



Article

A comparative study of deep learning models for dental segmentation in panoramic radiograph

Élisson da Silva Rocha¹  and Patricia Takako Endo^{1,*} ¹ Programa de Pós-Graduação em Engenharia da Computação, Universidade de Pernambuco; esr2@comp.poli.br, patricia.endo@upe.br

* Correspondence: patricia.endo@upe.br

Abstract: **Introduction:** Dental segmentation in panoramic radiograph has become very relevant in dentistry, since it allows health professionals to carry out their assessments more clearly and helps them to define the best possible treatment plan for their patients. **Objectives:** In this work, a comparative study is carried out with four segmentation algorithms (U-Net, DCU-Net, DoubleU-Net and Nano-Net) that are prominent in the medical literature on segmentation and we evaluate their results with the current state of the art of dental segmentation in panoramic radiograph. **Methods:** These algorithms were tested with a dataset consisting of 1,500 images, considering experiment scenarios with and without augmentation data. **Results:** DoubleU-Net was the model that presented the best results among the analyzed models, reaching 96.591% accuracy and 92.886% Dice using the dataset with data augmentation. Another model that stood out was Nano-Net using the dataset without data augmentation; this model achieved results close to that of the literature with only 235 thousand trainable parameters, while the literature model (TSASNet) contains 78 million. **Conclusions:** The results obtained in this work are satisfactory and present paths for a better and more effective dental segmentation process.

Keywords: Dental segmentation; deep learning; panoramic radiograph

1. Introduction

Radiography is a widely used examination in the field of dentistry, and aims to help the dentist in the diagnosis of some diseases that are more difficult to detect in a routine appointment, such as periapical lesions, bone anomalies, among others, especially in their initial state [1]. The radiograph can be obtained in two ways: either intraorally, which are the radiographs performed in the inner part of the patient's mouth that result in specific parts of the dentition; or extraorally, which are extrabuccal images that contain the entire oral region of the patient. Of the extraoral radiographs, the best known are panoramic radiographs.

Panoramic radiographs have advantages over other types of radiographs because they cause less exposure to radiation, provide greater comfort for the patient, are faster and easier to perform and have a large field of view. However, they present disadvantages such as having a complex structure and low-resolution images, which can affect the interpretation of the results by the health professional [2]. Another factor that interferes with the interpretation of the results is human subjectivity, since less experienced dentists may present more difficulties in diagnosing their patients using this type of examination [3].

When evaluating panoramic radiographs, a possible task that is susceptible to human failure is the identification of teeth, their exact shapes and boundaries. This process, called dental segmentation, is of paramount importance for the recognition of the visual dental pattern [4]. Today, dental segmentation plays a crucial role in supporting diagnosis and treatment plans [5].

With advances in artificial intelligence (AI), such as machine learning and deep learning, computer-aided diagnostic systems have been widely used in various medical areas, such as automated image-based diagnosis to detect lung and brain lesions [6–8], breast cancer in mammography images [9], segmentation of brain lesions [10] and segmentation



Citation: da Silva Rocha, E.; Endo, P.T. Title. *Preprints* **2022**, *1*, 0.
<https://doi.org/>

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

of pulmonary nodules of various densities [11], among others. In dentistry, these techniques have also been applied in various segments. For example, the authors [12] used a convolutional neural network (CNN) to diagnose dental caries in premolars, molars, or both, using periapical radiographs. In [13], a CNN was used to detect apical lesions in panoramic-rays. In [14], deep learning was used to segment the mandibular canal using computed tomography volumes. While in [5], segmentation of teeth in panoramic radiographs was performed. Although the results of previous research on AI have been extremely promising, the studies are still preliminary.

Starting from the idea that AI can aid dentists in identifying and segmenting teeth faster with the possibility of minimizing errors associated with human fatigue [3], this study aims to perform a comparison of segmentation of computational vision models and investigate their performance in segmenting panoramic radiograph images.

This paper is divided into five sections. After the introduction (Section I), Section II provides an overview of related work in the field of segmentation in panoramic radiograph images. Section III describes the materials and methods of the work, such as the dataset used, the proposed models, applied loss function and evaluation metrics. Section IV describes the experiments, presents the results and discussion. Finally, in Section V, conclusions and presentation of future work.

2. Related works

There are some works that already use panoramic radiograph image segmentation. The authors in [1] propose a fully automatic approach using a statistical model together with a deep learning model. The deep learning model generates a mask of the tooth area and the statistical model segments and labels the 28 teeth. The proposed model was evaluated in 14 test images with average accuracy and recall of 79.0% and 82.7%, respectively, and Dice of 74.4%. The authors in [15] proposed a region-based convolutional neural network (R-CNN) to detect and locate dental structures using 846 dental notes from 30 panoramic radiographs for training and 20 radiographs for validation and testing. The results obtained were 85.8% accuracy and 89.3% recall and an average intersection-over-union (IoU) of 87.7%.

