Accurate Localization of Dental Landmarks with a Hierarchical Deep Learning Model

George Vadakepurathan Jose

The University of Adelaide

a1895382

Abstract

Accurate landmark identification on dental casts is critical for evaluating palatal morphology, which is closely associated with orthodontic treatments and its outcomes. Current practices require clinicians to manually digitize more than 60 landmarks on 3D dental surface model. While digitization has enabled more precise evaluations, manual landmarking remains time-consuming, error-prone, and subject to observer variability. Automating the process of landmark identification could significantly enhance both the accuracy and efficiency of these measurements, benefiting clinicians and researchers alike.

This project addresses the challenge of accurately predicting seven key landmarks in 3D dental point cloud scans. The proposed methodology employs a two-stage prediction framework consisting of a coarse prediction followed by fine-grained localized prediction. The coarse prediction utilizes a point cloud encoder architecture with three PointNet++ set abstraction layers to generate an initial approximation of the landmark locations with 256 points. The finer prediction stage leverages a point cloud encoder-decoder network, incorporating three set abstraction layers and three feature propagation layers in the decoder for precise landmark refinement from a localized region. The final model achieves an average prediction error of less than 1.2 mm on 63 high-resolution dental scans, demonstrating high accuracy and robustness in dental landmark localization.

1. Introduction

Localization of dental landmarks is a critical first step in dental model analysis, particularly during treatment planning for patients with jaw and teeth deformities. These landmarks provide essential reference points for evaluating palatal morphology, assessing craniofacial structures, and monitoring treatment outcomes. With the advent of three-dimensional (3D) surface scanners, high-resolution digital dental models can now be generated, offering unprecedented accuracy and detail.

Despite the advancements in digitization, the current standard for landmark identification relies heavily on manual annotation by experts. This process is time-consuming, prone to fatigue-induced errors, and subject to intra- and inter-observer variability, which can affect the consistency and accuracy of the results. Automating landmark localization has the potential to overcome these challenges, making the process more efficient and reliable while reducing human dependency. Accurate and automated landmark prediction would not only benefit clinicians by streamlining workflows but also enable researchers to conduct large-scale studies with higher reproducibility. Processing high-resolution dental models, however, presents its own set of challenges. These models typically consist of more than 100,000 mesh cells, making them computationally expensive to analyse.

Recent advances in deep learning have demonstrated the effectiveness of neural networks in processing and analysing point cloud data—a structured representation of 3D surfaces. In particular, PointNet++ [1], a widely used architecture for point cloud processing, builds on its predecessor PointNet by introducing hierarchical learning of features from grouped points. The architecture divides the point cloud into local regions through a process called set abstraction, where points are grouped based on spatial proximity. Within each region, PointNet++ applies PointNet to extract local features and hierarchically combines them to capture both fine-grained local structures and global geometric context. This enables PointNet++ to handle variations in point density and spatial relationships effectively, making it particularly well-suited for tasks that require precise localization, such as dental landmark prediction.

This project aims to develop a deep learning framework capable of accurately predicting seven key landmarks in 3D dental point cloud scans. The proposed approach leverages a two-stage prediction framework, incorporating PointNet++ and its extensions to ensure both computational efficiency and high precision. The ultimate objective is to achieve a high degree of accuracy, ensuring clinically viable results.

2. Data and Methods

2.1 Dataset

This project utilizes a custom, in-house dataset comprising 324 high-quality 3D scans of the permanent dentition stage. Each dental scan represents a detailed surface of the maxilla, annotated with seven key landmark points. The landmark annotations were meticulously labelled by trained technicians and surgeons to ensure accuracy and consistency. The resolution of the scans varies, with each dental surface containing approximately 60,000 to 120,000 vertices. This level of detail captures the intricate morphology of the dental structures, making the dataset suitable for training deep learning models designed for precise landmark localization.

