End-to-End 3D Tooth Landmark Detection with Fuzzy Tooth Localization

Kaibo Shi[1], Hairong Jin[1,2], and Youyi Zheng[1](✉)

[1] State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China
22421210,hairong_jin,youyizheng@zju.edu.cn
[2] Computer and Computing Science, Hangzhou City University, Hangzhou, China

Abstract. Accurate detection of tooth landmarks is crucial for computer-aided orthodontic treatment. Previous methods often employ segmentation to isolate individual teeth, but rely heavily on segmentation accuracy and require annotated data. In this paper, we introduce a two-stage framework for tooth localization and landmark detection, eliminating the need for segmentation based on mesh deep learning. First, we define the fuzzy tooth regions based on landmark positions. Binary masks are generated for the tooth regions located from the original jaw mesh. By combining local features of individual teeth with the global features of the jaw model, our method predicts multiple heatmaps and the corresponding probabilities of potential landmarks for each tooth. Finally, we design a bipartite matching loss for both tooth localization and landmark detection to align the prediction set with the ground truth, thereby facilitating end-to-end inference throughout the entire process. Experimental results on the Teeth3DS+ dataset demonstrate that our method effectively detects a variable number of landmarks. Furthermore, it significantly outperforms existing baseline methods, exhibiting robust generalization and superior performance.
(The code will be released at https://github.com/sikingbo/ToothLDNet.)

Keywords: Tooth Landmark Detection · Tooth Localization · Two-stage · End-to-End · Multiple Heatmaps.

1 Introduction

In digital orthodontics, the automatic and accurate detection of tooth landmarks in 3D jaw models is crucial to helping orthodontists perform dental evaluations and plan treatments [20,26]. Traditionally, this process required manual segmentation and landmark marking by clinicians, which is time-consuming and laborious. The introduction of deep learning [3,6,12,14,10,17,24] has significantly advanced the development of semiautomatic and fully automatic methods for landmark detection. However, accurately detecting tooth landmarks remains challenging due to the complex geometry of jaw model surfaces, the differences among teeth, and the occasional presence of abnormal or missing teeth in certain patients. Consequently, many existing methods [22,25] employ

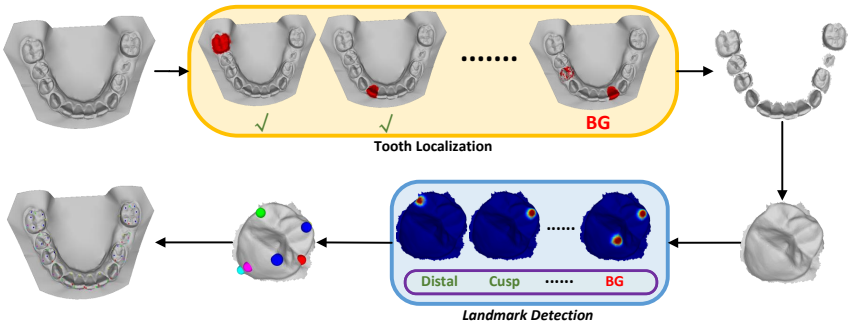

Fig. 1. The overall pipeline of ToothLDNet. **Top**: tooth localization performs multiple binary classification tasks, each corresponding to a tooth, and filters the 'background' (BG) masks. **Bottom**: Our method predicts multiple heatmaps with corresponding probabilities, filtering out the 'background' (BG) heatmaps.

tooth segmentation [8,11,27] as a preprocessing step for precise landmark detection.However, this approach necessitates segmentation labels for model training, which is typically a laborious and time-consuming process. Therefore, we investigate whether alternative strategies can achieve accurate landmark detection without the need for segmentation labels. Moreover, most existing landmark detection methods [25] generate a correspondence heatmap for each landmark within a segmented tooth. This approach limits the model's output to predicting a fixed number of landmarks, thereby restricting its generalizability. Some methods [22] attempt to extract a variable number of landmarks by predicting a shared heatmap, but they often rely on post-processing techniques such as clustering [16]. Furthermore, these approaches tend to analyze individual teeth in isolation, thereby neglecting the global features of the entire jaw model.

