IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 38, No. 4, August, 2024, pp. 1173-1181
Journal homepage: http://iieta.org/journals/ria

Deep Learning Based Teeth Segmentation

Husam Al-Behadili[1*], Omar A. Athab[2], Saddam K. Alwane[3]

[1] Electrical Engineering Department, University of Mustansiriyah, Baghdad 10052, Iraq
[2] Information and Communications Engineering Dept., University of Baghdad, Baghdad 10071, Iraq
[3] Computer Engineering Department, University of Technology, Baghdad 10066, Iraq

Corresponding Author Email: husam.albehadili@uomustansiriyah.edu.iq

https://doi.org/10.18280/ria.380411

Received: 25 October 2023
Revised: 1 March 2024
Accepted: 21 March 2024
Available online: 23 August 2024

Keywords:
deep learning, medical image segmentation, teeth segmentation, UNet, Res-Unet

ABSTRACT

This article discusses the importance of accurate teeth segmentation in dental diagnosis and treatment. Rapid advancements in Artificial Intelligence have led to the development of various approaches for example Res-UNet deep learning architecture. Res-UNet++ has been proposed as a refined version of the Res-UNet architecture to improve teeth segmentation performance. Res-UNet++ integrates three additional elements: squeeze and excitation block, atrous spatial pyramid pooling, and attention block. The purpose of these components is to enhance the performance of Res-UNet by improving the recalibration of features at both the channel and spatial levels, capturing multi-scale contextual information, and prioritizing the relevant regions of interest. Res-UNet++, UNet and Res-UNet were compared on two publicly available dental image datasets using evaluation criteria such as the dice coefficient and mean Intersection over Union (mIoU). The evaluation of these algorithms was implemented under the same experimental settings to statistically assess the significance of the enhancements. The result shows the superiority of Res-UNet++ over UNet and Res-UNet. The effectiveness of Res-UNet++ is demonstrated by its impressive assessment scores: the dice coefficient of 92.91% and 95.58% for the two databases, and the mean Intersection over Union (mIoU) of 88.68% and 88.72%.

1. INTRODUCTION

In clinical practice, clinicians frequently use radio-graphs as a standard imaging system for diagnosing and treating tooth loss, thanks to their cost-effectiveness. Additionally, the panoramic X-rays provide a rich detail due to its ability of capturing a broad range of the maxillomandibular region. Moreover, its radiation level is the lowest compared to other techniques [1]. The panoramic radio-graphs have been used by dentists to identify various of dental problems including: bone abnormalities, cavities, hidden dental structures, and post-traumatic fractures, which are difficult or almost impossible to detect through a visual examination [2]. Thus, the dentist will be able to prepare a suitable therapy strategy for each patient. In some situations, dentists may choose to outsource the task of analyzing X-rays due to its nature. This practice can be time-consuming. Requires a level of expertise to differentiate relevant dental features from non-essential ones such as jaw bones, nasal bones and spine bones [3]. Due to the dissimilarity of the dentist's experience levels by extracting the information from the radio-graph images, a different diagnosis for the same radio-graph may arise. Consequently, inappropriate treatments pathology for some cases could be happen [4].

Several algorithms and models are proposed to automate the diagnosis and solves this diagnosis diversity. Some of these systems utilize machine learning techniques like the contour model [5] and SVM [6] combined with handcrafted features. However, the performance of these methods is often limited by their reliance on handcrafted features [4]. On the other hand, breakthroughs have been achieved by developing automated systems through deep learning approaches that have surpassed machine learning algorithms in delivering superior results. Specialized architectures tailored for handling images across domains, such as UNet [7] and DeepMedic [8] have played a pivotal role in this field. These architectures have been redesigned into nested structures [9], or equipped with self-adaptability features [10], for enhanced performance. Because of its range of customization choices, UNet has been viewed as a framework, more than just an architecture. Making it a good fit for incorporating new techniques [8]. However, with these advancements, teeth segmentation technology encounters obstacles due to the nature of dental anatomy and the requirement for precise techniques. The current methods encounter difficulties, with teeth that overlap, shape variations and imaging imperfections. These issues indicate that the segmentation required more enhancements, particularly the teeth segmentation. Thus, Res-UNet++ has been suggested as a method to get improved performance.

Driven by the high-performance of Res-UNet++ [11] architecture in automatic polyp segmentation, this architecture has been adopted in the presented fundamental framework. Res-UNet++ is a novel architecture for medical image segmentation that improves upon the existing Res-UNet

model. Res-UNet is a combination of UNet and ResNet, two popular deep learning architectures for image segmentation and classification, respectively. Res-UNet uses residual blocks to enhance the feature extraction and skip connections to preserve the spatial information across the encoder-decoder network. It also allows for the propagation of information over layers, enabling the construction of deeper neural networks to address the degradation problem. Res-UNet++ extends Res-UNet by incorporating three additional components: squeeze and excitation block, Atrous Spatial Pyramid Pooling (ASPP), and attention block modules within Res-UNet architecture to enhance teeth segmentation. These components aim to improve the performance of Res-UNet by enhancing the channel-wise and spatial-wise feature recalibration, capturing multi-scale contextual information, and focusing on the relevant regions of interest, respectively. Which means further enhances the architecture's ability to capture intricate details and spatial dependencies within the images.

Res-UNet++ has been shown to achieve state-of-the-art results on various medical image segmentation tasks, such as polyp segmentation, brain tumor segmentation, and lung segmentation. Res-UNet++ is an advanced architecture that can handle the challenges of medical image segmentation, such as high inter-class similarity, intra-class variation, and low contrast. The model has been evaluated using two publicly accessible datasets. The TUFTS benchmark dataset serves as the first tool. The second dataset, used in the research [12], comprises de-identified panoramic dental x-ray photographs of 116 volunteers, sourced from Noor Medical Imaging Center, Qom, Iran. The experimental results display that the proposed model outperforms popular architectures like UNet [7] and its variant Res-UNet [13] efficiently with a notable performance increase.

