Mask RCNN with RESNET50 for Dental Filling Detection

S Aparna[1], Kireet Muppavaram[2], Chaitanya C V Ramayanam[3], K Satya Sai Ramani[4]

Department of Computer Science Engineering, GITAM School of Technology, Hyderabad, Telangana, India[1]
Department of Computer Science Engineering, JNTUH College of Engineering, Hyderabad, Telangana, India[2]
Student, Department of Computer Science Engineering, GITAM School of Technology, Hyderabad, Telangana, India[3, 4]

Abstract—**Teeth are very important for humans to eat food. However, teeth do get damaged for several reasons, like poor maintenance. Damaged teeth can cause severe pain and make it difficult to eat food. To safeguard the tooth from minor damages, an inert material is used to close the gap between the live part of the teeth or sometimes even the nerve and enamel. Although, long-time ignorance can increase the damage and inevitably result in root canal or tooth replacement. In the case of root canal, the gap between nerve and enamel is filled with an inert material. To check if the filling has been done properly, an X-ray is taken and verify. As technology is developing, robots are being introduced into many fields. In the medical field, there are instances where robots do surgery. For dental treatment, as an X-ray is taken to determine the filing, this work introduces a model to analyze the X-ray taken and estimate the level of filing done. The model is constructed using Mask RCNN with ResNet50 architecture. A dataset of different kinds of filings is taken and trained using the model. This model can be used to enable machines to perform dental operations as it works on pixel-level classification.**

Keywords—Dental x-rays; deep learning; mask RCNN; RESNET50

I. INTRODUCTION

Humans need food to survive. For eating food, teeth are essential and healthy teeth indicate healthy life. As medical sciences have advanced, dental research and learning also have advanced. People of the present generation have the most common issues related to teeth. As teeth are one of the most important parts of the body, which does most of the work helping in grinding the food consumed and supporting the digestive system, they need to be safeguarded from damage [28]. Not taking proper care of teeth may lead to some issues like cavities, swelling of gums, broken teeth, etc. When an issue arises, there are dentists who can attend the issue, and different problems need a different kind of treatment. Many issues may arise in teeth, and multiple solutions might be there to cure them. This project focuses on the external material that a dentist uses to fill, do the treatment, or close the gaps in teeth. This work mainly focuses on three things, namely "Endodontic", "Restoration" and "Implant" [9], [13]. "Endodontic" refers to the stick-like tools that a dentist generally uses to clean the inner part of teeth. "Restoration" refers to the filing that a dentist does on some break, some cavities, and then caps that are used after a root canal. "Implant" is a screw-like structure used to provide rigidity and structure to the teeth. Many studies were also conducted

regarding identification of human using dental X-rays [7], [14], [16]. The aim of their work was to identify teeth as a unique bio metric system [21], [22], [23]. Their studies help this work by providing ways in which the uniqueness of teeth was identified by using computer vision.

This work aims at helping dentists identify the objects in a dental radiograph. This work benefits dentists, but it also helps common people read their own dental radiograph and understand the treatment being done [1], [2], [3], [4], [5]. The model focuses on detecting the filings in dental x-rays, which have been generalized to three classes. Assuming a case of many cases arising, there will be many x-rays taken to identify the teeth damage. Automating the detection helps reduce time and improve efficiency [6],[8].

Also, as robotics is being advanced and are brought into different fields, the medical field has also conducted experiments. There are works demonstrating robots doing and assisting in surgeries. In the future dental operations can also be automated, and robots can perform minor operations. By identifying objects at pixel level, this work can help robots in real-time identification of objects in radiographs. The models trained in this work use the concept of transfer learning where "COCO" (Common Objects in Context) dataset trained weights are imported, and the new dataset updates those weights for results [9]. A Mask RCNN (regional convolutional neural network) model is used with RESNET50 (residual neural network) as its backbone architecture.

