Residual Attention U-Net for Tooth Segmentation in Panoramic Dental Radiographs with Auxiliary Lesion Segmentation

RAVELA.LATHA, K.CH.SRI KAVYA, SARAT K KOTAMARAJU

Department of ECE, Koneru Lakshmaiah Education Foundation, AP, India

Abstract

Accurate tooth segmentation in panoramic dental radiographs supports automated dental image analysis, but low contrast, overlapping structures, and indistinct boundaries complicate pixel-level delineation. This study evaluates a Residual Attention U-Net multi-task framework with tooth segmentation as the primary task and lesion segmentation as an auxiliary task. The Tufts Dental Database was partitioned into 700 training, 150 validation, and 150 independent test radiographs. The best improved multi-task checkpoint achieved tooth Dice of 91.14% and IoU of 83.73% at a threshold of 0.50. For lesion segmentation, the improved multi-task checkpoint at threshold 0.25 achieved Dice of 6.11%, whereas the separately fine-tuned lesion checkpoint at the validation-selected threshold of 0.05 achieved Dice of 13.86% and IoU of 7.44%. These results indicate stronger tooth delineation than lesion recovery and show that lesion performance depends on checkpoint and threshold selection.

Keywords: panoramic dental radiography; tooth segmentation; Residual Attention U-Net; muti-task learning; attention gates; lesion segmentation.

1. Introduction

Panoramic dental radiography provides a broad view of the dentition and maxillofacial region in a single image. Automated analysis of these images can support anatomical delineation and further image-analysis tasks. The Tufts Dental Database (TDD) provides 1,000 panoramic radiographs with expert annotations for dental image analysis [1].

Tooth segmentation is challenging because adjacent teeth can overlap, boundaries may be indistinct, restorations and missing teeth alter local appearance, and tooth morphology varies across images. Convolutional neural networks have become widely used for biomedical segmentation. U-Net combines encoder-derived contextual features with decoder reconstruction through skip connections [2].

Several adaptations have been investigated for panoramic dental images. Teeth U-Net uses multi-scale and attention-related processing [3], while other work has explored region-guided preprocessing and alternative U-Net variants [4]. Residual learning facilitates feature propagation through shortcut connections [5] and has been explored for dental segmentation [6]. Attention gates provide a mechanism for weighting spatially relevant features during decoding [7].

Pathological-region segmentation poses a separate challenge because lesion pixels may occupy only a small fraction of the image and can vary in size, contrast, and location. Prior studies

have investigated apical-lesion segmentation [8] and combined tooth/caries segmentation [9]. However, performance on anatomical tooth regions should not be assumed to transfer to small lesion regions.

This study investigates a Residual Attention U-Net with tooth segmentation as the primary task and auxiliary lesion segmentation as a second task. The framework uses residual blocks, attention-gated skip connections, and separate output heads. The tasks are evaluated independently, and the difference in performance is explicitly reported.

The contributions are: (1) an attention-gated residual encoder–decoder for tooth segmentation; (2) an auxiliary lesion output head and a two-stage training strategy; (3) independent test evaluation using pixel-level metrics and qualitative comparisons; and (4) an analysis of the limitations of lesion segmentation under severe class imbalance.

2. Materials and Methods

2.1 Dataset

The experiments used the Tufts Dental Database, comprising 1,000 panoramic dental radiographs [1]. The data were partitioned into 700 training images (70%), 150 validation images (15%), and 150 independent test images (15%). The training subset was used for parameter optimization; validation data supported checkpoint selection and lesion threshold selection; and the test subset was reserved for final evaluation. Tooth and lesion annotations were used as the corresponding segmentation targets.

Subset	Images	Proportion	Role
Training	700	70%	Network training and parameter optimization
Validation	150	15%	Model selection and lesion-threshold selection
Independent test	150	15%	Final quantitative and qualitative evaluation
Total	1,000	100%	Experimental cohort

Table 1. Dataset partition and use in model development.

