Automatic Teeth Segmentation From Panoramic X-ray Images Using Deep Learning Models

Shuaa S. Alharbi
Department of Information Technology
College of Computer, Qassim University
Buraydah, Saudi Arabia
Email: shuaa.s.alharbi@qu.edu.sa

***Abstract*—A dentist's primary objective when screening for X-ray problems is to determine the shape, number, and position of teeth. Computational tools have been proposed to aid specialists in making more accurate diagnoses rather than relying solely on the trained eyes of dentists. Teeth segmentation and object detection are the core functions of these tools when applied to X-ray images. Segmenting and detecting the teeth in images is actually the first step in enabling other automatic processing methods. Medical image segmentation, especially in dentistry field, has been transformed by Deep Learning (DL) in recent years. U-Net with its different extensions and modifications has been among the most popular deep networks developed for medical image segmentation. However, it is difficult to determine which one will work best for teeth segmentation. In this study, different semantic segmentation models are selected based on their common use in medical image segmentation. Models include:U-Net++, ResU-Net++ and MultiResU-Net. Using panoramic X-ray dataset, MultiResUNet architecture performed better than the other segmentation models with an accuracy of 97.16%.**

Keywords—Convolutional Neural Networks, Deep learning, Deep Neural Networks, Image Segmentation, Medical Image Processing, Semantic Segmentation.

I. INTRODUCTION

Artificial Intelligence (AI) is becoming increasingly popular and widely used in medical care. The application of AI in the field of dentistry has shown successful results in the dental clinic examining routine [1]. Thus, AI can be used in dentistry to detect and recognize different variables from images, such as segmenting teeth from other tissues. Even though the use of artificial intelligence has grown rapidly and widely in the health care field, its application in dental care has been relatively slow [2].

In recent years, there has been increasing interest in applying the Deep Learning (DL) models for medical image analysis. The deep learning, typically the Convolutional Neural Network (CNN, or ConvNet) has made a significant contribution to the medical images analysing tasks especially the segmentation. Semantic segmentation methods based on DL have demonstrated state-of-the-art performance over the past few years. It has been demonstrated that these techniques have been successful in classifying, segmenting, and detecting medical images. For these applications, the U-Net [3] deep learning technique has become very popular. The U-Net shape with its variations and extensions (*i.e.* U-net++ [4], Resunet++ [5]) has long been recognized as the dominant deep network architecture. In this regard, it is the most widely used architecture in the medical imaging segmentation field.

In computer-assisted procedures topically aim to applied in dental clinics, teeth segmentation is an essential step. By using this technique, it is possible to provide approximate outline images of doubtful regions in order to provide features that can distinguish tooth tissues from other types of tissues.

In this paper, we demonstrate the use of U-net shapes to improve the performance of automatic teeth segmentation from panoramic radiographs. We evaluate the performance and segmentation accuracy of these model using a pre-request dataset provided by Intelligent Vision Research Lab (Ivisionlab) alongside its ground truth [6]. Based on the results presented in this paper, these methods can be used to improve the detection and segmentation of teeth in panoramic X-ray images.

The first section discusses automatic tooth detection in panoramic images. In the second section, the methodology for evaluating the U-net algorithm is explained. Three and four sections describe the results of the evaluation experiment and the setup, respectively. In the last section, the findings of this study are summarized.

II. RELATED WORKS

In the last two decades, teeth segmentation has been the subject of research, mainly using threshold, region-based methods and and machine learning methods.

Jader *et al.,* [7] present the development of the first method for segmenting and recognizing teeth from panoramic X-ray images using a region-based CNN (R-CNN). This algorithm adds a branch for automatic recognition of object masks simultaneously with the branch for class classification and bounding box recognition [8].

In order to segment the teeth from 3D dental model, Xu *et al.,* [9] proposes an approach based upon deep convolutional neural networks for segmenting 3D dental models. Further, Tian *et al.,* [10] introduced an automated method of segmenting and describing teeth from 3D dental images by utilizing OCTREE sparse voxel technology, CRF model based on conditional random fields, and a 3D CNN named OCN [11].

