Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 33(12): 7736–7762.
DOI: 10.3934/era.2025342
Received: 26 October 2025
Revised: 28 November 2025
Accepted: 09 December 2025
Published: 23 December 2025

Research article

GFACNet: 3D dental segmentation from intraoral scans integrating geometric features and anatomical constraints

Gaofeng Zheng, Xiaodong Cui, Aibo Song* and Mingrui Lin

School of Computer Science and Engineering, Southeast University, Nanjing, China

* **Correspondence:** Email: absong@seu.edu.cn.

Abstract: Dental segmentation is a critical step in computer-aided orthodontic treatment planning, but accurate segmentation still faces numerous challenges due to complex tooth morphology, ambiguous gingival boundaries, and clinical issues such as malformed teeth, crowding, and malocclusion. This paper proposes GFACNet, a network that integrates geometric features and anatomical constraints for 3D dental segmentation from intraoral scan data. Our method comprises three key innovations: 1) a morphology-aware graph construction (MAGC) mechanism that adaptively constructs graph structures based on dental geometric characteristics, 2) a multi scale transformer (MST) feature integration module that processes features at different scales while capturing both local and global context, and 3) a hierarchical anatomical constraint loss (HACL) that incorporates multi level anatomical features to guide anatomically consistent segmentation. Experiments on real intraoral scanning datasets demonstrate that GFACNet significantly outperforms existing methods in handling complex dental morphologies, particularly in cases of malformed and missing teeth. Additionally, our method requires reduced computational resources while providing a more practical solution for clinical applications.

Keywords: 3D dental segmentation; intraoral scanning; geometric features; tooth segmentation

1. Introduction

Dental segmentation is a critical step in computer-aided orthodontic treatment planning, providing clinicians with precise geometric information for orthodontic appliance design, tooth movement simulation, and treatment planning [1–5]. However, accurate dental segmentation remains challenging due to complex tooth morphology [6], ambiguous gingival boundaries [7], and common clinical issues such as malformed teeth, crowding, and malocclusion [8]. In recent years, with the advancement of intraoral scanning technology, 3D mesh-based dental models have become essential tools in clinical practice, yet the quality of current segmentation methods often fails to meet clinical standards [9].

Traditional dental segmentation methods typically rely on geometric feature extraction and manually

designed segmentation rules, including contour-based approaches, projection-based methods, and curvature-based techniques [10]. These methods perform poorly when confronted with complex dental morphologies and depend heavily on expert knowledge, limiting their widespread application in clinical environments. With the development of deep learning technologies, automatic segmentation methods based on neural networks have gradually emerged as research focal points, including transformation-based methods that voxelize 3D meshes [11, 12], point cloud-based methods that process sampled point data [13, 14], and mesh-based methods that directly operate on dental surfaces [15, 16].

Despite recent progress in mesh-based dental segmentation, existing methods face three fundamental limitations that hinder their clinical deployment. First, uniform sampling strategies waste computational resources on geometrically simple regions while missing critical boundary details. Second, separate branch processing or sequential refinement architectures limit feature interaction and parallel computation efficiency. Third, purely data-driven approaches lack anatomical constraints, leading to clinically implausible segmentations in challenging cases with missing or malformed teeth. Additionally, these methods typically require numerous high-resolution mesh samples, creating barriers for resource-constrained clinical settings [17].

To address these challenges, we propose GFACNet (geometric features and anatomical constraints network) for 3D dental segmentation. Our method introduces three key innovations: 1) A morphology-aware graph construction (MAGC) mechanism that adaptively allocates computational resources via curvature-guided probabilistic sampling, achieving superior accuracy with 40% fewer sampling points (1500 vs. 2500–3000); 2) A multi scale transformer (MST) feature integration module that enables unified cross-scale attention for simultaneous multi scale interaction, reducing computational cost by approximately 50% compared to dual-branch and sequential refinement approaches; 3) A Hierarchical Anatomical Constraint Loss (HACL) that explicitly incorporates dental anatomical knowledge—including centroid positioning, shape consistency, and adaptive weighting for missing teeth—ensuring clinically plausible results where purely data-driven methods struggle.

Experimental results on real intraoral scanning datasets demonstrate that GFACNet significantly outperforms existing methods in handling complex dental morphologies, particularly for malformed and missing teeth, while requiring reduced computational resources. Our approach exhibits superior precision in processing dental boundary regions, which is crucial for subsequent orthodontic treatment planning.

The main contributions of this paper are:

1) A morphology-aware adaptive graph construction mechanism that dynamically constructs graph structures based on dental geometry, requiring significantly fewer sampling points than uniform sampling approaches while better capturing complex boundaries.

2) A multi scale transformer feature integration module that captures both local details and global context through unified cross-scale attention rather than separate branch processing or sequential refinement.

3) A hierarchical anatomical constraint loss that integrates dental anatomical knowledge into network training, providing robustness in clinically challenging scenarios where purely data-driven methods fail.

4) Extensive experiments that validate our method's effectiveness, achieving state-of-the-art performance with superior computational efficiency, particularly in handling anomalous dental conditions.

2. Related works

2.1. 3D point cloud segmentation

Due to their unordered nature, point cloud data cannot be directly processed using standard convolutional neural networks. Traditional methods typically convert 3D point clouds into processable forms, such as view-based approaches that project point clouds onto 2D planes, inevitably resulting in spatial information loss; or voxelization-based methods that discretize space into regular grids, introducing high computational complexity and quantization error problems.

In recent years, methods for directly processing point clouds have made significant progress. PointNet [13], as pioneering work, processes unordered point sets through symmetric operations, but its approach of treating each point independently limits the capture of local structural relationships. To address this limitation, PointNet++ [14] introduced multi scale hierarchical structures and regional feature aggregation mechanisms, enhancing the network's ability to understand spatial relationships between points. Other methods have explored applying convolution operations to point clouds, such as PointCNN [18], which applies traditional convolutions after normalizing point clouds through learned transformation matrices, while dynamic graph convolutional neural network (DGCNN) [19] treats point clouds as dynamic graph structures, capturing local geometric features through EdgeConv.

With the rise of attention mechanisms, transformer architectures have also been introduced to the field of point cloud processing. Point Transformer [20] and point cloud transformer (PCT) [21] proposed self-attention frameworks suitable for point clouds, effectively capturing long-range dependencies. PointMLP [22] adopts an efficient feed-forward network structure to process point cloud features, while PointNeXt [23] revisits and optimizes the design strategies of PointNet++. Although these methods perform well, most utilize only point coordinate information, neglecting the rich features that can be extracted from the original geometric surfaces, thus limiting their performance potential.

2.2. 3D dental segmentation

3D dental mesh models generated by intraoral scanning technology possess rich geometric characteristics. Traditional segmentation methods largely build upon geometric analysis: Curvature-based methods utilize local curvature variations to identify dental boundaries [24]; projection-based methods map 3D models to 2D planes for processing; and harmonic field-based methods construct potential fields by solving Laplacian equations to guide segmentation [25]. However, these methods often perform poorly when handling abnormal teeth and complex clinical scenarios [26].

Deep learning has brought new perspectives to dental segmentation. Early approaches convert 3D dental models into regular representations: Voxelization-based methods apply 3D Convolutional Neural Network (CNN) but may lose fine geometric details [27], while hierarchical methods are constrained by inter-layer information transfer, making early error correction difficult. Point cloud-based approaches downsample dental meshes for processing [28], but this simplification also results in loss of original geometric information.

Mesh-based approaches directly process mesh data, avoiding transformation-induced information loss. MeshSegNet [15] integrates graph-constrained learning modules for end-to-end dental labeling; iMeshSegNet [11] optimizes computational efficiency through improved architectures; and MPCNet [16] introduces local symmetric positional encoding to enhance feature representation.

Recent sophisticated architectures have explored various strategies to improve segmentation quality.

Dual-branch architectures: DBGANet [29] employs separate processing streams for coordinate and normal vector features, using geometric attention mechanisms to capture both local details and global structure before late-stage fusion. While effective at capturing multi level geometric information, eparate processing may lose fine-grained cross-scale correlations during early feature extraction, and the dual-branch design increases computational overhead (8.33 GFLOPs).

Progressive refinement frameworks: TSegLab [30] adopts a multi stage hierarchical processing approach that iteratively refines segmentation through successive stages. Although this progressive strategy improves quality, the sequential architecture limits parallel computation and increases inference time (51 ms), making real-time clinical deployment challenging.

Instance segmentation approaches: THISNet [31] highlights tooth regions using region proposal mechanisms for tooth instance segmentation, while Teeth-SEG [32] develops an efficient framework with multi scale aggregation specifically for orthodontic applications. These methods demonstrate strong performance but typically employ uniform sampling strategies that allocate equal computational resources across all regions—over-sampling geometrically simple areas while under-sampling critical boundary regions.

Common limitations: Despite their achievements, these contemporary methods share several fundamental constraints. Most rely purely on data-driven learning without explicit anatomical constraints, making them susceptible to producing segmentations that violate dental anatomical principles in challenging cases (e.g., missing teeth, severe malformations, extensive restorations). The lack of anatomical guidance is particularly problematic for clinical deployment, where anatomically consistent results are essential for treatment planning.

Beyond dental segmentation, related domains have demonstrated the value of incorporating domain-specific constraints and multi-scale learning. Hierarchical constraint mechanisms with key point sensitive loss have been successfully applied to sonar image classification [33], while transformer-based multi scale approaches show promise in weakly-supervised pavement crack segmentation [34]. These works validate the effectiveness of combining geometric analysis with learned features, though dental geometry presents unique challenges requiring specialized anatomical constraints.

3. Methods

3.1. Overall architecture

We propose a method based on a geometric features and anatomical constraints network (GFACNet) for 3D dental segmentation. Given a 3D dental model containing N mesh face, our goal is to label each mesh face as one of $C = 33$ semantic categories (32 tooth classes and one gingival class). GFACNet achieves accurate dental segmentation through three key modules:

1) Morphology-aware graph construction (MAGC): This module adaptively samples meshes based on dental morphology and constructs dynamic graph structures, paying special attention to high-curvature regions such as dental boundaries, ensuring important areas receive more refined representations.

