

Evaluating the Suitability of Different Intraoral Scan Resolutions for Deep Learning-Based Tooth Segmentation

Daron Marshall Weekley, Marshall University

Daron Weekley is a Graduate Student in the Department of Computer Science at Marshall University. He has a master's and bachelor's degree in biology from West Liberty University. Before entering the computer science field, Daron was published for his work in research labs specializing in neuroscience and microbiology.

Jace A McPherson-Duckworth, Marshall University

Jace Duckworth, Undergraduate at Marshall University with an interest in AI research

Anastasiia Sukhanova, Marshall Community & Technical College

Dr. Ananya Jana, Marshall University

Evaluating the Suitability of Different Intraoral Scan Resolutions for Deep Learning-Based Tooth Segmentation

Daron Weekley¹, Jace Duckworth¹, Anastasiia Sukhanova¹, Ananya Jana¹,

¹Department of Computer Sciences and Electrical Engineering,
Marshall University, Huntington, WV, USA

Corresponding author: jana@marshall.edu

Abstract

Intraoral scans are widely used in digital dentistry for tasks such as dental restoration, treatment planning, and orthodontic procedures. These scans contain detailed topological information, but manual annotation of these scans remains a time-consuming task. Deep learning-based methods have been developed to automate tasks such as tooth segmentation. A typical intraoral scan contains over 200,000 mesh cells, making direct processing computationally expensive. Models are often trained on downsampled versions, typically with 10,000 or 16,000 cells. Previous studies suggest that downsampling may degrade segmentation accuracy, but the extent of this degradation remains unclear. Understanding the extent of degradation is crucial for deploying ML models on edge devices. This study evaluates the extent of performance degradation with decreasing resolution. We train a deep learning model (PointMLP) on intraoral scans decimated to 16K, 10K, 8K, 6K, 4K, and 2K mesh cells. Models trained at lower resolutions are tested on high-resolution scans to assess performance. Our goal is to identify a resolution that balances computational efficiency and segmentation accuracy.

Introduction

Intraoral scanning is a digital technique that captures a 3D representation of an intraoral region. These scans are widely used in digital dentistry for dental restoration, orthodontic treatment planning, smile design, and surgery. However, many of these tasks still require manual processing, such as annotation. Annotating a single scan can take 45 to 60 minutes. Hence, automatic methods are necessary to reduce the manual efforts.

There are many deep learning-based methods that have been developed for the task of tooth segmentation, dental crown generation, and dental alignment, often adapting techniques from

general point cloud processing. Intraoral scans can contain over 200,000 mesh triangles, making direct processing computationally prohibitive. The existing works downsample the original scans to a lower resolution such as 16K or 10K mesh cells as shown in Table. 1. Prior studies have explored different topics such as the shortcoming of existing tooth segmentation methods [1] and possibility of learning from a single dental scan [2]. Prior studies [3, 4] also suggest that downsampling degrades the fine tooth features (for example, curvature information), but a systematic evaluation of its impact on segmentation accuracy has not been done. Our study attempts to fill in this gap by analyzing the effect of different resolution levels on model performance.

Deep-learning models trained on downsampled meshes are more suitable for deployment on devices that require methods with lower memory footprint. However, downsampling beyond a certain threshold can reduce segmentation accuracy. In this study, we evaluate tooth segmentation performance when downsampling intraoral scans across different resolutions. Using PointMLP [5], a deep learning model for point cloud processing, we assess segmentation accuracy on 10K and 16K scans. By evaluating segmentation performance across various resolutions, this study provides insights into the feasibility of using downsampled intraoral scans for deep learning-based segmentation.

Author/Paper	Faces/Points
Xu et al. [6]	40,000 points
Sun et al. [7]	10,000 points
Zhang et al. [8]	16,000 points
Cui et al. [9]	10,000 cells
Zhang et al. [10]	16,000 cells
Tian et al. [11]	16,000 cells
Zanjani et al. [12]	16,000 points
Wu et al. [13]	10,000 cells
Lian et al. [14]	10,000 points
Zhang et al. [15]	10,240 points
Benhamadou et al. [16]	10,000 points
Jana et al. [17]	16,000 points

Table 1: Intraoral scan resolutions used in various studies

Current deep learning-based approaches have shifted toward data-driven feature extraction. Xu et al. [18] introduced a multi-stage framework using CNNs for mesh cell labeling. Tian et al. [19] employed octree partitioning followed by a 3D CNN for hierarchical tooth segmentation. Jana et al. [20] developed a dual-branch segmentation method, leveraging sparse surface representation and curvature learning. 3D tooth segmentation methods primarily rely on fully supervised learning, though some, such as DArch [21] and TSegNet [22], use weak or semi-supervised learning. As such, these approaches typically operate on 16K or 10K mesh cells per scan. To our knowledge, no study systematically evaluates the impact of downsampling on tooth segmentation performance.

