Advancements in Dental Filling Detection Technologies and Strategies for Comprehensive Oral Health Care

Shivampeta Aparna
Gitam University, Hyderabad Campus, India
ashivamp@gitam.edu

Himabindu Gottumukkala
GITAM School of Technology, India
hgottumu@gitam.edu

Nitya Shivampet
Oak Park High School, Thousand Oaks, CA, USA
nitya.shivampet@gmail.com

Kireet Muppavaram
GITAM School of Technology, India
kireet04@gmail.com

Chaitanya C. V. Ramayanam
University of Southern California, Los Angeles, CA, USA
ramayana@usc.edu (corresponding author)

Received: 1 April 2024 | Revised: 15 April 2024 | Accepted: 16 April 2024

 DOI: https://doi.org/10.48084/etasr.7400

ABSTRACT

Many individuals face issues with their teeth, requiring the expertise of dentists to provide necessary care. Despite the advancements in dental techniques, there is a persistent shortage of dentists, prompting the development of tools to help the latter efficiently perform patient treatment. The current research focuses on refining the precision of the vital dental treatment known as dental fillings. The approach involves utilizing the Mask Region-based Convolutional Neural Network (MaskRCNN) with different variants of ResNET, such as ResNET50, ResNET101 C4, Dilated C5, and Feature Pyramid Network (FPN), to analyze diverse dental radiographs. By training on a broad range of tooth images, this methodology creates a pixel-based masking system, improving dentists' ability to precisely identify filling levels. Consequently, this innovation contributes significantly to expediting and refining the accuracy of dental treatments, ultimately benefiting individuals with tooth problems. Additionally, as a future prospect, this model can enable robots to perform dental operations as it provides pixel-level information necessary for the treatment.

Keywords-dental x-rays; RESNET; MaskRCNN; annotations; dental fillings; FPN; ROI pooling

I. INTRODUCTION

Dental healthcare relies significantly on precise diagnosis, treatment planning, and monitoring to ensure the oral health and well-being of patients. A crucial aspect of this process is the interpretation of dental radiographs, which provide vital insights into various dental conditions. However, accurately segmenting dental objects within radiographic images remains a complex challenge. Traditional segmentation methods often struggle to achieve the required pixel-level [1], leading to potential misdiagnoses, inadequate treatment planning, and suboptimal patient care. The central focus of this research is to advance dental radiograph analysis by developing an instance segmentation model [2, 3]. This model aims to meticulously and consistently delineate various dental objects. The ultimate goal is to significantly improve the accuracy, granularity, and clinical relevance of dental object segmentation [4]. By doing so, this study aims to empower dental practitioners with more robust tools for precise diagnosis, treatment planning, and patient management. Image segmentation of dental works is essential for proper patient treatment since there are various objects a medical practitioner would observe through this process, which would facilitate the former to successfully deal with the patient's issue. This study deploys the periapical view of the tooth from the detailed X-rays [5, 6]. The aforementioned objects have been generalized into three main classes, namely Endodontic, Restoration, and Implant. These three categories are selected among over 12 specific objects prevailing in dental fillings and implants. These categories have been generalized based on the opinion of an experienced dentist and by referring to the research done in [7]. Besides dentists, this research work may assist common people, who are not familiar with or do not have the expertise to understand a dental X-ray, to comprehend the treatment done, the work precision, and several other details that can be derived from dental radiographs. Also, as robotics have been introduced to the medical field [8], this research can provide a machine with the necessary pixel level information to perform a dental surgery.

The dataset used for training the neural network plays a critical part in the model effectiveness. A diverse dataset offers ample information for the neural network to learn from, leading to consistent performance even on real-world data. In this research, the dataset provided by [7], which comprises images collected from various dental clinics in Hyderabad, Telangana, India, was utilized. These images, being originally in jpg format with four channels, were converted to three-channel images to reduce computational complexity. The discussed images were resized to 416×416×3. Additionally, the annotations required for training the model were obtained from the same source, ensuring consistency with the dataset and the methodology followed in this study.