2) Multi scale transformer feature integration module (MST): This module employs a transformer architecture to directly process features at different scales, capturing both local dependencies and

global context, enabling effective integration of multi scale information.

3) Hierarchical anatomical constraint loss (HACL): This module considers multi level anatomical features of teeth, combining centroid constraints, shape constraints, and relative position constraints to guide the network in learning segmentation results that conform to dental anatomical structures.

GFACNet's processing flow is illustrated in Figure 1. First, we extract 24-dimensional features from the input 3D mesh model, where the first 12 dimensions represent the coordinates of the three vertices and the mesh face, and the latter 12 dimensions represent the normal vectors of the vertices and mesh face. Then, the MAGC module adaptively samples meshes based on curvature information and constructs dynamic graph structures. Next, the MST module processes the sampled features, extracting and integrating information across multiple scales. Finally, the network prediction results are optimized through the HACL module, ensuring conformity to dental anatomical principles.

Our method effectively handles complex dental morphologies, demonstrating excellent performance particularly in clinical common scenarios such as malformed teeth and missing teeth. Compared to existing methods, GFACNet achieves higher segmentation accuracy using fewer sampled meshes (reducing computation by approximately 40%), which is especially important for resource-constrained medical environments.

3.2. Morphology-aware graph construction (MAGC)

The morphology-Aware graph construction (MAGC) module is one of our core innovations in GFACNet, designed to address the limitations of traditional graph construction methods when processing complex dental morphologies. Unlike traditional methods that use static sampling strategies, MAGC can adaptively sample and construct graph structures based on the geometric characteristics of teeth, with special attention to key regions such as dental boundaries.

3.2.1. Curvature-aware sampling

Traditional farthest point sampling (FPS) algorithms often struggle to capture key regions when processing complex dental morphologies. To address this, we first calculate the principal curvatures K_1 and K_2 for each mesh, then define curvature intensity as:

$$C_i = \sqrt{K_{1i}^2 + K_{2i}^2}. \tag{3.1}$$

Based on curvature intensity, we design a probability function to guide the sampling process:

$$P(f_i) = \frac{C_i + \epsilon}{\sum_{j=1}^{N}(C_j + \epsilon)}, \tag{3.2}$$

where $\epsilon = 0.01$ is a small smoothing factor that prevents zero-curvature regions from being completely ignored. Through this strategy, high-curvature regions (such as dental boundaries) receive higher sampling probabilities, ensuring that the geometric details of these key regions can be fully learned by the network.

In the actual sampling process, we use rejection sampling to select M center meshes ($M = 1500$) according to the probability distribution $P(f_i)$, which is significantly reduced compared to the sampling quantity used by traditional methods (typically 2500–3000), while maintaining segmentation accuracy.

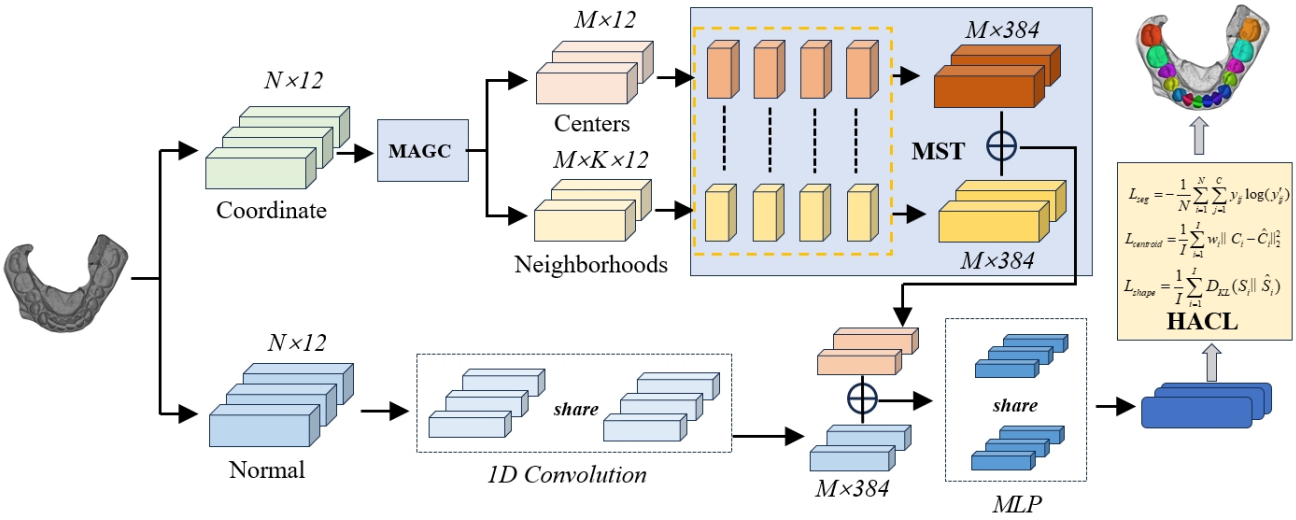

Figure 1. Overall architecture of GFACNet. The network takes a 3D dental mesh as input and extracts 24-dimensional features, which are processed through three key modules: morphology-aware graph construction (MAGC), multi scale transformer (MST), and hierarchical anatomical constraint Loss (HACL). The flowchart shows how coordinate features pass through MAGC to generate centers and neighborhoods which are then processed by MST to produce features of dimension $M \times 384$. Meanwhile, normal features are processed through 1D convolution and combined with MST output features. Finally, the results are optimized using the HACL module, which incorporates segmentation loss (L_{seg}), centroid constraint ($L_{centroid}$), and shape constraint (L_{shape}).

3.2.2. Dynamic neighborhood construction

For each sampled center mesh, traditional methods typically use k-nearest neighbors (KNN) algorithms with fixed K values to construct neighborhood relationships, which lack flexibility when processing dental regions of varying complexity. To address this, we designed a dynamic KNN algorithm that automatically adjusts the number of neighbors K based on local geometric complexity:

$$K_i = \min(K_{max}, \max(K_{min}, \lfloor \alpha \cdot C_i + \beta \rfloor)), \tag{3.3}$$

where $K_{min} = 16$ and $K_{max} = 48$ are predefined minimum and maximum neighbor quantity boundaries, and α and β are learnable parameters, initialized as $\alpha = 20$ and $\beta = 10$. This design enables the network to allocate more neighbors to geometrically complex regions and fewer neighbors to simple regions, thereby optimizing computational resource allocation.

3.2.3. Edge feature representation

In addition to node features, we also introduce edge feature representations to enhance the network's perception of dental boundaries. For each pair of connected meshes f_i and f_j, we define the edge feature as:

$$e_{ij} = \phi(f_i, f_j, \mathbf{d}_{ij}, \theta_{ij}), \tag{3.4}$$

where $\mathbf{d}_{ij} = \mathbf{c}_i - \mathbf{c}_j$ is the position difference vector between the centers of the two meshes, θ_{ij} is the angle between the normal vectors of the two meshes, and ϕ is a feature encoding function implemented

as a three-layer (multi layer perceptron) MLP network (input dimension 24 + 3 + 1 = 28, hidden layer dimensions 64 and 32). This edge feature design can capture the geometric relationships and topological structures between meshes, especially in key regions such as dental boundaries.

3.2.4. Adaptive feature aggregation

Finally, we use a graph attention mechanism to adaptively aggregate the features of the center mesh and its neighbors:

$$\mathbf{W}_A = \text{softmax}(\sigma([\mathbf{C} - \mathbf{N}; \mathbf{N}])), \tag{3.5}$$

$$\mathbf{N}' = \mathbf{W}_A \otimes \mathbf{N}, \tag{3.6}$$

where $\mathbf{C}$ is the center mesh feature, $\mathbf{N}$ is the neighbor mesh feature, σ is a lightweight MLP network, and $\otimes$ represents element-wise multiplication. This attention mechanism enables the network to automatically adjust the importance of different neighbors based on feature differences, thereby more effectively capturing local structural information.

3.3. Multi scale transformer feature integration module (MST)

The multi scale transformer feature integration module (MST) is a key component of GFACNet, designed to address the inefficiency of traditional methods that process local and global features separately before fusion. The MST module adopts a transformer-based architecture that can directly process features at different scales while capturing both local dependencies and global context, achieving effective integration of multi scale information.

3.3.1. Multi scale feature extraction

Traditional feature extraction methods often focus on information at a single scale, making it difficult to comprehensively capture the complex structure of teeth. We first construct multi scale feature representations.

Given the center mesh representation C and neighbor mesh representation N obtained from the MAGC module, we construct feature representations at $S = 4$ different scales:

$$X_s = \{X^1, X^2, X^3, X^4\}. \tag{3.7}$$

Each scale's feature representation is obtained by aggregating features within different radius ranges:

$$X^s = \text{Aggregate}(C, N, r_s), \tag{3.8}$$

where r_s is the aggregation radius for the s-th scale. We set $r_1 = 0.03$, $r_2 = 0.06$, $r_3 = 0.12$, $r_4 = 0.24$, values experimentally verified to effectively cover multi scale information from local microscopic structures to global macroscopic morphology. The Aggregate function is implemented as a graph convolution operation, specifically defined as:

$$\text{Aggregate}(C, N, r_s) = \sum_{j \in \mathcal{N}_i(r_s)} \frac{1}{|\mathcal{N}_i(r_s)|} \cdot \text{MLP}([C_i; N_j; C_i - N_j]), \tag{3.9}$$

where $\mathcal{N}_i(r_s)$ represents the set of neighbors of center mesh i within radius r_s, and MLP is a two-layer perceptron network with output dimensions $d_1 = 64$, $d_2 = 96$, $d_3 = 128$, and $d_4 = 192$ to accommodate the feature complexity needed for different scales.

