Tooth Keypoint Detection Using KeypointDETR

Guohua Zhu, Jaehan Joo, Dabin Hwang, Suk Chan Kim*
Pusan National Univ., Pusan National Univ., Pusan National Univ., *Pusan National Univ.
zhuguohua@pusan.ac.kr, jhjoo2018@pusan.ac.kr, dabinsor@pusan.ac.kr, *sckim@pusan.ac.kr

KeypointDETR를 이용한 치아 특징점 검출

주귀화, 주재한, 황다빈, 김석찬*
부산대학교, 부산대학교, 부산대학교, *부산대학교

Abstract

Accurate detection of anatomical landmarks on dental models is crucial for clinical applications such as occlusion analysis, orthodontic treatment, and automated tooth segmentation. In this study, we investigate the use of KeypointDETR, a transformer-based keypoint detection framework, for localizing cusp points on 3D dental meshes from the Teeth3DS dataset. We reformulate the landmark detection task as a set prediction problem, enabling end-to-end training without post-processing. Experiments demonstrate that our method achieves high accuracy and robustness, with mean Average Precision (mAP) and mean Average Recall (mAR). This approach lays a foundation for incorporating dental landmark detection into intelligent dental analysis systems.

Ⅰ. Introduction

Dental keypoints, such as cusp points, play a central role in understanding tooth morphology and occlusal surfaces. Traditional methods rely on surface curvature or expert-defined geometric heuristics, which are often limited by anatomical variability and mesh noise.

Recently, transformer-based detection models such as KeypointDETR have shown superior performance in object detection and keypoint localization in 2D and 3D domains[1]. Inspired by this, we apply KeypointDETR to detect only the cusp points—one of the most clinically relevant keypoint types—on the 3D dental surface models provided by the Teeth3DS dataset.

Previous studies typically relied on mesh processing or curvature analysis to detect landmarks, which can be sensitive to noise and limited in generalization. DETR and its variants (including KeypointDETR) cast detection as a set prediction problem. KeypointDETR in particular utilizes a learnable query-based mechanism to directly regress keypoint positions, avoiding the need for heatmap generation or post-matching. Methods such as PointNet++ have demonstrated effectiveness in extracting features from unstructured point clouds, making them suitable for dental mesh representations.

This study aims to:

Adapt KeypointDETR to process 3D dental meshes converted into point clouds.

Train the model to detect cusp points using only coordinate supervision.

Evaluate its performance using mean Average Precision (mAP) and mean Average Recall (mAR).

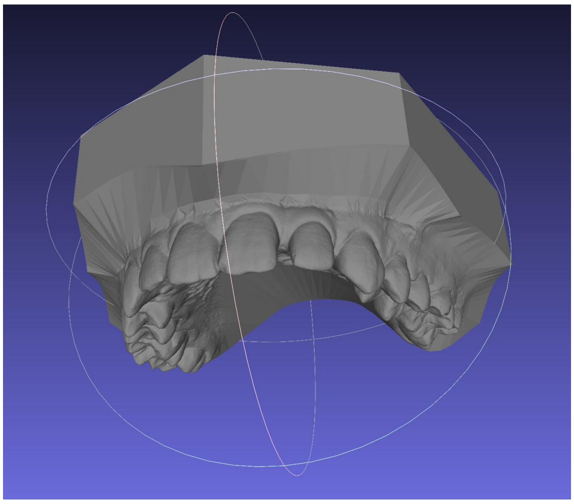

Figure 1. Teeth3DS dataset.

Ⅱ. Method

We use the Teeth3DS dataset, as Figure 1, which contains 3D dental models with annotated keypoints, including multiple cusp points on molars and premolars[2]. Each mesh is sampled into a 8192-point point cloud, normalized to a unit sphere. We retain only the cusp point annotations (3D coordinates) for each tooth.