The dataset for this project has been collected from different dental clinics in Hyderabad, Telangana, India. The dataset consists of several dental radiographs in "JPG" format containing RGBA (red, green, blue, alpha) channels. The images were processed and converted to RGB (red, green, blue) format with fixed dimensions of 416*416. Then the images were annotated using the "Labelme" tool with three classes. After annotations, the model was trained with 100 images, 250 images, 500 images, 750 images, 1000 images, and 1755 images to see the improvement in results as the dataset size increases [27].

In this work, the medical radiographs are collected, which is the dataset required, annotated the images obtained. Then trained a MASK RCNN model with RESNET50 architecture with increasing dataset sizes, epochs in two variants. Also, the model is tested with increasing validation sets and considering the average as the final result.

II. LITERATURE REVIEW

Chen H, et al, (2019) [6], have taken a dataset of 1250 dental X-ray films of periapical view and have trained a Faster RCNN model. They have done some post-processing to reduce the complexity of data for faster training of the model. Also, they suggest a filtering algorithm to delete overlapping boxes detected by their model. They have computed their results using IoU (Intersection over Union) score to calculate average recall. However, their work is limited to bounding boxes only which cannot be used for treatment with precision. The current work strives to improve the detection at pixel level to enable computer understand the exact position and region of treatment.

Jie Yang, et al, (2018) [27], have used a small dataset of 196 dental radiographs. They have said in their work that dental radiographs are important for clinical diagnosis and treatment. They have taken two classes for the condition of teeth. A CNN (convolutional neural network) model was constructed, and the F1 score was the measure of evaluation. Although, the author demonstrates results with a small dataset the accuracy is not satisfactory in terms of real time treatment. Taking the model into consideration, the current work intends to obtain the results which are more accurate and required for real time operations.

Kim J, et al, (2019) [19], have used a huge dataset of 12,179 panoramic dental radiographs to train a deep convolutional neural network. They have suggested a different measuring system where they have considered an increasing validation set count and finally computed the average F1 score. The mechanism to calculate results proposed by the author is unique. As such a similar way of validation system is implemented to summarize the end results of the current work.

III. PROPOSED MODEL

As shown in Fig. 1, the project has four steps:

- Data Collection and processing
- Data Annotations
- Model Training
- Testing

In this work, data is collected, i.e., dental periapical view radiographs from different dental clinics in Hyderabad, Telangana, India [10]. After the collecting step, the data is processed, as shown by Chen H, et al, (2019) [6]. The dimensionality of images has to be reduced for less computational complexity, and also resizing images is a must for a model to accept them as training input. Then proceed to the data annotations step, where annotations are done for required objects in an image. An image can contain more than one object. After the annotations are done, the corresponding files are passed to the model for training. The model is trained with certain parameters and is passed to the testing stage. The IoU threshold in the testing phase is set to 50% and AP50 (average precision) score is obtained.

Fig. 1. Proposed Model Workflow.

A. Data Collection

In this work, 1000 dental radiographs of periapical view have been collected from different dental clinics in Hyderabad, Telangana, India. The dental radiographs collected are images of "JPG" format with four channels RGBA. The images are processed by resizing to 416x416 dimensions to reduce the dimensional complexity. Also, the alpha channel is being eliminated, i.e., the images are converted from RGBA to RGB data. The images have a total of three classes: "endodontic", "restoration" and "implant".

B. Data Annotations

The images collected have to be annotated. Annotation indicates creating a mask around the desired object for the algorithm to identify and train on. Image annotations are done using "Labelme" tool. "Labelme" tool, as shown in Fig 2, is an open-source manual annotation creator. It provides us with the "create polygons option" that is used to create a mask around desired objects. A single image can have more than one desired object and of different class. A json file is generated containing the mask coordinates and the class identity name for every image annotated.
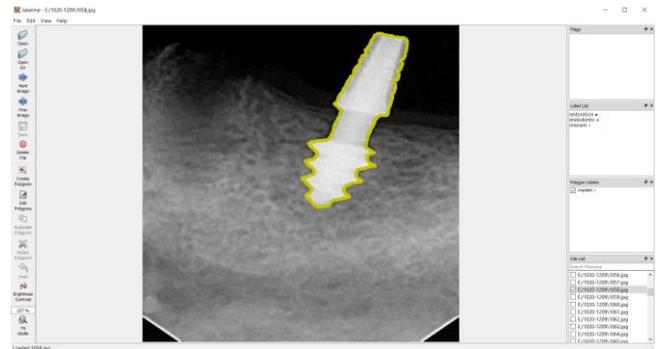

Fig. 2. Labelme Tool Interface.