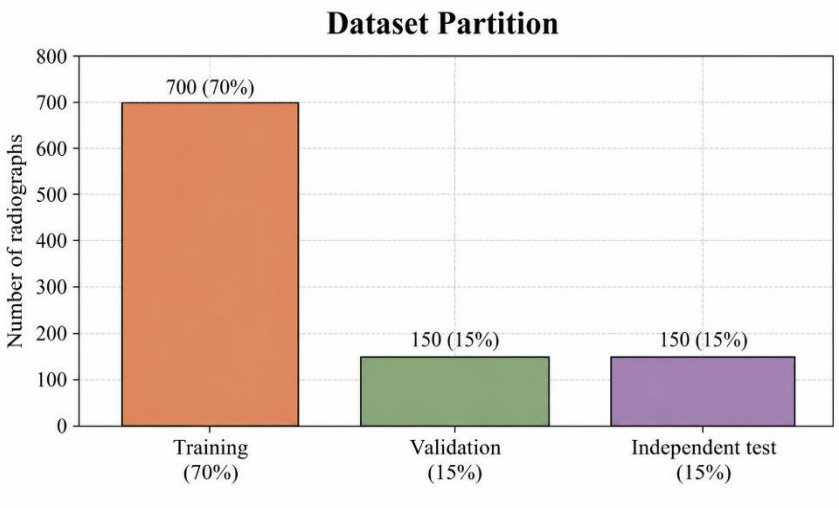


Fig. 1. Distribution of the panoramic radiographs across the training, validation, and independent test subsets

2.2 Image Preprocessing and Augmentation

Images were resized to 512 × 512 pixels. The corresponding tooth and lesion masks underwent the same spatial transformations to preserve pixel-level alignment. Masks were converted to grayscale and binarized using an intensity threshold of 127. Training-only augmentation included horizontal flipping ($p = 0.5$), vertical flipping ($p = 0.2$), rotation within ±20° ($p = 0.5$), brightness/contrast adjustment ($p = 0.5$), Gaussian noise ($p = 0.3$), and CLAHE ($p = 0.3$; clip limit 2; 8 × 8 tile grid). Input images were normalized using channel means (0.485, 0.456, 0.406) and standard deviations (0.229, 0.224, 0.225). Validation and test images were resized and normalized without stochastic augmentation.

2.3 Proposed Residual Attention U-Net Architecture

The network receives a three-channel 512 × 512 panoramic radiograph. Its encoder contains four residual stages with 64, 128, 256, and 512 channels, followed by a 1,024-channel bottleneck. Max pooling reduces spatial resolution between encoder stages. The decoder uses transposed-convolution upsampling and attention-gated skip connections to restore spatial resolution. Each gate uses the decoder representation as a gating signal and refines the corresponding encoder feature before concatenation.

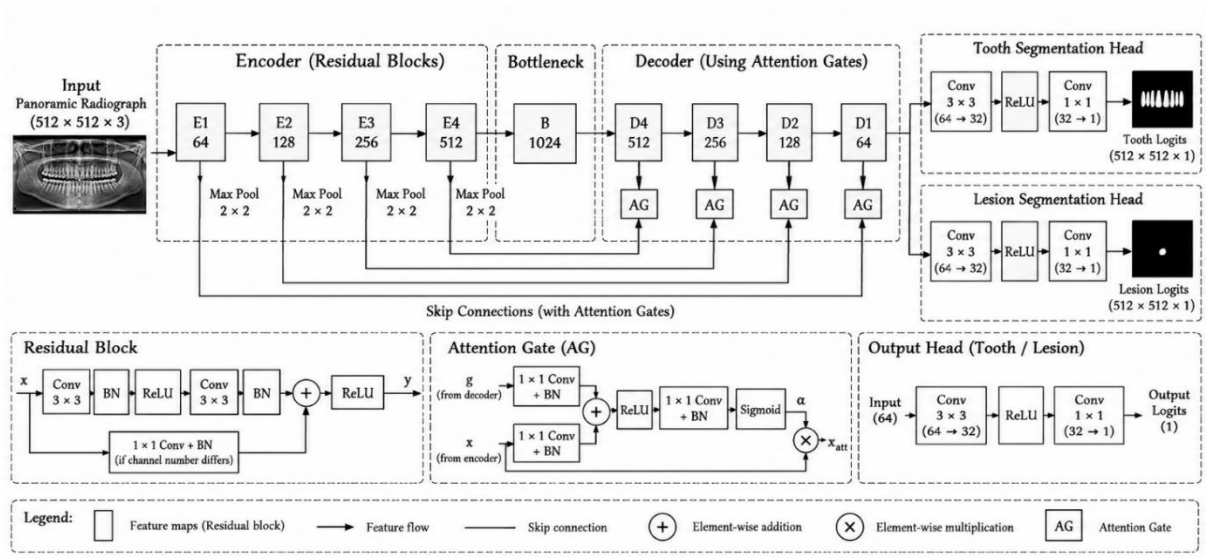


Fig. 2. Overall architecture of the proposed Residual Attention U-Net for multi-task tooth and auxiliary lesion segmentation.

Each residual block contains two convolutional layers with batch normalization and ReLU activation after the first convolution. A shortcut connection is added to the transformed representation; a 1 × 1 convolution and batch normalization align channel dimensions when needed. The decoder channel progression is 1,024 → 512 → 256 → 128 → 64.