3.3.2. Self-attention mechanism

For each scale's feature representation X^s, we apply a self-attention mechanism to capture the interrelationships between meshes:

$$A^s = \text{softmax}\left(\frac{Q^s(K^s)^T}{\sqrt{d_k}}\right), \tag{3.10}$$

$$Z^s = A^s V^s + X^s, \tag{3.11}$$

where Q^s, K^s, and V^s are the query, key, and value matrices, respectively, obtained from X^s through linear projections:

$$Q^s = X^s W_Q^s, \quad K^s = X^s W_K^s, \quad V^s = X^s W_V^s, \tag{3.12}$$

where W_Q^s, W_K^s, and W_V^s are learnable projection matrices, and d_k is the feature dimension. Note that we have added residual connections ($+ X^s$), which helps alleviate the vanishing gradient problem in deep networks.

To further enhance feature learning capability, we stack $L = 3$ self-attention layers at each scale, forming a deep self-attention network:

$$Z_l^s = \text{SelfAttention}(Z_{l-1}^s), \quad l = 1, 2, 3, \tag{3.13}$$

$$Z_l^s = \text{LayerNorm}(\text{FFN}(Z_l^s) + Z_l^s), \tag{3.14}$$

where FFN is a feed-forward neural network, containing two-layer MLP and gaussian error linear unit (GELU) activation functions; LayerNorm is a layer normalization operation that helps stabilize the training process.

3.3.3. Cross-scale feature fusion

Traditional methods often simply concatenate or sum features of different scales, which may lead to information redundancy or loss of key information. We design a cross-scale attention mechanism to adaptively fuse features of different scales:

First, we concatenate features of different scales.

$$X_{concat} = [Z^1, Z^2, Z^3, Z^4]. \tag{3.15}$$

Then, we calculate cross-scale attention:

$$Q_{cross} = X_{concat} W_{Q,cross}, \tag{3.16}$$

$$K_{cross} = X_{concat} W_{K,cross}, \tag{3.17}$$

$$V_{cross} = X_{concat} W_{V,cross}, \tag{3.18}$$

$$A_{cross} = \text{softmax}\left(\frac{Q_{cross}(K_{cross})^T}{\sqrt{d_k}}\right), \tag{3.19}$$

$$Z_{cross} = A_{cross} V_{cross}. \tag{3.20}$$

This design enables the network to learn the interrelationships between features at different scales, thereby achieving more effective feature fusion. Specifically, we construct a fully connected attention graph between different scales, where features at each scale can interact with features at all other scales.

Finally, we integrate all features through an MLP network:

$$F = \text{MLP}([Z^1, Z^2, Z^3, Z^4, Z_{cross}]), \tag{3.21}$$

where MLP is a three-layer perceptron network with hidden layer dimensions of 512 and 256, and an output dimension of 384.

3.3.4. Feature upsampling and integration

The feature representation after MST processing is based on sampled meshes and needs to be upsampled back to the original mesh size. We design a distance-weighted feature propagation algorithm:

$$F_i = \sum_{j \in \mathcal{S}} w_{ij} \cdot F_j, \tag{3.22}$$

$$w_{ij} = \frac{\exp(-\|p_i - p_j\|^2/\sigma^2)}{\sum_{k \in \mathcal{S}} \exp(-\|p_i - p_k\|^2/\sigma^2)}, \tag{3.23}$$

where $\mathcal{S}$ is the set of sampled meshes, p_i and p_j are the spatial coordinates of original mesh i and sampled mesh j respectively, and σ is a smoothing factor (set to twice the average nearest neighbor distance). This distance-based weighting strategy can maintain the spatial continuity of features, avoiding artifacts that might result from hard assignments.

3.4. Hierarchical anatomical constraint loss (HACL)

The hierarchical anatomical constraint loss (HACL) is one of our key innovations in GFACNet, aimed at integrating dental anatomical knowledge into the network training process to guide the model in learning segmentation results that better conform to medical principles. Unlike traditional methods that only use pixel-level segmentation losses, HACL considers dental anatomical features from multiple levels, including tooth centroid positions, shape features, and relative spatial relationships.

3.4.1. Multi level anatomical constraints

We design three levels of anatomical constraints to constrain segmentation results from different perspectives:

1) Centroid constraint: Constrains the centroid position of each tooth to ensure accurate overall spatial positioning.

2) Shape constraint: Constrains the tooth shape to conform to anatomical priors, preventing unreasonable segmentation results.

3) Relative position constraint: Constrains the relative position relationships between teeth, maintaining the overall structure of the dental arch.

This multi level constraint design stems from clinical observations: Dentists, when making diagnoses, consider not only the details of individual teeth but also the overall arrangement of teeth and their interrelationships.

3.4.2. Centroid constraint

The centroid constraint is based on a key observation: Even with individual variations in tooth morphology, the centroid positions of each tooth class remain relatively stable. First, we calculate pseudo-centroids for each tooth based on the segmentation results predicted by the network:

$$L_i = \arg\max(P_i), \tag{3.24}$$

where P_i is the class probability distribution of the i-th mesh, and L_i is the class with the highest probability, i.e., the pseudo-label. Then, based on these pseudo-labels, we extract meshes of different classes and calculate the centroid coordinates for each tooth:

$$C_i = \frac{1}{|\mathcal{M}_i|} \sum_{j \in \mathcal{M}_i} p_j, \tag{3.25}$$

where $\mathcal{M}_i$ is the set of meshes classified as the i-th tooth, and p_j is the center coordinate of mesh j.

To handle cases of missing teeth, we introduce an adaptive weighting mechanism. If the extraction result for a certain tooth is empty (indicating it might be missing), its centroid coordinate is set to a zero vector, and a lower weight is assigned in the loss calculation:

$$w_i = \begin{cases} 1.0, & \text{if } |\mathcal{M}_i| > \theta, \\ 0.2, & \text{otherwise}, \end{cases} \tag{3.26}$$

where θ is a set threshold (set to 10 in experiments).

The centroid constraint is defined as:

$$L_{centroid} = \frac{1}{I} \sum_{i=1}^{I} w_i \|C_i - \hat{C}_i\|_2^2, \tag{3.27}$$

where $\hat{C}_i$ is the ground truth value of the centroid of the i-th tooth, and I is the total number of tooth classes (32). Through this weighted Euclidean distance metric, the network can learn to accurately localize the position of each tooth while appropriately tolerating potentially missing teeth.

3.4.3. Shape constraint

In addition to centroid positions, tooth shapes also have class-specific characteristics. For example, incisors typically are flat-shaped, while molars are cube-shaped. The shape constraint aims to make the network learn these morphological features.

We first perform principal component analysis (PCA) on each tooth class to extract its characteristic shape representation:

$$S_i = \text{PCA}(\{p_j - C_i | j \in \mathcal{M}_i\}). \tag{3.28}$$

This representation captures the shape distribution of the tooth relative to its centroid. Then, we use the Kullback-Leibler divergence to measure the difference between the predicted shape and the true shape:

$$L_{shape} = \frac{1}{I} \sum_{i=1}^{I} D_{KL}(S_i \| \hat{S}_i), \tag{3.29}$$

where $\hat{S}_i$ is the true shape representation of the i-th tooth. To practically calculate the KL divergence, we model the shape representation as a multivariate Gaussian distribution:

$$D_{KL}(S_i \| \hat{S}_i) = \frac{1}{2}\left[\log\frac{|\hat{\Sigma}_i|}{|\Sigma_i|} + \mathrm{tr}(\hat{\Sigma}_i^{-1}\Sigma_i) + (\hat{\mu}_i - \mu_i)^T\hat{\Sigma}_i^{-1}(\hat{\mu}_i - \mu_i) - d\right], \tag{3.30}$$

where μ_i and Σ_i are the mean and covariance matrix of the predicted distribution, $\hat{\mu}_i$ and $\hat{\Sigma}_i$ are the parameters of the true distribution, and d is the dimension of the distribution (we take the first 8 principal components).

3.4.4. Total loss function

The total loss of HACL is combined by weighting the constraint terms:

$$L_{HACL} = \lambda_1 L_{centroid} + \lambda_2 L_{shape}, \tag{3.31}$$

where λ_1 and λ_2 are hyperparameters that balance the importance of different constraints. Through extensive experiments, we set $\lambda_1 = 0.5$ and $\lambda_2 = 0.3$, a combination that achieved the best performance in our experiments.

The final total loss function of the network combines the standard segmentation loss and HACL:

$$L_{total} = L_{seg} + \lambda L_{HACL}, \tag{3.32}$$

where L_{seg} is the negative log-likelihood segmentation loss:

$$L_{seg} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(y'_{ij}), \tag{3.33}$$

and λ is a weight parameter balancing the two types of losses, set to 0.4. In the training process, we adopt a staged strategy: For the first 100 epochs, only L_{seg} is used, then L_{HACL} is introduced for another 150 epochs of training. This strategy allows the network to first learn basic segmentation capabilities, then gradually incorporate anatomical knowledge.

3.5. Implementation details

In this work, we follow the FDI (Fédération Dentaire Internationale) [35] tooth numbering system as illustrated in Figure 2. The FDI system uses a two-digit notation: The first digit indicates the quadrant (1: upper right, 2: upper left, 3: lower left, 4: lower right) and the second digit indicates the tooth position (1–8, from central incisor to third molar). For example, tooth 11 is the upper right central incisor, and tooth 48 is the lower right third molar.

For network implementation, we extract 24-dimensional features as input, where the first 12 dimensions represent three vertices and mesh center coordinates, and the latter 12 dimensions represent the corresponding normal vectors. The network architecture details are as follows:

- MAGC module:
  - The number of sampling centers M is set to 1500.

