Coupling Bracket Segmentation and Tooth Surface Reconstruction on 3D Dental Models

Yuwen Tan[1], Xiang Xiang[1(✉)], Yifeng Chen[1], Hongyi Jing[1], Shiyang Ye[2], Chaoran Xue[2], and Hui Xu[2]

[1] Key Lab of Image Processing and Intelligent Control, Ministry of Education
School of Artificial Intelligence and Automation,
Huazhong University of Science and Technology, Wuhan, China
xex@hust.edu.cn

[2] State Key Lab of Oral Diseases; National Clinical Research Center for Oral Diseases; Department of Orthodontics,
West China Hospital of Stomatology, Sichuan University, Chengdu, China

Abstract. Delineating and removing brackets on 3D dental models and then reconstructing the tooth surface can enable orthodontists to premake retainers for patients. It eliminates the waiting time and avoids the change of tooth position. However, it is time-consuming and labor-intensive to process 3D dental models manually. To automate the entire process, accurate bracket segmentation and tooth surface reconstruction algorithms are of high need. In this paper, we propose a graph-based network named BSegNet for bracket segmentation on 3D dental models. The dynamic dilated neighborhood construction and residual connection in the graph network promote the bracket segmentation performance. Then, we propose a simple yet effective projection-based method to reconstruct the tooth surface. We project the vertices of the hole boundary on the tooth surface onto a 2D plane and then triangulate the projected polygon. We evaluate the performance of BSegNet on the bracket segmentation dataset and the results show the superiority of our method. The framework integrating the segmentation and reconstruction achieves a low reconstruction error and can be used as an effective tool to assist orthodontists in orthodontic treatment.

Keywords: 3D dental surface · bracket segmentation · surface reconstruction · deep learning · orthodontic treatment

1 Introduction

With the rapid development of 3D scanning devices, the application of computer-aided diagnosis in orthodontics has gradually developed. In orthodontic treatment, an essential step for patients is to wear retainers. Orthodontists need to

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43987-2_40.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 411–420, 2023.
https://doi.org/10.1007/978-3-031-43987-2_40

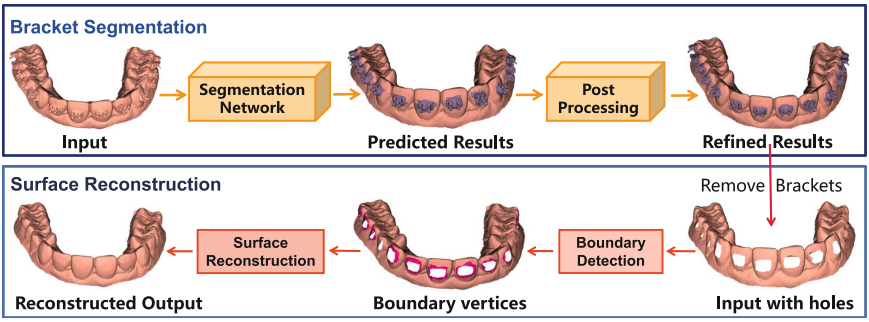

Fig. 1. The framework of bracket segmentation and tooth surface reconstruction. Integrating those two steps into a unified framework automates the whole process.

remove the brackets, utilize the intra-oral scanners to acquire the digital models, and then print models to fabricate retainers. However, there is a long waiting time for patients and the position of teeth would change, which affects the effectiveness of orthodontic treatment. As shown in Fig. 1, removing the brackets on 3D dental models and reconstructing the tooth surface can enable orthodontists to pre-make retainers. Patients can get retainers immediately after removing brackets which reduces the waiting time. When the retainer is lost or broken, a new one can be easily fabricated by using the archived digital model. The whole process helps to maintain the long-term stability of orthodontic treatment. However, it takes approximately 20 min for an orthodontist to precisely delineate brackets on a digital model. Besides that, the bracket segmentation requires a very high level of precision otherwise it will affect the reconstruction step. Therefore, an efficient and precise bracket segmentation method is crucial in orthodontic treatment. Orthodontists usually use CAD software (*i.e.*, Geomagic Studio) to reconstruct the tooth surface after removing the brackets due to the clinical applicability of its reconstructed results. However, it requires interactive manual operation and automatic segmentation algorithms cannot be deployed to the software. To automate the entire process, a 3D mesh reconstruction algorithm is also needed. Integrating segmentation and reconstruction into a unified framework can serve as an effective tool to assist orthodontic treatment.

