point Group Module. Point Group Module has a similar process to Point Group[10]. The pipeline of the Point Group Module is shown on the left side of Fig. 2.1 (b). The backbone network of the Point Group Module takes the sampled point cloud that contains coordinates and normals of n points. It predicts the tooth type labels and the offsets between each point and the center point of the corresponding tooth instance. Then, the shifted point cloud is obtained by moving each point according to offsets. Points that are predicted as gingiva are filtered out from this point cloud. Points closer to each other in the shifted point cloud are likely to belong to the same tooth instance. Therefore,

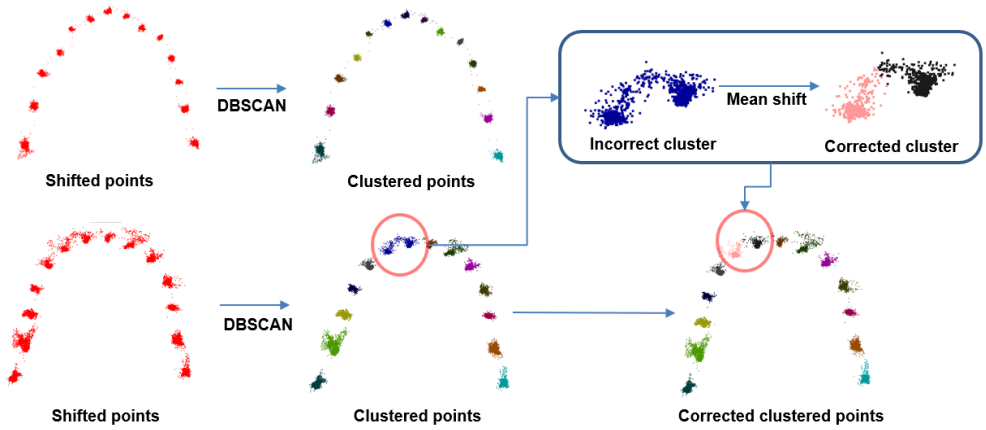
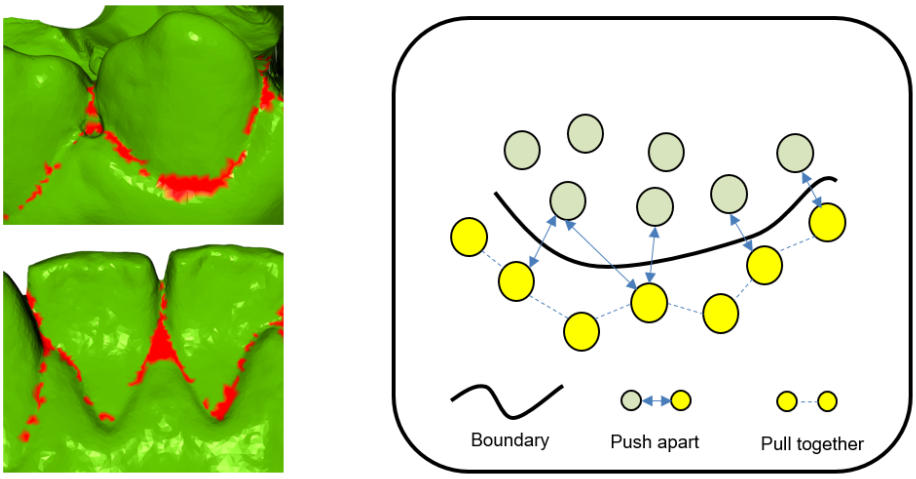


Figure 2.3: After DBSCAN[29], we calculate the variance of each cluster. If the variance of a certain cluster is more than 3 times higher than the average variance of all clusters, it can be considered that two or more tooth instances are clustered into one cluster. In this case, we applied the Mean Shift algorithm to correct these erroneous clusters.

we obtain the tooth instance label of each point by clustering the shifted point cloud with DBSCAN[29]. If there are some incorrectly predicted offsets or some tooth instances are small, more than one tooth instance can belong to the same cluster as you can see in Fig. 2.3. In this case, we compute the variances of each cluster. If the variance of a cluster is more than 3 times higher than the average variance of all clusters, it can be considered that two or more tooth instances are clustered into one cluster. We applied the Mean shift algorithm[30] to correct these erroneous clusters. The Point Group Module, which is based on clustering, is robust due to the cylindrical shape of each tooth. We show this in the ablation study(Chapter 3.4).

Mask Refinement module In order to obtain a more precise tooth segmentation result, we introduced the Mask Refinement Module. As shown in the



(a) Errorneous points in segmentation results (b) Contrastive Boundary Learning

Figure 2.4: (a) The red area means incorrect tooth instance labels while the green area means correct tooth instance labels. (b) visualization of the concept of Contrastive Boundary Learning[11]. Here, each color means ground truth tooth instance label.

right part of Fig. 2.1 (b), the point cloud is cropped around each center point of the predicted tooth instances. Then, each cropped point cloud is fed into the backbone network of the Mask Refinement Module to predict the tooth-gingiva mask. The tooth-gingiva mask is used to refine the tooth instance labels predicted in the Tooth Group Network. The Mask Refinement Module benefits from a cropping process that enables it to exploit more local information, in contrast to the Point Group Module which operates on the entire dental mesh. We also show the effectiveness of the Mask Refinement Module in the ablation study(Chapter 3.4).

