Scientific Research Publishing

Open Journal of Stomatology, 2024, 14, 316-326
https://www.scirp.org/journal/ojst
ISSN Online: 2160-8717
ISSN Print: 2160-8709

Single Tooth Segmentation on Panoramic X-Rays Using End-to-End Deep Neural Networks

Yu Sun[1,2], Jing Feng[1,2*], Huang Du[1], Juan Liu[1,2], Baochuan Pang[2], Cheng Li[2], Jinxian Li[2], Dehua Cao[2]

[1]Institute of Artificial Intelligence, School of Computer Science, Wuhan University, Wuhan, China
[2]Landing Artificial Intelligence Center for Pathological Diagnosis, Wuhan University, Wuhan, China
Email: *gfeng@whu.edu.cn

How to cite this paper: Sun, Y., Feng, J., Du, H., Liu, J., Pang, B.C., Li, C., Li, J.X. and Cao, D.H. (2024) Single Tooth Segmentation on Panoramic X-Rays Using End-to-End Deep Neural Networks. *Open Journal of Stomatology*, **14**, 316-326.
https://doi.org/10.4236/ojst.2024.146025

Received: May 15, 2024
Accepted: June 16, 2024
Published: June 19, 2024

Abstract

In dentistry, panoramic X-ray images are extensively used by dentists for tooth structure analysis and disease diagnosis. However, the manual analysis of these images is time-consuming and prone to misdiagnosis or overlooked. While deep learning techniques have been employed to segment teeth in panoramic X-ray images, accurate segmentation of individual teeth remains an underexplored area. In this study, we propose an end-to-end deep learning method that effectively addresses this challenge by employing an improved combinatorial loss function to separate the boundaries of adjacent teeth, enabling precise segmentation of individual teeth in panoramic X-ray images. We validate the feasibility of our approach using a challenging dataset. By training our segmentation network on 115 panoramic X-ray images, we achieve an intersection over union (IoU) of 86.56% for tooth segmentation and an accuracy of 65.52% in tooth counting on 87 test set images. Experimental results demonstrate the significant improvement of our proposed method in single tooth segmentation compared to existing methods.

Keywords

Single Tooth Segmentation, Teeth Counting, Panoramic X-Ray, Combinatorial Loss

1. Introduction

To diagnose oral diseases, dentists rely on X-ray images to analyze the structure of individual teeth. Panoramic X-ray images provide visualization of tooth structure and shape, playing a crucial role in dental diagnosis. However, the analysis of panoramic X-ray images presents significant challenges due to various factors such as image resolution, contrast brightness, noise, and the presence of dental restorations. Moreover, adjacent teeth can share overlapping regions, further complicating the analysis process. Manual analysis of X-ray images is time-consuming and can result in misdiagnosis or missed diagnoses [1]. With the advancements in deep learning [2], deep learning-based methods have shown promising results in various medical tasks [3]. Automatic segmentation of individual teeth in panoramic radiographs can aid dentists in diagnosis, reduce the time required for diagnosis, and lower medical costs, making it a critical task in panoramic X-ray analysis.

Traditional methods in dental segmentation primarily include region-based [4], threshold-based [5], clustering-based [6], boundary-based [7], and watershed-based approaches [8]. After the success of deep learning methods in major fields, neural networks are applied to various image segmentations. Considering the characteristics of X-ray images, convolutional neural networks may give higher accuracy, reliability and save diagnostic time [9]. Silva *et al.* [10] concluded that the results obtained using neural networks are better than traditional methods. Chen *et al.* [11] proposed a multiscale location-aware network and used MS-SSIM + dice loss + cross entropy (CE) with a combined loss to enhance the segmentation of tooth root boundaries. Nishitani *et al.* [12] improved the segmentation accuracy of tooth edges using a loss function weighted to the tooth edges by adding the CE of the tooth edges to the CE of the whole image. Koch *et al.* [13] used data augmentation, network integration, test time augmentation and bootstrapping strategies to improve the segmentation performance of U-Net for panoramic X-ray images.