Furthermore, Zhu *et al.,* [12], studied the tooth segmentation and detection of teeth using Mask R-CNN. On the basis of 100 images collected from a hospital, their method successfully distinguished between complex, crowded tooth structures. A fully (R-CNN) method for automatic tooth segmentation is being evaluated by Lee *et al.,* [13] to evaluate it using individual annotations of panoramic radiographs.

A novel feature-steered graph CNN has been developed by Sun *et al.,* in their study [14]. Using this network, individual teeth were segmented and identified from digital dental castings. Towards this goal, the framework constrains its segmentation and labeling based on the distribution of crown shapes and concave contours. This method is more accurate than other DL-based dental segmentation methods, such as PointNet [15], OCTREE-based CNN [11], and the two-phase cast segmentation methods.

In Zhao *et al.,* [16], attention networks were utilized for the segmentation of attention in a two-stage network for localizing multiple teeth from a publicly available panoramic X-ray image dataset. Silva *et al.,* [17] introduced TSAS-Net for the segmentation and localization of teeth from the panning dataset.

Using comprehensive semantic data, Cui *et al.,* [18] presented a comprehensive method for segmenting teeth based on Generative Adversarial Networks (GAN's) [19]. This paper presents a deep segmentation network based on an automatic pixel-level tooth segmentation method (ToothPix) using a conditional GAN structure (CGAN). A comparison of the ToothPix method with existing methods, such as Mask R-CNN and Pix2pix showed that it outperformed state-of-the-art methods. Furthermore, as part of their efforts to exploit 3D cone beam computed tomography (CBCT) images that are robust to metal artifacts generated by the procedure, Chung *et al.,* [20] have proposed the use of CNN for pixel-wise labeling of CBCT images.

A U-shaped deep CNN (U-Net) architecture was used by Zheng *et al.,* [2]. The authors propose a variant of the Dense U-Net that is anatomically constrained during design [3] that is designed to integrate oral anatomy knowledge with data-driven Dense U-Nets. They aim to provide an automated means of segmenting and detecting lesions in CBCT images.

In addition, the technique proposed by Yang *et al.,* in their work [21] aim to combines mathematical analysis (*i.e.*level set) with deep learning CNN in order to segment the teeth from CBCT images. Where Leite *et al.,* [22] described a methodology for combining different CNN models to assess the ability of a new AI-based tool to detect and segment teeth from panoramic radiographs. They developed a detector that detects teeth and fine-tunes the segmentation map by combining DeepLab-v3 [23] architecture and a pretrained ResNet-101 [24] to detect teeth.

A novel technique described by Chandrashekar *et al.,* [25] combines independent tooth segmentation and identification models obtained from panoramic radiographs. Through the use of tooth segmentation and identification models, their collaboration aims to improve collaborative learning. Through collaboration, segmented teeth are identified and numbered to enhance results.

A recent study by Hou *et al.,* [26] described a Teeth U-net model for the segmentation of dental panoramic X-ray images. The aim was to solve the problem of accurate segmentation of all teeth in dental panoramic images and the determination of clear boundaries between roots. As a means of recovering image features in the network, dense skip connections between the encoder and decoder are proposed through the use of multi-level connections. Where Duman *et al.,* [27] used 434 anonymized, mixed-size panoramic radiography images over the age of 13 years as data, they developed automatic tooth segmentation models using a Pytorch implementation of the U-Net model.

In terms of teeth segmentation, the transfer learning models show promising results. Haghanifar *et al.,* [28] aimed to automate the process of segmenting teeth and detecting dental caries in panoramic images by utilizing automatic diagnosis systems. Through transfer learning, the proposed model extracts relevant features from x-rays and draws predictions using a capsule network.

