ITC 2/55

Information Technology and Control
Vol. 55 / No. 2/ 2026
pp. 634-647
DOI 10.5755/j01.itc.55.2.44156

Semi-Supervised Tooth Instance Segmentation from Panoramic X-ray Images

Received 2026/01/26 | Accepted after revision 2026/05/15

HOW TO CITE: Zhang, K., Che, Q., Sun, J., Lin, S. (2026) Semi-Supervised Tooth Instance Segmentation from Panoramic X-ray Images. *Information Technology and Control*, 55(2), 634-647. https://doi.org/10.5755/j01.itc.55.2.44156

Semi-Supervised Tooth Instance Segmentation from Panoramic X-ray Images

Kang Zhang

School of Software, Shandong University, Jinan, Shandong, China, 250100; School of Information Engineering, Changji University, Changji, Xinjiang, China, 831100; e-mail: zhkangzhkang@163.com

Quanjiang Che, Jie Sun*

School of Information Engineering, Changji University, Changji, Xinjiang, China, 831100; e-mails: 21414kacar66@cjc.edu.cn; cjxyedu@139.com

Shiming Lin

School of Informatics, Xiamen University, Xiamen, Fujian, China 361104; School of Information Engineering, Changji University, Changji, Xinjiang, China, 831100; e-mail: 19036@cjc.edu.cn

Corresponding author: cjxyedu@139.com

Tooth image segmentation is a critical task in medical image analysis, where precise segmentation outcomes are essential for oral diagnosis and treatment planning. Dental X-ray images often present challenges such as boundary blurring, tooth misalignment, and complex anatomical variations, which hinder accurate tooth edge delineation, especially for wisdom tooth assessment. This study integrates semi-supervised learning with the nnU-Net deep learning framework and proposes a novel tooth image segmentation method that incorporates a dual-attention mechanism. By utilizing a limited set of labeled data alongside abundant unlabeled samples, the proposed approach effectively handles complex backgrounds and diverse tooth morphologies. Experimental evaluations show superior performance, with an average Dice coefficient of 92.05%, mean Intersection over Union (mIoU) of 86.27%, Normalized Surface Distance (NSD) of 94.72%, and Identification Accuracy (IA) of 85.48%. This method enhances diagnostic precision for oral clinicians, streamlines dental image analysis workflows, and holds significant potential for advancing digital dentistry in diagnosis and treatment.

KEYWORDS: Tooth segmentation, semi-supervised learning, nnU-Net, pseudo-label.

1. Introduction

With the advancement of medical standards and the widespread adoption of imaging equipment in grassroots hospitals, medical imaging data has become a cornerstone for pathological diagnosis, underpinning approximately 80% of clinical decisions [6]. Panoramic radiographs, widely utilized for dental condition assessments, play a crucial role in clinical dentistry for diagnosis, treatment, and surgical planning. The precise and automated segmentation of anatomical structures from these radiographs is a key enabler of AI-driven digitization in oral medicine. However, inherent challenges such as overlapping anatomical structures, blurred boundaries, and artifacts from the hard palate, cervical spine, or bite blocks pose significant difficulties, particularly in wisdom tooth evaluations, where boundary precision is essential to distinguish malformed or impacted teeth. These challenges often lead to misdiagnosis rates exceeding 20% in complex cases.

In oral medicine, dental image segmentation technology holds substantial clinical value for pathology diagnosis and treatment planning. Traditional approaches, such as Tuan et al.'s (2017) semi-supervised fuzzy clustering with spatial constraints for X-ray segmentation [15], Mahoor and Abdel-Mottaleb's (2004) automatic classification of bitewing images [10], and Im et al.'s (2022) deep learning-based tooth segmentation in digital dental models [5], have laid foundational work. Similarly, Sornam and Prabhakaran (2017) integrated linear adaptive swarm optimization with backpropagation neural networks to classify caries [14], while Geetha et al. (2020) employed backpropagation for caries diagnosis in periapical X-rays [3], and Vigil and Bharathi (2021) used 2D Otsu thresholding for periodontitis assessment via pocket depth in panoramic images [16]. However, these methods rely on manual feature extraction and simplistic thresholding, achieving limited generalization and unreliable edge delineation in clinical settings due to the intricate backgrounds and blurred boundaries of panoramic radiographs.

Recent progress in deep learning has revolutionized this domain, offering advanced segmentation paradigms. Wirtz et al. (2018) combined coupled shape models with neural networks, leveraging gradient features to enhance segmentation in panoramic radiographs [18], while Koch et al. (2019) improved robustness with a patch-based U-Net variant [7]. Pinheiro et al. (2021) advanced boundary precision using Mask R-CNN with PointRend [12], yet challenges persist with blurred boundaries and complex anatomical overlaps.Innovations like Lin et al.'s (2023) lightweight deep learning models for panoramic dental X-ray segmentation [8] and Zhang et al.'s (2022) BDU-Net with border guidance [23] have improved accuracy in complex structures, but their dependence on extensive labeled data, which is often scarce in dental datasets, leads to overfitting and suboptimal performance on low-contrast images. This study overcomes these limitations by integrating semi-supervised learning with nnU-Net and a dual-attention mechanism, enabling effective utilization of abundant unlabeled data and precise feature focusing to achieve superior boundary accuracy and robustness across diverse clinical scenarios.

The contributions of this work are as follows:

1. To address data scarcity and enhance model robustness, we propose a five-fold cross-validation strategy integrated with a semi-supervised learning framework, solving the problem of overfitting in limited labeled dental datasets.
2. To address complex backgrounds and blurred boundaries, we introduce tailored post-processing operations after network training, refining segmentation results and mitigating artifacts, thereby improving clinical applicability.
3. For precise tooth edge delineation in low-contrast regions, we incorporate innovative attention mechanisms (RFEM and CBAM) into the nnU-Net architecture, improving segmentation accuracy by focusing on critical dental features and overcoming the challenge of inaccurate boundary detection in baseline models.

2. Related Work

The evolution of medical image segmentation has transitioned from traditional techniques to advanced deep learning paradigms, laying the foundation for the innovations proposed in this study. This section reviews key advancements in four critical areas—2D segmentation, semi-supervised learning, nnU-Net, and attention mechanisms in dental image segmentation—highlighting their contributions and limitations to contextualize our proposed approach.

