Original Article - Swiss J. Rad. Nucl. Med. (2024) 12 (Vol.1): 5-14; https://doi.org/10.59667/sjoranm.v12i1.18

Automatic Joint Teeth Segmentation in Panoramic Dental Images using Mask Recurrent Convolutional Neural Networks with Residual Feature Extraction: Can it be useful in Oral Cancer Diagnosis and Management?

Raghavendra H. Bhalerao[1], Abhijeet Ashok Salunke[2,3], Shristi Sharan[1], Kamlesh Kumar[1], Priyank Rathod[2], Prince Kumar[1], Manish Chaturvedi[1], Nandlal Bharwani[2], Krupa Shah[1], Dhruv Patel[2], Keval Patel[2], Vikas Warikoo[2], Manisha Abhijeet Salunke[4], Shashank Pandya[2]

[2]Department of Surgical Oncology, The Gujarat Cancer and Research Institute, Ahemdabad, India
[1]Dept of Electrical Engineering, IITRAM, Ahemdabad, India
[4]Dental Surgeon, Ahmedabad, India

Swiss Journal of Radiology and Nuclear Medicine - www.sjoranm.com - SJORANM GmbH - CH-3072 Ostermundigen bei Bern - Switzerland

Abstract

Introduction

Panoramic dental images gives an in-depth understanding of the tooth structure, both lower and upper jaws, and surrounding structures throughout the cavity in our mouth. The Panoramic dental images provided have significance for dental diagnostics since they aid in the detection of an array of dental disorders, including oral cancer. We propose a novel approach to automatic joint teeth segmentation using the pioneer Mask Recurrent Convolutional Neural Network (MRCNN) model for dental image segmentation.

Material and Methods

In this study, a sequence of residual blocks are used to construct a 62-layer feature extraction network in lieu of ResNet50/101 in MRCNN. To evaluate the efficacy of our method, the UFBA-UESC and Tufts dental image dataset (2500 panoramic dental x-rays) were utilised. 252 x-rays were used in test set, rest of the x-rays were utilised as training (1800 images) and validation datasets (448 images) in ratio of 8:2 of the modified MRCNN model.

Results

Modified MRCNN achieved the final training and validation accuracies as 99.67% and 98.94%, respectively. The achieved accuracy of Dice coefficient (97.8%), Intersection over Union, (98.67%), and Pixel Accuracy (96.53%) respectively over the whole dataset. We also compare the performance of proposed model and other well established networks such as FPN, UNet, PSPNet, and DeepLabV3. The Modified MRCNN provides better results segmenting any two teeth which are close to each other.

Conclusion

Our proposed method will serve as a valuable tool for automatic segmentation of individual teeth for medical management. This current method leads to higher accuracy and precision. Segmented images can be used to evaluate periodic changes, providing valuable data for assessing the progression of oral cancer and the efficacy of management. Future research should focus on developing less complex, lightweight, and faster vision models while maintaining high accuracy.

Keywords: *Medical image segmentation, Image segmentation using MRCNN Medical Imaging, Teeth segmentation, AI in Oral Cancer, CNN in Oral Cancer*

[3]Corresponding author: Abhijeet Ashok Salunke - received: 19.09.2024 - published: 30.09.2024

Introduction

Panoramic dental images give an in-depth understanding of the tooth structure, both lower and upper jaws, and surrounding structures throughout the cavity in our mouth. The Panoramic dental images provided have significance for dental diagnostics since they aid in the detection of an array of dental disorders, including oral cancer. Panoramic radiographs are useful tools as they allow to evaluate the location, form, and structure of each tooth and confirm or reject a specific diagnosis such as infection, fracture, tooth loss, or simply identifying any previous dental treatment. The limitations of image capturing methods, poor contrast balance, limited resolution, and the presence of high amplitude noise in panoramic X-rays can make interpretation difficult and lead to diagnostic errors. As a result, providing automated analysis of panoramic X-ray ages has become crucial in the dentistry community since it can increase diagnostic accuracy while reducing screening time and hence medical costs. Medical image segmentation is one such tool for health care professionals, which entails identifying each picture pixel as a potential object of interest.

Literature review suggests several unsupervised pixel-wise segmentation algorithms for teeth have been developed, typically using intra-oral imaging (periapical or bitewing radiography). Silva et al published an intriguing overview of ten classic approaches to dental imaging segmentation (1). There have been very few investigations on panoramic X-ray images, using region-based, threshold-based, and boundary-based techniques (2-6). Silva et al. developed the UFBA-UESC Dental images Dataset, which contains 1500 panoramic X-ray pictures, and utilised it to test the aforementioned methodologies (1).

