浙江工业大学
ZHEJIANG UNIVERSITY OF TECHNOLOGY

计算机科学与技术学院、软件学院
College of Computer Science,Technology and Software

Transformer-based Tooth Alignment Prediction with Occlusion and Collision Constraints

Author：Zhenxing Dong，Jiazhou Chen

2025.10

CONTENT

CONTENT

01.

Background and Significance

Background and Significance

02.

Related Work

Related Work

03.

Prediction of Tooth Arrangement

Prediction of Tooth Arrangement

04.

Summary and Prospect

Summary and Prospect

01 Background and Significance

Background and Significance

01 Background and Significance

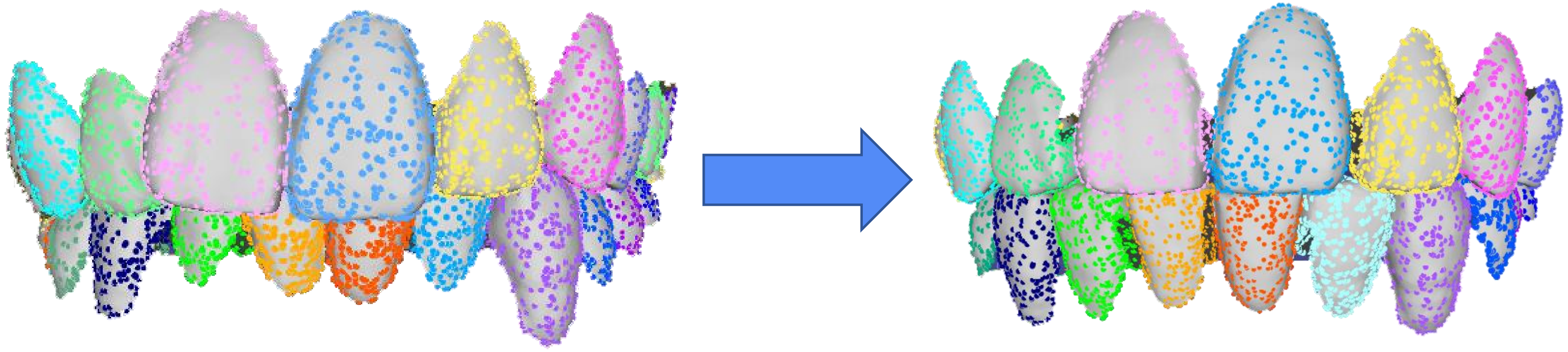
The cost of treatment in the field of stomatology has consistently remained high, with labor costs constituting a significant portion of orthodontic treatment procedures.

The dental medicine sector is undergoing industrial transformation and upgrading through digitalization. Many tasks previously reliant on manual labor can now be accomplished using computer technologies such as digital modeling and neural networks.

However, certain responsibilities, such as selecting orthodontic treatment plans for patients, still require dentists to apply professional knowledge and experience through careful observation.

01 Background and Significance

- This work aims to design a neural network architecture that takes intraoral scanner-derived segmented patient dental models as input
- Predict post-orthodontic treatment outcomes of fully aligned teeth.
- The system assists orthodontists in clinical decision-making for treatment planning while maximizing prediction accuracy.



02 Related Work

Related Work

02 Related Work

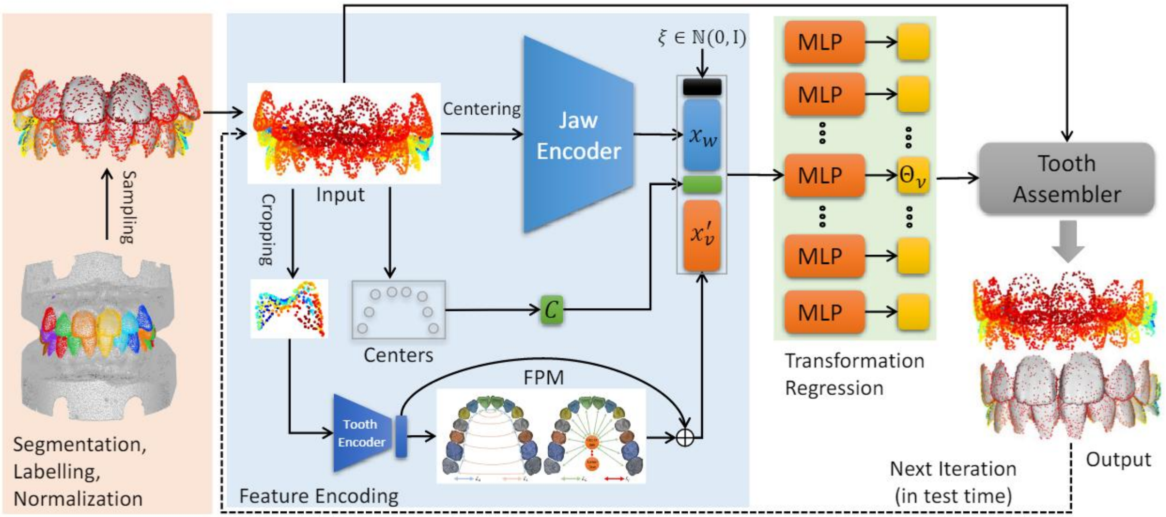
TANet(Graph Neural Network+PointNet)

Uses PointNet to encode point cloud features from intraoral scanner segmented models, including both global and local features.

Employs Graph Neural Networks to achieve connectivity and communication between local tooth features.

Incorporates tooth center points and positional encoding.

Utilizes an MLP decoder to regress 6-DoF information of teeth.



Limitations:

Unsatisfactory alignment prediction outcomes in complex cases.

Loss function fails to fully exert its intended effect.

PointNet encoder exhibits deficiencies in extracting local features.

input

predict

ground truth

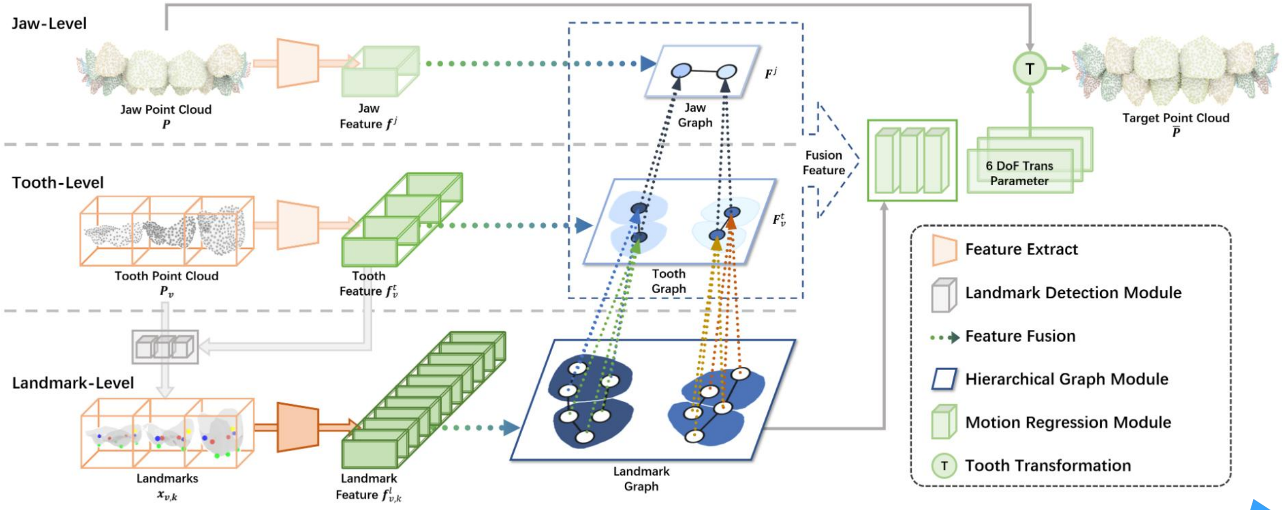
02 Related Work

TANet-Landmark Constraints and Hierarchical Graph Structure

Enhances TANet by computing four types of dental landmarks as key tooth features.

Uses DGCNN to extract point cloud features and constructs a three-level hierarchical graph neural network (Landmark → Tooth → Jaw) for bottom-up feature propagation.

Employs an MLP decoder to regress orthodontic transformation parameters.

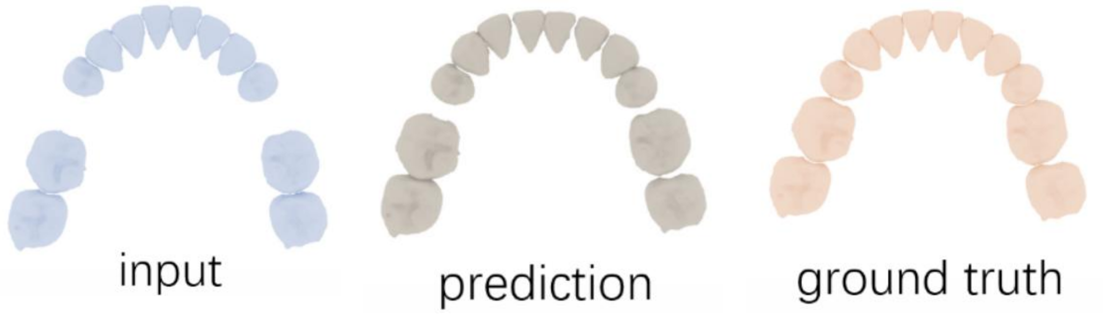


Limitations:

Performance heavily relies on landmark detection accuracy, which depends on the completeness of intraoral scans and segmentation.

Vulnerable to ambiguous landmark localization, significantly affecting results.

Complex architecture leads to long computational runtime.