Contrastive Boundary Learning The red area in Fig. 2.4 (a) means incorrect tooth instance labels while the green area means correct tooth instance

labels. As you can see in this figure, most erroneous tooth instance labels occur at the boundary between adjacent teeth or between teeth and gingiva. This result indicates that the neural network is having difficulty in accurately classifying tooth instance labels near the boundary. So, inspired by [11], we adopted Contrastive Boundary Learning to help our decoder layers predict contrastive features near the boundary. As shown in Fig. 2.4 (b), Contrastive Boundary Learning makes the features of the neighboring points with different labels distinct from one another. Conversely, neighboring points with the same labels are encouraged to have similar features. The Contrastive Boundary Learning is achieved by training the network with contrastive boundary loss, L_{cbl}. It is defined as follows.

$$L_{cbl} = \frac{-1}{|B_l|} \sum_{x_i \in B_l} log \frac{\sum\limits_{(x_j \in N_i) \wedge (l_j \in l_i)} exp(-d(f_i, f_j)/\tau)}{\sum\limits_{x_k \in N_i} exp(-d(f_i, f_k)/\tau)} \tag{2.1}$$

Here, f_i refers to the feature of x_i and B_l denotes boundary points. We consider certain points as boundary points if they have any different labels within their k neighbors. Note that only the boundary points are affected by contrastive boundary learning. **Total loss for training Tooth Group Network** The Tooth Group Network utilizes the following losses to train the model.

$$L_{Total} = L_{PGM\_off} + L_{PGM\_type} + L_{PGM\_cbl} + L_{MRM\_mask} + L_{MRM\_cbl} \tag{2.2}$$

Point Group Module is trained using L_{PGM_off}, L_{PGM_type} and L_{PGM_cbl}. L_{PGM_off} is regression loss to predict the offsets between the points and the corresponding center points. L_{PGM_type} is cross entropy loss used to generate the tooth type label of each point. In Mask Refine Module, the network learns the masks of the tooth and gingiva with L_{MRM_mask}. L_{PGM_cbl} and L_{MRM_cbl} are the contrastive boundary loss that is applied to Point Group Module and Mask Refine Module, respectively.

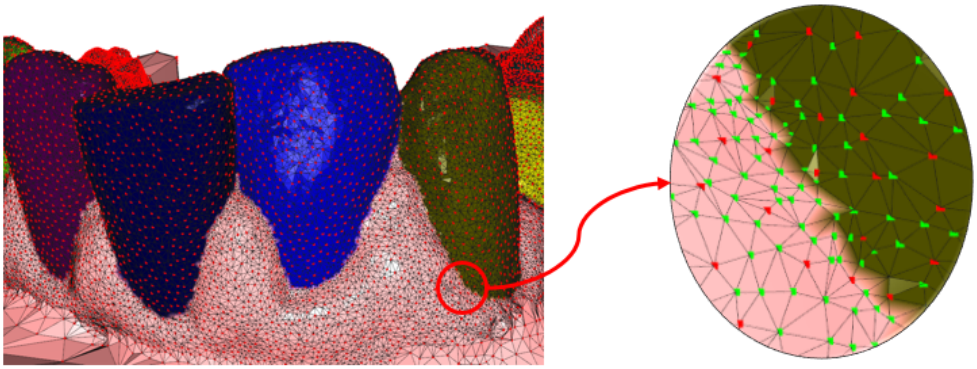


Figure 2.5: The red points represent the sampled points obtained using the Farthest Point Sampling, while the green points correspond to the unsampled points. It can be inferred that we cannot obtain smooth tooth instance segmentation result if we only predict the tooth instance labels of the sampled points, because there are many unpredicted and unsampled points around the boundary.

2.3 Boundary Aware Point Sampling

Because of the limited memory of GPU, we sampled the vertices of the dental mesh. However, all the instance labels of the dental mesh's vertices have to be predicted to obtain the final tooth segmentation result. Therefore, we cannot obtain the fine grained tooth instance labels in the vicinity of the boundary if we only predict the tooth instance labels of the points sampled via Farthest Point Sampling(see Fig. 2.5). In order to solve this problem, we propose the Boundary Aware Point Sampling. It is the point sampling method which samples more points near the predicted boundary. As shown in Fig 2.1 (a), we first obtain the tooth instance labels by feeding the points sampled by Farhest Point Sampling into the Tooth Group Network. Then, we identify the boundary area where predicted tooth instance labels change. The points around the

estimated boundary are sampled via Boundary Aware Point Sampling. Afterward, the sampled points are given as input to the Tooth Group Network. The tooth instance labels of the network output are more precise in the proximity of the boundary. We combine the predicted tooth instance labels obtained from both Farthest Point Sampling and Boundary Aware Point Sampling. The labels of the unsampled vertices are predicted the same as the labels of the nearest sampled vertices.

2.4 Tooth type classification

We classified the eight types of teeth based on the FDI Dental Numbering System from Central Incisior to Third Moral[31]. As can be seen in Fig. 2.1 (b), the Point Transformer in the Point Group Module outputs tooth type labels that can identify which tooth type each point belongs to. The tooth type of each tooth instance is defined as the most frequently occurring tooth type labels within that instance.

2.5 Training details

We use Stochastic Gradient Descent (SGD) as the optimizer during the network training process. The weight decay parameter is set to 1e-4 and the momentum parameter is set to 0.9. We use the cosine decay scheduler with an initial learning rate of 1e-1 and a minimum learning rate of 1e-5 during training. The batch size is 1 due to GPU memory limit. The hyperparameters of the Point Transformer network are the same as [11]. We trained the network for 40 epochs. To address the lack of data, we adopted augmentation techniques including random rotation (between -30 to 30 degrees), random scaling (between 0.85 to 1.15), and random translation (between -0.2 to 0.2). We set the sample point

number n to 24000 for both Farthest Point Sampling and Boundary Aware Point Sampling. In Boundary Aware Point Sampling, 4000 points are sampled by Farthest Point Sampling and 20000 points are sampled in the vicinity of the boundary. If any of the 30 neighboring points of each point has a different label, it is considered as a boundary. The number of points in the cropped point cloud for Mask Refinement Module is 3072. For contrastive boundary loss, the temparature parameter τ is set to 1, and the neighborhood number k is also set to 24.