Since 3D dental models can be transformed into point clouds, methods proposed for point cloud segmentation would provide guidance. Point cloud segmentation methods can be mainly summarized as MLP-based [10,12,13], graph-based [7,14], convolution-based [16], and attention-based [4]. Besides that, several methods have been proposed for computer-aided orthodontic processes such as tooth segmentation [3,9,17], landmark localization [5,6,15], and tooth completion [11,19] on 3D dental models. A two-stream graph-based network TSGCNet [17] is proposed for tooth segmentation. ToothCR [19] is proposed to recover the missing tooth which consists of point completion and surface reconstruction. The holes formed by the removal of the brackets are simple and regular polygons located on each tooth surface. An effective and fast method should be proposed

to fill these characteristic holes and reconstruct the tooth surface. Different from the network [8] proposed for single-tooth bracket separation, our work focuses on the more challenging task of segmenting entire dental models. To the best of our knowledge, there exists no work proposed to integrate the segmentation and reconstruction of 3D dental models to automate the overall process.

As graph-based network [6,9,15,17,18] shows its superiority on various tasks on 3D dental models, we analyze the performance of different local operations and modules in the graph network. Based on these analyses, we propose a network named BSegNet for bracket segmentation. After segmenting the brackets, we adopt a simple yet effective projection-based method to reconstruct the tooth surface, which converts 3D holes into a 2D plane and triangulates the projected polygons on the 2D plane. We then estimate the z coordinate of the new vertices through the neighbor information and transform the vertices to the original space. The proposed method can better recover the surface of teeth.

The contributions of this paper are as follows: 1) We propose a graph-based network named BSegNet for bracket segmentation on 3D dental models which can reduce the burden on orthodontists; 2) An effective method is proposed to reconstruct the tooth surface where brackets are removed; 3) The low reconstruction error of the automatically processed models suggests that the framework integrating the segmentation and reconstruction can be used as a powerful tool to assist orthodontists.

2 Method

We propose a network named BSegNet for bracket segmentation on 3D dental models. BSegNet regards each mesh cell as a graph node and updates the node-wise feature via several local modules. For dental models with brackets removed, a simple yet effective reconstruction method is adopted to reconstruct the tooth surface. We will describe the network architecture of BSegNet (Fig. 2) and the process of tooth surface reconstruction in detail.

2.1 Bracket Segmentation Network

Compared to point clouds, more geometric spatial information can be obtained from 3D mesh data. Different input features will have different effects on the subsequent network training. In this paper, we use a 24-dimension vector as the initial input. The 24-dimensional input vector corresponds to the coordinates of three vertices, the normal vectors of three vertices, the normal vectors of the mesh cell, and the coordinates of the mesh cell centroid.

We use a MLP to map the input feature $F_0 \in \mathbb{R}^{24}$ into $F_1 \in \mathbb{R}^{64}$. Then a transform net is adopted to transform the feature F_1 into a canonical space to improve the robustness. The transformed feature is fed to several local modules which are designed to encode the local features of each mesh cell in semantic space. In each local module, the first step is to construct the graph $\mathcal{G}(V, E)$ where $V = \{c_1, c_2, ..., c_N\}$ denotes the set of mesh cells and E represents the

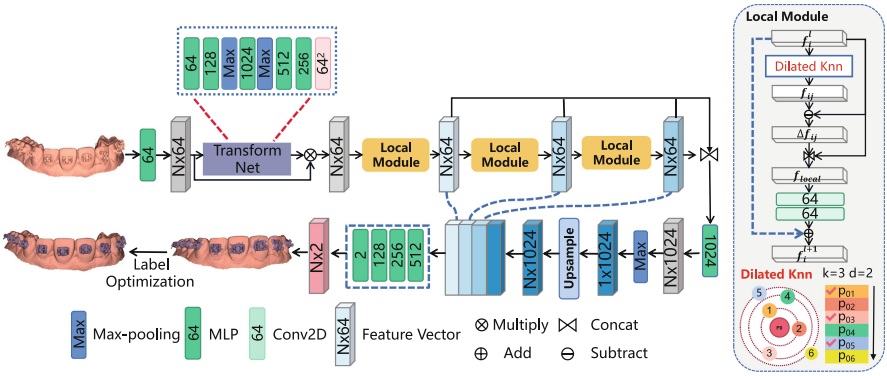

Fig. 2. The architecture of the bracket segmentation network (BSegNet).