III. METHOD

The dentist uses panoramic radiographs to obtain an overview of the entire mouth and jaw, including all the teeth, in dentistry. It has been used to detect larger concerns like infections, impacted teeth, and tumors. There is a low resolution in panoramic radiography X-ray images, which contributes to noise in the images. To process dental X-ray images, it is necessary to distinguish between the ROI and backgrounds [29]. In this research we compare 3 different CNN models that used regularly in medical image segmentation task and evaluate their results using a publicly available dataset. The following repositories contain all the code used in this paper: U-net++ [30], ResUNet++ [31] and MutiResU-Net [32].

A. Dataset and Ground Truth

It is noteworthy that panoramic X-ray images provide a greater degree of patient comfort than other radiographics, such as intraoral images (bitewing and periapical), and are less invasive, while examining a greater portion of the maxilla and mandible [33]. For dental image analysis, only a few datasets of panoramic X-ray images are publicly available.

Only a few sets of dental images were available for image analysis in the past, and almost all of these were intraoral X-rays (bitewing or periapical). The UFBA-UESC dental images dataset was published by Silva *et al.,* [17] to fill this gap, and it has proven to be a valuable resource for the community. The original data set was published with annotations for semantic segmentation only, which utilizes binary masks to distinguish teeth from non-teeth pixels. Jader *et al.,* [7] modified the UFBA-UESC Dental Images dataset to include instance segmentation information, and a total of 276 images containing 32 teeth were used for training and validation, with the remaining 1224 images being used for testing. Recently, Silva *et al.,* [6] from Ivisionlab they annotated 543 images with number information (including the 276 used by Jader *et al.,* [7]) to evaluate semantic segmentation.

 ISBN: 978-1-68558-049-0

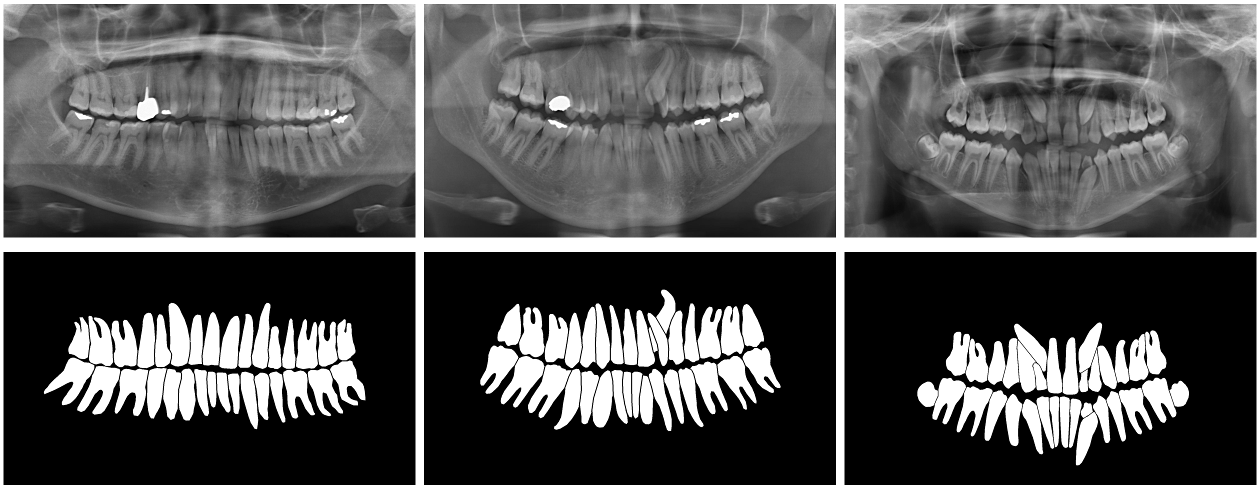

Figure 1: Three different simple panoramic X-ray images from Ivisionlab [6] alongside with their ground truth.