Chapter 3

Experimental Results

3.1 Dataset

We used the data provided by the MICCAI challenge[32] for the experiment. There are many cases that have missing teeth or braces in this dataset. Each case is composed of maxillary and mandibular meshes with 117415 vertices and 234700 faces on average. Out of 600 cases, 510 cases are used for training and 90 cases are used for testing tooth segmentation methods.

3.2 Evaluation metrics

To measure the performance of the tooth segmentation methods, we used the following three metrics.

Intersection over Union(IoU)

$$\text{IoU} = \frac{|L_{gt} \cap L_{pred}|}{|L_{gt} \cup L_{pred}|} \tag{3.1}$$

The Intersection over Union is commonly used to evaluate the performance of

segmentation algorithms, and this formula is modified for the tooth segmentation algorithm. Here, L_{gt} represents the ground truth instance labels of vertices, and L_{pred} represents the predicted instance labels obtained through the tooth segmentation network.

Boundary Intersection over Union(BIoU)

$$\text{BIoU} = \frac{|L_{b\_gt} \cap L_{b\_pred}|}{|L_{b\_gt} \cup L_{b_pred}|} \tag{3.2}$$

As discussed in Chapter 2.2 and 2.3, there are many incorrect tooth instance labels around the boundary. However, the boundary area is much smaller than the entire tooth area. Thus, we cannot effectively evaluate the accuracy of the predicted tooth instance labels near the boundary with the IoU we described earlier. Hence, the $BIoU$ metric is introduced to assess the precision of the tooth instance labels. $BIoU$ uses the same formula as IoU while $BIoU$ uses the tooth boundary labels instead of tooth instance labels. The label of each vertex is considered as the boundary vertex if one of the neighbor vertices has different label.

Tooth type classification accuracy

This score evaluates the tooth type classification accuracy. As mentioned earlier, the teeth can be categorized into 8 different types.

3.3 Comparison with other methods

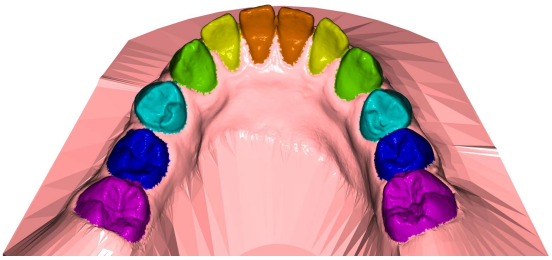
We compare our tooth segmentation method with other methods to assess the performance of our method.

Details for comparing methods We adapt existing point cloud segmentation methods such as PointNet[5], PointNet++[6], DGCNN[7] and Point Transformer[8] to perform tooth segmentation. Specifically, the point cloud sampled from the vertices is fed into the listed networks, and we extract the features

Ground Truth

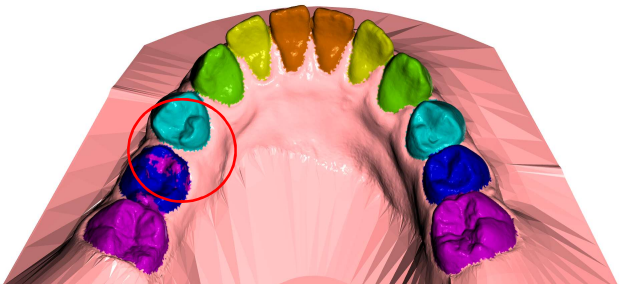
Tooth Group Network(Ours)