The remainder of the paper unfolds as follows: Section two delves into related work of teeth segmentation. Details concerning material and methodology appear in section three, while section four presents experimental results. Ultimately, Section five hosts a discussion regarding the conducted study and prospective improvements.

2. RELATED WORKS

The related work provides crucial insights into the evolution of dental image segmentation techniques leading up to the proposed Res-UNet++ architecture. Initially, Ronneberger et al. [7] announced the UNet architecture in 2015; since then, extensive investigation has been conducted on this topic and numerous studies confirmed its effectiveness in detecting and segmenting visual medical data. Over the years, diverse adaptations have transformed UNet: some included adding a batch normalization layer to its encoder component to improve stability, others applied innovative strategies aiming at boosting performance for designated tasks.

Machado et al. [14] managed to segment the mandible bone in panoramic X-ray photography successfully. Widyaningrum et al. [15] further extended the research landscape by pursuing a segmentation methodology for periodontitis staging; they utilized two distinct approaches Multi-Label UNet and Mask RCNN, pre-training the latter for other tasks via transfer learning. Almalki and Latecki [16] employed cutting-edge self-supervised deep learning algorithms such as Sim-MIM and UM-MAE; this approach improved the efficiency of their model in interpreting a finite set of dental radio-graphs. They leveraged the Swin Transformer an influential variation of the transformer model in their research study.

Attention techniques have also gained attention in teeth segmentation, as emphasized by Mahran et al. [17] and Harsh et al. [18], who utilized a channel-based Attention UNet models. These models incorporate attention blocks like Squeeze and Excitation (SE) to filter pertinent information, demonstrating superior performance compared to traditional methods.

The extendibility of channel-wise attention in UNet to other applications, often blended with spatial attention [19], gathers features on a global scale for the input. It adapts different types of attention mechanisms specifically for UNet architecture like grid-based attention gate [20]. This mechanism calculates local-scale coefficients of attentions that allow more detailed output and have demonstrated excellent performance in tasks such as pancreas segmentation [20], deforestation detection [21] and ischemic lesion segmentation within the brain [22]. However, its application remains absent from dental segmentation tasks thus far, according to our knowledge. Dayı et al. [23] assessed the diagnostic precision of deep learning models in segmenting occlusal, proximal, and cervical caries lesions seen on panoramic radio-graphs; their assessment used a dataset comprising 504 anonymized panoramic radio-graphs. Nafi'iyah et al. [24] trained an ensemble consists 3 models of Mobile-NetV2 on 106 panoramic images to alleviate the problem of mandibular segmentation; this approach also addressed the primary shortcomings of prior existing methods which did not fully represent the mandible. The same method has been followed by Arora et al. The authors [25] employed a model grounded in an encoder/decoder framework. The encoder segment incorporated numerous convolution neural network-based models, utilizing each network and combining their outputs to create fine grained contextual features to segment the teeth.

The proposed Res-UNet++ architecture builds upon these advancements, integrating elements from previous studies such as attention mechanisms and SE blocks while also addressing the specific challenges of teeth segmentation. By drawing upon the insights gained from prior research, Res-UNet++ aims to push the boundaries of dental image segmentation and contribute to improved diagnostic and treatment planning processes.

3. RES-UNET++

The Res-UNet++ network is based on the Deep Residual UNet (Res-UNet) which incorporates the well-known network UNet and ResNet network [7, 13]. ResNet provides a robust learning of the residual technique as mentioned by He et al. [26]. It has been widely recognized for its ability to address the degradation problem surpassing architectures, like Highway Networks and DiracNets. A study by Monti et al. [27], compared networks designed to tackle degradation highlighting the performance of ResNet in overcoming such challenges.

The proposed Res-UNet++ architecture includes components like blocks, squeeze excitation blocks, Atrous Spatial Pyramid Pooling (ASPP) and attention blocks. By integrating the residual block data flow smoothly across layers in this structure allowing for the creation of a neural network that effectively addresses degradation concerns, within each encoder. This process enhances channel inter-dependencies

while reducing the computational costs. The proposed Res-UNet++ framework composes 1 stem-block, 3 sequential encoder- blocks, an ASPP unit in-between followed by another set of three decoder blocks at the end. Figure 1 illustrates the schematic representation of Res-UNet++ framework. The residual unit, visible in this block diagram, integrates batch-normalization techniques, ReLU linear activation, and convolution segments. Two consecutive (3×3) convolution blocks, in conjunction with an identity mapping, compose every encoder block. Within each convolution blocks, a batch normalization layer, a ReLU activation layer and a convolution layer. Moreover, the identity mapping creates a linkage between the input and output of its respective encoder block. The initial convolution layer of the encoder-block employs a strided convolution layer to halve the spatial dimensions of the feature maps.

The squeeze-excitation-block processes the encoder block's output. Acting as a bridge, the ASPP expands the filters' field-of-view to encompass a broader context. Correspondingly, residual units also comprise the decoding path. Prior to each unit, the attention-block heightens the efficacy of the feature-maps. This is immediately followed by a nearest neighbor up-sampling of feature maps from the lower hierarchy. The feature-maps concatenated from their corresponding encoding pathway. Ultimately, the result of the decoder block undergoes ASPP processing and culminates in a 1×1 convolution with sigmoid activation, producing the segmentation map.