C. Model Training

After the annotations have been completed, combine all the json files to a single json file for the model to train on. A python script is used for the conversion. All the images along its corresponding json file are uploaded into google drive. The main model used to train is Mask RCNN which is capable of instance segmentation [11], [15]. The model was trained with base learning rate of 0.002, with 600 epochs and 10 images per batch. The model weights are the "COCO" instance model weights which are updated through transfer learning [24], [25].

IV. MODELS

A. Mask RCNN

Mask RCNN is a deep neural built to process instance segmentation i.e., it can identify objects at pixel level. Mask RCNN does both object detection and instance segmentation. As the work is aimed at identifying objects in real time at pixel level, the algorithm implemented is Mask RCNN. Mask RCNN has two stages of training. Initially, region proposals are generated where objects might be present in an input image. Now, in the second stage, these proposals are used to predict the object class and construct a bounding box around the object detected. The bounding box is refined and a mask is generated

at pixel level for the proposal in the first stage. Both these stages are connected to the backbone architecture.

The backbone architecture used in this project is "RESNET50" [17], [18]. A backbone architecture is a feature pyramid network-style deep neural network. "RESNET50" architecture is a bottom-up pathway that extracts features from the input raw images. Fig. 3 shows the architecture of Mask RCNN algorithm.

In the first stage, use a region proposal network to propose the regions where objects are expected. Now a feature map is generated through the proposals to bind the features to their locations in raw images. Now, use anchors, a set of boxes with predefined locations that are scaled with respect to the images. Use anchors of different size to bind the different levels of feature map, the region proposal network uses these anchors to Fig out the region of the feature map that should get an object and what dimension the bounding box should be. Convolving, down sampling and up sampling of feature map would retain the relative positions of objects in the original image.

In the second stage, take another neural network that takes proposed regions as input which is generated by the region proposal network in stage one [26]. The pixel level classification is done by fully connected network (FCN) as shown in Fig. 4. These proposals are now taken by a new neural network and assigned to several specific areas of a feature map level. These areas are scanned and object classes are generated with bounding boxes and masks. The main trick used here is the "ROI align" (Region of interest) used to locate the relevant areas of the feature map. A branch of "ROI align" is used to generate masks at pixel level for an image.
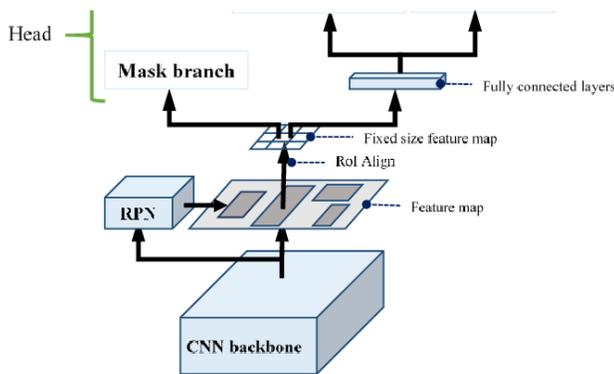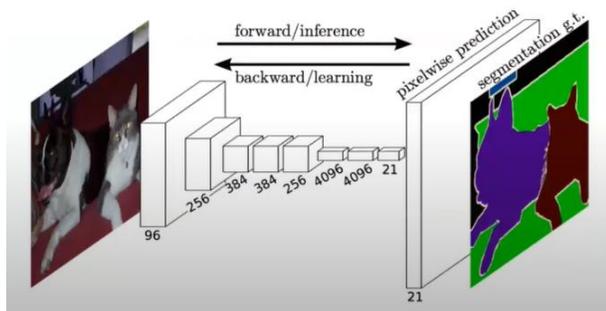

Fig. 3.   Mask RCNN Architecture.