To tackle these challenges, we introduce a two-stage framework, ToothLDNet, for tooth localization and landmark detection. ToothLDNet provides an end-to-end solution that can be applied directly to jaw models without segmentation. Initially, we observe that tooth landmarks can roughly indicate the positions of teeth. Based on these landmark locations, we calculate the bounding box to generate its localization label for each tooth. The first stage features a tooth localization network that estimates binary masks and probabilities for potential teeth, and extracts global features of jaw models for the next stage. The second stage landmark detection network extracts local features from individual teeth and combines them with global features. The fused features are then decoded to generate multiple heatmaps and probabilities for potential landmarks. We employ bipartite matching loss [2,7,13,18] for tooth localization and landmark detection networks. By comparing the predicted sets with the ground truth, these loss functions ensure that 'background' predictions are filtered out during inference, allowing the model to predict a variable number of targets end-to-end.

In summary, our contributions are as follows:

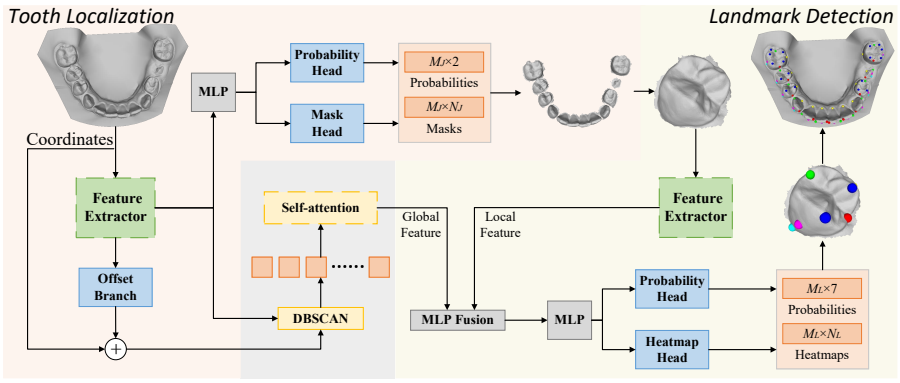

Fig. 2. The network architecture of ToothLDNet.

- We introduce ToothLDNet, a two-stage framework for accurately and efficiently detecting 3D tooth landmarks without relying on segmentation labels.
- We introduce a transformer-based network that extracts local geometric features and employs self-attention to capture global sequence features of teeth along the dental arch curve.
- We validated our method on the Teeth3DS+ dataset, achieving robust performance across a range of challenging dental cases and demonstrating excellent results.

2 Method

In this section, we provide a detailed overview of our method.ToothLDNet consists of two primary stages: tooth localization and landmark detection, as illustrated in Figure 2. We adopt a similar network architecture for both tooth localization and landmark detection. In addition to predicting the masks or heatmaps that highlight the locations of the teeth or landmarks, the network also generates corresponding probabilities to indicate the presence of each tooth or landmark. To ensure end-to-end inference, we design the bipartite matching loss to train both tooth localization and landmark detection networks.

2.1 Data Processing

Given a jaw model, we first determine the approximate region of each tooth based on the ground truth landmarks. As shown in Figure 3, a bounding box for tooth is generated using the Mesial, Distal, Inner, Outer, and Cusp points, which is then used to create a binary mask. Next, based on these masks, we calculate the offsets o from all face centroids to the corresponding tooth centroid, while setting the offsets for gingiva faces to zero vectors [28]. Furthermore, we crop tooth regions and calculate the geodesic distances from the face centroids to the landmarks. Finally, a Gaussian kernel is applied to generate the ground truth heatmaps [7].

