Pertanika J. Sci. & Technol. 33 (S3): 65 - 79 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Enhancing Yolov8 Backbone Using Gradient Descent-based Method for Dental Segmentation

Dhiaa Mohammed Abed[1,2], Shuzlina Abdul-Rahman[2]* and Sofianita Mutalib[2]

[1]*Faculty of Biomedical Engineering, University of Technology, Baghdad, Iraq*
[2]*School of Computer Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

ABSTRACT

This research addresses the challenge of dental segmentation in computer vision, a task focused on accurately outlining dental structures in images. The traditional methods, particularly convolution neural networks (CNNs), often suffer from suboptimal performance and computational inefficiency. Our study introduces an enhanced approach by applying the YOLOv8 algorithm, known for its effectiveness in object detection, for dental segmentation. Our proposed model improves YOLOv8's feature extraction capability by integrating additional layers into its backbone architecture, primarily focusing on the Coordinates-To-Features (C2f) module. This C2f-based feature extraction technique is designed to optimize gradient descent, reducing loss and maximizing prediction accuracy. By incorporating adaptive weights, the model effectively enhances the propagation of gradients, allowing for a more precise focus on dental structures. The adapted model, comprising 29 layers, is trained on a large-scale real-color dental dataset. Experimental evaluation demonstrates that the proposed model achieves exceptional performance, attaining 99.6% precision and recall in dental segmentation tasks. These results highlight the potential of YOLOv8 for specialized segmentation challenges and mark a significant contribution to automated dental analysis, offering direct benefits for clinical diagnostics and treatment planning in dentistry.

Keywords: Computer vision, deep learning, dental segmentation, image processing, YOLOv8

ARTICLE INFO

Article history:
Received: 22 August 2024
Accepted: 24 February 2025
Published: 24 April 2025

DOI: https://doi.org/10.47836/pjst.33.S3.05

E-mail addresses:
Dhiaa.M.Alfyadh@uotechnology.edu.iq (Dhiaa Mohammed Abed)
shuzlina@uitm.edu.my (Shuzlina Abdul-Rahman)
sofianita@uitm.edu.my (Sofianita Mutalib)
* Corresponding author

INTRODUCTION

Image segmentation is a major task in computer vision that divides an image into meaningful and semantically cohesive parts or segments. Its purpose is to distinguish pixels or regions with similar visual features, like color, texture, intensity, or spatial proximity, from the surrounding regions.

e-ISSN: 2231-8526

Image segmentation is essential for a variety of applications, including object recognition, scene interpretation, medical image analysis, autonomous cars, and more (Elyan et al., 2022).

The primary objective of image segmentation is to allocate a label or identifier to each pixel or region of the image, effectively partitioning the image into distinct regions based on the underlying visual characteristics. These labeled regions can then be used for further analysis, feature extraction, object detection, and other computer vision-related tasks (Bi et al., 2022).

Semantic segmentation assigns every pixel in an image to a definite class label, demonstrating the object or section it goes to. The goal is to achieve pixel-level labeling, where all pixels belonging to the same object share the same label. For example, in an image containing a person, a car, and a tree, semantic segmentation would label each pixel as belonging to the person, car, tree, or background (El Bsat et al., 2022).

Instance Segmentation is a particular type of segmentation. It differs from semantic segmentation by assigning a class label to each pixel and distinguishing between individual instances of a given class. In other words, it identifies and distinguishes various objects or instances of the same class. For example, in a photograph with numerous people, such segmentation would not just label all the pixels belonging to people but also assign a unique identifier to each person, separating them as individual instances (Almalki & Latecki, 2023).

Various approaches to image segmentation exist, ranging from traditional methods based on handcrafted features to advanced techniques leveraging deep learning models. Deep learning procedures in actual convolutional neural networks (CNNs) have recently gained popularity and demonstrated superior performance in image segmentation tasks. Models like U-Net, FCN, SegNet, and Mask R-CNN have become popular for semantic and instance segmentation responsibilities due to their aptitude to learn multifaceted spatial features and patterns from the data (Hou et al., 2023).

Despite the progress made with deep learning, image segmentation remains challenging, especially in complex scenes, occlusions, and varying lighting conditions. Researchers and developers continue to discover new techniques and architectures to improve image segmentation methods' precision, efficiency, and robustness for a wide range of uses (Wu et al., 2023).

Dental segmentation using deep learning is an advanced technique in the domain of medical image analysis, besides computer vision. The aim of dental segmentation is to recognize and delineate the boundaries of dental in an image, which can be suitable for numerous uses, including dental diagnosis, treatment planning, and orthodontic analysis. It is essential to note that dental image segmentation can be an inspiring task because of the differences in image quality, lighting, and occlusions. Consequently, having a robust and diverse dataset is crucial to training a successful deep-learning model. Additionally, in the medical domain, it is essential to ensure that the segmentation model is accurate and reliable before applying it to real-world applications (Lee & Kim, 2022).