1. Methods

1.1. PointMLP

In this work, we have used a point cloud segmentation method named PointMLP [5]. The PointMLP method uses a geometric affine module, which has demonstrated exceptional capability of capturing local geometry, even from sparse representation. Due to these characteristics, the PointMLP method seems to be a reasonably good choice. Given an input point cloud, PointMLP extracts local features step by step using residual point MLP blocks. First, a geometric affine module transforms the local points. Features are then extracted both before and after the aggregation process. By repeating these stages, PointMLP expands the receptive field and captures the full geometric structure of the point cloud. Figure 1 shows a visualization of the PointMLP method.

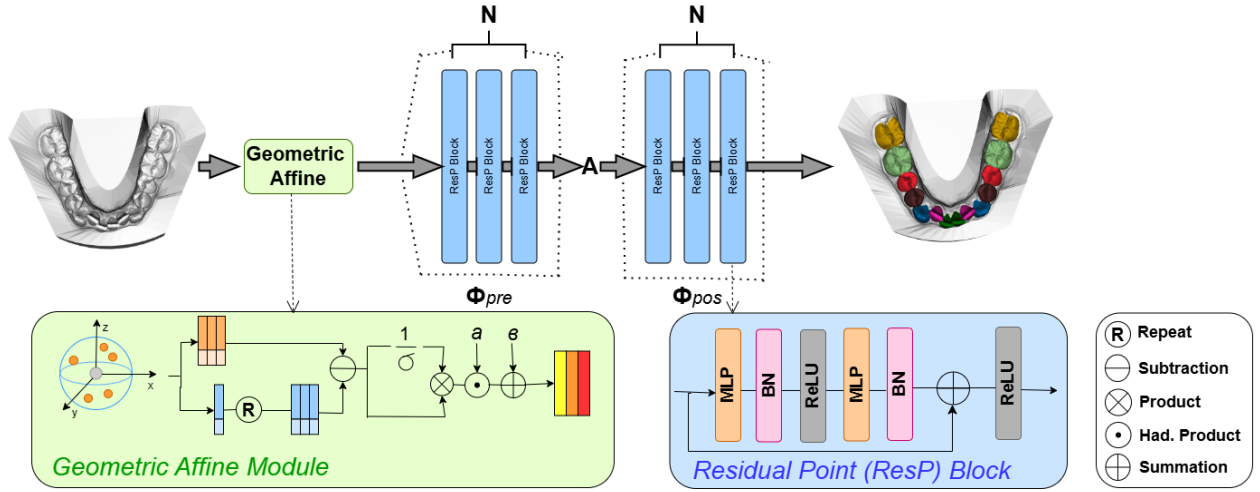


Figure 1: PointMLP method comprises of residual blocks (ϕ_{pre} and ϕ_{post}). α and β are the learnable parameters. σ denotes the feature deviation across all groups. The PointMLP method transforms the points via a normalization operation while maintaining original geometric properties. (Zoom in for better visibility)

1.2. Experimental Set-up

The original intraoral scans ($\sim 200K$ cells) were downsampled to the resolutions of 2K, 4K, 6K, 8K, 10K and 16K and a separate model was trained for each resolution. The results of the prediction were upsampled via KNN method to the 10K and 16K resolution to understand the effectiveness of the models trained at lower resolutions (2K, 4K, 6K and 8K). Examples of the differences in resolutions are shown in Figure 2.