2.1. 2D Segmentation

Two-dimensional (2D) image segmentation is a cornerstone of medical imaging, with applications ranging from organ to dental structure analysis. Early methods, such as thresholding, edge detection, and region growing, relied on traditional image processing but struggled with complex backgrounds, overlapping structures, and blurred boundaries—particularly in dental X-rays—due to their reliance on manual feature extraction [1, 5]. The advent of deep learning, particularly convolutional neural networks (CNNs), marked a significant leap forward. Ronneberger et al.'s (2015) U-Net, with its encoder-decoder architecture, effectively captures high-level semantic and low-level spatial features, becoming a benchmark for tasks like organ and dental segmentation [6]. In dental applications, Wirtz et al. (2018) enhanced U-Net with coupled shape models, leveraging prior tooth knowledge to improve segmentation accuracy in panoramic radiographs [18], while Koch et al. (2019) demonstrated enhanced robustness with a block-wise U-Net variant [7]. However, these methods face challenges with overlapping teeth and low-contrast regions, necessitating advanced techniques. This gap drives the exploration of semi-supervised learning to address data scarcity

2.2. Semi-Supervised Learning

Deep learning's dependence on extensive labeled data poses a challenge in medical imaging, where manual annotation is costly and time-intensive. Semi-supervised learning (SSL) addresses this by integrating labeled and unlabeled data, reducing annotation burdens. SSL methods include self-training, where pseudo-labels refine models iteratively, as shown by Lin et al. (2023) in their self-training U-Net with denoiser for tooth segmentation in X-rays [9], and consistency regularization, which ensures prediction stability under perturbations, as demonstrated by Wang et al. (2024) in the STS MICCAI 2023 challenge on 2D/3D tooth segmentation [20]. Recent advancements, such as adversarial training and entropy minimization, further enhance generalization, particularly for dental images with variable quality and anatomy [19]. Despite these gains, issues like pseudo-label noise and limited adaptability to the complexities of panoramic X-rays persist, which our post-processing and semi-supervised framework aim to overcome.

2.3. nnU-Net

nnU-Net introduced by Isensee et al. (2021), is an adaptive framework that automates hyperparameter optimization, excelling in medical image segmentation across diverse tasks [4]. Its encoder-decoder structure, enhanced by fully convolutional networks and skip connections, adapts to small datasets via automated architecture and preprocessing adjustments, outperforming traditional U-Net on various benchmarks [13]. Recent validations, such as Huang et al. (2025), reaffirm its robustness with high-resolution inputs [4]. In dental imaging, its flexibility suits panoramic X-rays, yet its reliance on labeled data and limited focus on boundary refinement hinder performance in complex cases. Our study leverages nnU-Net's adaptability, augmented with semi-supervised learning, to address these shortcomings.

2.4. Attention Mechanisms in Dental Image Segmentation

Attention mechanisms dynamically prioritize key features, significantly enhancing segmentation in dental imaging, especially for overlapping or malformed teeth. Spatial attention, as in Spatial Transformer Networks (STN), improves localization of tooth boundaries and roots [11], while channel attention, exemplified by the Convolutional Block Attention Module (CBAM), refines feature representations [17]. Pan et al. (2025) introduced a multi-scale conv-attention U-Net, enhancing segmentation performance in medical contexts [11], and Chen et al. (2025) applied attention-based U-Net for tooth and root canal segmentation, achieving precise edge detection [2]. Transformer-based approaches, like Wu et al.'s (2024) 3D tooth segmentation via point cloud partitioning [21], leverage global context but are less effective for 2D panoramic X-rays due to computational demands. More recent developments, such as Zhong et al. (2025) with grouped attention for automatic X-ray teeth segmentation [24] and Xu et al. (2025) with uncertainty-aware contrastive learning for tooth instance segmentation [22], demonstrate improved boundary precision and robustness. These methods enhance accuracy, yet struggle with low-contrast boundaries and data scarcity. Our approach extends these innovations with the Refined Feature Enhancement Module (RFEM) for feature compression and CBAM for optimized residual block representations, targeting precise delineation in challenging dental regions.

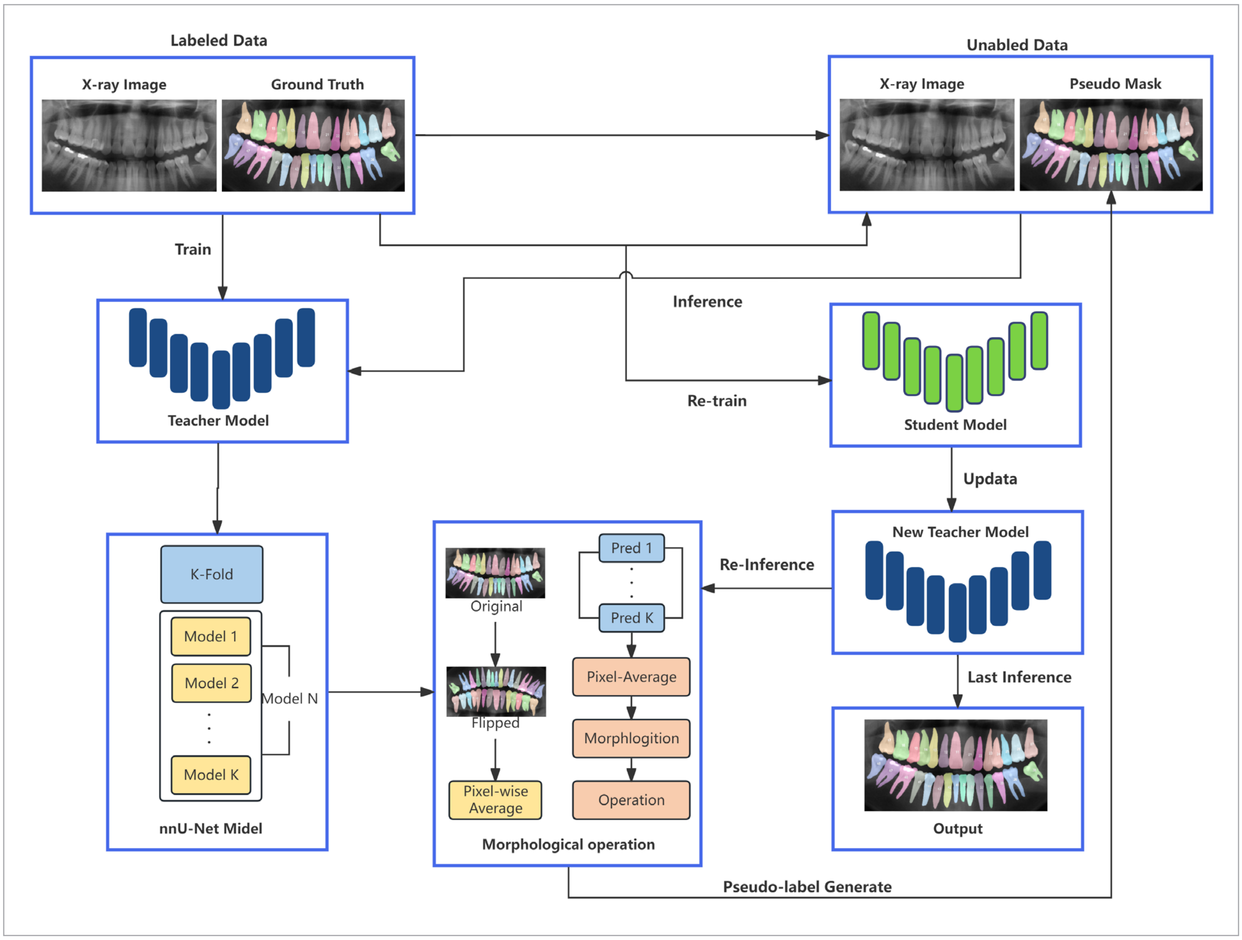
Despite these advancements, challenges such as data scarcity, blurred boundaries, and complex anatomical overlaps in panoramic X-rays remain unresolved. Our semi-supervised nnU-Net framework, enhanced by dual-attention mechanisms, addresses these gaps by effectively utilizing unlabeled data and refining boundary accuracy, offering a robust solution for clinical dental segmentation.

