U-Net for Segmentation of Dental Structures in Bitewing Radiographs

Murilo Salem
Federal University of Pelotas (UFPel)
Pelotas, Brazil
mcsalem@inf.ufpel.edu.br

Rafael Grasel
Federal University of Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
rgrasel0@gmail.com

Daniel Barretos
Federal University of Pelotas (UFPel)
Pelotas, Brazil
dhspbarretos@inf.ufpel.edu.br

Anderson Ferrugem
Federal University of Pelotas (UFPel)
Pelotas, Brazil
ferrugem@inf.ufpel.edu.br

Marcos Alves
Federal University of Pelotas (UFPel)
Pelotas, Brazil
mlalves@inf.ufpel.edu.br

Gabriel Porto
Federal University of Pelotas (UFPel)
Pelotas, Brazil
gmporto@inf.ufpel.edu.br

Samuel Starke
Federal University of Pelotas (UFPel)
Pelotas, Brazil
sstarke@inf.ufpel.edu.br

Jonas Almeida Rodrigues
Federal University of Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
jonasrodrigues@ufrgs.br

Abstract

Bitewing radiographs play an important role in dental diagnosis, especially in the detection of caries, evaluation of restorations, and follow-up of endodontic procedures. However, their manual interpretation is time-consuming and subject to variability, particularly in cases involving overlapping structures, low contrast, and restorative materials. In this context, deep learning methods have emerged as promising tools for assisting radiographic analysis. This paper presents an applied semantic segmentation pipeline for bitewing radiographs based on U-Net with a ResNet50 backbone. The proposed approach combines grayscale preprocessing, data augmentation, transfer learning, and multi-class segmentation to identify crowns, restorations, and root canal treatments. Experimental evaluation yielded an Intersection over Union (IoU) of 0.43, precision of 0.47, recall of 0.98, and F1-score of 0.44. These results indicate high sensitivity but limited precision, mainly due to false positives. Therefore, the present study should be interpreted as an initial feasibility investigation rather than as a comparative benchmark or a clinically deployable solution. Even so, the model was able to identify relevant dental structures and provides an applied basis for discussing bitewing radiographs as Smart Medical Media, in which AI-generated overlays may enrich visual interpretation, documentation, and communication in digital dentistry.

Keywords

dental radiography, semantic segmentation, U-Net, deep learning, bitewing radiographs, digital dentistry

1 Introduction

Dental radiographs are essential resources in clinical practice because they support diagnosis, treatment planning, and follow-up in several oral health conditions. Among the available radiographic modalities, bitewing radiographs are particularly relevant for detecting proximal caries, evaluating existing restorations, and monitoring structures associated with conservative and endodontic treatment [5, 15]. Their diagnostic value makes them one of the most frequently used imaging modalities in routine dental care.

Despite their usefulness, the interpretation of bitewing radiographs remains a demanding task. Anatomical overlap, image noise, variable contrast, and the visual similarity between distinct dental structures may complicate both human reading and automated analysis [1, 2]. In addition, manual interpretation depends on clinician experience and may be affected by interobserver variability, especially in subtle or ambiguous cases. These factors motivate the development of computational methods capable of assisting image interpretation in a more standardized and reproducible manner.

In recent years, artificial intelligence (AI), especially deep learning, has shown strong potential in medical image analysis. Convolutional neural networks have been successfully applied to classification, detection, and segmentation tasks, with segmentation being particularly relevant when precise localization of structures is required [3, 9]. In dentistry, segmentation has practical importance because it can support the delineation of restorations, crowns, endodontic materials, and other structures of clinical relevance, potentially contributing to visual explanation and decision support.

Among segmentation architectures, U-Net remains one of the most established approaches due to its encoder–decoder structure and skip connections, which allow the combination of contextual and spatial information [9]. When combined with pretrained backbones such as ResNet50, U-Net can benefit from transfer learning and improve feature extraction, especially in scenarios with limited datasets [3]. For applied research settings, this combination offers a robust baseline for investigating segmentation performance without requiring highly specialized model design.

This paper presents an applied deep learning pipeline for semantic segmentation of dental structures in bitewing radiographs using U-Net with a ResNet50 backbone. The target classes are crowns, restorations, and root canal treatments. The contribution of the work is primarily applied rather than methodological: instead of proposing a novel architecture, we adapt and evaluate a well-established segmentation pipeline in a specific dental imaging context and discuss the results from the perspective of Smart Medical Media. This perspective is relevant because radiographs are not only diagnostic records, but also digital media artifacts that can be computationally enriched with overlays, masks, and visual annotations.

The remainder of the paper is organized as follows. Section 2 presents the clinical and conceptual context of bitewing radiographs and Smart Medical Media. Section 3 describes the dataset, preprocessing steps, architecture, and training setup. Section 4 reports and discusses the experimental results. Section 5 outlines the main limitations and implications of the study. Finally, Section 6 concludes the paper and suggests directions for future work.