Preprocessing was performed to prepare the dataset for analysis. Each point cloud was first aligned by rotating it to ensure uniform orientation and then scaled to fit within a normalized bounding box of [−1,1] along the longest axis, which corresponds to the X-axis. Additionally, vertex normals were computed and appended to each vertex. As a result, each cell in the mesh is represented by a 6-dimensional vector, comprising three coordinates for the vertex positions and three components of the vertex normals. Finally, a Furthest Point Sampling (FPS) is applied to uniformly sample 24000 points from the point cloud. This enhanced representation provides geometric and surface orientation information, which aids the model in learning intricate spatial relationships more effectively.

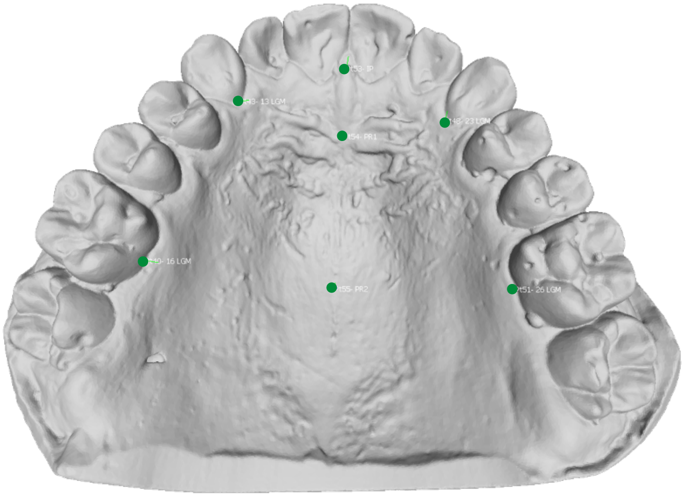

Figure 1: Visualization of a 3D dental scan with annotated landmark points.

2.2 Model Architecture

The proposed framework employs a two-stage architecture for accurate landmark localization, consisting of a coarse prediction stage and a fine prediction stage.

Coarse Estimation Stage

The first stage of the approach is a coarse predictor, designed to generate 256 points densely clustered around the seven target landmarks. This stage uses a PointNet++ encoder with three *Set Abstraction* (SA) layers. Each SA layer hierarchically groups the points within a defined query radius and extracts geometric features. The query radius increases progressively across the layers, starting from 0.025 in the first layer to 0.2 in the final layer.

The increasing query radius plays a critical role in feature extraction. Smaller radii in the initial layers capture fine-grained, local geometric details, while larger radii in the deeper layers enable the model to aggregate broader contextual information from the entire point cloud. This hierarchical feature learning allows the model to balance localized and global understanding, which is essential for accurately clustering points around the target landmarks.

Following the encoder, the coarse predictor employs two separate heads, each implemented as convolutional layers with a kernel size of 1. One head predicts the distance, and the other predicts the offset, with both outputs combined to generate the final 256-point cluster. The weights of both heads are initialized to zero at the start of training to prevent biased predictions during early learning stages.

Refinement Stage

The fine prediction stage refines the coarse results by focusing on localized regions of the point cloud around each predicted landmark. Using the coarse predictions, the model crops a geodesic ball with a radius of 6 mm (determined empirically) around each landmark. This cropped mesh patch contains sufficient localized information for each landmark while excluding irrelevant areas of the point cloud.

The fine prediction model employs an encoder-decoder PointNet++ architecture. The encoder consists of three SA layers, similar to the coarse stage, to extract features hierarchically from the cropped region. The decoder incorporates three *Feature Propagation (FP*) layers to upsample and refine these features back to the input resolution.

The fine prediction model outputs a *probability map* of the same size as the input point cloud. This probability map indicates the likelihood of each point being the target landmark. The ground truth probability map is generated as a Gaussian distribution centered at the true landmark position, with a standard deviation (σ) of 0.2. This approach ensures that the model learns a smooth and precise spatial representation of the landmark location and it has enough information regarding every point in the point cloud.

3. Implementation and Inference

The dataset is divided into three subsets: training, validation, and abnormal scans, consisting of 251, 63, and 10 samples, respectively. The training set is used for optimizing the model parameters, the validation set for monitoring performance during training, and the abnormal set for evaluating the model's robustness on challenging or outlier cases.