Fig. 4.   FCN Showing Pixel Level Classification.

In "Faster RCNN" it uses "ROI pooling" for the case quantization where the bounding box dimensions become the desired dimensions [12]. In case of "ROI align" there is no necessity for quantization for data pooling [6]. As shown in Fig. 5 the dimensionality has to be changed in mapping or pooling process requiring the use of bilinear interpolation.
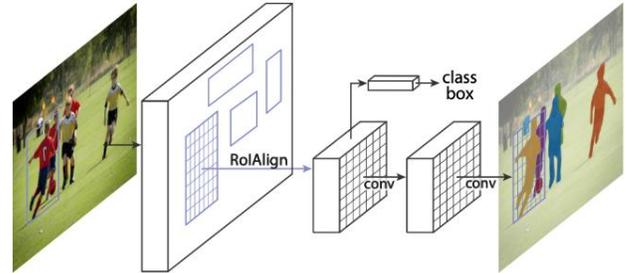

Fig. 5.   ROI Align in RESNET50 Architecture.

B. RESNET50

Resnet50 algorithm was made with the aim of solving vanishing gradient problem [20]. Vanishing gradient problem is a case in neural networks where the loss when sent back to update weights of the nodes, the differential value calculated becomes so small that the weights no longer get updated. In such a case the loss of the model does not decrease and the model is not capable of learning anymore. In case of RESNET models the input and processed input are both sent to the next stage of layers and the same happens when loss is propagated through the model. This concept is known as "skip connection".

If 'x' is the input and f(x) is the function output. As neural networks are good function approximators, they should be capable to identify a function where the output becomes input itself.

$$f(x) = x \qquad (1)$$

Assuming this if the input of the first layer of model is bypassed to be the output of the last layer of the model, the network should be able to predict the function it was learning before with the input added to it.

$$f(x) + x = h(x) \qquad (2)$$

With RESNET one can pass the gradients through skip connections to the next layers of a model and from end layers to initial layers. Resnet50 deals with 48 convolutional layers along with 1 "MaxPool" and 1 "Average Pool". It all starts with a convolutional layer with kernel size 7*7*3 with dimension of 64 and then stride 2. Then "MaxPooling" is done where for each feature map the maximum number is extracted to reduce the dimension of the image, which also reduces noise and then enter into the main Resnet layers in which they perform the convolutions with kernel size of 3*3*3 with the dimension set of 64, 128, 256 and 512. For every 2 convolutional blocks the skip connection takes place where the input and output value doesn't change.

The dimension changes with stride 2 during the skip connection, like from 64 to 128 then dimension change causes

problem at that particular time. In any case one convolutional block is added to the skip connection process for dimension change to take place without mismatch and errors. After all the RESNET layers are traversed, "Average pooling" is done where the average value is taken from a particular feature map and gives the output. This helps in smoothening of the image but this method cannot be used for sharp images. Fig. 6 showcases the concept of "skip connection".

Fig. 6.   RESNET Model Skip Connection.

After that, the obtained feature map is flattened and then passed to an artificial neural network. The process continues where certain weights are assigned with the inputs and sent to the successive layers, known as forward propagation. Loss is calculated at the output layer which is back propagated to the previous layers to update weights for improving the model. Fig. 7 shows a demonstration of RESNET50 architecture.
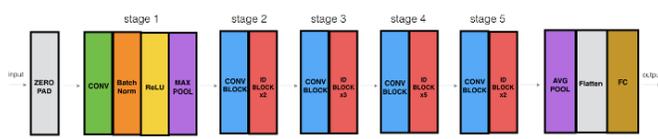
Fig. 7.   RESNET50 Architecture.

