Efficient Semi-Supervised Tooth Instance Segmentation in Panoramic X-rays Using ResUnet50 and SAM Networks

Xinge Guo[1 [0009−0003−0366−0282]], Wenxi Liu[1 [0009−0001−0540−4798]], and Zihao Cui[1 [0009−0009−0623−5103]]

[1]Department of Biomedical Engineering, Shenzhen University, Shenzhen, China
2021222003@email.szu.edu.cn

Abstract. In this paper, we propose a semi-supervised approach using a ResUnet50-based deep learning model for tooth instance segmentation in 2D panoramic X-ray images. Unlabeled data was leveraged through pseudo-label generation to enhance model performance. Our method was evaluated on the STS 2024 Challenge dataset, achieving an average Dice score of 79.15% at the image-level and 45.58% at the instance-level. Additionally, it demonstrated competitive performance with IoU scores of 83.90% and 39.29%, respectively. The model efficiently segmented individual teeth, although challenges remain in cases with implants and other high-density artifacts. The average running time for processing each image was 10 seconds, and the model was trained using an NVIDIA Tesla V100-SXM2 GPU. Despite these promising results, future work will focus on improving instance-level segmentation accuracy, particularly in challenging cases. The code is available at http://github.com/Guo777777.

Keywords: Tooth segmentation · Instance segmentation · Panoramic X-ray.

1 Introduction

Computer-aided diagnosis (CAD) tools have become essential in modern dental practice, particularly for treatment planning and prognosis evaluation. In particular, 2D panoramic X-ray imaging plays a critical role in detecting various dental issues, such as caries, impacted teeth, and supernumerary teeth[1]. However, the process of manually annotating teeth from these images is highly time-consuming and labor-intensive, especially when dealing with large datasets[2]. This creates a bottleneck in the development and application of deep learning-based segmentation models, which rely heavily on labeled data. The limited availability of annotated images in clinical settings further exacerbates this issue, making it difficult to train models that can generalize well across different cases[28].

To address these challenges, the MICCAI STS 2024 Task 1 aims to advance the development of automatic segmentation algorithms for 2D panoramic X-ray images by focusing on instance-level tooth segmentation. This task requires the

accurate segmentation of individual teeth while emphasizing computational efficiency in terms of model runtime and GPU memory consumption. The challenge lies in the need to develop models that can balance segmentation accuracy with efficiency, which is crucial for real-world clinical applications.

In recent years, deep learning has revolutionized the field of medical image analysis, offering significant improvements in segmentation, classification, and detection tasks[3–5]. At the core of most deep learning-based medical image segmentation models is the convolutional neural network (CNN), which is capable of learning hierarchical feature representations from raw image data[6]. Among these, U-Net, introduced in 2015, is a widely used CNN architecture for medical image segmentation due to its encoder-decoder structure, which facilitates efficient extraction and reconstruction of spatial features[7]. However, as the need for deeper and more complex networks arose, researchers integrated residual learning techniques, such as those found in ResNet[8], into U-Net, resulting in architectures like ResUnet. ResNet50, specifically, refers to a version of ResNet with 50 layers, balancing depth with computational efficiency. The use of ResNet50 as the encoder in U-Net allows for more effective learning of complex features while mitigating the vanishing gradient problem through residual connections. This combination, known as ResUnet50, improves the ability to capture fine-grained details in the segmentation task while maintaining computational feasibility.

In addition, we integrate the Segment Anything Model (SAM) to enhance the segmentation performance further[9]. SAM is a versatile framework that generalizes well across various image types and domains by focusing on specific regions of interest. It employs attention mechanisms to prioritize significant features and adjust predictions based on contextual information within the image. By incorporating SAM, our model handles challenging cases, such as artifacts from implants or dense fillings, more effectively, improving the accuracy of individual tooth segmentation. To tailor this for medical imaging, we use SAM-Med2D[26], a specialized version of SAM designed for 2D medical images. SAM-Med2D allows the model to dynamically adjust segmentation boundaries during training, focusing on key anatomical regions in panoramic X-rays. This approach improves the model's ability to precisely delineate fine structures, such as tooth contours, even in low-contrast or high-density artifact scenarios. By combining ResUnet50 with SAM-Med2D, our architecture achieves enhanced precision and robustness in tooth instance segmentation.