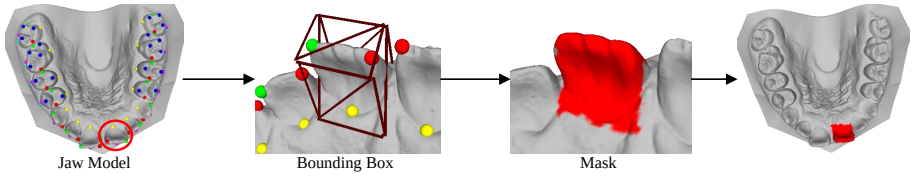

Fig. 3. The ground truth mask based on tooth landmarks. For these landmarks: green and red represent the Distal and Mesial points, yellow and cyan denote the Inner and Outer points, and magenta and blue indicate the Facial Axis and Cusp points.

2.2 Network Detail

Tooth Localization Existing high-accuracy 3D tooth landmark detection methods typically rely on pre-segmentation to crop individual teeth from the 3D jaw model, simplifying the landmark detection task. This pipeline necessitates segmentation annotations. However, we found that high-accuracy tooth segmentation does not significantly enhance landmark detection performance, prompting us to question whether segmentation is truly necessary as a preprocessing step for landmark detection on jaw models. In light of this, we observe that the landmark locations on the jaw model are sufficient to identify the approximate regions of each tooth. Therefore, we generate the ground truth for the tooth localization task based on landmarks. This approach only requires ensuring a sufficient recall rate for tooth localization in the jaw model, making it less sensitive to boundary issues.

Specifically, the high-resolution original jaw model is initially simplified using a mesh simplification algorithm [5], decreasing the number of faces from approximately 100,000 to around 10,000. From the simplified model with N_J faces, we extract face-wise feature vectors $X_J \in \mathbb{R}^{N_J \times 15}$ as input for the tooth localization network. X_J include face centroid coordinates, normals, and vectors from three vertices to the centroid. We then utilize the feature extractor from TeethGNN [28] to extract geometric features $F_J \in \mathbb{R}^{N_J \times D}$. Following this, we design an offset decoder branch to predict face-wise offsets $\hat{o}$. The offset branch consists of two MLP layers and a Conv1D layer, and is supervised by the ground truth offsets. Inspired by KeypointDETR [7], the query-wise decoder uses an MLP to decode F_J into M_J-dimensional representation, which are then transposed to obtain the mask features $F_m \in \mathbb{R}^{M_J \times N_J}$, where M_J is the preset number of queries for tooth localization. The mask features F_m are subsequently passed through the mask and probability heads to generate potential tooth mask and the corresponding probability. During inference, 'background' predictions are filtered based on probabilities, enabling end-to-end tooth localization. Ultimately, the results are projected back onto the original mesh using k-NN ($k = 1$), and individual tooth regions are cropped from the jaw model.

Global Feature Encoder Different types of teeth vary significantly in scale and shape, which results in substantial differences in the positions and local

geometric features of their landmarks. However, our pipeline does not provide semantic labels about the teeth. Given this consideration, and since the teeth are arranged along the dental arch curve, we use the self-attention mechanism to encode each tooth's potential semantic information based on its relative position along the arch, while also capturing its global features within the jaw model.

Given the face-wise geometric features F_J, centroid coordinates, and predicted offsets, we first add the centroids and offsets to obtain the shifted points P. We then apply DBSCAN [4] to P to identify tooth clusters, resulting in sets of tooth cluster features $\{F_J^{(i)}\}^T$ and corresponding cluster points $\{P^{(i)}\}^T$, where T represents the number of tooth clusters. Based on these sets, we compute the centroid coordinates $C_S \in \mathbb{R}^{T\times 3}$ and the max-pooled features $F_S \in \mathbb{R}^{T\times D}$ for tooth clusters. Finally, the tooth cluster features F_S and centroid coordinates C_S are fed into the self-attention block to encode the global features $F_G = \{f_G^{(i)}\}^T \in \mathbb{R}^{T\times D}$:

$$F_G = \text{Softmax}\left(\frac{\left([F_S + \text{MLP}(C_S)]\, W_Q\right)\left(F_S W_K\right)^\top}{\sqrt{D_k}}\right)[F_S + \text{MLP}(C_S)]\, W_V, \quad (1)$$

where W_Q,W_K, W_V are weight matrices, and D_k is the dimensionality of the keys in self-attention. The position encoding is derived from C_S through an MLP.