A progression from the fundamental Res-UNet++ is marked by the integration of squeeze-excitation-blocks depicted in Figure 1 by Cornflower Blue, the ASPP module emphasized in Flush Orange, and the attention-block highlighted in Hot Pink. Succinct elucidations of each constituent element are further expounded upon in the subsequent subsections.

3.1 Units of residue

He et al. [26] proposed the deep residual learning framework as a solution to the degradation problem that notoriously challenges training of deep neural networks due to increasing network depth. The Res-UNet [7] framework leverages full pre-activation residual units and skip connections, making the deep network training simpler and preventing information degradation. It offers distinct advantages such as reduced parameters, enhanced semantic segmentation performance, and stands comparable to more complex networks in terms of effectiveness [13]. Consequently, Res-UNet has been selected as a backbone architecture for the proposed model because of these benefits.

3.2 Units for squeezing and excitation

Hu et al. [28] emphasizes the integration of Squeeze-and-Excitation (SE) blocks into convolutional neural networks (CNNs) using both empirical evidence and theoretical frameworks. Empirically, extensive tests have been demonstrated the effectiveness of SENets, which include SE blocks, in achieving state-of-the-art performance across diverse datasets such as ImageNet, CIFAR-10, and CIFAR-100. Notably, incorporating SE blocks into established architectures like ResNet-50 and DenseNet has led to significant performance gains. Theoretical analysis underscores the potential of SE blocks: they enhance representational power and generalization capabilities indeed, through explicit modeling of inter-channel dependencies; this facilitates selective feature emphasis and global information utilization. Squeeze and excitation techniques enhance the network's information representation capacity by efficiently representing interdependencies between channels and re-calibrating features response. make the network more sensitive to important features and less sensitive to unimportant features. It does this in two phases: squeezing and exciting. The initial phase, referred to as the squeeze phase, it integrates global average pooling with each channel to compress them; thereby generating channel-specific statistics for information embedding on a large scale. Subsequently, the excitation phase ensues with a goal to thoroughly encapsulate each channel's inherent dependencies [28]. Within the proposed framework, a residual-block has been combined with a squeeze and excitation block; this serves as an effective strategy that increases generalization capabilities across various datasets and improves network performance overall.

3.3 Atrous spatial pyramidal pooling utilizes

Atrous convolutions, often referred to as dilated convolutions, deviate from standard convolutions by incorporating empty spaces within the convolutional kernel. The presence of these gaps, or holes, enables the convolutional operation to extract input signals with a wider receptive field, so extracting information from a more extensive context without augmenting the number of parameters or computational expense. The dilation rate in atrous convolutions regulates the distance between kernel elements and defines the effective stride inside the input feature map. Atrous convolutions may capture multi-scale information at different resolutions by modifying the dilation rate. This allows the network to extract features at several scales simultaneously. Atrous convolutions are highly effective in semantic images segmentation as they allow the network to include contextual information from a broader variety of spatial settings, therefore capturing multi-scale information more efficiently. The Res-UNet++ design combines features from paralleled convolution layers with different dilation rate to collect the multi-scale data efficiently. This process is achieved by incorporating the atrous convolutions in the ASPP structure within Res-Unet++. The procedure enables the model to be adjusted to various object scales and situations to enhance its capacity of precise teeth segmentation [29, 30].

The ASPP in the proposed architecture serves as a connection between the encoder and decoder components, as described in Figure 1. The ASPP model provides promising outcomes in segmentation challenges by supplying essential multi-scale information. The multi-scale information has been leveraged to gather precious multi-scale information for semantic segmentation tasks.

3.4 Units of attention

Attention mechanisms, a concept widely embraced in Natural Language Processing (NLP) [18], concentrate on a subset of the input. These mechanisms have demonstrated their usefulness in semantic segmentation tasks such as pixel-wise prediction [31]. Intrinsically, they identify sections of the neural network that need increased focus which subsequently reduces the computational load involved in transforming data from each tooth image into a fixed-dimensional vector. Attention mechanisms stand out due to their simplicity, adaptability to diverse input sizes, and ability to enhance

feature quality-all factors that elevate results. Unlike previous methods such as UNet [7] and Res-UNet [13], which directly concatenate the encoder's feature maps with those of the decoder, the presented architecture draws inspiration from attention mechanisms' successes in both NLP and computer vision contexts by introducing an attention block within its decoder segment. This augmentation allows for an intense focus on vital areas within the feature maps.