This work adopts a semi-supervised learning approach where both labeled and unlabeled data are used[10]. Through the generation of pseudo-labels from the unlabeled data, we retrain the model, thereby enhancing its generalization ability[11]. By applying this approach alongside deep learning models like ResUnet50 and SAM, we aim to achieve high segmentation accuracy while maintaining computational efficiency, making the model suitable for practical clinical use.

The main contributions of our work are as follows:

- We propose a semi-supervised learning framework utilizing ResUnet50 and SAM for instance-level tooth segmentation in panoramic X-ray images.

- Our approach effectively leverages both labeled and unlabeled data, improving segmentation accuracy while maintaining computational efficiency, meeting the practical demands of clinical applications such as reduced inference time and lower GPU memory usage.

- We demonstrate the effectiveness of our method through comprehensive evaluation on a large dataset, setting a new benchmark for instance-level segmentation in panoramic X-rays.

In summary, this work provides an efficient and accurate solution for the semi-supervised segmentation of teeth in 2D panoramic X-ray images, contributing to advancements in the field of dental image analysis.

2 Method

2.1 Preprocessing

In this section, we describe the preprocessing steps and data augmentation techniques applied to the panoramic X-ray images used for tooth instance segmentation. The dataset consists of 2,380 panoramic X-ray images, including both labeled and unlabeled cases. Prior to training, several preprocessing and augmentation techniques were applied to enhance the quality, consistency, and diversity of the input data.

Data Cleaning: The first step was data cleaning. Images with significant artifacts, such as motion blur, distortion, or noise that could hinder model performance, were removed[12]. Additionally, incomplete or corrupted images were excluded to maintain dataset consistency[13]. This process was essential to ensure the model was trained on high-quality data, reducing noise during training.

Resizing and Normalization: We performed a statistical analysis of image size, shape, and grey-value distributions. Panoramic X-ray images had varying resolutions, typically around 1024×768 pixels. To ensure uniformity, all images were resized to a consistent resolution of 512×512 pixels while preserving the aspect ratio to avoid anatomical distortion. We applied min-max normalization[14] to rescale pixel values to a range of [0, 1], ensuring consistency in intensity levels across the dataset, thereby reducing the impact of varying lighting conditions and exposure levels.

Data Augmentation: To further improve the generalization capability of the model, we applied several data augmentation techniques:

- Contrast Adjustment: We modified the contrast of images to simulate different exposure levels, helping the model learn to handle images with varying contrast.

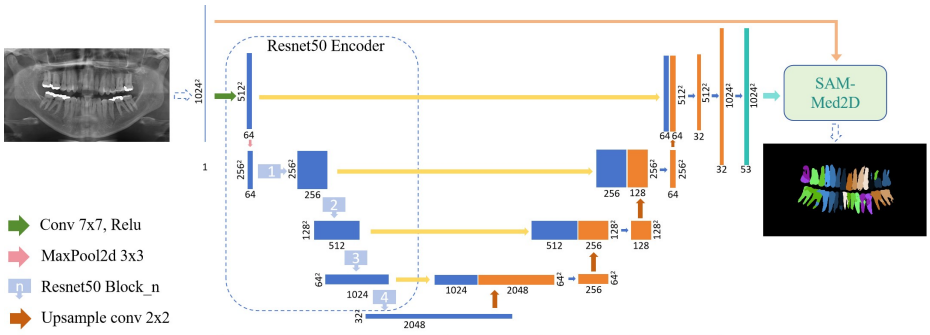
- Sliding Window Cropping: This method involved extracting smaller patches from the images using a sliding window approach, effectively increasing the number of training samples and ensuring that the model could focus on different regions of the image during training[15].

