Automatic Joint Teeth Segmentation in Panoramic Dental Images using Mask R-CNN with Residual Feature Extraction:

Can it be useful in Oral Cancer Diagnosis and Management?

Raghavendra H. Bhalerao[1], Abhijeet Ashok Salunke[2,*], Shristi Sharan[1], Kamlesh Kumar[1], Priyank Rathod[2], Prince Kumar[1], Manish Chaturvedi[1], Nandlal Bharwani[2], Krupa Shah[1], Dhruv Patel[2], Keval Patel[2], Vikas Warikoo[2], Manisha Abhijeet Salunke[3], Shashank Pandya[2]

[1] Department of Electrical Engineering, IITRAM, Ahmedabad, India

[2] Department of Surgical Oncology, The Gujarat Cancer and Research Institute, Ahmedabad, India

[3] Dental Surgeon, Ahmedabad, India

* Corresponding author: Abhijeet Ashok Salunke

Abstract

Introduction. Panoramic dental radiographs provide an overview of the teeth, jaws, and surrounding structures. Automated segmentation may support the quantitative analysis of dental images. This study evaluated a modified Mask R-CNN approach with residual feature extraction for the segmentation of individual teeth in panoramic radiographs.

Material and Methods. A sequence of residual blocks was used to construct a reported 62-layer feature extraction network for a modified Mask R-CNN (MRCNN). The UFBA-UESC and Tufts dental image datasets comprised 2,500 panoramic radiographs: 1,800 were assigned to training, 448 to validation, and 252 to testing.

Results. The modified model achieved a reported training accuracy of 99.67% and validation accuracy of 98.94%. Averaged across validation runs, the author-reported corrected Dice score was 98.67%, intersection over union (IoU) was 97.8%, and pixel accuracy was 96.53%. On the separate test set of 252 radiographs, the Dice scores were 96.0% for the modified model and 83.0% for the original model. Training, validation, and test metrics are reported separately.

Conclusion. The findings support the potential of the proposed approach for automated tooth segmentation in panoramic dental radiographs. Accurate delineation of individual teeth may provide a useful component for future computer-assisted dental image analysis. Its possible contribution to oral cancer assessment, treatment planning, or follow-up requires dedicated clinical validation, as these applications were not evaluated in the present study.

Keywords: Panoramic radiography; tooth segmentation; Mask R-CNN; residual learning; medical image analysis.

Received: 19 September 2024 · Published: 25 September 2024 · Open Access · CC BY 4.0

Introduction

Panoramic radiographs provide an overview of the teeth, jaws, and surrounding structures and are widely used in dental image assessment. Automated tooth segmentation may support image analysis by delineating individual teeth and providing a basis for quantitative measurements. In the context of oral oncology, such tools could potentially contribute to future image-analysis workflows. However, segmentation of teeth does not by itself establish the ability to detect malignancy or assess treatment response. The present study therefore evaluates technical tooth-segmentation performance rather than clinical performance in oral cancer diagnosis or management.

Poor contrast, limited resolution, and noise in panoramic radiographs can make image interpretation and segmentation difficult. Medical image segmentation assigns image pixels to objects or regions of interest. The aim here is to delineate individual teeth, including adjacent teeth that may be difficult to separate.

Silva et al. reviewed conventional approaches to tooth segmentation and introduced the UFBA-UESC dataset of 1,500 panoramic radiographs (1). Earlier studies used region-based, threshold-based, and boundary-based techniques for dental or jaw-image analysis (2-6). Separation of teeth from surrounding bone can remain challenging.

Deep-learning methods have been investigated for semantic and instance segmentation. Jader et al. used Mask R-CNN for instance segmentation of teeth (6,7). Koch et al. used U-Nets for segmentation of panoramic dental radiographs (8,9).

This study examines a modified Mask R-CNN with residual feature extraction. The original report also includes results labelled FPN, U-Net, U-Net++, PSPNet, DeepLabV3, and DeepLabV3+; limitations of these comparisons are described below.

Material and Methods

A. Data

The UFBA-UESC dental image dataset (1) contributed 1,500 images, reported as 1991 × 1127 pixels, with tooth-number annotations in FDI notation and COCO format. The original report describes eight groups according to tooth presence, restorations, and appliances. The Tufts Dental Database (19) contributed 1,000 annotated panoramic radiographs. The combined dataset contained 2,500 images.