connectivity relationship. To construct the $E \in |V| \times |V|$, we use the *dilated knn* algorithm to define the connectivity relationship where a mesh cell only connects to its nearest k neighbors. For a given mesh cell c_i, we calculate the feature distance between other mesh cells and select the k nearest neighbors within the range of the $k \times d$ neighbors, where d is the dilation factor to determine the dilation range. *Dilated knn* can get a large receptive field than *knn*. Let $\mathcal{N}_i$ denotes the set of neighbors, $D_s = \{c_1, c_2, ..., c_{k\times d}\}$ are the nearest $k \times d$ cells (except c_i). The dilated k neighbors of c_i is defined as

$$\mathcal{N}_i^k = \{c_1, c_{1+d}, c_{1+2\times d}, ...c_{1+(k-1)\times d}\} \tag{1}$$

After building the graph, we construct the local feature of each cell and update the feature through the graph convolution layers. Let the $f_i^l \in \mathbb{R}^{64}$ denote the feature vector of the mesh cell c_i in the l-th layer and $\triangle f_{ij} = \{f_l^i - f_l^j\}|_{j\in\mathcal{N}_i^k}$ denotes the edge feature which is used to capture the geometric relationship between each cell and its neighbors. Then the local feature $f_{local} = (\triangle f_{ij}^l \oplus f_i^l)$ goes through two Conv2D layers to further encode the feature. The update process of the mesh cell features (graph node) is defined as

$$f_i^{l+1} = R(h_\vartheta(\triangle f_{ij}^l \oplus f_i^l))|_{j\in\mathcal{N}_i^k} + f_i^l \tag{2}$$

where h_ϑ denotes the Conv2D layers and $R(.)$ stands for the feature aggregation function. To avoid the over-smoothing problem in the graph network, we use the residual connection in each local module the same as DeepGCN [7]. We concatenate the features of each local module and use an MLP layer to form a global high-dimension feature. Then the high-dimension global feature is concatenated with local features, forming the node-wise feature. Finally, several projection layers and one classifier layer are used to predict an N $\times$2 probability matrix.

As the prediction results of the network may have some isolated labeled mesh cells, we use the graph-cut method to refine the results. The post-processing stage minimizes an energy function by combining the probability term and the

smoothness term. The energy function to be optimized is defined as

$$E = \sum_{i=1}^{N} -log(max(p_i(l_i), \epsilon)) + \lambda \sum_{i}^{N} \sum_{j \in \mathcal{N}_i} S(p_i, p_j, l_i, l_j) \tag{3}$$

where $p_i(l_i)$ denotes the probability belongs to the l_i, ϵ is the minimal probability threshold, and λ denotes the smooth parameter. The local smoothness term is defined as

$$S(p_i, p_j, l_i, l_j) = \begin{cases} 0, & l_i = l_j \\ -log(\frac{\theta_{ij}}{\pi}) d_{ij}, & l_i \neq l_j \end{cases} \tag{4}$$

where θ_{ij} denotes the dihedral angle of two adjacent facets and d_{ij} denotes the distance between the centroids of two adjacent facets.

2.2 Tooth Surface Reconstruction

First, we identify all the holes and extract their boundaries. Then we project the vertices of the boundary into the 2D plane. We triangulate the projected polygon without inserting new vertices. If the line segment inside the polygon exceeds the preset length, it should be n-equally divided. To get more uniform triangles, we use the optimal delaunay triangulation algorithm [1] to iteratively optimize the position of vertices. The optimization process of vertices is as follows

$$p^* = \frac{1}{\sum |T_j|} \sum_{T_j \in \Omega(p)} |T_j| c_j \tag{5}$$

where p is the vertex needs to be optimized, $\Omega(p)$ denotes the set of first-ring neighborhood mesh cells of p, $|T_j|$ denotes the area of mesh cell T_j, c_j is the circumcentre of T_j. After optimizing the positions of the vertices, the triangles have almost the same angles and the distribution is closer to the original models.