The final 64-channel decoder representation is passed to two independent output heads. Each head uses a 3 × 3 convolution to reduce the feature channels to 32, ReLU activation, and a 1 × 1 convolution to produce a single-channel logit map. Sigmoid is applied during inference to obtain pixel probabilities. Tooth masks use a threshold of 0.50 the lesion threshold is selected using validation predictions.

2.4 Training Strategy and Optimization

Training consisted of joint multi-task optimization followed by lesion-specific fine-tuning. During joint training, the tooth objective combined binary cross-entropy and Dice loss. The lesion objective combined weighted binary cross-entropy, Dice loss, and focal Tversky loss, with positive-class weighting of 20 and an additional task-level weight of 2.5 for the lesion objective.

The joint model was trained for a maximum of 40 epochs using AdamW with batch size 2. Reduce LR On Plateau monitored validation loss and reduced the learning rate by a factor of 0.5 after three epochs without improvement. Early stopping used a patience of seven epochs. The checkpoint with the lowest validation loss was retained.

The best multi-task checkpoint initialized lesion-specific fine-tuning. The lesion fine-tuning objective combined weighted binary cross-entropy, Dice loss, and focal binary cross-entropy with weights 0.40, 0.40, and 0.20, respectively, and positive-class weighting of 50. Fine-tuning ran for up to 20 epochs with early-stopping patience five. The checkpoint with the highest validation lesion Dice was retained. The lesion threshold of 0.05 was selected using validation data and fixed before test evaluation.

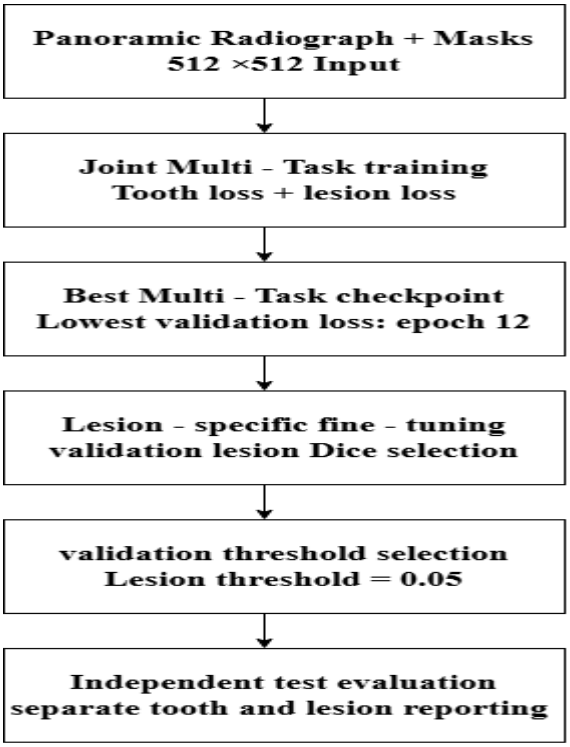


Fig. 3. Two-stage training and evaluation workflow.

Setting	Joint training	Lesion fine-tuning
Maximum epochs	40	20
Optimizer	AdamW	AdamW
Batch size	2	Confirm in notebook
Early-stopping patience	7	5
Checkpoint criterion	Lowest validation loss	Highest validation lesion Dice
Positive-class weight	20 (lesion BCE)	50 (lesion BCE)
Threshold	Tooth: 0.50	Lesion: 0.05 (validation-selected)

Table 2. Main training settings and checkpoint-selection criteria.

2.5 Evaluation Metrics

Accuracy, precision, recall, F1-score, intersection over union (IoU), and Dice coefficient were calculated from binary prediction masks. Accuracy measures overall pixel-wise correctness and precision measures the proportion of predicted foreground pixels that are true positives, recall measures the proportion of reference foreground pixels recovered; F1-score is the harmonic mean of precision and recall, IoU measures intersection divided by union and Dice measures twice the intersection divided by the sum of predicted and reference foreground pixels. Because lesion pixels are sparse, accuracy was interpreted alongside overlap and sensitivity metrics.

3. Results

3.1 Training and Validation Behaviour

The improved joint-training run completed 19 epochs. The lowest recorded validation loss was 1.3345 at epoch 12, and this checkpoint was retained as the best multi-task model. The training/validation loss plot should be generated from the saved improved_train_loss_history.npy and improved_val_loss_history.npy arrays; the chart below is not fabricated from unobserved per-epoch values.

The notebook loss plot records 19 epochs, with the lowest validation loss of 1.3345 at epoch 12. The saved loss curve is reproduced in Fig. 4.