As with any medical-related AI application, it's crucial to validate the model's performance with the help of dental professionals and conduct extensive testing and validation to ensure its accuracy and safety in clinical practice (Chandrashekar et al., 2022).

Related Works

Many previous research works have been based on dental segmentation, where many different algorithms have been used to obtain high accuracy. The most important of this research are:

(a) Im et al. (2022) assessed the effectiveness and accuracy of deep learning-based automatic teeth segmentation in digital dental models. Consuming 516 digital dental models, it created a method for autonomous teeth segmentation and classification based on a dynamic graph convolutional neural network (DGCNN). Thirty digital dental models were segmented using three different methods for comparison: (1) landmark-based tooth segmentation (LS) using OrthoAnalyzer software, (2) tooth designation and segmentation (DS) using Autolign software, and (3) automatic teeth segmentation (AS) utilizing the DGCNN-based algorithm from LaonSetup software. The clinical crown height (CCH), segmentation duration, mesiodistal (MD) width, and segmentation success rate were all assessed. A digital dental model's automatic tooth segmentation utilizing deep learning has a high segmentation achievement rate, precision and efficiency, making it suitable for orthodontic diagnosis and usage production.

(b) Tian et al. (2019) suggested a novel method for segmenting and classifying tooth kinds on 3D dental models based on 3D convolution neural networks (CNNs) and the sparse voxel octree. A dental classification approach constructed on two-level hierarchical features learning is first suggested to address the misclassification issue in extremely similar tooth classes. Second, individual teeth-gingiva and interteeth segments are segmented using an upgraded three-level hierarchical segmentation approach based on deep convolution features. The interteeth fusion zone and gingival margin borders are refined using the conditional random field model. The experiment results indicate that the Level_1 network has a classification accuracy of 95.96%, the Level_2 network has an average classification accuracy of 88.06%, and the tooth segmentation accuracy of 89.81%. The suggested method has more potential for application in computer-assisted orthodontic treatment diagnosis and is more widely applicable and exact than current state-of-the-art methodologies.

(c) Rashid et al. (2022) delivered a diverse sample for the dental (colored or X-ray) images and implemented a deep learning method to improve carious area localization. They also applied a comprehensive program that uses basic dental scans to automatically identify carious areas. In order to identify dental carious areas, the instantiation principally uses a pre-trained hybrid Mask RCNN in conjunction with a heterogeneous dataset of dental photos (colored photographs or X-rays) obtained from diverse causes. Dentists' examinations revealed that annotated datasets have an accuracy of up to 96%, while the precision of the suggested method ranges from 78% to 92%. Furthermore, the approach received more than 80% general acceptance among dentists.

(d) Fatima et al. (2023) suggested model has two parts: (1) for periapical disease localization on a limited dataset, a region-based network (RBN) and (2) a lightweight modified MobileNet-v2 backbone. The lightweight Mask-RCNN is assessed on a specially marked dataset that includes pictures of five distinct kinds of periapical wounds to gauge the efficacy of the suggested model. The results show that the model has a total precision of 94%, a mean average precision of 85%, and a mean inspection above a separation of 71.0% in identifying and identifying periapical lesions. Compared to current techniques, the suggested model greatly increases recognition and classification, and localization precision consumes fewer photos and outperforms state-of-the-art approaches.

(e) Zhao et al. (2020) proposed a Two-Stage Attention Segmentation Network (TSASNet) used for dental panoramic X-ray images to improve tooth boundary and root segmentation due to small differences and uneven intensity supply. First, a framework for attention in the initial stage and employing global and local attention modules will be employed to approximate the tooth placement. The attention model can automatically detect coarse tooth borders and gather contextual information pixel by pixel without an interaction operator. Second, they utilize a fully convolutional network to extract the real tooth area from the responsiveness maps gained in the first stage, obtaining more accurate final boundary information. The usefulness of TSASNet is demonstrated using a benchmark dataset of 1,500 dental panoramic X-ray images. Their suggested solution outperforms the state-of-the-art methods, achieving 96.94% precision, 92.72% dice, and 93.77% recall.

(f) Al Nassan et al. (2022) modified CNNs to segment the bitewing image in this investigation. Before segmenting, the bitewing radiographs are imported into MATLAB and improved to produce a binary mask image that removes the background from the original images. This image will be fed into the CNN model. Those masks are the aim of the deep learning model. We were able to attain the greatest accuracy on hidden images after training the provided method with 456 bitewing shots, 88.27% Fl-score and 97.3% accuracy.

(g) Sivagami et al. (2020) UNet architecture using convolutional neural networks was presented to accurately separate dental panoramic x-ray pictures, which radiographic pictures assist medical professionals in accurately identifying and diagnosing diseases. Radiographic pictures contain X-rays, computed tomography (CT), and magnetic resonance imaging (MRI). In general, X-ray images are complicated. Noise makes it more difficult to distinguish between different portions of the teeth. This makes the segmentation process extremely tough. Dental image segmentation assists dentists in detecting damaged teeth, determining the proper position for dental implant insertion, and determining the direction of the tooth structure. The UNet architectural model is a contemporary method for medical image segmentation. In this paper, they achieved a Dice score of 94% and an accuracy of 97% in dental X-ray picture segmentation using UNet architecture. Furthermore, a comparison is made between the performance of the UNet architecture and other image segmentation techniques for the segmentation of dental X-ray pictures.