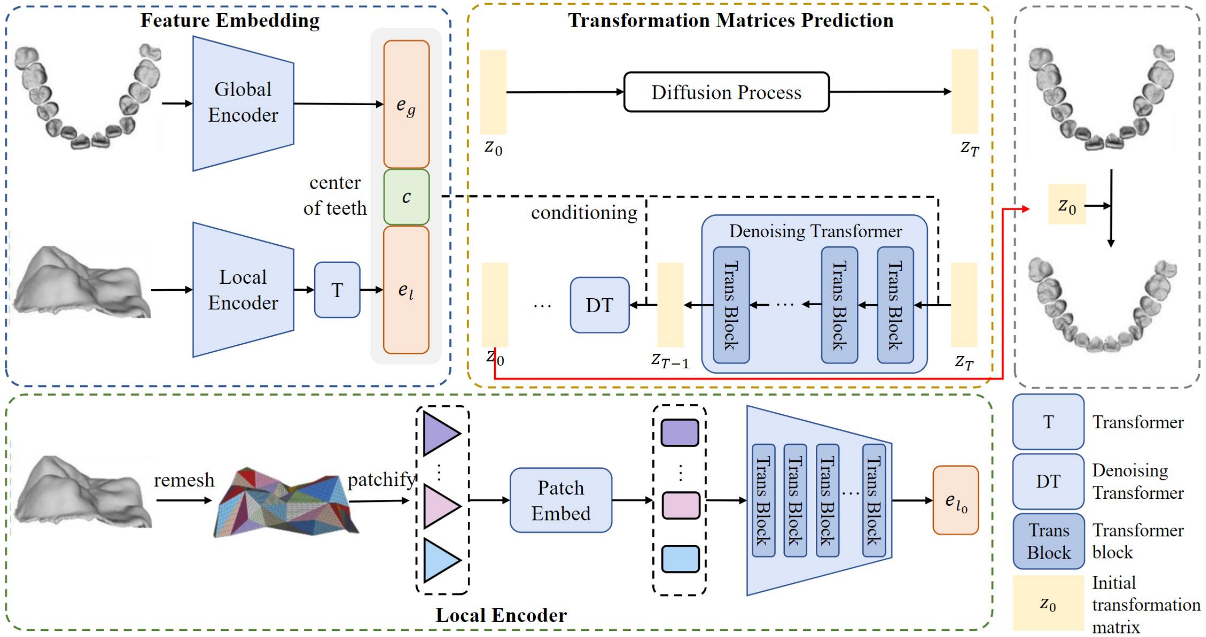
02 Related Work

TADPM(Diffusion+Mesh-MSE)

Employs diffusion probabilistic models to learn the transformation matrix distribution from malocclusion to normal occlusion through a gradual denoising process of random variables.

Utilizes standard self-attention modules for intermediate feature propagation.

Final orthodontic transformation parameters are regressed via MLP.



Limitations:

The multi-stage denoising structure leads to model complexity, requiring extensive computational resources for both training and inference.

Both MSE preprocessing and TADPM single-round training are time-consuming, resulting in high computational costs.

input
predict
ground truth

02 Related Work

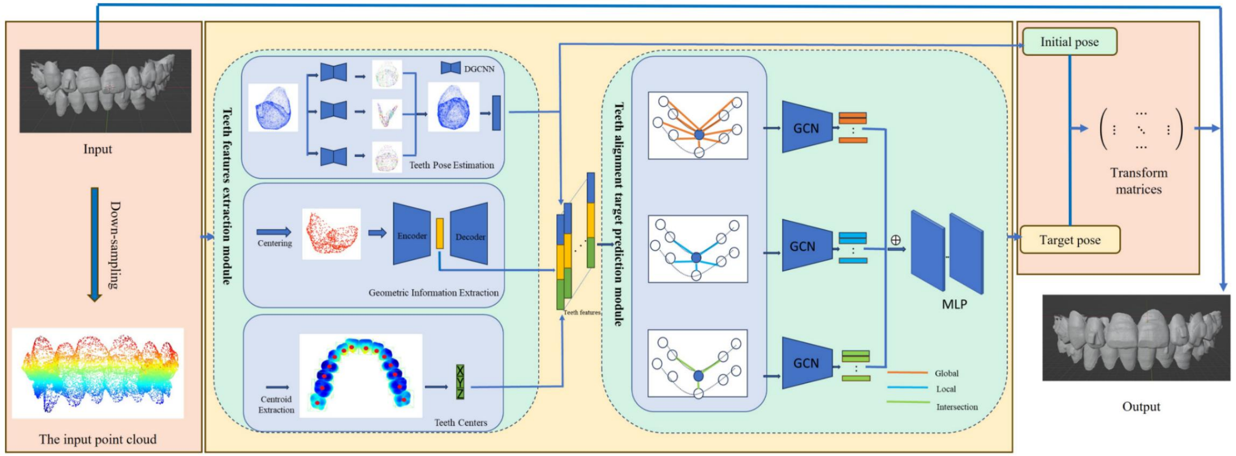
TAPoseNet(DGCNN+PointNet+GCN)

DGCNN for tooth pose learning

PointNet for local features + centroid features

Multi-scale GCN for spatial relationships

MLP regression

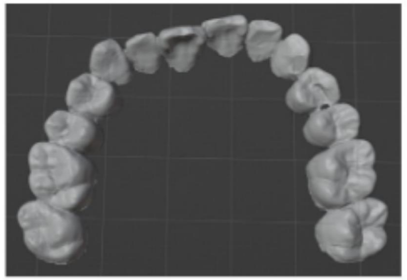


Limitations:

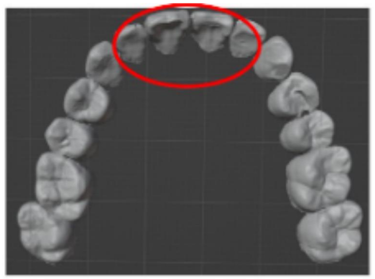
Small dataset with no augmentation

Oversimplified loss function

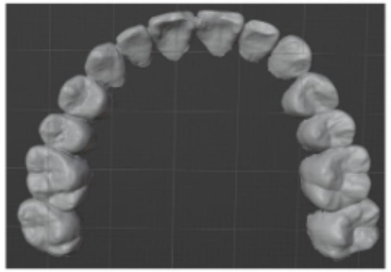
Dependent on tooth axis prediction



input



predict



ground truth

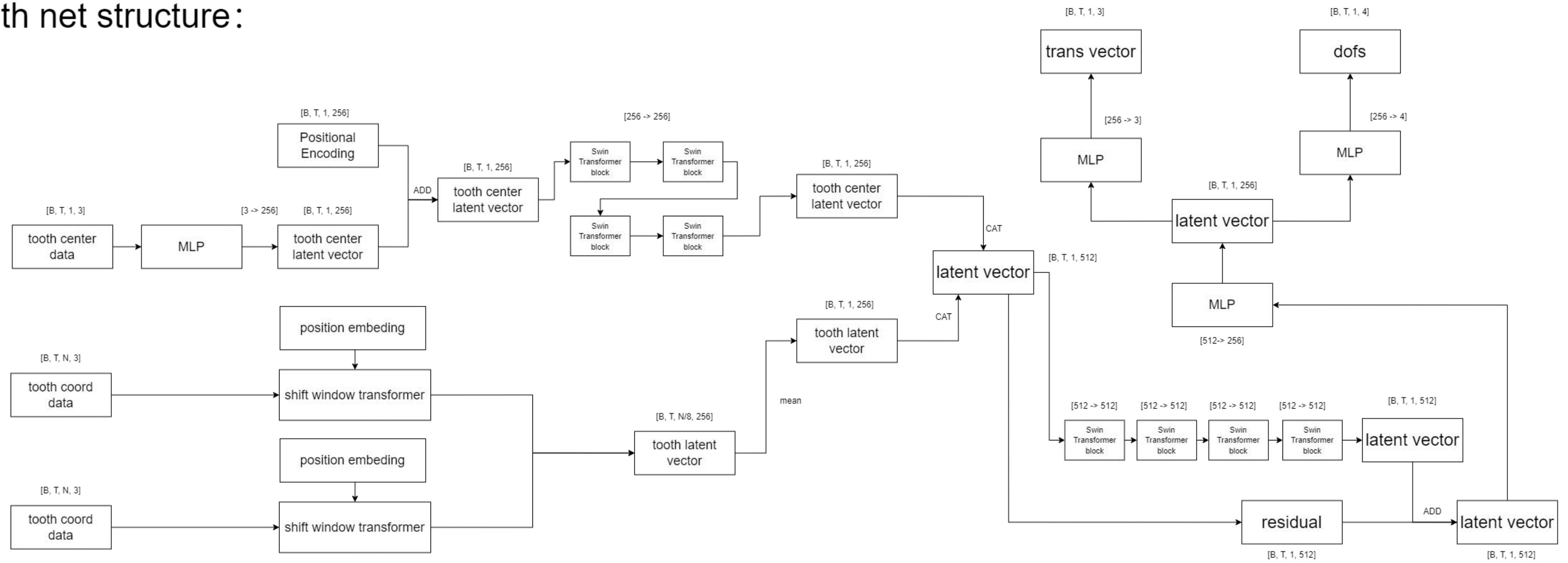
03 Prediction of Tooth Arrangement

Prediction of Tooth Arrangement

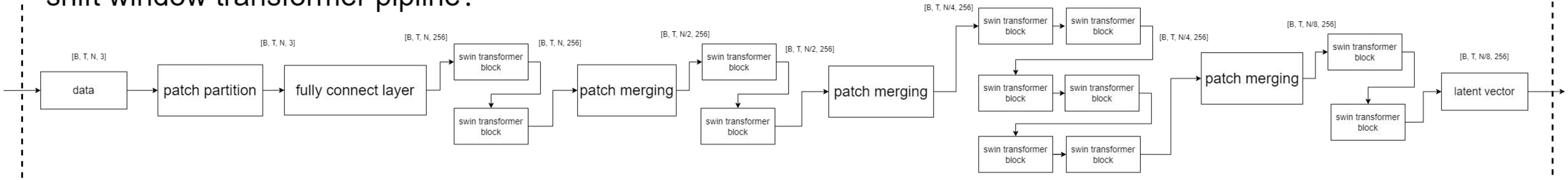
03 Prediction of Tooth Arrangement

Detail Network

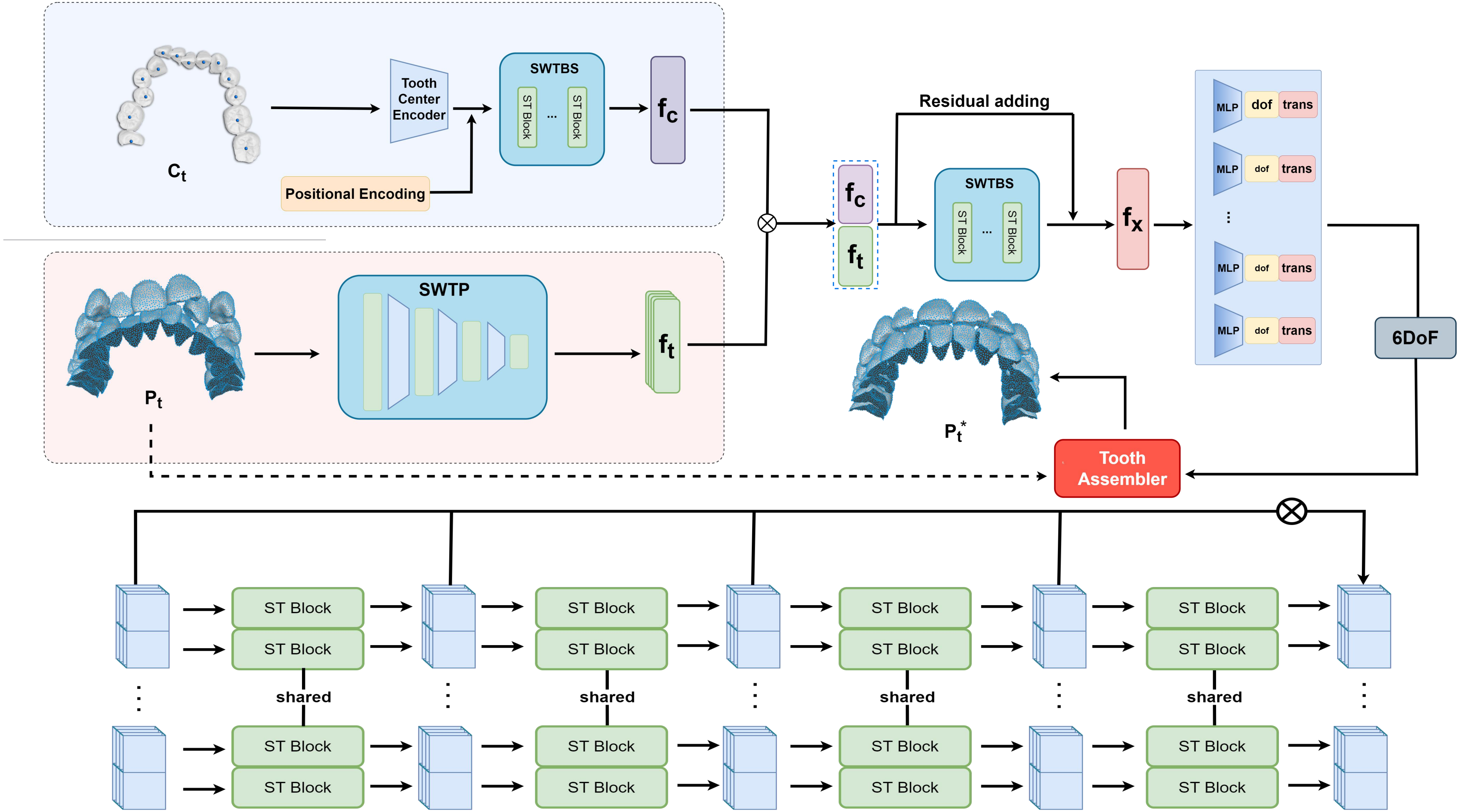
orth-tooth net structure：



shift window transformer pipline：



03 Prediction of Tooth Arrangement



03 Prediction of Tooth Arrangement

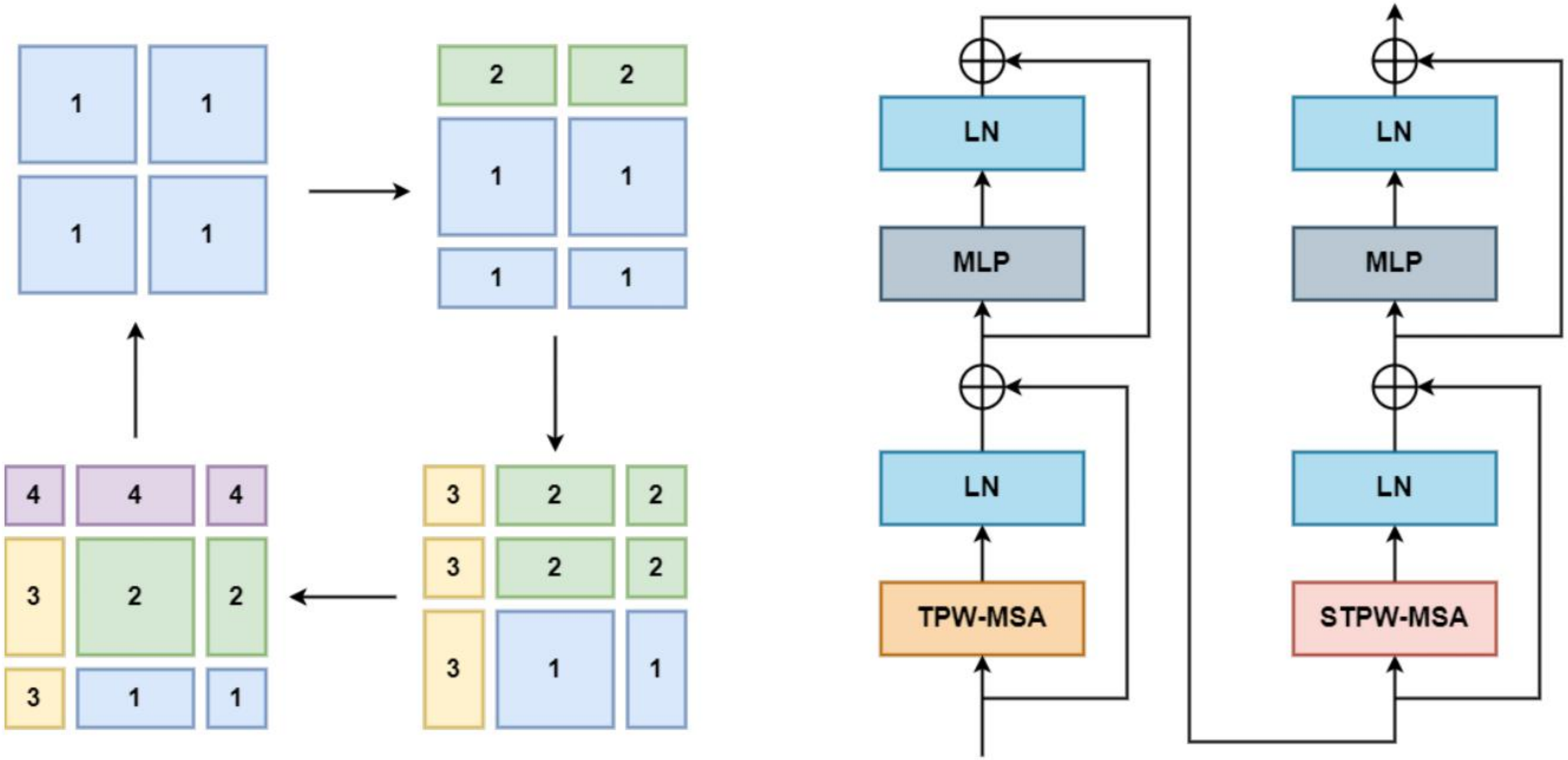


Figure 3-2. The evolved sliding window and the optimized ST Block structure after improvement

03 Prediction of Tooth Arrangement

Feature Extraction & Upsampling

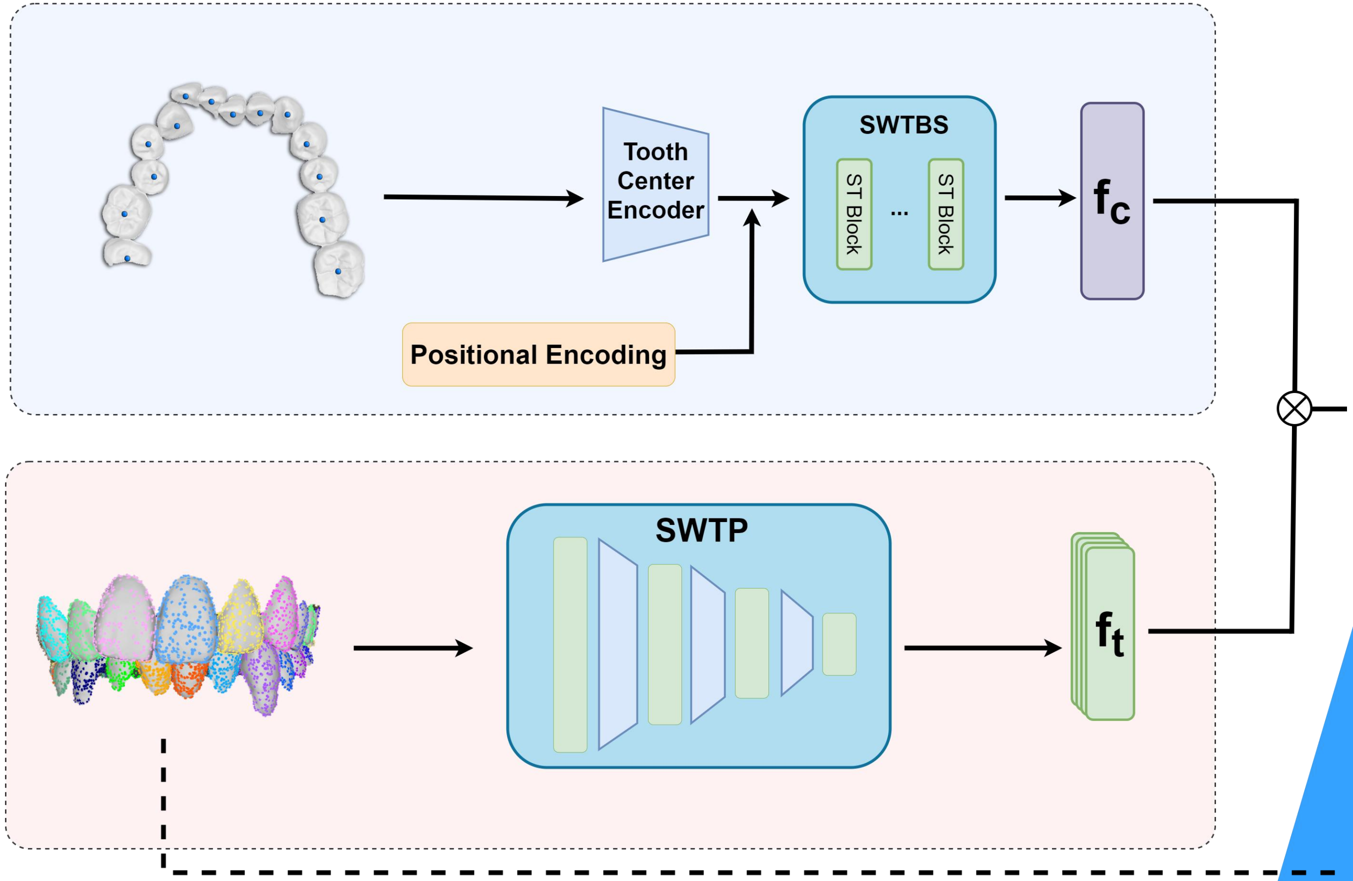
Dual feature extraction modules (global + local):