Fig. 4. Training and validation loss during joint multi-task training the minimum recorded validation loss was 1.3345 at epoch 12.

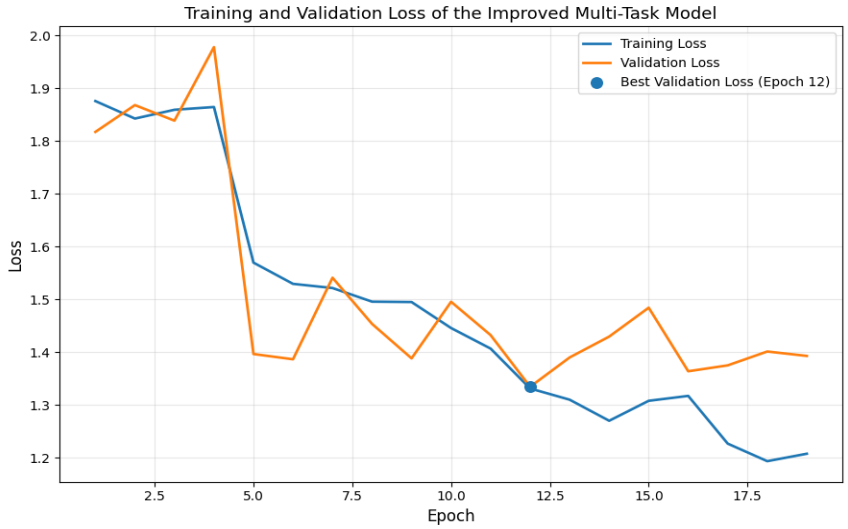


Fig. 5. Training and validation loss during joint multi-task training

3.2 Lesion Threshold Selection

The fine-tuned lesion model was assessed at candidate probability thresholds using validation predictions. The validation Dice score was highest at threshold 0.05 (0.2429) and decreased slightly at higher thresholds, reaching 0.2403 at threshold 0.50. Therefore, 0.05 was selected

using validation data before independent test evaluation. This threshold selection improved lesion overlap relative to the separately evaluated improved multi-task checkpoint at threshold 0.25, but the final test Dice remained limited at 13.86%.

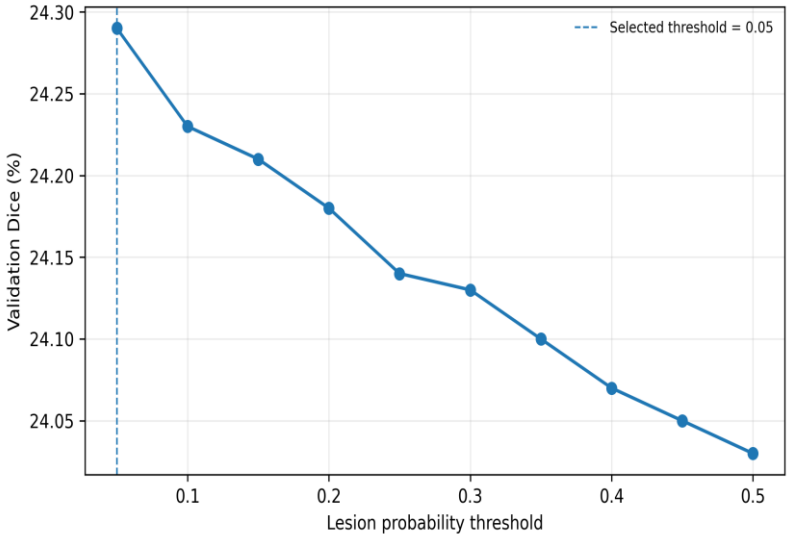


Fig. 6. Validation lesion Dice across probability thresholds for the fine-tuned model.

3.3 Independent Test Performance

The best improved multi-task checkpoint achieved tooth accuracy of 98.00%, precision of 91.52%, recall of 90.77%, F1-score and Dice of 91.14%, and IoU of 83.73% at a tooth threshold of 0.50. The same checkpoint achieved lesion accuracy of 99.58%, precision of 50.43%, recall of 3.25%, F1-score and Dice of 6.11%, and IoU of 3.15% at a lesion threshold of 0.25. Separately, the lesion-fine-tuned checkpoint achieved accuracy of 99.55%, precision of 36.04%, recall of 8.58%, F1-score and Dice of 13.86%, and IoU of 7.44% at the validation-selected threshold of 0.05.