Landmarks Detection Similar to the tooth localization network, given the i-th tooth region cropped from the jaw model, we first use the TeethGNN encoder [28] to extract geometric features $F_L \in \mathbb{R}^{N_L\times D}$ from input features $X_L \in \mathbb{R}^{N_L\times 15}$ of this tooth region, where N_L represents the number of faces. Since F_L contains only local information within tooth region, it is concatenated with the global features $f_G^{(i)}$, resulting in the fused heatmap features:

$$F_h = F_L \oplus f_G^{(i)}, \quad (2)$$

where global features $f_G^{(i)}$ originate from the tooth cluster with the highest IoU in the i-th tooth region. Next, we decode the heatmap features F_h into M_L dimension using two MLP layers, followed by transposing to obtain $F_h \in \mathbb{R}^{M_L\times N_L}$, where M_L is the preset number of queries for tooth landmark detection. F_h is then passed through separate heads for heatmap and probability prediction, generating both the heatmap and corresponding probability. During inference, we filter out the 'background' heatmaps based on the predicted probabilities and extract the landmarks with the highest heat values from the predicted heatmaps.

2.3 Loss Function

The overall loss consists of three parts: tooth localization loss $\mathcal{L}_J$, landmark detection loss $\mathcal{L}_L$, and offset loss $\mathcal{L}_O$:

$$\mathcal{L} = \mathcal{L}_J + \mathcal{L}_L + \lambda_o \mathcal{L}_O \quad (3)$$

$$\mathcal{L}_O = \frac{1}{N_J} \sum \|o_i - \hat{o}_i\| \tag{4}$$

where o and $\hat{o}$ are ground truth and predicted offsets. $\mathcal{L}_J$ and $\mathcal{L}_L$ are designed as bipartite matching loss. The loss starts by constructing a cost matrix $\{c_{i,j}\} \in \mathbb{R}^{N \times M}$ that defines the relationship between ground truth and prediction sets:

$$c_{i,j} = \lambda_m \cdot \sum \|m_i - \hat{m}_j\| + \lambda_p \cdot (1 - \log \hat{p}_j)\,, \tag{5}$$

where m_i represents the i-th ground truth map, and $\hat{m}_j$ signifies the j-th prediction map, which means mask for tooth localization and heatmap for landmark detection; $\hat{p}_j$ denotes the corresponding probability of prediction map.

With the Hungarian algorithm [9], we determine the optimal bipartite matching between prediction sets and ground truth. Only predictions that match the ground truth can calculate map loss and probability loss, otherwise only probability loss is calculated. We use cross-entropy for probability loss and MSE loss for predicted map loss. For tooth localization, ground truth p_i of 0 or 1 indicates whether the predicted mask matches the ground truth mask. For landmark detection, p_i represents the corresponding landmark category. The bipartite matching loss $\mathcal{L}_b$ is computed as follows:

$$\mathcal{L}_b = \sum_{i=1}^{M} \left(-\lambda_{ce} p_i \log(\hat{p}_i) + \lambda_{mse} \mathbb{1}_{\{p_i \neq 0\}} \frac{1}{N} \sum \|m_i - \hat{m}_i\|^2 \right). \tag{6}$$