MHA layers with positional encoding for tooth centroids

Swin-T block sequence (SWTBS) for upsampling

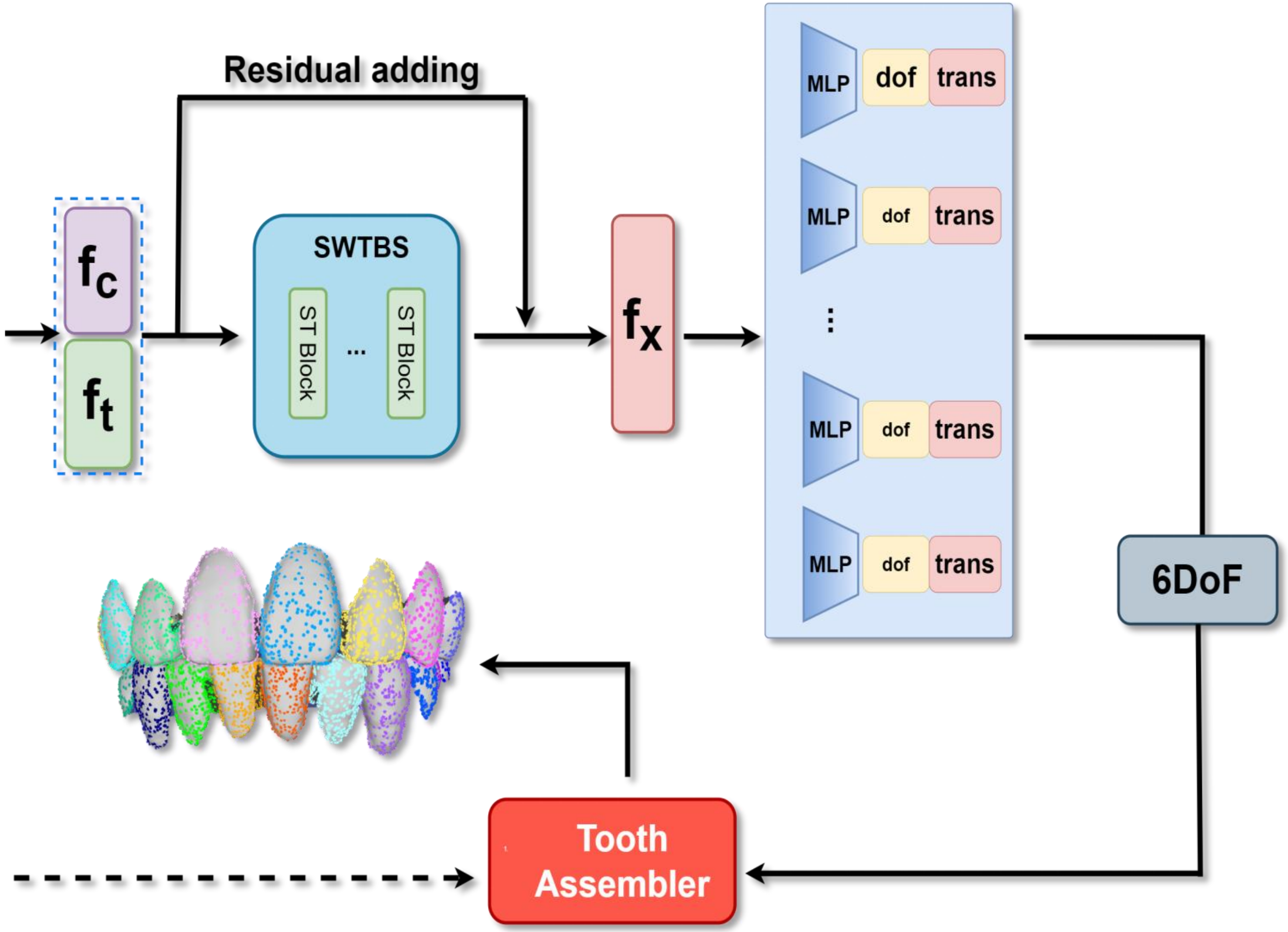
Swin Transformer Pipeline (SWTP) for 3D point clouds

Hierarchical downsampling via sliding windows



03 Prediction of Tooth Arrangement

全局和局部相结合的两个特征提取模块：

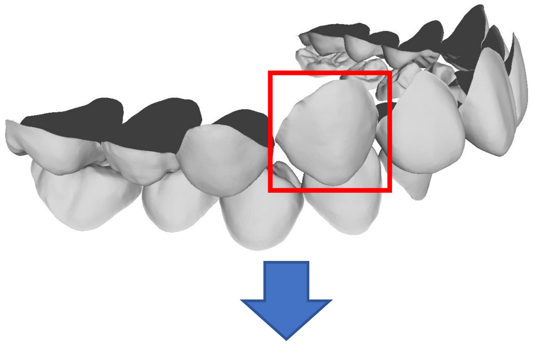


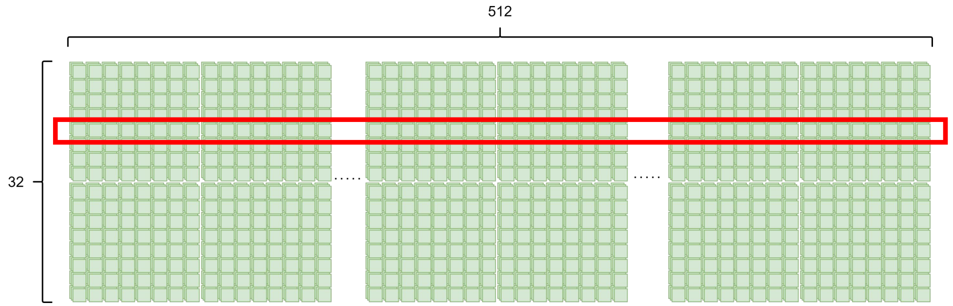
Feature Transfer & Parameter Regression

Tooth cloud features (f_t) merged with jaw features (f_c)

- Further feature exchange via SWTBS
- Residual collection for training optimization
- MLP decoder for 6-DoF parameter regression
- Tooth assembler module for final prediction

03 Prediction of Tooth Arrangement





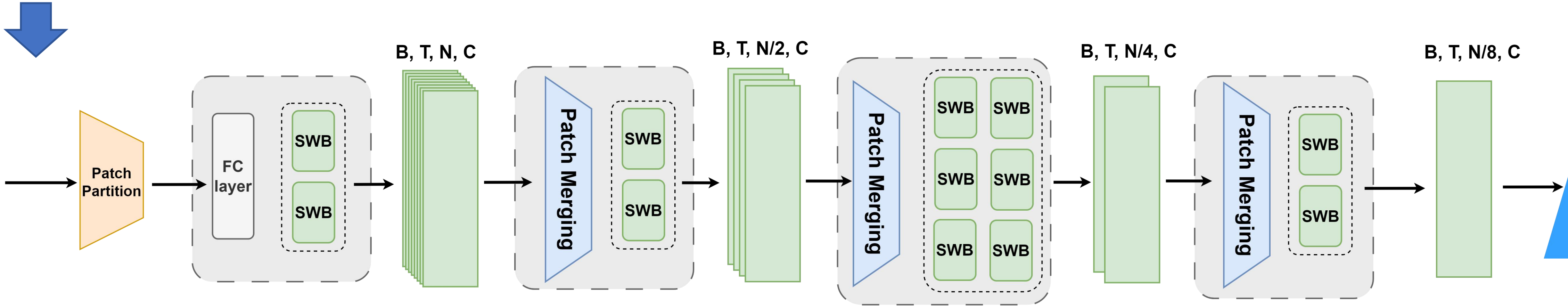
Local Feature Extraction

The maxillary and mandibular dental data each contain up to 16 teeth, with each tooth sampled to 512 points. Each point has XYZ coordinate values, forming a 32 * 512 matrix with 3 channels.