Through previous research and deep learning algorithms, note that increasing segmentation accuracy depends on several criteria: extracting the largest number of features from the image, selecting the most important features, and deleting the rest. One of the most important of these algorithms is CNN, which is one of the most important deep learning algorithms. This algorithm is characterized by high accuracy in segmenting, detecting, and classifying images due to layers of extracting features from images and selecting the best features, which greatly affect the decision. Some research has added a technique with the CNN algorithm to increase the number of features extracted; the structure of the U-Net or Mask R-CNN algorithms was combined with the CNN architecture to increase accuracy by increasing the number of features. Note that the greater the amount of features extracted from the images and selecting the best among them leads to an increase in the accuracy of the segmentation. However, these layers may lead to slow execution due to their size compared to the entered images, and to obtain a proposed model. The accuracy of the segmentation and the execution time must be taken into account.

MATERIALS AND METHODS

The YOLO algorithm is considered one of the modern algorithms as it has shown its efficiency in many areas. For this reason, it was chosen as a proposal for dental segmentation. This research used and modified YOLOv8 of the algorithm to display its increased accuracy in dental segmenting. Modifications were made to the model's layers by increasing the cumulative number of backbone layers to enhance the number of features extracted from the image. Additionally, the sizes of the layers were modified to make the proposed model suitable for the process of the dental segmentation and isolating it from the rest of the mouth. Figure 1 illustrates the proposed YOLOv8 segmentation model.

The proposed model consists of 29 layers, and the specifications of each layer have been accurately determined to suit the process of dental segmentation. Table 1 illustrates the specification of the proposed YOLOv8 segmentation model.

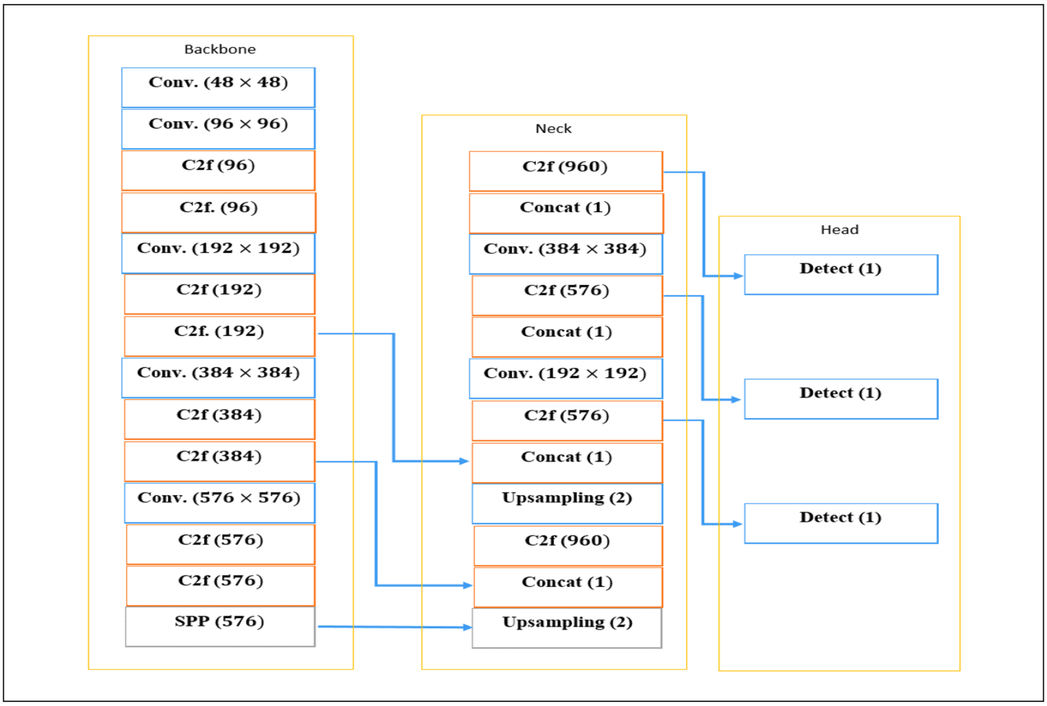

Figure 1. Proposed YOLOv8 segmentation model

Table 1
Specification of proposed YOLOv8 segmentation model

Layer No.	Part	Layer	Specification
1	Backbone	Conv.	[3, 48, 3, 2]
2		Conv.	[48, 96, 3, 2]
3		C2f	[96, 96, 2, True]
4		C2f	[96, 96, 2, True]
5		Conv.	[96, 192, 3, 2]
6		C2f	[192, 192, 4, True]
7		C2f	[192, 192, 4, True]
8		Conv.	[192, 384, 3, 2]
9		C2f	[384, 384, 4, True]
10		C2f	[384, 384, 4, True]
11		Conv.	[384, 576, 3, 2]
12		C2f	[576, 576, 2, True]
13		C2f	[576, 576, 2, True]
14		SPP	[576, 576, 5]