According to the original study report, 1,800 images were assigned to training, 448 to validation, and 252 to testing. The training-to-validation ratio among the 2,248 non-test images was approximately 8:2.

The authors confirm that the corrected validation metrics were averaged across validation runs: Dice 98.67%, IoU 97.8%, and pixel accuracy 96.53%. Individual validation-run results are not available. Test-set performance is reported separately and was not included in these validation means.

The original report described four groups of 112 validation images, but the available records do not establish whether these represented independent training folds or subdivisions of a fixed validation set. A reproducible four-fold cross-validation protocol therefore cannot be specified.

The original report also states that different training, validation, and test proportions were explored before selecting the final allocation. The selection criterion and whether test outcomes informed this choice are not documented. The test results are consequently described as results on the reported test set, without asserting an independently protected holdout.

Patient-level separation, random seeds, and split files were not available for verification in this revision. This limits assessment of generalisability; it does not itself establish that overlap occurred.

Material and Methods

B. Mask R-CNN architecture

The architecture is based on Mask R-CNN (7), illustrated in Figure 1A. Region-based convolutional neural networks detect candidate objects and refine their class assignments and bounding boxes. R-CNN used convolutional features with a classifier (10); Fast R-CNN introduced ROI pooling (11), and Faster R-CNN introduced a region proposal network (RPN) (12). Feature pyramid networks combine information at multiple scales (13).

The RPN predicts objectness and bounding-box coordinates. Mask R-CNN adds a mask-prediction branch alongside classification and box regression. Fully convolutional approaches provide the broader context for pixel-level prediction (15).

C. Proposed residual CNN

The original study describes a 62-layer feature extraction network consisting of 20 residual blocks. The diagram divides it into six stages, S1-S6 (Figure 1B). The baseline backbone is inconsistently named ResNet50 and ResNet101 in the original materials; its exact configuration cannot be established from the available implementation records.

Figure 1B shows 4, 4, 6, 3, and 3 residual blocks in stages S2-S6, respectively, giving 20 blocks in total. Each illustrated bottleneck block contains three convolutions: 1 × 1, 3 × 3, and 1 × 1. The original description of three blocks per stage has been corrected to distinguish convolutional layers from repeated residual blocks.

A 1 × 1 projection on a shortcut is described for matching dimensions at stage transitions. Adam was reported as the optimiser.