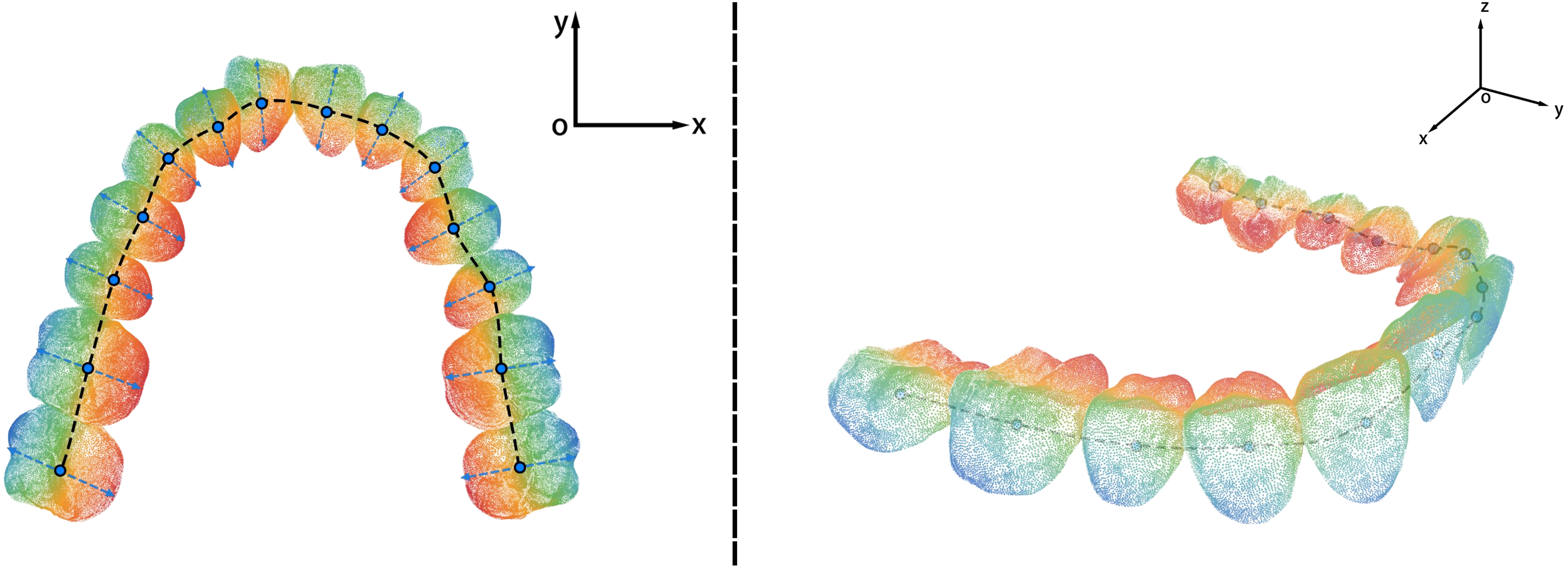
SWTP is used for local feature extraction of tooth point clouds, consisting of four stages. Each stage requires feature merging and downscaling.

During feature merging, only the data columns (the dimension of tooth point cloud count) are merged, while the data rows remain unmerged.

This is because the height dimension represents the number of teeth, and there is no shared feature space for merging across different teeth, as the final predictions of rotation and translation are performed individually for each tooth rather than collectively for multiple teeth.



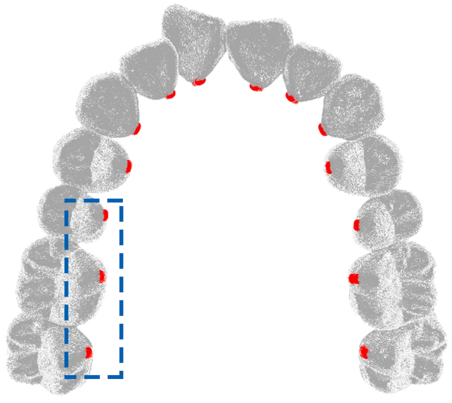
03 Prediction of Tooth Arrangement



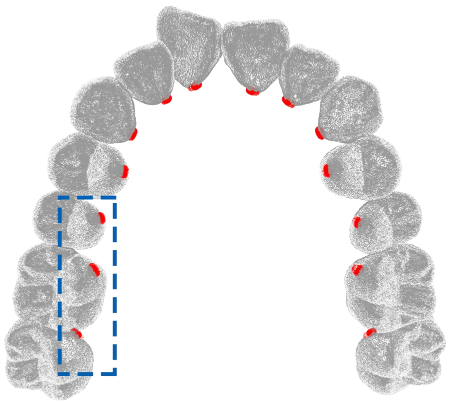
Tooth Arch-Based Point Cloud Serialization

- Hermite curve-interpolated arch line from centroids
- Point sorting by signed distance to arch (labial: +, lingual: -)
- Maintains consistent relative positions among 512 points/tooth

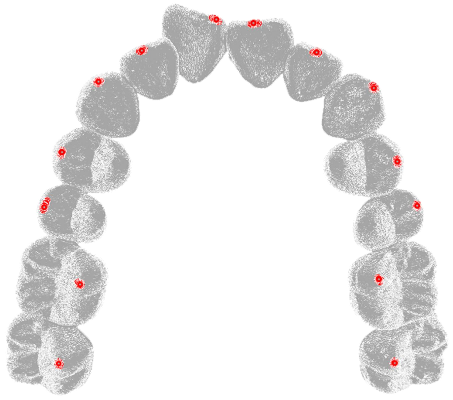
03 Prediction of Tooth Arrangement



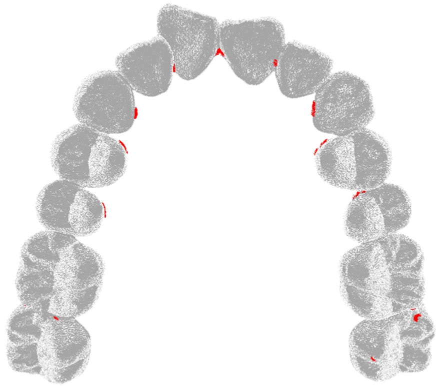
Based on dental arch core line



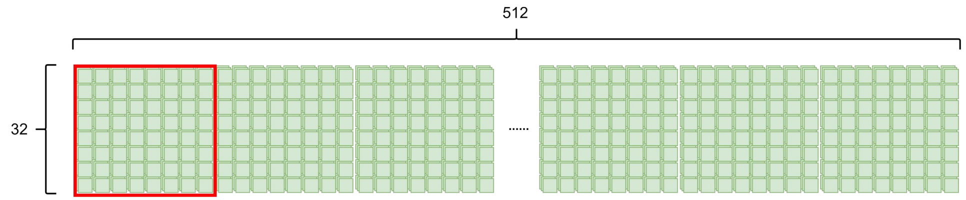
Based on dental arch center

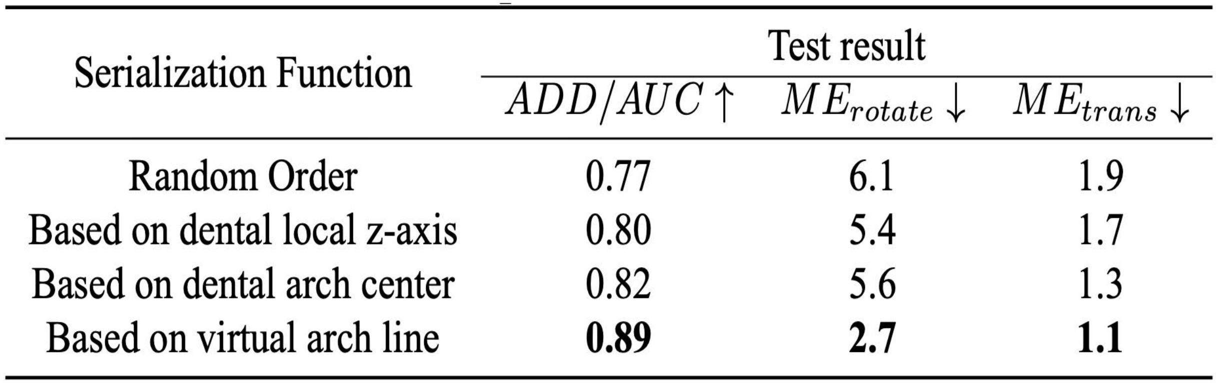


Based on distance to crown



Random order





Serialization Function	Test result		
	$ADD/AUC\uparrow$	$ME_{rotate}\downarrow$	$ME_{trans}\downarrow$
Random Order	0.77	6.1	1.9
Based on dental local z-axis	0.80	5.4	1.7
Based on dental arch center	0.82	5.6	1.3
Based on virtual arch line	**0.89**	**2.7**	**1.1**

Point Cloud Serialization Rules & Rationale

Uses n×n sampling windows (n teeth × n points/tooth)

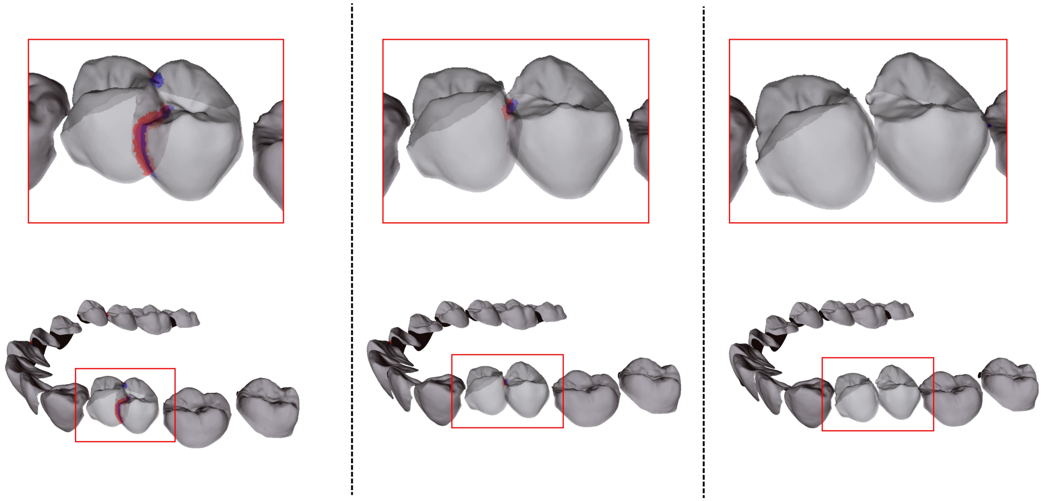
Sub-point clouds preserve global positional relationships

Sliding windows maintain inter-tooth spatial features

Ablation studies validate arch-line-based serialization superiority (see table)

03 Prediction of Tooth Arrangement

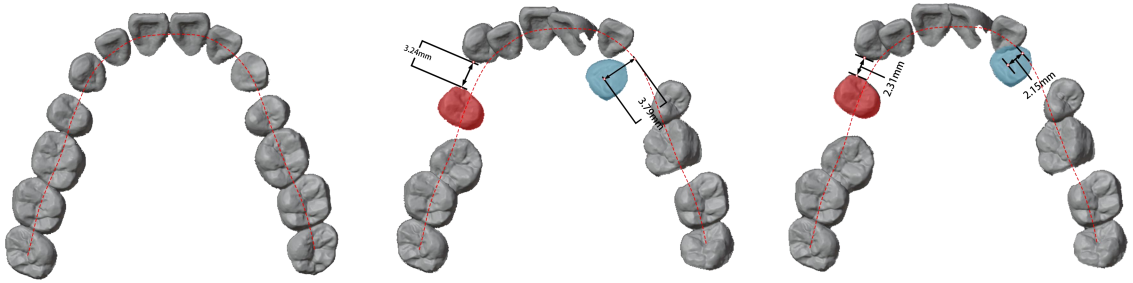
Two constraints



Uses BVH collision detection to check inter-tooth collisions

Employs simulated dental arch line as avoidance trajectory (preserves arch morphology + ensures normal transformation range)