Silva and Oliveira [16] conducted a literature review on segmentation methods applied in dental images. In the study, they noticed a gap in the panoramic radiograph images and, to fill this gap, the authors presented a new data set containing 1,500 panoramic radiograph images with great variability and proposed a deep learning model to segment images in this set of images. The initial results obtained were 84% precision, 76% recall and 92% accuracy. Subsequently, using this same database, the authors in [5] proposed the TSASNet (two-stage attention segmentation network) which is divided into two stages. The first stage contains an attention network, which can be global and/or local, to obtain preliminary information on the radiograph and, in the second stage, in fact, a segmentation network. This method was evaluated and reached 96.94% accuracy, 94.97% precision, 93.77% recall and 92.72% Dice. The authors in [17] proposed a fully convolutional neural network (FCN) model based on the U-Net architecture to improve segmentation performance. The final result was 93.4%, 93.5% and 93.7% of Dice, precision and recall, respectively. Other models using ensemble strategies were proposed, in which the best model obtained results of 93.6%, 93.3% and 94.3% of Dice, precision and recall, respectively.

This work is very similar to the works of [5,17], but it uses other segmentation models that are gaining prominence in the literature with other medical segmentation datasets [18–21].

3. Materials and methods

3.1. Dataset

In this work, we used a dataset created by Silva and Oliveira [16] and made publicly available. This dataset is the largest freely available dataset for research and contains 1,500 panoramic radiograph images divided into 10 categories. The images were acquired at the Imaging Diagnostic Center of the Universidade Estadual do Sudoeste da Bahia (UESB)

using an X-ray camera model ORTHOPHOS XG 5/XG 5 DS/Ceph, manufactured by Sirona Dental Systems GmbH (please see [16] for more details such as: focus-sensor distance, scale of images, widescreen sensor resolution). The teeth mask was defined globally in each buccal image, rather than per tooth. The images were classified according to the variety of structural features of the teeth. This information is shown in detail in Table 1. We selected 80% of the images for the training set, 10% for the validation set and 10% for the test set.

Table 1. Patient categories

Category	1	2	3	4	5	6	7	8	9	10
32 teeth	Yes	Yes	Yes	Yes	+32	-	No	No	No	No
Filling	Yes	Yes	No	No	-	-	Yes	Yes	No	No
Braces	Yes	No	Yes	No	-	-	Yes	No	Yes	No
Dental implant	No	No	No	No	No	Yes	No	No	No	No

Images have been resized from 1,127 px height by 1,991 px width (originally) to 256 px height by 256 px width. We used the panoramic radiograph images in RGB colors, that is, with three color channels, and the masks only in gray scale, that is, one color channel. Before passing the data to the models, we normalized each pixel that initially ranged from 0 to 255 to 0 to 1.

3.2. Data augmentation

This work applied data augmentation in the training set to increase data diversity and test whether the models present a good generalization. For the data augmentation, two techniques were used: random rotation and horizontal inversion. With this, the training data set has increased from 1,200 to 3,600 images. The validation and test set did not change.

3.3. Segmentation models

3.3.1. U-Net

U-Net is a semantic segmentation model that uses convolutional layers, and was proposed by Ronnerberger et al. [18]. The architecture of this network is symmetrical and consists of an encoder path (encoding) and a decoder path (decoding). The encoder follows the typical architecture of a convolutional network, which is used to extract features from images. The encoder consists of four convolution blocks. The decoder is used to build the segmentation map based on the encoder's features. Since this architecture is symmetrical, as stated earlier, the decoder is also formed by four blocks. At the end, a convolution operation is performed to generate the final segmentation. With this, the U-Net reaches the end with 23 convolutional layers, as shown in Figure 1. For more details on the model and the convolution blocks in the encoder and decoder, please see [18].

3.3.2. DCU-Net

The DCU-Net (Dual-Channel U-Net) is a model that is proposed by Lou et al. [19], which is based on the U-Net architecture and MultiResUnet architecture (MultiResUnet is another proposed model that is also described in the paper by [19]). Based on the comparisons made between U-Net and MultiResUnet, authors noted that MultiResUnet obtained better outputs than the U-Net model, but in some cases, MultiResUnet may not work well in relation to feature extraction. With this, the author developed a new architecture, DC-Unet, in which a two-channel block is used (Dual-Channel block) with formations used in U-Net and MultiResUnet.

The U-Net is represented by the DC blocks, where each block contains 3 convolutional layers and the MultiResUnet is represented by the ResPath. Each ResPath also contains a block of convolutional, being ResPath 1 with 8 convolutional layers, ResPath 2 with 6 convolutional layers, Respath 3 with 4 convolutional layers and finally ResPath 4 with 2 convolutional layers (see Figure 2). For more details on the model, please see [19].