Task/checkpoint	Threshold	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	IoU (%)	Dice (%)
Tooth / best improved multi-task	0.50	98.00	91.52	90.77	91.14	83.73	91.14
Lesion / improved multi-task	0.25	99.58	50.43	3.25	6.11	3.15	6.11
Lesion / fine-tuned checkpoint	0.05	99.55	36.04	8.58	13.86	7.44	13.86

Table 3. Independent test-set performance. Tooth and lesion metrics come from different selected checkpoints and are reported separately.

The two lesion rows correspond to different model checkpoints and probability thresholds they should not be interpreted as a same-checkpoint comparison.

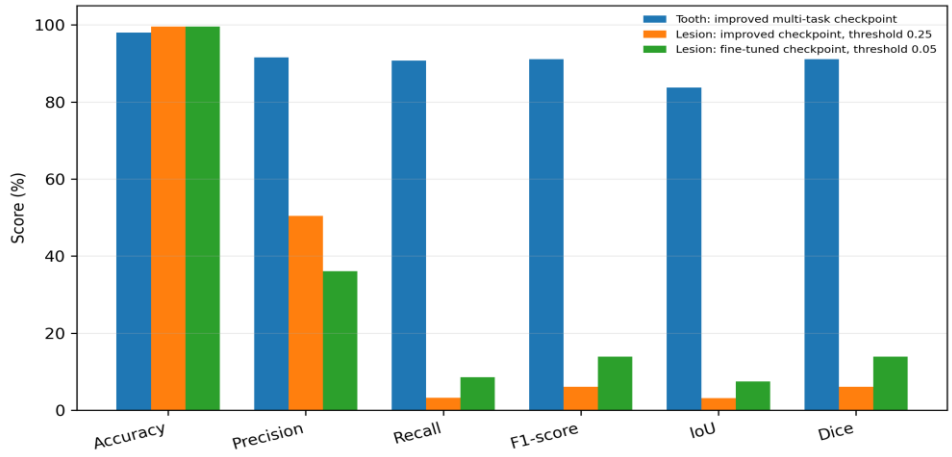


Fig. 7. Test-set metrics from the improved multi-task checkpoint and the separately fine-tuned lesion checkpoint, using the thresholds stated in Table 3.

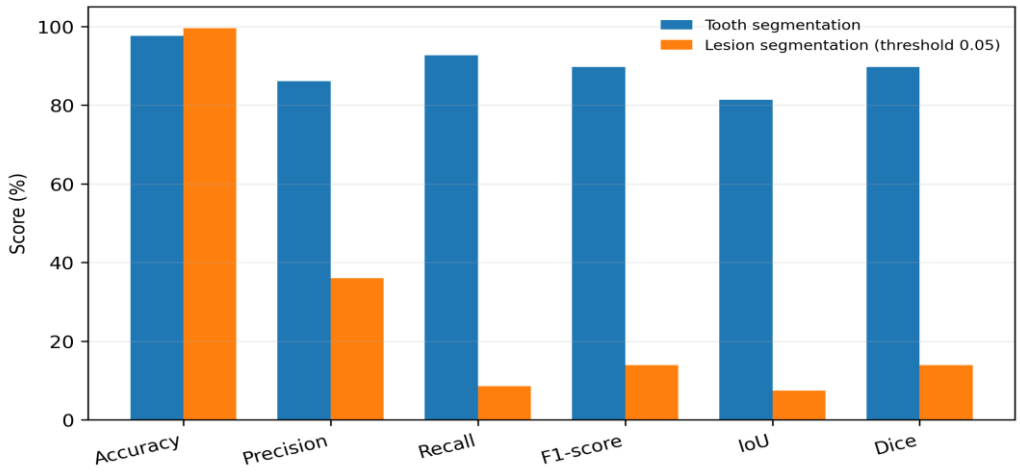


Fig. 8. Independent test metrics for tooth segmentation and lesion segmentation at threshold 0.05, using the separately reported task-specific test outputs.

