A COMPARATIVE STUDY OF CLASSICAL AND DEEP LEARNING MODELS FOR ANATOMICAL POINT CLOUD REGISTRATION

Zou Han

King's College London

ABSTRACT

This project focuses on rigid 3D point cloud registration for high-resolution dental scans, aiming to support accurate and robust alignment for tracking anatomical changes over time in the oral cavity. We implemented and evaluated several registration methods, including the traditional Procrustes analysis and three deep learning models: Multi-Layer Perceptrons (MLP), PointNet++, and Graph Convolutional Networks (GCN). The dental data consisted of intraoral scans collected from real patients at multiple time points, with anatomical landmarks provided based on input from dental experts. For the Procrustes method, we tested various landmark selection strategies to assess its performance under different spatial configurations, and compared the results with those of the learning-based models. Using Chamfer Distance and Mean Squared Error (MSE) as evaluation metrics, our experiments showed that while the Procrustes method performed acceptably under ideal annotations, the PointNet++ model demonstrated superior generalizability and alignment accuracy across most landmark settings. All models, datasets, and analysis code from this project have been made openly available, providing practical guidance and technical reference for 3D point cloud registration in medical applications, particularly within the field of dentistry.

1. INTRODUCTION

In medical image analysis, 3D point cloud registration is a fundamental and essential task [1]. This is especially true when multiple scans of the same anatomical region are taken over time and need to be compared. In such cases, the accuracy and stability of the alignment directly affect the quality of subsequent analysis [2]. In dental practice, the application of this technique is particularly important [3,4]. For example, accurately aligning intraoral scans collected at different time points can help track tooth movement, monitor disease progression, and support the design of prosthetics or orthodontic treatment plans.

This project focuses on rigid point cloud registration for high-resolution dental scans, with the goal of identifying an alignment method that is both accurate and robust enough to handle common challenges in clinical data, such as anatomical variation and inconsistent landmark annotations. We implemented and compared four different registration methods: the traditional Procrustes method [5], and three deep learning-based models — Multi-Layer Perceptrons (MLP), PointNet++, and Graph Convolutional Networks (GCN) [6]. These methods represent two distinct approaches: classical geometric techniques and modern machine learning strategies.

The data used for training and evaluation consisted of intraoral 3D scans collected from real patients at multiple time points. All anatomical landmarks were annotated with input from dental experts to ensure medical relevance. For the Procrustes method, we tested multiple landmark selection strategies to evaluate its performance under different spatial configurations, and compared its results against those of the deep learning models.

To evaluate registration accuracy, we used two key metrics: Chamfer Distance [7] and Mean Squared Error (MSE). These metrics provide insight from different perspectives — one focusing on overall shape alignment, and the other on point-to-point accuracy at key landmarks. The experimental results showed that while the Procrustes method achieved acceptable performance under ideal annotations, the PointNet++ model demonstrated superior generalizability and more stable performance across varied conditions.

All models, data processing pipelines, and evaluation code have been released openly to support reproducibility and enable future research. We hope this project can serve as a useful reference for improving and applying 3D point cloud registration techniques in medical imaging, particularly within the field of dentistry.

Code:
https://github.com/Champagnehz/Code_POINT_CLOUD_REGISTRATION

2. METHODOLOGY

This section describes the methodology I followed in the project, with a focus on the implementation and evaluation of rigid 3D point cloud registration techniques applied to dental

data. My main contributions included 3D landmark annotation, design of landmark selection strategies based on the methodology proposed in paper [8], initial implementation of the Procrustes algorithm, development and training of the GCN model, and the organization and analysis of experimental results. These tasks covered key stages from data preparation to model evaluation, ensuring a complete and interpretable pipeline.

2.1 3D Annotation and Data Preprocessing

In this project, I was primarily involved in the 3D annotation of dental data and worked closely with my teammates to complete several key tasks, including data preprocessing, manual annotation, and strategy design. The data we used came from intraoral scans of real patients at multiple time points, stored in point cloud format, which provided significant value for longitudinal analysis.

Through extended hands-on exploration, I discovered that the issue of annotation points not moving with the model during rotation in Blender could be effectively resolved using the "empty object anchoring" method. I shared this solution with my teammate Shichao Zhang, and together we completed the annotation of multiple sets of scan data, ensuring consistency in position.

To balance registration accuracy with the goal of minimizing the number of annotated landmarks, we designed and evaluated eight different landmark selection strategies for the Procrustes registration method. By comparing the performance of these strategies side by side, we were able to identify the most effective configuration. This optimal setup was then used as a consistent baseline for subsequent comparisons with deep learning models, allowing for clearer and more meaningful evaluation across methods.

In addition, I participated in experiments evaluating the impact of removing specific landmarks on registration performance. We completed a full annotation of 116 landmarks on sample A1 (see Fig. 1), which served as the reference baseline and the parent set for all subsequent subset strategies (Thea_01 to Thea_08). Based on this foundation, we systematically tested how the removal of different types of feature points affected registration error, allowing us to identify which landmarks played a critical role in maintaining geometric integrity during rigid alignment.