The dataset for this paper was obtained from Ivisionlab [6], [34] in order to perform our experiments . In this dataset, total of 1500 panoramic X-ray images with high variability have been grouped into ten categories in this dataset. It has also been updated with more instance annotations and includes information regarding numbering. A combination of panoramic X-ray images and ground truth images is included in this dataset. Figure 1 shows some samples of the dataset alongside ground truth.

B. Models Architectures Overview

1) U-net++ Architecture: The U-net++ architecture [4] in terms of medical image segmentation, is a more powerful architecture. There are several nested, dense skip pathways connecting the encoder and decoder sub-networks in this architecture. As a result of the redesign of the skip pathways, the semantic gap between the feature maps of the encoder and decoder sub-networks is reduced.

2) ResUNet++ Architecture: The ResUNet++ Architecture [5] is based on the Deep Residual U-Net (ResUNet) [35], which is a deep residual learning concept combined with an U-Net. There are three encoder blocks and three decoder blocks comprised of the ResUNet++ architecture. An encoder block comprises two successive convolutional blocks of 3×3 and an identity mapping. Using residual blocks, a deeper neural network that can solve the degradation problem in each encoder using residual blocks that propagate information over layers. Consequently, channel interdependencies are improved while computational costs are reduced.

3) MutiResU-Net Architecture: In MutiResU-Net architecture [36], a MultiRes block is proposed as a replacement for two convolutional layers. The number of filters in the convolutional layers is controlled by a parameter within each MultiRes block. A MultiRes block has been proposed in order to enhance U-Net's capability to analyze and assess data at multiple resolutions. In some cases, there is a discrepancy between the features propagated through the encoder network and the features propagated through the decoder network. In order to balance these two incompatible feature sets, MutiResU-Net offers some additional processing (*i.e.*Res paths).

IV. Results and Discussion

In this section, we present quantitative and qualitative validations using panoramic X-ray images, then compare the results with different CNNs approaches.

A. Evaluation and Assessment Metrics

In order to evaluate the predictive performance of each detection model, we use F1 and F2 scores. F1 scores are calculated by combining precision and recall, and therefore provide a more accurate measure of predictive performance than simply the percentage of correct predictions. Where F2 is defined as the weighted average mean of precision and recall (given a threshold value). A F2 score places more emphasis on recall than precision, in contrast with the F1 score, in which precision and recall are equally weighted.

A measurement of accuracy is used to determine how closely a measurement is to a standard or known value. Moreover, segmentation tasks are commonly evaluated by Dice scores and Jaccard indices in medical imaging. It is common for convolutional neural networks to be optimized for cross-entropy (weighted) when they are trained to segment images. The purpose of this measurment is essentially to quantify the overlap between our prediction output and the target mask.

The relevant mathematical expressions are as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall (Sensitivity)} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{F1-Measure} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{3}$$

$$\text{F2-Measure} = \frac{TP}{TP + 0.2 \cdot 0.8FN} \tag{4}$$

TABLE I: QUANTITATIVE COMPARISON OF DIFFERENT CNN MODULES APPLIED TO IVISIONLAB DATASET [6].

CNN Model	Evaluation Matrix					
	Jaccard index	*Recall*	*Precision*	*Accuracy*	*F1-Measure*	*F2-Measure*
U-Net++ [4]	**0.8591**	**0.9228**	0.9273	0.9715	**0.9218**	**0.9217**
ResU-Net++ [5]	0.8501	0.9098	0.9283	0.9703	0.9161	0.9115
MultiResU-Net [36]	0.8588	0.9162	**0.9339**	**0.9716**	**0.9218**	0.9176

* Bold font indicates the best value.