  - Dynamic neighborhood with $K_{min} = 16$ and $K_{max} = 48$.
  - Edge feature encoder: three-layer MLP (28→64→32→16).
  - Curvature-based sampling with $\alpha = 20$ and $\beta = 10$.
- MST module:
  - Four scales with radii $r_1 = 0.03$, $r_2 = 0.06$, $r_3 = 0.12$, $r_4 = 0.24$.
  - Feature dimensions: $d_1 = 64$, $d_2 = 96$, $d_3 = 128$, $d_4 = 192$.
  - Three stacked self-attention layers per scale with 8 attention heads.
  - Cross-scale attention with projection dimension 256.
  - Final feature fusion: three-layer MLP (480 + 256→512→256→384).
- HACL module:
  - Centroid constraint with threshold $\theta = 10$ and weight $\lambda_1 = 0.5$.
  - Shape constraint using PCA with 8 principal components and weight $\lambda_2 = 0.3$.
  - Total loss weight $\lambda = 0.4$.
- Prediction head: Four-layer MLP (384→256→128→64→33).
- Training strategy: First 100 epochs with only L_{seg}, followed by 150 epochs with combined L_{total}.

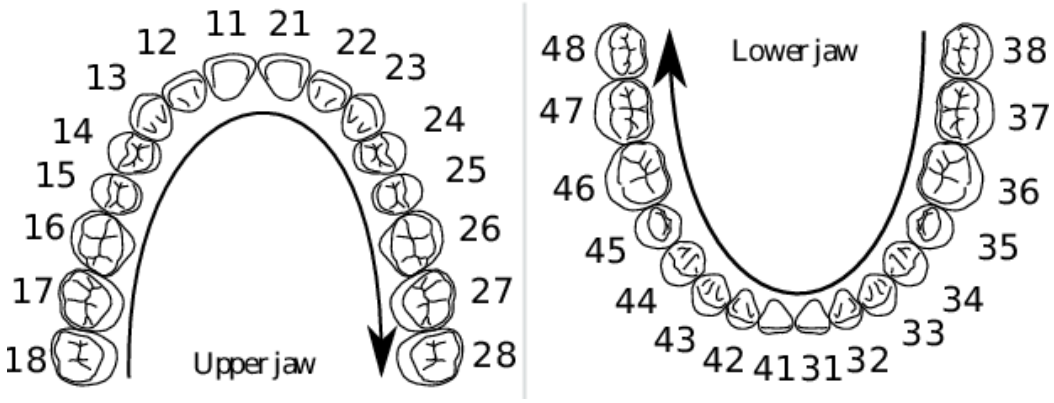

Figure 2. FDI tooth numbering system used in our work. The upper jaw contains teeth numbered from 11 to 28 (from the central incisor to the third molar), while the lower jaw contains teeth numbered from 31 to 48, following the same anatomical sequence.

4. Experimental

4.1. Dataset and experimental setup

4.1.1. Dataset

In this study, we utilized the 3DTeethSeg22 dataset comprising 1800 high-quality 3D intraoral scans from 900 patients across collaborative clinics in France and Belgium. Patient demographics include equal gender distribution (50% male, 50% female) with 70% patients under 16 years, 27% between 16 and 59 years, and 3% over 60 years. Half received orthodontic treatment while the other half underwent restorative treatment. Data acquisition was performed by experienced orthodontists and dentists using Primescan (Dentsply), Trios3 (3Shape), and iTero Element 2 Plus scanners with 10–90 micrometers

accuracy and 30–80 points per square millimeter resolution. Each scan is manually annotated following FDI numbering system with 33 semantic labels.

Preprocessing includes mesh cleaning, PCA-based coordinate centering, occlusal plane alignment, and manual tooth cropping with UV mapping through harmonic parameterization. For missing teeth, the annotation workflow supplements positional information through adjacent tooth arrangement patterns. Malformed teeth are handled via manual boundary annotation with curvature feature identification, while occlusion cases undergo occlusal plane alignment and clinical validation. All annotations are validated by experienced specialists with iterative refinement for accuracy. We employed 5-fold cross-validation with patient-level separation (70% training, 15% validation, 15% testing) to prevent data leakage.

4.1.2. Evaluation metrics

To comprehensively evaluate our proposed GFACNet and baseline methods, we employed three widely-used evaluation metrics:

1) *Overall accuracy (OA)*: measures the percentage of correctly classified mesh face across all classes, calculated as:

$$\mathrm{OA} = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i + FP_i)}. \tag{4.1}$$

2) *Mean intersection over union (mIoU)*: quantifies the average IoU across all tooth classes, calculated as:

$$\mathrm{IoU}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c + \mathrm{FN}_c}, \tag{4.2}$$

$$\mathrm{mIoU} = \frac{1}{C}\sum_{c=1}^{C} \mathrm{IoU}_c, \tag{4.3}$$

where TP_c, FP_c, and FN_c represent true positives, false positives, and false negatives for class c, respectively, and C is the number of classes.

3) *Mean accuracy (mAcc)*: represents the average per-class classification accuracy, calculated as:

$$\mathrm{Acc}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}, \tag{4.4}$$

$$\mathrm{mAcc} = \frac{1}{C}\sum_{c=1}^{C} \mathrm{Acc}_c. \tag{4.5}$$

These metrics comprehensively reflect clinical usability: OA indicates reliability for automated workflows with minimal manual correction; mIoU measures boundary precision critical for appliance design requiring < 0.3 mm tolerance, directly impacting prosthetic fit accuracy; mAcc ensures balanced performance across all tooth types, preventing systematic errors in specific regions (e.g., molars, malformed teeth) that could compromise treatment planning.

4.1.3. Training configuration

All experiments were conducted on an NVIDIA RTX 3090 GPU with 24 GB memory. For network training, we employed the Adam optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$, an initial learning rate of 1×10^{-3} and weight decay of 1×10^{-4}. We adopted a cosine annealing learning rate schedule with a minimum

learning rate of 1×10^{-5}. Additionally, we implemented a learning rate decay strategy that reduced the rate by a factor of 0.5 every 20 epochs when validation loss plateaued.

The network was trained with a batch size of 4 dental models for a total of 250 epochs, following a two-stage training strategy: The first 100 epochs utilized only the segmentation loss (L_{seg}), while the subsequent 150 epochs incorporated the full loss function (L_{total}), including our proposed hierarchical anatomical constraint loss. This approach allowed the network to first learn basic segmentation capabilities before gradually incorporating anatomical knowledge.

To enhance model robustness and generalization, we applied data augmentation during training including random translation along X, Y, Z axes within ranges of $[-6, 6]$, $[-8, 8]$, $[-5, 5]$ mm respectively, random rotation around Z-axis within $[-\pi/10, \pi/10]$ radians, random scaling with factors between 0.9 and 1.1, and Gaussian noise injection with $\sigma = 0.01$ added to vertex coordinates.

4.2. Comparison with existing methods

We compared our proposed GFACNet with eight state-of-the-art 3D dental segmentation methods, including point cloud-based methods (PointNet, PointNet++, Point Transformer) and mesh-based methods (MeshSegNet, TSGCNet, SGTNet, DBGANet, TSegLab). To ensure fair comparison, all methods were retrained and evaluated on the same dataset.

From Table 1, we can observe that our proposed GFACNet significantly outperforms existing methods on all evaluation metrics. Compared to the current best performer SGTNet [36], GFACNet improves OA, mIoU, and mAcc by 1.54%, 2.87%, and 0.55%, respectively. When compared with the recently proposed DBGANet [29] and TSegLab [30], our method achieves improvements of 1.32%/1.89% and 0.98%/1.64% in OA and mIoU, respectively, demonstrating superior performance over these contemporary approaches. Notably, the Chamfer distance evaluation demonstrates our method's superior boundary precision, achieving 0.241 mm compared to SGTNet's 0.285 mm, DBGANet's 0.267 mm, and TSegLab's 0.273 mm, representing a 15.4%, 9.7%, and 11.7% improvement in boundary accuracy, respectively.

Our method achieves 4.73 G FLOPs while delivering 88.15% mIoU, significantly outperforming competing methods: 47.0% and 51.1% FLOPs reduction compared to TSGCNet and SGTNet respectively, 43.2% reduction compared to DBGANet, and 74.4% reduction compared to Point Transformer while achieving higher accuracy. Notably, although TSegLab achieves competitive accuracy (86.51% mIoU), our method outperforms it by 1.64% while requiring 38.5% fewer FLOPs (4.73 G vs. 7.69 G), demonstrating better efficiency. These improvements are statistically significant, demonstrating the effectiveness of our method. Furthermore, GFACNet also shows advantages in inference speed, being approximately 22% faster than SGTNet and 31% faster than DBGANet, which is particularly important for clinical applications. Table 1 shows a comprehensive performance comparison of different methods on the test set, including segmentation accuracy and computational efficiency. To analyze the performance of different methods on various tooth categories in more detail, Table 2 shows the IoU values for all tooth categories.

Table 2 shows that GFACNet achieves the best results across almost all tooth categories. Particularly in challenging anterior regions (such as T13/T43) and certain posterior regions (such as T15/T45), our method improves over SGTNet by 3.2%/2.4% and 5.6%/7.0%, respectively. Compared to DBGANet, our method shows consistent improvements across all tooth categories, with particularly notable gains in challenging regions such as T15/T45 (3.2%/3.9%) and T28/T38 (0.9%/1.3%). Similarly, when

compared to TSegLab, GFACNet demonstrates superior performance in complex tooth types, especially for T15/T45 (2.4%/3.3%) and boundary-sensitive regions. This indicates that our method has better adaptability when processing different types of teeth.

Table 1. Quantitative comparison with existing methods (5-fold cross-validation average results).