The annotation workflow mainly included the following steps:

- Centering: Make sure each scan was translated to the origin.
- Manual annotation: This process was guided by relevant literature and input from dental experts.
- Format conversion: I converted the Blender-exported annotation data into standardized model input formats (such as NumPy arrays or CSV files) and modularized the code for reuse by the rest of the team.
- Landmark position processing: I wrote code to extract the one-to-one matched landmark coordinates across different years, which allowed direct input into downstream registration models.
- Strategy foundation: The eight landmark selection strategies used in the project (Thea_01 to Thea_08) were all generated based on this standardized landmark dataset, as detailed in **Appendix A**.

This phase of the work provided a stable and reliable data foundation for the subsequent comparison experiments using the Procrustes registration method and for training deep learning models.

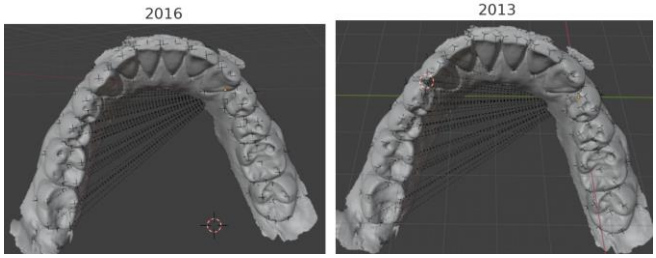

Fig. 1. The complete landmark set on sample A1 between 2016 and 2013 scans (116 landmarks)

2.2 Initial Implementation of Procrustes Algorithm

I implemented a basic version of the Procrustes rigid registration algorithm in Python to align 3D dental landmark point clouds collected at different time points. This method uses translation and rotation (without scaling) to minimize the Euclidean distance between corresponding points. The objective of the alignment is to solve [9]:

$$min_{R,t} \sum_{i=1}^{n} | RX_i + t - Y_i|^2$$

where X_i and Y_i represent corresponding landmarks from the source and target point clouds, respectively.

To solve this optimization problem, both point sets are first centered by subtracting their centroids. The optimal rotation matrix R is then obtained by computing the covariance matrix $H = X^T Y$, yielding:

$$H = U\Sigma V^T \, , R = VU^T$$

The translation vector t is then obtained as:

$$t = \bar{Y} - R\bar{X}$$

where $\bar{X}$ and $\bar{Y}$ are the centroids of the input point sets.

Input data consisted of 3D landmark coordinates stored in JSON format. Each point cloud was first centered by subtracting its centroid, and normalized when necessary to reduce scale bias. After registration, the centroid was added back to restore the original spatial position. The output included two aligned point sets (*mtx1* and *mtx2*) and a disparity value representing the registration error. The

disparity can be understood as a normalized residual defined by:

$$Disparity = \frac{\sum_{i=1}^{n} |X_i' - Y_i'|^2}{\sum_{i=1}^{n} |Y_i'|^2}$$

where X_i' and Y_i' are the centered and scaled versions of the input point sets.

To verify the registration result, I programmed code to generate 3D visualizations of the aligned point clouds. I also wrote auxiliary scripts for data loading, coordinate matrix processing, and error output, and tested multiple data formats to ensure that the implementation could be smoothly integrated into the subsequent strategy selection and evaluation pipeline. In addition, to explore the effect of centering and transformation sequence on registration accuracy, MSE and disparity were used for evaluation. As the first completed registration method in this project, this implementation provided a stable and reliable baseline for evaluating the eight landmark selection strategies within the Procrustes framework, as well as for comparison with deep learning models.

A simplified workflow of this implementation is shown below (see Fig. 2), outlining the main steps from preprocessing to output.

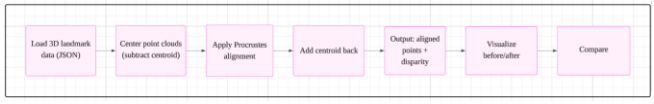

Fig. 2. Workflow of the initial procrustes registration pipeline

2.3 GCN Model Implementation and Training

In this project, I was responsible for implementing and training a Graph Convolutional Network (GCN) model for rigid registration of 3D dental point clouds. The model was designed to predict rigid transformations—specifically a rotation matrix R and a translation vector t—to achieve accurate alignment between point clouds captured at different time points.

2.3.1 Model Architecture and Key Components

The model was built in PyTorch and follows a three-layer structure named GCNRegistrationModel, where each layer contains an EdgeConvAttention module (Fig.3 shows the idea of EdgeConv [10]). This module uses message passing over edge features, which is a common mechanism in Graph Convolutional Networks (GCNs) and particularly effective for learning representations from point cloud data. The attention mechanism enables the network to adaptively aggregate features from local neighborhoods based on spatial context, improving its ability to capture fine-grained geometric structures. The outputs from all three layers are concatenated and passed through a Multi-Layer Perceptron (MLP) for global feature extraction. A set of fully connected (FC) layers then predicts the transformation parameters, including a 3×3 rotation matrix and a 3D translation vector.