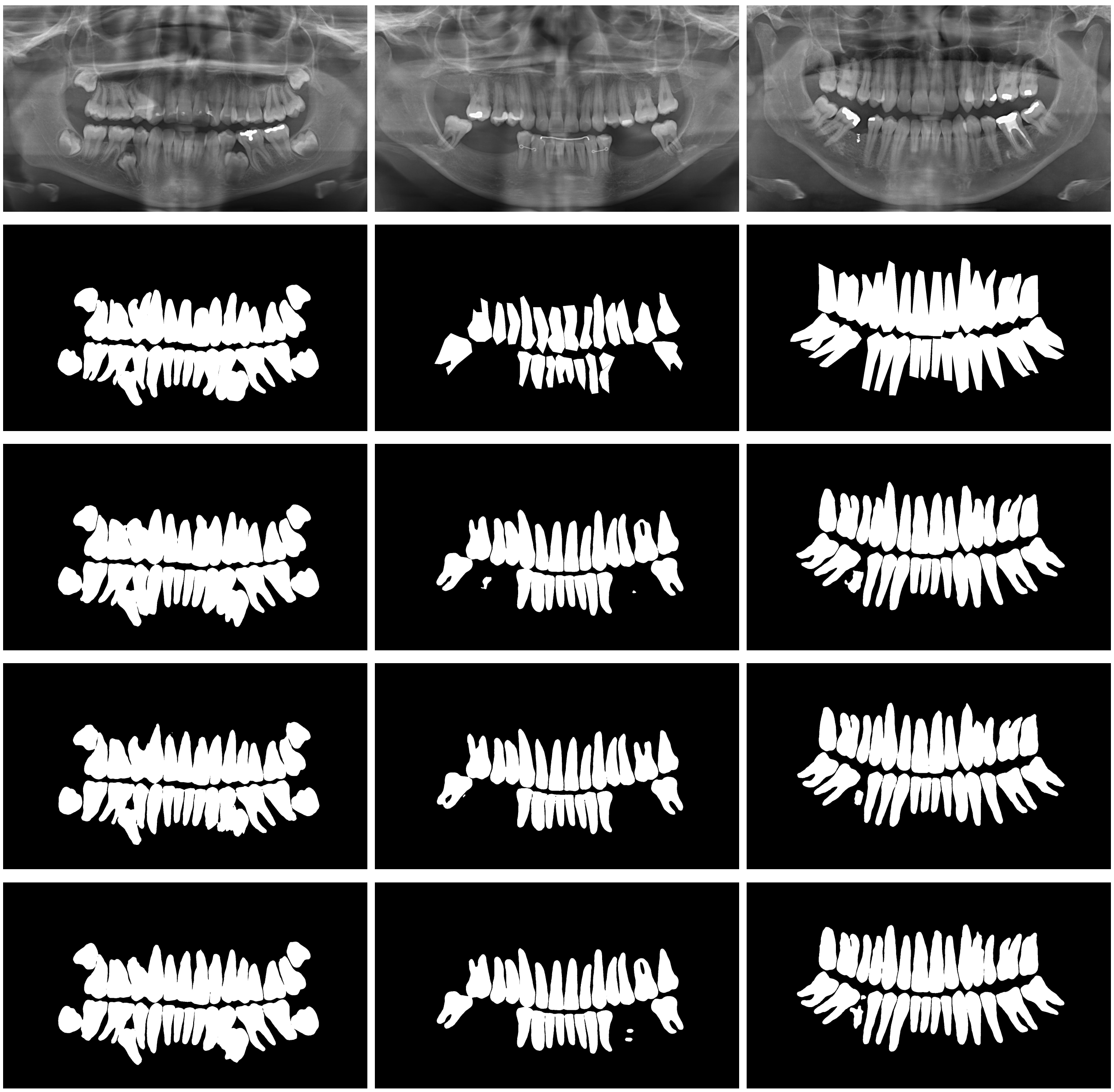

Figure 2: Qualitative analysis and comparison of the different CNN models using sample of panoramic X-ray images from IvisionLab dataset [6]. (First row): shows the original images, (Second row): ground truth, (Third row): U-Net++ [4], (Fourth row): ResU-Net++ [5] and (Fifth row): MultiResU-Net [36] segmentation results.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$\text{Jaccard index} = \frac{TP}{TP + FN + FP} \tag{6}$$

Where TP is true positive, TN—true negative, FP—false positive, and FN—false negative cases.