However, researchers discovered that typical segmentation processes failed to completely separate the teeth from surrounding bone parts in the oral cavity. The published studies, motivated by their recent success on various computer vision tasks, proposed Convolutional Neural Networks (CNN) for tooth segmentation. Silva et al suggested utilising deep learning to separate the whole tooth arch from the background (1). Jader et al. used a MRCNN method to perform instance segmentation on a modified version of the previously described dataset to explore tooth segmentation and detection (6,7). Koch et al used the pioneer U-Net to train a semantic segmentation network using a patching scheme method and found excellent results in recognising the foreground class (presence of teeth) (8,9).

Oral cancer often manifests with changes in the structure and appearance of teeth and surrounding tissues. Automated teeth segmentation using Mask RCNNs can assist in the early detection of these changes by enhancing image analysis, consistency and efficiency and in monitoring disease progression. Segmented images can be used to monitor changes over time, providing valuable data for assessing the progression of oral cancer and the effectiveness of treatments.

We propose a novel approach to automatic joint teeth segmentation using the pioneer Mask Recurrent Convolutional Neural Network (MR CNN) model for dental image segmentation. We propose an enhanced MRCNN model, in an attempt to enhance detection accuracy in images with dense objects and a complex backdrop. This study also evaluates the performance of the prodded model and models that include FPN, UNet, UNet++, PSPNet, DeepLabV3, and DeepLabV3+.

Material and Methods

A) Data

Datasets Ivision LAB (1) DNS Panoramic x-ray dataset, also known as the UFBA-UESC Dental images Dataset, has been utilised for this research. This dataset comprises 1500 images of 1991 x 1127 pixels in size that have been tagged with teeth numbering (FDI notation, i.e. location label) in the COCO format. The images are classified into eight groups based on the presence or absence of all teeth, as well as the presence or absence of restorations and appliances. We have also used another public dataset called Tufts Dental Database (19). This dataset comprises 1000 panoramic dental radiographs with annotations. We have merged both the datasets for training and testing of our model, therefore, in total we have used 2500 panoramic x-ray images for this study.

The available dataset is divided into three parts for training, cross validation, and testing of the modified Mask R-CNN model. In all, 252 x-rays were preserved to construct the test set, and rest of the x-rays were settled as training and validation datasets in approximate ratio of 8:2, which implies 1800 images for training and 448 for four-fold cross-validation (112 per fold) of the modified

MRCNN model. However, for experimental purposes, we tried different proportions of the whole dataset for allocation of training, validation, and testing datasets. And the proportion mentioned above is the best fit ratio of the three categories which is being used to present the final results of this study.

B) MRCNN Architecture

Our architecture is based on the Mask-RCNN model, which is frequently used for medical image segmentation because of its ease of application and significantly high performance while requiring comparably a little amount of training data.

this two-stage detection method (11-14). The likelihood that a target object is present in an anchor box is then represented by mapping each sample with four coordinate values and a probability value. Finally, RPN unites coordinate regression and binary classification losses. The FCN is a groundbreaking algorithm semantic segmentation that continues to serve as a model for the majority of contemporary methods (15). Designing network as a collection of convolutional layers with differentiable down-sampling (convolution) and up-sampling (transpose convolution) inside the network is the fundamental concept of the FCN (16-18).