2 Bitewing Radiographs and Smart Medical Media

Bitewing radiographs occupy a central place in digital dentistry because they provide a compact but information-rich view of posterior teeth and surrounding structures. They are commonly used to reveal proximal caries, recurrent caries under restorations, coronal adaptations, and radiopaque evidence associated with previous procedures [15, 16]. Their clinical value is therefore well established, but their interpretation remains visually demanding.

From a computational perspective, bitewing radiographs are also challenging images. Relevant structures may appear with subtle boundaries, low local contrast, or partial overlap, while metallic or restorative materials may create regions with strong intensity variations [1]. In addition, anatomical diversity between patients and acquisition variability between exams reduce the uniformity of the visual patterns available for learning. These characteristics make bitewing radiographs a relevant but difficult application domain for semantic segmentation.

The literature on AI in dentistry has grown considerably, including studies on lesion detection, radiographic classification, and image segmentation [10, 12]. Still, many contributions focus primarily on predictive performance and less on how outputs can be incorporated into broader digital workflows. In this sense, the notion of Smart Medical Media offers a useful conceptual lens. Under this view, medical images are not only passive data records, but media objects that may be enriched, reorganized, and interpreted through computational layers such as overlays, annotations, and explanatory visual cues [6, 13].

Applied to dental radiography, this perspective suggests that segmentation masks may serve not only as algorithmic outputs, but also as visual resources that support communication, documentation, and interpretation. For instance, predicted masks could highlight structures of interest during image review, facilitate explanation in educational or clinical contexts, and integrate with digital dental systems that manage records, visual assets, and patient-facing media. However, such possibilities depend on the reliability of the underlying segmentation process. For this reason, in the present work the Smart Medical Media perspective is treated cautiously: the current results support the feasibility of segmentation-assisted visualization, but do not justify stronger claims of immediate clinical deployment.

3 Methodology

3.1 Dataset and Target Classes

The dataset is composed of bitewing radiographs and their corresponding segmentation masks. The segmentation task is defined over three target classes of clinical relevance: crowns, restorations, and root canal treatments. These classes were selected because they represent structures and interventions that are visually important in radiographic interpretation and can benefit from explicit delineation.

Rather than treating the problem as a generic binary segmentation task, the proposed pipeline adopts a multi-class configuration. This choice increases the practical relevance of the experiment by distinguishing structures with different diagnostic meanings, while also increasing task complexity. Multi-class segmentation is especially challenging in dental radiographs because some classes may occupy small regions, present visual similarity, or appear less frequently in the dataset.

3.2 Preprocessing

All radiographs are read in grayscale and resized to 512×512 pixels. Pixel intensities are normalized to the $[0, 1]$ interval to improve numerical stability during training. The corresponding masks are resized to the same spatial resolution and stacked channel-wise, resulting in a multi-class target tensor of shape $(512, 512, 3)$.

This preprocessing choice aims to preserve the radiographic content while standardizing the input dimensions. Because radiographic datasets often contain images with different sizes and acquisition settings, resizing and normalization are important to ensure a consistent input distribution for the neural network.

3.3 Data Augmentation and Partitioning

To reduce overfitting and improve variability during training, data augmentation is applied to both images and masks in a synchronized manner. The augmentation pipeline includes horizontal flipping, brightness and contrast adjustment, small translations, and rotations restricted to 0° or 180° [7, 11]. These transformations were chosen to increase diversity while preserving the spatial correspondence between image content and segmentation labels.

The dataset is randomly divided into training and validation subsets using a 70%/30% split. Input loading, shuffling, batching, and preprocessing are implemented with TensorFlow pipelines, allowing the training procedure to remain consistent and reproducible.

3.4 Model Architecture

The proposed model is based on U-Net with a ResNet50 encoder pretrained on ImageNet. U-Net is a well-established segmentation architecture composed of an encoder–decoder structure with skip connections that merge low-level spatial information with high-level contextual features [9]. This design is especially useful in medical image segmentation, where accurate localization and preservation of fine details are important.

The use of ResNet50 as encoder introduces the advantages of transfer learning. Residual connections support deeper feature extraction, while ImageNet pretraining provides a strong initialization for visual representations [3]. Since the input radiographs are grayscale, a preliminary 1×1 convolution maps the single-channel input into a 3-channel representation compatible with the pretrained backbone. The decoder reconstructs the segmentation masks progressively and combines encoder features through skip connections. The final layer employs sigmoid activation over three channels, enabling pixel-wise multi-label prediction.