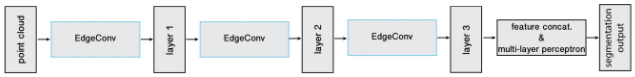

Fig. 3. Schematic neural network architecture

The predicted rotation matrix is orthogonalized during training using Singular Value Decomposition (SVD[11]) to ensure it remains a valid rigid transformation (i.e., with determinant = 1).

2.3.2 Data Preprocessing and Loading

The training data came from intraoral scans of real patients taken at multiple time points (e.g., 2013 and 2016). Each scan pair was represented by manually annotated landmarks and stored in *.json* format. I developed two scripts, *T_NPY.py* and *T_NPY_combined_multiple.py*, to convert the landmark data into *.npy* format. The resulting tensors had a standardized shape of [1,N,3]. Preprocessing steps included centering and normalization to remove positional and scale bias.

2.3.3 Loss Function Design

The loss function was designed to jointly optimize local and global alignment accuracy while preserving the rigidity of the transformation. It combined three terms: the Mean Squared Error (MSE), which measured the point-wise Euclidean distance between predicted and target landmarks; the Chamfer Distance (CD), which evaluated the overall shape similarity between the two point clouds; and an orthogonality loss that constrained the predicted rotation matrix to remain rigid and valid. The total loss was formulated as:

$$\text{Loss} = \text{MSE} + 0.1 \times \text{CD} + 0.01 \times \text{Orthogonality Loss}$$

This multi-term design ensured that the network not only focused on precise landmark alignment but also maintained structural coherence and transformation integrity during training.

2.3.4 Training Procedure

The model was trained on a GPU for 1000 epochs using the *ReduceLROnPlateau* scheduler for dynamic learning rate adjustment. During early training, I manually verified gradient flow to confirm that all modules were being updated correctly. Model checkpoints were saved automatically, and the final output was selected based on the lowest validation error.

2.3.5 Post-Processing: ICP Refinement

To further improve local registration accuracy, I applied an Iterative Closest Point (ICP [12,13]) refinement step using Open3D, based on point-to-point distance minimization. The output included both the GCN-aligned result and the ICP-refined version, which were exported as .ply files for visualization and analysis.

2.3.6 Evaluation and Visualization

Using *Validation_model.py* and *Validation_T_NPY.py*, I generated visualizations of the point clouds before registration, after GCN alignment, and after ICP refinement. The results demonstrated that the model achieved good overall alignment, while local accuracy was further improved with ICP optimization.

2.4 Result Organization and Evaluation Metrics

To ensure comparability across different registration methods under a consistent evaluation standard, all model outputs were organized by method category (Procrustes, GCN, MLP, and PointNet++). I compiled and categorized these results systematically to support analysis of model performance within a unified evaluation framework. For each sample, the outputs included *.ply* files of the aligned point clouds and automatically computed registration error metrics, which facilitated both quantitative analysis and visual inspection.

During the evaluation process, we applied a standardized workflow across all models. This included pre- and post-alignment error analysis, point cloud format normalization, and the extraction of key evaluation metrics. The primary quantitative indicators remained Mean Squared Error (MSE) and Chamfer Distance.

For the Procrustes method, we tested eight different landmark selection strategies and compared their performance. For learning-based models, we also included results after refinement using ICP, enabling a more comprehensive assessment of the model's performance.

In addition to the numerical metrics, we also performed 3D deviation visualization of the aligned point clouds for intuitive interpretation. Some samples were further analyzed using surface deviation evaluation, where spatial deviation from the reference structure was calculated and visualized using color maps to highlight regional alignment accuracy [14]. This type of analysis complemented the numerical results and helped validate registration quality from an anatomical and geometric perspective.

3. RESULTS AND DISCUSSION

This section presents and interprets the registration results obtained from both classical and learning-based methods. We evaluate their performance using quantitative metrics (Mean Squared Error and Chamfer Distance), structural robustness, and visual interpretability. The analysis is divided into four parts: first, we assess the performance of the Procrustes method under different landmark selection strategies; second, we compare three learning-based models (MLP, PointNet++, GCN); third, we examine the performance gap between traditional and deep learning approaches; and finally, we validate the alignment outputs using clinical surface deviation analysis. Together, these results provide a comprehensive understanding of how each method performs under varying anatomical and data conditions.

3.1 Performance of Procrustes Algorithm

To evaluate the performance of classical rigid registration methods, we applied the Procrustes algorithm to manually annotated 3D dental point clouds using eight different landmark selection strategies (Thea_01 to Thea_08), with the complete 116-point set used as a baseline. These strategies (see **Appendix A**) were constructed by selectively excluding landmarks from anatomical regions such as the gingival junction, occlusal surface, or posterior molars, in order to analyze the contribution of different structures to registration accuracy.