3 Experiment

3.1 Implementation Details and Dataset

All experiments are conducted on an RTX 2080Ti GPU. Both the tooth localization network and the landmark detection network are trained for 50 epochs. The learning rate is set at 0.001 for tooth localization and 0.0001 for landmark detection. The number of queries, M_J and M_L, are both preset to 50. In the data preparation phase for heatmap generation, we set $\sigma = 0.75$. In the cost matrix, λ_m and λ_p are assigned values of 1 and 100, respectively. For the loss function, the coefficients λ_{ce} and λ_{mse} are set to 1 and 5 for tooth localization, and 1 and 30 for landmark detection. Additionally, λ_O is set to 0.5. Additionally, We conduct experiments on the publicly available large-scale benchmark dataset, Teeth3DS+ [1],

3.2 Metrics

We evaluate the performance of 3D landmark detection by calculating the average precision (AP) and average recall (AR) across different types of landmarks. Additionally, landmark localization accuracy is evaluated using Chamfer Distance (CD) between the predicted set and the ground truth set [7].

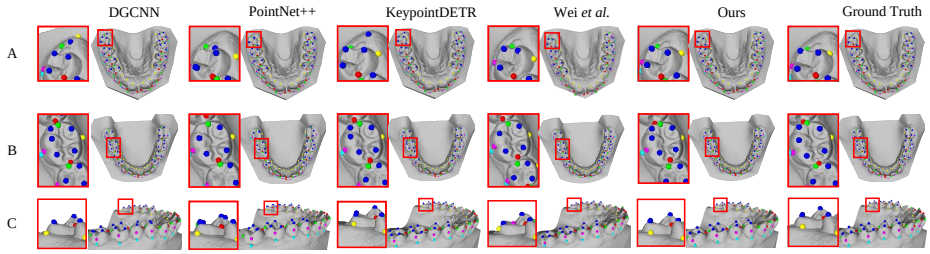

Fig. 4. Qualitative comparison of a upper jaw model (case A), a lower jaw model (case B), and a jaw model with the side view (case C). The red boxes highlight the landmark positions where our method significantly outperforms the comparison methods.

Table 1. Results of AP(%), AR(%) and CD for landmark detection.

	Cusp		Facial Axis		Inner&Ouuter		Mesial&Distal		All
	AP	AR	AP	AR	AP	AR	AP	AR	CD
PointNet++ [15]	56.4	48.6	69.2	53.7	56.2	43.2	60.0	48.0	2.32
DGCNN [21]	68.3	56.2	69.6	54.4	57.5	45.1	69.6	56.8	1.82
Wei *et al.* [23]	69.5	53.1	70.4	55.5	66.5	55.5	69.3	55.7	1.53
KeypointDETR [7]	69.7	56.7	72.2	58.3	70.1	57.5	70.7	**58.1**	1.46
Ours	**69.7**	**57.5**	**72.3**	**58.5**	**70.4**	**57.8**	**71.2**	57.6	**1.44**

3.3 Results

To ensure a fair comparison, we consistently use our tooth localization network to crop individual teeth from the jaw model, and then apply comparison methods to extract landmarks from these cropped teeth. We compare ToothLDNet with two point cloud feature extraction algorithms (DGCNN [21] and PointNet++ [15]) and two keypoint detection methods (KeypointDETR [7] and Wei *et al.* [23]). In both feature extraction algorithms, the decoder outputs a heatmap for each type of landmark, with cusp landmarks extracted through clustering within the shared heatmap. For evaluation, we calculate the AP and AR for each landmark at thresholds ranging from 0 to 3, with an interval of 0.1. The mean is then computed across all thresholds. Additionally, we use CD to evaluate the performance across all landmarks. The quantitative results are presented in Table 1, where our method outperforms the comparison approaches in terms of accuracy, recall, and localization. As shown in Figure 4, clustering-based post-processing methods (DGCNN [21], PointNet++ [15], Wei *et al.* [23]) inevitably result in extra or missing cusp landmarks. This is mainly due to the fact that clustering methods rely heavily on prior knowledge for hyperparameter tuning.

3.4 Ablation Study

Effects of tooth localization. In our framework, tooth localization serves as the foundation for landmark detection. We compare our method with two

Table 2. Results of AP(%), AR(%) and CD for the variations of our method.