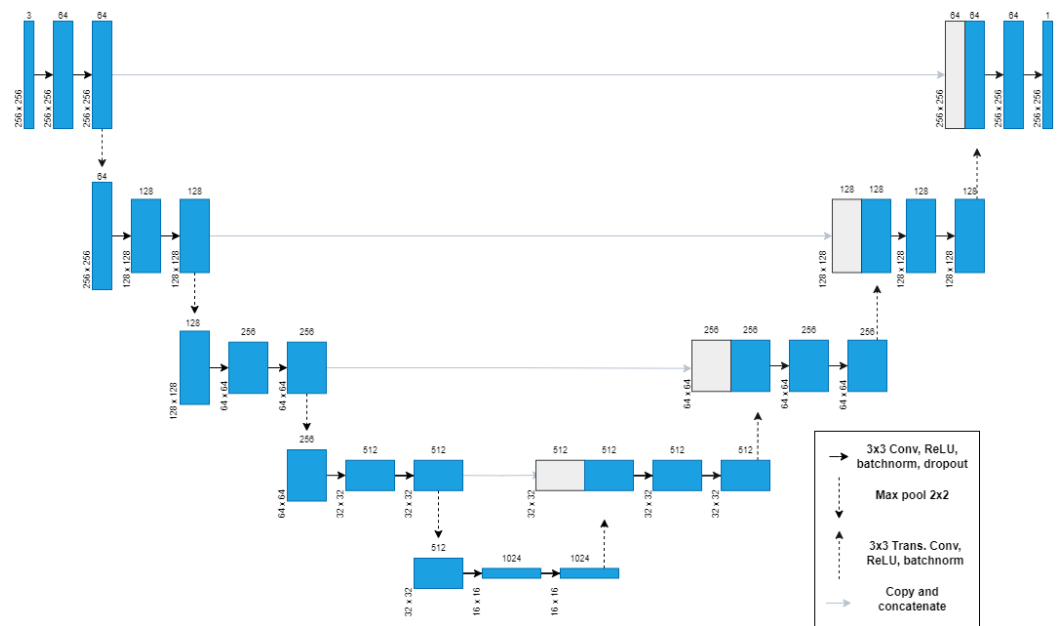


Figure 1. U-Net architecture (adapted from [18])

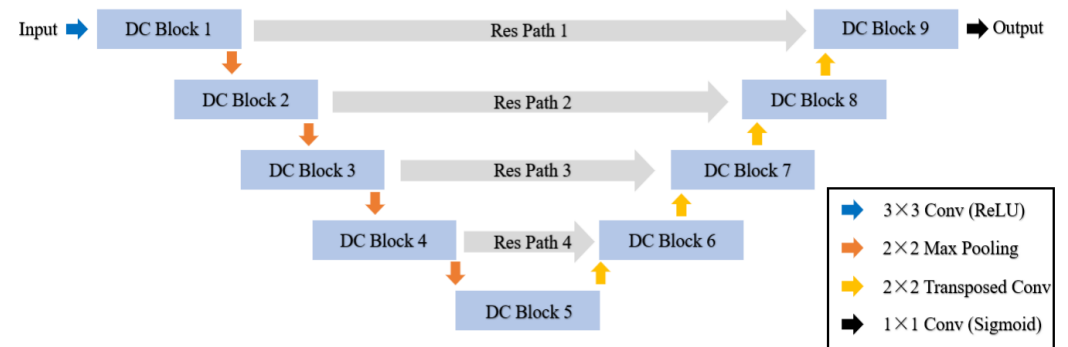


Figure 2. DCU-Net architecture (adapted from [19])

3.3.3. DoubleU-NET

The DoubleU-Net model was proposed in the paper by [20] and was validated using four medical image segmentation databases. The proposed architecture of DoubleU-Net¹ (Figure 3) consists of two large networks.

The first network (NETWORK 1) starts with a VGG-19 as an encoder, which passes through an Atrous Spatial Pyramid Pooling (ASPP), which aims to capture contextual information within the network, and is followed by four decoder blocks to reach the output of the first network (output 1). Next, we multiply the input image by the output of NETWORK 1 (output 1), which acts as input for the second model. The second model consists of four encoder blocks, which passes to the ASPP and follows on to the four decoder blocks that generate the output of the second network (output 2). For more details on the model, see the paper by [20].