3. Methods

This study presents a semi-supervised nnU-Net instance segmentation framework that integrates pseudo-label generation with a multi-model ensemble strategy to enhance segmentation performance in the presence of limited labeled data. The process begins by training a baseline nnU-Net model on a small labeled dataset. Multiple sub-models are generated through K-fold cross-validation to introduce diversity. Unlabeled data is then augmented using vertical flipping before being passed through the sub-models for prediction. The results are fused via pixel-wise averaging. Post-processing steps, including binarization and morphological operations, are applied to refine the pseudo-labels. These pseudo-labels are combined with the original labeled

Figure 1

Semi-supervised nnU-Net instance segmentation workflow diagram.



data to further refine the final nnU-Net model, enabling precise instance segmentation of unlabeled samples. The framework leverages data augmentation, consistency regularization, and ensemble techniques to enhance pseudo-label accuracy and model robustness. These features make it particularly suitable for medical image segmentation tasks, including dental panoramic image analysis. The complete segmentation workflow is illustrated in Figure 1.

3.1. Data Pre-processing

Data pre-processing is a critical step to ensure robust model training and optimal performance. The pipeline combines image processing, label generation, and management of unlabeled images, all while resizing images to enhance model accuracy, training efficiency, and generalization.

In the image processing phase, original X-ray images are converted to 8-bit single-channel grayscale format. This simplifies the task of delineating tooth morphology and structure. The images are then resized to a standardized resolution, with coordinates adjusted to maintain proportions. Standardization ensures uniform input dimensions, preventing discrepancies due to varying image sizes.

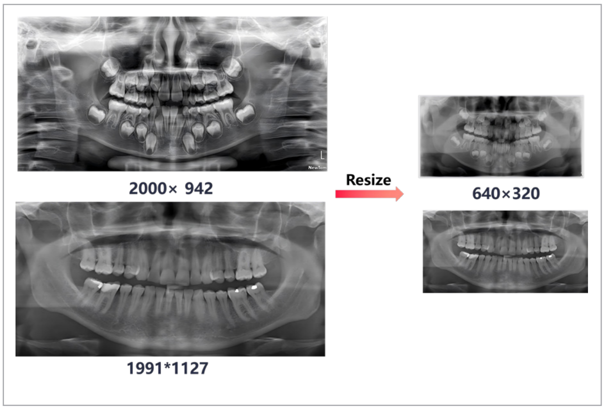
Label generation involves extracting contour coordinates and classification labels from JSON annotations to create label maps. An empty 8-bit grayscale image is initialized with pixel values of 0 representing the background. Then, true instance segmentation is formulated by treating each tooth as a separate class. Tooth classification labels are mapped onto the image, where the corresponding tooth regions are filled with unique integer values (e.g., 1 to 32, matching their specific tooth numbering identifiers). This multi-class encoding creates accurate instance-level label maps for training, distinct from binary segmentation.

For unlabeled images, label maps are initialized as grayscale images with zero pixel values, facilitating their integration into unsupervised learning. This helps expand the training dataset and enhances the model's adaptability to unlabeled samples.

To reduce computational burden and improve training efficiency, all images and labels were uniformly resized to 320 × 640, as shown in Figure 2. This step significantly decreases computational resource consumption, optimizes memory usage, and accelerates

Figure 2

Resize both the image and mask to a reduced dimension.



the training process. The resized dimensions are more compatible with the network input requirements, which helps improve both training efficiency and the final segmentation accuracy.

3.2. Pseudo-Label Generation

Pseudo-label generation stands as a pivotal innovation that harnesses unlabeled data to mitigate scarcity of annotated samples in dental segmentation. The process initiates with training sub-models via K-fold cross-validation on labeled data. Each sub-model predicts on augmented unlabeled images. Predictions fuse through pixel-wise averaging to produce initial maps. Binarization and morphological operations refine these maps into high-fidelity pseudo-labels. This strategy diminishes noise and elevates label quality beyond conventional self-training methods by leveraging ensemble diversity and augmentation consistency.

3.3. Refined Feature Extraction Module (RFEM)

In order to enhance the network's ability to capture key clinical information and suppress irrelevant background noise, we designed the Refined Feature Extraction Module (RFEM). Rather than utilizing multi-scale convolutions, the RFEM is constructed as a channel attention mechanism that compresses effective features and promotes the stability of feature distribution through adaptively learned channel weights.

To ensure complete reproducibility, the exact structural topology and hyperparameters of the RFEM are defined as follows. Given an intermediate feature map $F \in R^{B\times C\times H\times W}$ (where B is the batch size, C is the number of channels, and $H\times W$ is the spatial dimension), the module first compresses the spatial dimensions via an adaptive average pooling operation (Avg). This aggregates the spatial features to obtain a global information descriptor of size 1×1. This descriptor encapsulates the global distribution of channel features. The tensor shape is then adjusted (R_{s_1}) to a flattened vector to improve operational efficiency.

Subsequently, the feature vector undergoes a dimensionality compression and expansion process governed by two linear (fully connected) layers. The first linear layer compresses the channel dimension by a strict reduction ratio of r=16, followed by a ReLU activation function to introduce non-linearity and remove redundant information. The second linear layer restores the original channel dimension C. A Sigmoid activation is then applied, and the tensor shape is readjusted (R_{s_2}) to $B\times C\times1\times1$ to serve as the channel weight allocation mechanism. Finally, the input feature F is multiplied by these generated channel weights element-wise to obtain the enhanced result F_{RF}.

The process of computing the refined features $F_{refined}$ can be described as follows:

$$F_c = linear_1\left(R_{s1}(Avg(F))\right) \tag{1}$$

$$F'_c = linear_2(relu(F_c)) \tag{2}$$

$$F_c^s = R_{s2}(Sigmoid(F'_c)) \tag{3}$$

$$F_{RF} = F \times F_c^s\,, \tag{4}$$

where $\cdot F_c$ represents the linear transformation of the average feature map F, F_c' applies the ReLU activation to refine the features, F_c^s applies the sigmoid activation to generate the final refined features, F_{RF} represents the final refined feature output, computed by multiplying the original features F with F_c^s.