Sequential processing: incisors → molars with iterative avoidance if constraints are violated



origin ground truth

after simple augmention

after constraint

Applies jaw regularization constraints via dental arch line during augmentation

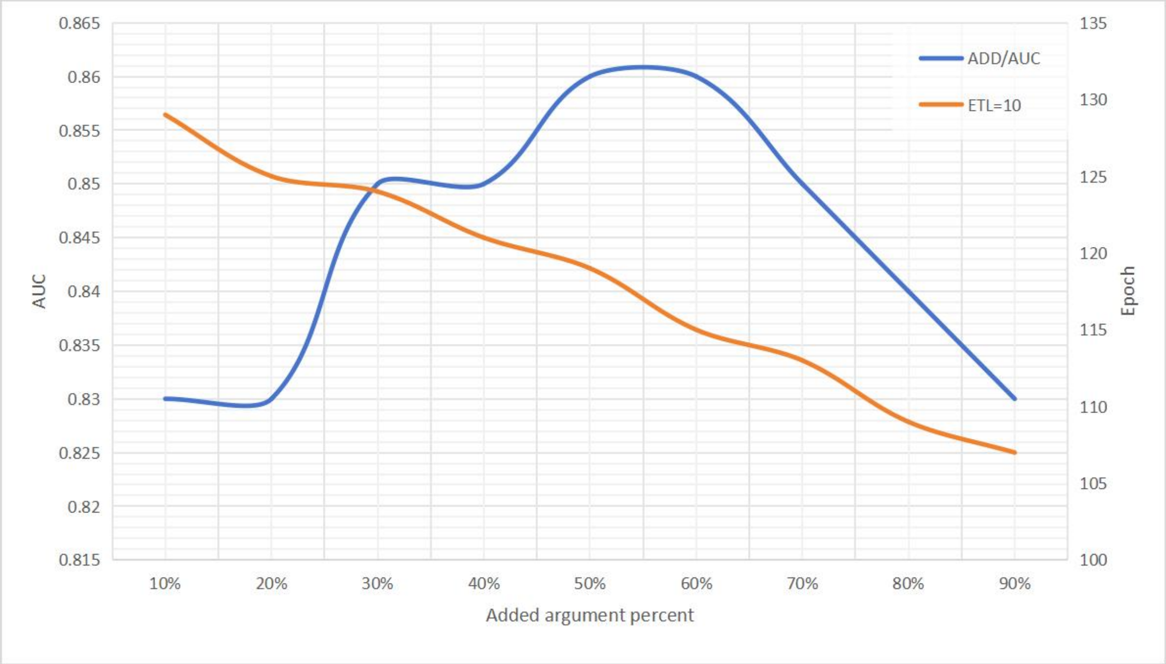
Pulls adjacent teeth closer if gap exceeds normal range (excluding missing teeth)

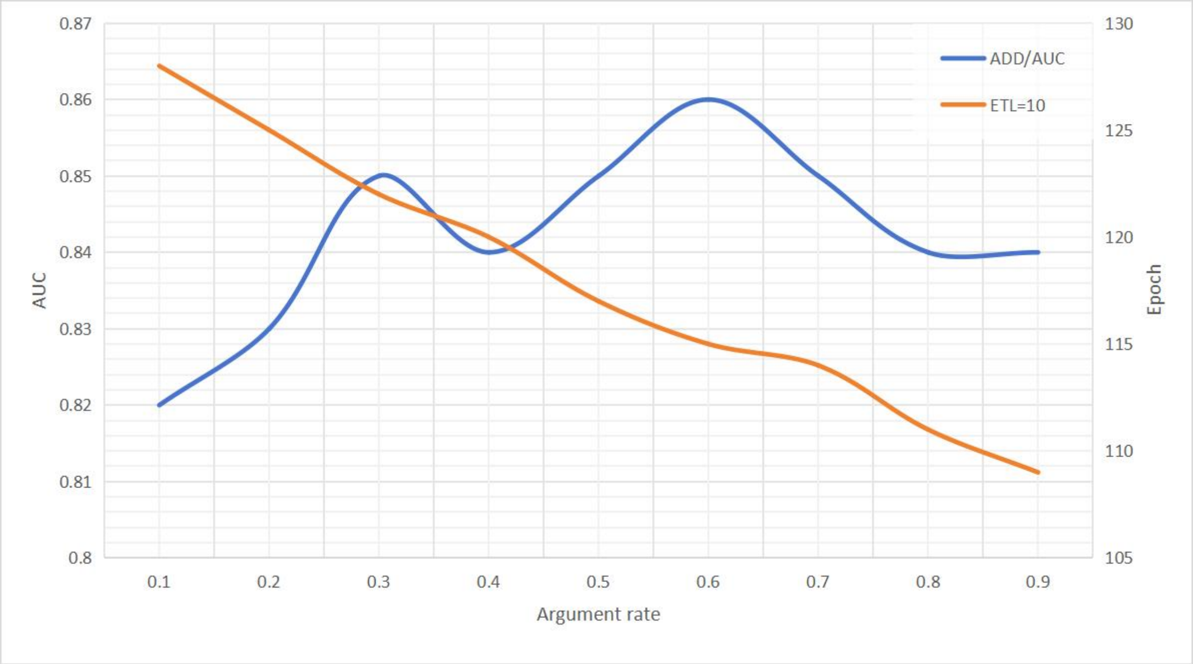
Repositions teeth beyond acceptable arch range inward to normal position

Sequential processing: incisors → molars

03 Prediction of Tooth Arrangement

Best Ratio to argument





Insufficient augmented data fails to fulfill the purpose of supplementing the dataset, but more augmentation is not always better. Excessively high ratio of augmented data reduces the network's exposure to real data, leading to network distortion.

To explore the optimal degree of data augmentation, we conducted ablation comparison experiments under consistent parameters including epochs, batch size, and test set, using two methods: standard augmentation (during training phase) and constrained augmentation (during pre-processing phase).

The left side shows constrained augmentation, where the independent variable is the ratio of augmented data to the original total data volume. The right side shows standard augmentation, where the independent variable is the triggering probability of augmentation during training. The dependent variable is the final test accuracy after training.

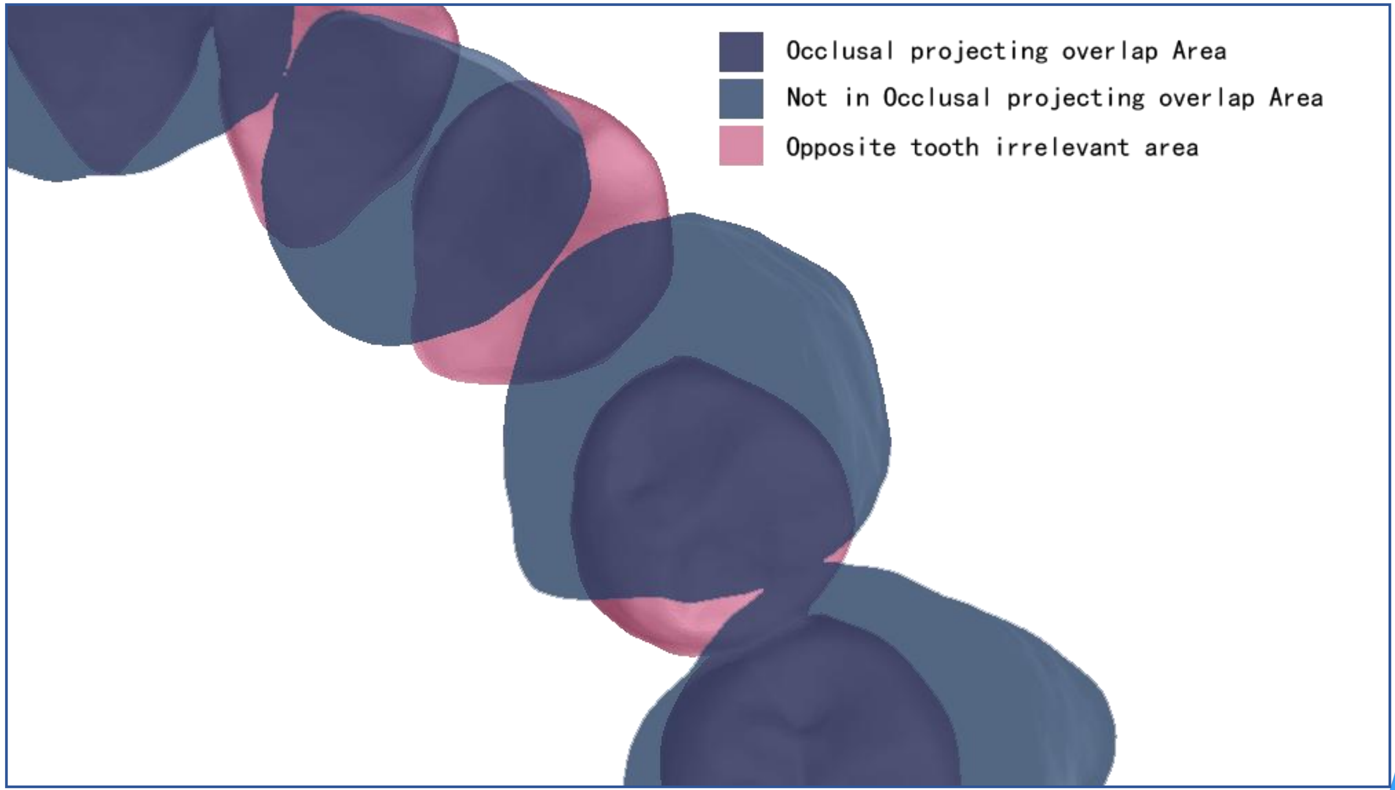
03 Prediction of Tooth Arrangement

Occlusal projecting overlap Loss

Represents whether the occlusal area between the predicted upper and lower jaws matches the ground truth