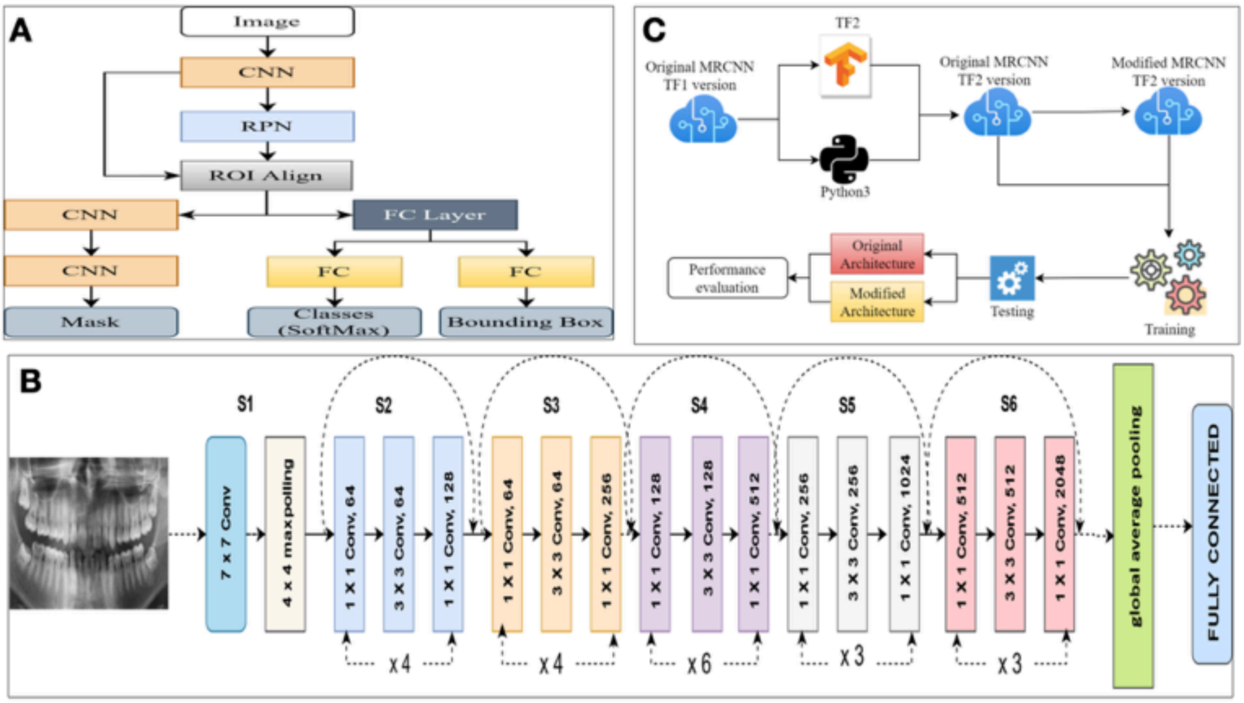

Figure 1
A: Training flow of original MRCNN architecture for image segmentation. From top to bottom: Panoramic x-ray (Image); Convolutional Neural Network (CNN); Region Proposal Network (RPN); Region of Interest (ROI); ROI align; Fully connected layers (FC). Mask is provided as an output by the Convolutional Network, whereas, bounding box coordinates (Bounding Box) and class scores (Classes) are given by FC using the SoftMax activation function.
B: Modified backbone architecture of modified MRCNN, where (xn) shows the occurrence of a block.
C: Modified MRCNN architecture for dental image segmentation.

The MRCNN architecture is depicted in Figure 1A. Generic deep learning techniques for object identification yield several object boxes, which are then refined using classifiers and regressors.

The region based convolutional neural network (RCNN) is a typical method for detecting objects since it uses CNN (Convolutional Neural Network) to identify visual data along with a support vector machine (SVM) that categorises object boxes (10-14). Other approaches, a like the ROI pooling used by FRCNN (11), region proposal network (RPN) and FRCNN (12), and feature pyramid network (FPN) (13), have refined

C) Proposed CNN (Convolutional Neural Network)

In order to reduce the training parameters, a 62-layer residual network is proposed to replace the baselines Net101 in the original Mask RCNN model to reduce the network parameters and ensure the accuracy of dental image object detection. Additionally, the Adam optimiser has been used to replace SGD optimiser to further improve the performance of the model.
Using the aforementioned residual block as a starting point, we construct a 62-layer network consisting of 20 residual nodes. The over-fitting issue, brought on by the excessive number of network layers, may be greatly alleviated by the

usage of modular design to manage network layers. On top of that, our suggested network is divided into 6 stages (labelled S1, S2, S3, S4, S5, and S6), as opposed to the single stage utilised by ResNet101, and a 1 x 1 convolution is added to the shortcut link in the first residual block of residual blocks from the previous stage, and the feature map is 1/4 the size of the input picture. The next five stages are residual block stacking. In addition, the network culminates in a global average pooling. There is no need to tune any parameters in order to avoid over-fitting when