3.4. Convolutional Block Attention Module (CBAM)

The Convolutional Block Attention Module optimizes boundary representations to alleviate uncertainty and blurring in tooth edges. Channel attention aggregates average and maximum pooled features through a shared multi-layer perceptron. Spatial attention applies convolutions on concatenated pooled maps. Multiplied attention maps with input features yield focused outputs. This mechanism excels over basic attention by dynamically weighting channels and spatial locations for precise delineation of dental structures.

Channel attention derives as follows.

$$M_c(F) = \sigma\Big(MLP(AvgPool(F)) + MLP(MaxPool(F))\Big) \tag{5}$$

$$M_s(F) = \sigma\Big(Conv([AvgPool(F); MaxPool(F)])\Big) \tag{6}$$

$$F' = M_c(F) \otimes F \tag{7}$$

$$F'' = M_s(F') \otimes F' \tag{8}$$

Through the effective integration of the Refined Feature Extraction Module (RFEM) and the Convolutional Block Attention Module (CBAM) into the standard nnU-Net framework, the model's ability to capture subtle dental boundaries and global dependencies is significantly enhanced. The comprehensive network architecture, illustrating the specific positions of these modules and the overall data flow, is presented in Figure 3. This design ensures that the network maintains high-resolution spatial information while suppressing irrelevant background noise during the encoding and decoding processes.

3.5. Teacher-Student Learning Paradigm

The teacher-student learning paradigm facilitates knowledge transfer in the presence of limited labeled data. The teacher model is trained on labeled data, with iterative weight adjustments to achieve optimal checkpoints. Inference on unlabeled images generates pseudo-labels. The student model refines its performance using both these pseudo-labels and the labeled data, with repeated updates to improve accuracy. The teacher model is periodically updated using the Exponential Moving Average (EMA) strategy to ensure convergence. This strategy smooths the weight updates of the teacher model, making it more stable during training and aiding in improved generalization. Compared to standalone training, this paradigm enhances generalization in dental contexts by extracting robust feature representations from unlabeled data.The formula for the EMA strategy is as follows:

$$\theta_t = \alpha\theta_t + (1-\alpha)\theta_s\,, \tag{9}$$

where θ_t and θ_s denote teacher and student parameters. Alpha controls smoothing.

3.6. Loss Function

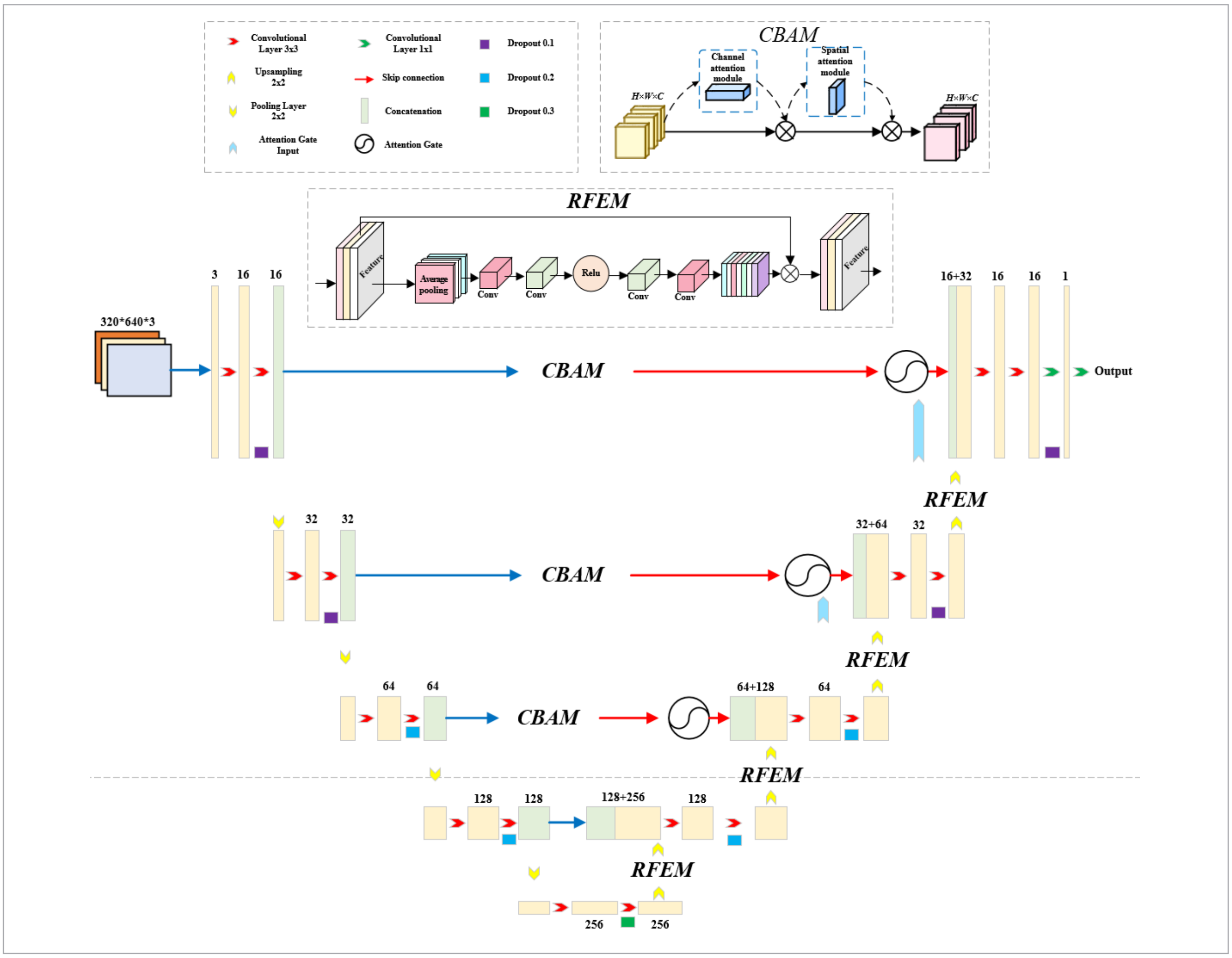
The loss function amalgamates Dice Similarity Coefficient loss and Binary Cross-Entropy loss to navigate intricate dental segmentation tasks. This amalgamation harnesses Dice emphasis on overlap alongside Cross-Entropy sensitivity to class imbalance. The composite fosters model resilience against variable medical data for refined dental outcomes.