II. LITERATURE SURVEY

Authors in [7] utilized a dataset of several dental radiographs and employed the Mask RCNN model with the ResNET 50 architecture to generate masks on new dental radiographs. They proved that this model is beneficial in helping a dentist see the exact level of filling that has been performed. Also, during a medical operation conducted by a machine, the machine gets knowledge on the extent of the filling that has to be done. The concept of utilizing an instance segmentation model in the medical x-rays is fascinating, and in this work the impact of layers is determined. The current work strives to build on the model mentioned above, improve the results even by the slightest margin, and experiment with the addition of layers to the architecture. By doing so, the current model can be ameliorated and become convenient for practical and real-time purposes. Authors in [9] implemented a faster RCNN, which is a model used for object detection and classification to train on dental X-ray films of periapical view. The size of the dataset used is 1250. Post-processing of data helps in the faster training of the model while an edge has been provided by adding a filter which eliminates the overlapping bounding boxes predicted by the model. The filtering algorithm and the post-processing steps make this work unique. However, this work is limited to predicting bounding boxes only and does not extend to instance segmentation. The measure for average recall is still found through the Intersection over Union (IoU) score. In addition to the model, the evaluation metrics acquired from this work are of great interest. Authors in [8] demonstrated that dental disease detection using machine learning techniques is a possibility. They took several radiographs of patients and claimed that disease detection in teeth can be identified through these X-rays. They then utilized the classification and segmentation techniques to identify and process the disease. By utilizing instance segmentation, this work aims to identify the disease. In the specific work however, the extent of filling is being verified to assure proper treatment and automation to aid the dentists. It was shown that an automated system can be devised to identify dental diseases, which aids both people by providing faster and precise care and dentists by assisting them in swiftly dealing with patients.

III. DATASET

The dataset plays the most crucial role in the training of a neural network. It needs a uniform image set and the corresponding annotation file with the coordinates of the mask of the objects. Some processing of data to make it training-ready is necessary.

A. Description

For this research study, since several models are going to be compared for improved results, the dataset of [7] was utilized. The dataset contains 1755 images of 416×416×3 pixels in jpg format. As per this study's generalization the same three classes Endodontic, Restoration, and Implant were used. The 416 in the dimensions represent the height and width of the image and the number 3 represents the RGB channels. This dataset was collected from different dental clinics located in Hyderabad, Telangana, India.

B. Annotation

As the recommended model is required to find the pixel-level mask of the object, the neural network deployed in this study needs to be initially informed of how a mask is present within the radiograph. For this to be accomplished, there are several tools to manually create a mask of the desired objects on the images and export it as a json file. This file contains the coordinates of all points present within the mask along with the label of the mask. There is one json file per image, which contains multiple masks of the same and different categories. The tool Labelme was employed for annotating the images.

IV. MODELS

Neural networks are the arrangement of logic units (neurons) in a layered structure to mimic the learning procefure of an organism. For this research, the predefined ResNET model is utilized with variations to test how the learning changes with respect to the number and arrangement of layers. For the pixel-based masking, Mask RCNN is applied and the combination of Mask RCNN and ResNET will be the model to generate the masks in dental radiographs.

A. Mask RCNN

Mask RCNN is a deep learning technique designed for instance segmentation on object detection. Real-world problems have more objects and more complex forms of images for CNN to solve [10]. In order to tackle this problem RCNN (Region-based CNN) is put into service. In this research, several regional objects present in the images are exploited and then CNN is applied to every proposed regional bounding box. This makes the model capable of predicting multiple objects of numerous classes within the same image. Mask RCNN is an instance segmentation model that is capable of classifying every pixel into a category and this classification results in the mask, which is the accurate outline of the object. Firstly, Mask RCNN uses ROI align to create proposals for regional bounding boxes on the images. Then, image segmentation takes place to separate objects from the images. Finally, instance segmentation, which classifies each pixel to the right object creating a mask, occurs. Three outputs, the object category, a bounding box, and the mask, are acquired. The mask, which can provide accurate information regarding the treatment being done on the dental fillings, is the goal of this project. Authors in [9] utilized faster RCNN, whereas authors in [7] used Mask RCNN. In this work, the latter is deployed, but the architecture behind it is modified. This change in model architecture could lead to better training, leading to improved results.