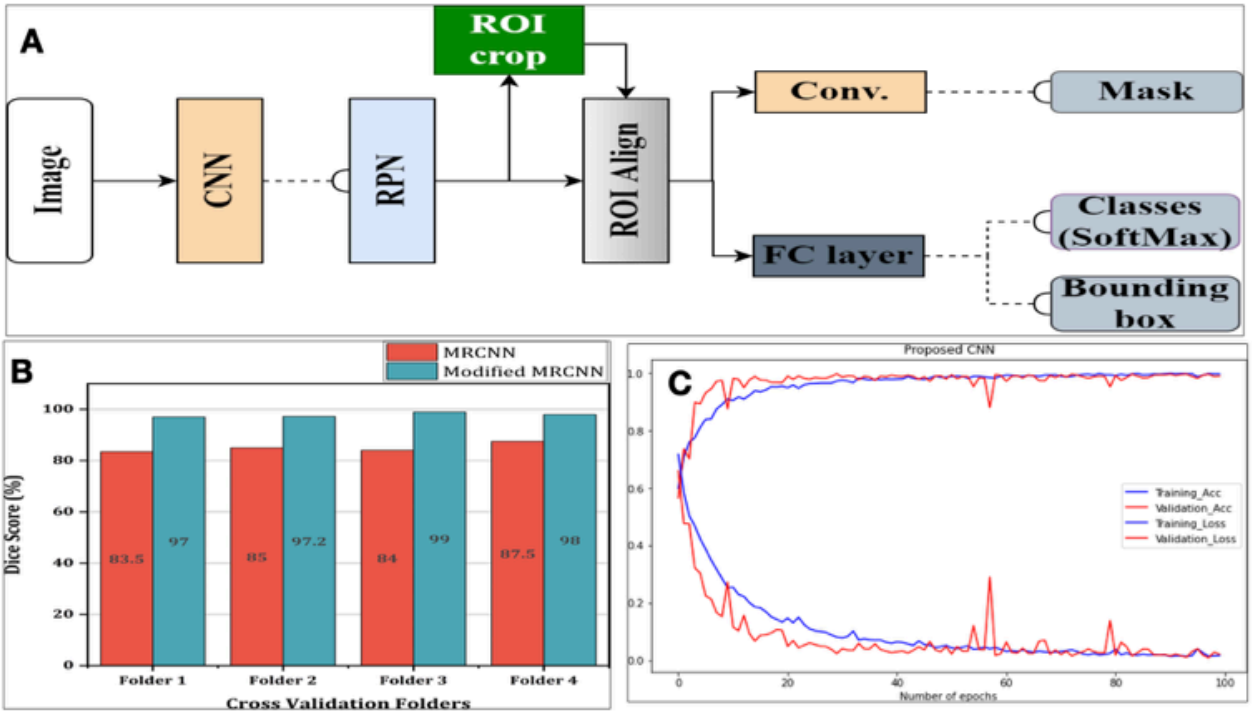

Figure 2
A: Pipeline representing the work done throughout this research including all the steps such as implementation, conversion, training, testing, and performance evaluation for both, MRCNN and Modified MRCNN, architectures, using Tensorflow-2 (TF2) and Python3.
B: Dice score values and comparison for both MRCNN and Modified MRCNN on 4-fold cross validation.
C: Live performance track of Modified-MRCNN training in terms of training- and validation- Accuracy (upper two curves) and Losses (lower two curves).

each stage in order to match dimensions. The schematic is shown in Figure 2. Each stage, from S2 to S6, consists three residual blocks amongst which the middle one uses a 3 x 3 convolution and 1 x 1 in the other two. Additionally, feature maps of each residual block have been tuned and decided based on the neighbour layer input and output. The complete architecture (Figure 2) can be visualised in two network-1 (from stage S1-S5), network-2 (last stage S6). Network-1 uses the conventional convolutions whereas the three-level structure of the last stage S6 (Network-2) uses dilated convolutions for all elements, the former uses ordinary convolutions throughout. We will take a closer look at how well these two networks recognise targets in next section.

This study uses Network-1 to illustrate the proposed network's individual row. As the network's input, the picture is randomly divided into squares measuring 512 by 512 pixels; in the first convolution stage (S1), the convolution kernel size is 7 by 7; and in the second convolution stage (S2), a 4 by 4 MaxPooling is used no stride. The current stage uses three using global average pooling, and the method also makes feature mapping and classifications more consistent and is more resistant to spatial variations in input and output. When it comes to categorisation, the complete connection layer is ultimately put to use. Surprisingly, foundational network of MRCNN, network-1, does not make use of the complete connection layer. The major responsibility of the feature extractor is to examine the input sonar picture for useful features and then pass those features along to the remainder of the MRCNN.

D) Proposed Modifications

The backbone network of modified M-RCNN architecture is proposed in Figure 1B & 1C. The model feeds an input imager-by-layer to the modified CNN backbone network to extract the feature maps through the FPN structure to fuse multi-scale information through the modified-convolution structure to obtain self-calibrated feature maps. According on the collected feature maps, the model's backend executes classifi-cation, bounding-box regression, and instance segmentation. Each boundary box in our model is subdivided into tooth and no tooth areas.

E) Loss function of modified MRCNN and Training

To optimise our model parameters, we utilise Dice loss as the criteria. The loss function utilised for this research is defined as: To test the utility of proposed strategy, we ran trials with two distinct model settings: a traditional MRCNN model, and the Modified MRCNN. We down scaled the images to 512x512 pixels while pre-processing our dataset. We utilised 64 feature maps at the highest level for both the MRCNN setups. With each down-sampling phase, the number of feature maps doubles until it reaches 512 in the fourth stage and the bottleneck. Furthermore, 0.0001 is being used as initial learning rate, the Adam optimiser is employed to train our model. With batch size of two, the number of training epochs was 100. If the validation performance did not improve, the learning rate was cut in half every 15 epochs. For training purposes we have used NVIDIA RTX 3060 as the Graphical Computing Unit with 12 Gigabytes of dedicated memory along with 32 Gigabytes of total Random Access Memory, which took average 85 seconds to run each epoch.