Method	OA (%)	mIoU (%)	mAcc (%)	CD (mm)	Params (M)	FLOPs (G)	Inference time (ms)
PointNet	90.29 ± 0.04	75.68 ± 0.16	82.08 ± 0.21	0.453	1.82	2.84	47
PointNet++	80.61 ± 0.22	56.61 ± 0.47	69.59 ± 0.44	0.521	0.97	3.67	53
Point Transformer	88.13 ± 0.42	70.60 ± 0.73	78.55 ± 0.81	0.389	19.39	18.45	65
MeshSegNet	88.30 ± 0.15	67.89 ± 0.29	76.05 ± 0.19	0.367	1.79	5.21	78
TSGCNet	93.76 ± 0.09	86.53 ± 0.25	88.38 ± 0.15	0.298	4.13	8.92	52
SGTNet	93.24 ± 0.05	85.28 ± 0.09	87.85 ± 0.08	0.285	5.90	9.67	49
DBGANet	93.46 ± 0.11	86.26 ± 0.17	88.15 ± 0.13	0.267	6.35	8.33	55
TSegLab	93.80 ± 0.08	86.51 ± 0.12	88.42 ± 0.10	0.273	5.87	7.69	51
GFACNet (Ours)	94.78 ± 0.07	88.15 ± 0.13	88.93 ± 0.11	0.241	5.09	4.73	38

Table 2. Detailed IoU values comparison for all tooth categories (%).

Tooth category	DBGANet	TSegLab	PointNet	PointNet++	MeshSegNet	TSGCNet	SGTNet	GFACNet
T11/T41	84.8/86.2	85.3/86.7	73.4/74.2	56.8/52.5	60.8/55.2	83.2/85.7	83.4/85.3	86.7/87.9
T12/T42	84.5/85.1	84.9/85.6	72.8/72.5	55.2/54.3	62.5/57.1	82.8/84.6	83.0/84.9	86.1/86.4
T13/T43	91.2/91.8	91.6/92.1	70.6/84.3	56.6/59.4	65.7/71.5	90.9/89.2	89.5/91.0	92.7/93.4
T14/T44	86.8/88.5	87.1/89.0	71.5/79.6	52.4/57.0	66.5/73.8	84.3/86.1	85.2/87.1	88.3/90.7
T15/T45	81.3/85.9	82.1/86.5	69.1/77.7	47.8/55.4	67.4/75.7	78.1/83.7	78.9/82.8	84.5/89.8
T16/T46	90.5/86.8	91.1/87.3	80.2/78.5	60.2/62.1	83.7/80.9	88.2/81.5	89.4/85.2	92.0/88.7
T17/T47	92.1/89.2	92.5/89.8	83.6/80.8	64.4/65.8	86.3/83.1	90.1/83.0	91.6/87.9	93.1/90.5
T18/T48	83.7/82.8	84.2/83.5	77.2/76.0	58.1/60.2	76.5/74.2	79.8/77.2	82.3/81.4	85.1/84.2
T21/T31	92.5/87.3	93.0/87.9	84.1/74.4	63.5/51.0	71.2/74.9	91.8/87.1	92.0/85.0	93.6/88.7
T22/T32	91.2/86.1	91.7/86.8	79.8/73.2	61.3/52.7	70.3/72.5	89.4/88.2	90.3/84.9	92.3/87.6
T23/T33	92.8/93.3	93.2/93.7	80.6/80.2	54.1/56.5	64.6/83.3	91.9/92.2	90.8/92.2	93.8/94.5
T24/T34	89.8/90.9	90.3/91.4	78.3/76.5	53.3/53.8	65.2/72.0	88.7/90.3	87.5/89.6	91.2/92.1
T25/T35	92.1/85.6	92.7/86.2	74.0/78.2	56.0/52.0	70.0/78.4	90.3/82.7	91.1/85.0	93.5/86.9
T26/T36	87.9/86.5	88.5/87.1	77.5/75.8	58.7/59.6	77.2/76.8	84.9/82.0	86.7/85.5	89.3/87.8
T27/T37	83.2/86.9	83.7/87.5	76.4/82.8	59.0/61.5	75.0/81.6	81.8/80.3	82.3/85.8	84.1/88.3
T28/T38	81.8/83.3	82.3/83.8	74.5/77.1	56.3/58.9	73.6/78.2	78.5/76.9	80.1/82.4	82.7/84.6

Figure 3 presents a visual comparison of segmentation results between our GFACNet and other methods. As shown, our method produces results closest to the ground truth across three different dental models. Particularly in tooth boundary regions and areas with complex morphology, GFACNet demonstrates superior accuracy. In contrast, PointNet shows significant limitations in detail processing, while MeshSegNet and TSGCNet, though improved, still exhibit segmentation errors at certain tooth junctions. SGTNet performs well overall but lacks precision on some teeth with complex morphology. DBGANet and TSegLab, while achieving competitive results, occasionally exhibit over-segmentation in tightly crowded regions and under-segmentation in areas with subtle geometric variations. These visualization results further validate the superior performance of our method in 3D dental segmentation tasks.

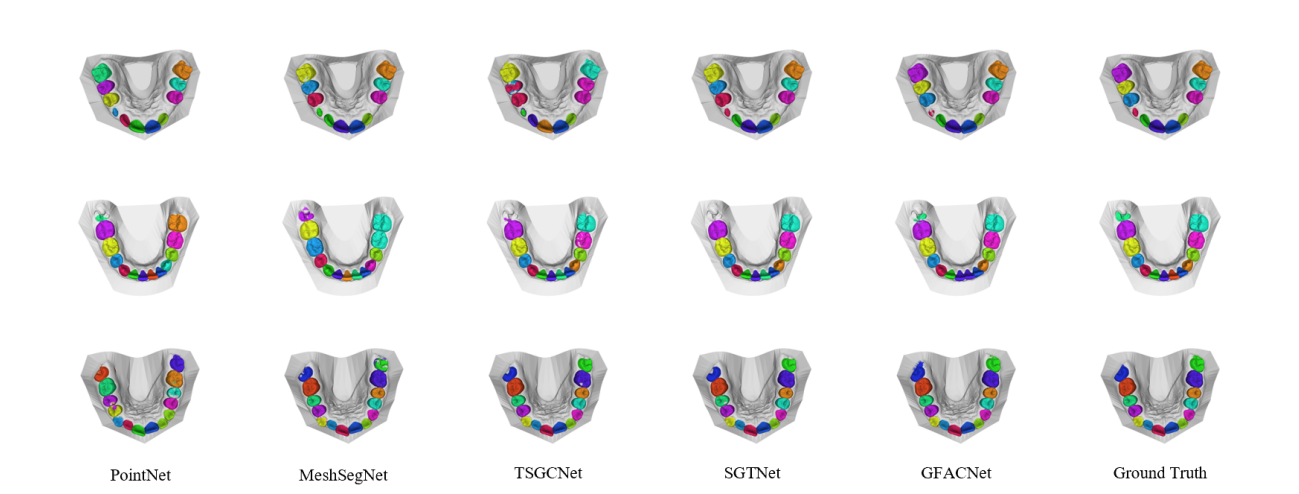

Figure 3. Qualitative comparison of segmentation results.

Figure 4 provides zoomed-in comparisons of challenging regions to better demonstrate GFACNet's advantages. The left case shows crowded anterior teeth where DBGANet exhibits boundary ambiguity, while GFACNet maintains precise boundary delineation through curvature-aware sampling. The right case presents a malformed tooth where DBGANet produces fragmented segmentation, whereas GFACNet achieves anatomically consistent boundaries through anatomical constraints.

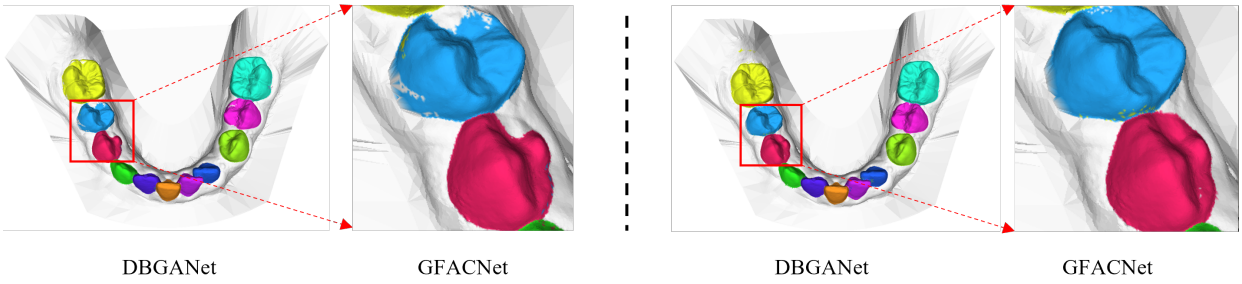

Figure 4. Zoomed-in comparison in challenging regions. Left: Crowded teeth. Right: Malformed tooth. Red boxes indicate magnified regions. GFACNet demonstrates superior boundary accuracy and anatomical consistency.

4.3. Ablation studies

To validate the effectiveness of each component in GFACNet, we conducted comprehensive ablation experiments examining both segmentation performance and computational efficiency. We designed the following variants for comparison:

1) Baseline: Using original point features and traditional GCN structure without specialized modules.
2) GFACNet w/ FPS+KNN: Replacing the MAGC module with traditional farthest point sampling (FPS) and fixed K-Nearest Neighbors (KNN).
3) GFACNet w/ self-attention: Replacing the MST module's multi scale feature integration with standard self-attention mechanism.

4) GFACNet w/ GAT: Replacing the MST module with graph attention network (GAT).

5) GFACNet w/ CE loss: Using only cross-entropy loss without additional anatomical constraints.

6) GFACNet (full): Complete model.

4.3.1. Performance analysis

The ablation study results in Tables 3 and 4 demonstrate that each innovative component of GFACNet contributes significantly to both segmentation accuracy and computational efficiency. The baseline variant, while capable of basic dental segmentation, exhibits limited accuracy with an mIoU of 84.61%. This establishes the necessity for specialized architectural designs in dental segmentation tasks.