	Cusp		Facial Axis		Inner&Ouuter		Mesial&Distal		All
	AP	AR	AP	AR	AP	AR	AP	AR	CD
DGCNN [21]	67.0	55.6	62.3	48.3	64.2	51.7	69.6	57.0	1.58
TeethGNN [28]	68.8	55.7	63.4	49.8	66.3	52.6	70.5	57.5	1.50
PointNet++ Encoder [15]	57.3	49.3	70.1	54.2	63.4	54.7	60.5	48.8	2.22
DGCNN Encoder [21]	68.6	57.1	72.0	58.2	70.3	57.5	69.9	56.6	1.46
w/o Self-attention [19]	68.4	56.3	71.3	57.4	70.1	57.4	71.0	57.5	1.49
w/o Position Encoding	68.6	57.2	71.7	57.5	70.2	57.6	70.9	57.1	1.48
Ours	**69.7**	**57.5**	**72.3**	**58.5**	**70.4**	**57.8**	**71.2**	**57.6**	**1.44**

segmentation networks, DGCNN [21] and TeethGNN [28], as shown in the first sub-table of Table 2. The experimental results indicate that our framework outperforms tooth segmentation in landmark detection. This is primarily because the fuzzy localization of the teeth allows for greater tolerance in landmark detection, while tooth segmentation requires higher boundary quality and may result in incomplete tooth segmentations.

Effects of encoder. We use the TeethGNN [28] encoder, which consists of static EdgeConv layers, to extract geometric features. To evaluate its effectiveness, we compare it with the DGCNN [21] and PointNet++ [15] encoders. The second sub-table shown in Table 2 confirms that the TeethGNN encoder achieves superior performance in 3D tooth landmark detection.

Effects of self-attention module. The self-attention module is designed to extract global features of the teeth from the jaw model. To evaluate its effectiveness, we conduct experiments with and without the self-attention module. Furthermore, positional encoding is essential in the self-attention mechanism as it encodes the relative positions of teeth along the dental arch, embedding potential semantic information. To assess its impact, we compare results with and without positional encoding. As shown in the third sub-table of Table 2, the results indicate that incorporating positional encoding within the self-attention module significantly enhances the model's ability to capture global dependencies and refine sequential features, ultimately improving tooth landmark detection.

4 Conclusion

This paper introduces ToothLDNet, a framework comprising a tooth localization network and a landmark detection network. The tooth localization network crops the fuzzy regions of the teeth from the jaw model by predicting multiple binary masks. The landmark detection network then identifies the landmark positions. Both networks are trained using a bipartite matching loss. During end-to-end inference, 'background' masks and landmarks are filtered based on the corresponding predicted probabilities. Extensive experiments on the Teeth3DS+ dataset validate the efficiency and robustness of ToothLDNet.