Table 1 *(continue)*

Layer No.	Part	Layer	Specification
15		Upsampling	[None, 2, 'nearest']
16		Concat	[1]
17		C2f	[960, 384, 2]
18		Upsampling	[None, 2, 'nearest']
19		Concat	[1]
20		C2f	[576, 192, 2]
21	Neck	Conv.	[192, 192, 3, 2]
22		Concat	[1]
23		C2f	[576, 384, 2]
24		Conv.	[384, 384, 3, 2]
25		Concat	[1]
26		C2f	[960, 576, 2]
27		Conv.	[1]
28	Head	Conv.	[1]
29		Conv.	[1]

The C2f layer is important in extracting features from images, significantly contributing to teeth identification. While the C2f layer is highly efficient in determining teeth, it can be improved from the mathematical side. In the proposed model to improve Adaptive Cross-Stage Partial Connections, this layer has been improved based on the introduction of adaptive weights, through which the grid is allowed to modify and integrate features more efficiently and improve the flow of the gradient through the next layers. Let x_i represent the output of the i^{th} layer, and w_i be an adaptive weight associated with the connection from layer *i*. The output of C2f is y can be as shown in Equation 1:

$$y = \sum x_i \times w_i \qquad [1]$$

Now, the summation of weight must be 1, and the SoftMax activation function must be used to ensure Equation 2.

$$w_i = SoftMax(z_i) \qquad [2]$$

where z_i is a learnable parameter for each connection.

Minimizing the loss function, which measures the discrepancy between the predictions made by the model, maximizes the model's performance by optimizing weights, which can be described in Equation 3:

$$L(w) = Loss\big(y(w)\big) \qquad [3]$$

where Loss is a suitable loss function, such as mean squared error or cross-entropy loss.

Gradient descent-based methods can be employed to optimize the weights. The gradient of the loss function for the weights can be calculated using the chain rule in Equation 4:

$$\frac{dL}{dw_i} = \frac{dL}{dy} \times \frac{dy}{dw_i} \quad [4]$$

Where dL/dy is the gradient of the loss function with output y, and dy/dw_i is the derivative of the output weight w_i.

This improvement allows the proposed model to focus more on dental features and more efficiently facilitate the spread of gradients through adaptive weights.

The process of building a dataset is of high importance for building an intelligent system. AI algorithms need a number of different types of images to obtain high accuracy in training. In this dissertation, the dataset (top view) was taken from a dental clinic where 100 images were taken of dental patients from different patients. The images taken from the dental clinic are in different sizes and were divided into two parts: (1) 70% for training and (2) 30% for testing. Figure 2 illustrates samples of the dataset. These images were taken through real cases in a dental clinic in Iraq.

Two main parts of preparing the dataset before the proposed system uses it are dental dataset augmentation and dental dataset annotation. Image augmentation was used to transform the original images to increase the size of the dental dataset. This helps improve the model's generalization and robustness by exposing it to a wider range of variations in the input data. Image augmentation is particularly useful when the available labeled dataset is limited (Figure 3). Here are image augmentation techniques used to increase dental dataset: Rotate the image by a certain angle, flip the image horizontally or vertically, zoom in or out of the image, apply shift transform to the left or right, apply shift transform to the up or down, and apply a shear transformation to the image.

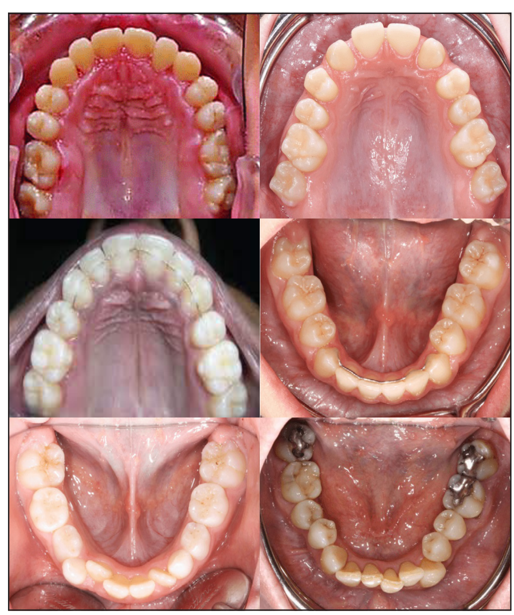

Figure 2. Samples of dataset

This step increased the number of images in the dataset from 100 to 600, which is highly suitable for training and testing the enhanced YOLOv8 algorithm and increasing its accuracy in dental segmentation.