MAGC module effectiveness: Replacing our MAGC module with traditional FPS+KNN sampling leads to notable performance degradation, with mIoU dropping from 88.15% to 86.25%. More importantly, this traditional approach significantly increases computational cost, requiring 5.91 G FLOPs compared to our method's 4.73 G FLOPs (20.0% reduction) and consuming 10.84 GB GPU memory versus 9.16 GB (15.5% improvement). This demonstrates that our morphology-aware adaptive graph construction not only improves segmentation accuracy by focusing on geometrically important regions such as dental boundaries, but also enhances computational efficiency.

MST module optimization: Our multi scale transformer feature integration proves superior to both standard self-attention and GAT approaches. The MST module achieves 88.15% mIoU compared to 86.73% with self-attention and 87.32% with GAT, while maintaining competitive computational cost (4.73 G FLOPs vs. GAT's 5.47 G FLOPs). This indicates that multi scale feature integration is crucial for simultaneously understanding local details and global structures in complex dental morphologies.

HACL Impact: The hierarchical anatomical constraint loss provides significant accuracy improvement, boosting mIoU from 87.69% to 88.15% with minimal computational overhead. Since HACL primarily affects the training phase through anatomical constraints rather than inference computation, it delivers improved segmentation consistency for dental anatomical structures without compromising efficiency.

To further validate the effectiveness of individual components within the MAGC module, we conducted detailed ablation studies on its two key subcomponents: edge feature representation and adaptive feature aggregation. Table 5 presents the results of this analysis.

As shown in Table 5, both subcomponents contribute significantly to the overall performance. Removing the edge feature representation results in a 0.46% drop in mIoU, demonstrating its importance in capturing geometric relationships at tooth boundaries. The adaptive feature aggregation mechanism, when removed, causes a 0.15% mIoU decrease, indicating its role in selectively weighting neighbor contributions. The full MAGC module, combining both components, achieves optimal performance across all metrics, validating our design choices.

4.3.2. Computational efficiency analysis

The efficiency analysis reveals that each component contributes to both accuracy and computational optimization. Our complete GFACNet achieves the best performance while requiring only 4.73 G FLOPs and 9.16 GB GPU memory, with an inference time of 38 ms. This represents substantial improvements over alternative approaches, particularly in handling challenging cases such as malformed teeth and complex dental boundaries where traditional methods often fail.

The results confirm that our architectural innovations work synergistically—the MAGC module's intelligent sampling reduces computational load while preserving critical geometric information, the MST module efficiently integrates multi scale features, and the HACL ensures anatomically consistent results without inference overhead.

Table 3. Ablation study results with computational efficiency analysis.

Method	OA (%)	mIoU (%)	mAcc (%)	FLOPs (G)	GPU memory (GB)	Time (ms)
Baseline	92.38	84.61	86.27	4.89	8.73	42
GFACNet w/ FPS+KNN	93.42	86.25	87.45	5.91	10.84	46
GFACNet w/ Self-attention	93.85	86.73	87.96	5.28	9.45	43
GFACNet w/ GAT	94.10	87.32	88.24	5.47	9.72	44
GFACNet w/ CE Loss	94.45	87.69	88.61	4.73	9.16	38
GFACNet (Full)	94.78	88.15	88.93	4.73	9.16	38

Table 4. Component-wise performance and efficiency comparison.

Component comparison	OA (%)	mIoU (%)	mAcc (%)	FLOPs (G)	Memory (GB)	Time (ms)
Standard graph construction	93.42	86.25	87.45	5.91	10.84	46
MAGC (Ours)	94.12	87.58	88.21	4.85	9.22	39
Vanilla transformer	93.85	86.73	87.96	5.28	9.45	43
MST (Ours)	94.31	87.94	88.67	4.79	9.18	39
Cross-entropy loss	94.45	87.69	88.61	4.73	9.16	38
HACL (Ours)	94.78	88.15	88.93	4.73	9.16	38

Table 5. Ablation study on MAGC sub-components.

Configuration	OA (%)	mIoU (%)	mAcc (%)	CD (mm)
w/o Edge feature	93.89	87.12	88.35	0.259
w/o Adaptive aggregation	94.05	87.43	88.52	0.253
MAGC (Full)	94.12	87.58	88.21	0.248

4.3.3. Comparison with alternative multi scale fusion strategies

To validate the superiority of our MST module, we compared it against classical multi scale fusion architectures: feature pyramid network (FPN) with top-down pathways and UNet++ with nested skip connections. We replaced MST with these methods while keeping MAGC and HACL unchanged. Table 6 presents the quantitative comparison.

As shown in Table 6, MST outperforms FPN by 1.26% in mIoU with 8.7% lower FLOPs, and surpasses UNet++ by 0.94% in mIoU with 16.1% fewer FLOPs. The inference time is also 13.6% and 19.1% faster than FPN and UNet++ respectively.

The superiority stems from architectural differences: FPN's sequential top-down processing limits cross-scale information flow, while UNet++'s dense connections introduce computational redundancy. Our MST enables simultaneous interaction across all scales through unified cross-scale attention (Eqs (3.16)–(3.20)), capturing richer dependencies while maintaining efficiency for complex dental geometry segmentation.

Table 6. Performance comparison of different multi scale fusion strategies.

Fusion strategy	OA (%)	mIoU (%)	mAcc (%)	FLOPs (G)	Time (ms)
FPN	93.96	86.89	88.07	5.18	44
UNet++	94.08	87.21	88.31	5.64	47
MST (Ours)	94.78	88.15	88.93	4.73	38

Table 7. Analysis of attention intensity in different regions.

Scale	Boundary	Non-boundary	Ratio
$r_1 = 0.03$	0.78	0.45	1.73
$r_2 = 0.06$	0.72	0.51	1.41
$r_3 = 0.12$	0.59	0.54	1.09
$r_4 = 0.24$	0.53	0.52	1.02

4.4. Attention mechanism analysis

To gain deeper insights into the role of the attention mechanism in the MST module, we analyzed the attention intensity at different scales, focusing on dental boundary regions and non-boundary regions. The results are shown in Table 7.

Table 7 shows that attention at smaller scales has significantly higher intensity in boundary regions compared to non-boundary regions, with a ratio of 1.73 at the smallest scale ($r_1 = 0.03$). This ratio decreases as the scale increases, becoming nearly equal (1.02) at the largest scale ($r_4 = 0.24$). These findings indicate that the MST module automatically focuses on key regions in segmentation, with smaller scales attending to detailed boundary features and larger scales capturing overall tooth structures. This adaptive attention allocation mechanism is a key factor enabling GFACNet to accurately process complex dental boundaries while maintaining global structural awareness.

To further validate the training stability of the dynamic KNN parameters α and β, we visualize their learning trajectories throughout the training process in Figure 5. As shown, both parameters converge smoothly after approximately 150 epochs, with α stabilizing around 22.3 and β around 11.7. The convergence behavior aligns well with the validation mIoU improvement, demonstrating that the learned parameters effectively capture optimal neighborhood size allocation based on local geometric complexity. Notably, the parameters exhibit stable convergence without oscillation, confirming the robustness of our adaptive graph construction mechanism.

4.5. Comparison of different sampling strategies

To further validate the effectiveness of the sampling strategy in the MAGC module, we compared the impact of different sampling methods on segmentation performance, with results shown in Table 8.

As shown in Table 8, our proposed curvature-aware sampling strategy significantly outperforms other sampling methods under the same number of sampling points (1500). The mIoU of our MAGC module reaches 89.87%, which is 4.14%, 2.56%, and 1.45% higher than random sampling, FPS, and density-based sampling, respectively. Notably, even compared to FPS with twice the number of sampling points (3000), our method still achieves better segmentation performance (89.87% vs. 88.95%) while substantially reducing computational overhead (38 ms vs. 58 ms inference time and 10,719 MB vs. 16,573 MB GPU memory).

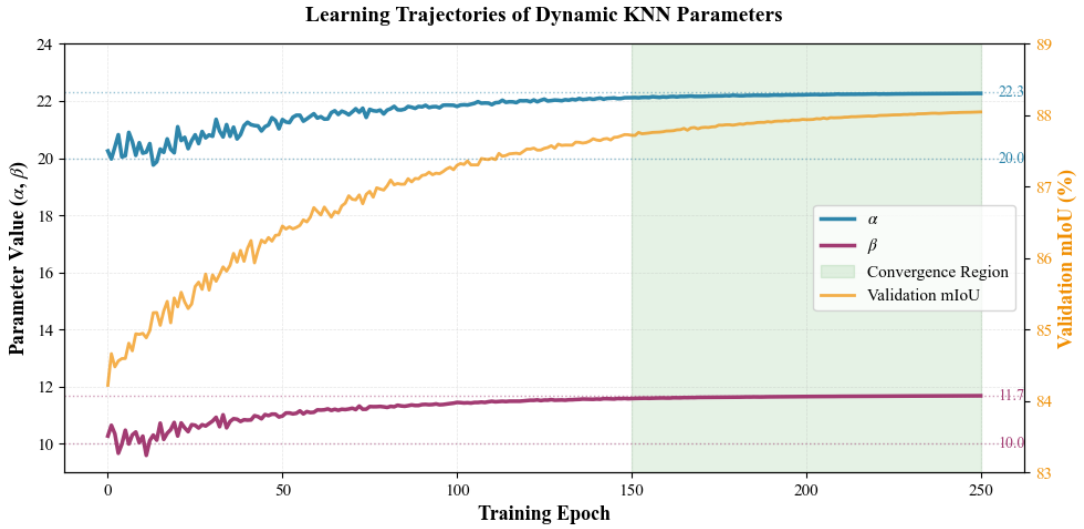

Figure 5. Learning trajectories of dynamic KNN parameters α and β during training.

Table 8. Performance comparison of different sampling strategies.