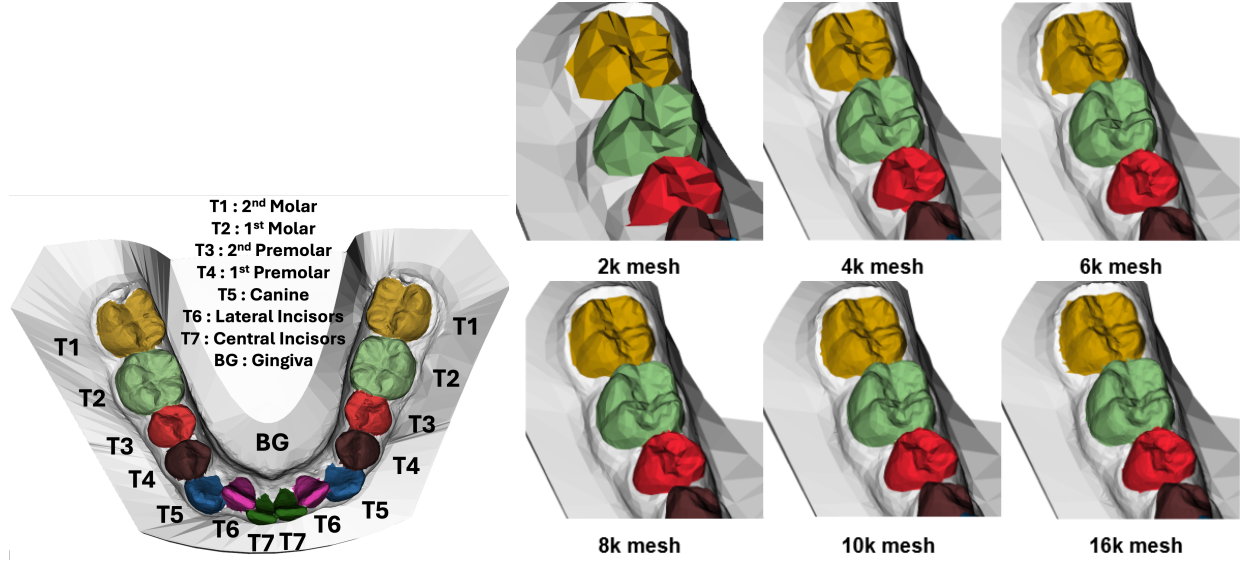


Figure 2: A labeled tooth mesh (left) and different resolutions of the same region in the intraoral scan.(Zoom in for better visibility)

1.3. Dataset & Evaluation Metrics

We use the public dataset *3D Teeth Seg Challenge 2022* [23]. The intraoral scans provided the ground truth annotations of different teeth and gum regions, including central incisors (T7), lateral incisors (T6), canine/cuspids (T5), 1st premolars (T4), 2nd premolars (T3), 1st molars (T2), 2nd molars (T1), and background/gingiva (BG)(Figure 2). Tooth segmentation is the task of separating 3D dental models into these unique semantic parts. We use Dice Score (DSC), Overall Accuracy (OA), Sensitivity (SEN) and Positive Predictive Value (PPV) to evaluate the performance of our model.

1.4. Data Pre-Processing

In our experiments, we used a subset of the public dataset consisting of 571 subjects from the training set of the dataset. The lower jaw scans of the subjects with a maximum of 14 teeth were selected. Each of these resolutions have been created from the original mesh by the process of mesh quadric decimation. Each mesh cell can be described with four points - three vertices of the mesh triangle and the barycenter of the mesh triangle along with the normals at each of these four points. With these four points, a 24 dimensional vector is constructed comprising 12 coordinate vectors and 12 normal vectors at the four points, respectively. The intraoral scans have a base which is not part of the gum. In our current work, we crop a portion of the dental mesh from the scan. Afterward, we manually verified the meshes to ensure that we have teeth and gingiva for all the models. To efficiently process and visualize 3D intraoral scans, we integrate the Vedo and VTK libraries into our workflow. Vedo is utilized for interactive rendering, surface manipulation, and mesh analysis, facilitating an intuitive exploration of the segmentation outputs. Meanwhile, VTK is employed for advanced 3D processing, geometric transformations, and volumetric

rendering of dental structures.

1.5. Data Augmentation

For better generalizability of the model, we augment the training and validations sets by combining 1) random rotation, 2) random translation, and 3) random rescaling of each 3D dental surface in reasonable ranges. Specifically, along each of the three axes in the 3D space, a training/validation surface has 50% probability of translation with a displacement and zoom with a ratio uniformly sampled between $[-10, 10]$ and $[0.8, 1.2]$, respectively. Each training/validation surface has 50% probability to be rotated along the x, y, and z-axes with angles uniformly sampled between $[-\pi, \pi]$. The combination of these random operations simulated 4 “new” cases from each original surface. We split the data at an 80:20 ratio for training and testing. Next, we allocate 20% of the training data for the purpose of validation. To ensure consistency across augmented samples, we apply the same transformation parameters to the corresponding labels and annotations. Data augmentation is performed statically during the data preprocessing stage to reduce training time.