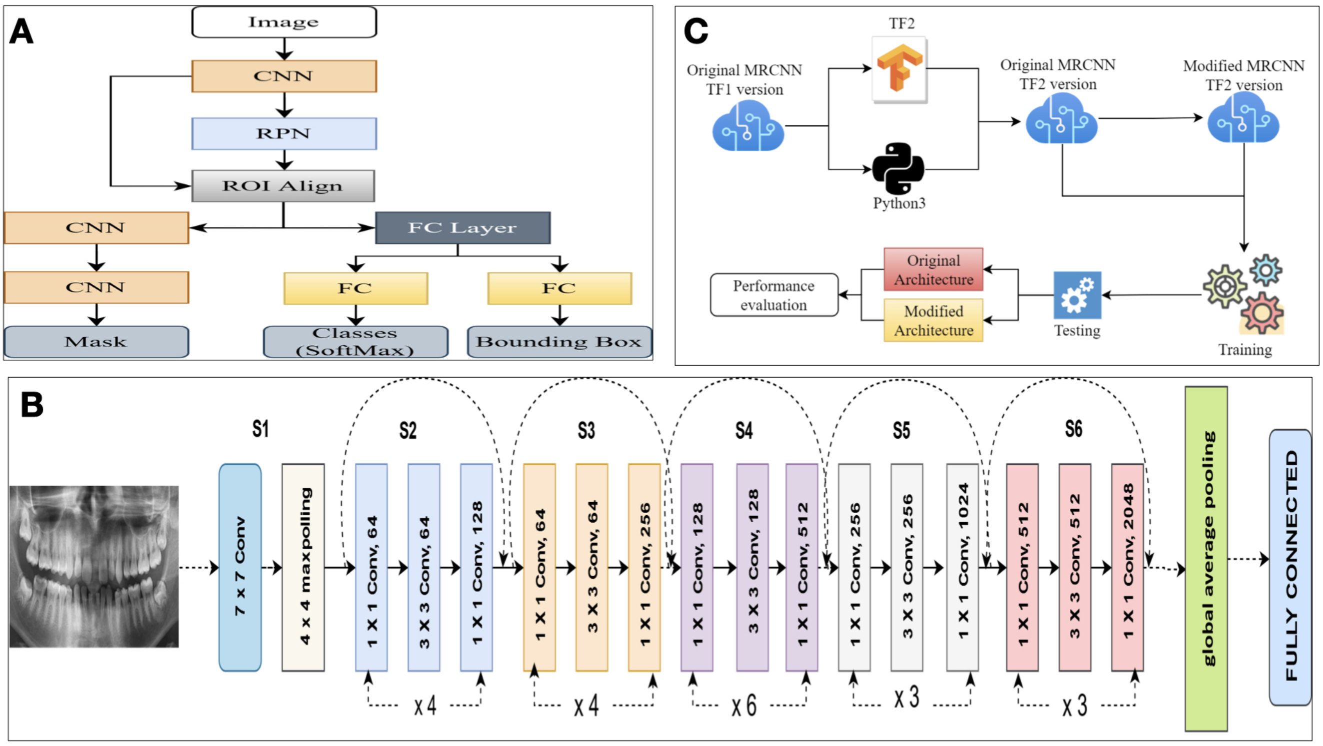

Figure 1. A: Original Mask R-CNN schematic, showing the CNN, RPN, ROI Align, mask branch, and fully connected (FC) outputs for classes and bounding boxes. **B:** Reported modified residual backbone; ×n denotes block repetitions. **C:** Implementation workflow: conversion from TensorFlow 1 to TensorFlow 2, training, testing, and evaluation. Original image panels are retained.

Material and Methods

C. Proposed residual CNN (continued)

The original description divides the network into Network-1 (S1-S5), using conventional convolutions, and Network-2 (S6), described as using dilated convolutions. Figure 1B shows a 7 × 7 convolution and 4 × 4 max pooling before the residual stages, followed by global average pooling and a fully connected layer. The text states that the feature-extraction portion does not use the final fully connected layer.

The original records describe both 512 × 512 image resizing and random square input patches of that size. Stride settings and the precise relation between these preprocessing descriptions are not documented consistently. No unverified settings have been reconstructed.

D. Proposed modifications

The residual backbone and modified model are illustrated in Figures 1B and 2A. Input dental radiographs are processed layer by layer to obtain feature maps, with an FPN structure combining multi-scale information. The backend performs classification, bounding-box regression, and instance segmentation. Within each predicted box, the mask distinguishes tooth from background.

ROI cropping is proposed to introduce location information into the segmentation pathway (Figure 2A). These architectural features describe the proposed design; their individual contribution was not established through a reported ablation study. The mapping of all diagram components to the trained implementation remains a reproducibility limitation.