We used Mean Squared Error (MSE, in mm²) and Chamfer Distance (CD, in mm) to assess registration error before and after alignment. As shown in Table 1, all strategies achieved a significant reduction in landmark-based error (with pre-alignment MSEs above 1000 and post-alignment MSEs generally below 10), indicating that the Procrustes algorithm provides stable and effective alignment for rigid point-based registration.

Procrustes	Selection Strategy	MSE				Chamfer Distance	
		On landmarks		On meshes			
		Original	Transformed	Original	Transformed	Original	Transformed
A1	116 Landmarks	1057.491789	70.288451	284.783674	0.260321	122676122.1	382053.5117
	Thea_01	1090.785176	85.049766		0.393316		462526.7048
	Thea_02	983.940069	6.999802		0.066684		263734.9832
	Thea_03	1005.098441	9.348067		0.055408		281416.2232
	Thea_04	982.421944	3.335057		0.087575		260070.5624
	Thea_05	1018.688897	5.105764		0.059716		274501.4936
	Thea_06	1031.520097	7.481238		0.059102		282058.7768
	Thea_07	993.026497	0.312319		0.082786		268696.4898
	Thea_08	988.088348	0.270888		0.121223		271794.3087
	A28	963.41453	126.346371	298.316506	2.984646	88788880.4	1043625.949
	A30	1135.36406	0.175575	459.384635	0.174814	140635710.6	123463.1811
	A32	325.799037	0.183023	116.957962	0.145693	44674259.18	426464.5959
	A35	837.383366	0.234234	312.771307	0.40447	102189198.5	171140.9805
	A37	774.57795	0.271899	190.682994	0.361113	67776278.86	188626.5454

Table 1. Results of Procrustes method

*All reported Mean Squared Error (MSE) values are in mm², while Chamfer Distance (CD) values are in mm.

Among all strategies, the full 116-point set served as a stable baseline. Removing gingival junction (Thea_01) or occlusal surface points (Thea_02) slightly increased RMSE but barely affected Chamfer Distance, suggesting these regions contribute less to geometric alignment. Thea_03, which removed both, caused a moderate rise in both metrics, indicating loss of key constraints.

Notably, Thea_08 achieved the lowest MSE (0.270 mm) using only 42 landmarks, demonstrating excellent local alignment. Overall, Thea_02 (Fig. 4), Thea_04 (Fig. 5), and Thea_08 (Fig. 6), with 114, 70, and 42 landmarks respectively, struck a balance between accuracy and annotation cost. Thea_06, with only two landmarks per tooth, had the highest error due to inadequate spatial constraints.

Thea_04 and Thea_05 showed degraded performance when posterior molars were excluded, underscoring their importance. Thea_06 and Thea_07 further illustrate that

sparse landmark configurations lead to reduced registration accuracy.

Post-registration performance on meshes level is particularly important, as it reflects the algorithm's ability to align the overall dental arch geometry. Thea_02 and Thea_04 achieved a more balanced trade-off among landmark accuracy, mesh-level precision, and global Chamfer Distance, making them more suitable for robust alignment. Although the Procrustes algorithm generally demonstrated stable performance across most cases, its notably poor results on sample A28 revealed its strong dependence on the quality and density of the landmark set. This also highlights its sensitivity to outliers, limited use of contextual or structural information, and inherent inability to directly handle dense point clouds.

Based on these findings, we focus our subsequent comparisons on Thea_02, Thea_04, and Thea_08 Procrustes strategies against learning-based registration models.

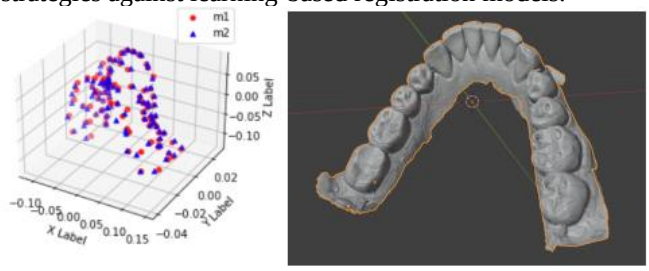

Fig. 4. Registration result using Thea_02 (114 landmarks)
* Removal of central points on the occlusal surface

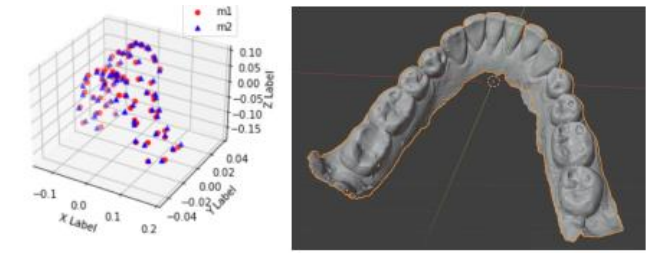

Fig. 5. Registration result using Thea_04 (70 landmarks)
* Removal of occlusal surface points and additional four posterior molar landmarks while retaining two points along the alignment direction

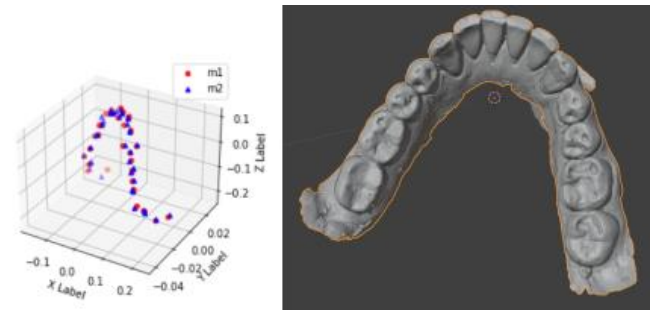

Fig. 6. Registration result using Thea_08 (70 landmarks)
*Retains only outer central landmarks while adding the midpoints on the left and right sides of the teeth.