The dental annotation step includes the work of image segmentation annotation and involves labeling individual regions within an image to delineate dental. This

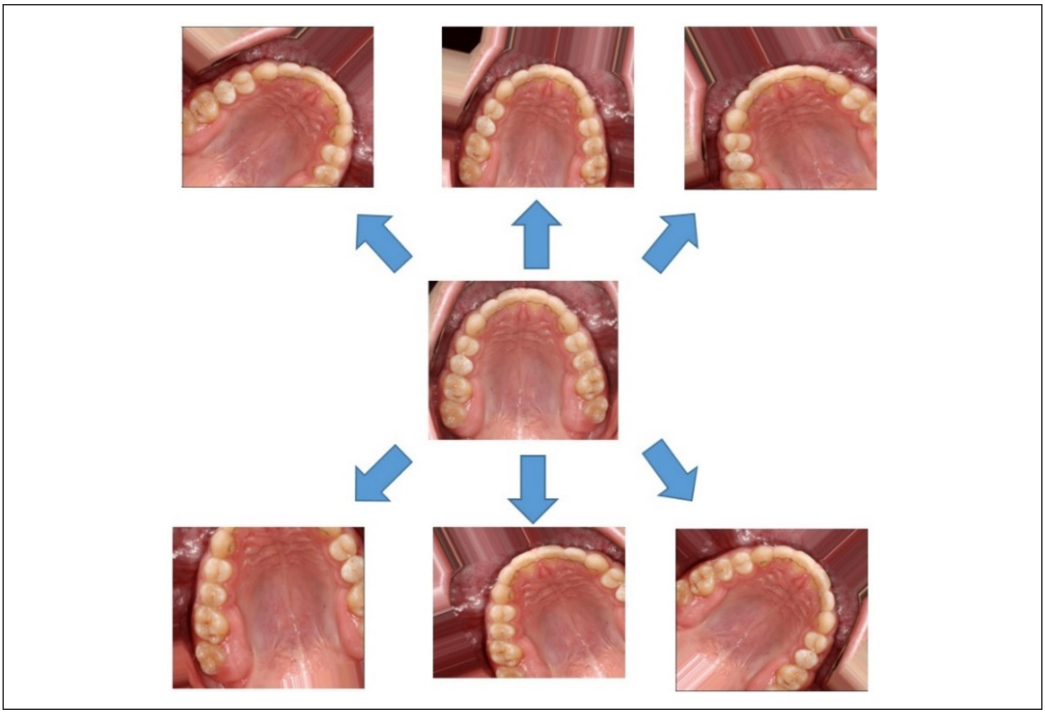

Figure 3. Dental dataset augmentation

annotation is commonly used in computer vision tasks, such as training models for semantic segmentation. Annotators draw polygons around the dental. Each polygon defines a segment of the image. The vertices of the polygon represent the boundary of the annotated dental. Masks are binary images where pixels inside the dental are labeled as foreground (1) and pixels outside as background (0). A LabelMe program was used to create dental annotation (Figure 4).

RESULTS AND DISCUSSION

The practical experiments conducted to train the proposed YOLOv8 segmentation model in dental segmentation reached high accuracy in segmenting dental from oral parts. Many experiments have been conducted to demonstrate the precision of the proposed model. The experiment's trained model uses an epoch of 50, with Figure 5 illustrating the performance measurement.

The train/box_loss scale of the lower value indicates better performance, as it represents the loss associated with the model's ability to accurately predict the boxes surrounding the detected teeth. Here, its value reached 0.38524 because the shapes of the teeth are not rectangular. For this reason, the determination will be the tooth and part of its surroundings. As for the train/seg_loss scale, it is related to the loss of dental segmentation, and note

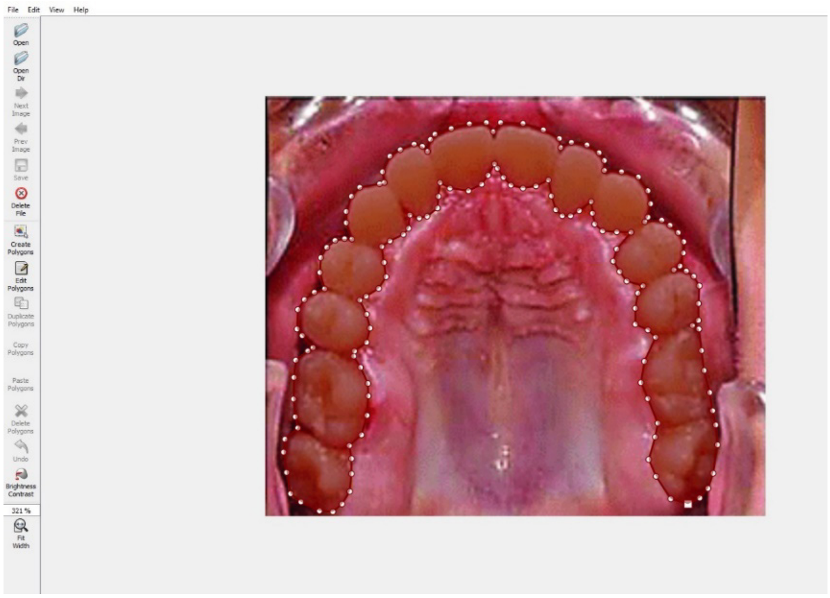

Figure 4. Dental annotation

that its value reached 0.35353 because it takes a determination of the tooth in a way that surrounds the tooth and does not take from the surrounding parts. Its value is less than the scale train/box_loss. Train/cls_loss scale represents the loss represented by the classification of the teeth, and we note that its value is high, 0.87865, because the model is for the segmentation of the teeth and not their classification. Train/dfl_loss measure is related to the process of augmentation of dental images, and in the absence of repetition, its value rises because the dental images are not duplicated, and its value reaches 1.0683. Everything mentioned is specific to the images of the model's training.