F) The model

Pipeline representing the work done throughout this research including all the steps such as implementation, conversion, training, testing, and performance evaluation for both, MRCNN and Modified MRCNN, architectures, using Tensorflow-2 (TF2) and Python3 (Figure 2A). This work has been carried out using the open-source library Tensorflow, created by Google, with supporting library as Keras in Python3. Also, to tune the original architecture of MRCNN, we implemented the whole MRCNN model that supports Tensorflow version 2.7.0 (20), which was the first MRCNN development supporting the latest Tensorflow version while this research was carried out. This development of original MRCNN was later on used for the modification and research purposes. Using the Modified MRCNN architecture, we have followed a two step process for tooth segmentation. In the first stage, we trained the modified MRCNN model to detect the teeth, on the annotated dataset consisting 1800 panoramic x-rays and 448 x-rays for the validation purpose. Then the trained model was tested on the testing dataset for performance evaluation in the second stage. Simultaneously, training of original MRCNN model was also executed with same procedure for comparison between its performance and the results achieved by modified architecture.

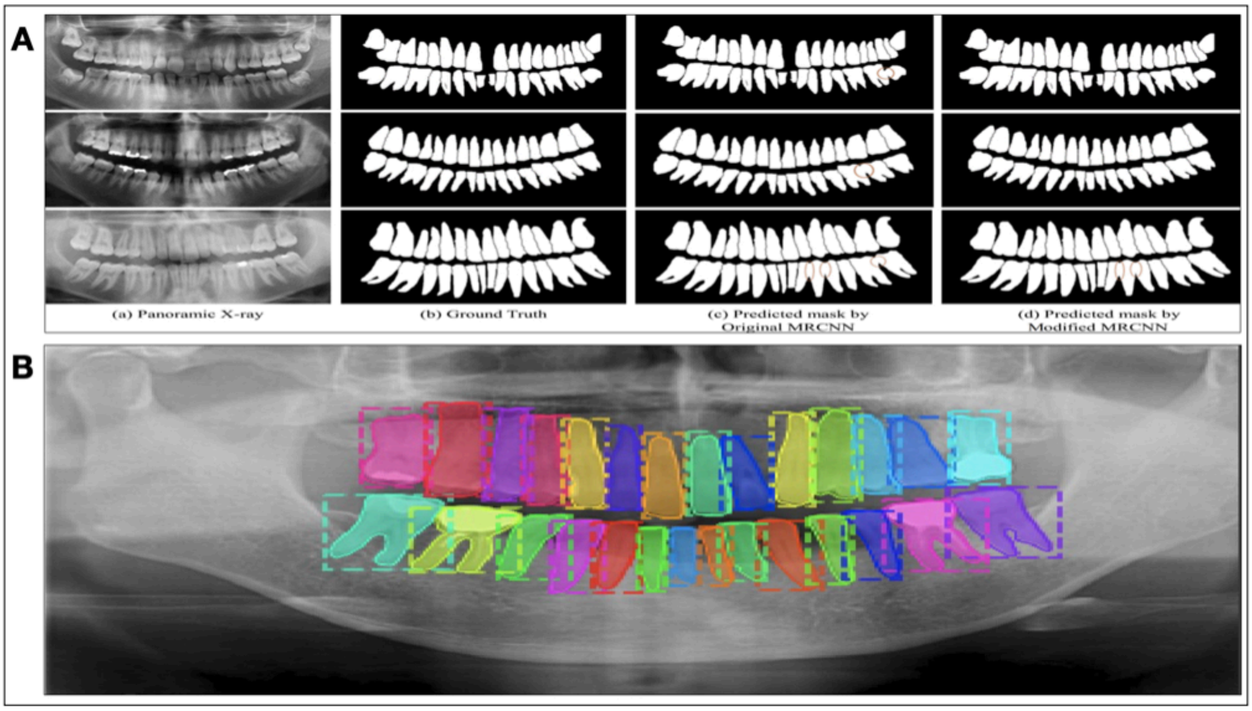

Figure 3
A: The following are some examples of the segmentation results obtained by using the original MRCNN (c), and Modified MRCNN (d) architectures
B: Detection results, using proposed model, shown as individual mask on each tooth with color differences generated by the system

G) Proposed method pipeline

Performance metrics:

For the assessment of the performance of the proposed method of medical image segmentation, several metrics are utilised which are as following:

- Dice Score: For evaluation of our proposed method we are using Dice Score (DS) as one of the performance metrics.

- Pixel Accuracy: Pixel Accuracy is defined as the proportion of correctly categorised pixels divided by the total number of pixels.

- Intersection over union: The projected segmentation map P's area of intersection with the ground truth map G's area of union, which is measured in units of 0 to 1, is defined as intersection over union (IoU). Whereas, average of IoUs over all the defined classes can be defined as mean-IoU.