These results suggest that the density and anatomical distribution of landmarks play a critical role in determining registration accuracy. While the Procrustes method is capable of fitting well under ideal conditions, its performance remains highly sensitive to the chosen landmark strategy when applied to clinically diverse data. Therefore, designing a reasonable annotation scheme is essential for ensuring the robustness and generalizability of classical registration methods.

3.2 Performance of Learning-Based Methods

After completing the benchmark evaluation of the traditional method, we further assessed three learning-based point cloud registration models: MLP, PointNet++, and GCN. All models were trained and tested using a standardized input format and landmark annotation strategy, and were evaluated on the same dataset and with the same metrics as the Procrustes algorithm to ensure fair comparison. The analysis focused on their performance across different samples, consistency in accuracy, and generalization ability when facing complex dental structures.

Among the three models, PointNet++ consistently achieved the best performance in terms of both MSE and Chamfer Distance (see Table 2). For example, in sample A1, PointNet++ reduced the MSE from 284.78 to 1.13 and decreased the Chamfer Distance from over 122 million to 271,796—comparable to the Procrustes result and significantly outperforming both MLP and GCN. Similar trends were observed across other samples from A28 to A37, where PointNet++ maintained low error values even in anatomically complex cases.

	MSE					Chamfer distance				
	Original	Procrustes	MLP	GCN	PointNet	Original	Procrustes	MLP	GCN	PointNet
A1	284.783674	0.121223	0.108525	0.289925	1.130998	122676122.1	271794.3087	276156.6329	341684.7744	271795.6058
A28	298.316506	2.984646	2.702364	2.707173	0.580558	88788880.4	1043625.949	971862.3492	974246.8933	215659.4562
A30	459.384635	0.174814	0.341319	0.488649	0.174583	140635710.6	123463.1811	214803.8113	217341.4013	123512.8669
A32	116.957962	0.145693	0.463114	0.774355	2.026585	44674259.18	426464.5959	581108.4093	753770.2542	427535.8218
A35	312.771307	0.40447	0.867851	0.703831	0.586316	102189198.5	171140.9805	338638.2756	296735.7023	170595.0791
A37	190.682994	0.361113	0.384394	0.639277	0.637884	67776278.86	188626.5454	200043.0929	348963.6619	189025.2041

Table 2. Results of learning-based models

GCN showed moderate improvements, performing better than MLP in most cases, particularly in samples with greater geometric variation such as A30. Its edge-based message passing mechanism helped preserve local structural information more effectively, although its overall accuracy was still slightly behind PointNet++.

MLP, lacking the ability to model spatial structure, produced less stable results. While it performed reasonably well in simpler cases like A1, it showed limited generalizability in anatomically irregular samples such as A28 and A35, where error values fluctuated more significantly.

These findings are consistent with the theoretical design of each model—PointNet++'s hierarchical feature extraction enables stronger robustness and consistency across diverse

dental morphologies. As such, PointNet++ emerges as the most reliable learning-based registration method in this study.

3.3 Comparison Between Classical and Learning-Based Methods

In the previous two sections, we evaluated the performance of the Procrustes algorithm under various landmark selection strategies and compared it with three deep learning-based models using a unified annotation scheme. To better understand the strengths and weaknesses of both approaches, this section presents a direct comparison, focusing on accuracy, robustness, and adaptability to complex anatomical structures.

In terms of overall performance, PointNet++ demonstrated greater consistency and stability across different samples. Even in cases with fewer landmarks or more complex anatomical geometry, it maintained low registration errors. For example, in sample A35, which features irregular dental morphology, PointNet++ delivered strong error control, whereas the Procrustes method often exhibited local misalignment when landmarks were sparsely or unevenly distributed. In contrast, PointNet++ adapted more robustly to variations in landmark distribution and structural complexity, indicating stronger resilience and more predictable performance under real-world clinical conditions.

Additionally, the Procrustes method showed a strong dependence on the choice and distribution of landmarks, with its performance varying noticeably depending on anatomical regions and point density. Learning-based models, by contrast, learned spatial features directly from the data, making them less sensitive to manual landmark design and better equipped to generalize across diverse shapes and clinical presentations.

That said, Procrustes remains highly accurate under ideal conditions, especially when landmark quality is high, and its deterministic nature provides strong interpretability.
In summary, Procrustes is preferable when dealing with dense and consistently annotated landmark sets in well-defined anatomical regions. However, when facing more diverse clinical data—such as missing landmarks, structural anomalies, or patient variability—learning-based methods like PointNet++ offer superior performance.