Tooth Group Network w/o BAPS



TSegNet

Point Transformer



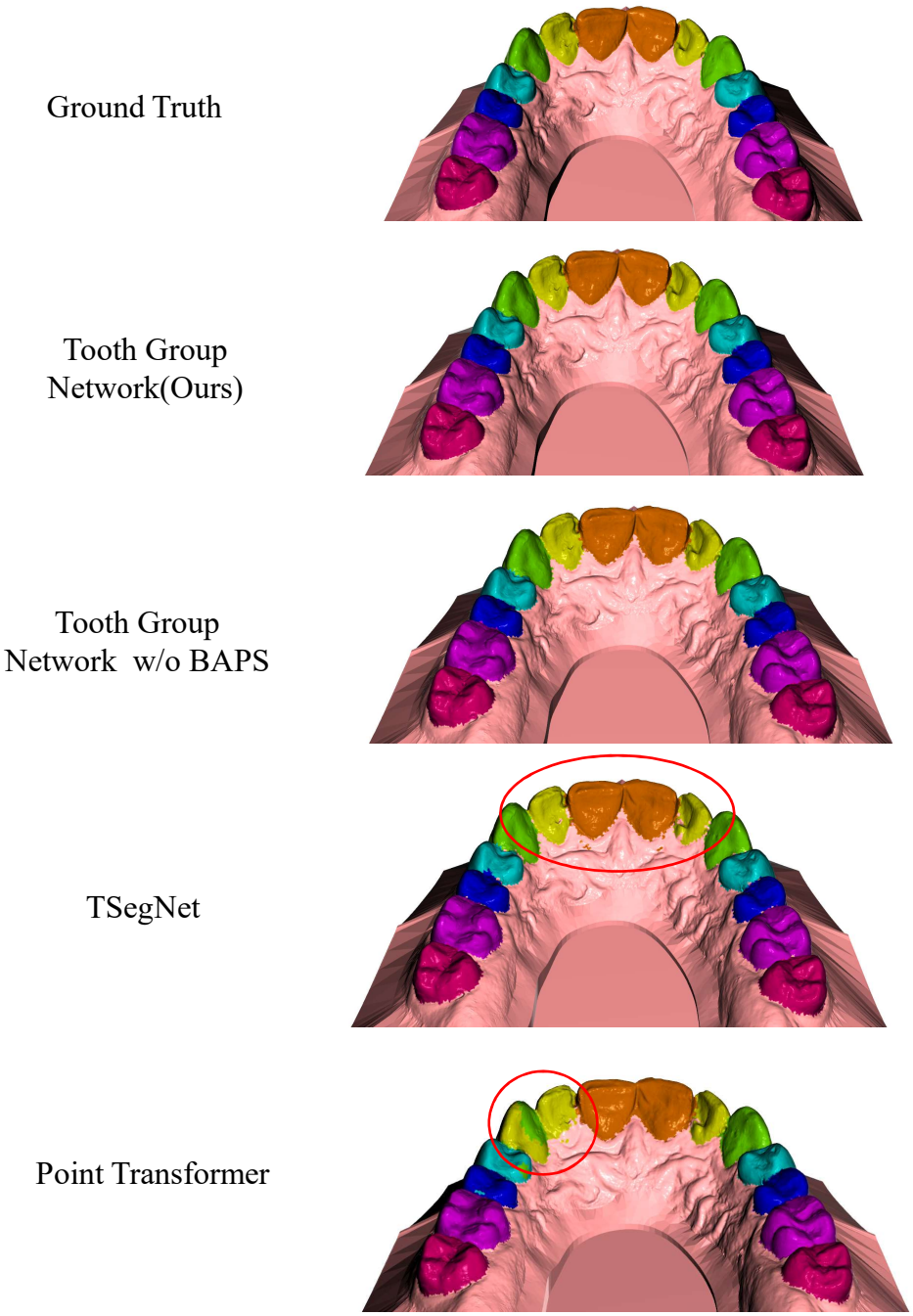


Figure 3.1: Qualitative comparison results.

Method	Metrics		
	$IoU(\%)$	$BIoU(\%)$	$CLS(\%)$
PointNet[5]	78.38%	10.33%	92.63%
PointNet++[6]	89.27%	27.33	96.54%
DGCNN[7]	88.28%	26.34%	95.54%
Point Transformer[8]	91.09%	38.47%	97.08%
TSegNet[9]	91.87%	40.47%	97.58%
Ours w/o boundary learning	93%	44.18%	98.67%
Ours	**95.99** %	**62.3**%	**98.67**%

Table 3.1: Quantitative comparisons on the test dataset.

of each point. We input the extracted features into the segmentation head to obtain the label for each point in the point cloud. These networks are trained by the cross entropy loss. For unsampled vertices, we assign labels based on the label of the nearest sampled vertex. For a fair comparison, we adjust the number of the network parameters to 8M approximately which is also the number of the backbone network parameters of the Tooth Group Network. In addition, we compare our method with TSegNet[9]. TSegNet is proposed for the tooth segmentation task. We have chosen this paper for its demonstration of the method's performance using over 1000 dental cases, which is larger than any other papers in this field. The TSegNet's centroid prediction module uses an offset regression method to predict the center point of each tooth instance. Then, it crops the dental mesh around each center point, and the cropped point cloud is input to the tooth segmentation module. We have implemented TSeg-

Net as described in [9]. The unsampled vertices are predicted using the similar method described above.

Comparison results We quantitively and qualitatively compare the performance of our tooth segmentation method with other methods. Due to the additional points sampled via Boundary Aware Point Sampling, our method takes a long time and uses more GPU memory compared to other methods. Therefore, we also evaluate a variation version of our tooth segmentation method without the Boundary Aware Point Sampling for a fair comparison. As shown in Table 3.1, our method without Boundary Aware Point Sampling outperforms other methods in terms of all of the metrics. This means that our method is better suited for the tooth segmentation task than other networks, not simply because it samples more points, but because it is more efficient for the tooth segmentation task. After applying Boundary Aware Point Sampling, our method showed significantly better performance than other networks. Furthermore, our method has higher tooth type classification accuracy than other methods. Figure 13.3 shows a visual comparison of our method with other tooth segmentation methods.

3.4 Ablation studies

We conducted extensive experiments to evaluate the effectiveness of each component in the Tooth Group Network.

Effectiveness of Boundary Aware Point Sampling As shown in Table 3.2, the IoU score increases to 2.99% when more points are sampled via Boundary Aware Point Sampling. In particular, the $BIoU$ score, which is adopted to evaluate the boundary segmentation accuracy, increases to as much as 20.12%. This result demonstrates that the Boundary Aware Point Sampling achieves

Method	Metrics		
	$IoU(\%)$	$BIoU(\%)$	$CLS(\%)$
Ours w/o Boundary Aware Point Sampling	93%	44.18%	98.67%
Ours w/o Tooth Group Module	93.36%	55.44%	98.1%
Ours w/o Mask Refinement Module	95.08%	59.51%	98.67%
Ours w/o Contrastive Boundary Learning	95.36%	61.3%	98.67%
Ours	**95.99** %	**62.3**%	**98.67**%

Table 3.2: Results of Ablation studies.

the purpose to increase the tooth instance label accuracy in the vicinity of the boundary. As can be seen in the Fig. 3.2, we can obtain the smooth tooth instance labels using Boundary Aware Point Sampling.