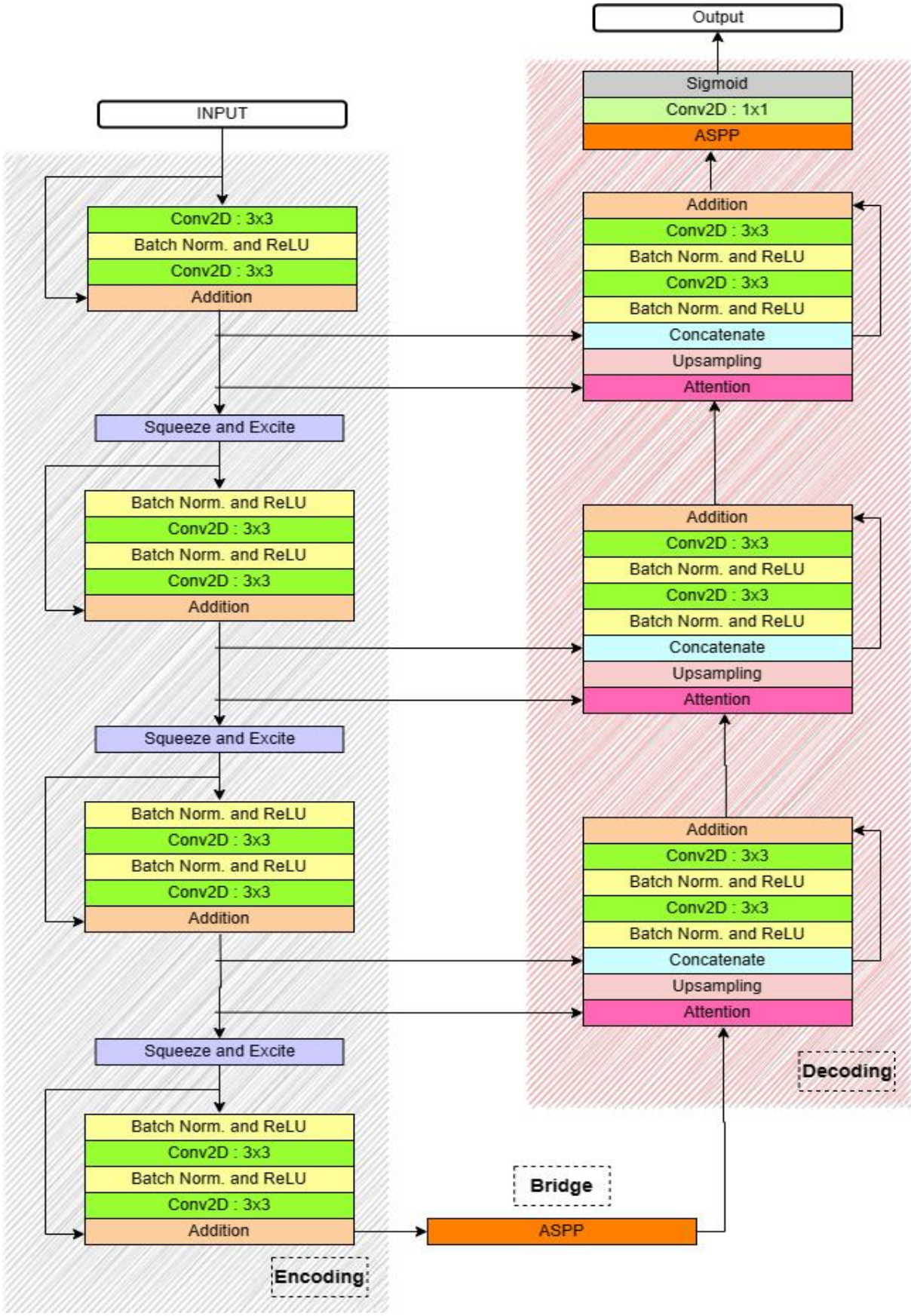


Figure 1. Block diagram of the proposed Res-UNet++ architecture

4. EXPERIMENT

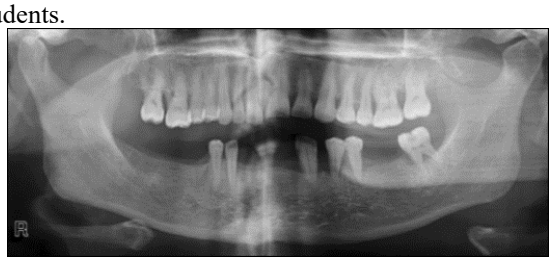
The investigation has been conducted through a comprehensive series of steps that encompasses algorithm training, validate the parameters, and test the model on unseen images. To evaluate the Res-Unet++ algorithm, the algorithm performance has been precisely compared with the well-known Unet and Res-Unet networks. By these comparative analyses, we can holistically understand both the strengthens and weakness of Res-Unet++ architecture. The algorithms have been applied on two publicly available datasets. These data sets consist a diverse set of samples to simulate the reality, which has a diverse of teeth cases. Thus, these datasets were carefully selected to thorough evaluate the performance of the proposed algorithm. The model's performance has been thoroughly evaluated by analyzing important metrics including the dice coefficient, mean Intersection over Union (mIoU), and Pixel Accuracy (PA).

Furthermore, deeper investigation was probed into nuanced evaluations: the focus extends to the architectures' robustness when faced with challenging or previously unseen samples. Conducting these analyses on different datasets, the conclusion is corroborated and remain sturdy and relevant across a myriad of scenarios.

4.1 Datasets

4.1.1 Multimodal dataset from TUFTS university

The multimodal dataset from TUFTS University [32] consists of 1,000 de-identified images of panoramic radio-graphs, Figure 2 (a), and five additional major components which include: i) teeth masks Figure 2 (b), ii) maxillomandibular masks, iii) eye tracker-generated maps in both grey and quantized formats, iv) detailed text descriptions of each radio-graph, and finally, v) two forms of abnormality segmentation mask-expertly annotated and student-level. Expert-annotated images may display a higher level of experience, expertise, and accuracy in identifying and labeling abnormalities in panoramic radio-graphs compared to their student-level annotated counterparts. The dental professionals involved likely provide annotations that are more accurate and reliable due to their extensive experience and specialized knowledge; they possess an advanced understanding of dental pathology as well as radio-graphic interpretation - this results in the finer identification plus labeling of abnormalities. As a gold standard for comparison and benchmarking purposes, expert annotations excel. Conversely, the accuracy and consistency of student annotations may fluctuate based on their dental training level and experience; limited proficiency in radio-graph interpretation can lead to discrepancies or errors within these notes. Student annotations, however, do more than just offer a different perspective on radio-graphic interpretation; they actually provide valuable insights into the learning process and diagnostic skill development among dental students.