Some studies have focused on instance segmentation of teeth. Jader *et al.* [14] used Mask R-CNN [15] for tooth instance segmentation. Silva *et al.* [16] used four different networks, Mask R-CNN, PANet [17], HTC [18], ResNeSt [19] for instance segmentation, detection and numbering of teeth, respectively. PANet achieved the best results. However, these methods impose strict requirements on tooth numbering and labeling, necessitate large datasets with high-quality annotations, and suffer from sample imbalance issues when datasets are limited, affecting the model's performance, especially when teeth overlap significantly.

Current semantic segmentation methods often yield teeth masks that are adhered to each other, influenced by the degree of tooth overlap in the original image. This poses challenges for subsequent operations on individual teeth, such as tooth counting. Helli *et al.* [20] employed morphological methods, including erosion, sharpening, filtering, and contour area thresholding, to separate adhered teeth and estimate the number of teeth based on the count of connected components. However, this approach requires manual adjustment of post-processing operations for different x-ray images, resulting in a tedious process. Moreover, it fails to separate teeth with severe overlap. For incisors with relatively small areas, the post-processing may shrink the area, making it difficult to determine an appropriate threshold.

In this work, we propose a novel approach to encourage the model to learn tooth boundary features using a combinatorial loss function. This approach allows for the separation of tooth contours in the original image even when teeth are closely spaced, unlike existing semantic segmentation methods that output masks with connected tooth boundaries. This facilitates easier subsequent processing of individual teeth and significantly improves tooth counting accuracy.

2. Method

2.1. Model Architecture

We adopt the U-Net [21] model as our baseline architecture, which is widely used in medical image segmentation after being proposed in 2015 because it can achieve very good segmentation results on small datasets. U-Net uses a classical encoder-decoder architecture, where each stage of the encoder consists of two consecutive convolution blocks, each consisting of a 3 × 3 convolution, batch normalization, and ReLU activation. The decoder stages involve upsampling and two convolution blocks to recover the original image resolution while incorporating semantic and positional information through skip connections. The final layer employs a 1 × 1 convolution operation. The overall model architecture is shown in **Figure 1**. Attention U-Net [22] adds attention gates to the U-Net model, so that the model learns to suppress irrelevant regions and focus on useful features during training. The attention gates add only a small amount of computation and can be easily integrated into other models. The decoder's feature