Effectiveness of Tooth Group Module The tooth instance segmentation can be executed without the Tooth Group Module. Instead of predicting the offset between a point and its corresponding tooth instance center point, tooth instance labels can be predicted directly by feeding the output features of the Point Transformer into a segmentation head trained with cross-entropy loss. When the tooth instance labels are predicted using the segmentation head instead of the Tooth Group Module, there is a significant decrease in both IoU (by 2.64%) and $BIoU$ (by 6.9%)(Table 3.2. This suggests that the Tooth Group Module is crucial for accurate tooth instance segmentation. As mentioned earlier, the clustering-based method is robust because of the cylinder shape of the tooth instance, which can be clustered even without any predicted offset to some extent. As shown in Fig. 3.3, the tooth instance segmentation accuracy is increased near the boundary between adjacent teeth.

Effectiveness of Mask Refinement Module We conducted an ablation

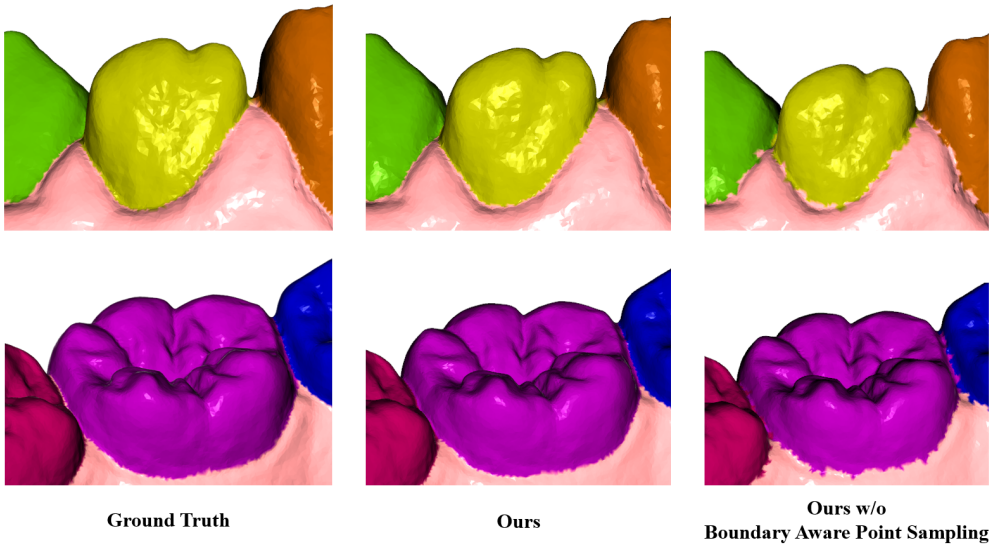


Figure 3.2: Visualization of the tooth segmentation results with or without Boundary Aware Point Sampling.

study to assess the efficacy of the Mask Refinement Module. The Tooth Group Network is compared with the modified network that has the same condition except for Mask Refinement Module. Upon removing the Mask Refinement Module, the results indicate a decrease in IoU to 0.91% and $BIoU$ to 2.7%, as shown in 3.2. As illustrated in the left part of Fig. 3.4, our approach yields accurate results in cases with missing or extracted teeth. As depicted in the right segment of Fig. 3.4, Mask Refinement Module utilizes local features enabling the Tooth Group Network to segment the teeth and gingiva.

Effectiveness of Contrastive Boundary Learning The performance comparison results of the Tooth Group Network with or without Contrastive Boundary Learning or not are presented in Table 3.2. The IoU is increased by 0.6% and $BIoU$ is increased by 1% through the use of Contrastive Boundary Learning. Fig. 3.5 illustrates the effectiveness of Contrastive Boundary Learn-

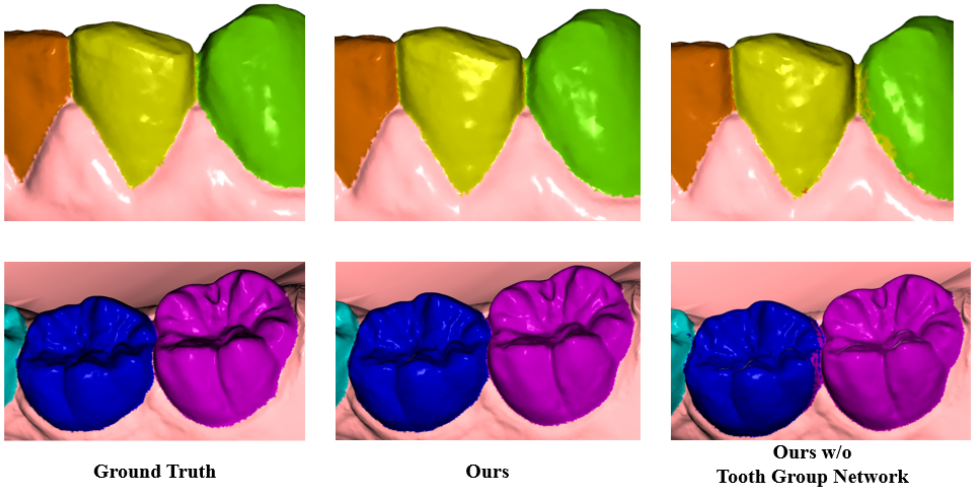


Figure 3.3: Visualization of the tooth segmentation results with or without Tooth Group Network.

ing. Contrastive Boundary Learning can improve instance segmentation performance in the vicinity of the boundary.