B. Residual Networks

ResNET is a deep learning neural network with a skip connection feature that enables faster and better learning of the model. For the introduced model, several different forms of the ResNET are to be employed to provide a different architecture compared to previous works as follows:

- ResNET 50
- ResNET 101 Dilated C5
- ResNET 101 C4
- ResNET 101 FPN

Models can be trained in different ways. One way is to apply transfer learning where the weights that have been saved after training the model on huge datasets, such as the COCO dataset, are used, which helps in faster and generalized training of the model. Another way is to train the model from scratch without transferring any weights. Even though this technique makes the model-specific to the data, it usually extracts better results. When dealing with medical data, the model is not expected to perform or detect other category objects and hence it can be trained as data-specific. From [7], it has been noticed that the absence of transfer learning produces better results and thus the same approach is adopted in this work. Since there are only three categories targeted by the model and they are medical objects, which are not found in the generic ImageNet training data, the training without transfer learning generates better results.

C. ResNET 50

ResNET 50 is a 50-layered neural network that consists of 48 convolutional layers and 2 pooling layers and which aims to eliminate the vanishing gradient problem [11, 12]. When the loss is calculated in the backpropagation of values, the number estimated in the differential function becomes so low that it becomes insignificant. Due to this low value, the weights of the neurons fail to update and the loss of the model does not change. This means that the model is unable to learn from the data anymore. This issue is known as the vanishing gradient problem. To tackle this challenge, ResNET 50 has a "skip connection" after every two convolutional layers, which sends the input to the layers ahead [13]. With the processed input and the raw input the layers ahead receive more information to train upon. ResNET 50 has a max pool layers right after the first block of convolutional layers to remove the noise and unwanted features from the data as the layers from the point get both raw and processed data. The architecture to train the Mask RCNN model was implemented in [7].

D. ResNET 101 Dilated C5

ResNET 101 Dilated C5 is a combination of the ResNET 101 architecture, which consists of 101 layers and dilated convolutional layers in the fifth stage of ResNET architecture [14, 15]. Dilated convolutional layers, also known as atrous convolutions, are a modified version of the convolutional layers that inserts gaps in the kernels which perform on the images. Each gap provides a wider reception to the kernel enabling it to extract more features without the increase in parameters. This way the model is able to obtain a wider area of data from images. C5 denotes the last layers of the ResNET 101 architecture. Dilated C5 means that the dilated convolutional layers are being applied on the last block of the ResNET 101 architecture while the remaining blocks stay unchanged. With this architectural change, the suggested model is anticipated to have a wider view for better training.

E. ResNET 101 C4

ResNET 101 C4 is the deep neural network and a variant of ResNET with a total of 101 layers. This ResNET variant also has a skip connection to help eliminate the vanishing gradient problem. The additional layers assist in dealing with the increasing size of the datasets. The results will disclose the benefit of an increase in the layers while the size of the dataset remains the same. The C4 represents a block of layers present within the ResNET 101 architecture. This means that some changes have been made, specifically in the C4 block of the architecture, to upgrade model performance. In the proposed model, the C4 architecture provides additional layers compared to previous works. This set of additional layers can promote the better learning of the model.

F. ResNET 101 FPN

ResNET 101 FPN is the combination of ResNET 101 and Feature Pyramid Network [16]. ResNET 101, has 101 layers consisting of several blocks and skip connections, helping tackle the vanishing gradient problem. FPN adopts the concept of deploying feature maps at multiple scales. FPN extracts the feature maps at different scales and arranges them in a pyramid structure. This entails low resolution, yet high semantic images and high resolution with low semantic images. After the pyramid structure arrangement, FPN utilizes a top-down pathway and lateral connections that bring together features at different scales and create a multi-scale representation of feature maps. This has helped improve ResNET 101 as the main challenge to tackle in the processing of data (images) is size. Due to the robust features of the FPN, the ResNET architecture can tackle the change in sizes and perform better.