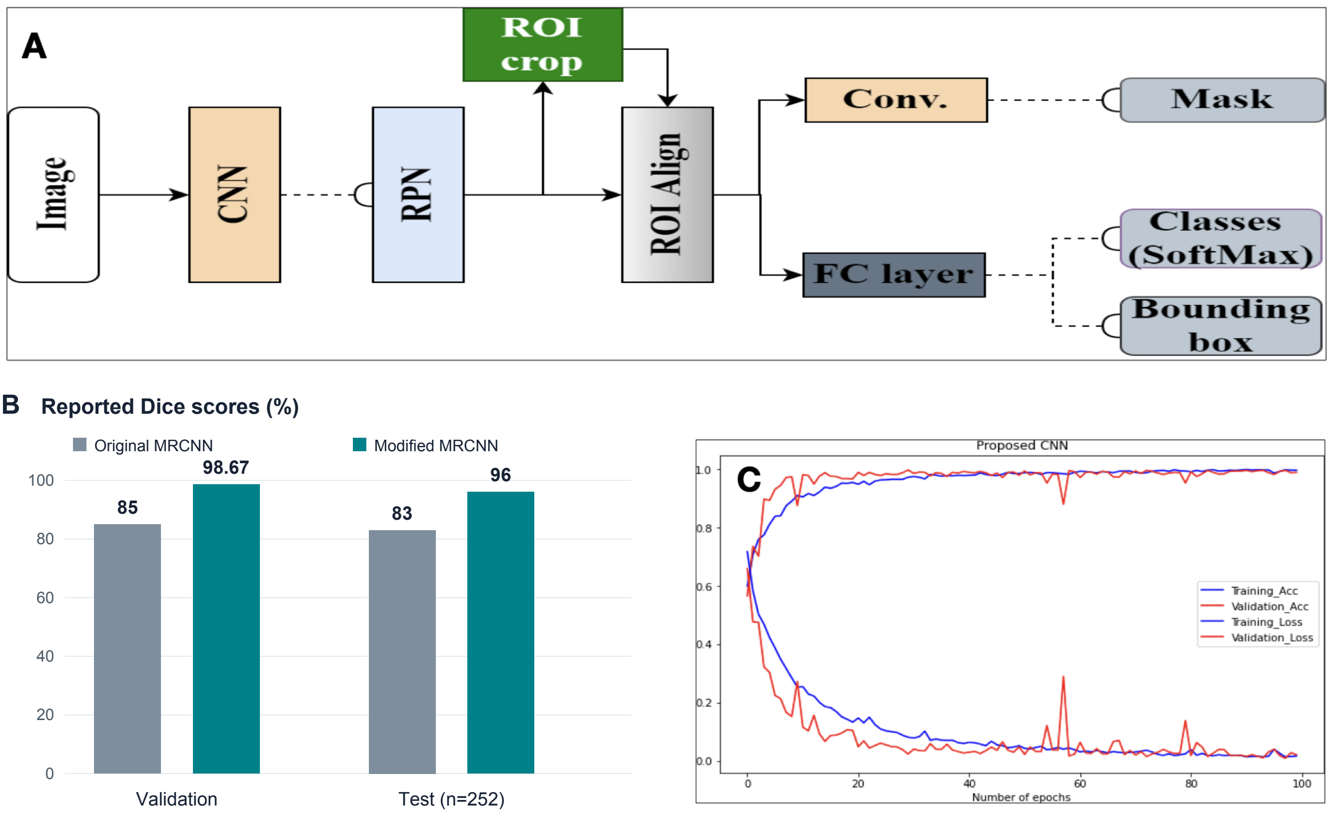

Figure 2. A: Modified Mask R-CNN with ROI cropping. **B:** Aggregate Dice scores, separated into validation and test results. The modified-model validation mean is corrected to 98.67%; the other three scores are retained from the original report. No run-level values or error bars are available. **C:** Original training/validation accuracy (upper curves) and loss (lower curves), blue/red respectively. Curves are retained as reported; the accuracy definition and loss implementation cannot be verified. Panel B replaces the earlier four-group chart.

Material and Methods

E. Loss function and training

The authors confirm that Dice loss was used for mask-segmentation optimisation. The exact loss implementation is not available for verification in this revision; consequently, its mathematical variant, smoothing, reduction and weighting, and combination with the other model-loss components cannot be specified retrospectively. No reconstructed formula is presented as the training objective used in this study.

The original and modified models were compared. Images were reported as resized to 512 × 512 pixels. The training description specifies 64 feature maps initially, doubling with downsampling to 512. Their precise correspondence to the channel counts up to 2048 in Figure 1B is not documented.

The reported initial learning rate was 0.0001, with Adam optimisation, batch size 2, and 100 epochs. When validation performance did not improve, the learning rate was halved every 15 epochs.

F. Implementation and evaluation workflow

The reported hardware was an NVIDIA RTX 3060 with 12 GB GPU memory and 32 GB system RAM; mean epoch duration was approximately 85 seconds.

The workflow is shown in Figure 1C. TensorFlow 2.7.0, Keras, and Python 3 were reported, with a public Mask R-CNN implementation cited as the starting point (20). This reference does not identify a verified commit, trained weights, or the complete implementation of the modified 62-layer network.

The modified model was trained using 1,800 annotated panoramic radiographs, with 448 images assigned to validation. The trained model was then evaluated on the reported test set of 252 images. The original model was reported as trained and evaluated using the same procedure.

Figure 2C preserves the original training curves. The reported training-process accuracy values are kept distinct from Dice, IoU, and pixel accuracy. Exact definitions and underlying logs for these curves are unavailable.