Occlusal projection range refers to the point cloud of the overlapping region between upper and lower teeth from a top-down perspective

Larger discrepancy between the predicted occlusal projection range and GT results in greater loss value


Occlusal projecting overlap Area
Not in Occlusal projecting overlap Area
Opposite tooth irrelevant area

$$m_i = \underset{p_j \epsilon P^f_{\beta_t}}{\mathrm{Argmin}} \left\| p_i - p_j \right\|_2, \ p_i \epsilon P^f_t$$

$$X_t(i) = \begin{cases} 1 \ (m_i < \tau) \\ 0 \ (m_i \geq \tau) \end{cases}$$

$$L_{fit} = \underset{t \epsilon T}{Ave} \left(\sum_{i=0}^{n-1} \left| \bar{X}_t(i) - X_t^*(i) \right| \right)$$

03 Prediction of Tooth Arrangement

Occlusal distance uniformity Loss

- Consistency and similarity of connection vectors between corresponding points in the occlusal region

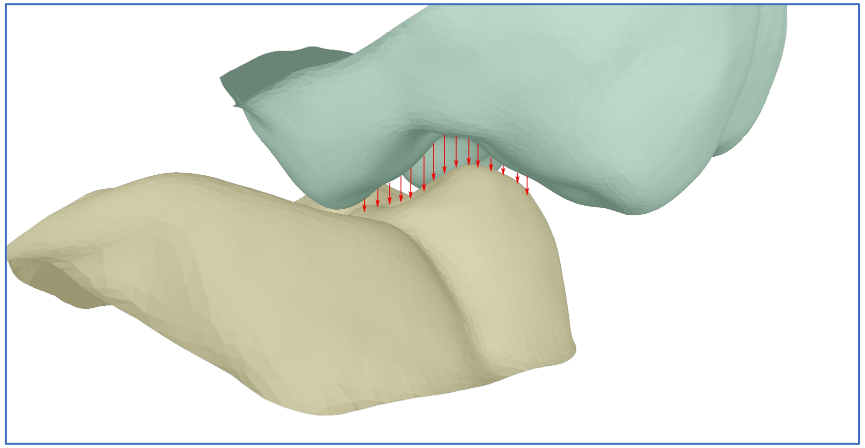
- During normal occlusion, the distance between corresponding points in the upper and lower dental occlusal regions should be essentially consistent to comply with the post-contact stress distribution rule

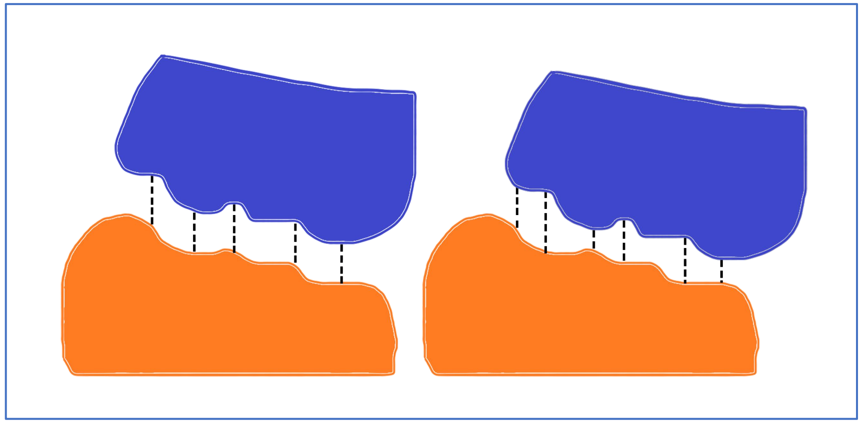
- Greater disparity in distances between corresponding points within the occlusal projection area correlates with higher functional loss

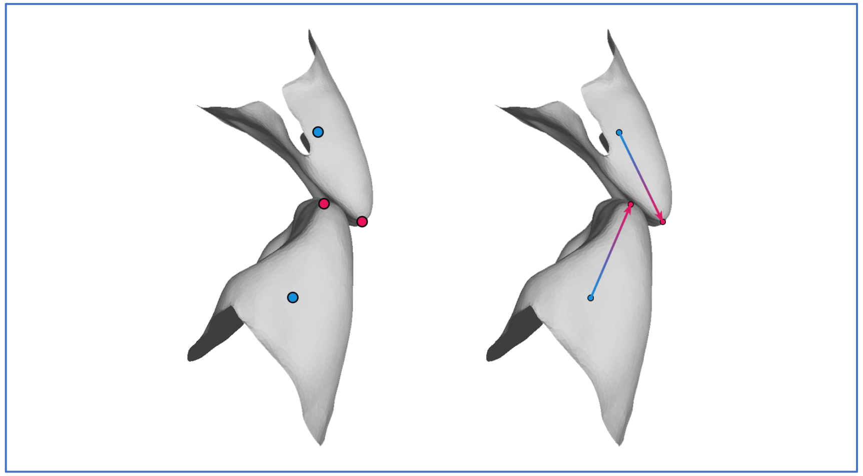
- For the four anterior teeth, the uniformity of occlusion is directly described using vectors formed by the dental centroid and the crown vertex

$$L_{uni}^{pior} = \sum_{t\in T_{pior}} \underset{X_t(i)=1}{Var}\left(\min_{X_{\beta_t}(j)=1} \|p_i - p_j\|_2\right)$$

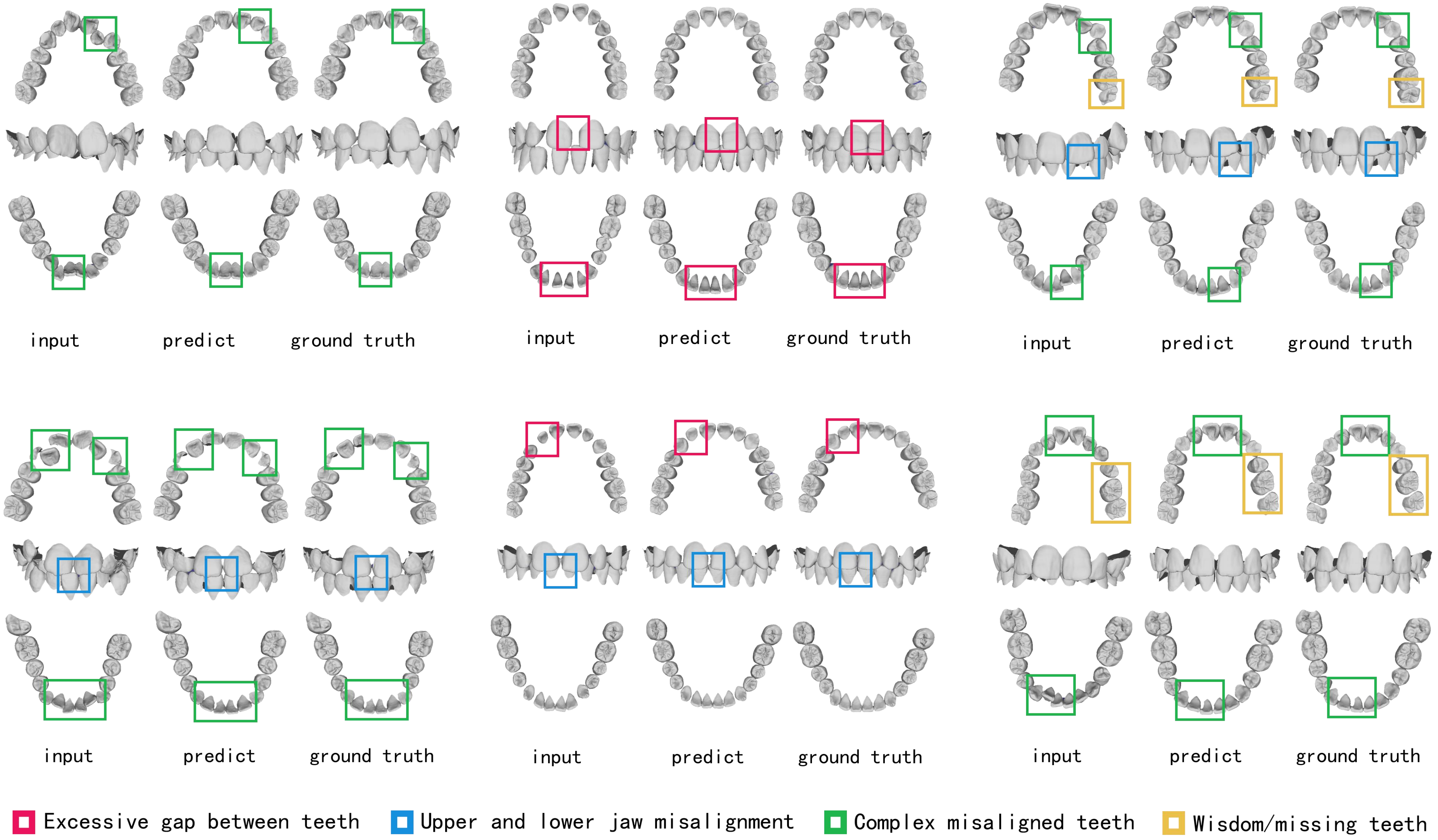
$$L_{uni}^{ant2} = \sum_{t\in T_{ant}} arccos\left(\frac{(\overline{Peak_t} - \bar{c}_t)\cdot(Peak_t^* - c_t^*)}{\|\overline{Peak_t} - \bar{c}_t\|_2 * \|Peak_t^* - c_t^*\|_2}\right)$$