3.3.4. Nano-NET

The Nano-Net model was proposed in the paper by [21], and its objective is to perform accurate segmentation in real time. As a result, Nano-Net uses the encoder-decoder model (see Figure 4), but in the encoder a pre-trained model is used so that the convergence of the model is faster. The pre-trained model used in Nano-Net was MobileNetV2 [22]. The

¹ code: <https://github.com/DebeshJha/2020-CBMS-DoubleU-Net>

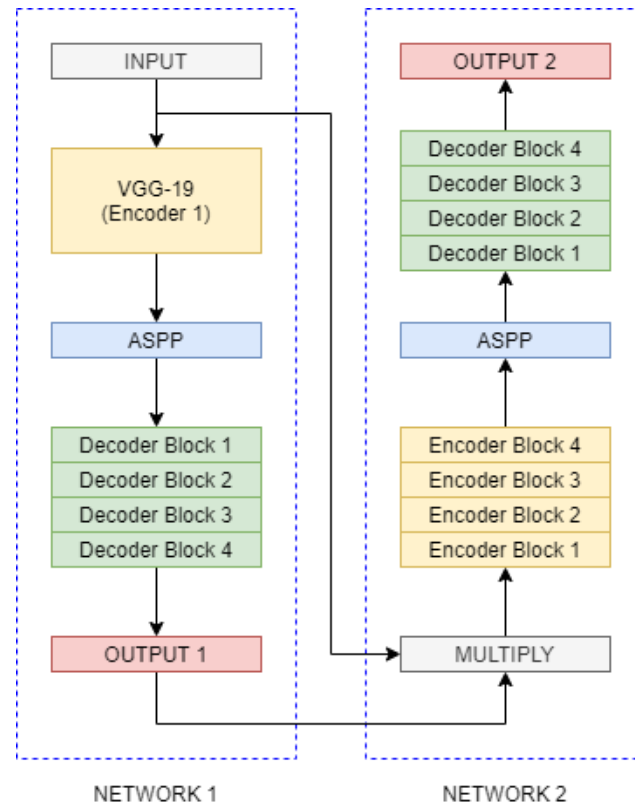


Figure 3. DoubleU-Net architecture – adapted from [20]

decoder, in turn, is formed by three blocks. In the paper by [21], three configurations of Nano-Net are proposed (Nano-Net-A, Nano-Net-B and Nano-Net-C), the difference between the models is the number of feature channels used, where A is 32, 64 and 128; B is 32, 64, and 96; and C is 16, 24, and 32. In the present paper, Nano-Net-A was used because it was the model that obtained the best results. For more details about Nano-Net, see the paper by [21].

3.4. Loss function

The loss function used in the segmentation models of this paper was a weighted equation, which is described as follows:

$$L = \lambda_1 * L_{CE} + \lambda_2 * L_{IoU}$$

Where the L_{CE} is the cross entropy (CE) function multiplied by weight 1 and the L_{IoU} is a loss function as proposed by [23], which calculates the intersection-over-union (IoU) region multiplied by 2. 1 and 2 are 0.4 and 0.6, respectively. The calculation of L_{IoU} is shown below:

$$L_{IoU} = \frac{|P \cap G|}{|P \cup G|}$$

In which P represents the predicted area and G represents the actual area (ground truth).

3.5. Evaluation criterion

In the field of medical image segmentation, the Dice coefficient [24] is widely used to evaluate segmentation models, and is defined as follows:

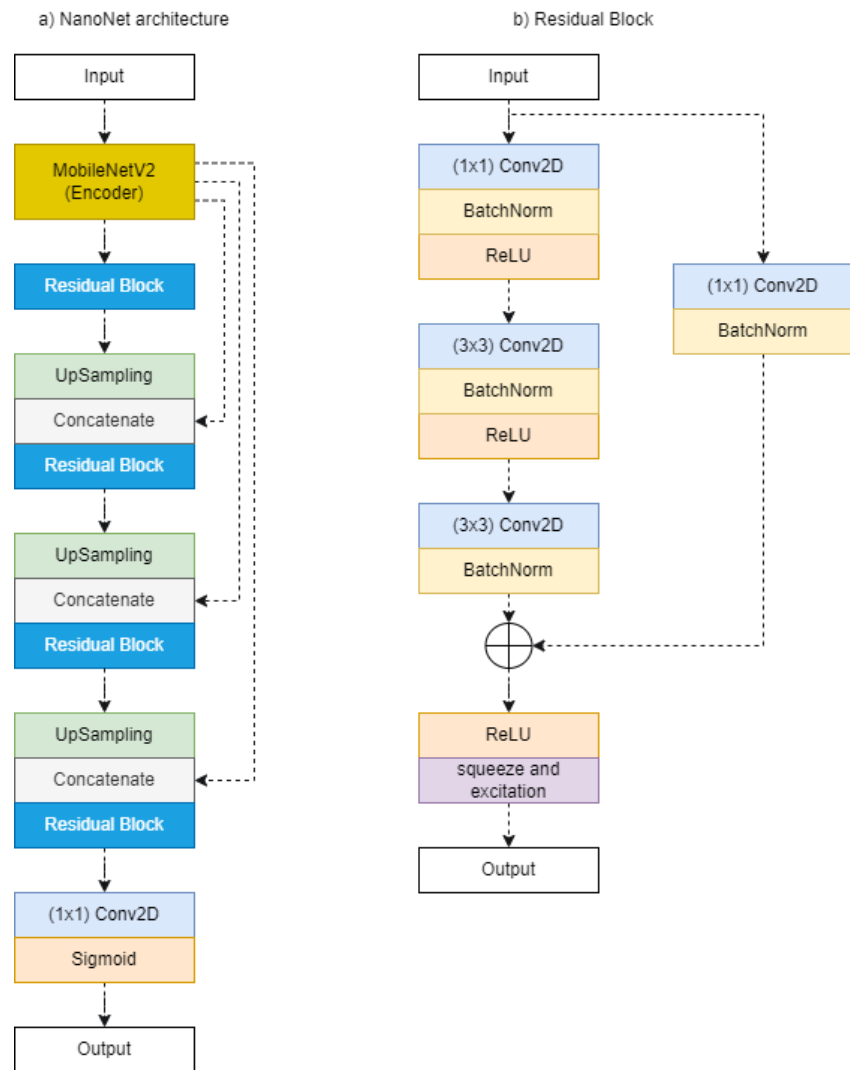


Figure 4. NanoNet architecture (adapted from [21])

$$dice = \frac{2|P \cap G|}{|P| + |G|}$$

in which P is the predicted area and G is the real area (ground truth). The value of the Dice coefficient varies between 0 and 1, in which 0 is the lowest possible accuracy and 1 the highest possible accuracy.

Segmentation models can also be evaluated pixel by pixel, and for this type of evaluation, the following metrics can be used:

$$acuracia = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precisao = \frac{TP}{TP + FP}$$

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

in which:

- TP: are the amount of pixels correctly predicted from the tooth area;
- TN: are the amount of pixels correctly predicted from the background area;

- FP: are the amount of incorrectly predicted pixels of the tooth area;
- FN: are the amount of pixels incorrectly predicted from the background area.

Finally, we add the amount of trainable parameters (param) in the segmentation model for comparison between the models.

4. Experimental results and discussion

4.1. Comparison setting

The results are divided into three subsections: segmentation models using a dataset with data augmentation (Subsection 4.2), segmentation models using a dataset without data augmentation (Subsection 4.3) and the third step of comparing the models with the state of the art (Subsection 4.4).

In the training settings of the models, the ADAM optimizer with the learning rate of 0.0001 was used, the loss function described in Section 3.4. The models were trained with 30 epochs. Only Nano-Net was trained with 50 because it is a model that uses a pre-trained layer, the training process was faster and its convergence used 50 epochs on average. The other models used 30 epochs based on empirical tests, see Table 2. The batch size was different for datasets, with data augmentation and without data augmentation. Models trained with a dataset without data augmentation used a batch size of size 4. The models trained with the dataset that includes the data augmentation had the following sizes: 32 for the U-Net and Nano-Net, 16 for the DoubleU-Net and 10 for the DCU-Net. The batch size for the U-Net and Nano-Net were not affected, but the DoubleU-Net and DCU-Net had their batch size reduced due to limitations of the training machine.

Table 2. Model training settings

Model	Batch_size with data augmentation	Batch_size without data augmentation	Epochs
U-Net	32	4	30
DCU-Net	10	4	30
DoubleU-Net	16	4	30
Nano-Net	32	4	50

All models were trained and tested using the Google Collaboratory platform. This platform provided a machine with an nVidia195Tesla P100 video card and 27.3GB of RAM.

4.2. Results using data augmentation

The results of the four metrics (Dice, accuracy, precision and recall) for the models trained with the dataset containing the data augmentation can be seen in Table 3.

Table 3. Result (in %) of segmentation models using data augmentation

Model	Dice (%)	Accuracy (%)	Precision (%)	Recall (%)	Parameter
U-Net	91.033	95.966	95.342	87.122	31,031,745
DCU-Net	91.616	96.208	96.107	87.547	10,069,928
DoubleU-NET	92.886	96.591	93.095	92.705	29,264,930
Nano-Net	89.855	95.576	96.326	84.222	235,425

The Dice metric obtained 92.886% with the DoubleU-Net model, followed by DCU-Net, U-Net and, finally, Nano-Net with 89.855%. For accuracy and recall, the order remained the same, with DoubleU-Net with 96.591% and 92.705%, respectively. Nano-Net was 95.576% and 84.222%. For precision, the Nano-Net Model obtained 96.326%, followed by DCU-NET, U-NET and DoubleU-Net with 93.095%.

DoubleU-net performed better in three of the four metrics analyzed, falling behind Nano-Net only in the precision metric, 3% less. Nano-Net obtained the worst recall result, well behind the other models, and reached an 8% difference compared to DoubleU-Net.

Analyzing the number of parameters trained by the model, Nano-Net used 235 thousand while U-Net reached 31 million parameters.