(a)

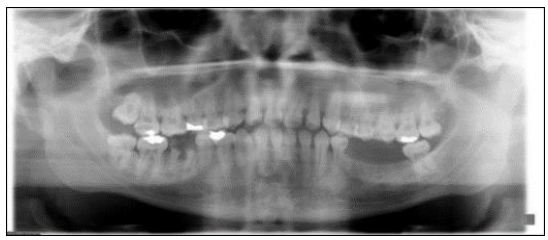
4.1.2 Noor Medical Imaging Centre (NMIC)

A de-identified dataset has been used, consisting of 116 panoramic dental x-ray radio-graphs from volunteers at Noor Medical Imaging Center (NMIC), Qom, Iran [12]. Although this dataset includes manual segmentations of mandibles, these segmented images were deemed irrelevant to presented study; therefore, only the original images served for the research purposes. The Soredex CranexD digital panoramic x-ray unit took the images. All image widths vary between 2,600 and 3,138 pixels, while their heights range from 1,050 to 1,380 pixels. During data preprocessing, all panoramic dental x-ray images were resized to a standard size of 1,500x1,500 pixels. In Figure 2 (f), an example of an input image has been provided. A teeth mask has been obtained for each panoramic dental x-ray image in the dataset through manual labeling. Figure 2 (g) labels all the teeth in this mask, as described.

4.2 Implementation details

The model was implemented and trained on an NVIDIA RTX 3080 GPU using the PyTorch framework version 1.6.0. and tensorboardX library. Initially, the developed model trained using 16 batch size. Moreover, in conjunction with 32GB RAM, Adam algorithms has been used the for-optimization tasks. Algorithm's learning rate was programmed at 0.0001. Lower learning rate generally is preferred, even though it may decelerate the speed of convergence. Conversely, adopting a higher learning rate often impedes achieving convergence effectively.

In this study, the variability in image sizes within and across datasets have been taken into account to develop a strategy that effectively utilizes the GPU and minimizes training time. The images displayed varying resolutions, which led us to resize them consistently to a resolution of 1,500×1,500 pixels. For further precision, the training dataset has been enhanced by cropping each image with a margin of 224×224 before have been used in the frameworks. The proposed approach integrates also a diverse range of techniques that have been used to augment the data, including center and random cropping, horizontal and vertical flipping, scale and illumination level augmentation, cutout, and random rotation. The rotated samples have been rotated at random angle from a range spanning 0 to 90°. To balance model performance assessment, the data were allocated 80%, 10%, and 10% for training, validation, and testing respectively. Training all models over 75 epochs with a reduced learning rate facilitated generalization. In line with specific requirements, the batch size, epoch count, and learning rate has been adjusted accordingly. To mitigate the risk of overfitting, which could potentially compromise accuracy-an issue that often arises with smaller batches-a larger batch size has been chosen. Further enhancing the proposed model's performance, The Stochastic-Gradient-Descent has been integrated with Restart (SGDR) into the training strategy.



(f)