A

B

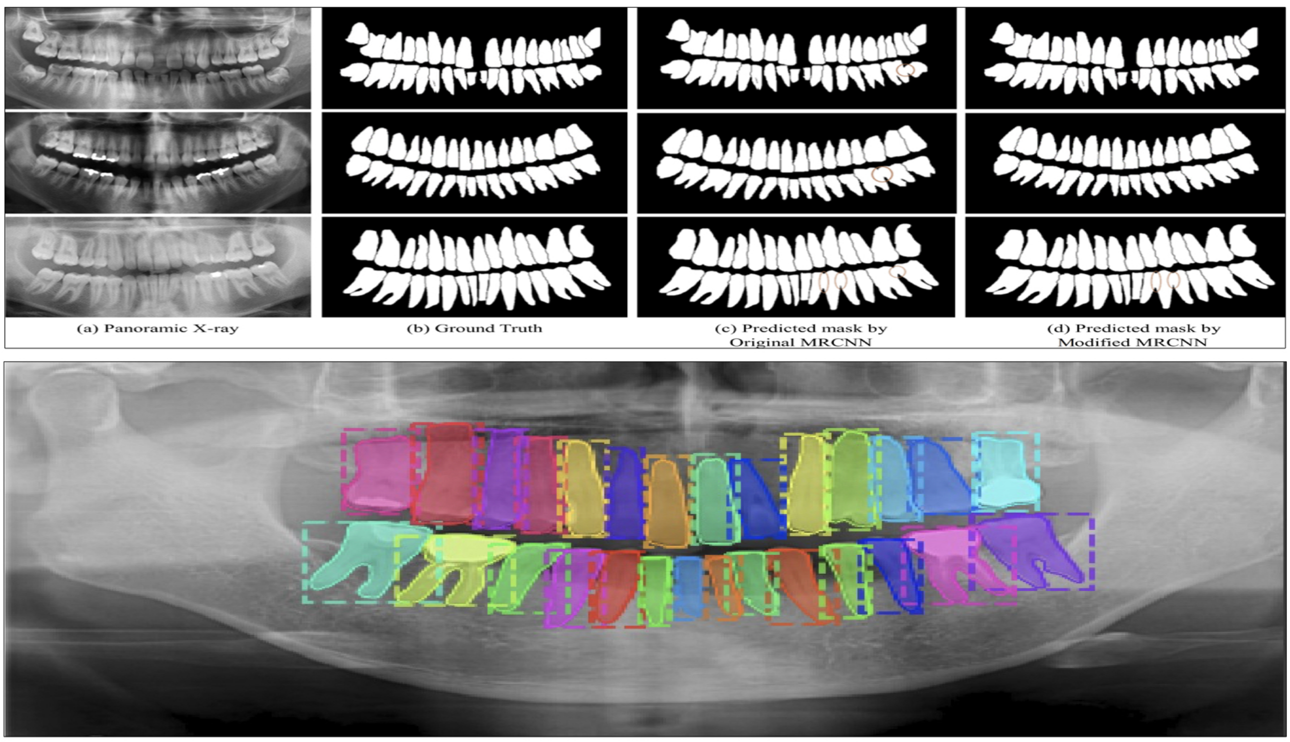

Figure 3. A: Examples of (a) panoramic input images, (b) ground-truth masks, (c) original-model predictions, and (d) modified-model predictions. Circles identify segmentation errors, including an error in the third modified-model example. **B:** Individual tooth masks and bounding boxes generated by the proposed model, displayed in different colours. Original panels are retained.

Material and Methods

G. Performance metrics

Dice score measures overlap between a predicted segmentation and a reference mask. Pixel accuracy is the proportion of correctly classified pixels among evaluated pixels. Intersection over union (IoU) is the intersection of predicted and reference foreground regions divided by their union. Mean IoU refers to an average over specified classes.

These descriptions identify the intended metrics. The exact class inclusion, handling of empty masks, reduction, and weighting used in the reported calculations cannot be established retrospectively. The corrected validation figures are author-reported means across validation runs.

H. Residual blocks

Deep networks such as VGG (16) and GoogLeNet (17) motivated research into greater depth while controlling computational cost and optimisation difficulties. He et al. introduced residual learning (18), in which a block learns a residual mapping added to a shortcut input.