3.5 robustness to abnormal case

To evaluate the robustness of our tooth instance segmentation method, we conducted tests on abnormal cases. The results presented in Fig. 3.6 (a) and (b) demonstrate the robustness of our tooth segmentation method in cases of abnormal tooth arrangements due to missing or misaligned teeth. In addition, Ours can perform robustly even if there is an unpredictable noisy gingiva area(Fig. 3.6 (c)) or there are teeth equipped with braces(Fig. 3.6 (d)).

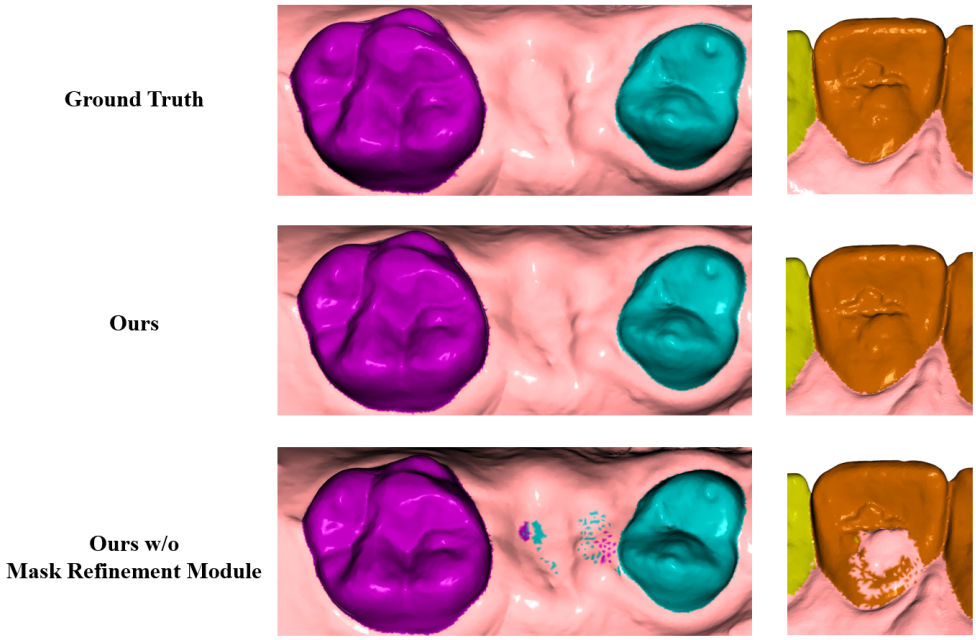


Figure 3.4: Visualization of the tooth segmentation results with or without Mask Refinement Module.

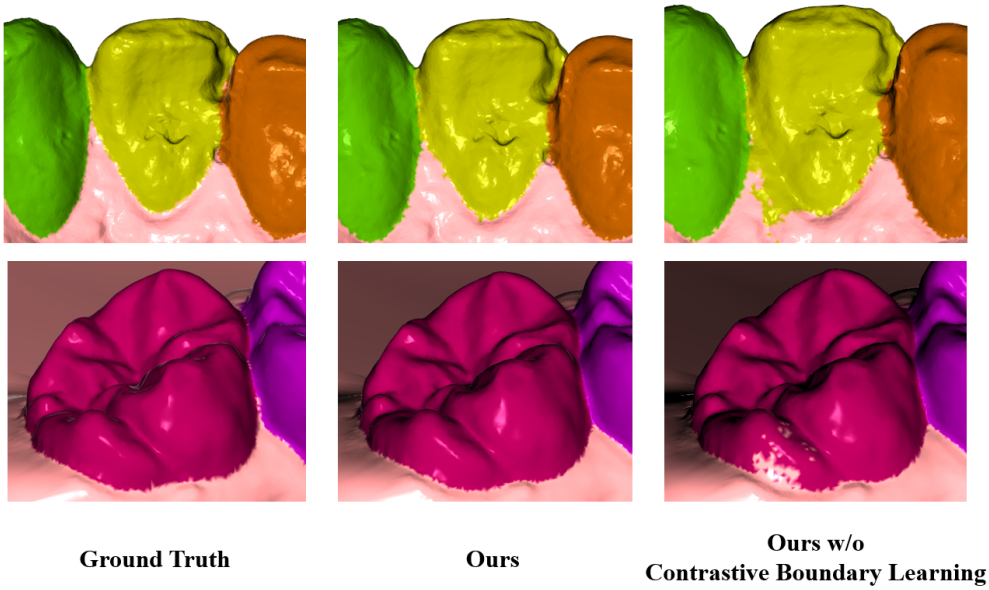


Figure 3.5: Ablation study for Boundary Aware Point Sampling.