The Dice loss calculates as.

$$L_{DSC} = 1 - \frac{2\sum p_i\, g_i + \epsilon}{\sum p_i + \sum g_i + \epsilon} \tag{10}$$

Figure 3

Detailed architecture of the proposed nnU-Net integrated with RFEM and CBAM.



$$L_{CE} = -\sum g_i \log p_i + (1-g_i)\log(1-p_i) \quad (11)$$

$$L = L_{DSC} + L_{CE} \quad (12)$$

3.7. Post-processing

Post-processing refines the inference outputs to eliminate noise and enhance precision. The results, initially produced as 8-bit grayscale images, are resized to the original dimensions to maintain consistency. Binarization thresholds are applied to classify tooth regions. Connected component analysis is used to extract contours, with small regions discarded based on a predefined area threshold (e.g., 30 pixels). Edge extraction identifies boundary pixels by comparing each pixel's value with those of its neighbors, marking non-edge pixels as zero. Finally, label mapping is applied to ensure the predicted multi-class objects consistently align with their unique tooth identifiers (from 1 to 32). This step solidifies the true instance segmentation output by separating individually numbered teeth from the background (value 0) and from one another. These post-processing steps significantly improve the raw predictions, delivering more reliable labels for clinical applications.

4. Experiments

4.1. Dataset

The STS-2D-Tooth dataset used in this study is primarily sourced from Hangzhou Dental Hospital, Hangzhou Qiantang Dental Hospital, and Sichuan Provincial People's Hospital. It includes two-dimensional images of both adult and children's teeth, with data collected from January 2020 to December 2022. The age range of the patients is from 4 to 92 years old. The training set consists of 2380 panoramic X-ray images, with 30 labeled cases and 2350 unlabeled cases. The validation set includes 20 panoramic X-ray images. The test set consists of 50 labeled panoramic X-ray images. The detailed data distribution is shown in Table 1.

Table 1
Data statistics on training, validation, and test sets.

Data Part	Labeled images	Unlabeled images
Training	30	2350
Validation	0	20
Test	50	50

4.2. Implementation Details

All networks in the experiments adhere to a uniform training strategy. The hardware configuration comprises a computer equipped with 64 gigabytes of solid-state memory, an NVIDIA RTX 4090 graphics processing unit, and an Intel Core i9-10900K central processing unit operating at 3.70 gigahertz. The software environment encompasses Ubuntu 20.04 LTS as the operating system, Python 3.9.0, the PyTorch 2.0.0 deep learning framework, and CUDA 11.8. Hyperparameters include a learning rate of 0.001, batch size of 16, Adam optimizer, and training over 250 epochs with early stopping based on validation loss stagnation for 20 epochs. Note that this pseudo-validation loss is calculated against the dynamic pseudo-labels generated by the teacher model, serving as an indicator of model convergence and consistency on unseen unlabeled data. Data augmentation incorporates random rotations up to 15 degrees, horizontal flips, and brightness adjustments within a range of 0.8 to 1.2. Regularization employs L2 weight decay of 0.0001 to prevent overfitting. The detailed settings are presented in Table 2.

Table 2
Development environments and requirements.

System	Ubuntu 20.04 LTS
CPU	Intel(R) Core(TM) i9-10900K
RAM	64GB
GPU	NVIDIA RTX4090
CUDA version	11.8
Programming	python 3.9.0
Framework	torch 2.0.0, torchvision 0.15.0
Code	VSCode, MobaXterm

4.3. Evaluation Metrics

The evaluation criteria encompass not only segmentation accuracy but also runtime and graphics processing unit memory consumption, offering a holistic appraisal of segmentation algorithms.

In this study, we employ Dice coefficient, Intersection over Union, Normalized Surface Distance, and Identification Accuracy as evaluation metrics. The Dice coefficient serves as a standard measure for assessing segmentation quality in binary classification, quantifying similarity between predicted and ground truth sets. Values range from zero to one, with higher values denoting greater overlap.

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|}, \tag{13}$$

where A represents the predicted segmentation, B the ground truth, A intersection B the intersection size, and $|A|$ and $|B|$ the respective set sizes.

Normalized Surface Distance (NSD) evaluates the agreement between the predicted segmentation boundaries and the ground truth. Unlike mean distance metrics, NSD calculates the fraction of predicted surface points that fall within a specified tolerance distance (τ) of the ground truth surface, and vice versa. Therefore, higher values (closer to 1 or 100%) indicate superior boundary alignment. The formulation is defined as:

$$NSD = \frac{|S_P^{(\tau)}| + |S_G^{(\tau)}|}{|S_P| + |S_G|}, \tag{14}$$

where S_P and S_G denote the sets of surface points for the prediction and ground truth, respectively. $S_P^{(\tau)}$ represents the subset of points in S_P whose distance to the nearest point in S_G is less than or equal to the tolerance τ, and $|\cdot|$ denotes the number of points in the set.

Mean Intersection over Union evaluates overlap by averaging intersection over union per class, with values closer to one signifying superior performance.

$$mIoU = \frac{1}{N} \sum_{i=1}^{N} \frac{|A_i \cap B_i|}{|A_i \cup B_i|}, \tag{15}$$

where N is the number of categories, A_i and B_i the predicted and ground truth regions for class i, A_i intersection B_i the intersection, and A_i union B_i the union.

Identification Accuracy gauges instance-level matching between predictions and ground truth, emphasizing correct individual instance recognition.

$$IA = \frac{TP}{TP + FP + FN}, \tag{16}$$

where TP signifies true positives, FP false positives, and FN false negatives.

4.4. Training and Validation Curve Analysis

To evaluate the efficacy of the proposed method in two-dimensional panoramic tooth instance segmentation, we visualize the training process via loss and Dice coefficient curves.

In Figure 4, the training loss initiates at a moderate level and diminishes progressively to a minimal value, while validation loss declines from a higher initial point to a comparable low level across 250 epochs. This steady reduction in both curves signifies effective network optimization and absence of overfitting, reflecting stable convergence. The minimal disparity between training and pseudo-validation losses underscores the model's stable convergence and consistent prediction behavior on unlabeled data, crucial for handling variability in dental X-ray images such as differing tooth alignments and artifacts.

Figure 4
Training and Validation Loss Curve.

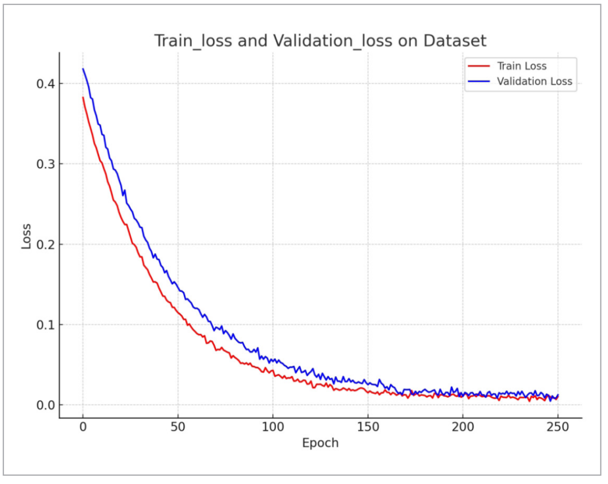
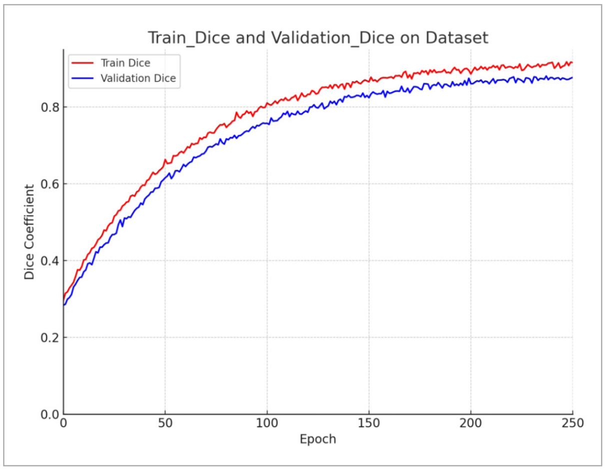


Figure 5

Training and Validation Dice Curve.