If the desired mapping is $F(x)$, the residual can be written $R(x) = F(x) - x$, so that the block output is $R(x) + x$. A two-layer block can contain two 3×3 convolutions; a bottleneck block uses 1×1, 3×3, and 1×1 convolutions. The 1×1 layers control the channel dimension around the spatial convolution.

The proposed design uses three-convolution bottleneck blocks. Residual connections assist optimisation; they do not establish a clinical benefit or guarantee freedom from overfitting.

Results

Training performance

The final training accuracy reported for the modified model was 99.67%. The training and validation curves are shown in Figure 2C. These training-process metrics are distinguished from the segmentation metrics reported below.

Validation performance

Averaged across validation runs, the author-confirmed corrected values for the modified model were 98.67% for the Dice score, 97.8% for intersection over union (IoU), and 96.53% for pixel accuracy (Table 1). Individual run-level values are unavailable; variability and confidence intervals cannot be reconstructed. The original model had a reported validation Dice score of 85.0%. The difference between the reported validation Dice scores was therefore 13.67 percentage points.

The validation accuracy reported during training was 98.94% (Figure 2C). This value is reported separately and should not be interpreted as the Dice score or pixel accuracy in Table 1.

Test performance

On the reported test set of 252 panoramic radiographs, the Dice score was 96.0% for the modified model and 83.0% for the original model. The difference was 13.0 percentage points. These test results are distinct from the validation results and form the basis for describing performance on the reported test set.

Interpretation

The reported Dice scores were numerically higher for the modified model on the validation and test data. These are aggregate author-reported estimates; the unavailable implementation details and individual validation-run results limit independent verification and interpretation of comparative performance. The findings concern tooth segmentation and do not establish diagnostic accuracy for oral cancer, treatment-response assessment, or improvement in patient outcomes. The unresolved compatibility of IoU and pixel accuracy is addressed in the limitations.

Results and Discussion

Table 1. Reported segmentation metrics. The corrected modified-model row contains validation means across runs. Comparator values are retained from the original report, which associated the comparison with Tufts data; their exact split and evaluation protocol are not established. Rows should not be treated as a verified like-for-like ranking.

Model	IoU (%)	Pixel accuracy (%)	Dice (%)
FPN (13)	86.37	95.17	92.24
U-Net (9)	86.67	95.22	92.46
U-Net++ (21)	86.62	95.19	92.47
PSPNet (22)	86.08	95.00	91.77
DeepLabV3 (23)	86.21	95.00	92.00
DeepLabV3+ (24)	86.41	95.13	91.80
Modified MRCNN - validation	97.8	96.53	98.67

Qualitative results

Figure 3 shows input radiographs, reference masks, and predictions from both models. The displayed modified-model examples separate some adjacent teeth that are merged in original-model predictions. Errors nevertheless remain, including the circled example in the third row. These selected examples illustrate behaviour and do not quantify its frequency.

Discussion

The study examines a modified residual backbone and ROI cropping for tooth segmentation in panoramic radiographs. The reported Dice scores are numerically higher for the modified model: by 13.67 percentage points in validation and 13.0 percentage points on the reported test set. These differences are descriptive and do not establish statistical or independently verified superiority.

The proposed ROI pathway may help convey location information, but the available comparison does not isolate this component from other architecture and training changes. A causal attribution of the observed difference to ROI cropping alone is therefore not supported.

Panoramic tooth segmentation is challenging because of noise, uneven contrast, overlapping structures, restorations, and variability in tooth position. Similar tooth shapes and grey-level intensities can make boundaries difficult to distinguish. Missing teeth and changes in neighbouring tooth positions can reduce the usefulness of location priors.

Jader et al. investigated Mask R-CNN for instance segmentation of teeth (6). The present results can be considered alongside this work, but differences in datasets, splits, training, and evaluation prevent an unqualified comparison. Table 1 retains the originally reported comparator values with their provenance limitations. The segmentation head associated with the FPN row and complete common comparison protocols are not documented.

The observed segmentation performance provides a basis for further research into computer-assisted dental image analysis. Any application to oral cancer assessment would require disease-specific evaluation beyond the technical segmentation experiments reported here.