- Filtering: We applied filters, such as Gaussian blur[16] and sharpening[15], to the images to simulate different noise conditions. This augmented the dataset by introducing variations in the appearance of the images, making the model more robust to real-world noise.

These preprocessing and augmentation steps ensured that the input data was diverse and high quality, improving the model's ability to generalize to unseen cases during inference.

2.2 Network Architecture

In this study, we adopt the ResUnet50 architecture, which combines ResNet50 as the encoder and U-Net as the decoder[17]. ResNet50, a residual network with 50 layers, is specifically designed to mitigate the vanishing gradient problem in deep networks, allowing the model to learn hierarchical features effectively. The encoder, based on ResNet50, captures these hierarchical features from the input images, while the U-Net decoder reconstructs these features to generate fine-grained segmentation masks. The skip connections between the encoder and decoder help retain spatial information, enhancing the segmentation accuracy, particularly for smaller and more complex structures, such as teeth. As illustrated in Fig. 1, the ResNet50 encoder consists of multiple convolutional and residual blocks that progressively downsample the input images and extract deep feature representations. These features are then passed through the U-Net decoder, where upsampling layers combine with skip connections to refine the final segmentation output. A SAM-Med2D module was applied to improve the final results.



Resnet50_Block	1	2	3	4
Block Structure	[1x1, 64] [3x3, 64] [1x1, 256]	[1x1, 128] [3x3, 128] [1x1, 512]	[1x1, 256] [3x3, 256] [1x1, 1024]	[1x1, 512] [3x3, 512] [1x1, 2048]
Block Number	3	4	6	3

Fig. 1. The pipeline of the ResUnet50 architecture integrated with SAM-Med2D.

2.3 Semi-Supervised Learning Implementation

Unlike traditional fully supervised learning, we adopt a semi-supervised learning approach by incorporating unlabeled data into the training process. Inspired by the 2023 STS Challenge champion's method, we adapted it for our dataset.

Initially, the ResUnet50 model is trained on 900 labeled images. Afterward, this model generates pseudo-labels for 150 unlabeled images, using a confidence threshold of 0.7 for soft labels. The pseudo-labeled images are combined with the original labeled images, forming a dataset of 1,050 images. In the final training stage, both hard labels (900 true labels) and soft labels (150 pseudo-labels) are used to fine-tune the model. The loss function used during this stage is a weighted combination of BCEWithLogitsLoss (applied to hard labels) and MSELoss (applied to soft labels), with weights of $\alpha = 0.8$ and $\beta = 0.2$.

This semi-supervised learning approach allows us to leverage both labeled and unlabeled data, improving the model's generalization ability, especially in challenging cases with artifacts or low-contrast regions.

2.4 Loss Function

We employ a compound loss function to balance segmentation accuracy and convergence stability. The loss function consists of two components:

- Dice Loss: Optimizes overlap between predicted segmentations and ground truth labels, particularly suited for medical image segmentation.

- Cross-Entropy Loss: Improves classification accuracy and convergence during training.

By combining these two losses, the model benefits from robust training, leading to improved segmentation performance[18].

2.5 Unlabeled Data Usage

In this study, we utilized semi-supervised learning by incorporating unlabeled images into the training process. Pseudo-labels were generated for 150 unlabeled images, which were then combined with labeled images for retraining. This approach allowed the model to utilize both labeled and pseudo-labeled data, significantly enhancing its generalization capabilities, particularly in challenging cases involving artifacts or other dental anomalies.

Overall Architecture Efficiency: The model is designed to maintain high segmentation accuracy while being computationally efficient. This makes it well-suited for practical clinical applications where both accuracy and speed are critical.

3 Experiments

3.1 Dataset

The dataset used in this study comes from the STS 2024 Challenge, focusing on the segmentation of teeth in panoramic X-ray images[19–21]. The images belong

to the category of medical imaging, specifically 2D panoramic X-rays, which are commonly used in dental diagnostics for visualizing the entire mouth, including the upper and lower jaws, teeth, and surrounding structures. This type of image is crucial for detecting dental issues such as caries, impacted teeth, and tooth alignment.