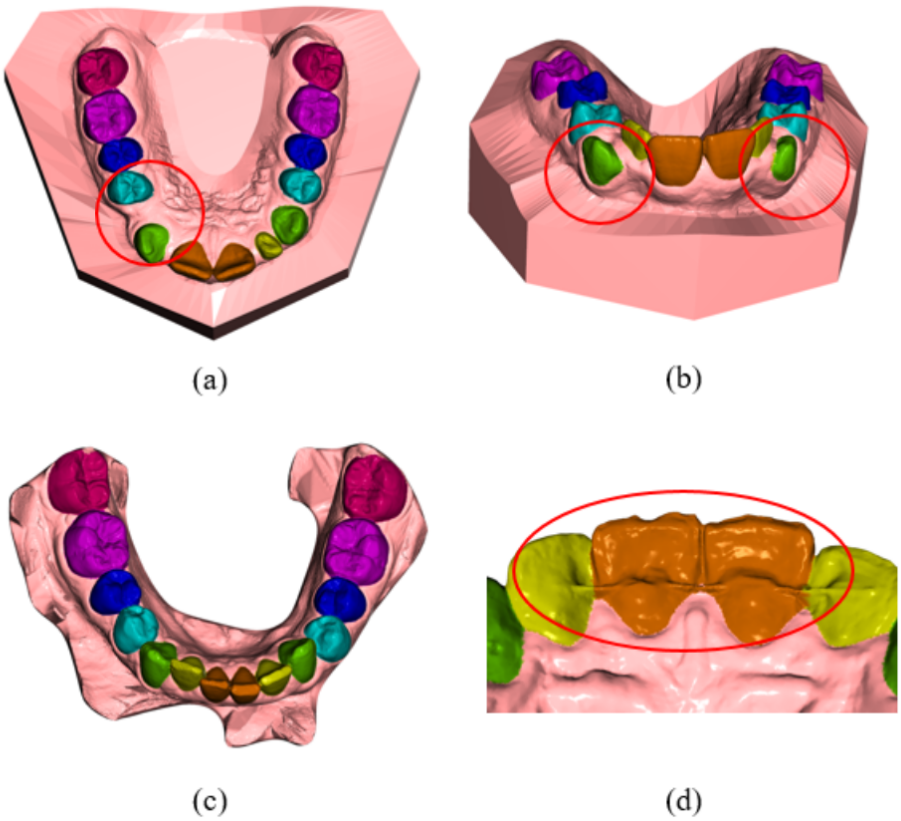


Figure 3.6: Abnormal dental cases. (a) missing teeth. (b) misaligned teeth. (c) raw dental mesh without preprocessing. (d) Teeth equipped with braces.

Chapter 4

Discussion

Tooth segmentation methods cannot reliably perform if the dental mesh differs significantly from the assumed normal mesh. It also has limitations in improving accuracy because only part of the points are sampled due to the GPU memory limitations. Although we adopted several components to address these issues, there remain opportunities for further improvement.

The processing time for each sampling and clustering exceeds 1 second. It is longer than the inference time for the backbone network(Point Transformer[8] which takes less than 200ms for processing 24000 points. Sampling and clustering algorithms are currently implemented to use CPU. Therefore, it can be optimized if we implement the same algorithms using GPU.

The tooth classification of third morals(wisdom tooth) is another limitation. Out of 600 dental cases, only 6 dental cases have third molars. Because deep learning networks have to be trained with a lot of data, it can be inferred that the network cannot predict third morals reliably. One possible solution to address this issue could be to adopt weighted cross entropy loss in the

training process or provide additional third moral dental cases. Other solutions could be to incorporate domain-specific knowledge about Dentistry.

The limited availability of diverse labeled dental datasets hampers the improvement of the deep learning-based approach. There are a lot of different types of 3D dental scanners, and the appearance of the inside of the oral cavity or teeth varies greatly among individuals. Thus, there are significant differences between the scanned dental meshes. Our algorithm can help the process of acquiring labeled scanned data.

Chapter 5

Conclusion

In this work, we propose a robust and accurate tooth segmentation method. We utilized a clustering-based tooth instance segmentation method, which predicts the offset between each point and its corresponding tooth center point. Contrastive boundary learning is introduced for our network to generate more distinctive features near the boundary. Moreover, we propose the Boundary Aware Point Sampling to sample more points in the vicinity of the boundary. Our method has higher MIoU and tooth type classification accuracy compared to other tooth instance segmentation methods. Notably, we observed a significant improvement in the BIoU score, which is used to evaluate the accuracy of tooth segmentation along the boundary. Several ablation studies verified the effectiveness of each component. Furthermore, we demonstrated that our tooth segmentation method can perform in abnormal dental cases. The dentist can work efficiently with our method because of the smoothed segmentation result near the boundary between adjacent teeth and or teeth and gingiva.

Bibliography

[1] Marco Attene, Sagi Katz, Michela Mortara, Giuseppe Patané, Michela Spagnuolo, and Ayellet Tal. Mesh segmentation-a comparative study. In *IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06)*, pages 7–7. IEEE, 2006.

[2] Ariel Shamir. A survey on mesh segmentation techniques. In *Computer graphics forum*, volume 27, pages 1539–1556. Wiley Online Library, 2008.

[3] Kan Wu, Li Chen, Jing Li, and Yanheng Zhou. Tooth segmentation on dental meshes using morphologic skeleton. *Computers & Graphics*, 38:199–211, 2014.

[4] Sheng-hui Liao, Shi-jian Liu, Bei-ji Zou, Xi Ding, Ye Liang, and Jun-hui Huang. Automatic tooth segmentation of dental mesh based on harmonic fields. *BioMed research international*, 2015, 2015.

[5] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[6] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

[7] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019.

[8] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16259–16268, 2021.

[9] Zhiming Cui, Changjian Li, Nenglun Chen, Guodong Wei, Runnan Chen, Yuanfeng Zhou, Dinggang Shen, and Wenping Wang. Tsegnet: An efficient and accurate tooth segmentation network on 3d dental model. *Medical Image Analysis*, 69:101949, 2021.

[10] Li Jiang, Hengshuang Zhao, Shaoshuai Shi, Shu Liu, Chi-Wing Fu, and Jiaya Jia. Pointgroup: Dual-set point grouping for 3d instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and Pattern recognition*, pages 4867–4876, 2020.

[11] Liyao Tang, Yibing Zhan, Zhe Chen, Baosheng Yu, and Dacheng Tao. Contrastive boundary learning for point cloud segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8489–8499, 2022.

[12] R. Hanocka, A. Hertz, N. Fish, R. Giryes, S. Fleishman, and D. Cohen-Or. Meshcnn: A network with an edge. *ACM Transactions on Graphics*, 38:90:1–90:12, 2019.