Future technical work could improve detection and mask quality, investigate tooth numbering and missing teeth, and evaluate robustness across imaging conditions. These remain research objectives.

Limitations

The results should be interpreted as author-reported aggregate estimates. The exact loss implementation and individual validation-run outputs are unavailable, limiting reproducibility and preventing reconstruction of between-run variability or confidence intervals. The averaging and weighting procedures within each metric cannot be independently verified.

The reported validation IoU exceeds the reported pixel accuracy. For binary masks evaluated on the same pixels with consistent weighting, IoU cannot exceed pixel accuracy; averaging over the same runs does not remove this constraint. Thus, the compatibility of these aggregate values cannot be established from the available records. Differences in metric computation or weighting might account for the discrepancy, but this explanation cannot be confirmed retrospectively. The reported figures should therefore not be treated as independently verified evidence of comparative superiority.

The study assessed tooth segmentation. It did not evaluate oral cancer detection, treatment response, recurrence, or patient outcomes. Potential clinical applications require dedicated validation with appropriate reference standards.

The exact backbone configuration, preprocessing sequence, split-selection procedure, patient-level separation, and common comparator protocols are not fully documented. The original descriptions and diagrams are retained only to the extent that they can be identified as reported design information. They do not substitute for executable code, trained weights, or evaluation records.

The reported test results have not been newly recalculated or independently confirmed in this revision. No uncertainty estimates, significance tests, or external-validation claims can be derived from the available aggregate values.

Clinical interpretation

Accurate tooth segmentation may be useful as a supporting component of computer-assisted dental image analysis. Potential applications include anatomical delineation, quantitative comparison of dental images, and the development of tools for planning or follow-up. These possibilities should be distinguished from the outcomes evaluated in the present study, which concern segmentation performance.

The present results do not establish diagnostic performance for oral cancer, the detection of recurrence, assessment of treatment response, or improvements in patient outcomes. No dedicated disease-specific or longitudinal evaluation of these endpoints was reported. Accordingly, the proposed method should be considered a technical foundation for further investigation rather than a clinically validated tool for these purposes.

Before use in oral oncology, the method would require evaluation in representative patient cohorts with appropriate clinical reference standards. Such studies should assess its added value within a clinician-led workflow, including robustness across imaging conditions and the consequences of segmentation errors. Potential applications to treatment planning, prosthetic design, or follow-up remain subjects for future research and should not be interpreted as demonstrated benefits of the current model.

Conclusion

This study evaluated a modified Mask R-CNN approach with residual feature extraction for automated tooth segmentation in panoramic dental radiographs. The reported results suggest potential for delineating individual teeth and support further development of computer-assisted dental image-analysis methods.

The clinical implications should be interpreted within the scope of the study. Oral cancer detection, treatment response, recurrence, and patient outcomes were not evaluated; therefore, no conclusions about clinical effectiveness in these areas can be drawn. Future work should address external validation, reproducibility, computational efficiency, and disease-specific clinical evaluation before such applications are considered.

Correspondence

Dr. Abhijeet Ashok Salunke
ORCID: 0000-0003-0103-8599
The Gujarat Cancer & Research Institute
Civil Hospital Campus, Asarwa
Ahmedabad 380 016, Gujarat, India
Phone: +91-79-2268 8000
Fax: +91-79-2268 5490

Conflict of interest

The authors declare that there were no conflicts of interest within the meaning of the recommendations of the International Committee of Medical Journal Editors when the article was written.

Data and implementation records

The public datasets are identified in references 1 and 19, and the cited software repository in reference 20. Complete study-specific split files, the exact modified-model implementation, and individual validation-run results are not available for verification in this revision.

Publisher's note

The statements, opinions, and data are those of the individual authors and contributors, and not of the Swiss Journal of Radiology and Nuclear Medicine or its editors. The publisher and editors disclaim responsibility for injury to people or property resulting from ideas, methods, instructions, or products referred to in the content.