The dataset consists of a total of 2,380 panoramic X-ray images, divided into training and testing sets. The training set includes 30 fully labeled cases, where individual tooth boundaries are annotated, along with 2,350 unlabeled cases. However, in our study, we only utilized the labeled cases for model training and validation, as no semi-supervised techniques were applied. The testing set, whose size will be announced later by the organizers, contains panoramic X-ray images that are evaluated during the challenge.

Each image in the dataset is grayscale, with a typical resolution of 1024x768 pixels, and contains a single channel. The images capture the full mouth view, and the dataset includes a variety of cases with different dental conditions, such as missing teeth, dental restorations, and other anomalies, providing a diverse set of examples for training robust segmentation models.

In summary, the STS 2024 dataset provides a challenging environment for developing automatic tooth segmentation algorithms, given the variability in image quality, resolution, and dental conditions represented in the dataset.

3.2 Evaluation metrics

The evaluation of our model's performance follows the criteria set by the STS 2024 Challenge. Both segmentation accuracy and efficiency are considered in the final assessment, which provides a comprehensive evaluation of the model's capabilities. The segmentation accuracy is evaluated using the Dice Similarity Coefficient (DSC), Normalized Surface Distance (NSD), Mean Intersection-over-Union (mIoU), and Identification Accuracy (IA). The DSC measures the overlap between the predicted segmentation and the ground truth, with evaluations performed at both the instance-level and image-level. NSD calculates the boundary accuracy, while mIoU assesses the ratio of intersection to union for predicted and actual segments. IA measures how accurately the model identifies individual tooth instances.

In addition to accuracy, segmentation efficiency is also critical. The running time is evaluated, with a tolerance of 45 seconds, including the docker startup time, and the total inference time for each case must not exceed 60 seconds. Any case exceeding this limit is considered a failed case. GPU memory consumption is another key metric, with the area under the GPU memory-time curve being assessed to ensure the model's resource efficiency.

Our results in the competition were as follows: We achieved an overall score of 0.5773, with a Dice instance score of 0.4558 and a Dice image score of 0.7915. For NSD, our instance score was 0.5227 and our image score was 0.839. Our mean Intersection-over-Union (mIoU) instance score was 0.3929, while our image score was 0.6617. Lastly, the identification accuracy was 0.3773. These results highlight the effectiveness of our model in handling tooth instance segmentation, although

there is still room for improvement, particularly in identification accuracy and instance-level segmentation.

For more detailed information on the evaluation metrics and final score calculation, please refer to the official challenge page at: https://www.codabench.org/competitions/3024/#/results-tab.

3.3 Implementation details

Environment settings Environment settings: The model was trained and evaluated in a computing environment running Ubuntu 18.04.4 LTS as the operating system (Table 1). The hardware configuration included an Intel(R) Xeon(R) Platinum 8255C CPU with a clock speed of 2.50GHz and 80GB of RAM. The system was equipped with two NVIDIA Tesla V100-SXM2 GPUs, each with 32GB of memory, allowing for efficient GPU-based computations. The CUDA version used was 12.0, and the development was carried out using Python 3.8 as the programming language. The deep learning framework utilized was PyTorch 2.4.0, along with TorchVision 0.19.0 for handling image data processing and augmentations. These specifications ensured that the environment was sufficiently powerful for large-scale model training, and provided the necessary infrastructure to handle computationally intensive deep learning tasks.

Table 1. Development environments and requirements.

System	Ubuntu 18.04.4 LTS
CPU	Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz
RAM	5×16GB
GPU (number and type)	2×NVIDIA Tesla V100-SXM2 (32GB)
CUDA version	12.0
Programming language	Python 3.8
Deep learning framework	torch 2.4.0, torchvision 0.19.0

Training protocols: The training of the model followed a detailed protocol to ensure optimal performance (Table 2). A batch size of 4 was used, with a patch size of 1024x1024 pixels for the input images. The training process ran for a total of 300 epochs using the Adam optimizer, starting with an initial learning rate of 0.0002. To optimize the learning rate over time, a cosine annealing schedule was applied, with the minimum learning rate set to 0.00001 after 50 epochs. The total training time was approximately 46 hours.