Acknowledgments. This work is partly supported by the National Natural Science Foundation of China (No. 62172363).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ben-Hamadou, A., Neifar, N., Rekik, A., Smaoui, O., Bouzguenda, F., Pujades, S., Boyer, E., Ladroit, E.: Teeth3ds+: An extended benchmark for intraoral 3d scans analysis (2022) 6
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on computer vision. pp. 213–229. Springer (2020) 2
3. Dascalu, T., Ibragimov, B.: Assignment theory-augmented neural network for dental arch labeling. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 295–304. Springer (2023) 1
4. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd. vol. 96, pp. 226–231 (1996) 5
5. Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques. pp. 209–216 (1997) 4
6. Huang, T., Shi, J., Jin, G., Li, J., Wang, J., Du, J., Shi, J.: Topological gcn for improving detection of hip landmarks from b-mode ultrasound images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 692–701. Springer (2024) 1
7. Jin, H., Shen, Y., Lou, J., Zhou, K., Zheng, Y.: Keypointdetr: An end-to-end 3d keypoint detector. In: European Conference on Computer Vision. pp. 374–390. Springer (2024) 2, 3, 4, 6, 7
8. Jin, H., Shen, Y., Lou, J., Zhou, K., Zheng, Y.: Tsrnet: A dual-stream network for refining 3d tooth segmentation. IEEE Transactions on Visualization and Computer Graphics (2024) 2
9. Kuhn, H.W.: The hungarian method for the assignment problem. Naval research logistics quarterly **2**(1-2), 83–97 (1955) 6
10. Lang, Y., Chen, X., Deng, H.H., Kuang, T., Barber, J.C., Gateno, J., Yap, P.T., Xia, J.J.: Dentalpointnet: landmark localization on high-resolution 3d digital dental models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 444–452. Springer (2022) 1
11. Li, P., Gao, C., Liu, F., Meng, D., Yan, Y.: Thisnet: Tooth instance segmentation on 3d dental models via highlighting tooth regions. IEEE Transactions on Circuits and Systems for Video Technology (2023) 2
12. Liu, Y., Li, W., Wang, C., Chen, H., Yuan, Y.: When 3d partial points meets sam: Tooth point cloud segmentation with sparse labels. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 778–788. Springer (2024) 1
13. Ngo, T.D., Hua, B.S., Nguyen, K.: Isbnet: a 3d point cloud instance segmentation network with instance-aware sampling and box-aware dynamic convolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13550–13559 (2023) 2

14. Pfister, T., Charles, J., Zisserman, A.: Flowing convnets for human pose estimation in videos. In: Proceedings of the IEEE international conference on computer vision. pp. 1913–1921 (2015) 1
15. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017) 7, 8
16. Rodriguez, A., Laio, A.: Clustering by fast search and find of density peaks. science **344**(6191), 1492–1496 (2014) 2
17. Rodríguez-Ortega, J., Tabik, S.: Charnet: Conditioned heatmap regression for robust dental landmark localization. arXiv preprint arXiv:2501.13073 (2025) 1
18. Sun, J., Qing, C., Tan, J., Xu, X.: Superpoint transformer for 3d scene instance segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2393–2401 (2023) 2
19. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017) 8
20. Wang, C., Wei, G., Wei, G., Wang, W., Zhou, Y.: Tooth alignment network based on landmark constraints and hierarchical graph structure. IEEE Trans. Vis. Comput. Graph. **30**(2), 1457–1469 (2022) 1
21. 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) 7, 8
22. Wei, G., Cui, Z., Zhu, J., Yang, L., Zhou, Y., Singh, P., Gu, M., Wang, W.: Dense representative tooth landmark/axis detection network on 3d model. Computer Aided Geometric Design **94**, 102077 (2022) 1, 2
23. Wei, G., Ma, L., Wang, C., Desrosiers, C., Zhou, Y.: Multi-task joint learning of 3d keypoint saliency and correspondence estimation. Computer-Aided Design **141**, 103105 (2021) 7
24. Wu, H., Wang, C., Mei, L., Yang, T., Zhu, M., Shen, D., Cui, Z.: Cephalometric landmark detection across ages with prototypical network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 155–165. Springer (2024) 1
25. Wu, T.H., Lian, C., Lee, S., etal: Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3d intraoral scans. IEEE Transactions on Medical Imaging **41**(11), 3158–3166 (2022) 1, 2
26. Xia, J.J., Gateno, J., Teichgraeber, J.F.: New clinical protocol to evaluate craniomaxillofacial deformity and plan surgical correction. Journal of oral and maxillofacial surgery **67**(10), 2093–2106 (2009) 1
27. Zhang, L., Zhao, Y., Meng, D., Cui, Z., Gao, C., Gao, X., Lian, C., Shen, D.: 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) 2
28. Zheng, Y., Chen, B., Shen, Y., Shen, K.: Teethgnn: semantic 3d teeth segmentation with graph neural networks. IEEE Transactions on Visualization and Computer Graphics **29**(7), 3158–3168 (2022) 3, 4, 5, 8