1.6. Training and Evaluation

We train the segmentation model using PointMLP with multi-resolution intraoral scans. Training is conducted for 200 epochs using the Adam optimizer with an initial learning rate of 0.001. The batch size is set to 16 and each epoch took approximately 5 minutes to complete. Training is conducted on RTX 6000 GPUs. The entire pipeline, including preprocessing, training, and evaluation, is implemented in PyTorch.

2. Results and Analysis

2.1. Segmentation Performance

We evaluate the segmentation performance of PointMLP across different resolutions and report the results in Table 2. The distribution of the segmentation labels is shown in Figure 5. The model trained at each resolution is tested against the standard 10K and 16K resolution scans to analyze performance degradation. The overall results from these experiments are listed in Table 3. We analyze segmentation performance for background (BG) and each tooth class (T1–T7) and report the Dice scores in Table 3. We measure per-class segmentation accuracy and report the results in Table 2. The qualitative results of our experiments is shown in Figure 3 and Figure 4. We also assess segmentation consistency across different test cases. Our findings highlight the trade-offs between resolution and segmentation accuracy, providing insights into model robustness and practical deployment considerations.



Figure 3: Qualitative results of predictions from the original models and their upsampled versions.
(Zoom in for better visibility)

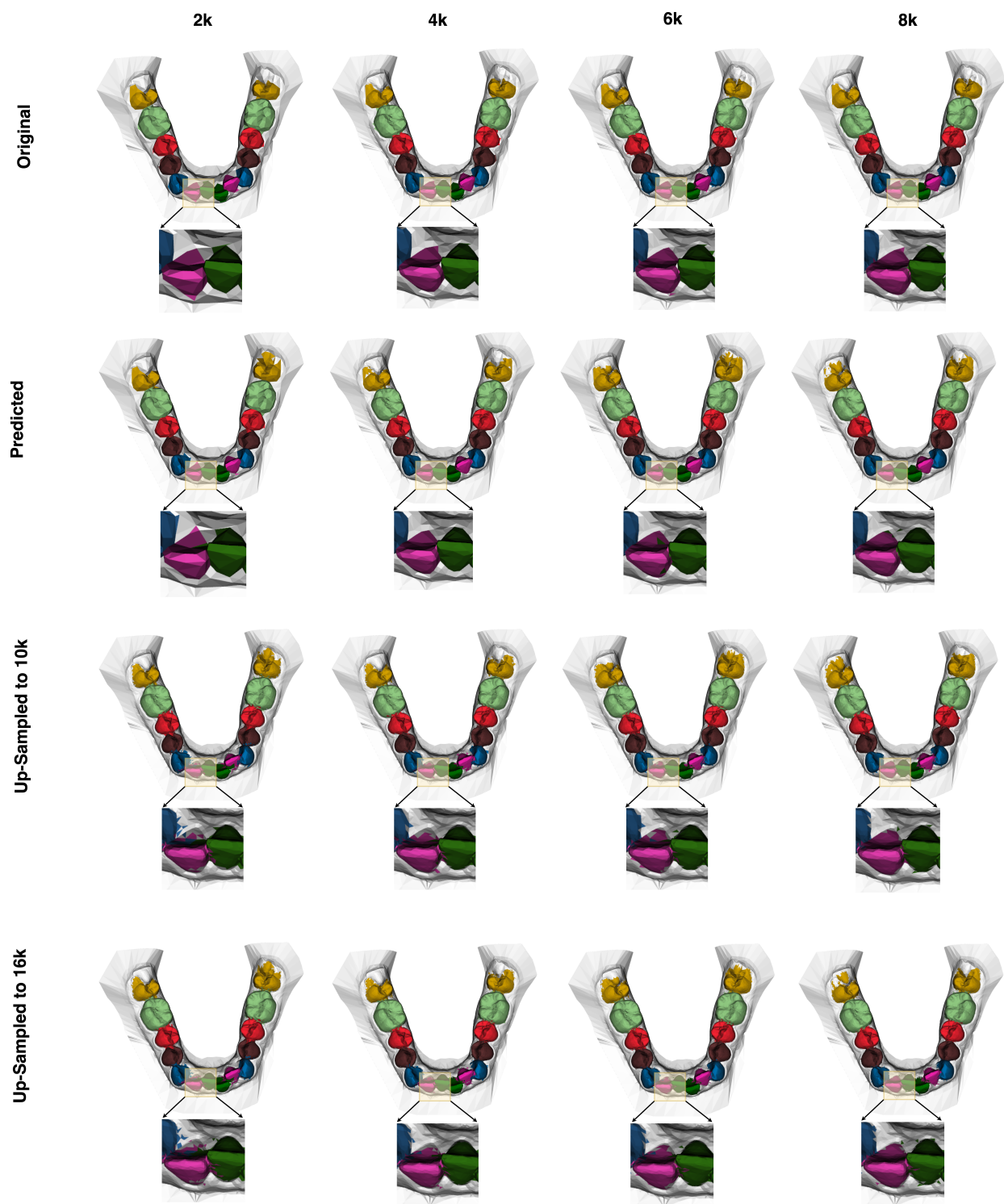


Figure 4: Qualitative results of predictions from the original models and their upsampled versions (continued). (Zoom in for better visibility)

Table 2: Per-class DSC score across different resolutions and their upsampled versions.

Resolution	BG	T1	T2	T3	T4	T5	T6	T7
16K	0.9646	0.9566	0.9457	0.9443	0.9402	0.9265	0.9163	0.8606
10K	0.9701	0.9605	0.9504	0.9500	0.9442	0.9267	0.9171	0.8435
8K	0.9728	0.9642	0.9527	0.9527	0.9440	0.9252	0.9232	0.8441
8K (to 16K)	0.9468	0.9181	0.9164	0.9068	0.8964	0.8929	0.8927	0.8389
8K (to 10K)	0.9569	0.9283	0.9256	0.9121	0.9070	0.9048	0.9013	0.8422
6K	0.9752	0.9686	0.9573	0.9564	0.9488	0.9250	0.9177	0.8533
6K (to 16K)	0.9405	0.9112	0.9075	0.8999	0.8873	0.8784	0.8777	0.8344
6K (to 10K)	0.9502	0.9209	0.9155	0.9044	0.8965	0.8885	0.8849	0.8373
4K	0.9786	0.8793	0.9430	0.9437	0.9738	0.9673	0.9566	0.9680
4K (to 10K)	0.8319	0.7536	0.6679	0.6626	0.6498	0.6426	0.6145	0.5569
4K (to 16K)	0.8208	0.7510	0.6647	0.6585	0.6481	0.6397	0.6116	0.5558
2K	0.9738	0.8760	0.9410	0.9347	0.9636	0.9495	0.9371	0.9504
2K (to 10K)	0.8213	0.7578	0.6905	0.6838	0.6546	0.6408	0.6303	0.5651
2K (to 16K)	0.8320	0.5697	0.6923	0.6435	0.6876	0.6571	0.6317	0.7586

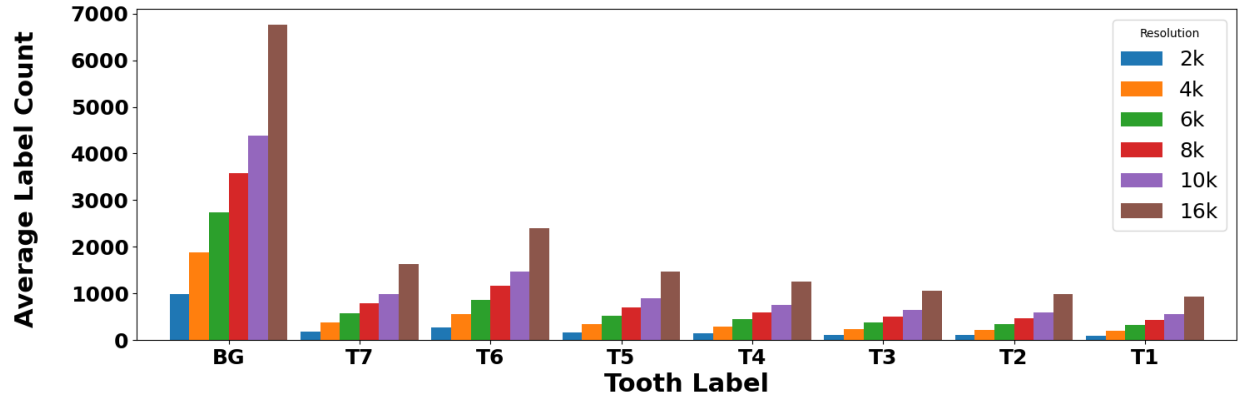


Figure 5: The distribution of the labels across different classes. (Zoom in for better visibility)