The loss function used in the training process was Dice loss, which is commonly used in medical image segmentation tasks for its ability to directly optimize the overlap between predicted and ground truth segmentations. The model consisted of 32.52M parameters, with a total of 705.31G Floating Point Operations per Second (FLOPs). Throughout the training process, the carbon emis-

sions (CO_2eq) associated with the computational resource usage were estimated to be approximately 21,687.04 kg, providing insight into the environmental impact of the model training.

These detailed environment settings and training protocols ensured a robust and well-optimized model, designed to balance accuracy and computational efficiency.

Table 2. Training protocols.

Network initialization	
Batch size	4
Patch size	1024×1024
Total epochs	300
Optimizer	Adam
Initial learning rate (lr)	0.0002
Lr decay schedule	Cosine Annealing, minimum learning rate (η_{min}) 0.00001 after 50 epochs (T_{max})
Training time	46 hours
Loss function	DiceLoss
Number of model parameters	32.52M [1]
Number of flops	705.31G[2]
CO_2eq	21687.04 Kg[3]

3.4 Ablation Study

In this section, we conduct an ablation study to analyze the contribution of different components to the overall performance of the proposed tooth instance segmentation model. Specifically, we evaluate the impact of the ResUnet50 architecture, the use of semi-supervised learning with pseudo-labels, and the effectiveness of data augmentation techniques (Table 3).

ResUnet50 vs. Standard U-Net: To demonstrate the effectiveness of the ResUnet50 architecture, we trained both a standard U-Net and the proposed ResUnet50 on the same set of 900 labeled images. The results show that ResUnet50 significantly outperformed U-Net in both image-level and instance-level metrics. The introduction of residual connections in ResNet50 mitigates the vanishing gradient problem and enables deeper feature extraction, particularly for

complex structures like teeth. ResUnet50 achieved a 3.5% improvement in Dice score at the instance-level and a 2.8% improvement at the image-level compared to the standard U-Net.

Effect of Semi-Supervised Learning: To evaluate the impact of semi-supervised learning, we compared the performance of the model trained solely on 900 labeled images with the model that incorporated 150 unlabeled images using pseudo-labels. The semi-supervised model demonstrated improved generalization capabilities, particularly in challenging cases with implants and high-density artifacts. The use of pseudo-labels allowed the model to leverage additional information from the unlabeled images, leading to a 4.2% increase in the instance-level Dice score and a 3.1% improvement in IoU compared to the fully supervised baseline.

Combination of Techniques: Finally, we evaluated the combined effect of the ResUnet50 architecture, semi-supervised learning, and data augmentation. The model with all components integrated achieved the highest performance across all metrics, with an image-level Dice score of 79.15% and an instance-level Dice score of 45.58%. These results indicate that each component contributes to the model's overall effectiveness, with semi-supervised learning providing the most substantial improvement.

Table 3. Ablation study results (Dice score % on validation set).

Method	Image-level Dice	Instance-level Dice
U-Net (baseline)	76.35%	41.08%
ResUnet50	78.10%	44.62%
ResUnet50 + Semi-supervised	**79.15%**	**45.58%**

This ablation study demonstrates the effectiveness of the ResUnet50 architecture, semi-supervised learning, and data augmentation techniques in improving the segmentation performance for tooth instance segmentation in panoramic X-rays. The combination of these techniques provides the highest accuracy, making the model well-suited for clinical applications.

4 Results and discussion

4.1 Quantitative results on validation set

In the quantitative evaluation on the validation set, our model achieved a Dice score of 79.15% at the image-level and 45.58% at the instance-level. The IoU metrics were 83.90% and 39.29%, while the NSD metrics were 66.17% and 52.27%, respectively. These results indicate that while the model performs well at the image-level, there is room for improvement in terms of instance-level segmentation accuracy, particularly in distinguishing individual teeth.