Sampling strategy	Points	mIoU (%)	Memory (MB)	Time (ms)
Random	1500	85.73	9782	37
FPS	1500	87.31	10,215	38
Density-based	1500	88.42	10,564	38
Curvature-aware	1500	89.87	10,719	38
FPS	3000	88.95	16,573	58

This demonstrates the efficiency of the MAGC module, which intelligently allocates computational resources to regions with rich geometric information while avoiding redundant computation. The results indicate that sampling strategy has a significant impact on 3D dental segmentation performance, and our proposed MAGC module effectively improves both segmentation accuracy and computational efficiency through curvature-aware sampling.

4.6. Parameter sensitivity analysis

To validate the robustness of our method to hyperparameter variations and justify our parameter choices, we conducted sensitivity analysis on key parameters including the curvature smoothing factor ϵ, dynamic KNN range, number of scales in MST module, attention heads, and loss weights in HACL module. The results are presented in Table 9.

The sensitivity analysis demonstrates that our parameter choices are well-justified. The curvature smoothing factor $\epsilon = 0.01$ achieves optimal performance by effectively balancing noise reduction and geometric feature preservation. Values too small ($\epsilon < 0.01$) fail to adequately smooth geometric noise, while larger values ($\epsilon > 0.02$) over-smooth important geometric details. Similarly, the dynamic KNN range $K_{min} = 16$, $K_{max} = 48$ provides the best trade-off between local detail capture and computational efficiency.

For the MST module, using 4 scales ($S = 4$) achieves the best performance, as it effectively captures

multi scale information from local details to global structure. Fewer scales (2 and 3) miss important geometric variations, while more scales (5) introduce redundant information without significant gains. The choice of 8 attention heads balances model capacity and computational cost, with performance degrading slightly at both lower (4, 6) and higher (12) values.

Table 9. Parameter sensitivity analysis.

Parameter	Setting	Value	OA (%)	mIoU (%)	mAcc (%)
ϵ	Setting 1	0.005	94.52	87.89	88.67
	Setting 2	0.008	94.71	88.02	88.81
	Ours	0.01	94.78	88.15	88.93
	Setting 3	0.02	94.69	87.98	88.79
	Setting 4	0.05	94.43	87.66	88.45
K range	Setting 5	[8, 24]	94.21	87.43	88.29
	Setting 6	[12, 36]	94.56	87.92	88.71
	Ours	[16, 48]	94.78	88.15	88.93
	Setting 7	[20, 60]	94.69	87.98	88.79
Scales S	Setting 8	2	94.23	87.31	88.42
	Setting 9	3	94.51	87.76	88.68
	Ours	4	94.78	88.15	88.93
	Setting 10	5	94.61	87.89	88.74
Attn heads	Setting 11	4	94.45	87.68	88.59
	Setting 12	6	94.62	87.91	88.78
	Ours	8	94.78	88.15	88.93
	Setting 13	12	94.71	87.98	88.82
λ_1, λ_2	Setting 14	0.3, 0.2	94.59	87.81	88.71
	Setting 15	0.4, 0.3	94.67	87.93	88.79
	Ours	0.5, 0.3	94.78	88.15	88.93
	Setting 16	0.6, 0.4	94.63	87.87	88.76

Regarding the HACL loss weights, our selection of $\lambda_1 = 0.5$ and $\lambda_2 = 0.3$ provides optimal balance between centroid and shape constraints. Lower values ($\lambda_1 = 0.3-0.4$) underweight anatomical guidance, while higher values ($\lambda_1 = 0.6$, $\lambda_2 = 0.4$) over-constrain the model, limiting its flexibility in handling anatomical variations. These results confirm that our hyperparameter configuration is robust and well-tuned for dental segmentation tasks.

4.7. Failure case analysis

While GFACNet achieves strong overall performance, analyzing failure cases reveals important limitations. Figure 6 illustrates two representative failure modes observed in our validation set.

The left case shows severe dental crowding with overlapping anterior teeth. GFACNet's prediction exhibits boundary ambiguity in the overlapped region (marked with red circles), where the model struggles to accurately delineate the junction between tightly crowded teeth. This under-segmentation occurs because extreme crowding (< 0.5 mm inter-tooth spacing) reduces geometric distinctiveness. The right case presents extensive dental restorations (marked with red circles). The metallic or ceramic

restorations introduce artificial geometric patterns that deviate from natural tooth morphology, causing the MAGC module to incorrectly interpret restoration-induced surface variations as tooth boundaries, leading to segmentation inconsistencies.

Quantitatively, these challenging scenarios account for 8.3% of validation cases where mIoU drops below 80%. Extreme crowding cases (2.1% of dataset) achieve 76.4% mIoU, while extensive restoration cases (6.2% of dataset) achieve 78.9% mIoU, compared to our overall 88.15% performance.

Despite these limitations, GFACNet maintains clinically acceptable performance (> 85% mIoU) for 91.7% of cases, confirming its practical utility while highlighting specific scenarios requiring human oversight.

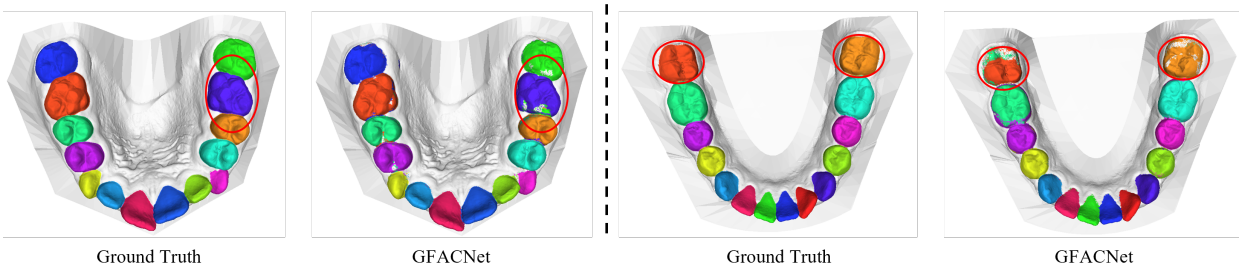

Figure 6. Representative failure cases of GFACNet. Left: Severe crowding causes boundary detection failures (red circles). Right: Extensive restorations introduce segmentation inconsistencies (red circles). Ground truth vs. GFACNet prediction for each case.

4.8. Clinical application analysis

Beyond segmentation accuracy, practical clinical deployment requires consideration of computational efficiency and workflow integration.

Our method's computational efficiency directly benefits clinical practice. With 4.73 G FLOPs and 38 ms inference time, GFACNet enables real-time processing during patient consultations, completing dental arch segmentation within seconds. The reduced computational requirements allow deployment on mid-range workstations (e.g., NVIDIA RTX 3060), lowering hardware barriers for smaller clinics compared to methods like Point Transformer (18.45 G FLOPs) or SGTNet (9.67 G FLOPs).

GFACNet integrates seamlessly with existing workflows. It accepts standard mesh formats (STL, OBJ, PLY) from mainstream intraoral scanners including Primescan (Dentsply), Trios3 (3Shape), and iTero Element, requiring no additional preprocessing. The FDI-based output (Figure 2) enables direct use in orthodontic appliance design, treatment simulation, and patient communication tools.

The method handles common challenging scenarios effectively, including missing teeth, malformations, and crowding, as demonstrated by strong performance on difficult categories (T15/T45, T28/T38). However, performance may degrade with extensive dental restorations or metal artifacts, which were underrepresented in training data. Future work will address these limitations through dataset expansion and clinic-specific adaptation techniques.

5. Conclusions

In this paper, we presented GFACNet, a novel deep learning approach for 3D dental scan segmentation and labeling that integrates geometric features and anatomical constraints. Our method tackles the

challenges of complex tooth morphology, ambiguous gingival boundaries, and patient-specific anomalies through a structured three-stage pipeline.

First, our morphology-aware graph construction mechanism dynamically adapts to tooth geometry, focusing computational resources on critical regions such as tooth boundaries. Second, our multi scale transformer feature integration effectively captures both local details and global context, enabling precise crown delineation. Finally, our hierarchical anatomical constraint loss incorporates dental domain knowledge into the learning process, producing segmentation results that respect natural tooth arrangements and anatomical properties. Extensive experiments on the Teeth3DS benchmark demonstrated that GFACNet significantly outperforms state-of-the-art methods across all evaluation metrics. Particularly, our approach achieved higher accuracy while requiring approximately 40% fewer sampling points, making it more computationally efficient for clinical applications. The performance advantages were especially pronounced when handling challenging cases such as crowded dentition, malformed teeth, and missing teeth, which are common in clinical practice. Future research directions include extending the model to handle additional clinical scenarios, such as dental implants and orthodontic appliances, as well as exploring the integration of our method with other dental imaging modalities for comprehensive treatment planning. While GFACNet is specifically designed for dental segmentation, the core principles of our MAGC mechanism—curvature-guided adaptive sampling and dynamic neighborhood construction—have potential applications in other non-rigid surface segmentation tasks. For instance, organs with complex geometries (e.g., brain cortex, cardiac structures) or irregular anatomical structures (e.g., bone surfaces, vascular networks) could benefit from morphology-aware graph construction. Future work will explore extending this approach to broader medical imaging domains and general 3D shape analysis tasks.

Data availability

The 3DTeethSeg22 dataset used in this study is publicly available at the official challenge website (https://osf.io/xctdy/overview).