V.   RESULTS

The results for the instance segmentation problems are measured using the average precision scores. For measuring any classification object, preliminarily calculate "True Positive" (TP), "True Negative" (TN), "False Positive" (FP) and "False Negative" (FN). In instance segmentation, calculate them by using Intersection over Union (IoU) score as shown in Fig. 8. As a mask is being generated around the objects in the input image and have the ground truth of the object, try to find the area of intersection of the two masks and calculate the area union of the two masks. These values are divided to get Intersection over Union score.

The Intersection over Union score is now used as along with a threshold where if the IoU score is greater than threshold then it is considered as "True Positive" and if the IoU score is less than threshold then it is considered as "False Positive". If an object is not being recognized, it is a "False Negative" and a wrong classification is considered "False Positive". Fig. 9 is a sample of area covered by bouding boxes and ground truth to calculate IoU score for classification.

Using these values, the precision values are calculated. Precision is the measure of true positive detections.

$$Precision = \frac{TP}{TP+FP} \qquad (3)$$

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{}{}$$

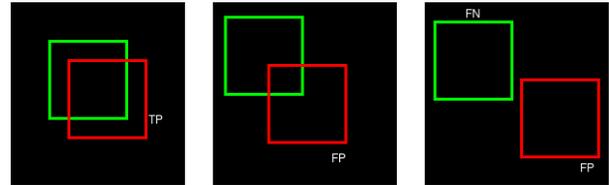Fig. 8.   Intersection over Union Calculation.

Fig. 9.   IoU Score Classification.

In evaluations, 50% threshold is being considered. AP50 means the average precision at 50% IoU threshold.

The dataset has been split into parts of increasing order to observe the change in results as the dataset used for training is being improved. Taking the data sizes of 100, 250, 500, 750 and 1000 into consideration and testing over validations sets of 20, 50, 100, 150 and 200 [27], [29]. After calculating the results, the average value was calculated and was considered the final AP50 score. Starting with dataset of 100 the results obtained for bounding boxes are: 55.963, 54.636, 54.12, 46.881, 63.141 and for mask segmentation are: 58.186, 51.193, 51.649, 37.723, 56.268 respectively as shown in Fig. 10. The dataset was trained for 600 epochs as the loss was minimum and stable at that point.

Moving on to the dataset with 250 images; the results for bounding box AP50 scores are: 59.571, 55.252, 54.14, 51.654, 55.503 and for mask segmentation are: 59.571, 50.282, 49.546, 40.736, 46.59 as shown in Fig. 11. The dataset was trained for 600 epochs as well.
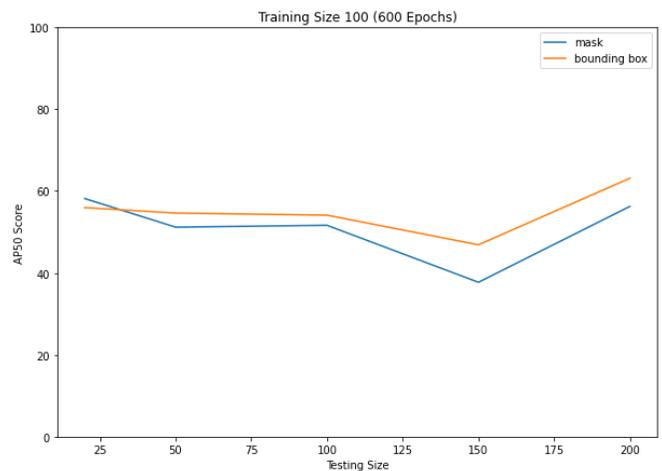