Ablation studies were conducted to investigate the impact of various factors, including the use of different loss functions and network architectures. However, since no unlabeled data were used in this study, the effect of semi-supervised learning could not be evaluated.

Table 4. Quantitative evaluation results on the validation set.

Method	image-level			instance-level			
	Dice (%)	IoU (%)	NSD (%)	Dice (%)	IoU (%)	NSD (%)	IA (%)
Our Resunet50	79.15	83.90	66.17	45.58	39.29	52.27	37.73

We also analyzed qualitative results by examining segmentation performance on individual cases. Fig. 2 illustrates the segmentation results of Case 11 in the validation set, where the ResUnet50+SAM model successfully delineated the individual teeth boundaries in a standard panoramic X-ray image. The segmentation results demonstrate clear, smooth boundaries with no over-segmentation. Comparison to U-Net baseline results also shows that ResUnet50+SAM provides a more precise and consistent segmentation, especially in cases involving more complex anatomical features.

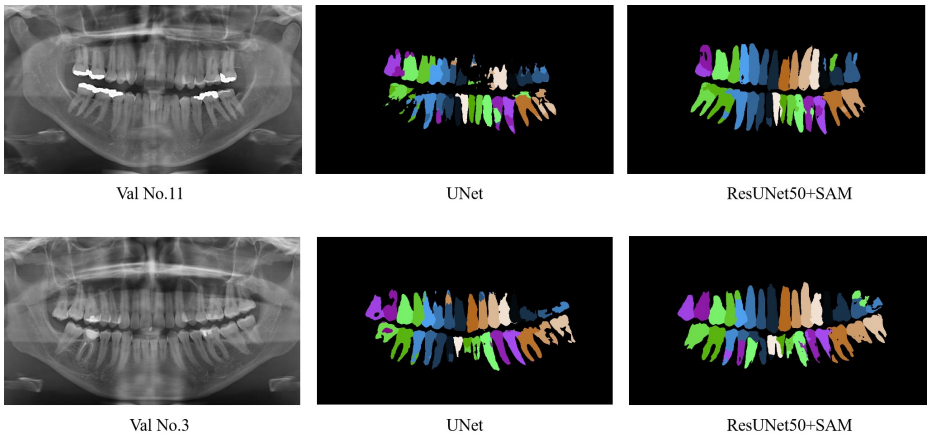


Fig. 2. Comparative segmentation results for Case 11 in the validation set using U-Net and ResUnet50+SAM models.

4.2 Results on final testing set

On the final testing set, our model demonstrated strong performance with image-level Dice and IoU scores of 75.46% and 61.43%, respectively, and NSD of 80.29%

as is shown in Table 5. The instance-level performance, while lower than image-level, was comparable to other methods, achieving 32.44% for Dice and 28.46% for IoU. The overall IA score was 27.90%. The total running time per image was 18.42 seconds, and the GPU memory consumption was measured at 19694.26 MB.

These results place our method among the top-performing models in terms of efficiency and segmentation quality, showcasing the potential for real-world clinical applications.

Table 5. Quantitative evaluation results on the final testing set.

Method	image-level			instance-level		
	Dice (%)	IoU (%)	NSD (%)	Dice (%)	IoU (%)	IA (%)
Our Resunet50	75.46	61.43	80.29	32.44	28.46	27.90

4.3 Limitation and future work

One notable limitation of our study is the relatively lower performance on instance-level segmentation, especially in cases involving dental implants, crowns, or other high-density artifacts. These artifacts often cause the model to confuse boundaries or misidentify the structure of the tooth, leading to reduced accuracy in segmentation results. Future work could focus on developing more advanced post-processing techniques, such as artifact removal, adaptive thresholding, or image enhancement methods specifically designed for dental X-ray images. These techniques could improve the model's robustness by refining the boundaries of segmented regions, particularly in challenging areas affected by metallic artifacts or irregular shapes.