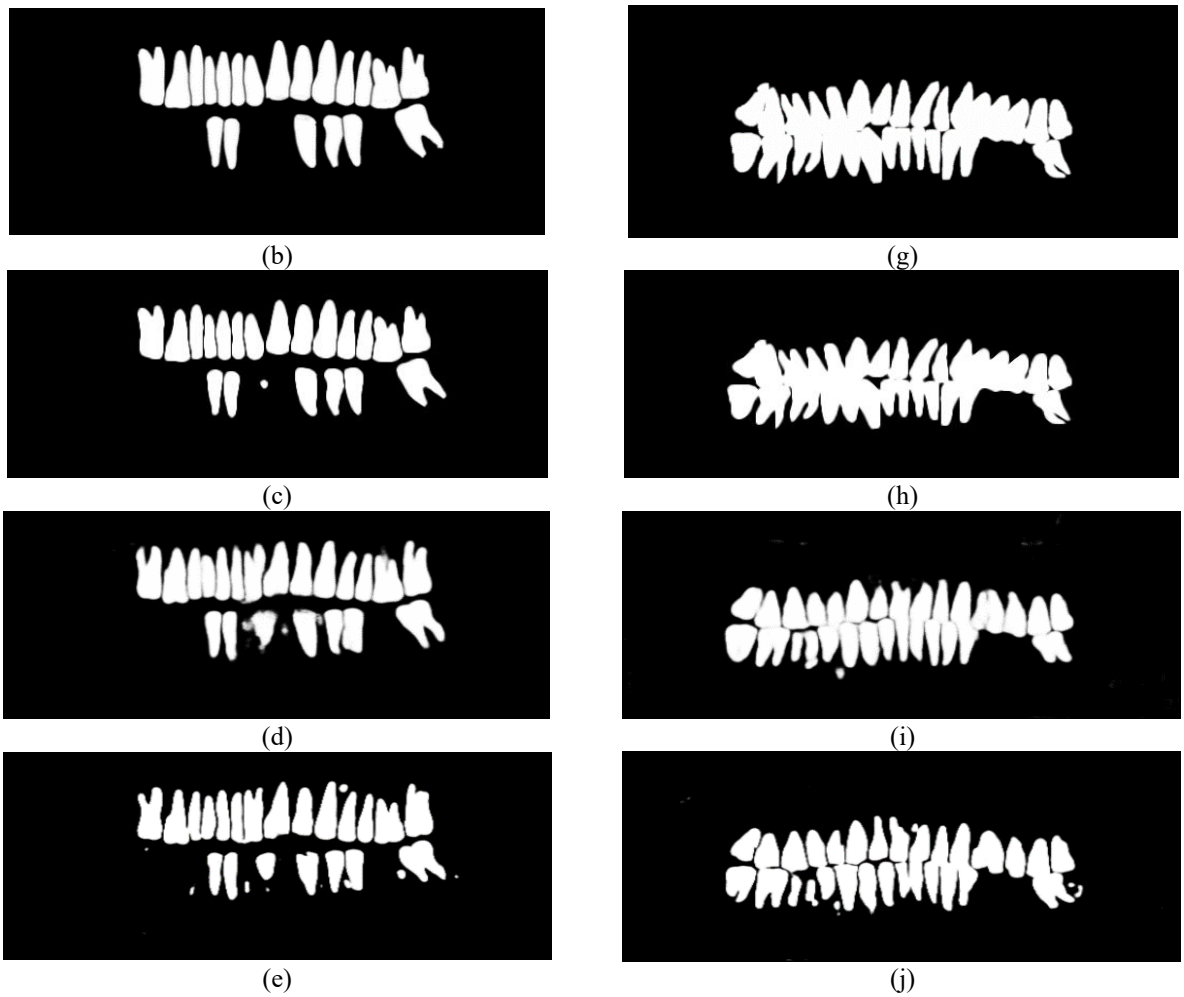


Figure 2. A sample of the datasets and predictions done on the test set. The first row (a & f) are samples of TUFTS and NMIC, respectively. (b & g) are the ground truth of first row. (c & h) are the prediction results of Res-Unet++, (d & i) is the prediction results of Unet, and (e & j) is the prediction results of Res-Unet

5. OBATAINED RESULTS

The effectiveness of the Res-UNet++ architecture has been demonstrated through a dual set of experiments conducted on the TUFTS and NMIC datasets. To compare models, the outcomes derived from the proposed Res-Unet++ were juxtaposed, original UNet, and original Res-UNet architectures-both conventionally favored choices for semantic segmentation undertakings.

5.1 Results derived from the TUFTS dataset

We meticulously fine-tuned the hyper-parameters-elements such as learning rate, epoch count, optimizer, batch size, and filters size-to refine the Res-UNet++ model. This endeavor necessitated us to train frameworks with diverse hyper-parameter configurations and then rigorously evaluate their performance. Table 1 presents the results for Res-UNet++, Res-UNet [32], and UNet [32]. The proposed model significantly outperformed in achieving the highest scores for dice coefficient and mean Intersection over Union (mIoU) on the TUFTS dataset. Despite UNet displaying higher Pixel Accuracy (PA), it produced less competitive scores for dice coefficient and mIoU essential metrics for any semantic segmentation task

Table 1. Models' evaluation results on TUFTS dataset

Method	Dice	MIoU	PA
ResNet++	**92.91**	**88.68**	95.12
ResNet	92.36	86.49	95.13
UNet	92.46	86.67	**95.22**

The recommended architecture demonstrated remarkable dominance over baseline models considering both dice coefficient and mIoU measures. Figure 2 confirms that results as Figure 2 (c)(d)(e) represent the obtained segmented images from Unet, Res-Unet, and Res-Unet++, respectively.

5.2 Results derived from the NMIC dataset

To comprehensively assess the automatic teeth segmentation performance, supplementary experiments were conducted to evaluate the model's generalizability. then the adaptability of this model has been examined within the proposed architecture by testing its performance on an alternative dataset. This crucial step towards generalization signifies a significant stride in building a medically viable model. Table 2 summarizes the outcomes of each architecture on the NMIC datasets. In terms of dice coefficient and mIoU, the presented architecture remained competitively positioned; Table 2 offers further qualitative outcomes from all models.

Table 2. Results of the tested Models on NMIC dataset

Method	Dice	MIoU	PA
ResNet++	**95.58**	**88.72**	97.44
ResNet	93.87	87.35	97.19
UNet	94.32	88.13	**97.48**

Res-UNet++ provides preferable results over the baseline models as shown in Figure 2. The outcomes display that this method is finer in quantitate, and qualitative. Moreover, it is clear that this approach is exceedingly efficient in the case of TUFTS dataset as well as NMIC. In medical images segmentation applications, the Res-UNet++ technique is the model of choice because of its mIoU and dice coefficient potency. It is necessary to apply the proposed architecture to TUFTS and NMIC datasets to determine its competence in clinical and medical image segmentation applications. This holds true in general and even with the scarcity of images in the NMIC dataset. The fact that Res-UNet++ demonstrates exceptional results on independent and different datasets, shows its ability to be generalized to other datasets including real-life clinical applications. The dataset consists of deviations in image disorders, teeth attributes, and patient maxillomandibular demographics, demonstrating practical diversity experienced in clinical practice. As a result, this will have a great influence on patient care via promoting the accuracy of diagnosis.

The dice coefficient is defined as the extent of the overlap amidst the predicted segmentation mask and the ground truth mask. It weighs the similitude between these masks via regarding false positives and false negatives. In contracts, the Pixel accuracy evaluates the percentage of accurately classified pixels in the segmentation mask without regarding the class imbalance between tooth and non-tooth pixels.

Res-Unet++ in both experiments achieved a high dice coefficient while preserving a competitive pixel accuracy. Thus, resulting in superior capturing of the intricate details of teeth and perform well in classifying most pixels correctly.

A higher Dice coefficient or mIoU means a more promising model accuracy for the segmentation of anatomical structures and pathological areas. This leads to an improvement in clinical missions including diagnosis, treatment planning, and patient observation. An effective treatment planning could be achieved by accurate pathology identifier. The latest need an accurate teeth segmentation that helps the specialist identify the condition correctly. Hence, improving the accuracy of teeth segmentation will significantly improve the dental care outcomes. From this relation, we can estimate the potential of the teeth segmentation in practical use at the dental clinics.