H) Residual Block

In Visual Geometry Group (VGG) (7), there are 19 layers in the network, and in GoogleNet (8), there are 22 layers. The question that arises here is about the relation between accuracy and number of layers. When there are more network layers in deep learning, there are usually two problems:

1) the use of computer resources; and

2) gradient vanishing or gradient explosion.

A residual learning framework has been proposed by He et al. (9) to make network training easier. A number of residual blocks make up the residual network. By using shortcut and skip connection the residual block is created. This straightforward change can significantly speed up training and enhance the effectiveness of the model's training without adding any new parameters or computational burden to the network. This straightforward structure can also effectively address the deterioration issue. In this paper, residual blocks form the core of the network we propose.

Identity mapping makes it simple to implement shortcut connections. Let's assume desired output of network is F(x). We are getting R(x) = F(x) − x, where x is the input provided at the beginning of layers. To meet the requirements need to recast the original mapping R(x) + x. The residual block R(x) is flexible in nature. It can be used as a two or three layer network. A two-layer residual block is formed by two 3 x 3 convolution modules. Whereas the three-layer residual block is made up of 1 x 1, 3 x 3, and 1 x 1 convolutions blocks. In this case, the 1 x 1 layer's calculation is first reduced by the middle 3 x 3 convolution layer, which then restores under another 1 x 1 layer while maintaining accuracy and reducing calculation. We have adopted the 3 layers residual block for this work.

Results

Quantitative results

The total findings of the suggested technique are described in terms of its the average Dice score for each tooth location based on the cross validation stage (Figure 2B). The plots attached in shows the performance of modified-MRCNN while training and illustrate the final training and validation accuracies as 99.67% and 98.94%, respectively (Figure 2C). Additionally, Table 1 shows the comparison between various existing model performances on Tufts Dental Database (19), based on the performance metrics defined earlier. We note that the performance of the Modified MRCNN configuration consistently beats that of the original MRCNN and other models for tooth class segmentation. This demonstrates the significance of removing unnecessary information and feeding the target information only, which not only reduces the data size but also helps the model to learn fast and with better accuracy The significant discrepancy in DS between the original MRCNN and the Modified MRCNN seems to be because of the plethora of unimportant information provided while training. The incorporation of past information through ROI cropping on the skip connections improves the accuracy of the dice by over 12% on tooth segmentation in panoramic dental x-rays. This is due to the fact that the original MRCNN model has an incorrect classification problem with panoramic X-ray images including damaged teeth, noisy pixels, and missing teeth may also be the cause. The positioning information conveyed by the masks and bounding boxes ensures that each tooth is positioned close to its proper spot. The average Dice accuracy of the modified MRCNN and MRCNN designs is 97.8% and 85%, respectively. Additionally, using the test data set and the same technique, we calculated the Dice metric. The modified MRCNN performed best with dice score of 96%.
Following the validation scores, the unmodified MRCNN model has the lower dice score of 83%.

Qualitative Results (Figure 3A, 3B)

We display segmentation results generated by the MRCNN and our suggested Modified MRCNN. The Modified MRCNN provides better results segmenting any two teeth which are so close that the distance between them is negligible which may lead to count single object while covering two or more tooth in same prediction, typically, this happens when a tooth structures are not proper or grown adjacently. Though, this issue has been addressed with significantly high accuracy by the modifications suggested through this research, there are few instances like the one demonstrated in third row of inaccurate mask prediction (denoted within circles) on individual tooth that also occurs with the Modified MRCNN.

Discussion

The present study approaches the key issue of determining of training samples needed to obtain results in tooth identification and segmentation concentrate on periapical or intraoral radiography. As a direct modification of its standard backbone ResNet50, we have adopted a different backbone structure than the typical MRCNN network, with dilated layers replacing the few conventional layers of the network. Similar to the original MRCNN, our model is able to perform object identification and segmentation simultaneously. This model uses a modified CNN based on a modified backbone network to gather additional discriminating feature information and then adds some size of dilated convolutions to increase the impact of instance/semantic segmentation (25-27). Clinicians will benefit from AI's role in managing oral cancer by using automatic segmentation to help them manage these medical conditions.

Most of studies on tooth identification and segmentation concentrate on periapical or intraoral radiography and try improve the performance of their methods by proposing large changes in the existing methodologies. It is challenging to create an image processing algorithm that can correctly recognise and locate each of the 32 teeth on panoramic X-rays, as such radiographs frequently exhibit significant variability due to high amplitude noise or even low contrast.

Furthermore, non-experts may have difficulty distinguishing between the teeth within each group (incisors, canines, premolars, and molars). Their morphologies are very similar, as are their grey level amplitudes, and the position information may be misleading because a missing tooth can induce its neighbour to migrate and occupy its place. Aside from the tooth arrangement, which varies widely from person to person, two consecutive teeth may share some restorations and seem close together, if not overlapping, on an X-ray image.