Table 3: Tooth segmentation results using PointMLP with different resolutions

Input Size (#mesh cells)	OA	DSC	SEN	PPV	Inference Time (ms)
16K	0.9476	0.9318	0.9451	0.9353	158.93
10K	0.9515	0.9328	0.9467	0.9377	248.78
8K	0.9532	0.9348	0.9495	0.9396	207.89
8K (upsample to 10K)	0.2454	0.9097	0.9172	0.9210	258.26
8K (upsample to 16K)	0.2454	0.9011	0.9061	0.9150	209.05
6K	0.9549	0.9377	0.9490	0.9442	163.45
6K (upsample to 10K)	0.2372	0.8997	0.9015	0.9166	298.59
6K (upsample to 16K)	0.2372	0.8921	0.8916	0.9115	190.06
4K	0.9642	0.9513	0.9552	0.9591	279.69
4K (upsample to 10K)	0.2575	0.6724	0.6836	0.7340	302.09
4K (upsample to 16K)	0.2575	0.6687	0.6788	0.7317	192.33
2K	0.9577	0.9408	0.9465	0.9491	102.22
2K (upsample to 10K)	0.2956	0.6840	0.6942	0.7400	233.36
2K (upsample to 16K)	0.2956	0.6805	0.6904	0.7376	413.72