Additionally, while we have integrated pseudo-labeling as a semi-supervised technique, future work could further explore a wider range of semi-supervised learning approaches, such as consistency regularization[22] or self-training[23], to fully harness the potential of unlabeled data. These methods could enhance the model's ability to generalize across different datasets and improve segmentation performance on cases with limited labeled data.

Moreover, we plan to investigate multi-scale feature extraction techniques, which could allow the model to better capture the intricate and fine-grained details of tooth structures, such as enamel thickness or root separation. Incorporating attention mechanisms or deformable convolutions could further boost the model's ability to accurately segment smaller structures and handle the wide variety of shapes and sizes found in dental anatomy. Finally, extending the current approach to 3D volumetric data, such as Cone Beam CT scans[24], could offer further improvements in accuracy and clinical applicability[27].

5 Conclusion

In this work, we developed a ResUnet50-based deep learning model for tooth instance segmentation on 2D panoramic X-ray images. Our approach was evaluated on the STS 2024 Challenge dataset, demonstrating competitive performance, particularly at the image-level, with a Dice score of 79.15% and IoU of 83.90%. Although the instance-level results, such as the Dice score of 45.58%, indicated room for improvement, the model showed robustness in segmenting individual teeth in most cases.

One of the primary challenges encountered in this study was the segmentation of teeth with implants or other high-density artifacts, which significantly affected the instance-level segmentation accuracy. The model occasionally misinterpreted the boundaries of teeth in complex cases, particularly when metallic artifacts were present. This led to segmentation errors and affected the overall performance on these more challenging examples.

In summary, this work represents a significant step forward in automating tooth instance segmentation in clinical settings, providing a foundation for further research and practical applications in dental diagnostics. The combination of ResUnet50 and SAM demonstrated potential for improving accuracy and efficiency, making it a viable approach for real-world clinical use.

Acknowledgements. The authors of this paper declare that the segmentation method they implemented for participation in the STS 2024 challenge has not used any additional datasets other than those provided by the organizers. The proposed solution is fully automatic without any manual intervention. We thank all the data owners for making the X-ray images and CT scans publicly available and Codebench [25] for hosting the challenge platform.

References

[1] Różyło-Kalinowska I. Panoramic radiography in dentistry[J]. Clinical Dentistry Reviewed, 2021, 5(1): 26. 1

[2] Wang X, Alqahtani K A, Van den Bogaert T, et al. Convolutional neural network for automated tooth segmentation on intraoral scans[J]. BMC Oral Health, 2024, 24(1): 804. 1

[3] Razzak M I, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future[J]. Classification in BioApps: Automation of decision making, 2018: 323-350. 2

[4] Chan H P, Samala R K, Hadjiiski L M, et al. Deep learning in medical image analysis[J]. Deep learning in medical image analysis: challenges and applications, 2020: 3-21.

[5] Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis[J]. Medical image analysis, 2022, 79: 102444. 2

[6] Chauhan R, Ghanshala K K, Joshi R C. Convolutional neural network (CNN) for image detection and recognition[C]//2018 first international conference on secure cyber computing and communication (ICSCCC). IEEE, 2018: 278-282. 2

[7] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241. 2

[8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778. 2

[9] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 4015-4026. 2

[10] Learning S S. Semi-supervised learning[J]. CSZ2006. html, 2006, 5: 2. 2

[11] Li Y, Yin J, Chen L. Informative pseudo-labeling for graph neural networks with few labels[J]. Data Mining and Knowledge Discovery, 2023, 37(1): 228-254. 2

[12] Amarù S, Marelli D, Ciocca G, et al. DALib: A Curated Repository of Libraries for Data Augmentation in Computer Vision[J]. Journal of Imaging, 2023, 9(10): 232. 3

[13] Ou Y, Zöllei L, Retzepi K, et al. Using clinically acquired MRI to construct age-specific ADC atlases: Quantifying spatiotemporal ADC changes from birth to 6-year old[J]. Human brain mapping, 2017, 38(6): 3052-3068. 3