6. DISCUSSION

The Res-UNet++ approach shows favorable outcomes for both the TUFTS and NMIC datasets. Figure 1 illustrates a segmentation map by Res-UNet++, demonstrating an enhanced ability to capture shape information effectively, compared to other architectures within the same datasets. This suggests that, contrasted with currently dominant state-of-the-art models, the segmentation mask generated by Res-Unet++ bears a more striking resemblance to the ground truth. Notably, the UNet architecture also produces highly competitive segmentation masks.

Res-Unet++ utilize a spectrum of available loss functions, encompassing binary cross entropy, dice loss, and mean square loss during the model training process. The empirical results underscored an upward trend in the Dice coefficient values upon evaluation, highlighting commendable performance by the presented model. When the Dice coefficient has been excluded from this array of loss functions; all other options yielded significantly lower mIoU values. The dice coefficient loss function has been selected empirically. Also, the variables like the number of filters, batch size, optimizer and chosen loss function have been observed how considerably wield influence over outcomes. Increasing dataset size, supplementing with additional augmentation techniques and integrating post-processing steps could potentially improve model performance. The augmentation introduced by the proposed architecture may embraced, despite an increase in parameter count, to train the proposed model for superior results. The Res-UNet++ architecture may apply to the pattern classification as well as the segmentation process in the medical field or beyond that. These applications warrant further comprehensive validation. This code has been diligently optimized to a considerable extent using the available expertise and knowledge. Avenues for further refinement continue to exist, potentially influencing architectural outcomes. The presented code has been executed exclusively on an Nvidia RTX 3080 GPU machine; this raises the important consideration that image resizing during such a process might result in losing critical information. Moreover, integrating more parameters into Res-UNet++ prolongs the training duration. The approximate time to train the models Unet, Res-Unet, and Res-Unet++ using 2,000 Samples was 11, 12, and 13 minutes, respectively. However, the testing time was almost same and negligible.

7. CONCLUSION

In this study, Res-Unet++ has been presented to improve the accuracy of teeth segmentation. The presented architecture achieves optimal performance by harnessing components like residual-block, squeeze-and-excitation block, ASPP, and attention-block. Residual blocks enable the direct flow of information through shortcut connections, facilitating the training of deeper neural networks and enhancing the model's ability to capture complex patterns and details in teeth images. Teeth segmentation needs to focus on some details in the images rather than others. This attention could be competently achieved by utilizing Squeeze and excitation blocks. Squeeze and excitation blocks recalibrate the feature responses according to the channel by emphasizing informative details and suppressing others. Additionally, the teeth and details scale may differ from image to image and from tooth to tooth, therefore the ASPP module has been chosen in the proposed algorithm. The ASPP can handle the different details scales by utilizing dilated convolution with different rates. Moreover, some regions in the maxillomandibular radio-graphs are more important rather than other regions. This spatial selectivity has been implemented by using Attention blocks. Thus, the segmentation accuracy has been promoted. Residual, squeeze and excitation, Attention, and ASPP blocks together widely enhanced the segmentation accuracy and reduced the effect of the surrounding tissues.

Res-UNet++, UNet, and Res-UNet were compared on two publicly available dental image datasets TUFTS and NMIC using evaluation criteria such as the dice coefficient and mean Intersection over Union (mIoU) and precision accuracy (PA).

The evaluation of these algorithms was implemented under the same experimental settings to statistically assess the significance of the enhancements. The result shows the excellence of Res-UNet++ over UNet and Res-UNet. The effectiveness of Res-UNet++ is demonstrated by its impressive assessment scores: the dice coefficient of 92.91% and 95.58% for the two databases, and maintaining high mean Intersection over Union (mIoU) of 88.68% and 88.72%. The experiment results are summarized in Tables 1 and 2. The images masks and the segmented images in Figure 2, shows the advantages of the proposed model over the UNet and Res-UNet models. The study attributes the success of Res-UNet++ to its adoption of elements including residual blocks, ASPP, squeeze and excitation blocks and attention blocks. These advancements have the potential to enhance practices by increasing accuracy reducing errors and potentially streamlining dental examinations. Despite its performance there are considerations to take into account when balancing performance enhancements, with computational complexity. This becomes more crucial when considering the training times resulting from higher parameter numbers underscoring the importance of understanding these trade-offs for actual use, in clinical settings.

ACKNOWLEDGMENT

The authors would like to acknowledge the College of Engineering at AL-Mustansiriyah university, Baghdad, Iraq, https://uomustansiriyah.edu.iq/ for their support in organizing this research.

REFERENCES

[1] Rajpoot, V., Dubey, R., Khan, S.S., Maheshwari, S., Dixit, A., Deo, A., Doohan, N.V. (2022). Orchard Boumans algorithm and MRF approach based on full threshold segmentation for dental X-ray images. Traitement du Signal, 39(2): 737-744. https://doi.org/10.18280/ts.390239

[2] Wang, C.W., Huang, C.T., Lee, J.H., Li, C.H., Chang, S.W., Siao, M.J., Lai, T.M., Ibragimov, B., Vrtovec, T., Ronneberger, O., Fischer, P., Cootes, T.F., Lindner, C. (2016). A benchmark for comparison of dental radiography analysis algorithms. Medical Image Analysis, 31: 63-76. https://doi.org/10.1016/j.media.2016.02.004

[3] Sivasankaran, P., Dhanaraj, K. (2022). A rapid advancing image segmentation approach in dental to predict cryst. Traitement du Signal, 39(1): 239-246. https://doi.org/10.18280/ts.390124.

[4] Brady, A., Laoide, R.Ó., McCarthy, P., McDermott, R. (2012). Discrepancy and error in radiology: Concepts, causes and consequences. The Ulster Medical Journal, 81(1): 3-9.