Figure 5 shows three examples of images tested by the four models, comparing them with the original image and the ground truth.

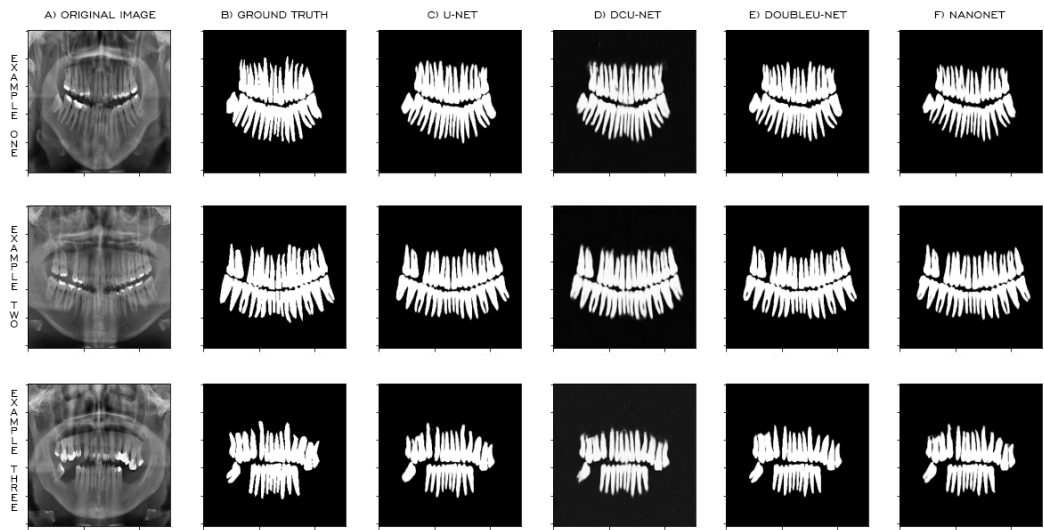


Figure 5. Example of three panoramic radiograph images tested by segmentation models with data augmentation

In general, the models presented outputs that were very close to ground truth, though with a few observations for some models: the DCU-Net model was the one that presented more failures in its outputs, as can be seen more easily in Example 1 and Example 3. The U-Net model was the model that presented some flaws in detail, especially in the roots of the teeth. As shown in Example 1 and 3, the circled areas have root detection failures. The DoubleU-Net and Nano-Net models were the ones that presented the best details in the roots.

4.3. Results without using data augmentation

The results of the four metrics (Dice, accuracy, precision and recall) for models trained with the dataset without data augmentation can be seen in Table 4.

Table 4. Segmentation models without augmentation data

Model	Dice (%)	Accuracy (%)	Precision (%)	Recall (%)	Parameter (%)
U-Net	91.681	96.191	94.898	88.700	31,031,745
DCU-Net	91.451	96.123	95.390	87.846	10,069,928
DoubleU-Net	92.695	96.552	94.184	91.283	29,264,930
Nano-Net	91.739	96.173	93.783	89.815	235,425

The DoubleU-Net model obtained 92.695% of Dice followed by Nano-Net, U-Net and, finally, the DCU-Net with 91.451%. This same order was seen for the recall metric, with 91.283% for DoubleU-Net and 87.846% for DCU-Net. Regarding accuracy, the order was changed to DoubleU-net, Nano-Net, DCU-Net and U-Net. With regard to precision, DCU-Net presented the best result with 95.390%, followed by U-NET, DoubleU-Net and Nano-Net with 93.783%.

DoubleU-Net continues to show great prominence, as it also obtained the best results in three of the four metrics analyzed in this experiment. However, another model that stood out was Nano-Net, which for training without data augmentation presented results close to DoubleU-Net with fewer trainable parameters.

Figure 6 shows the same three examples of images tested in the previous experiment (with data augmentation) with the outputs of the models that were trained without data augmentation.