After the polygon triangulation, the x-y coordinates are determined and a layer-by-layer procedure is employed to estimate the corresponding z-values. To avoid the vertices at the gingiva, we only use the information of the first-ring neighborhoods of the boundary vertices. We first compute the slopes of the boundary vertices which are defined as

$$k_i^b = \frac{1}{|\mathcal{N}^1(i)|} \sum_{j \in \mathcal{N}^1(i)} \frac{\triangle z_{ij}}{\sqrt{\triangle x_{ij}^2 + \triangle y_{ij}^2}} \tag{6}$$

where $\mathcal{N}^1(i)$ denotes the set of first-ring neighborhood vertices of the boundary point b_i. Then the calculation of the z-coordinate value is denoted as

$$z_{new}^i = \frac{1}{\mathcal{N}^b(i)} \sum_{j \in \mathcal{N}^b(i)} (z_j^b + k_j^b \sqrt{\triangle x_{ij}^2 + \triangle y_{ij}^2}) \tag{7}$$

where $\mathcal{N}^b(i)$ denotes the set of adjacent boundary points and k_j^b denotes the slope of boundary vertex b_j. Then we regard the added points as new boundary vertices and repeat the above process until the z-values of all vertices are calculated.

Before transforming back to the original space, the extreme z values are removed by median filtering. Then we use a rotation matrix to obtain the coordinates in the original space. Finally, we employ Laplacian smoothing to enhance the smoothness of the reconstructed surface. To make the reconstructed surface blend better with the boundary, we need to reduce the effect of smoothing on the first-ring neighborhood of the boundary. The calculation equation is as follows

$$p_{new} = up_b + (1-u)p_s, \quad u = \frac{1}{1 + exp(-\frac{d_h}{d_b})} \tag{8}$$

where p_b and p_s denote the vertex before and after smoothing, d_b is the average distance between the vertex and the adjacent boundary vertices, and d_h is the average length of the hole boundary.

3 Experiments and Results

3.1 Datasets and Implementation Details

We collect 80 dental mesh models in STL format from different patients. The number of mesh cells in each dental model is approximately 100,000 and all the dental models are down-sampled to nearly 24,000 mesh cells. The ground truth segmentations are annotated by professional orthodontists on the down-sampled dental models. We divide the dataset into a training, validation, and test set which consists of 45, 14, and 21 subjects, respectively. The performance of bracket segmentation is evaluated by mean Intersection-over-Union (mIoU) and Overall Accuracy (OA). The performance of reconstruction is evaluated by the Mean Distance (MD) and Standard Deviation (SD) of the distance between the models reconstructed by our method and by Geomagic Studio. We also evaluate the reconstruction error of models processed by the automatic framework and manually by orthodontists. We train all the networks by minimizing the cross-entropy loss for 400 epochs except for MeshSegNet which minimizes the dice loss. We use the Adam optimizer and set the mini-batch as 5. The initial learning rate is 0.001, and we anneal the learning rate using the cosine functions.

Table 1. The segmentation results of the test dataset in terms of both OA and mIoU. (w/p) denotes the segmentation results after label optimization. **Bold** is the best.

Method	Input	OA	mIoU	Bracket	Other	OA(w/p)	mIoU(w/p)
PointNet [12]	4p,4n	90.21	81.78	78.80	84.75	91.83	84.52
PointNet++ [13]	4p,4n	94.44	89.26	87.78	90.74	95.87	91.86
DGCNN [14]	4p,4n	96.14	92.42	91.42	93.41	97.47	94.96
PointConv [16]	4p,4n	95.30	90.83	89.65	92.00	96.81	93.66
PCT [4]	4p,4n	96.44	92.99	92.13	93.85	97.56	95.13
PointMLP [10]	4p,4n	96.51	93.15	92.29	94.00	97.60	95.23
MeshSegNet [9]	4p,1n	95.62	91.46	90.39	92.52	96.89	93.84
TSGCNet [17]	4p,4n	93.00	86.68	84.96	88.41	95.00	90.30
Ours	4p,4n	**97.28**	**94.60**	**93.95**	**95.26**	**98.13**	**96.25**

3.2 Experimental Evaluation

Comparison Results. We compare our method with SOTA point cloud segmentation methods and teeth segmentation methods. All the results are shown in Table 1 and BSegNet achieves the best performance. Since PointNet lacks the feature encoding of local regions, it performs worst among all the methods. The performance of PointNet++ and PointConv are close and outperform PointNet by a large margin. Although PointMLP has the best performance among all the compared methods, the mIoU of BSegNet is higher (94.60 vs. 93.15). We also compare our method with the attention-based method PCT and the mIoU of our method is higher than PCT (94.60 vs. 92.99). The tooth segmentation network on 3D dental models performs worse than several point segmentation methods in the bracket segmentation task, especially for TSGCNet which uses a two-stream network to encode the coordinates and normal vectors respectively. MeshSegNet performs better than TSGCNet but the proposed graph-constrained learning modules cause high computation complexity. Our method is based on the DGCNN but the mIoU of our method is much higher (94.60 vs. 92.42).