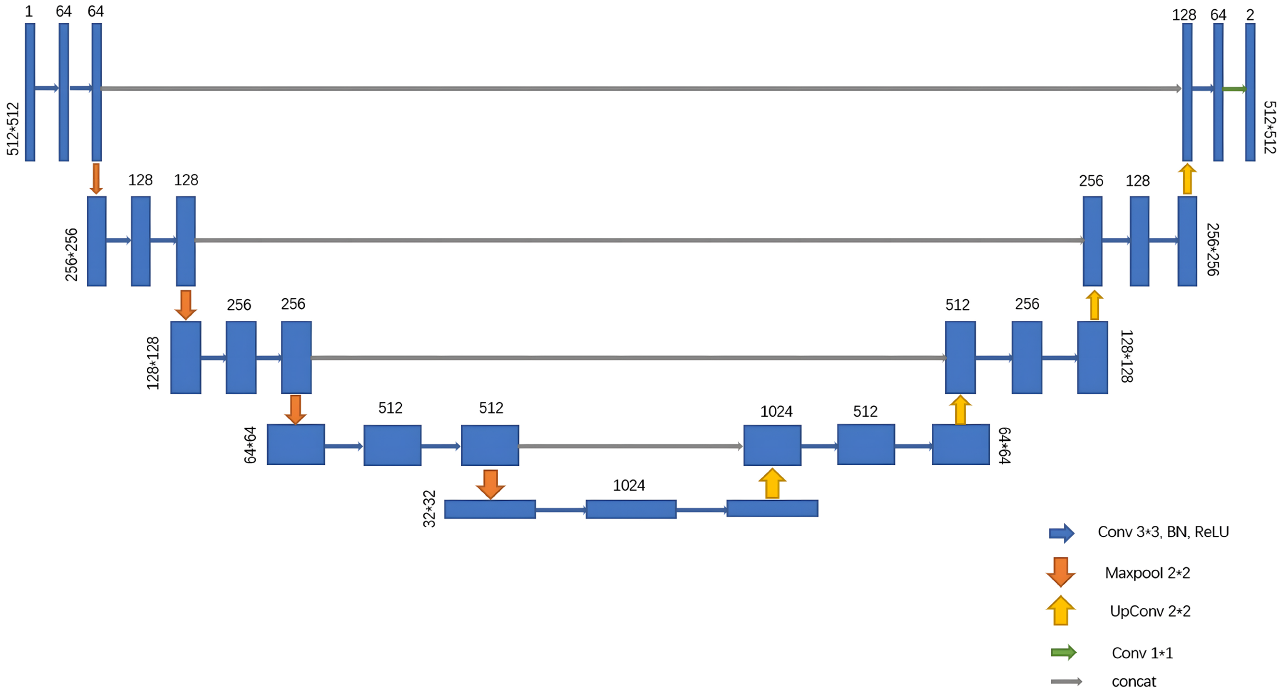
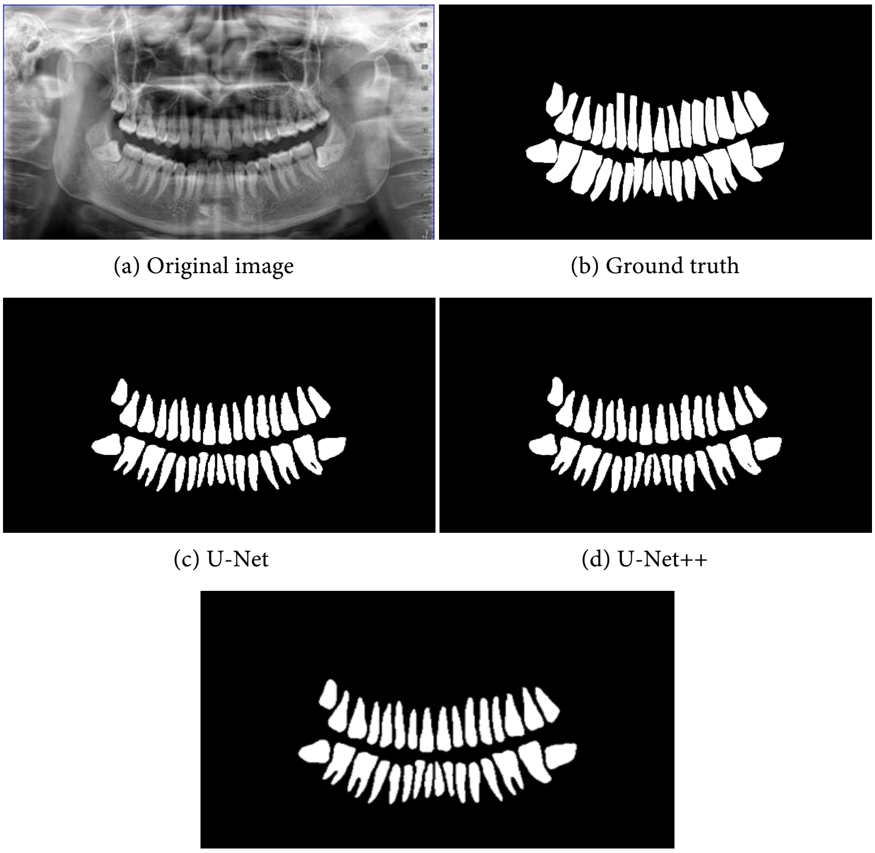


Figure 1. The U-Net model architecture used in this study. The number of channels is denoted on top of the feature map (blue boxes). The shape is denoted on the edge of the feature map.

map and its upper encoder's feature map are fed into the attention gates, and the output results are stitched with the decoder's upsampling results and then fed into the decoder's convolution block to increase the sensitivity of the model to foreground pixels. The U-Net++ [23] network adds a Dense-like structure to the U-Net skip connection and incorporates features from the next stage of convolution, applying this strategy at each stage to reduce the semantic differences of skip connections. In addition, U-Net++ uses a pruning strategy that can balance the accuracy and speed of segmentation. The segmentation results can be seen in **Figure 2**.