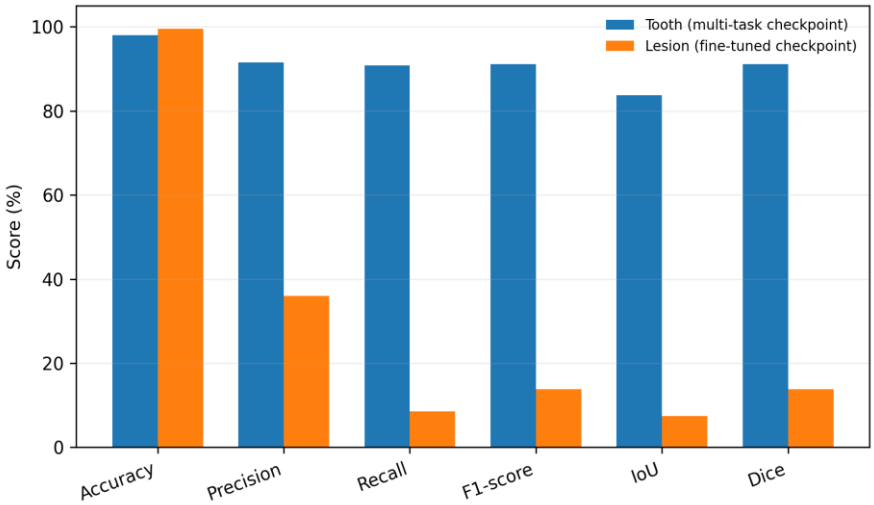


Fig. 9. Comparison of test-set metrics for tooth and lesion segmentation. Different checkpoints were used.

The lesion test confusion counts for the selected threshold were TP = 14,221, TN = 39,130,552, FP = 25,242, and FN = 151,585. Despite high pixel accuracy, low recall and overlap metrics show that the fine-tuned lesion model recovered only a small proportion of annotated lesion pixels.

3.4 Qualitative Results

Representative qualitative predictions from the independent test set are shown in Figs. 10–13. Each figure presents the original panoramic radiograph, ground-truth tooth mask, predicted tooth mask, ground-truth lesion mask, and predicted lesion mask. The examples illustrate relatively close agreement between predicted and reference tooth regions, whereas lesion predictions show variable and often incomplete recovery of annotated lesion regions.

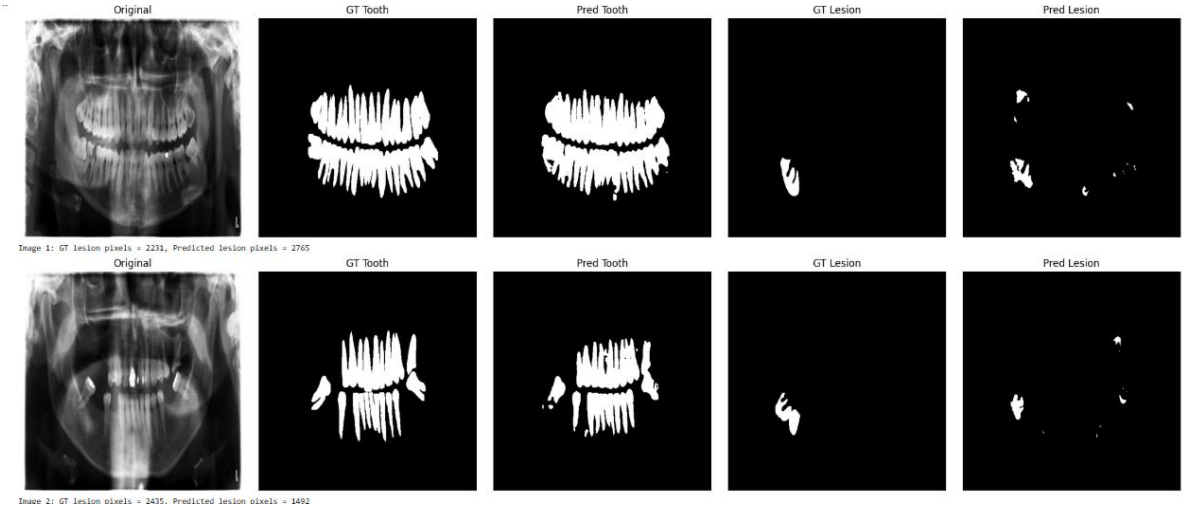


Fig. 10. Qualitative tooth and lesion segmentation results on representative independent test images. Columns show the original radiograph, ground-truth tooth mask, predicted tooth mask, ground-truth lesion mask, and predicted lesion mask.