Coarse Estimation Module

The coarse prediction module is trained using the training dataset, where each input data point is represented as a point cloud of size [24000,6], including vertex positions and normals. The model processes this input to predict a cluster of 256 points, each described by [256,3], which represent the coarse approximations of the landmarks.
The training objective for the coarse prediction module is defined as a combination of three loss terms:

1. **Distance Loss**
   This term measures the Euclidean distance between each predicted point and its nearest ground truth landmark. Minimizing this loss ensures that the predictions are spatially close to the actual landmarks.
2. **Chamfer Loss**
   The chamfer loss minimizes the bidirectional distance between predicted points and landmarks, enforcing two key properties:
   - **Surjection**: Every landmark should correspond to at least one predicted point.
   - **Injection**: Every predicted point should correspond to exactly one landmark. This ensures a complete and unique mapping between predicted points and the ground truth landmarks.
3. **Separation Loss**
   The separation loss encourages the predicted points to be associated with the correct landmarks. It is defined as the ratio of the distance to the closest landmark over the distance to the second closest landmark. Minimizing this term ensures that each predicted point is tightly clustered around the correct landmark, avoiding ambiguity.

The total loss for the coarse prediction module is computed as:

$$L_{coarse} = L_{distance} + L_{chamfer} + 0.5 \times L_{separation}$$

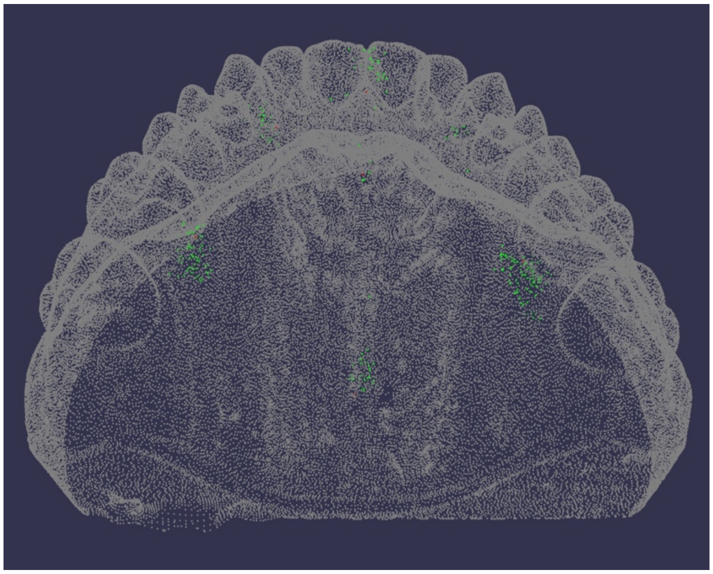

Figure 2: Visualization of 256 coarse predictions

Refinement Module

The fine prediction module refines the coarse predictions by focusing on localized regions around each predicted landmark. The localized regions are generated based solely on the output of the coarse prediction module, ensuring that no ground truth information is used during this process, thus maintaining the integrity of the data.

To form the input for the fine prediction module, the 256 predicted points from the coarse prediction stage are used as the basis. *DBSCAN* (Density-Based Spatial Clustering of Applications with Noise) clustering is applied to these points, dividing them into seven clusters corresponding to the landmarks. A geodesic ball with a radius of 6 mm is then cropped around each cluster centre, which contains the localized information for each landmark. This approach ensures that each patch contains relevant information specific to each landmark, while irrelevant points outside the region of interest are excluded.

This module is trained using a mean squared error (MSE) loss between the predicted probability map and the ground truth probability map. The ground truth probability map is modeled as a Gaussian distribution centered around the true landmark position with a standard deviation (σ=0.2). The predicted point is the point with the highest probability.