B. Quantitative and Qualitative Comparison

In the following Table I and Figure 2, U-Net shape CNN models are quantitatively analyzed using IvisionLab data [6]. It is notable that MutiResU-Net outperformed compare with other methods with accuracy of 97.16%. This is because MutiResU-Net performs better on heterogeneous datasets than classical U-Net [36].

C. Experimental Setup

The experiments were conducted using Python, more specifically Python3 [37]. Where in order to construct the network models, Keras [38] was used with Tensorflow [39] as the backend.

V. Conclusion

The results of our study indicate that MutiResU-Net may succeed the other U-Net architectures in the future, particularly when it comes to segmenting teeth from panoramic X-ray images. This experiment and assumption relied on a single dataset for the evaluation of different models, which could explain why MultiResU-Net performed better. Future research should conduct experiments with different datasets to see whether this claim holds.

Acknowledgment

The author extend her appreciation to the Deputyship for Research & Innovation, Ministry of Education, and Qassim University, Saudi Arabia, for funding this research work. The author also thank Qassim University for technical support.

References

[1] D. L. Duong, M. H. Kabir, and R. F. Kuo, "Automated caries detection with smartphone color photography using machine learning," *Health Informatics Journal*, vol. 27, no. 2, p. 14604582211007530, 2021.

[2] Z. Zheng, H. Yan, F. C. Setzer, K. J. Shi, M. Mupparapu, and J. Li, "Anatomically constrained deep learning for automating dental cbct segmentation and lesion detection," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 603–614, 2020.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[4] Z. Zhou, M. M. Rahman 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: International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, 2018, pp. 3–11.

[5] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen, "Resunet++: An advanced architecture for medical image segmentation," in *2019 IEEE International Symposium on Multimedia (ISM)*, 2019, pp. 225–2255.

[6] B. Silva, L. Pinheiro, L. Oliveira, and M. Pithon, "A study on tooth segmentation and numbering using end-to-end deep neural networks," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2020, pp. 164–171.

[7] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic x-ray images," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018, pp. 400–407.

[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[9] X. Xu, C. Liu, and Y. Zheng, "3d tooth segmentation and labeling using deep convolutional neural networks," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 7, pp. 2336–2348, 2018.

[10] S. Tian, N. Dai, B. Zhang, F. Yuan, Q. Yu, and X. Cheng, "Automatic classification and segmentation of teeth on 3d dental model using hierarchical deep learning networks," *IEEE Access*, vol. 7, pp. 84 817–84 828, 2019.

[11] P.-S. Wang, Y. Liu, Y.-X. Guo, C.-Y. Sun, and X. Tong, "O-cnn: Octree-based convolutional neural networks for 3d shape analysis," *ACM Transactions On Graphics (TOG)*, vol. 36, no. 4, pp. 1–11, 2017.

[12] G. Zhu, Z. Piao, and S. C. Kim, "Tooth detection and segmentation with mask r-cnn," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 070–072.

[13] J.-H. Lee, S.-S. Han, Y. H. Kim, C. Lee, and I. Kim, "Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 129, no. 6, pp. 635–642, 2020.

[14] D. Sun, Y. Pei, G. Song, Y. Guo, G. Ma, T. Xu, and H. Zha, "Tooth segmentation and labeling from digital dental casts," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 669–673.

[15] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[16] Y. Zhao, P. Li, C. Gao, Y. Liu, Q. Chen, F. Yang, and D. Meng, "Tsasnet: Tooth segmentation on dental panoramic x-ray images by two-stage attention segmentation network," *Knowledge-Based Systems*, vol. 206, p. 106338, 2020.

[17] 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.