[5] Lin, P.L., Lai, Y.H., Huang, P.W. (2010). An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information. Pattern Recognition, 43(4): 1380-1392. https://doi.org/10.1016/j.patcog.2009.10.005

[6] Al-Behadili, H., Grumpe, A., Migdadi, L., Wöhler, C. (2016). Semi-supervised learning using incremental support vector machine and extreme value theory in gesture data. In 2016 UKSim-AMSS 18th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK, IEEE, pp. 184-189. https://doi.org/10.1109/UKSim.2016.5

[7] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, Proceedings, Part III 18. Springer International Publishing, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[8] Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis, 36: 61-78. https://doi.org/10.1016/j.media.2016.10.004

[9] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, Proceedings 4. Springer International Publishing, pp. 3-11. https://doi.org/10.1007/978-3-030-00889-5_1

[10] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., Maier-Hein, K.H. (2018). Nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv Preprint arXiv: 1809.10486. https://doi.org/10.48550/arXiv.1809.10486

[11] Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H.D. (2019). Resunet++: An advanced architecture for medical image segmentation. In 2019 IEEE International Symposium on Multimedia (ISM), Diego, CA, USA, IEEE, pp. 225-2255. https://doi.org/10.1109/ISM46123.2019.00049

[12] Abdi, A.H., Kasaei, S., Mehdizadeh, M. (2015). Automatic segmentation of mandible in panoramic x-ray. Journal of Medical Imaging, 2(4): 044003-044003. https://doi.org/10.1117/1.JMI.2.4.044003

[13] Zhang, Z., Liu, Q., Wang, Y. (2018). Road extraction by deep residual U-NET. IEEE Geoscience and Remote Sensing Letters, 15(5): 749-753. https://doi.org/10.1109/LGRS.2018.2802944

[14] Machado, L.F., Watanabe, P.C.A., Rodrigues, G.A., Junior, L.O.M. (2023). Deep learning for automatic mandible segmentation on dental panoramic x-ray images. Biomedical Physics & Engineering Express, 9(3): 035015. https://doi.org/10.1088/2057-1976/acb7f6

[15] Widyaningrum, R., Candradewi, I., Aji, N.R.A.S., Aulianisa, R. (2022). Comparison of Multi-Label U-Net and Mask R-CNN for panoramic radiograph segmentation to detect periodontitis. Imaging Science in Dentistry, 52(4): 383. https://doi.org/10.5624%2Fisd.20220105

[16] 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.
[17] Mahran, A.M.H., Hussein, W., Saber, S.E.D.M. (2023). Automatic teeth segmentation using attention U-Net. Preprints.org. https://doi.org/10.20944/preprints202306.1468.v2
[18] Harsh, P., Chakraborty, R., Tripathi, S., Sharma, K. (2021). Attention U-Net architecture for dental image segmentation. In 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, IEEE, pp. 1-5. https://doi.org/10.1109/CONIT51480.2021.9498422
[19] Biswas, M., Pramanik, R., Sen, S., Sinitca, A., Kaplun, D., Sarkar, R. (2023). Microstructural segmentation using a union of attention guided U-Net models with different color transformed images. Scientific Reports, 13(1): 5737. https://doi.org/10.1038/s41598-023-32318-9
[20] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv Preprint arXiv: 1804.03999. https://doi.org/10.48550/arXiv.1804.03999
[21] John, D., Zhang, C. (2022). An attention-based U-Net for detecting deforestation within satellite sensor imagery. International Journal of Applied Earth Observation and Geoinformation, 107: 102685. https://doi.org/10.1016/j.jag.2022.102685
[22] Karthik, R., Radhakrishnan, M., Rajalakshmi, R., Raymann, J. (2021). Delineation of ischemic lesion from brain MRI using attention gated fully convolutional network. Biomedical Engineering Letters, 11: 3-13. https://doi.org/10.1007/s13534-020-00178-1
[23] Dayı, B., Üzen, H., Çiçek, İ.B., Duman, Ş.B. (2023). A novel deep learning-based approach for segmentation of different type caries lesions on panoramic radiographs. Diagnostics, 13(2): 202. https://doi.org/10.3390/diagnostics13020202
[24] Nafi'iyah,, N., Fatichah, C., Herumurti, D., Astuti, E.R., Putra, R.H. (2023). MobileNetV2 ensemble segmentation for mandibular on panoramic radiography. International Journal OF Intelligent Engineering & System, 16(2): 546-548.
[25] Arora, S., Tripathy, S.K., Gupta, R., Srivastava, R. (2023). Exploiting multimodal CNN architecture for automated teeth segmentation on dental panoramic X-ray images. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 237(3): 395-405. https://doi.org/10.1177/09544119231157137
[26] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.
[27] Monti, R.P., Tootoonian, S., Cao, R. (2018). Avoiding degradation in deep feed-forward networks by phasing out skip-connections. In Artificial Neural Networks and Machine Learning-ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, Proceedings, Part III 27. Springer International Publishing, pp. 447-456.
[28] Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132-7141.
[29] Chen, L.C., Papandreou, G., Schroff, F., Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv Preprint arXiv: 1706.05587. https://doi.org/10.48550/arXiv.1706.05587
[30] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4): 834-848. https://doi.org/10.1109/TPAMI.2017.2699184
[31] Li, H., Xiong, P., An, J., Wang, L. (2018). Pyramid attention network for semantic segmentation. arXiv Preprint arXiv: 1805.10180. https://doi.org/10.48550/arXiv.1805.10180
[32] Panetta, K., Rajendran, R., Ramesh, A., Rao, S.P., Agaian, S. (2021). Tufts dental database: A multimodal panoramic x-ray dataset for benchmarking diagnostic systems. IEEE Journal of Biomedical and Health Informatics, 26(4): 1650-1659. https://doi.org/10.1109/JBHI.2021.3117575