The following metrics are specific to validation and these scales are similar to the previous scales. However, they are applied to data on which the proposed model has not been trained, where val/box_loss is similar to train/box_loss, but val/box_loss represents the loss of the boxes that surround the images of teeth that the model has not been trained on. This applies to other measures where train/seg_loss , train/cls_loss, and train/dfl_loss like val/seg_loss , val/cls_loss, and val/dfl_loss. metrics/precision(B)، metrics/recall(B)، metrics/mAP50(B)، metrics/mAP50-95(B). These metrics evaluate the model's performance in tooth detection. The metrics/precision(B) scale reached 0.9761, representing the ratio of true tooth detection to the total number of expected teeth. The metrics/recall(B) reached 0.996, representing the ratio of true tooth detection to the total number of true teeth. As for metrics/mAP50(B), it reached 0.995, which is the average resolution at the IoU threshold by 50%. In addition to scale metrics/mAP50-95(B), the average accuracy calculated on IoU thresholds, ranging

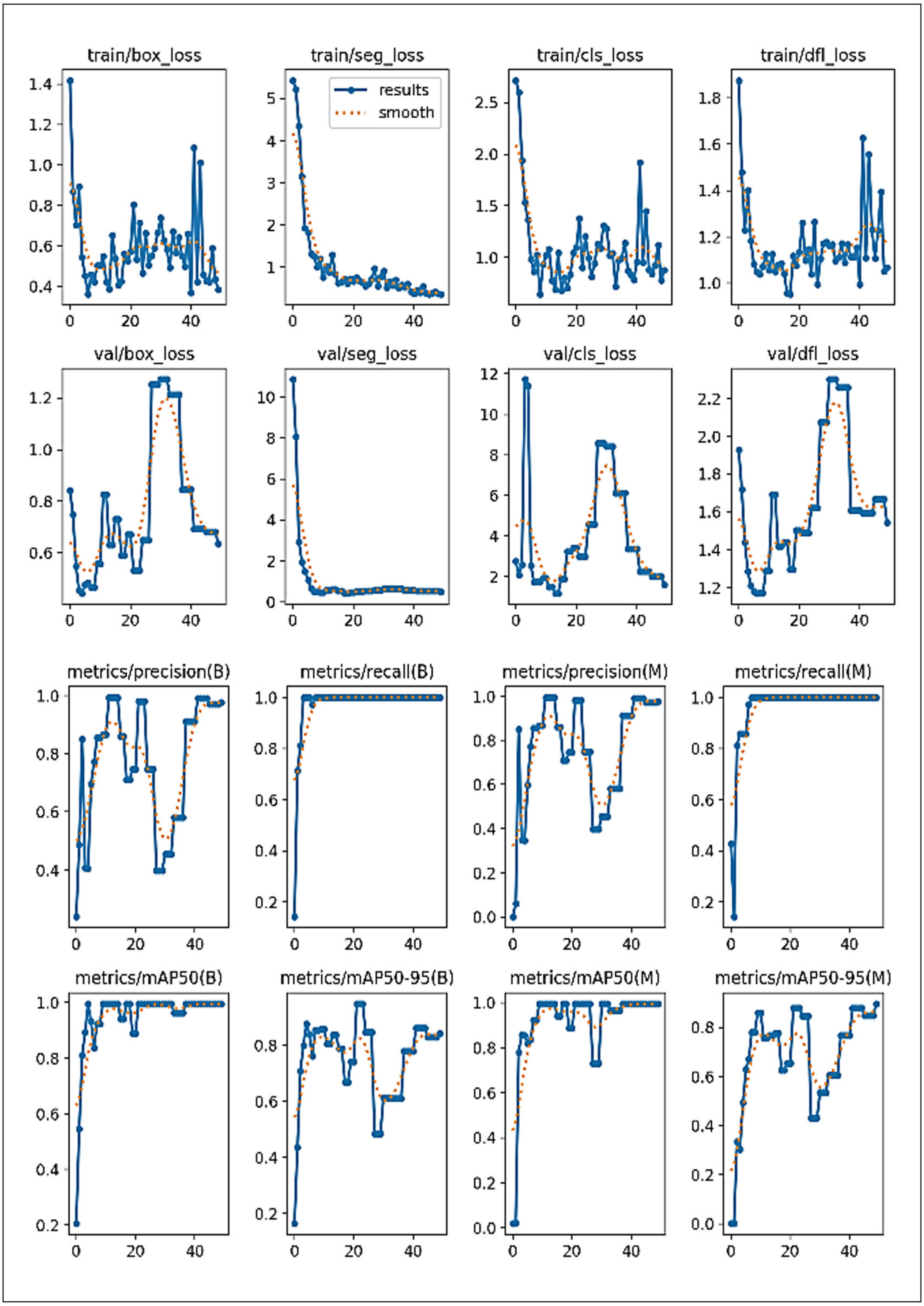

Figure 5. Illustrate performance measurement

from 50% to 95%, reached 0.84186. metrics/precision(M)، metrics/recall(M)، metrics/mAP50(M)، metrics/mAP50-95(M). These scales are like the detection scales, but these are for segmentation. For this reason, note that the values of the segmentation scales are higher. Table 2 illustrates the values of measurement.

Many images were tested to evaluate the proposed model, as the images were of different lighting and sizes. Figure 6 illustrates the mask detection results, and Figure 7 explains the segmentation results.

The proposed model for dental segmentation will be compared with some segmentation algorithms. Many algorithms can be referred to and get good results in dental segmentation. The dataset used in the

Table 2
Values of measurement

Measure	Value
train/box_loss	0.38524
train/seg_loss	0.35353
train/cls_loss	0.87865
train/dfl_loss	1.0683
metrics/precision(B)	0.9761
metrics/recall(B)	0.996
metrics/mAP50(B)	0.995
metrics/mAP50-95(B)	0.84186
metrics/precision(M)	0.9761
metrics/recall(M)	0.996
metrics/mAP50(M)	0.995
metrics/mAP50-95(M)	0.8955
val/box_loss	0.63554