3.4 Visual Validation and Surface Deviation Analysis

To complement the quantitative evaluation with visual and geometric insights, we imported all registration outputs into ***Geomagic Control*** for clinical-grade deviation analysis. This process assessed the surface alignment between the registered (new) scan and the baseline (old) scan across the entire dental arch. In the resulting deviation maps, yellow areas indicate regions where the new scan exceeds the reference surface, while blue areas denote surface depression relative to the baseline.

In a real clinical context, blue deviations may suggest tooth structure loss due to individual wear patterns, occlusal habits, or longer time intervals between scans. Conversely, yellow deviations may result from local skeletal deformation, alignment inaccuracy, or algorithmic overfitting during registration. As such, this analysis not only reflects the technical performance of the models but also provides clinically meaningful insights into underlying anatomical changes.

From the overall results, the PointNet++ model achieved the most accurate and geometrically consistent registration across most samples (see Fig. 7, Table 3). For example, in sample A28, it produced an average deviation of just 0.0029 mm and an RMS error of 0.1922—the lowest among all methods—with deviations tightly distributed around zero, indicating strong global and local alignment performance.

Case Number	Max Distance (+)	Max Distance (-)	Average Distance	Avg Dist (+)	Avg Dist (-)	Standard Deviation	RMS Estimate
A01	3.1511	-2.3075	0.0978	0.2515	-0.1803	0.2714	0.2884
A28	1.3430	-1.3202	0.0029	0.1404	-0.1729	0.1900	0.1922
A30	3.0162	-2.7574	0.0039	0.1451	-0.1405	0.1862	0.1862
A32	3.2342	-1.7614	-0.0547	0.1083	-0.1559	0.1795	0.1876
A35	3.2731	-2.0407	0.0382	0.1453	-0.1153	0.1665	0.1709
A37	2.2082	-1.6547	0.0589	0.1391	-0.0991	0.1521	0.1631

Table 3. Geomagic surface deviation metrics for PointNet++

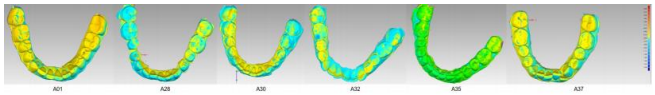

Fig. 7. Surface deviation map of PointNet++

In contrast, MLP (see Fig. 8, Table 4) and GCN (see Fig. 9, Table 5) exhibited larger average deviations and greater variability in standard deviation, particularly in anatomically complex samples such as A28 and A32. While GCN benefits from edge-based message passing to enhance local structural continuity, its overall accuracy still falls short of PointNet++.

Case Number	Max Distance (+)	Max Distance (-)	Average Distance	Avg Dist (+)	Avg Dist (-)	Standard Deviation	RMS Estimate
A01	1.8862	-0.9339	0.1175	0.2387	-0.1135	0.2569	0.0.2825
A28	3.2297	-3.3000	0.1950	1.1492	-1.0167	1.3478	1.3618
A30	2.9304	-2.3300	0.1895	0.4644	-0.2354	0.4573	0.4938
A32	2.5722	-3.1845	-0.2309	0.2718	-0.5410	0.5434	0.5904
A35	3.3341	-3.4654	0.0314	0.4854	-0.48191	0.6302	0.6310
A37	1.2873	-1.8093	-0.0653	0.1872	-0.1429	0.1973	0.2078

Table 4. Geomagic surface deviation metrics for MLP

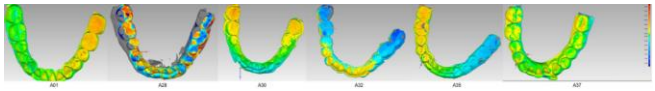

Fig. 8. Surface deviation map of MLP

Case Number	Max Distance (+)	Max Distance (-)	Average Distance	Avg Dist (+)	Avg Dist (-)	Standard Deviation	RMS Estimate
A01	3.1324	-1.4729	0.1520	0.4023	-0.2704	0.4319	0.4578
A28	3.1900	-3.1900	0.0725	1.2464	-1.2141	1.4829	1.4847
A30	2.7462	-3.1385	0.2012	0.3702	-0.6314	0.6835	0.7125
A32	3.3901	-3.3832	-0.1324	0.4890	-0.7360	0.8453	0.8684
A35	3.6100	-3.6266	0.0670	0.4900	-0.3251	0.5303	0.5345
A37	2.4882	-3.0499	-0.1008	0.3190	-0.4952	0.5571	0.5662

Table 5. Geomagic surface deviation metrics for GCN

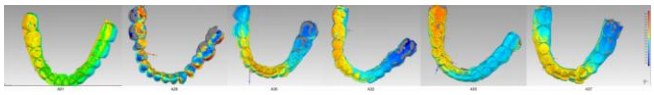

Fig. 9. Surface deviation map of GCN

The Procrustes method performed well in certain strategies (e.g., A30, A32; see Fig. 10, Table 6), but showed less stability across samples. Due to its reliance on a limited set of discrete anatomical landmarks, it struggled in cases where landmarks were sparse or unevenly distributed—often resulting in noticeable misalignment in the posterior region or occlusal surfaces. For instance, in sample A1, although Thea_08 achieved a relatively low RMS (0.2696), the actual alignment outcome was poor.