V. RESULTS

To compare the results, a metric called Average Precision is deployed. To calculate this, both the predicted mask coordinates and the true mask coordinates are demanded. From the annotations this study already has the true co-ordinates of the mask. By using the proposed model, the predicted mask coordinates are acquired. Next, the Intersection over Union (IoU) score for the masks is calculated. Employing the two mask shapes obtained, the total area in union of the two masks, divided by the area of intersection, is estimated. The IoU score is then compared with a threshold, in this case 50%. If the IoU is greater than the threshold then it is considered as a "true" value. Assume there is an object and the suggested model predicts "true", then it is a "True Positive". If there is an object but the introduced model fails to even detect it or misclassifies it, then such a case is "False Positive". If there is no object but the model predicts an object, then it is "False Negative". And when there is no object and the model recommended does not show any object, it is "True Negative". Precision and recall values are computed utilizing the count of these four values. Now, for each class, a precision-recall curve is made, and the area interpolated by the area under this curve is known as the Average Precision (AP):

$$AP = \sum_{k=1}^{n} (P(k) \ * \triangle \ recall_k) \quad (1)$$

where P(k) is the precision at point k, delta recall at k represents the change of recall from k-1 to k and n is the total number of points. The mean of the APs is known as the Mean Average Precision (MAP) score, which is a standard metric used in comparing instance segmentation models because it effectively captures both the localization accuracy and the ability to distinguish between different object instances.

$$MAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \quad (2)$$

where N is the total number of classes and AP_i is the AP score of class i. For bounding boxes, the results with the Mask RCNN algorithm with different sets of the ResNet architecture are shown in Table I.

There is a need to check the segmentation scores to see their difference and how the model gets affected by different architectures. Clearly, when compared to the previous models, more layers have been enabled within the architecture with additional operations, such as dilation, to improve the learning of the proposed model.

From the results obtained, it is obvious that the outcomes acquired by the Mask RCNN with ResNET 101 FPN are the best, since bounding boxes have achieved a total of 88.481 AP50 score and Segmentation has achieved 76.423 AP50 score. It is closely followed by the ResNET50 architecture [7]. In the medical field a minor difference in metrics could have a great difference in treatment. Such an improvement shows that with an increase of the dataset and layers in the architecture, much better results can be obtained.

TABLE I. COMPARISON OF MASK RCNN MODELS WITH DIFFERENT RESNET BACKBONES FOR BOUNDING BOXES

Results (Mask RCNN)	AP	AP50	AP75	APs	APm	APl
Resnet 50	53.994	88.337	56.108	22.135	50.195	64.57
Resnet 101 Dilated C5	55.072	85.38	68.619	24.436	56.651	61.044
Resnet 101 C4	55.567	84.705	65.453	22.228	60.17	67.01
Resnet 101 FPN	50.96	88.481	55.434	41.46	54.222	57.297

TABLE II. COMPARISON OF MASK RCNN MODELS WITH DIFFERENT RESNET BACKBONES FOR SEGMENTATION

Results (Mask RCNN)	AP	AP50	AP75	APs	APm	APl
Resnet 50	46.46	74.279	53.791	1.894	38.568	71.376
Resnet 101 Dilated C5	49.855	73.314	58.359	2.052	39.921	74.161
Resnet 101 C4	49.16	73.94	58.336	1.456	48.836	73.479
Resnet 101 FPN	50.412	76.423	60.462	20.787	49.115	72.677

VI. CONCLUSION

This work is mainly focused on testing and enhancing the dental filling detection model. The need for creating a system in the dental field, where a machine can understand the level of the filling in treatment is essential. By developing a system, the time devoted by a medical practitioner to read a dental radiograph is eliminated. Additionally, people with no previous medical knowledge may better understand what is being done in the dental filling treatment. This study may, also, enhance a system, which can aid a robot perform a surgery on a patient using image segmentation, which provides information at pixel level.

From the results obtained, it is thought that Mask RCNN with ResNET 101 with FPN has the best results, obtaining an AP50 score of 88.481 on bounding box with object classification and AP50 score of 76.423 on image segmentation with pixel level precision. Although this result is the best acquired one, it has only a minor difference with previous models. That said, a minor improvement in the medical field can have a drastic change in real time results.