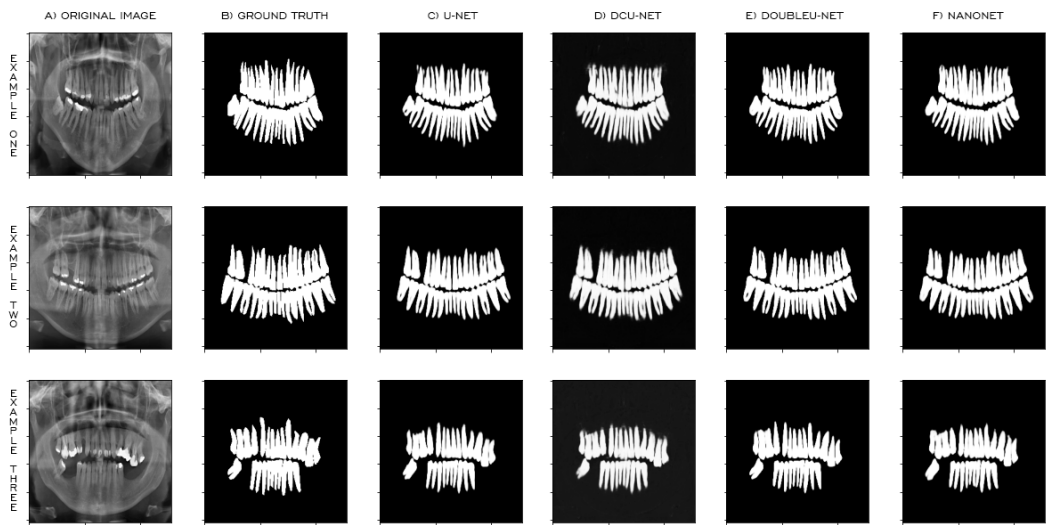


Figure 6. Example of three panoramic radiograph images tested by segmentation models without data augmentation

The conclusions are the same as the experiment with data augmentation, but with some negative points for the DCU-Net Model, The U-Net showed some flaws in the details of the roots, the Nano-Net also showed some flaws and the DoubleU-Net was the model with more satisfactory performance.

4.4. Comparison with the state of the art

From the results described above, we highlight the results of the DoubleU-Net and Nano-Net models, both for training using data augmentation, as well as for training without data augmentation. Thus, this section presents a comparison of these models with others in the literature [5,17].

Table 5 presents the results obtained by the models proposed by [5] and [17], using the same dataset tested in this work.

Table 5. Comparison with literature models

Model	Dice (%)	Accuracy (%)	Precision (%)	Recall (%)	Parameter
TSASNET [5]	92.72	96.94	94.97	93.77	78,270,000
U-NET (Ensemble) [17]	93.58	95.17	93.69	93.91	-
DoubleU-NET w/data augmentation	92.886	96.591	93.095	92.705	29,264,930
Nano-Net w/data augmentation	89.855	95.576	96.326	84.222	235,425
DoubleU-NET wo/data augmentation	92.695	96.552	94.184	91.283	29,264,930
Nano-Net wo/data augmentation	91.739	96.173	93.783	89.815	235,425

The U-Net* model proposed by [17] obtained the best Dice coefficient with 93.58%, followed by DoubleU-Net with data augmentation with 92.886%. TSASNet obtained the best accuracy, with 96.94%, While U-Net*[17] with 95.17% had the worst. Precision was a highlight of Nano-Net (True) with 96.326%, but obtained the worst recall. For Recall, the best result was U-Net*[17] with 93.91%, followed by 93.77% of TSASNet and 92.705% for DoubleU-Net (True).

Regarding the number of parameters, TSASNet requires more than 78 million trainable parameters, which shows how costly this architecture can be for training. DoubleU-Net managed to lower this number to 29 million and the greatest prominence is for Nano-Net, which uses 235 thousand. The number of parameters trained by the U-Net architecture*[17] was not presented, but by analyzing the work of [17], their proposal uses an ensemble of two U-Net models, in which each model corresponds to approximately 31 million parameters. Thus, their proposal is expected to be close to 62 million trainable parameters.

It is observed that U-Net* [17] has better results in two metrics, while TSASNet has 1 and Nano-Net (True) with another. With the analysis, we see DoubleU-Net (True) has competitive results in almost all metrics and Nano-Net is competitive even with the number of trainable parameters being far below any other model.

5. Conclusion and future work

Panoramic radiographs can be a very effective tool to the dentist support to diagnosis and to define a treatment plan. The use of segmentation models to detect teeth and their exact limits can be of paramount importance for eliminating a task that is quite susceptible to human failure. In this work, we analyzed four models that are being used in the literature for other purposes, in order to investigate their performance using a panoramic radiograph database.

The results obtained in this work are satisfactory and present paths for a better and more effective dental segmentation process. Among the four models tested, two achieved good results. DoubleU-Net presented itself as a competitive model with the current literature, with results very close to the cited works, but with a number of parameters trained by the model being well below, going from 78 million to 29 million.

Nano-Net, in turn, attracted attention for its low number of parameters trained by the model while still achieving interesting results of accuracy, precision, recall and Dice; thus showing that it is a model that can be exploited for better results.

A limitation of this work is in relation to the database used. As described in the Section 3.1, all images were recorded by the same machine, so that the deep learning models can have this performance only for the parameters used by that given machine.

As the next steps of this research, it is planned to further evaluate the learning process of DoubleU-Net and Nano-Net to mold the creation of a new segmentation model that can better detail and evaluate dental panoramic radiograph images. Another way is to perform an analysis of the images generated as the output of each model in order to propose an improvement in ground truth. When analyzing the database used, it is noticed that some dental roots are square rather than more rounded. This can be a characteristic of human fatigue, since we are dealing with a dataset that is too large for the area. Another step of this research is to look for datasets that contain more variable data, obtained from different machines with different parameters for a greater generalization of the trained models.