2.2. Proposed Hybrid Loss

We summarize the advantages of the methods proposed in previous studies and design a new combined loss for segmenting tooth boundaries. Traditionally, tooth semantic segmentation employs a 1 × 1 convolution in the final layer, followed by sigmoid activation and binary cross-entropy loss calculation. However, this approach results in masks that tightly adhere to each tooth. Inspired by BCNet [24], they modeled the region of interest (ROI) as two overlapping layers, with the top layer detecting the occluded object and the bottom layer detecting the occluded object. But this method obviously increases the complexity of the



(a) Original image

(b) Ground truth

(c) U-Net

(d) U-Net++

(e) Attention U-Net

Figure 2. Segmentation results of different networks.

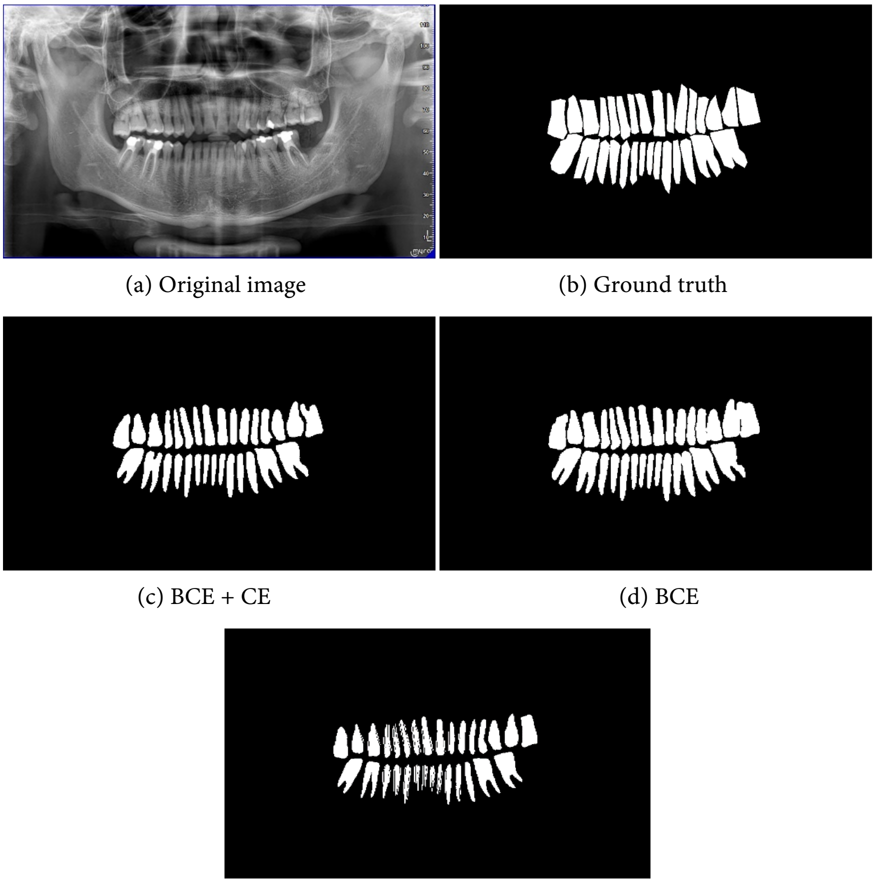
model. Considering that our dataset is relatively small, increasing the number of parameters too much may make the model less effective. Therefore, we modify the model's output to consist of two channels and compute the cross-entropy loss (CE) with the ground truth. This modification aims to increase the model's sensitivity to tooth boundary pixels, as shown in **Figure 3**. Considering the potential loss of information during image resizing, we convert the labels into a two-channel tensor using one-hot encoding and calculate binary cross-entropy loss (BCE) with the model's output. This step encourages the model to distinguish teeth from background pixels and incorporates it into the loss computation, enabling the model to differentiate between boundary pixels of adjacent teeth and teeth from the background. Consequently, the model's ability to accurately segment individual teeth in panoramic X-ray images is improved. The final loss function is shown as follows.