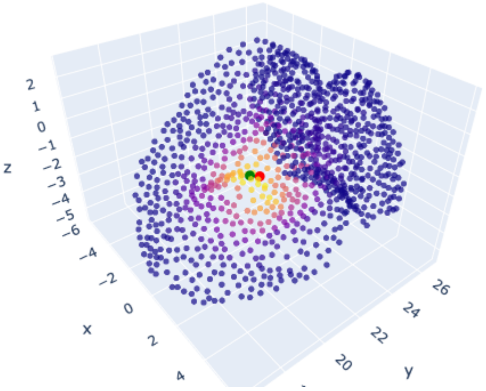

Figure 3: Visualization of Refinement probability map (Green: Ground Truth, Red: Prediction)

All training and evaluation procedures were implemented in Python using the PyTorch framework, and the experiments were conducted on an NVIDIA RTX 2080 Ti GPU. The coarse prediction module was trained for 50 epochs using the Adam optimizer with an initial learning rate of 1e-3. The refinement module was trained for 50 epochs as well, with a higher learning rate of 2e-3 and a gradient clipping threshold of 0.5 to stabilize training.

4. Results

The proposed model achieves a mean distance of less than 1.2 mm from the ground truth for each of the seven landmarks. This level of precision demonstrates the model's ability to consistently and accurately localize dental landmarks in 3D point cloud scans. The success of the model can be attributed to its two-stage architecture, which combines coarse localization with fine refinement. The coarse prediction module effectively clusters points around the landmarks, ensuring that the fine prediction module processes localized regions with minimal noise. By focusing on smaller, geodesically cropped regions, the fine prediction module can refine landmark positions with greater accuracy, leveraging the detailed geometric information present in these localized patches. This hierarchical approach balances efficiency and precision, allowing the model to process high-resolution point clouds without sacrificing accuracy.

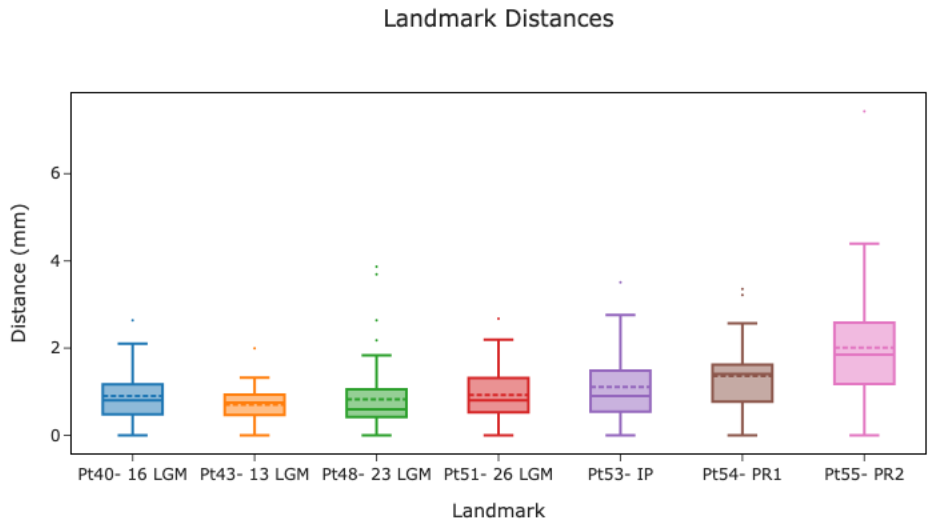

Figure 4: Distance from ground truth for seven landmarks

Upon further analysis, it was observed that the landmarks associated with *Palatal Rugae (PR)* generally exhibit higher mean distances and greater standard deviation compared to other landmarks. This discrepancy can be attributed to the location of the PR landmarks on the dental palate. These landmarks are positioned in the middle of the palate, where the geometric structure provides limited distinctive information. The absence of strong structural cues makes it more challenging for the refinement module to precisely localize these points.