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Author contributions

Conceptualization: G.Z. and A.S.; Methodology: G.Z. and X.C.; Software: G.Z. and M.L.; Validation: G.Z., X.C. and M.L.; Formal analysis: G.Z.; Investigation: G.Z. and A.S.; Data curation: G.Z. and M.L.; Writing—original draft: G.Z.; Writing—review and editing: all authors; Visualization: G.Z. and M.L.; Supervision: A.S.; Project administration: A.S. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors would like to thank the organizers of the 3DTeethSeg'22 Challenge for providing the benchmark dataset and the anonymous reviewers for their valuable comments.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. V. Majanga, S. Viriri, A survey of dental caries segmentation and detection techniques, *Sci. World J.*, **2022** (2022), 8415705. https://doi.org/10.1155/2022/8415705
2. Z. Cui, C. Li, N. Chen, G. Wei, R. Chen, Y. Zhou, et al., TSegNet: An efficient and accurate tooth segmentation network on 3D dental model, *Med. Image Anal.*, **69** (2020), 101949. https://doi.org/10.1016/j.media.2020.101949
3. S. Hou, T. Zhou, Y. Liu, P. Dang, H. Lu, H. Shi, Teeth U-Net: A segmentation model of dental panoramic X-ray images for context semantics and contrast enhancement, *Comput. Biol. Med.*, **152** (2023), 106296. https://doi.org/10.1016/j.compbiomed.2022.106296
4. T. J. Jang, K. C. Kim, H. C. Cho, J. K. Seo, A fully automated method for 3D individual tooth identification and segmentation in dental CBCT, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2021), 6562–6568. https://doi.org/10.1109/TPAMI.2021.3086072
5. T. J. Jang, H. S. Yun, C. M. Hyun, J. Kim, S. Lee, J. K. Seo, Fully automatic integration of dental CBCT images and full-arch intraoral impressions with stitching error correction via individual tooth segmentation and identification, *Med. Image Anal.*, **93** (2024), 103096. https://doi.org/10.1016/j.media.2024.103096
6. S. Kondo, W. Morita, H. Ohshima, The biological significance of tooth identification based on developmental and evolutional viewpoints, *J. Oral Biosci.*, **64** (2022), 287–302. https://doi.org/10.1016/j.job.2022.05.004
7. S. Zhuang, G. Wei, Z. Cui, Y. Zhou, Robust hybrid learning for automatic teeth segmentation and labeling on 3D dental models, *IEEE Trans. Multimedia*, **27** (2025), 792–803. https://doi.org/10.1109/TMM.2023.3289760
8. T. Ari, H. Sağlam, H. Öksüzoğlu, O. Kazan, İ. Ş. Bayrakdar, S. B. Duman, et al., Automatic feature segmentation in dental periapical radiographs, *Diagnostics*, **12** (2022), 3081. https://doi.org/10.3390/diagnostics12123081
9. B. M. Elgarba, S. Van Aelst, A. Swaity, N. Morgan, S. Shujaat, R. Jacobs, Deep learning-based segmentation of dental implants on cone-beam computed tomography images: A validation study, *J. Dent.*, **137** (2023), 104639. https://doi.org/10.1016/j.jdent.2023.104639
10. B. Y. Tekin, C. Ozcan, A. Pekince, Y. Yasa, An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs, *Comput. Biol. Med.*, **146** (2022), 105547. https://doi.org/10.1016/j.compbiomed.2022.105547

11. T. Wu, C. Lian, S. Lee, M. Pastewait, C. Piers, J. Liu, et al., Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3D intraoral scans, *IEEE Trans. Med. Imaging*, **41** (2022), 3158–3166. https://doi.org/10.1109/TMI.2022.3180343
12. Z. Li, T. Liu, J. Wang, C. Zhang, X. Jia, Multi-scale bidirectional enhancement network for 3D dental model segmentation, in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, (2022), 1–5. https://doi.org/10.1109/ISBI52829.2022.9761556
13. C. R. Qi, H. Su, K. Mo, L. J. Guibas, Pointnet: Deep learning on point sets for 3D classification and segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 652–660.
14. C. R. Qi, L. Yi, H. Su, L. J. Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space, in *Advances in Neural Information Processing Systems*, **30** (2017).
15. C. Lian, L. Wang, T. Wu, F. Wang, P. Yap, C. Ko, et al., Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3D intraoral scanners, *IEEE Trans. Med. Imaging*, **39** (2020), 2440–2450. https://doi.org/10.1109/TMI.2020.2971730
16. R. Xie, Y. Yang, Z. Chen, WITS: Weakly-supervised individual tooth segmentation model trained on box-level labels, *Pattern Recognit.*, **133** (2023), 108974. https://doi.org/10.1016/j.patcog.2022.108974
17. A. Haghanifar, M. M. Majdabadi, S. Haghanifar, Y. Choi, S. Ko, PaXNet: Tooth segmentation and dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier, *Multimed. Tools Appl.*, **82** (2023), 27659–27679. https://doi.org/10.1007/s11042-023-14435-9
18. Y. Li, R. Bu, M. Sun, W. Wu, X. Di, B. Chen, Pointcnn: Convolution on x-transformed points, in *Advances in Neural Information Processing Systems*, **31** (2018).
19. Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, J. M. Solomon, Dynamic graph CNN for learning on point clouds, *ACM Trans. Graph.*, **38** (2019), 1–12. https://doi.org/10.1145/3326362
20. H. Zhao, L. Jiang, J. Jia, P. H. S. Torr, V. Koltun, Point transformer, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), 16259–16268.
21. M. Guo, J. Cai, Z. Liu, T. Mu, R. R. Martin, S. Hu, PCT: Point cloud transformer, *Comput. Vis. Media*, **7** (2021), 187–199. https://doi.org/10.1007/s41095-021-0229-5
22. X. Ma, C. Qin, H. You, H. Ran, Y. Fu, Rethinking network design and local geometry in point cloud: A simple residual MLP framework, preprint, arXiv:2202.07123.
23. G. Qian, Y. Li, H. Peng, J. Mai, H. Hammoud, M. Elhoseiny, et al., Pointnext: Revisiting pointnet++ with improved training and scaling strategies, *Advances in Neural Information Processing Systems*, **35** (2022), 23192–23204.
24. Y. Ma, Z. Li, Computer aided orthodontics treatment by virtual segmentation and adjustment, in *2010 International Conference on Image Analysis and Signal Processing*, (2010), 336–339. https://doi.org/10.1109/IASP.2010.5476100
25. Y. Kumar, R. Janardan, B. Larson, J. Moon, Improved segmentation of teeth in dental models, *Comput. Aided Des. Appl.*, **8** (2011), 211–224. https://doi.org/10.3722/cadaps.2011.211-224
26. N. Wongwaen, C. Sinthanayothin, Computerized algorithm for 3D teeth segmentation, *Proc. Int. Conf. Electron. Inf. Eng.*, **1** (2010), 277–280. https://doi.org/10.1109/ICEIE.2010.5559877

27. B. Graham, M. Engelcke, L. Van Der Maaten, 3D semantic segmentation with submanifold sparse convolutional networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 9224–9232.

28. J. Zhang, C. Li, Q. Song, L. Gao, Y. Lai, Automatic 3D tooth segmentation using convolutional neural networks in harmonic parameter space, *Graph. Models*, **109** (2020), 101071. https://doi.org/10.1016/j.gmod.2020.101071

29. Z. Lin, Z. He, X. Wang, B. Zhang, C. Liu, W. Su, et al., DBGANet: Dual-branch geometric attention network for accurate 3D tooth segmentation, *IEEE Trans. Circuits Syst. Video Technol.*, **34** (2024), 4285–4298. https://doi.org/10.1109/TCSVT.2023.3331589

30. A. Rekik, A. Ben-Hamadou, O. Smaoui, F. Bouzguenda, S. Pujades, E. Boyer, TSegLab: Multi-stage 3D dental scan segmentation and labeling, *Comput. Biol. Med.*, **185** (2025), 109535. https://doi.org/10.1016/j.compbiomed.2024.109535

31. P. Li, C. Gao, F. Liu, D. Meng, Y. Yan, THISNet: Tooth instance segmentation on 3D dental models via highlighting tooth regions, *IEEE Trans. Circuits Syst. Video Technol.*, **34** (2023), 5229–5241. https://doi.org/10.1109/TCSVT.2023.3341805

32. B. Zou, S. Wang, H. Liu, G. Sun, Y. Wang, F. Zuo, et al., Teeth-SEG: An efficient instance segmentation framework for orthodontic treatment based on multi-scale aggregation and anthropic prior knowledge, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2024), 11601–11610.

33. X. Chen, H. Tao, H. Zhou, P. Zhou, Y. Deng, Hierarchical and progressive learning with key point sensitive loss for sonar image classification, *Multimedia Syst.*, **30** (2024), 380. https://doi.org/10.1007/s00530-024-01590-8

34. Y. Apedo, H. Tao, A weakly supervised pavement crack segmentation based on adversarial learning and transformers, *Multimedia Syst.*, **31** (2025), 380. https://doi.org/10.1007/s00530-025-01850-1

35. F. Naz, M. A. Anwar, M. Qasim, S. S. Bavabeedu, S. U. Khateeb, A. A. Khalil, et al., Clinical evaluation of composite restorations using the Fédération Dentaire Internationale (FDI) criteria: A retrospective cross-sectional study at University Dental Hospital, The University of Lahore, Pakistan, *Cureus*, **17** (2025), e83582. https://doi.org/10.7759/cureus.83582

36. S. Liu, J. Chi, C. Wu, F. Xu, X. Yu, SGT-Net: A transformer-based stratified graph convolutional network for 3D point cloud semantic segmentation, *Comput. Mater. Continua*, **79** (2024). https://doi.org/10.32604/cmc.2024.049450

37. L. Qiu, C. Ye, P. Chen, Y. Liu, X. Han, S. Cui, Darch: Dental arch prior-assisted 3D tooth instance segmentation with weak annotations, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), 20752–20761.

38. Y. Zhao, Y. Liu, Y. Tan, L. Zhang, T. Huang, C. Gao, 3D dental model segmentation with graph attentional convolution network, *Pattern Recognit. Lett.*, **152** (2021), 79–85. https://doi.org/10.1016/j.patrec.2021.09.005

39. M. Fey, J. E. Lenssen, F. Weichert, H. Müller, SplineCNN: Fast geometric deep learning with continuous B-spline kernels, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 869–877.

40. Y. Kumar, R. Janardan, B. Larson, J. Moon, Improved segmentation of teeth in dental models, *Comput. Aided Des. Appl.*, **8** (2011), 211–224. https://doi.org/10.3722/cadaps.2011.211-224

AIMS Press