$$\text{loss} = \lambda_1 * \text{CE} + \lambda_2 * \text{BCE}$$

3. Experiments and Results

3.1. Dataset

We used a challenging dataset consisting of 202 panoramic X-ray radiographs


(a) Original image

(b) Ground truth

(c) BCE + CE

(d) BCE

(e) CE

Figure 3. Segmentation results of single loss function and proposed loss function.

with teeth labeled by experts. The images had a resolution of 1118 × 606 and included various categories such as dentures, missing teeth, and restorations. The dataset contained a substantial number of overlapping pixels between adjacent teeth. We divided the dataset into a training set of 115 images, employing 5-fold cross-validation, and a test set of 87 images. To enhance the model's generalization ability, we applied various data augmentation techniques such as random cropping, affine transformations, blurring, Gaussian noise, random contrast and brightness variations, horizontal flipping, and vertical flipping.

3.2. Training and Evaluation Detail

We employed three network architectures: 5-layer U-Net, Attention U-Net, and U-Net++. These architectures were used for semantic segmentation of teeth, with 64 feature maps in the top layer and 1024 feature maps in the bottom layer. We use the Adam optimizer with an initial learning rate of 0.0001. When the validation loss did not decrease for 5 consecutive epochs, we reduced the learning rate to 0.2 times its original value. And the following metrics are used to evaluate the performance of the model, where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 score} = \frac{2 * \text{recall} * \text{precision}}{\text{recall} + \text{precision}}$$

$$\text{iou} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}$$

In addition, the performance is further evaluated using the number of teeth accuracy and the average number of teeth error. T denotes the total number of images with correct tooth count output, ALL denotes the number of all images in the test set, and TS denotes the sum of tooth count errors for each image. The number of teeth is calculated by first binarizing the mask output from the model, converting it into a grayscale image to obtain the connected domain component, and finally calculating the number of contours with a connected domain area greater than 500. To compare the performance of our proposed method, the method proposed by Helli [20] was used. We modified the number of iterations for the morphological open operation to 2 and the number of iterations for the erosion to 1, and also set the area threshold to 500 to compare the accuracy of the number of teeth obtained by the two methods.

$$\text{teeth accuracy} = \frac{\text{T}}{\text{ALL}}$$

$$\text{teeth MAE} = \frac{\text{TS}}{\text{ALL}}$$

The experimental results, shown in Table 1, show that the IoU results of the three networks are almost the same, while the U-Net network has a higher accuracy in the number of teeth, a smaller average error in the number of teeth, and better segmentation results for a single tooth than the other two networks.

3.3. Ablation Study

To verify the effectiveness of our proposed method, we perform ablation study on the proposed loss function. We evaluate the performance of the model using only CE and BCE when the output channel is 2. The experiment result, as shown in Table 2, shows that using CE alone loses precision which is expected because a part of the pixels of the tooth boundaries are classified as background. But it will over-segment each tooth boundary while using BCE. The result shown in Figure 3 shows that the teeth in the output mask are adhered to each other. By combining these two loss functions, compared to using BCE and morphology method, there is a significant increase in the number of teeth accuracy on the test dataset. We set several different sets of parameters to explore their effects on the experimental results, as shown in Table 3. We found that increasing the penalty of BCE increases IoU, but the separation of single teeth becomes worse. Conversely, increasing the penalty of CE decreased IoU but did not improve the number accuracy of teeth. This observation was likely due to the high number of overlapping pixels between adjacent teeth in our dataset, making it challenging to completely separate overlapping teeth.

Additionally, we validated our method on a publicly available dataset [25] containing panoramic oral X-ray images of 116 patients, which were anonymized and labeled by experts. We used the first 96 images for training and validation

Table 1. Evaluation results of different networks.