Such issues raise doubts about CNN's capacity to discriminate between two distinct teeth that have the same form and pixel values and require an efficient method for segmenting each tooth and localising it separately for further processing. In order to automatically segment, identify, and identify the teeth, Jader et al. (6) introduced instance segmentation techniques that used four distinct architectures: PANet, HTC, ResNet, and UNet. The algorithms produced remarkably good results.Our suggested strategy achieved substantially higher accuracy and precision when compared to existing methods. When prior data is included via ROI cropping on skip connections, the accuracy of dicing for tooth segmentation in panoramic dental x-rays is improved by around 12%.

Panoramic X-rays are also rarely the subject of publications because they are frequently of poor quality (excessive noise, uneven contrasts, etc.). On the other hand, through this research, we are putting forth a Modified MRCNN model-based automatic segmentation technique that links regions of interest to skip connections and incor-

Model	IoU (%)	P_A (%)	DS (%)
FPN (13)	86.37	95.17	92.24
UNet (9)	86.67	95.22	92.46
UNet++ (21)	86.62	95.19	92.47
PSPNet (22)	86.08	95.00	91.77
DeepLabV3 (23)	86.21	95.00	92.00
DeepLabV3+ (24)	86.41	95.13	91.80
Modified MRCNN	**98.67**	**96.53**	**97.8**

Table 1
Comparison between performance of various model with proposed modified MRCNN model, in terms of Pixel Accuracy, Intersection over union, and Dice Score.

porates location prior. effectively illustrating the need for the MRCNN to include some geo-metric localization information in order to address this issue, improve performance, and avoid incorrectly segmenting these X-rays with noisy surrounding pixels. A modified MRCNN model's average gain in the Dice score value per tooth class lends credence to the idea that even small adjustments to current procedures can result in a noticeable improvement in efficiency. In terms of the Dice score, the suggested segmentation strategy shows better results than the original MRCNN model.

Alteration in the shape, texture, and overall appearance of teeth as well as adjacent structures can frequently be among the initial signs of oral cancer.

The following methods can help in the early detection of these changes:

Automated tooth segmentation utilising Mask RCNNs.

Improving Image Analysis: The model can assist in finding abnormal growths or lesions that may signal oral cancer by precisely segmenting teeth and surrounding structures.

Efficiency and Consistency: Automated systems can analyse images efficiently and consistently, eliminating the need for human interpretation, which can be laborious and prone to error.

Tracking the Course of the Disease: Segmented pictures can be utilised to track changes over time, giving important information for evaluating the course of oral cancer and the efficacy of treatment.

Clinicians will benefit from AI's role in managing oral cancer by using automatic segmentation to help them manage this difficult condition. Apart from diagnosis, the comprehensive analysis offered by Mask RCNNs and automatic tooth segmentation can aid in the treatment of oral cancer by:

- *Treatment Planning:* Precise segmentation aids in the planning of surgical procedures, guaranteeing the preservation of good structures and the removal of diseased tissues.

- *Prosthetic Design:* Using dental prosthesis is a common part of post-surgical rehabilitation. Creating prosthetics that fit and operate better can be aided by precise segmentation.

- *Follow-up Care:* Automated analysis methods can be utilised during routine follow-up visits to identify any cancer recurrence early on and enable timely action.

As the Modified MRCNN relies on mask created, either manually or by any other annotation method, further stages include using or inventing an efficient object detection technique that reduces the difference between detected and ground-truth masks. Also the future researches may focus on further analysis of dental x-rays while detecting existing disease or previously done treatment in each tooth along with numbering and looking for missing teeth as well.

Conclusion

We present a novel deep learning approach to medical image analysis that addresses the challenge of segmenting teeth and evaluates the ability to segment teeth in panoramic images. This will be a useful tool for both a novel segmentation methodology for medical pictures and automatic segmentation of individual teeth for medical diagnosis in dental treatment. Integrating Mask RCNNs with residual feature extraction for automatic joint teeth segmentation in panoramic dental pictures constitutes a major advance in dental imaging technology. Its application is essential for the early identification, diagnosis, and management of oral cancer in along with enhancing both the precision and efficacy of dental diagnostics. This kind of innovation has an opportunity to boost patient outcomes by providing more accurate and promptly delivered health services.

Our proposed method will serve as a valuable tool for automatic segmentation of individual teeth for medical management. This current method leads to higher accuracy and precision. Segmented images can be used to evaluate periodic changes, providing valuable data for assessing the progression of oral cancer and the efficacy of management. Future research should focus on developing less complex, lightweight, and faster vision models while maintaining high accuracy.

Correspondence to:
Dr. Abhijeet Ashok Salunke
https://orcid.org/0000-0003-0103-8599
The Gujarat Cancer & Research Institute
Civil Hospital Campus, Asarwa,
Ahmedabad-380 016.
Gujarat, INDIA
Phone :+91-79- 2268 8000
Fax : +91-79-2268 5490

Conflict of interest:

The authors declare that there were no conflicts of interest within the meaning of the recommendations of the International Committee of Medical Journal Editors when the article was written.