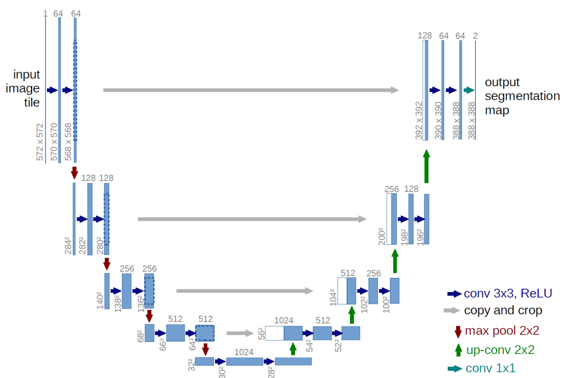

Figure 1: U-Net-based segmentation architecture with a ResNet50 encoder.

3.5 Training Setup

Training is performed using the Adam optimizer with an initial learning rate of 10^{-4} [4]. To improve convergence and reduce overfitting, the training process also uses checkpointing, early stopping, and learning rate scheduling. These strategies are common in applied deep learning workflows and are particularly useful when working with limited datasets [14].

Model performance is evaluated through Intersection over Union (IoU), precision, recall, and F1-score. Together, these metrics provide a broad view of segmentation quality. IoU and F1-score indicate the overlap between prediction and reference masks, while precision and recall highlight the balance between false positives and false negatives.

4 Results and Discussion

Table 1 summarizes the quantitative results obtained on the validation subset. Figure 2 presents representative qualitative examples comparing predicted masks and reference masks.

The results reveal a clear imbalance between sensitivity and selectivity. The recall of 0.98 indicates that the model is highly sensitive to the presence of relevant structures, meaning that most target regions tend to be captured by the predictions. On the other hand, the precision of 0.47 shows that a considerable proportion of

Table 1: Segmentation performance on the validation set.

IoU	Precision	Recall	F1-score
0.43	0.47	0.98	0.44

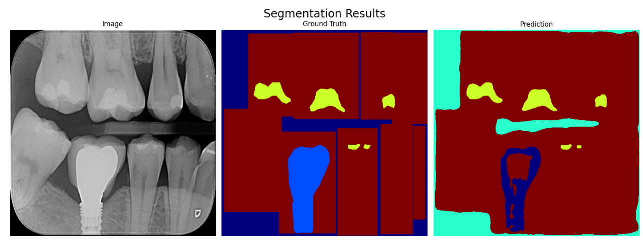

Figure 2: Representative segmentation outcomes. The examples illustrate both successful delineation of dental structures and failure cases with false positives.

predicted positive pixels do not correspond to the reference masks. This behavior is consistent with the observed false positives and helps explain the relatively low IoU (0.43) and F1-score (0.44).

From a practical standpoint, this pattern suggests that the model is more effective at broadly identifying candidate regions than at producing accurate and compact delineations. In exploratory settings, this behavior may still be useful because high recall reduces the risk of completely missing relevant structures. However, the current precision level limits the reliability of the segmentation output for more demanding use cases. In other words, the model appears to capture meaningful radiographic patterns, but still lacks sufficient selectivity to support stronger claims of robustness.

The qualitative examples reinforce this interpretation. Some predictions successfully highlight structures associated with the target classes, indicating that the model learned relevant visual cues from the radiographs. At the same time, other cases exhibit over-segmentation and spurious responses, particularly in regions whose grayscale patterns may resemble the target categories. This combination of useful detections and noisy predictions suggests that the present configuration is promising as an initial applied study, but not yet mature enough to support claims of precise clinical automation.

These findings are also important for the Smart Medical Media discussion. Even with modest segmentation scores, the predicted masks can still function as visual overlays that enrich image interpretation in a human-in-the-loop context. Their current value lies less in replacing specialist judgment and more in supporting exploratory analysis, digital documentation, and visual explanation. Therefore, the experimental results do not validate a high-confidence diagnostic tool, but they do support the feasibility of using segmentation as a computational enrichment layer over bitewing radiographs.

5 Limitations and Implications

This study has several limitations that should be made explicit. First, the dataset is limited in size, which restricts the diversity of visual patterns available during training and reduces the model's capacity to generalize. Second, class imbalance likely affects the learning process, especially for structures that appear less frequently or occupy smaller regions in the images. Third, the work does not include direct comparisons against baseline approaches or alternative segmentation architectures, which limits the strength of the technical validation. Fourth, per-class metrics were not reported, preventing a more detailed analysis of which structures were segmented more or less effectively.

These limitations are directly reflected in the interpretation of the results. The current performance is better understood as evidence of feasibility than as a definitive technical benchmark. The study shows that a U-Net + ResNet50 pipeline can learn patterns associated with crowns, restorations, and root canal treatments in bitewing radiographs, but it also shows that false positives remain a relevant challenge. As a consequence, the system should not be interpreted as a standalone clinical tool.