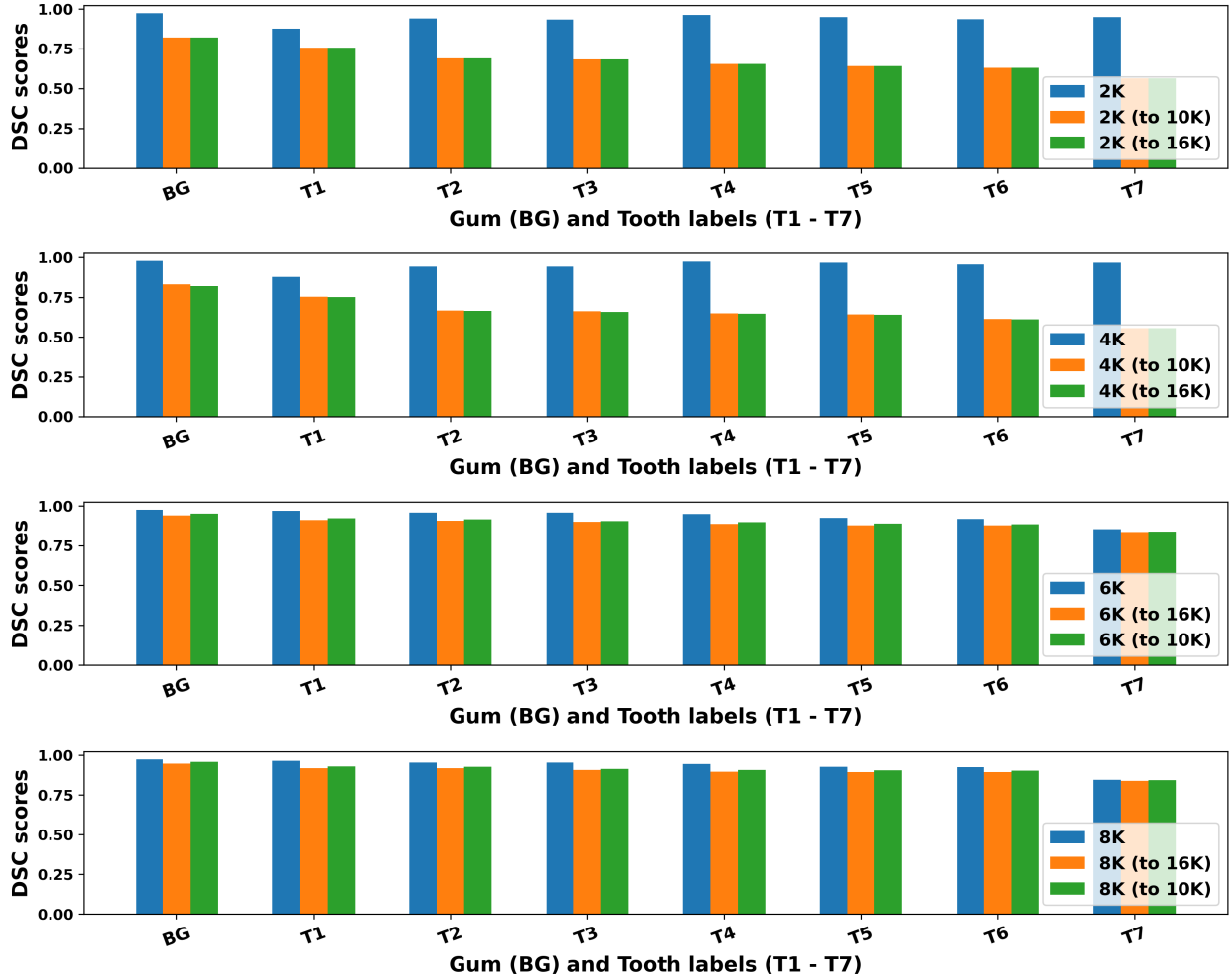


Figure 6: A graphical visualization of per-class DSC score across different resolutions compared to their upsampled cohorts. Each resolution and corresponding upsampled version is represented by one of the four graphs and DSC scores are compared.

3. Discussion

There is a clear visual difference between the labels of the decimated intraoral scans and the predicted labels from the model trained on that resolution as can be seen from Figures 3, 4. The predicted labels appear to correct for errors present in the original decimated mesh labels at all resolutions. These corrections result in cleaner models with well-defined labels that are more accurately positioned than their decimated counterparts. We also observe that the performance of the lower resolution model when upsampled to 10K and 16K show only slight differences in DSC scores as shown in Fig. 6 and Table 3. Our experiments showed that the lower resolutions, such as 6K, can be still considered for edge devices as the predictions retain performance ~ 90 even when upsampled to 10K or 16K models.

4. Conclusion and Limitations

This study analyzed how intraoral scan resolution affects the performance of deep learning-based tooth segmentation. The results indicate that while models trained on lower resolutions retain reasonable segmentation accuracy, performance degradation becomes significant at resolutions below 6K. Background gingiva segmentation remain stable across resolutions, but fine-grained tooth boundary recognition deteriorates at lower resolutions. Certain tooth classes, particularly premolars and canines, exhibit higher misclassification rates when trained on lower resolution meshes.

We acknowledge that the experiments were limited to a single segmentation architecture, PointMLP. For better generalization, more experiments need to be conducted. Our study focuses on only a particular aspect, which is the trade-off between resolution and segmentation accuracy. Although, the validity of these findings needs to be investigated in clinical settings, we were able to provide a preliminary quantitative assessment of segmentation performance across different intraoral scan resolutions.

References

- [1] A. Jana, A. Maiti, and D. N. Metaxas, "A critical analysis of the limitation of deep learning based 3d dental mesh segmentation methods in segmenting partial scans," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2023, pp. 1–7.
- [2] A. Jana, H. M. Subhash, and D. Metaxas, "Automatic tooth segmentation from 3d dental model using deep learning: A quantitative analysis of what can be learnt from a single 3d dental model," *arXiv preprint arXiv:2209.08132*, 2022.
- [3] F. G. Zanjani, D. A. Moin, B. Verheij, F. Claessen, T. Cherici, T. Tan *et al.*, "Deep learning approach to semantic segmentation in 3d point cloud intra-oral scans of teeth," in

International Conference on Medical Imaging with Deep Learning. PMLR, 2019, pp. 557–571.