Our approach is based on KeypointDETR and consists of a point cloud encoder followed by a transformer encoder-decoder. This structure enables the model to detect 3D cusp points directly from raw point cloud input.

We use a PointNet++ inspired encoder to extract local and global geometric features[3]. It performs farthest point sampling, neighborhood grouping, and MLP-based feature aggregation. The output is a set of point-wise features used as tokens for the transformer.

The transformer encoder applies self-attention to model relationships among point features. The decoder takes a fixed number of learnable queries, each predicting a 3D coordinate and a confidence score for a potential cusp point.

Predicted keypoints are matched with ground truth using the Hungarian algorithm. The training loss combines L1 regression and binary confidence loss. During inference, top-scoring keypoints are selected as final cusp point predictions.

The dataset was split into training, validation, and test sets using a 6:2:2 ratio. All splits were performed at the tooth level to ensure no data leakage. We ensured that samples from the same tooth model were not shared across different subsets.

The model was trained using the AdamW optimizer with an initial learning rate of 1e-4 and weight decay of 1e-2. The batch size was set to 8 and the training ran for 100 epochs. Cosine annealing was used to schedule the learning rate. We used a combination of L1 loss for keypoint localization and binary cross-entropy loss for confidence scores. The Hungarian matching algorithm was applied to match predicted and ground-truth keypoints during training.

We evaluated performance using two standard metrics: mean Average Precision (mAP) and mean Average Recall (mAR). A prediction was considered correct if the predicted keypoint was within a predefined Euclidean distance threshold from the corresponding ground truth. We report mAP and mAR to reflect different levels of clinical tolerance. Visualization results are showed as Figure 2.

The model achieved a mean Average Precision (mAP) of 0.51 and a mean Average Recall (mAR) of 0.39. These results suggest that the model is reasonably precise when predicting cusp points, but still misses a notable number of ground truth landmarks, as reflected in the lower recall. Improvements in recall may require enhanced keypoint localization strategies or better handling of ambiguous regions.

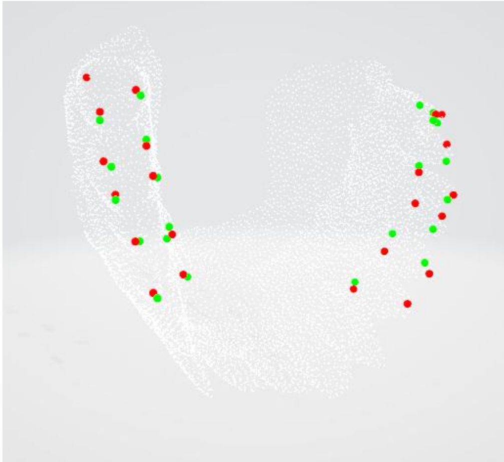

Figure 2. Visualization results. Red indicates the ground truth (GT), and green indicates the predicted values.

Ⅲ. Conclusion

We proposed a novel application of KeypointDETR to detect cusp points on 3D dental models. Using only point clouds and coordinate supervision, our approach achieves high mAP and mAR on the Teeth3DS dataset. This demonstrates the feasibility of transformer-based models in high-precision dental landmark detection and opens the door to broader adoption in digital dentistry.

ACKNOWLEDGMENT

This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program) (00144500, Development of 3D ultrasonic sensor technology for vehicles based on meta structure) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea)

“This work is financially supported by Korea Ministry of Land, Infrastructure and Transport(MOLIT) as 「Innovative Talent Education Program for Smart City」".

REFERENCES

[1] Jin H, Shen Y, Lou J, et al. KeypointDETR: An End-to-End 3D Keypoint Detector[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 374-390.

[2] Ben-Hamadou A, Smaoui O, Chaabouni-Chouayakh H, et al. Teeth3ds: a benchmark for teeth segmentation and labeling from intra-oral 3d scans[J]. arXiv e-prints, 2022: arXiv: 2210.06094.

[3] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.