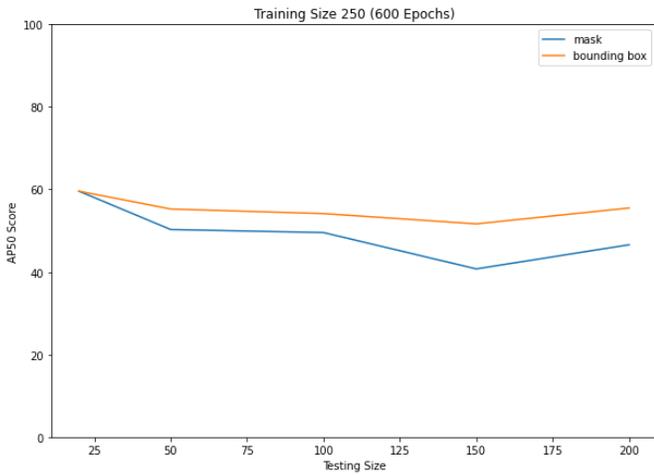Fig. 10. AP50 Scores Obtained for Dataset of 100 Images.

Fig. 11. AP50 Scores Obtained for Dataset of 250 Images.

Moving on to the dataset with 500 images; the results for bounding box AP50 scores are: 75.596, 71.929, 76.32, 60.471, 76.283 and for mask segmentation are: 69.166, 64.899, 72.075, 47.209, 66.472 as shown in Fig. 12. The dataset was trained for 600 epochs as well.
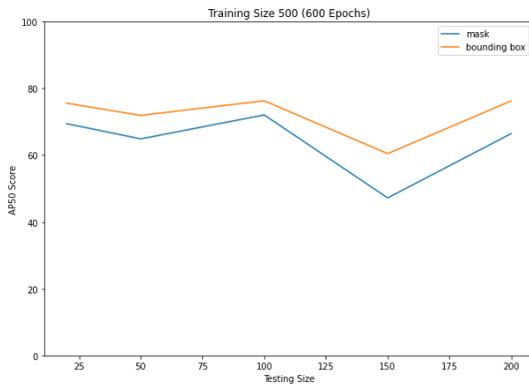

Fig. 12. AP50 Scores Obtained for Dataset of 500 Images.

Moving on to the dataset with 750 images; the results for bounding box AP50 scores are: 83.168, 81.3, 79.203, 68.126, 80.022 and for mask segmentation are: 85.269, 78.019, 75.605, 57.17, 71.194 as shown in Fig. 13. The dataset was trained for 600 epochs as well.
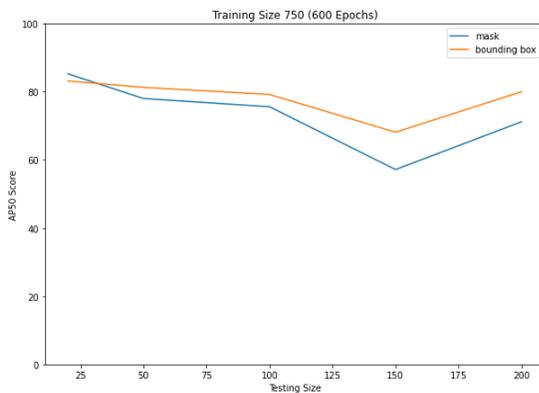

Fig. 13. AP50 Scores Obtained for Dataset of 750 Images.

Finally, the dataset with 1000 images; the results for bounding box AP50 scores are: 84.334, 82.489, 78.091, 84.083, 83.28 and for mask segmentation are: 86.883, 76.708, 75.186, 72.115, 76.029 as shown in Fig. 14. Although, the dataset was trained for 600 epochs there was still trainable parameters.
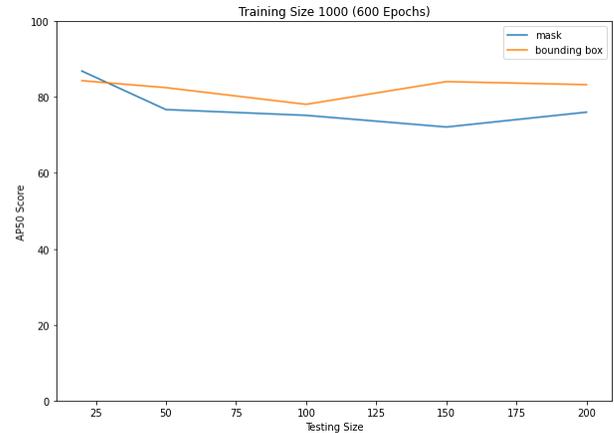

Fig. 14. AP50 Scores Obtained for Dataset of 1000 Images for 600 Epochs.