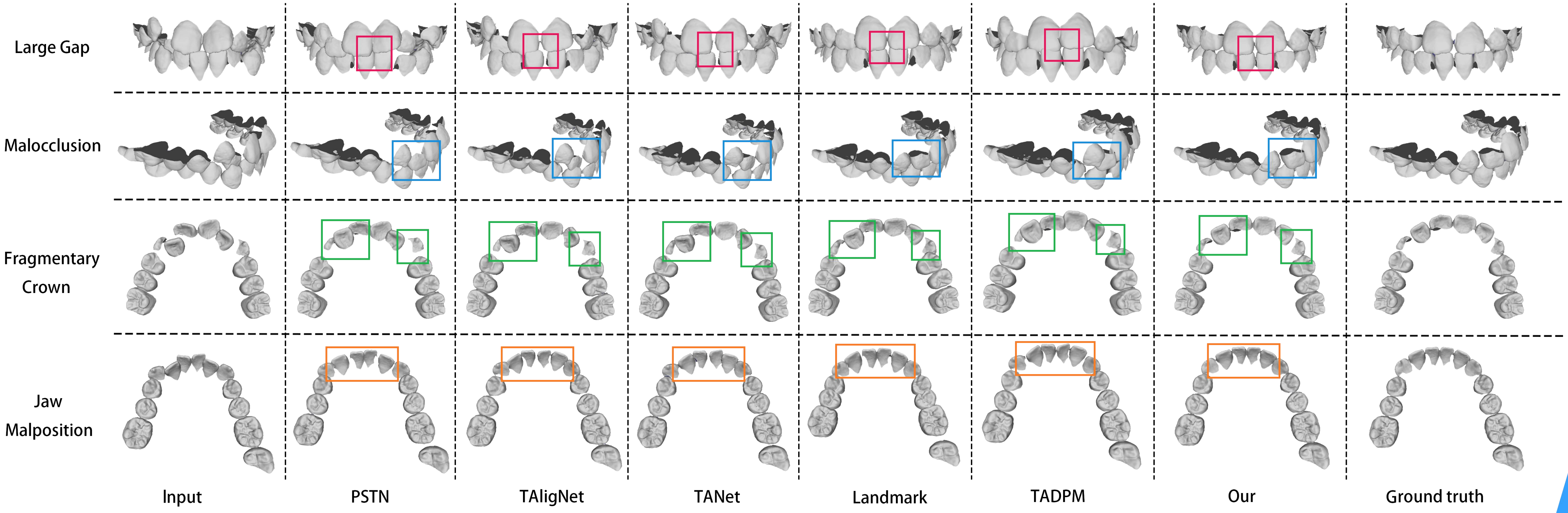




03 Prediction of Tooth Arrangement



03 Prediction of Tooth Arrangement



Misalignment of upper and lower incisors and malocclusion between the upper and lower jaws.
Solution: These can be effectively addressed using occlusion projection range consistency loss and occlusal distance uniformity loss constraints.

Incomplete tooth models, trident-shaped misalignment, and single-jaw interproximal misalignment.
Solution: Data serialization combined with a sliding window Transformer enables accurate identification of relative positions between teeth.

03 Prediction of Tooth Arrangement

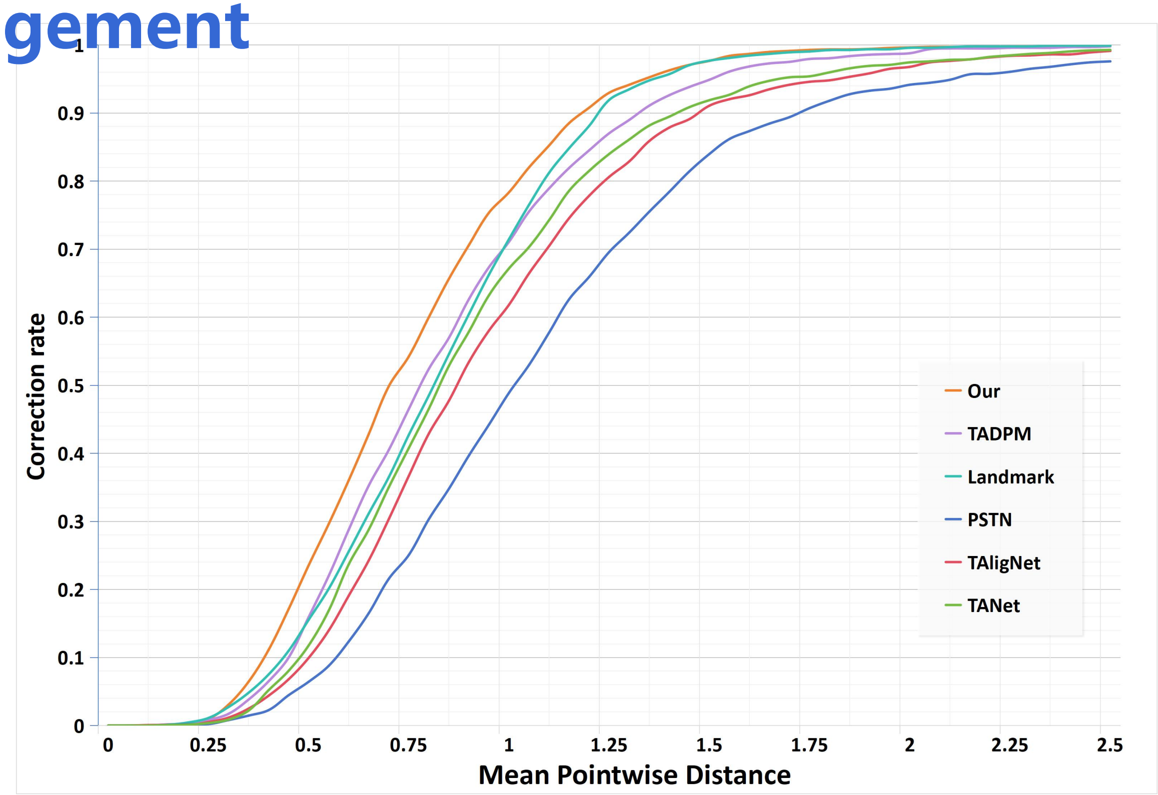
Comparisons with SOTA methods

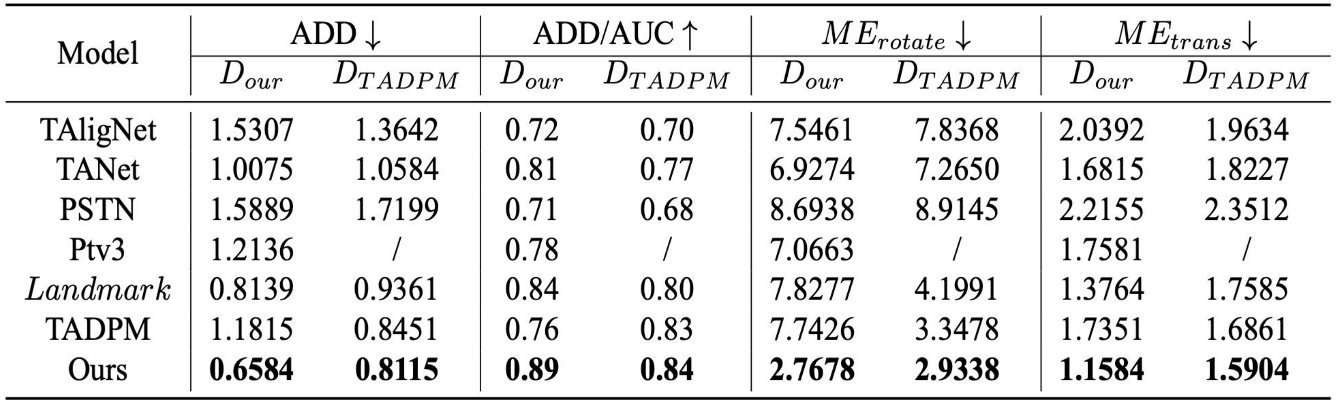
- We reproduced the methods of TANet, PSTNet, TAligNet, and Landmark, and conducted training and testing.

- We compared the AUC, mean rotation error, and mean translation error of other methods.

- We compared the curves of mean point distance. After the mean point distance exceeds 2.5, all curves approach 1, so only curves with k ≤ 2.5 are shown in the graph.

- Under different definitions of mean point distance, the accuracy of our method is the highest.





Model	ADD ↓		ADD/AUC ↑		ME_{rotate} ↓		ME_{trans} ↓	
	D_{our}	D_{TADPM}	D_{our}	D_{TADPM}	D_{our}	D_{TADPM}	D_{our}	D_{TADPM}
TAligNet	1.5307	1.3642	0.72	0.70	7.5461	7.8368	2.0392	1.9634
TANet	1.0075	1.0584	0.81	0.77	6.9274	7.2650	1.6815	1.8227
PSTN	1.5889	1.7199	0.71	0.68	8.6938	8.9145	2.2155	2.3512
Ptv3	1.2136	/	0.78	/	7.0663	/	1.7581	/
Landmark	0.8139	0.9361	0.84	0.80	7.8277	4.1991	1.3764	1.7585
TADPM	1.1815	0.8451	0.76	0.83	7.7426	3.3478	1.7351	1.6861
Ours	**0.6584**	**0.8115**	**0.89**	**0.84**	**2.7678**	**2.9338**	**1.1584**	**1.5904**

06 Summary and Prospect

Summary and Prospect

06 Summary and Prospect

Based on digital orthodontic solutions and related fields, this paper conducts in-depth and comprehensive research and discussion. The work is summarized as follows:

Orthodontic Prediction Method:

A novel high-precision and efficient neural network method for tooth alignment prediction is proposed. Using a Swin-T multi-level feature fusion architecture as the core, the method introduces tooth point cloud serialization rules, improves the data augmentation mechanism, and designs an occlusion evaluation loss function. Experimental results demonstrate its effectiveness and high prediction accuracy.

Technical Implementation:

The proposed framework effectively processes dental point cloud data and optimizes alignment prediction through structured feature fusion and enhanced training strategies.

Experimental Validation:

Rigorous testing confirms that the method achieves superior performance in terms of prediction precision and operational efficiency compared to existing approaches.

06 Summary and Prospect

Directions for Improvement of Orthodontic Treatment Plan Construction System:

Limitations and Improvement Ideas of Orthodontic Prediction Algorithm: Teeth with severe lateral distortion affect the accuracy of serialization, and the algorithm does not support extraction prediction, focusing only on the final orthodontic outcome.

Future improvements may involve refining the serialization method, adding extraction site prediction, and leveraging iterative characteristics to optimize the network for predicting the entire treatment cycle.

浙江工业大学
ZHEJIANG UNIVERSITY OF TECHNOLOGY

计算机科学与技术学院、软件学院
College of Computer Science,Technology and Software

THANKS

Author：Zhenxing Dong，Jiazhou Chen