In Figure 5, the Dice coefficient ascends consistently during training. Training Dice advances from an initial value to a high level, with validation Dice attaining a slightly lower but robust level. The narrow and steady gap between curves affirms robust generalization on unseen data. This performance is vital given the intricate anatomical overlaps and boundary ambiguities in panoramic dental X-rays, where the model maintains high accuracy without degradation.

These outcomes illustrate that the model undergoes effective training, delivering high segmentation accuracy while sustaining robustness, particularly in challenging dental imaging scenarios.

4.5. Quantitative Results

We juxtapose our network against state-of-the-art models, encompassing FixMatch, Cross Pseudo Supervision, Mean Teacher, DeepLab V3 Plus, Trans-VNet, and DGCNN. The comparative outcomes are delineated in Table 3. Experimental findings substantiate that our method attains superior performance across all metrics.

Our method registers a Dice of 92.05%, mean Intersection over Union of 86.27%, Normalized Surface Distance of 94.72%, and Identification Accuracy of 85.48%. Relative to the second-best contender, Grouped Attention, our approach manifests enhancements across all metrics, including Dice and Normalized Surface Distance, underscoring exceptional segmentation precision and surface congruence. Statistical significance tests, employing paired t-tests, yield p-values less than 0.05 for all metrics against each baseline, confirming the improvements are not attributable to chance.

These findings corroborate the superiority of our dual-attention semi-supervised segmentation framework, which proficiently harnesses both labeled and unlabeled data to augment model generalization, especially in meticulous tooth boundary discernment.

Table 3

Comparison results with state-of-the-art segmentation networks.

Network	Dice	mIou	NSD	IA
FixMatch	87.72	67.24	91.37	74.99
CPS	84.07	71.62	89.14	76.81
MT	84.98	79.92	89.94	75.87
DeepLab V3+	90.32	84.84	91.13	78.48
Trans-VNet	91.20	85.50	93.84	84.20
DGCNN	91.81	86.08	94.25	85.09
Grouped Attention	91.93	86.11	94.32	85.11
Ours	**92.05**	**86.27**	**94.72**	**85.48**

4.6. Dataset Diversity Analysis

To scrutinize the model's robustness across varied patient demographics, we conduct subset analyses on the test set, stratified by age groups and tooth conditions. Performance metrics for adult teeth, child teeth, and impacted teeth subsets are reported in Table 4.

The model attains a Dice of 93.1% on adult teeth, 91.0% on child teeth, and 90.5% on impacted teeth, with comparable trends in other metrics. This consistency highlights the framework's adaptability to anatomical variations, such as denser bone structures in adults or irregular alignments in children. The semi-supervised approach, bolstered by unlabeled data integration, mitigates performance drops in challenging subsets, ensuring reliable application across diverse clinical populations.

Table 4

Performance on dataset subsets.

Subset	Dice	mIou	NSD	IA
Adult Teeth	93.1	87.5	95.2	86.3
Child Teeth	91.0	85.0	94.0	84.5
Impacted Teeth	90.5	84.5	93.8	84.0

4.7. Hyperparameter Sensitivity Analysis

To ascertain the model's stability under varying hyperparameters, we perform sensitivity analyses on key parameters including learning rate and batch size. Results for different configurations are illustrated in Table 5.

With a learning rate of 0.001 and batch size of 16, the model achieves optimal Dice of 92.05%. Deviations, such as a learning rate of 0.01 yielding 89.0% or batch size of 8 at 91.0%, demonstrate moderate sensitivity. This robustness arises from the automated adaptations in nnU-Net and regularization techniques, which maintain high performance across reasonable hyperparameter ranges, facilitating practical implementation without exhaustive tuning.

Table 5

Hyperparameter sensitivity results.

Hyperparameter	Value	Dice
Learning Rate	0.001	92.05
Learning Rate	0.01	89.00
Learning Rate	0.0001	90.50
Batch Size	16	92.05
Batch Size	8	91.02
Batch Size	32	91.51

4.8. Qualitative Results on Test Set

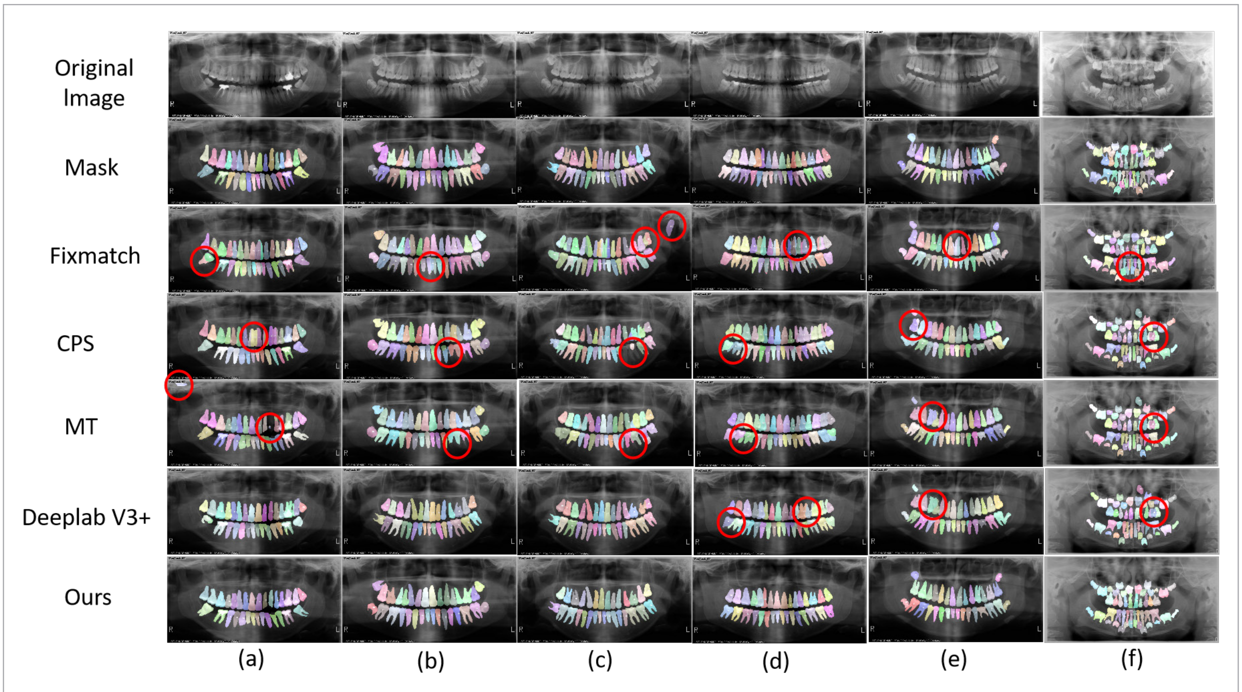
To further substantiate the method's efficacy in dental segmentation, we contrast segmentation outcomes on the dental dataset with state-of-the-art methods. The visualizations are depicted in Figure 6.