val/seg_loss	0.48883
val/cls_loss	1.5975
val/dfl_loss	1.5415

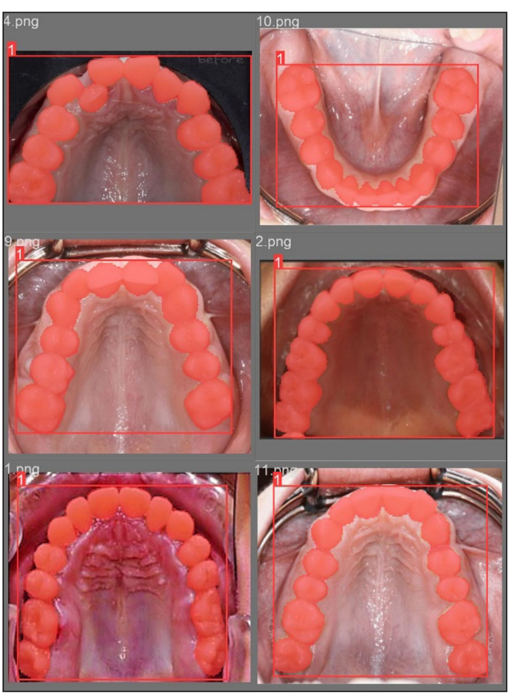

Figure 6. Show the mask detection results

Figure 7. Illustrate segmentation results using the YOLOv8 segmentation model

paper will be applied to compare its results with the results of the model proposed in this paper. The comparison results are summarized in Table 3.

Table 3 shows the difference between the previous algorithms compared to the enhanced model, where the proposed model outperformed the rest of the algorithms by comparing it in terms of accuracy in dental segmentation due to the equation that has been improved to reduce loss and reach the best features. In addition, it takes less time to segment the teeth because the algorithm of YOLOv8 has been reduced to complex equations, which take a long time to calculate, as in U-Net and R-CNN.

Table 3
Comparison with segmentation algorithms

Method	Accuracy	Precision	Recall	F1-Score	Time Per Frame
U-Net	0.8150	0.8247	0.8421	0.8333	100 ms
R-CNN	0.8667	0.8900	0.8725	0.8812	87 ms
ASF-YOLO	0.9512	0.9674	0.9468	0.9570	78 ms
YOLO2U-Net	0.9571	0.9674	0.9570	0.9622	76 ms
Our Work	0.9950	0.9761	0.9964	0.9861	66 ms

CONCLUSION

Dental isolation remains a critical challenge in modern dentistry, with various technological advancements aiding dentists in diagnosing dental conditions. Tooth segmentation is particularly challenging due to the similarity in characteristics among dental structures, necessitating the development of a fast and highly precise system. In this study, an enhanced YOLOv8-based model was proposed and optimized for dental segmentation, incorporating a gradient descent-based approach to refine feature extraction and improve segmentation accuracy. The proposed model achieved an unprecedented accuracy of approximately 99.6% when tested on real-world dental images. These results demonstrate the model's high efficiency and reliability in accurately isolating teeth. Given its superior performance, the enhanced YOLOv8 algorithm presents a promising automated dental analysis and segmentation solution. Future research can explore further refinements and adaptations of this model to improve its generalizability across diverse dental imaging datasets and clinical applications.