The main observation made in this study is that with the same size of dataset, the results with different layers did not show a significant difference. This means that even though ResNET tackles the vanishing gradient problem, the size of the dataset still has a huge impact. As the number of epochs starts to increase, the threshold on learning decreases if the dataset available has fewer samples. This means that the model presented in this work could perform much better with a larger dataset. Also, as a future scope, different models available can be used on the same dataset and compared. In addition to that, a larger dataset can be applied to compare the results with respect to the model architecture and the dataset size.

REFERENCES

[1] V. Sa-ing, K. Wangkaoom, and S. S. Thongvigitmanee, "Automatic dental arch detection and panoramic image synthesis from CT images," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, Jul. 2013, pp. 6099–6102, https://doi.org/10.1109/EMBC.2013.6610944.

[2] F. Shamsafar, "A new feature extraction method from dental X-ray images for human identification," in *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, Zanjan, Iran, Sep. 2013, pp. 397–402, https://doi.org/10.1109/IranianMVIP.2013.6780018.

[3] A. Ajaz and D. Kathirvelu, "Dental biometrics: Computer aided human identification system using the dental panoramic radiographs," in *2013 International Conference on Communication and Signal Processing*, Melmaruvathur, India, Apr. 2013, pp. 717–721, https://doi.org/10.1109/iccsp.2013.6577149.

[4] L. Čular, M. Tomaić, M. Subašić, T. Šarić, V. Sajković, and M. Vodanović, "Dental age estimation from panoramic X-ray images using statistical models," in *Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis*, Ljubljana, Slovenia, Sep. 2017, pp. 25–30, https://doi.org/10.1109/ISPA.2017.8073563.

[5] N. C. Kundur, B. C. Anil, P. M. Dhulavvagol, R. Ganiger, and B. Ramadoss, "Pneumonia Detection in Chest X-Rays using Transfer Learning and TPUs," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11878–11883, Oct. 2023, https://doi.org/10.48084/etasr.6335.

[6] N. Kumar, A. Hashmi, M. Gupta, and A. Kundu, "Automatic Diagnosis of Covid-19 Related Pneumonia from CXR and CT-Scan Images," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7993–7997, Feb. 2022, https://doi.org/10.48084/etasr.4613.

[7] S. Aparna, K. Muppavaram, C. C. V. Ramayanam, and K. S. S. Ramani, "Mask RCNN with RESNET50 for Dental Filling Detection," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 10, 54/31 2021, https://doi.org/10.14569/IJACSA.2021.0121079.

[8] S. Divakaran, K. Vasanth, S. D, and S. V, "Classification of Digital Dental X-ray Images Using Machine Learning," in *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, Chennai, India, Mar. 2021, pp. 1–3, https://doi.org/10.1109/ICBSII51839.2021.9445171.

[9] H. Chen *et al.*, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Scientific Reports*, vol. 9, no. 1, Mar. 2019, Art. no. 3840, https://doi.org/10.1038/s41598-019-40414-y.

[10] M. Salemdeeb and S. Erturk, "Multi-national and Multi-language License Plate Detection using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 5979–5985, Aug. 2020, https://doi.org/10.48084/etasr.3573.

[11] S. Mukherjee, "The Annotated ResNet-50," *Towards Data Science*, Aug. 18, 2022. https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758.

[12] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv, Apr. 10, 2015, https://doi.org/10.48550/arXiv.1409.1556.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 580–587, https://doi.org/10.1109/CVPR.2014.81.

[14] F. Yu, V. Koltun, and T. Funkhouser, "Dilated Residual Networks." arXiv, May 28, 2017, https://doi.org/10.48550/arXiv.1705.09914.

[15] S.-H. Tsang, "Review: DRN — Dilated Residual Networks (Image Classification & Semantic Segmentation)," *Towards Data Science*, Mar. 20, 2019. https://towardsdatascience.com/review-drn-dilated-residual-networks-image-classification-semantic-segmentation-d527e1a8fb5.

[16] S.-H. Tsang, "Review: FPN — Feature Pyramid Network (Object Detection)," *Towards Data Science*, Mar. 20, 2019. https://towardsdatascience.com/review-fpn-feature-pyramid-network-object-detection-262fc7482610.