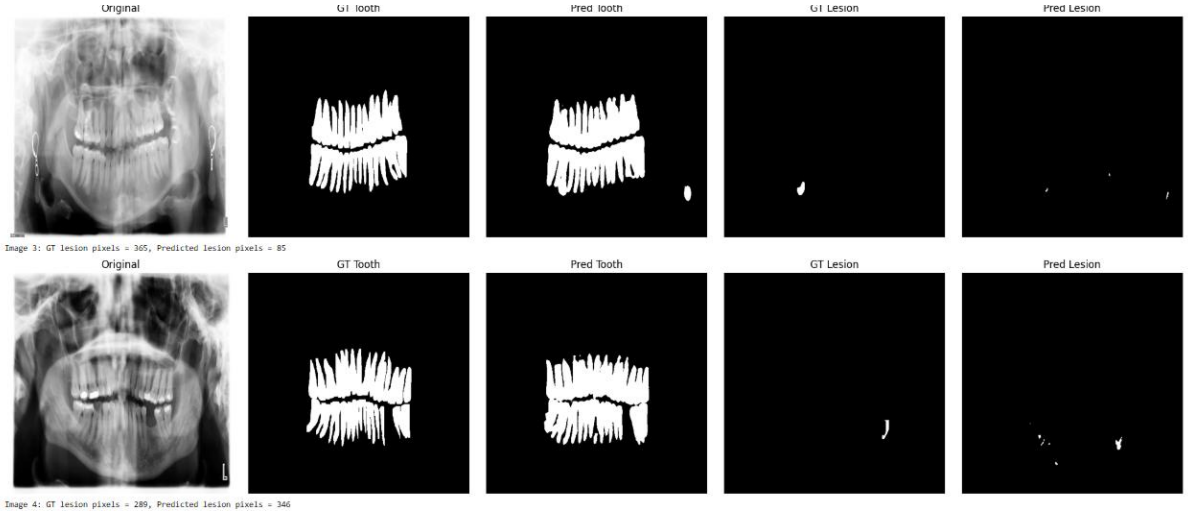


Fig. 11. Qualitative tooth and lesion segmentation results on representative independent test images. Columns show the original radiograph, ground-truth tooth mask, predicted tooth mask, ground-truth lesion mask, and predicted lesion mask.

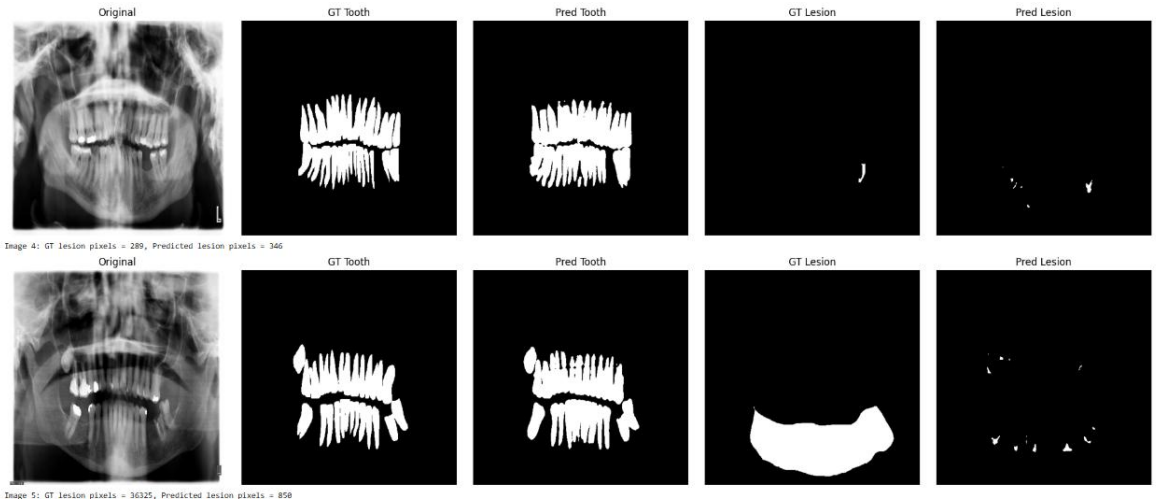


Fig. 12. Qualitative tooth and lesion segmentation results on representative independent test images. Columns show the original radiograph, ground-truth tooth mask, predicted tooth mask, ground-truth lesion mask, and predicted lesion mask.

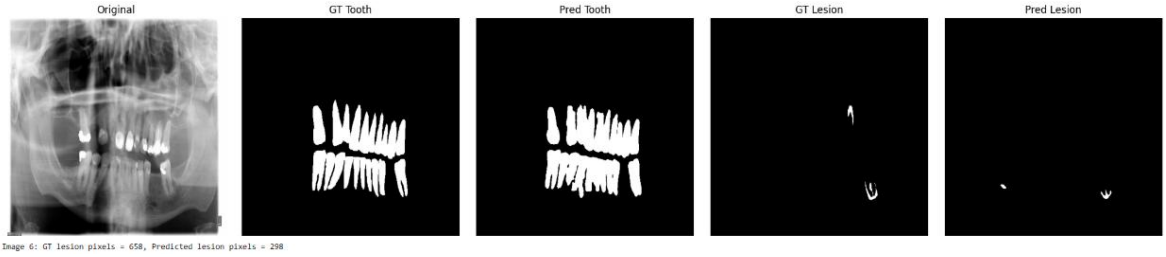


Fig. 13. Qualitative tooth and lesion segmentation results on representative independent test images. Columns show the original radiograph, ground-truth tooth mask, predicted tooth mask, ground-truth lesion mask, and predicted lesion mask.