- [4] X. Chen, N. Ma, T. Xu, and C. Xu, “Deep learning-based tooth segmentation methods in medical imaging: A review,” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 238, no. 2, pp. 115–131, 2024.
- [5] X. Ma, C. Qin, H. You, H. Ran, and Y. Fu, “Rethinking network design and local geometry in point cloud: A simple residual mlp framework,” *arXiv preprint arXiv:2202.07123*, 2022.
- [6] X. Xu, Y. Liu, G. Wang, Z. Zhao, S. Li, and Z. Zhang, “3d tooth segmentation and labeling using deep convolutional neural networks,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2603–2615, 2018.
- [7] Y. Sun, Y. Wang, Y. Wang, Z. Zhang, and Z. Zhao, “Model adaptive tooth segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2020, pp. 442–450.
- [8] Z. Zhang, Y. Wang, Y. Sun, and Z. Zhao, “Darch: Deep architectures for automated tooth segmentation,” *IEEE Access*, vol. 7, pp. 118 712–118 721, 2019.
- [9] Z. Cui, Z. Li, S. Song, Y. Zhou, S. Zhuang, G. Wei, G. Li, and G. Zheng, “Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3d intraoral scanners,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2019, pp. 837–845.
- [10] Z. Zhang, Y. Wang, Y. Sun, and Z. Zhao, “Robust hybrid learning for automatic teeth segmentation and labeling on 3d dental models,” *IEEE Access*, vol. 8, pp. 4561–4571, 2020.
- [11] Y. Tian, Y. Liu, G. Wang, Z. Zhao, S. Li, and Z. Zhang, “3d dental model segmentation with graph attentional convolution network,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2020, pp. 841–849.
- [12] F. P. Zanjani, A. Sadr, A. Fedorov, S. Kapila, Y. Kapila, B. Paniagua, and M. Styner, “Tsegnet: An efficient and accurate tooth segmentation network on 3d dental model,” *IEEE Access*, vol. 8, pp. 52 307–52 317, 2020.
- [13] T.-H. Wu, C. Lian, S. Lee, Y. Wang, Y. Sun, Z. Zhang, and Z. Zhao, “Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3d intraoral scans,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 1, pp. 123–134, 2023.
- [14] C. Lian, Y. Wang, Y. Sun, Z. Zhang, and Z. Zhao, “Unsupervised pre-training improves tooth segmentation in 3-dimensional intraoral mesh scans,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2315–2326, 2021.
- [15] Z. Zhang, Y. Wang, Y. Sun, and Z. Zhao, “Transformer-based 3d tooth segmentation via point cloud region partition,” *Scientific Reports*, vol. 13, no. 1, pp. 1–12, 2023.

- [16] A. Ben-Hamadou, O. Smaoui, A. Rekik, S. Pujades, E. Boyer, H. Lim, M. Kim, M. Lee, M. Chung, Y.-G. Shin *et al.*, “Automatic 3d teeth semantic segmentation with mesh augmentation network,” *arXiv preprint arXiv:2305.18277*, 2023.
- [17] A. Jana, H. M. Subhash, and D. N. Metaxas, “3d tooth mesh segmentation with simplified mesh cell representation,” *arXiv preprint arXiv:2301.10531*, 2023.
- [18] X. Xu, C. Liu, and Y. Zheng, “3d tooth segmentation and labeling using deep convolutional neural networks,” *IEEE transactions on visualization and computer graphics*, vol. 25, no. 7, pp. 2336–2348, 2018.
- [19] S. Tian, N. Dai, B. Zhang, F. Yuan, Q. Yu, and X. Cheng, “Automatic classification and segmentation of teeth on 3d dental model using hierarchical deep learning networks,” *IEEE Access*, vol. 7, pp. 84 817–84 828, 2019.
- [20] A. Jana, H. M. Subhash, and D. Metaxas, “3d tooth mesh segmentation with simplified mesh cell representation,” in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–5.
- [21] L. Qiu, C. Ye, P. Chen, Y. Liu, X. Han, and S. Cui, “Darch: Dental arch prior-assisted 3d tooth instance segmentation with weak annotations,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 752–20 761.
- [22] Z. Cui, C. Li, N. Chen, G. Wei, R. Chen, Y. Zhou, and W. Wang, “Tsegnet: An efficient and accurate tooth segmentation network on 3d dental model,” *Medical Image Analysis*, vol. 69, p. 101949, 2021.
- [23] A. Ben-Hamadou, O. Smaoui, H. Chaabouni-Chouayakh, A. Rekik, S. Pujades, E. Boyer, J. Strippoli, A. Thollot, H. Setbon, C. Trosset *et al.*, “Teeth3ds: a benchmark for teeth segmentation and labeling from intra-oral 3d scans,” *arXiv preprint arXiv:2210.06094*, 2022.