License

References

1. G. Silva, L. Oliveira, and M. Pithon, “Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives,” Expert Systems with Applications, vol. 107, pp. 15 – 31, 2018. doi: https://doi.org/10.1016/j.eswa.2018.04.001

2. A. Lurie, G. M. Tosoni, J. Tsimikas, and F. Walker Jr., “Recursive hierarchic segmentation analysis of bone mineral density changes on digital panoramic images,” Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, vol. 113(4), pp. 549–558.e1, 2012. https://doi.org/10.1016/j.oooo.2011.10.002

3. Y.Y. Amer and M. J. Aqel, “An efficient segmentation algorithm for panoramic dental images,” Procedia Computer Science, vol. 65, pp. 718–725, 2015, international Conference on Communications, management, and Information technology (ICCMIT’2015). https://doi.org/10.1016/j.procs.2015.09.016

4. M. K. Alsmadi, “A hybrid fuzzy c-means and neutrosophic for jaw lesions segmentation,” Ain Shams Engineering Journal, vol. 9, no. 4, pp. 697–706, 2018. https://doi.org/10.1016/j.asej.2016.03.016

5. M. R. M. Razali, N. S. Ahmad, R. Hassan, Z. M. Zaki, and W. Ismail, “Sobel and canny edges segmentations for the dental age assessment,” in 2014 International Conference on Computer Assisted System in Health, 2014, pp. 62–66. https://doi.org/10.1109/CASH.2014.10

6. G. Jader, J. Fontinele, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, “Deep instance segmentation of teeth in panoramic x-ray images,” in 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2018, pp. 400–407. https://doi.org/10.1109/SIBGRAPI.2018.00058

7. K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp.2980–2988. https://arxiv.org/abs/1703.06870

8. T. L. Koch, M. Perslev, C. Igel, and S. S. Brandt, “Accurate segmentation of dental panoramic radiographs with U-Nets,” in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 15–19. https://doi.org/10.1109/ISBI.2019.8759563

9. O. Ronneberger, P.Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention (MICCAI), ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

References (continued)

10. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142–158, Jan. 2016, https://doi.org/10.1109/TPAMI.2015.2437384

11. R. Girshick, “Fast R-CNN,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169

12. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. https://doi.org/10.1109/TPAMI.2016.2577031

13. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2117–2125. https://doi.org/10.1109/CVPR.2017.106

14. F. Deng, H. Hu, S. Chen, Q. Guan, and Y. Zou, “Rich feature hierarchies for cell detecting under phase contrast microscopy images”, in 2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP), Nov. 2015, pp. 348–353. https://doi.org/10.1109/ICICIP.2015.7388195

15. N. Atif, M. Bhuyan, and S. Ahamed, “A Review on Semantic Segmentation from a Modern Perspective,” in 2019 International Conference on Electrical, Electronics and Computer Engineering (UPCON), Nov. 2019, pp. 1–6. https://doi.org/10.1109/UPCON47278.2019.8980189

16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014) https://doi.org/10.48550/arXiv.1409.1556

17. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015) https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html

18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016) https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

19. 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", in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 4, pp. 1650-1659, April 2022, https://doi.org/10.1109/JBHI.2021.3117575

20. Mask-RCNN implementation for Tensorflow 2.7.0 and Keras 2.7.0. https://github.com/Kamlesh364/Mask-RCNN-TF2.7.0-keras2.7.0

21. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested U-net architecture for medical image segmentation,” in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Berlin, Germany: Springer, 2018, pp. 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

22. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2881–2890. https://openaccess.thecvf.com/content_cvpr_2017/html/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.html

23. L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” 2017, https://doi.org/10.48550/arXiv.1706.05587.

24. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-Decoder with atrous separable convolution for semantic image segmentation,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 801–818. https://openaccess.thecvf.com/content_ECCV_2018/html/Liang-Chieh_Chen_Encoder-Decoder_with_Atrous_ECCV_2018_paper.html

25. Su, B., Zhang, Q., Gong, Y. et al. Deep learning-based classification and segmentation for scalpels. Int J CARS. 2023;18(5):855-864. https://doi.org/10.1007/s11548-022-02825-7

26. Wang, H., Xiao, N., Luo, S. et al. Multi-scale dense selective network based on border modeling for lung nodule segmentation. Int J CARS. 2023;18(5):845-853. https://doi.org/10.1007/s11548-022-02817-7

27. Manjunatha, Y., Sharma, V., Iwahori, Y. et al. Lymph node detection in CT scans using modified U-Net with residual learning and 3D deep network. Int J CARS. 2023;18(4):723-732. https://doi.org/10.1007/s11548-022-02822-w