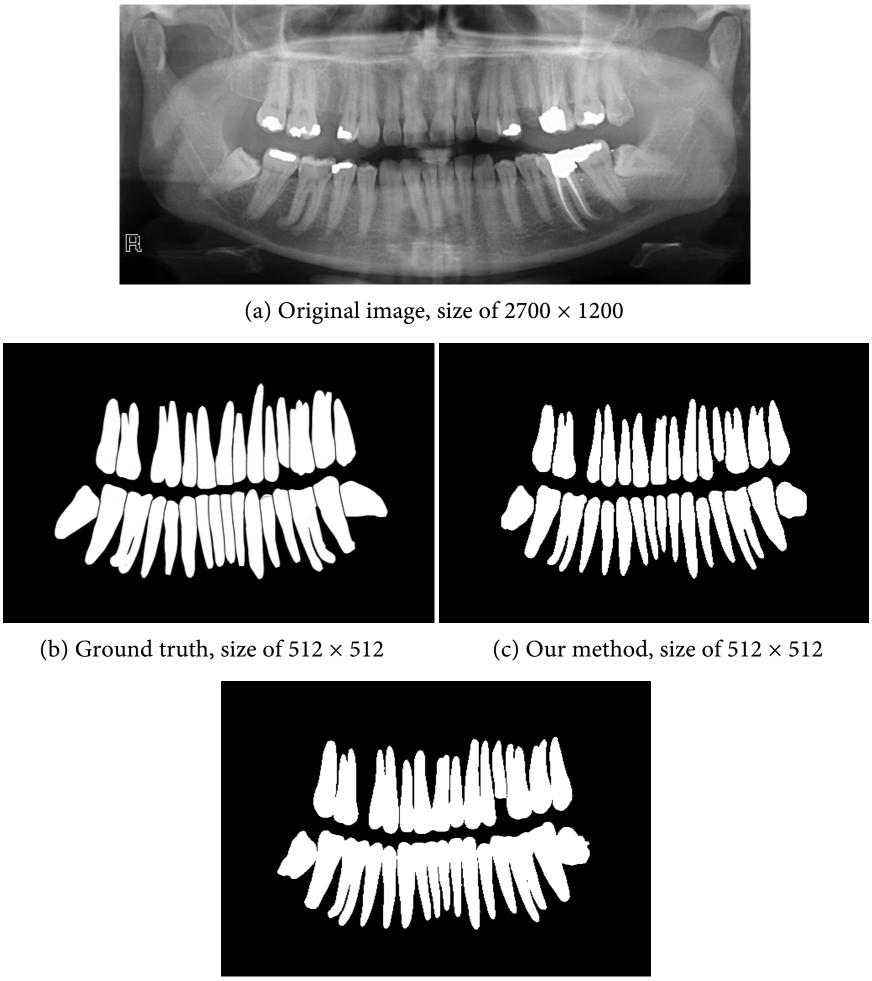
	Accuracy	Precision	Recall	F1 score	IoU	Teeth accuracy	Teeth MAE
U-Net	0.9695	0.8896	0.9707	0.9283	0.8656	0.6552	0.51
U-Net++	0.9684	0.8844	0.9710	0.9256	0.8605	0.5517	0.75
Attention U-Net	0.9693	0.8883	0.9707	0.9276	0.8643	0.6207	0.57

Table 2. Evaluation results of ablation study.

	IoU	Teeth accuracy	Teeth MAE
BCE + CE	0.8656	0.6552	0.51
BCE	0.9006	0.01	5.40
CE	0.7864	0.4483	1.07
BCE + morphology		0.1149	3.59
BCE + CE	0.8656	0.6552	0.51

Table 3. Comparison of the ratio of two loss functions.

	IoU	Teeth accuracy	Teeth MAE
$\lambda_1 = 2, \lambda_2 = 1$	0.8511	0.6092	0.54
$\lambda_1 = 1.5, \lambda_2 = 1$	0.8633	0.6552	0.47
$\lambda_1 = 1, \lambda_2 = 1$	0.8656	0.6552	0.51
$\lambda_1 = 1, \lambda_2 = 1.5$	0.8797	0.5057	0.99
$\lambda_1 = 1, \lambda_2 = 2$	0.8801	0.4253	1.17



(a) Original image, size of 2700 × 1200

(b) Ground truth, size of 512 × 512

(c) Our method, size of 512 × 512

(d) BCE, size of 512 × 512

Figure 4. Segmentation results on a publicly available dataset.

and the remaining 20 images for testing. The experimental results, as shown in Figure 4, demonstrated the effectiveness of our proposed method in segmenting individual teeth to a certain extent.