Disclaimer/Publisher's Note:

The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Swiss J. Radiol. Nucl. Med. and/or the editor(s). Swiss J. Radiol. Nucl. Med. and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

License Policy:

References

1. 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. doi: https://doi.org/10.1016/j.eswa.2018.04.001

2. A. Lurie, G. M. Tosoni, J. Tsimikas, and W. Fitz, "Recursive hierarchic segmentation analysis of bone mineral density changes on digital panoramic images," Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, vol. 113(4), pp. 549–558, 2012. https://doi.org/10.1016/j.oooo.2011.10.002

3. Y.Y. Amer and M. J. Aqel, "An efficient segmentation algorithm for panoramic dental images," Procedia Computer Science, vol. 65, pp. 718–725, 2015, international Conference on Communications, management, and Information technology (ICCMIT'2015). https://doi.org/10.1016/j.procs.2015.09.016

4. M. K. Alsmadi, "A hybrid fuzzy c-means and neutrosophic for jaw lesions segmentation," Ain Shams Engineering Journal, vol. 9, no. 4, pp. 697–706, 2018. https://doi.org/10.1016/j.asej.2016.03.016

5. M. R. M. Razali, N. S. Ahmad, R. Hassan, Z. M. Zaki, and W. Ismail, "Sobel and canny edges segmentations for the dental age assessment," in 2014 International Conference on Computer Assisted System in Health, 2014, pp. 62–66. https://doi.org/10.1109/CASH.2014.10

6. G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic x-ray images," in 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2018, pp. 400–407. https://doi.org/10.1109/SIBGRAPI.2018.00058

7. K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in ´ 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp.2980–2988.

8. T. L. Koch, M. Perslev, C. Igel, and S. S. Brandt, "Accurate segmentation of dental panoramic radiographs with U-Nets," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 15–19. https://doi.org/10.1109/ISBI.2019.8759563

9. O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

10. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142–158, Jan. 2016, https://doi.org/10.1109/TPAMI.2015.2437384

11. R. Girshick, “Fast R-CNN,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169

12. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. https://doi.org/10.1109/TPAMI.2016.2577031

13. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2117–2125. https://doi.org/10.1109/CVPR.2017.106

14. F. Deng, H. Hu, S. Chen, Q. Guan, and Y. Zou, “Rich feature hierarchies for cell detecting under phase contrast microscopy images”, in 2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP), Nov. 2015, pp. 348–353. https://doi.org/10.1109/ICICIP.2015.7388195

15. N. Atif, M. Bhuyan, and S. Ahamed, “A Review on Semantic Segmentation from a Modern Perspective,” in 2019 International Conference on Electrical, Electronics and Computer Engineering (UPCON), Nov. 2019, pp. 1–6. https://doi.org/10.1109/UPCON47278.2019.8980189

16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)https://doi.org/10.48550/arXiv.1409.1556

17. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015) https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html

18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016) https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

19. K. Panetta, R. Rajendran, A. Ramesh, S. P. Rao and S. Agaian, "Tufts Dental Database: A Multimodal Panoramic X-Ray Dataset for Benchmarking Diagnostic Systems", in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 4, pp. 1650-1659, April 2022, https://doi.org/10.1109/JBHI.2021.3117575

20. Mask-RCNN implementation for Tensorflow 2.7.0 and Keras 2.7.0. https://github.com/Kamlesh364/Mask-RCNN-TF2.7.0-keras2.7.0

21. Z. Zhou, M. M. R. 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, Berlin, Germany: Springer, 2018, pp. 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

22. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2881–2890. https://openaccess.thecvf.com/content_cvpr_2017/html/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.html

23. L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” 2017, arXiv:1706.05587.

24. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoderdecoder with atrous separable convolution for semantic image segmentation,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 801–818. https://openaccess.thecvf.com/content_ECCV_2018/html/Liang-Chieh_Chen_Encoder-Decoder_with_Atrous_ECCV_2018_paper.html

25. Su, B., Zhang, Q., Gong, Y. et al. Deep learning-based classification and segmentation for scalpels. Int J CARS (2023). https://doi.org/10.1007/s11548-022-02825-7

26. Wang, H., Xiao, N., Luo, S. et al. Multi-scale dense selective network based on border modeling for lung nodule segmentation. Int J CARS (2023). https://doi.org/10.1007/s11548-022-02817-7

27. Manjunatha, Y., Sharma, V., Iwahori, Y. et al. Lymph node detection in CT scans using modified U-Net with residual learning and 3D deep network. Int J CARS (2023). https://doi.org/10.1007/s11548-022-02822-w