[18] W. Cui, L. Zeng, B. Chong, and Q. Zhang, "Toothpix: Pixel-level tooth segmentation in panoramic x-ray images based on generative adversarial networks," in *International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 1346–1350.

[19] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[20] M. Chung, M. Lee, J. Hong, S. Park, J. Lee, J. Lee, I.-H. Yang, J. Lee, and Y.-G. Shin, "Pose-aware instance segmentation framework from cone beam ct images for tooth segmentation," *Computers in biology and medicine*, vol. 120, p. 103720, 2020.

[21] Y. Yang, R. Xie, W. Jia, Z. Chen, Y. Yang, L. Xie, and B. Jiang, "Accurate and automatic tooth image segmentation model with deep convolutional neural networks and level set method," *Neurocomputing*, vol. 419, pp. 108–125, 2021.

[22] A. F. Leite, A. V. Gerven, H. Willems, T. Beznik, P. Lahoud, H. Gaêta-Araujo, M. Vranckx, and R. Jacobs, "Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs," *Clinical oral investigations*, vol. 25, no. 4, pp. 2257–2267, 2021.

[23] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[24] Y. Rao, L. He, and J. Zhu, "A residual convolutional neural network for pan-shaprening," in *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)*. IEEE, 2017, pp. 1–4.

[25] G. Chandrashekar, S. AlQarni, E. E. Bumann, and Y. Lee, "Collaborative deep learning model for tooth segmentation and identification using panoramic radiographs," *Computers in Biology and Medicine*, vol. 148, p. 105829, 2022.

[26] S. Hou, T. Zhou, Y. Liu, P. Dang, H. Lu, and H. Shi, "Teeth u-net: A segmentation model of dental panoramic x-ray images for

context semantics and contrast enhancement," *Computers in Biology and Medicine*, vol. 152, p. 106296, 2023.

[27] S. Duman, E. F. Yılmaz, G. Eşer, Ö. Çelik, I. S. Bayrakdar, E. Bilgir, A. L. F. Costa, R. Jagtap, and K. Orhan, "Detecting the presence of taurodont teeth on panoramic radiographs using a deep learning-based convolutional neural network algorithm," *Oral Radiology*, vol. 39, no. 1, pp. 207–214, 2023.

[28] A. Haghanifar, M. M. Majdabadi, S. Haghanifar, Y. Choi, and S.-B. Ko, "Paxnet: Tooth segmentation and dental caries detection in panoramic x-ray using ensemble transfer learning and capsule classifier," *Multimedia Tools and Applications*, pp. 1–21, 2023.

[29] S. M. Kahaki, M. Nordin, N. S. Ahmad, M. Arzoky, W. Ismail *et al.*, "Deep convolutional neural network designed for age assessment based on orthopantomography data," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9357–9368, 2020.

[30] MrGiovanni. (2019) U-net++. [Online]. Available: https://github.com/MrGiovanni/UNetPlusPlus

[31] DebeshJha. (2022) Resunetplusplus. [Online]. Available: https://github.com/DebeshJha/ResUNetPlusPlus

[32] Nibtehaz. (2022) Multiresunet. [Online]. Available: https://github.com/nibtehaz/MultiResUNet

[33] J. Kim, H.-S. Lee, I.-S. Song, and K.-H. Jung, "Dentnet: Deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.

[34] silva. (2020) Ivisionlab, dns panoramic. [Online]. Available: https://github.com/IvisionLab/dns-panoramic-images

[35] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.

[36] N. Ibtehaz and M. S. Rahman, "Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.

[37] G. Van Rossum *et al.*, "Python programming language." in *USENIX annual technical conference*, vol. 41, no. 1, 2007, pp. 1–36.

[38] N. Ketkar and N. Ketkar, "Introduction to keras," *Deep learning with python: a hands-on introduction*, pp. 97–111, 2017.

[39] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: a system for large-scale machine learning." in *Osdi*, vol. 16, no. 2016, 2016, pp. 265–283.