ACKNOWLEDGEMENT

The authors express their gratitude to the College of Computing, Informatics and Mathematics, the Institute of Postgraduate Studies (IPSis) and the Research Management Centre, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia, for supporting the publication of this paper.

REFERENCES

Al Nassan, W., Bonny, T., Obaideen, K., & Hammal, A. A. (2022). A customized convolutional neural network for dental bitewing images segmentation. In *2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA)* (pp. 347-351). IEEE. https://doi.org/10.1109/ICECTA57148.2022.9990564

Almalki, A. & Latecki, L. J. (2023). Self-supervised learning with masked image modeling for teeth numbering, detection of dental restorations, and instance segmentation in dental panoramic radiographs. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 5594-5603). IEEE. https://doi.org/10.1109/WACV56688.2023.00555

Bi, Y., Xue, B., Mesejo, P., Cagnoni, S., & Zhang, M. (2022). A survey on evolutionary computation for computer vision and image analysis: Past, present, and future trends. *IEEE Transactions on Evolutionary Computation, 27*(1), 5-25. https://doi.org/10.1109/TEVC.2022.3220747

Chandrashekar, G., AlQarni, S., Bumann, E. E., & Lee, Y. (2022). Collaborative deep learning model for tooth segmentation and identification using panoramic radiographs. *Computers in Biology and Medicine, 148,* 105829. https://doi.org/10.1016/j.compbiomed.2022.105829

El Bsat, A. R., Shammas, E., Asmar, D., Sakr, G. E., Zeno, K. G., Maxari, A. T., & Ghafari, J. (2022). Semantic segmentation of maxillary teeth and palatal Rugae in two-dimensional images. *Diagnostics, 12*(9), 2176. https://doi.org/10.3390/diagnostics12092176

Elyan, E., Vuttipittayamongkol, P., Johnston, P., Martin, K., McPherson, K., Moreno-Garcia, C. F., Jayne, C., & Sarker, M. M. K. (2022). Computer vision and machine learning for medical image analysis: Recent advances, challenges, and way forward. *Artificial Intelligence Surgery, 2*, 24-45. https://doi.org/10.20517/ais.2021.15

Fatima, A., Shafi, I., Afzal, H., Mahmood, K., Diez, I. D. L. T., Lipari, V., Ballester, J. B., & Ashraf, I. (2023). Deep learning-based multiclass instance segmentation for dental lesion detection. *Healthcare 2023, 11,* 347. https://doi.org/10.3390/healthcare11030347

Hou, S., Zhou, T., Liu, Y., Dang, P., Lu, H., & Shi, H. (2023). Teeth U-Net: A segmentation model of dental panoramic X-ray images for context semantics and contrast enhancement. *Computers in Biology and Medicine, 152,* 106296. https://doi.org/10.1016/j.compbiomed.2022.106296

Im, J., Kim, J. Y., Yu, H. -S., Lee, K. -J., Choi, S. -H., Kim, J.-H., Ahn, H-K., & Cha, J. -Y. (2022). Accuracy and efficiency of automatic tooth segmentation in digital dental models using deep learning. *Scientific Reports, 12*(1), 9429. https://doi.org/10.1038/s41598-022-13595-2

Lee, S., & Kim, J. E. (2022). Evaluating the precision of automatic segmentation of teeth, gingiva and facial landmarks for 2d digital smile design using real-time instance segmentation network. *Journal of Clinical Medicine, 11*(3), 852. https://doi.org/10.3390/jcm11030852

Rashid, U., Javid, A., Khan, A. R., Liu, L., Ahmed, A., Khalid, O., Saleem, K., Meraj, S., Iqbal, U., & Nawaz, R. (2022). A hybrid mask RCNN-based tool to localize dental cavities from real-time mixed photographic images. *PeerJ Computer Science, 8*, e888. https://doi.org/10.7717/peerj-cs.888

Sivagami, S., Chitra, P., Kailash, G. S. R., & Muralidharan, S. R. (2020). UNet architecture based dental panoramic image segmentation. In *2020 International Conference on Wireless Communications*

Signal Processing and Networking (WiSPNET) (pp. 187-191). IEEE. https://doi.org/10.1109/WiSPNET48689.2020.9198370

Tian, S., Dai, N., Zhang, B., Yuan, F., Yu, Q., & Cheng, X. (2019). Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. *IEEE Access, 7*, 84817-84828. https://doi.org/10.1109/ACCESS.2019.2924262

Wu, Y., Li, Z., Wang, J., He, M., Zhou, W., & Lin, P. (2023). Application study on the recognition of oral obstructed tooth images using semantic segmentation. *Second International Conference on Biomedical and Intelligent Systems (IC-BIS 2023)* (Vol. 12724, pp. 610-614). SPIE. https://doi.org/10.1117/12.2687389

Zhao, Y., Li, P., Gao, C., Liu, Y., Chen, Q., Yang, F., & Meng, D. (2020). TSASNet: Tooth segmentation on dental panoramic x-ray images by two-stage attention segmentation network. *Knowledge-Based Systems, 206,* 106338. https://doi.org/10.1016/j.knosys.2020.106338