Informed Consent Statement: Not applicable.

Acknowledgments: Authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE); and Universidade de Pernambuco (UPE), an entity of the Government of the State of Pernambuco focused on the promotion of Teaching, Research and Extension.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wirtz, A.; Mirashi, S.G.; Wesarg, S. Automatic teeth segmentation in panoramic X-ray images using a coupled shape model in combination with a neural network. *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 712–719.
2. Kim, J.; Lee, H.S.; Song, I.S.; Jung, K.H. DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. *Scientific reports* **2019**, *9*, 1–9.
3. Leite, A.F.; Van Gerven, A.; Willems, H.; Beznik, T.; Lahoud, P.; Gaëta-Araujo, H.; Vranckx, M.; Jacobs, R. Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs. *Clinical oral investigations* **2021**, *25*, 2257–2267.
4. Barboza, E.B.; Marana, A.N.; Oliveira, D.T. Semiautomatic dental recognition using a graph-based segmentation algorithm and teeth shapes features. 2012 5th IAPR International Conference on Biometrics (ICB). IEEE, 2012, pp. 348–353.
5. Zhao, Y.; Li, P.; Gao, C.; Liu, Y.; Chen, Q.; Yang, F.; Meng, D. TSASNet: Tooth segmentation on dental panoramic X-ray images by Two-Stage Attention Segmentation Network. *Knowledge-Based Systems* **2020**, *206*, 106338.
6. Akkus, Z.; Galimzianova, A.; Hoogi, A.; Rubin, D.L.; Erickson, B.J. Deep learning for brain MRI segmentation: state of the art and future directions. *Journal of digital imaging* **2017**, *30*, 449–459.

7. Song, Q.; Zhao, L.; Luo, X.; Dou, X. Using deep learning for classification of lung nodules on computed tomography images. *Journal of healthcare engineering* **2017**, 2017.
8. Wang, H.; Zhou, Z.; Li, Y.; Chen, Z.; Lu, P.; Wang, W.; Liu, W.; Yu, L. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images. *EJNMMI research* **2017**, 7, 11.
9. Becker, A.S.; Marcon, M.; Ghafoor, S.; Wurnig, M.C.; Frauenfelder, T.; Boss, A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. *Investigative radiology* **2017**, 52, 434–440.
10. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis* **2017**, 36, 61–78.
11. Kubota, T.; Jerebko, A.K.; Dewan, M.; Salganicoff, M.; Krishnan, A. Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models. *Medical Image Analysis* **2011**, 15, 133–154.
12. Lee, J.H.; Kim, D.H.; Jeong, S.N.; Choi, S.H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of dentistry* **2018**, 77, 106–111.
13. Ekert, T.; Krois, J.; Meinhold, L.; Elhennawy, K.; Emara, R.; Golla, T.; Schwendicke, F. Deep learning for the radiographic detection of apical lesions. *Journal of endodontics* **2019**, 45, 917–922.
14. Jaskari, J.; Sahlsten, J.; Järnstedt, J.; Mehtonen, H.; Karhu, K.; Sundqvist, O.; Hietanen, A.; Varjonen, V.; Mattila, V.; Kaski, K. Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes. *Scientific reports* **2020**, 10, 1–8.
15. Lee, J.H.; Han, S.S.; Kim, Y.H.; Lee, C.; Kim, I. 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* **2020**, 129, 635–642.
16. Silva, G.; Oliveira, L.; Pithon, M. Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Systems with Applications* **2018**, 107, 15–31.
17. Koch, T.L.; Perslev, M.; Igel, C.; Brandt, S.S. Accurate segmentation of dental panoramic radiographs with U-Nets. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, pp. 15–19.
18. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
19. Lou, A.; Guan, S.; Loew, M.H. DC-UNet: rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation. Medical Imaging 2021: Image Processing. International Society for Optics and Photonics, 2021, Vol. 11596, p. 115962T.
20. Jha, D.; Riegler, M.A.; Johansen, D.; Halvorsen, P.; Johansen, H.D. Doubleu-net: A deep convolutional neural network for medical image segmentation. 2020 IEEE 33rd International symposium on computer-based medical systems (CBMS). IEEE, 2020, pp. 558–564.
21. Jha, D.; Tomar, N.K.; Ali, S.; Riegler, M.A.; Johansen, H.D.; Johansen, D.; de Lange, T.; Halvorsen, P. NanoNet: Real-Time Polyp Segmentation in Video Capsule Endoscopy and Colonoscopy. *arXiv preprint arXiv:2104.11138* **2021**.
22. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.
23. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. Proceedings of the 24th ACM international conference on Multimedia, 2016, pp. 516–520.
24. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 fourth international conference on 3D vision (3DV). IEEE, 2016, pp. 565–571.