Case Number	Max Distance (+)	Max Distance (-)	Average Distance	Avg Dist (+)	Avg Dist (-)	Standard Deviation	RMS Estimate
A1_02	1.6378	-2.9855	0.0590	0.1467	-0.1102	0.1638	0.1741
A1_04	2.6135	-2.9148	0.1167	0.2025	-0.1258	0.2037	0.2348
A1_08	2.4203	-2.9629	0.1075	0.2469	-0.1588	0.2472	0.2696
A28	3.1000	-3.1000	0.0086	1.1710	-1.1539	1.4194	1.4237
A30	1.5816	-1.3950	0.0011	0.1420	-0.1455	0.1773	0.1733
A32	2.9962	-1.9530	-0.0557	0.1061	-0.1544	0.1764	0.1850
A35	3.3406	-2.3117	0.0397	0.1445	-0.1156	0.1632	0.1679
A37	2.9077	-2.3252	0.0650	0.1367	-0.0912	0.1437	0.1577

Table 6. Geomagic surface deviation metrics for Procrustes method

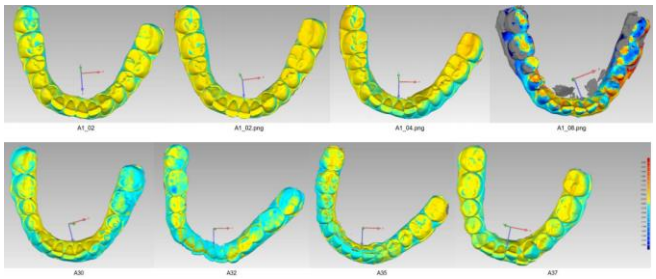

Fig. 10. Surface deviation map of Procrustes method

The figures presented highlight typical surface deviation maps for four registration models across selected samples. PointNet++ demonstrated the most compact and consistent error bands, followed by carefully curated Procrustes strategies. In contrast, MLP and GCN showed wider error ranges, with irregular deviation patterns particularly evident in anatomically critical areas, indicating limitations in their local registration capabilities.

In summary, the results of this section further confirm the advantages of deep learning models, particularly PointNet++, in capturing complex geometric structures. At the same time, they reaffirm that with high-quality annotations, the Procrustes algorithm can still provide interpretable and accurate alignment in localized regions.

4. EQUALITY, DIVERSITY AND INCLUSION

This project actively considered the principles of Equality, Diversity, and Inclusion (EDI) at both the team and technical levels. As a team with diverse academic and cultural backgrounds, we maintained inclusive communication and allocated tasks according to individual strengths, ensuring that every member could participate meaningfully.

On the technical side, we used real intraoral scan data collected from different patients across multiple time points, which inherently captured anatomical variations due to differences in age, sex, and dental condition. Although the dataset did not include explicit labels for ethnicity or socio-economic status, our evaluation of landmark selection strategies and registration robustness reflects an awareness of the importance of individual variation in the design of medical tools.

We also recognized that dental morphology can vary significantly across individuals, and registration methods should not assume anatomical uniformity. To address this, we evaluated the impact of different landmark configurations on alignment performance, aiming to improve model adaptability in diverse clinical contexts.

Finally, all code and data from this project have been openly released to support equitable access for researchers from different regions and institutions, contributing to a more inclusive research ecosystem.

5. CONCLUSIONS

This project focused on rigid 3D point cloud registration for high-resolution dental scans, systematically evaluating the performance of both classical geometric algorithms and modern deep learning models under varying landmark strategies and anatomical complexities. We implemented and assessed Procrustes analysis, Multi-Layer Perceptrons (MLP), PointNet++, and Graph Convolutional Networks (GCN), using metrics such as Mean Squared Error (MSE), Chamfer Distance, and 3D surface deviation visualization for comprehensive comparison.

The results demonstrated that PointNet++ consistently achieved the best registration accuracy and structural consistency across most tested samples. It maintained reliable performance even in challenging cases with sparse annotations or complex morphology, highlighting its strong generalization capability and robustness. In contrast, the

traditional Procrustes method, while stable under ideal landmark conditions, showed significant sensitivity to landmark density and spatial distribution, limiting its applicability to more diverse clinical data.

Through comparative analysis of eight landmark selection strategies, we also identified the anatomical contribution to registration accuracy. Specifically, Thea_02, Thea_04, and Thea_08 achieved a practical balance between precision and annotation effort. Additionally, clinical surface deviation analysis using Geomagic Control provided an intuitive validation of each model's alignment quality at the dental arch level, complementing traditional numerical metrics.