Even so, the work remains relevant from an applied and media-oriented perspective. In digital dentistry, segmentation outputs may be useful as interpretative layers that highlight structures of interest, support documentation workflows, and facilitate communication in settings where human review remains central [8, 12]. In this sense, the main implication of the study is not immediate diagnostic automation, but the demonstration that radiographic segmentation can contribute to the broader notion of Smart Medical Media, especially when used as an assistive visual resource rather than as a final authority.

6 Conclusion

This paper presented an applied deep learning pipeline for multi-class segmentation of bitewing radiographs using U-Net with a ResNet50 backbone. The model was trained to identify crowns, restorations, and root canal treatments through a workflow involving grayscale preprocessing, data augmentation, transfer learning, and semantic segmentation.

The results indicate that the model is able to identify relevant dental structures, but also that false positives substantially reduce precision and overall overlap quality. For this reason, the study should be interpreted as an initial feasibility investigation in a specific dental imaging context, rather than as a comparative benchmark or a clinically ready solution. Its main contribution lies in demonstrating that segmentation of bitewing radiographs is viable as an applied research direction and in framing the resulting masks as potential enrichment layers within Smart Medical Media.

Future work should expand the dataset, report per-class performance, and compare the proposed approach with baseline models and alternative segmentation architectures. Such steps are necessary to strengthen the technical contribution and to better assess the role of AI-assisted segmentation in digital dentistry.

References

[1] Abdulbadea Altukroni, Omar Ezz El-Deen, and Sadaf Jabeen. 2023. Enhancing the Quality of Dental Radiographic Images: A Review on Panoramic and Periapical Radiograph Enhancement Techniques. 13, 5 (2023), 1–9. doi:10.4172/2161-0681.23.13.457 Publisher: OMICS International.

[2] Burak Dayı, Hüseyin Üzen, İpek Balıkçı Çiçek, and Şuayip Burak Duman. 2023. A Novel Deep Learning-Based Approach for Segmentation of Different Type Caries Lesions on Panoramic Radiographs. 13, 2 (2023), 202. doi:10.3390/diagnostics13020202 Publisher: Multidisciplinary Digital Publishing Institute.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778. doi:10.1109/CVPR.2016.90

[4] Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)* (2015).

[5] J. H. Lee, D. H. Kim, S. N. Jeong, and S. H. Choi. 2023. Detection and diagnosis of dental caries using artificial intelligence and panoramic radiography: A systematic review. *Oral Radiology* 39, 2 (2023), 183–191. doi:10.1007/s11282-022-00616-7

[6] V. Patel, P. Kumar, and A. Sharma. 2021. Smart media in dental practice: Enhancing clinical workflows with digital integration. *International Journal of Dental Technology* 12, 3 (2021), 145–157.

[7] Luis Perez and Jason Wang. 2017. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621* (2017).

[8] Rata Rokhshad, Maxime Ducret, Akhilanand Chaurasia, Teodora Karteva, Miroslav Radenkovic, Jelena Roganovic, Manal Hamdan, Hossein Mohammad-Rahimi, Joachim Krois, Pierre Lahoud, and Falk Schwendicke. 2023. Ethical considerations on artificial intelligence in dentistry: A framework and checklist. *Journal of Dentistry* 135 (2023), 104593. doi:10.1016/j.jdent.2023.104593

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2015), 234–241.

[10] James-Andrew Sarmiento, Liushifeng Chen, and Prospero Naval. 2023. Multi-class Semantic Segmentation of Tooth Pathologies and Anatomical Structures on Bitewing and Periapical Radiographs. In *2023 18th International Conference on Machine Vision and Applications (MVA)* (2023-07). 1–5. doi:10.23919/MVA57639.2023.10215653

[11] Connor Shorten and Taghi M Khoshgoftaar. 2019. A survey on image data augmentation for deep learning. *Journal of Big Data* 6, 1 (2019), 60.

[12] L. D. Slashcheva, K. Schroeder, L. J. Heaton, H. J. Cheung, B. Prosa, N. Ferrian, J. Grantz, D. Jacobi, J. J. O'Malley, M. Helgeson, and E. P. Tranby. 2025. Artificial intelligence-produced radiographic enhancements in dental clinical care: provider and patient perspectives. *Frontiers in Oral Health* 6 (2025), 1473877. doi:10.3389/froh.2025.1473877

[13] R. Smith and L. Jones. 2022. The role of multimedia and smart technologies in modern dentistry. *Journal of Digital Dentistry* 5, 1 (2022), 33–47.

[14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, and et al. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.

[15] Ann Wenzel. 2021. Radiographic modalities for diagnosis of caries in a historical perspective: from film to machine-intelligence supported systems. 50, 5 (2021). doi:10.1259/dmfr.20210010 Publisher: Oxford Academic.

[16] S. C. White and M. J. Pharoah. 2018. *Oral Radiology: Principles and Interpretation* (7th ed.). Elsevier.