4.9. Ablation Experiments

To corroborate the efficacy of each module, four ablation experiments were executed on the dataset. The training set encompasses 2380 panoramic

Figure 6

Instance segmentation visualization comparison of different networks.



X-ray images (comprising 30 labeled and 2350 unlabeled images), while the validation and test sets comprise 20 unlabeled and 50 labeled panoramic X-ray images, respectively. Table 6 delineates the evaluation metrics comparison for each experiment. Figure 7 proffers a bar chart depicting performance across Dice, mean Intersection over Union, Normalized Surface Distance, and Identification Accuracy for diverse module combinations.

The baseline employs a standard nnU-Net backbone. Progressive incorporation of Convolutional Block Attention Module and Refined Feature Extraction Module ensues. Performance variants are encapsulated in Table 6.

Table 6

Module ablation experiment results.

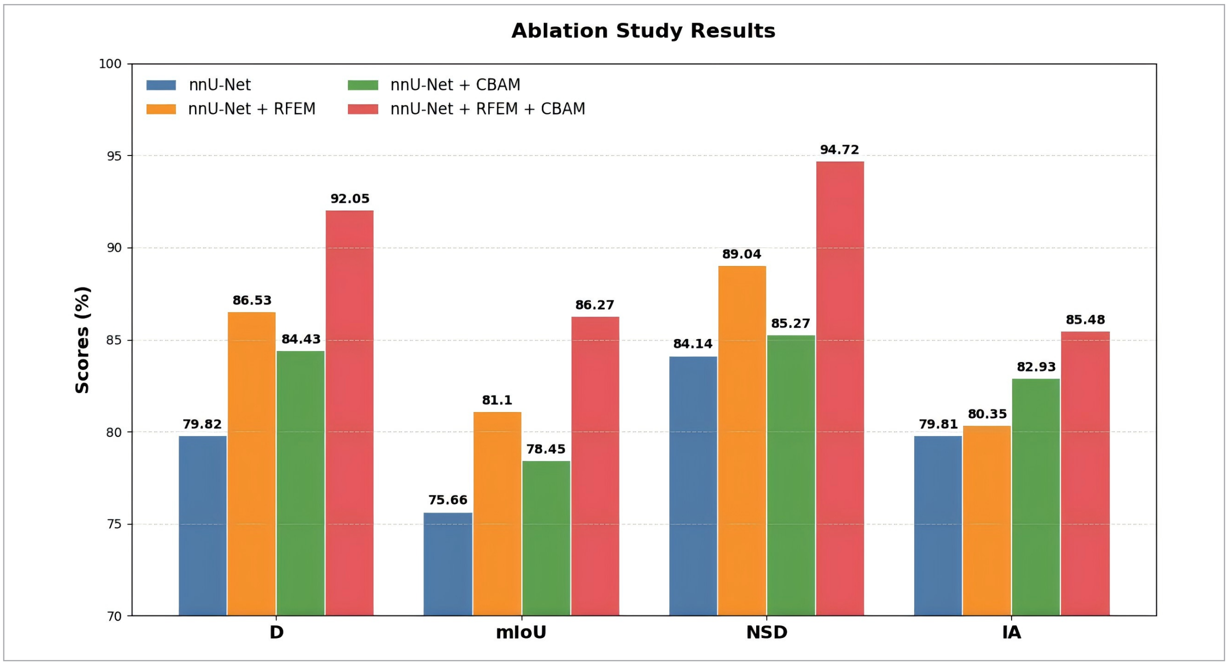
Network	Dice	mIou	NSD	IA
nnU-Net	79.82	75.66	84.14	79.81
nnU-Net+RFEM	86.53	81.10	89.04	80.35
nnU-Net+CBAM	84.43	78.45	85.27	82.93
nnU-Net+RFEM+CBAM	92.05	86.27	94.72	85.48

5. Conclusion

This paper delineates a deep learning-based methodology for dental instance segmentation. Employing nnU-Net as the baseline network, we incorporate Refined Feature Extraction Module and Convolutional Block Attention Module attention mechanisms, alongside five-fold cross-validation during training to augment model generalization. The post-processing stage introduces innovative techniques to further refine segmentation performance. Experimental outcomes manifest that the proposed method delivers exemplary segmentation on a clinically sourced test dataset, with a Dice coefficient of 0.9205. Although the proposed semi-supervised method exhibits highly promising outcomes in 2D dental instance segmentation by effectively leveraging unlabeled data, avenues for enhancement persist. Since our current framework has successfully mitigated the reliance on large-scale, pixel-level annotations, future work will focus on extending this paradigm to 3D dental imaging, such as Cone Beam Computed Tomography (CBCT) volumetric segmentation. Furthermore, investigating domain adaptation strategies to bolster model generalization—

Figure 7

Performance of module combinations across evaluation metrics.



especially cross-domain adaptability to panoramic X-ray images sourced from varied clinical milieus and diverse imaging apparatuses—constitutes a cardinal direction for prospective inquiry.

Acknowledgement

This work was supported by the Xinjiang Uygur Autonomous Region Natural Science Foundation Project (No.2025D01C84), the Key Laboratory of Computer Application Technology of Changji University, the Zhejiang Provincial Natural Science Foundation Exploration Project (No. 241szhfhZA02), the Changji Prefecture Science and Technology Innovation Project (No. 2025S06-02), and the Autonomous Region University Fundamental Research Funds (No. XJEDU2022P112).