4. Conclusion and Discussion

The results show that semantic segmentation of single teeth with end-to-end deep neural networks is feasible. Among the tested network architectures, U-Net achieved the best results, with a segmentation IoU of 86.56% and an accuracy of 65.52% in terms of the number of teeth. Importantly, our proposed method is easily transferable to other models and suitable for end-to-end semantic segmentation of adherent objects. However, there are still opportunities for further optimization and improvement in our work. The performance of segmentation can be enhanced through additional research and development. Future work in this area will focus on segmenting and numbering instances of teeth in panoramic dental X-rays, as well as detecting abnormalities in the oral cavity. This advancement will enable the automatic generation of diagnostic reports and contribute to reducing diagnostic time.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Chen, H., Zhang, K., Lyu, P., Li, H., Zhang, L., Wu, J., *et al.* (2019) A Deep Learning Approach to Automatic Teeth Detection and Numbering Based on Object Detection in Dental Periapical Films. *Scientific Reports*, **9**, Article No. 3840. https://doi.org/10.1038/s41598-019-40414-y

[2] Shen, D., Wu, G. and Suk, H. (2017) Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, **19**, 221-248. https://doi.org/10.1146/annurev-bioeng-071516-044442

[3] Liu, L., Xu, J., Huan, Y., Zou, Z., Yeh, S. and Zheng, L. (2020) A Smart Dental Health-Iot Platform Based on Intelligent Hardware, Deep Learning, and Mobile Terminal. *IEEE Journal of Biomedical and Health Informatics*, **24**, 898-906. https://doi.org/10.1109/jbhi.2019.2919916

[4] Lurie, A., Tosoni, G.M., Tsimikas, J. and Walker, F. (2012) Recursive Hierarchic Segmentation Analysis of Bone Mineral Density Changes on Digital Panoramic Images. *Oral Surgery*, *Oral Medicine*, *Oral Pathology and Oral Radiology*, **113**, 549-558.e1. https://doi.org/10.1016/j.oooo.2011.10.002

[5] Indraswari, R., Arifin, A.Z., Navastara, D.A. and Jawas, N. (2015). Teeth Segmentation on Dental Panoramic Radiographs Using Decimation-Free Directional Filter Bank Thresholding and Multistage Adaptive Thresholding. 2015 *International Conference on Information & Communication Technology and Systems* (*ICTS*), Surabaya, 16 September 2015, 49-54. https://doi.org/10.1109/icts.2015.7379870

[6] Alsmadi, M.K. (2018) A Hybrid Fuzzy C-Means and Neutrosophic for Jaw Lesions Segmentation. *Ain Shams Engineering Journal*, **9**, 697-706. https://doi.org/10.1016/j.asej.2016.03.016

[7] Ali, R.B., Ejbali, R. and Zaied, M. (2015) GPU-Based Segmentation of Dental X-Ray Images Using Active Contours without Edges. 15*th International Conference on Intelligent Systems Design and Applications*, Marrakech, 14-16 December 2015, 505-510.

[8] Li, H., Sun, G., Sun, H. and Liu, W. (2012) Watershed Algorithm Based on Morphology for Dental X-Ray Images Segmentation. *IEEE* 11*th International Conference on Signal Processing*, Vol. 2, 877-880.