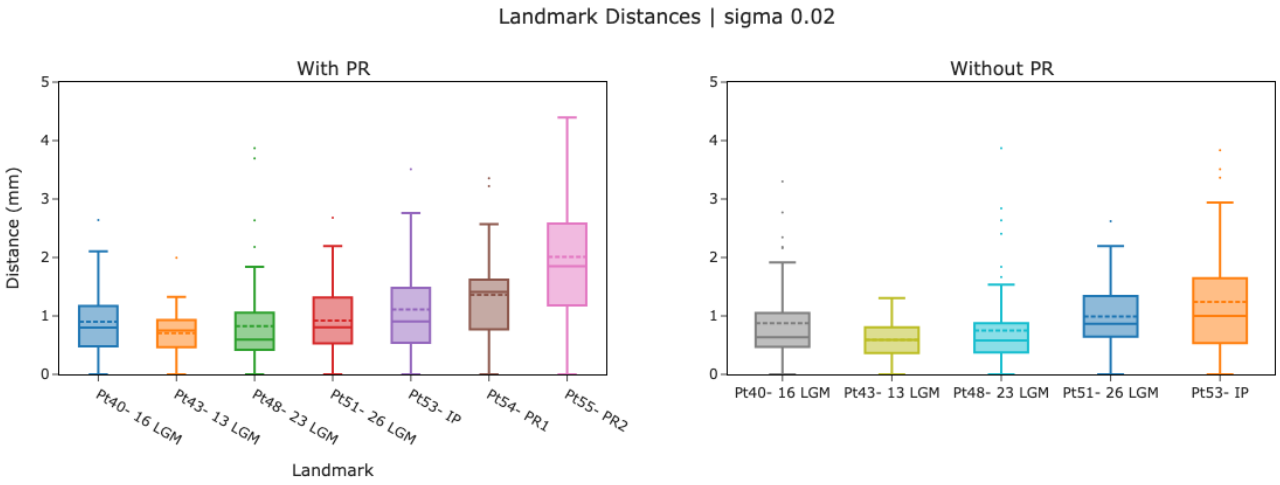

Figure 5: Landmark distances including and excluding Palatal Rugae (PR)

To evaluate the impact of these challenging landmarks, experiments were conducted with the PR points excluded from the analysis. The results in Figure 4 show a noticeable improvement, with reduced standard deviation in the distances from the ground truth. This indicates that the predictions are more tightly clustered around the true landmark positions, highlighting the sensitivity of the model to the structural context of each landmark.

5. Conclusion

This project successfully developed a deep learning framework for predicting seven key landmarks in 3D dental point cloud scans, achieving an average error of less than *1.2 mm* from the ground truth. The two-stage architecture, combining coarse and fine prediction modules, proved highly effective in addressing the challenges of high-resolution dental scans, streamlining the localization process while maintaining accuracy. The results demonstrate significant clinical potential by automating landmark annotation, reducing human error and inter-observer variability, and saving time in orthodontic treatment planning and craniofacial analysis. However, challenges remain, particularly in improving predictions for landmarks located in geometrically flat regions like the Palatal Rugae. Future work could focus on incorporating additional contextual features, integrating the model into clinical workflows, and exploring advanced architectures such as attention mechanisms. This study highlights the feasibility and impact of deep learning-based approaches for automated dental landmark localization.

6. References List

Balder Croquet, Matthews, H, Mertens, J, Fan, Y, Nele Nauwelaers, Mahdi, S, Hanne Hoskens, Sergani, AE, Xu, T, Vandermeulen, D, Bronstein, M, Marazita, M, Weinberg, S & Claes, P 2021, 'Automated landmarking for palatal shape analysis using geometric deep learning', *Orthodontics and Craniofacial Research*, vol. 24, Wiley, no. S2, pp. 144–152.

Cui, Z, Li, C, Chen, N, Wei, G, Chen, R, Zhou, Y, Shen, D & Wang, W 2021, 'TSegNet: An efficient and accurate tooth segmentation network on 3D dental model', *Medical Image Analysis*, vol. 69, p. 101949.

Lang, Y, Chen, X, Deng, HH, Kuang, T, Barber, JC, Gateno, J, Yap, P-T & Xia, JJ 2022, 'DentalPointNet: Landmark Localization on High-Resolution 3D Digital Dental Models', *Lecture notes in computer science*, Springer Science+Business Media, pp. 444–452, viewed 22 January 2025, <https://conferences.miccai.org/2022/papers/150-Paper0632.html>.

Lang, Y, Deng, HH, Xiao, D, Lian, C, Kuang, T, Gateno, J, Yap, P-T & Xia, JJ 2021, 'DLLNet: An Attention-Based Deep Learning Method for Dental Landmark Localization on High-Resolution 3D Digital Dental Models', *Lecture notes in computer science*, Springer Science+Business Media, pp. 478–487.