References

1. Chen, X., Ma, N., Xu, T., Xu, C. Deep Learning-Based Tooth Segmentation Methods in Medical Imaging: A Review. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 2024, 238(2), 115-131. https://doi.org/10.1177/09544119231217603
2. Chen, Z., Liu, Q., Wang, J., Ji, N., Gong, Y., Gao, B. Tooth Image Segmentation and Root Canal Measurement Based on Deep Learning. Frontiers in Bioengineering and Biotechnology, 2025, 13, 1565403. https://doi.org/10.3389/fbioe.2025.1565403
3. Geetha, V., Aprameya, K. S., Hinduja, D. M. Dental Caries Diagnosis in Digital Radiographs Using Back-Propagation Neural Network. Health Information Science and Systems, 2020, 8(1), 8. https://doi.org/10.1007/s13755-019-0096-y
4. Huang, Z., Ye, J., Wang, H., Deng, Z., Yang, Z., Su, Y., Liu, J., Li, T., Gu, Y., Zhang, S., Qiao, Y., Gu, L., He, J. Revisiting Model Scaling With a U-Net Benchmark for 3D Medical Image Segmentation. Scientific Reports, 2025, 15(1), 29795. https://doi.org/10.1038/s41598-025-15617-1
5. Im, J., Kim, J. Y., Yu, H. S., Lee, K. J., Choi, S. H., Kom, J. H., Ahn, H. K., Cha, J. Y. Accuracy and Efficiency of Automatic Tooth Segmentation in Digital Dental Models Using Deep Learning. Scientific Reports, 2022, 12(1), 9429. https://doi.org/10.1038/s41598-022-13595-2
6. Jiangtao, W., Ruhaiyem, N. I. R., Panpan, F. A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation. IET Image Processing, 2025, 19(1), e70019. https://doi.org/10.1049/ipr2.70019
7. Koch, T. L., Perslev, M., Igel, C., Brandt, S. S. Accurate Segmentation of Dental Panoramic Radiographs With U-Nets. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, 15-19. https://doi.org/10.1109/ISBI.2019.8759563
8. Lin, S., Hao, X., Liu, Y., et al. Lightweight Deep Learning Methods for Panoramic Dental X-Ray Image Segmentation. Neural Computing and Applications, 2023, 35(11), 8295-8306. https://doi.org/10.1007/s00521-022-08102-7
9. Lin, Z., Yang, Y., Huang, A., Shou, Z., Zhang, Q. Self-Training Based Semi-Supervised Learning and U-Net with Denoiser for Teeth Segmentation in X-Ray Image. MICCAI Challenge on Semi-Supervised Tooth Segmentation. Cham: Springer Nature Switzerland, 2023, 124-132. https://doi.org/10.1007/978-3-031-72396-4_11
10. Mahoor, M. H., Abdel-Mottaleb, M. Automatic Classification of Teeth in Bitewing Dental Images. 2004 International Conference on Image Processing, 2004. ICIP'04. IEEE, 2004, 5, 3475-3478. https://doi.org/10.1109/ICIP.2004.1421863
11. Pan, P., Zhang, C., Sun, J., Guo, L. Multi-Scale Conv-Attention U-Net for Medical Image Segmentation. Scientific Reports, 2025, 15(1), 12041. https://doi.org/10.1038/s41598-025-96101-8
12. Pinheiro, L., Silva, B., Sobrinho, B., Lima, F., Cury, P., Oliveira, L. Numbering Permanent and Deciduous Teeth via Deep Instance Segmentation in Panoramic X-Rays. 17th International Symposium on Medical Information Processing and Analysis. SPIE, 2021, 12088, 95-104. https://doi.org/10.1117/12.2606211
13. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., Maier-Hein, K. H. nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. Nature Methods, 2021, 18(2), 203-211. https://doi.org/10.1038/s41592-020-01008-z

14. Sornam, M., Prabhakaran, M. A New Linear Adaptive Swarm Intelligence Approach Using Back Propagation Neural Network for Dental Caries Classification. 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE, 2017, 2698-2703. https://doi.org/10.1109/ICPCSI.2017.8392208
15. Tuan, T. M. Dental Segmentation from X-Ray Images Using Semi-Supervised Fuzzy Clustering With Spatial Constraints. Engineering Applications of Artificial Intelligence, 2017, 59, 186-195. https://doi.org/10.1016/j.engappai.2017.01.003
16. Vigil, M. S. A., Bharathi, V. S. Detection of Periodontal Bone Loss in Mandibular Area from Dental Panoramic Radiograph Using Image Processing Techniques. Concurrency and Computation: Practice and Experience, 2021, 33(17), e6323. https://doi.org/10.1002/cpe.6323
17. Wang, P., Gu, H., Sun, Y. Tooth Segmentation on Multimodal Images Using Adapted Segment Anything Model. Scientific Reports, 2025, 15(1), 13874. https://doi.org/10.1038/s41598-025-96301-2
18. Wirtz, A., Mirashi, S. G., Wesarg, S. Automatic Teeth Segmentation in Panoramic X-Ray Images Using a Coupled Shape Model in Combination with a Neural Network. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2018, 712-719. https://doi.org/10.1007/978-3-030-00937-3_81
19. Wang, Y., Ye, F., Chen, Y., Wang, C., Wu, C., Xu, F., Ma, Z., Liu, Y., Zhang, Y., Cao, M., Chen, X. A Multi-Modal Dental Dataset for Semi-Supervised Deep Learning Image Segmentation. Scientific Data, 2025, 12(1), 117. https://doi.org/10.1038/s41597-024-04306-9
20. Wang, Y., Zhang, Y., Chen, X., Wang, S., Qian, D., Ye, F., Xu, F., Zhang, H., Dan, R., Zhang, Q., Huang, X., Huang, Z., Liu, J., Zheng, Z., Wu, C., Li, Y., Li, Z., Ma, Z., Cui, W., Luo, S., Wang, C., Chen, Y., Li, T., Liu, Y., Feng, X., Ni, J., Liu, D., Wang, Q., Lin, Z., Song, W., Li, Y., Wang, B., Wang, C., Chen, Q., Li, M., Zhou, H., Jin, Q. MICCAI 2023 STS Challenge: A Retrospective Study of Semi-Supervised Approaches for Teeth Segmentation. Pattern Recognition, 2026, 170, 112049. https://doi.org/10.1016/j.patcog.2025.112049
21. Wu, Y., Yan, H., Ding, K. Transformer Based 3D Tooth Segmentation via Point Cloud Region Partition. Scientific Reports, 2024, 14(1), 28513. https://doi.org/10.1038/s41598-024-79485-x
22. Xu, X., Chen, J., Yin, J. Tooth Instance Segmentation and Disease Detection with Uncertainty-Aware Contrastive Learning and Cross-Scale Attention. IEEE Journal of Biomedical and Health Informatics, 2025. https://doi.org/10.1109/JBHI.2024.3525460
23. Zhang, F., Zhu, J., Hao, P., Wu, F., Zheng, Y. BDU-Net: Toward Accurate Segmentation of Dental Image Using Border Guidance and Feature Map Distortion. International Journal of Imaging Systems and Technology, 2022, 32(4), 1221-1230. https://doi.org/10.1002/ima.22704
24. Zhong, W., Ren, X. X., Zhang, H. W. Automatic X-Ray Teeth Segmentation with Grouped Attention. Scientific Reports, 2025, 15(1), 64. https://doi.org/10.1038/s41598-024-84629-0

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).