[14] Patro S. Normalization: A preprocessing stage[J]. arXiv preprint arXiv:1503.06462, 2015. 3

[15] Rashid K M, Louis J. Window-warping: a time series data augmentation of imu data for construction equipment activity identification[C]//ISARC. Proceedings of the international symposium on automation and robotics in construction. IAARC Publications, 2019, 36: 651-657. 3, 4
[16] Gedraite E S, Hadad M. Investigation on the effect of a Gaussian Blur in image filtering and segmentation[C]//Proceedings ELMAR-2011. IEEE, 2011: 393-396. 4
[17] Aboussaleh I, Riffi J, Fazazy K E, et al. Efficient U-Net architecture with multiple encoders and attention mechanism decoders for brain tumor segmentation[J]. Diagnostics, 2023, 13(5): 872. 4
[18] Azad R, Heidary M, Yilmaz K, et al. Loss functions in the era of semantic segmentation: A survey and outlook[J]. arXiv preprint arXiv:2312.05391, 2023. 5
[19] Zhang Y, Ye F, Chen L, et al. Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection[J]. Scientific Data, 2023, 10(1): 380. 5
[20] Cui W, Wang Y, Zhang Q, et al. CTooth: a fully annotated 3d dataset and benchmark for tooth volume segmentation on cone beam computed tomography images [C]//International Conference on Intelligent Robotics and Applications. Cham: Springer International Publishing, 2022: 191-200.
[21] Cui W, Wang Y, Li Y, et al. CTooth+: A Large-scale Dental Cone Beam Computed Tomography Dataset and Benchmark for Tooth Volume Segmentation[C]// MICCAI Workshop on Data Augmentation, Labelling, and Imperfections. Cham: Springer Nature Switzerland, 2022: 64-73. 5
[22] Zhang H, Zhang Z, Odena A, et al. Consistency regularization for generative adversarial networks[J]. arXiv preprint arXiv:1910.12027, 2019. 11
[23] Rosenberg C, Hebert M, Schneiderman H. Semi-supervised self-training of object detection models[J]. 2005. 11
[24] Wang Y, Zhang Y, Chen X, et al. STS MICCAI 2023 Challenge: Grand challenge on 2D and 3D semi-supervised tooth segmentation[J]. arXiv preprint arXiv:2407.13246, 2024. 11
[25] Xu, Z., Escalera, S., Pavao, A., Richard, M., Tu, W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns 3(7) (2022). 12
[26] Cheng J, Ye J, Deng Z, et al. Sam-med2d[J]. arXiv preprint arXiv:2308.16184, 2023. 2
[27] Dai J, Guo X, Zhang H, et al. Cone-beam CT landmark detection for measuring basal bone width: a retrospective validation study[J]. BMC Oral Health, 2024, 24(1): 1091. 11
[28] Zhang H, Wang C W, Muzakky H, et al. Deep Learning Techniques for Automatic Lateral X-ray Cephalometric Landmark Detection: Is the Problem Solved?[J]. arXiv preprint arXiv:2409.15834, 2024.
1

Table 6. Checklist Table. Please fill out this checklist table in the answer column.

Requirements	Answer
A meaningful title	Yes
The number of authors (≤6)	3
Author affiliations and ORCID	Yes
Corresponding author email is presented	Yes
Validation scores are presented in the abstract	Yes
Introduction includes at least three parts: background, related work, and motivation	Yes
A pipeline/network figure is provided	Figure 1
Pre-processing	Page 2
Strategies to use the partial label	Page 3
Strategies to use the unlabeled images.	Page 4
Strategies to improve model inference	Page 3
Post-processing	Page 4
The dataset and evaluation metric section are presented	Page 5
Environment setting table is provided	Table 1
Training protocol table is provided	Table 2
Ablation study	Page 8
Efficiency evaluation results are provided	Table 3
Visualized segmentation example is provided	Figure 2
Limitation and future work are presented	Yes
Reference format is consistent.	Yes