4. Discussion

The improved multi-task checkpoint yielded a tooth Dice coefficient of 91.14% and IoU of 83.73%. Its lesion output at threshold 0.25 achieved Dice of 6.11% and IoU of 3.15%. The separately fine-tuned lesion checkpoint performed better on overlap metrics at the validation-selected threshold of 0.05, achieving Dice of 13.86% and IoU of 7.44%, although recall remained low at 8.58%.

Lesion segmentation remained substantially weaker than tooth segmentation. At threshold 0.25, the improved multi-task checkpoint achieved lesion Dice of 6.11%, IoU of 3.15%, and recall of 3.25%. The separately fine-tuned checkpoint at the validation-selected threshold of 0.05 achieved Dice of 13.86%, IoU of 7.44%, and recall of 8.58%. The high pixel accuracies should not be interpreted as successful lesion segmentation because background pixels dominate the images. These findings indicate incomplete lesion-region recovery; the lesion branch should be considered exploratory rather than clinically reliable.

A critical protocol detail is that tooth metrics were obtained from the best joint multi-task checkpoint, whereas lesion metrics were obtained from the separately fine-tuned checkpoint.

Since fine-tuning updates model parameters, the reported results are not predictions from one identical final checkpoint. Future experiments should evaluate both tasks on the same checkpoint in addition to task-specific fine-tuning, conduct ablation studies, examine lesion-focused sampling and losses, and validate on external datasets.

The comparison with prior studies must be interpreted cautiously because datasets, annotation protocols, image preprocessing, data partitions, and evaluation procedures can differ. The present results support the suitability of the tested framework for tooth segmentation on this dataset but do not establish external generalization or clinical utility.

5. Conclusion

This study evaluated a Residual Attention U-Net multi-task framework for tooth segmentation with auxiliary lesion segmentation in panoramic dental radiographs. The best improved multi-task checkpoint achieved tooth Dice of 91.14% and IoU of 83.73%. Lesion segmentation remained limited: the improved checkpoint achieved Dice of 6.11% at threshold 0.25, while a separately fine-tuned checkpoint achieved Dice of 13.86% and IoU of 7.44% at the validation-selected threshold of 0.05. Further lesion-specific optimization, same-checkpoint evaluation of both tasks, ablation studies, and independent external validation are required before clinical applicability can be assessed.

References

[1] K. Panetta, R. Rajendran, A. Ramesh, S. P. Rao, and S. Agaian, “Tufts Dental Database: A multimodal panoramic X-ray dataset for benchmarking diagnostic systems,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 4, pp. 1650–1659, 2022, doi: 10.1109/JBHI.2021.3117575.

[2] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. MICCAI, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[3] S. Hou, T. Zhou, Y. Liu, P. Dang, H. Lu, and H. Shi, “Teeth U-Net: A segmentation model of dental panoramic X-ray images for context semantics and contrast enhancement,” Computers in Biology and Medicine, vol. 152, Art. no. 106296, 2023, doi: 10.1016/j.compbiomed.2022.106296.

[4] N. Sogur and N. K. Cercevik, “Region-guided preprocessing and U-Net variants for improved panoramic dental radiograph segmentation,” Biomedical Signal Processing and Control, vol. 120, Part B, Art. no. 110068, 2026, doi: 10.1016/j.bspc.2026.110068. Verify bibliographic details before submission.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. CVPR, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[6] S. A. Almalki, S. Alsubai, A. Alqahtani, and A. A. Alenazi, “Denoised encoder-based residual U-Net for precise teeth image segmentation and damage prediction on panoramic radiographs,” Journal of Dentistry, vol. 137, Art. no. 104651, 2023, doi: 10.1016/j.jdent.2023.104651.

[7] O. Oktay et al., “Attention U-Net: Learning where to look for the pancreas,” arXiv:1804.03999, 2018.

[8] I. S. Song et al., “Deep learning-based apical lesion segmentation from panoramic radiographs,” Imaging Science in Dentistry, vol. 52, no. 4, pp. 351–357, 2022, doi: 10.5624/isd.20220078.

[9] A. C. Mărginean, S. Mureşanu, M. Hedeşiu, and L. Dioşan, “Teeth segmentation and carious lesions segmentation in panoramic X-ray images using CariSeg, a networks’ ensemble,” Heliyon, vol. 10, no. 10, Art. no. e30836, 2024, doi: 10.1016/j.heliyon.2024.e30836.