As shown in Table 2, the reconstruction error of the models predicted by BSegNet is significantly lower than PointMLP. When reconstructing the models processed by doctors, our reconstruction method achieves low values of SD and MD which reveals it can replace the interactive reconstruction operation to some extent. We also compare our method with another reconstruction method Meshfix and our method has a lower value of SD (0.032 vs. 0.049). Compared to the manual process by doctors, the reconstruction error of our automatic framework is clinically acceptable and it can assist in orthodontic treatment. Figure 3 displays the segmentation and tooth surface reconstruction results.

Ablation Study and Analysis. In this section, we analyze different operations and modules in the BSegNet. As shown in Table 3, the performance of the 24-dimension input is better than the 15-dimension input which suggests the

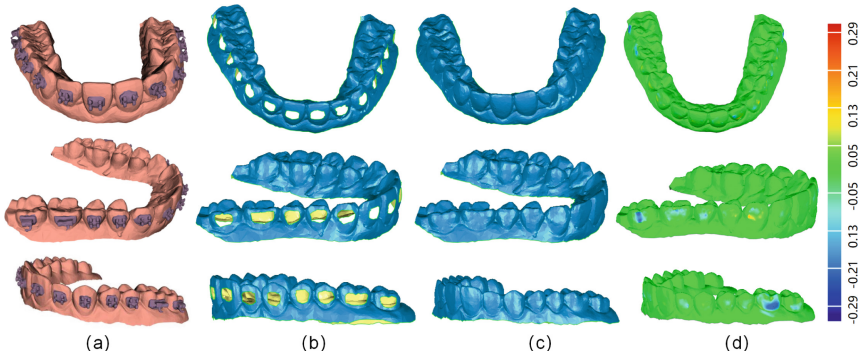

Fig. 3. (a) The bracket segmentation results after post-processing; (b) Dental models with brackets removed; (c) Models after tooth surface reconstruction; (d) Color-coded error map of automatically processed models compared with ground truth.

Table 2. The tooth surface reconstruction results compared to the ground truth in terms of MD(mm) and SD(mm).

Method	**Seg.**	**Recon.**	SD(mm)	MD+(mm)	MD-(mm)	Aver.(mm)
Semi.	PointMLP [10]	Geo	0.060	0.014	0.005	0.0095
Semi.	BSegNet	Geo	0.026	0.004	0.005	0.0045
Semi.	Doctors	Ours	0.015	0.001	0.004	0.0025
Auto.	BSegNet	Meshfix [2]	0.049	0.001	0.018	0.0095
Auto.	BSegNet	Ours	0.032	0.002	0.009	0.0055

normal vectors of the vertices help the subsequent network training. However, curvature information which is helpful for the teeth segmentation task does not improve and even hurts the bracket segmentation performance. An essential step in the graph-based network is to construct the local features of the center node. Many methods have been proposed to construct complex local features. However, our results show that using the edge feature with the central feature is sufficient for effectively representing local regions. We also analyze the results of different aggregation operations, and the results indicate that Max pooling is the most effective method for aggregation. As shown in Table 4, adopting the dynamic graph which updates the neighbor information in each local module, the mIoU has an improvement (+0.88) compared to the fixed graph. Under the same network structure, using the residual connection would further improve model performance by (+0.70). The use of *dilated knn* only brings slight performance improvement but does not require extra computation costs. With the post-processing, the predicted results can be further refined. Overall, using all the operations mentioned above acquires the best performance with 94.60 of mIoU and 96.25 after post-processing.

Table 3. Analysis experiments of different inputs, feature aggregation functions, and local feature constructions.

Input	OA	mIoU	R(.)	OA	mIoU	Feature	OA	mIoU
1p,1n	95.93	92.02	Att	95.40	91.03	$\{\triangle p_{ij}, f_j\}$	95.02	90.31
4p,1n	96.74	93.55	Max	**97.12**	**94.29**	$\{\triangle f_{ij}, f_i\}$	**97.12**	**94.29**
4p,4n	**97.12**	**94.29**	Sum	96.72	93.52	$\{\triangle f_{ij}, f_i, f_j\}$	96.56	93.20
4p,1n,3c	95.50	91.26	Mean	96.72	93.51	$\{\triangle p_{ij}, \triangle f_{ij}, f_i, f_j\}$	96.75	93.57

Table 4. Ablation study of different operators in the BSegNet.