[13] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *IEEE International Conference on Computer Vision*, pages 945–953, 2015.

[14] D. Maturana and S. Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 922–928, 2015.

[15] M. H. Guo, J. X. Cai, Z. N. Liu, T. J. Mu, R. R. Martin, and S. M. Hu. Pct: Point cloud transformer. *Computational Visual Media*, 7:187–199, 2021.

[16] N. Engel, V. Belagiannis, and K. Dietmayer. Point transformer. *IEEE Access*, 9:134826–134840, 2021.

[17] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1534–1543, 2016.

[18] J. Hou, A. Dai, and M. Nießner. 3d-sis: 3d semantic instance segmentation of rgb-d scans. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4421–4430, 2019.

[19] B. Yang, J. Wang, R. Clark, Q. Hu, S. Wang, A. Markham, and N. Trigoni. Learning object bounding boxes for 3d instance segmentation on point clouds. *Advances in neural information processing systems*, pages 6737–6746, 2019.

[20] S. Shi, X. Wang, and H. Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 770–779, 2019.

[21] W. Wang, R. Yu, Q. Huang, and U. Neumann. Sgpn: Similarity group proposal network for 3d point cloud instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2569–2578, 2018.

[22] M Zhong, X Chen, X Chen, G Zeng, and Y Wang. Maskgroup: Hierarchical point grouping and masking for 3d instance segmentation. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2022.

[23] T. Vu, K. Kim, T. M. Luu, T. Nguyen, and C. D. Yoo. Softgroup for 3d instance segmentation on point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2708–2717, 2022.

[24] S. Chen, J. Fang, Q. Zhang, W. Liu, and X. Wang. Hierarchical aggregation for 3d instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15467–15476, 2021.

[25] X. Xu, C. Liu, and Y. Zheng. 3d tooth segmentation and labeling using deep convolutional neural networks. *IEEE transactions on visualization and computer graphics*, 25:2336–2348, 2018.

[26] F. G. Zanjani, D. A. Moin, F. Claessen, T. Cherici, S. Parinussa, A. Pourtaherian, and S. Zinger. Mask-mcnet: Instance segmentation in 3d point cloud of intra-oral scans. pages 128–136, 2019.

[27] C. Lian, L. Wang, T. H. Wu, F. Wang, P. T. Yap, C. C. Ko, and D. Shen. Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3d intraoral scanners. *IEEE Transactions on medical imaging*, 39:2440–2450, 2020.

[28] T. H. Wu, C. Lian, S. Lee, M. Pastewait, C. Piers, J. Liu, F. Wang, L. Wang, C. Y. Chiu, and W. Wang. Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3d intraoral scans. *IEEE Transactions on Medical Imaging*, 2022.

[29] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21, 2017.

[30] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence*, 17(8):790–799, 1995.

[31] S Peck and L Peck. A time for change of tooth numbering systems. *Journal of dental education*, 57(8):643–647, 1993.

[32] Achraf Ben-Hamadou, Oussama Smaoui, Houda Chaabouni-Chouayakh, Ahmed Rekik, Sergi Pujades, Edmond Boyer, Julien Strippoli, Aurélien Thollot, Hugo Setbon, Cyril Trosset, et al. Teeth3ds: a benchmark for teeth segmentation and labeling from intra-oral 3d scans. *arXiv preprint arXiv:2210.06094*, 2022.

초록

치과에서 환자의 치아를 교정하거나 수복하는 치료에 대한 계획을 세우기 위해, 치기공사들은 CAD 툴을 이용하여 3차원 구강 스캐너를 통해 수집된 치아 매시 데이터에서 각 치아 영역을 분리해야 한다. 3차원 치아 매시 데이터에서 각 치아를 분리해내는 일은 시간이 많이 소요되는 작업이며, 이를 보완하기 위해 자동/반자동 치아 분할 알고리즘이 연구됐다. 하지만, 이전까지 제안된 알고리즘들은 상실된 치아가 있는 경우나 교정기가 부착되어있는 경우 등 특이 케이스에 대해 잘 동작하지 않는 경우가 있었으며, 평범한 구강 스캔 데이터라도 불안정한 분할 결과를 내놓는 경우가 많았다. 또한, 치아와 잇몸 사이의 경계나 치아와 치아 사이의 경계와 가까운 부분에서 잘못된 분할 결과를 내놓는 경우가 많았다.

본 논문에서는 기존 치아 분할 알고리즘의 단점을 보완하기 위한 인공지능 네트워크를 제안한다. 구체적으로, 클러스터링 방식의 인스턴스 분할 방법을 적용하여 상실 치아가 있는 경우나 교정기가 부착된 케이스 등 다양한 케이스에서 올바른 치아 분할 결과를 도출할 수 있도록 네트워크를 구성하였다. 그리고, 대조 경계 학습을 적용하여 치아와 치아사이의 경계와 치아와 잇몸 사이의 경계 부분에서 치아 분할성능을 높였다. 또한, 경계 인식 포인트 샘플링 기법을 제안하고, 이를 이용해 경계 부분에서 더 많은 매시의 정점을 샘플링함으로써 치아 분할 성능을 높였다.

본 논문에서 제안한 치아 분할 방법이 기존에 제안된 다른 방법보다 더 좋은 성능을 가짐을 실험적으로 확인하였다. 또한, 추가 실험을 통해 우리가 치아 분할 알고리즘에 도입한 각 요소들이 효과가 있음을 보였다.

주요어: 딥러닝, 3차원 포인트 클라우드 분할, 3d 구강 스캔 데이터 분할, 의료 데이터 분석

학번: 2021-26211