[9] Krois, J., Ekert, T., Meinhold, L., Golla, T., Kharbot, B., Wittemeier, A., *et al.* (2019) Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Scientific Reports*, **9**, Article No. 8495. https://doi.org/10.1038/s41598-019-44839-3

[10] Silva, G., Oliveira, L. and Pithon, M. (2018) Automatic Segmenting Teeth in X-Ray Images: Trends, a Novel Data Set, Benchmarking and Future Perspectives. *Expert Systems with Applications*, **107**, 15-31. https://doi.org/10.1016/j.eswa.2018.04.001

[11] Chen, Q., Zhao, Y., Liu, Y., Sun, Y., Yang, C., Li, P., *et al.* (2021) Mslpnet: Multi-Scale Location Perception Network for Dental Panoramic X-Ray Image Segmentation. *Neural Computing and Applications*, **33**, 10277-10291. https://doi.org/10.1007/s00521-021-05790-5

[12] Nishitani, Y., Nakayama, R., Hayashi, D., Hizukuri, A. and Murata, K. (2021) Segmentation of Teeth in Panoramic Dental X-Ray Images Using U-Net with a Loss Function Weighted on the Tooth Edge. *Radiological Physics and Technology*, **14**, 64-69. https://doi.org/10.1007/s12194-020-00603-1

[13] Koch, T.L., Perslev, M., Igel, C. and Brandt, S.S. (2019) Accurate Segmentation of Dental Panoramic Radiographs with U-Nets. 2019 *IEEE* 16*th International Symposium on Biomedical Imaging* (*ISBI* 2019), Venice, 8-11 April 2019, 15-19. https://doi.org/10.1109/isbi.2019.8759563

[14] Jader, G., Fontineli, J., Ruiz, M., Abdalla, K., Pithon, M. and Oliveira, L. (2018) Deep Instance Segmentation of Teeth in Panoramic X-Ray Images. 2018 31*st SIBGRAPI Conference on Graphics*, *Patterns and Images* (*SIBGRAPI*), Paraná, 29 October-1 November 2018, 400-407. https://doi.org/10.1109/sibgrapi.2018.00058

[15] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2961-2969. https://doi.org/10.1109/iccv.2017.322

[16] Silva, B., Pinheiro, L., Oliveira, L. and Pithon, M. (2020) A Study on Tooth Segmentation and Numbering Using End-to-End Deep Neural Networks. 33*rd SIBGRAPI Conference on Graphics*, *Patterns and Images*, Recife/Porto de Galinhas, 7-10 November 2020, 164-171.

[17] Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J. (2018) Path Aggregation Network for Instance Segmentation. 2018 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8759-8768. https://doi.org/10.1109/cvpr.2018.00913

[18] Chen, K., Ouyang, W., Loy, C.C., Lin, D., Pang, J., Wang, J., *et al.* (2019) Hybrid Task Cascade for Instance Segmentation. 2019 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 4974-4983. https://doi.org/10.1109/cvpr.2019.00511

[19] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., *et al.* (2022) Resnest: Split-Attention Networks. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), New Orleans, 18-24 June 2022, 2736-2746. https://doi.org/10.1109/cvprw56347.2022.00309

[20] Helli, S. and Hamamci, A. (2022) Tooth Instance Segmentation on Panoramic Dental Radiographs Using U-Nets and Morphological Processing. *Düzce University Journal of Science & Technology*, **10**, 39-50.

[21] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015 *Medical Image Computing and Computer-Assisted Intervention*, Munich, 5-9 October 2015, 234-241.

[22] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y. and Kainz, B. (2018) Attention U-Net: Learning Where to Look for the Pancreas. 1*st Conference on Medical Imaging with Deep Learning*, Amsterdam, 4-6 July 2018, 1-10.

[23] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J. (2018) Unet++: A

Nested U-Net Architecture for Medical Image Segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*: 4*th International Workshop*, *DLMIA* 2018, *and* 8*th International Workshop*, *ML*-*CDS* 2018, Vol. 4, 3-11.

[24] Ke, L., Tai, Y. and Tang, C. (2021) Deep Occlusion-Aware Instance Segmentation with Overlapping Bilayers. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 4019-4028. https://doi.org/10.1109/cvpr46437.2021.00401

[25] Abdi, A.H., Kasaei, S. and Mehdizadeh, M. (2015) Automatic Segmentation of Mandible in Panoramic X-Ray. *Journal of Medical Imaging*, **2**, Article ID: 044003. https://doi.org/10.1117/1.jmi.2.4.044003