In terms of individual contributions, I led the annotation of 3D dental point clouds, implemented the initial Procrustes registration, designed and trained the GCN model, developed the error evaluation pipeline, and compiled all results and visualizations. This end-to-end workflow—from data preprocessing to algorithmic analysis—ensured both the reproducibility and reliability of our results.

Overall, this project offers a viable technical framework for automated registration in digital dentistry, and establishes a benchmarking foundation for future research in multimodal medical imaging, point cloud learning, and clinically-assisted diagnostics.

6. REFERENCES

[1] X. Huang, G. Mei, J. Zhang, and R. Abbas, 'A comprehensive survey on point cloud registration', Mar. 05, 2021, *arXiv*: arXiv:2103.02690. doi: 10.48550/arXiv.2103.02690.

[2] S. Monji-Azad, J. Hesser, and N. Löw, 'A review of non-rigid transformations and learning-based 3D point cloud registration methods', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 58–72, Feb. 2023, doi: 10.1016/j.isprsjprs.2022.12.023.

[3] Y. Zhou, L. Yuan, Y. Li, and J. Yu, 'Digital Dental Biometrics for Human Identification Based on Automated 3D Point Cloud Feature Extraction and Registration', *Bioengineering*, vol. 11, no. 9, p. 873, Aug. 2024, doi: 10.3390/bioengineering11090873.

[4] D. Wu, J. Jiang, J. Wang, S. Zhou, and K. Qian, 'Accuracy evaluation of dental CBCT and scanned model registration method based on pulp horn mapping surface: an in vitro proof-of-concept', *BMC Oral Health*, vol. 24, no. 1, p. 827, Jul. 2024, doi: 10.1186/s12903-024-04565-3.

[5] W. W. Tian, C. J. Tian, and L. Han, 'An improved ICP algorithm based on Generalized Procrustes Analysis', in *Proceedings of the 2016 3rd International Conference on Mechatronics and Information Technology*, Shenzhen, China,: Atlantis Press, 2016. doi: 10.2991/icmit-16.2016.160.

[6] M. Liu, X. Li, J. Liu, W. Liu, and Z. Yu, 'TUCNet: A channel and spatial attention-based graph convolutional network for teeth upsampling and completion', *Computers in Biology and Medicine*, vol. 166, p. 107519, Nov. 2023, doi: 10.1016/j.compbiomed.2023.107519.

[7] T. Wu, L. Pan, J. Zhang, T. Wang, Z. Liu, and D. Lin, 'Density-aware Chamfer Distance as a Comprehensive Metric for Point Cloud Completion', 2021, *arXiv*. doi: 10.48550/ARXIV.2111.12702.

[8] A. Ben-Hamadou *et al.*, 'Teeth3DS+: An Extended Benchmark for Intraoral 3D Scans Analysis', Nov. 11, 2024, *arXiv*: arXiv:2210.06094. doi: 10.48550/arXiv.2210.06094.

[9] M. Rudemo, 'Statistical Shape Analysis. I. L. Dryden and K. V. Mardia, Wiley, Chichester 1998. No. of pages: xvii+347. Price: £60.00.ISBN 0-471-95816-6', *Statist. Med.*, vol. 19, no. 19, pp. 2716–2717, Oct. 2000, doi: 10.1002/1097-0258(20001015)19:19<2716::AID-SIM590>3.0.CO;2-O.

[10] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, 'Dynamic Graph CNN for Learning on Point Clouds', Jun. 11, 2019, *arXiv*: arXiv:1801.07829. doi: 10.48550/arXiv.1801.07829.

[11] K. S. Arun, T. S. Huang, and S. D. Blostein, 'Least-Squares Fitting of Two 3-D Point Sets', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987, doi: 10.1109/TPAMI.1987.4767965.

[12] P. J. Besl and N. D. McKay, 'A method for registration of 3-D shapes', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992, doi: 10.1109/34.121791.

[13] Q.-Y. Zhou, J. Park, and V. Koltun, 'Open3D: A Modern Library for 3D Data Processing', Jan. 30, 2018, *arXiv*: arXiv:1801.09847. doi: 10.48550/arXiv.1801.09847.

[14] 'Geomagic Control X'. [Online]. Available: https://oqton.com/geomagic-controlx/

APPENDICES

Strategy	Description
Thea_01	Removes all gingival junction landmarks while retaining most tooth structure points.
Thea_02	Removes all occlusal surface landmarks; tests the impact of excluding biting surface features.
Thea_03	Combines Thea_01 and Thea_02; removes both gingival and occlusal points.
Thea_04	Further removes posterior molar points; retains only two alignment points per tooth.
Thea_05	Builds on Thea_03 and Thea_04; keeps only directional alignment points.
Thea_06	Most sparse: removes nearly all non-essential points, keeping only two per tooth.
Thea_07	Retains only outer central landmarks; removes all other structural points.
Thea_08	Based on Thea_07, adds left-right midpoints for each tooth to enhance symmetry representation.

Appendix A. Summary of Landmark Selection Strategies (Thea_01 – Thea_08)