Dynamic	Residual Con.	Dilated-knn	OA	mIoU	OA(w/p)	mIoU(w/p)
			96.28	92.71	97.55	95.14
✓			96.76	93.59	97.79	95.58
✓	✓		97.12	94.29	97.99	95.97
✓	✓	✓	**97.28**	**94.60**	**98.13**	**96.25**

4 Conclusion

In this paper, we propose a network named BSegNet for bracket segmentation on 3D dental models which can reduce the burden on orthodontists. BSegNet is a graph-based network that employs dynamic dilated neighborhood construction and residual connections to improve segmentation results. With label optimization, the segmentation results can be further refined. Experimental results on a clinical dataset demonstrate our method significantly outperforms related state-of-the-art methods. We also propose a simple yet effective method to reconstruct the tooth surface which can better recover the feature of the teeth. The whole framework achieves a low reconstruction error and can be used as a powerful tool to assist doctors in orthodontic diagnosis.

Acknowledgement. This research was supported by Sichuan Univ. Interdisciplinary Innovation Res. Fund (RD-03-202108), Natural Science Fund of Hubei Province (2022CFB823), HUST Independent Inno. Res. Fund (2021XXJS096), Alibaba Innovation Research (AIR) program (CRAQ7WHZ11220001-20978282), and grants from MoE Key Lab of Image Processing and Intelligent Control.

References

1. Alliez, P., Cohen-Steiner, D., Yvinec, M., Desbrun, M.: Variational tetrahedral meshing. In: ACM SIGGRAPH 2005 Papers, pp. 617–625 (2005)
2. Attene, M.: A lightweight approach to repairing digitized polygon meshes. Vis. Comput. **26**, 1393–1406 (2010)
3. Cui, Z., et al.: A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nat. Commun. **13**(1), 2096 (2022)

4. Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., Hu, S.M.: PCT: point cloud transformer. Comput. Vis. Media **7**, 187–199 (2021)
5. Lang, Y., et al.: DentalPointNet: landmark localization on high-resolution 3D digital dental models. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science, vol. 13432, pp. 444–452. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16434-7_43
6. Lang, Y., et al.: DLLNet: an attention-based deep learning method for dental landmark localization on high-resolution 3D digital dental models. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12904, pp. 478–487. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87202-1_46
7. Li, G., Muller, M., Thabet, A., Ghanem, B.: DeepGCNs: can GCNs go as deep as CNNs? In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9267–9276 (2019)
8. Li, R., et al.: Deep learning for separation and feature extraction of bonded teeth: tool establishment and application (2022)
9. Lian, C., et al.: Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3D intraoral scanners. IEEE Trans. Med. Imaging **39**(7), 2440–2450 (2020)
10. Ma, X., Qin, C., You, H., Ran, H., Fu, Y.: Rethinking network design and local geometry in point cloud: a simple residual MLP framework. arXiv preprint arXiv:2202.07123 (2022)
11. Ping, Y., Wei, G., Yang, L., Cui, Z., Wang, W.: Self-attention implicit function networks for 3D dental data completion. Comput. Aided Geom. Des. **90**, 102026 (2021)
12. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652–660 (2017)
13. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
14. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. **38**(5), 1–12 (2019)
15. Wu, T.H., et al.: Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3D intraoral scans. IEEE Trans. Med. Imaging **41**(11), 3158–3166 (2022)
16. Wu, W., Qi, Z., Fuxin, L.: PointConv: deep convolutional networks on 3D point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9621–9630 (2019)
17. Zhang, L., et al.: TSGCNet: discriminative geometric feature learning with two-stream graph convolutional network for 3D dental model segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6699–6708 (2021)
18. Zheng, Y., Chen, B., Shen, Y., Shen, K.: TeethGNN: semantic 3D teeth segmentation with graph neural networks. IEEE Trans. Vis. Comput. Graph. **29**(7), 3158–3168 (2022)
19. Zhu, H., Jia, X., Zhang, C., Liu, T.: ToothCR: a two-stage completion and reconstruction approach on 3D dental model. In: Gama, J., Li, T., Yu, Y., Chen, E., Zheng, Y., Teng, F. (eds.) Advances in Knowledge Discovery and Data Mining. PAKDD 2022. Lecture Notes in Computer Science, vol. 13282, pp. 161–172. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-05981-0_13