The model trained utilized the transfer learning concepts and the weights were imported from the "COCO dataset" model. Transfer learning is the concept where the weights are imported by training a model on a different dataset. This should help the new model update weights in less epochs. For the model trained with 600 epochs, observations seen were best bounding box and mask segmentation average AP50 score with 1000 images dataset.
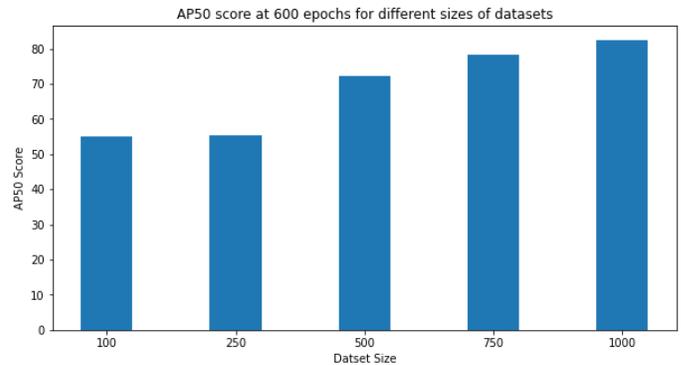

Fig. 15. Comparing AP50 Score of different Dataset Sizes with respect to Bounding Boxes.

As continuing with observations, the 1000 images dataset has obtained AP50 score of 82.4554 for bounding boxes and 77.3842 for mask segmentation at 600 epochs. Fig. 15 compares the AP50 scores of different datasets and it is observed that 1000 images dataset has the best results yet, however it is also inferred that improvement in results with respect to dataset size declines exponentially.

Further, the 1000 images dataset was in two variations:

- With transfer learning (1200 epochs).

- Without transfer learning (1600 epochs).

When trained with transfer learning for 1200 epochs on 1000 images dataset the model gave AP50 scores of 88.4248 and 84.2662, respectively for bounding boxes and mask segmentation as shown in Fig. 16.

Also, for training without transfer learning, the weights are initiated and updated from scratch. Due to which the model needed to be trained for 1600 epochs where the loss stabilizes. The results obtained at this stage are 76.1696 and 72.123 for bounding boxes and mask segmentation as shown in Fig. 17.

In Fig. 18, the results of "with transfer learning" and "without transfer learning" are compared and it is observed that "with transfer learning" model gives better results. Additionally, as observed that with an increase in the dataset size, the results also improve. So, another dataset consisting of 1755 images is considered.

The 1755 images dataset was trained for 1200 epochs with transfer learning and obtained an AP50 score of 88.2416 and 82.1224 respectively for bounding boxes and mask segmentation as shown in Fig. 19.
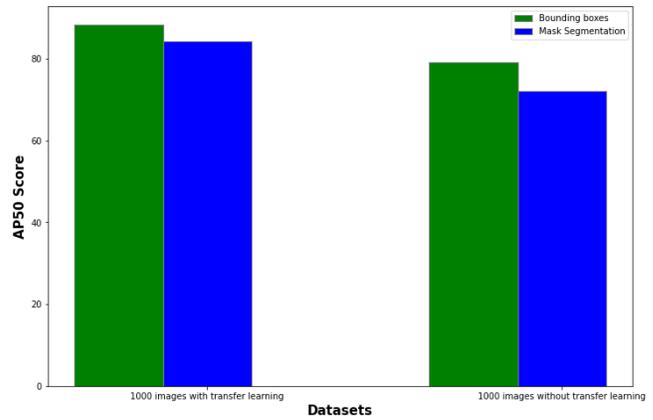

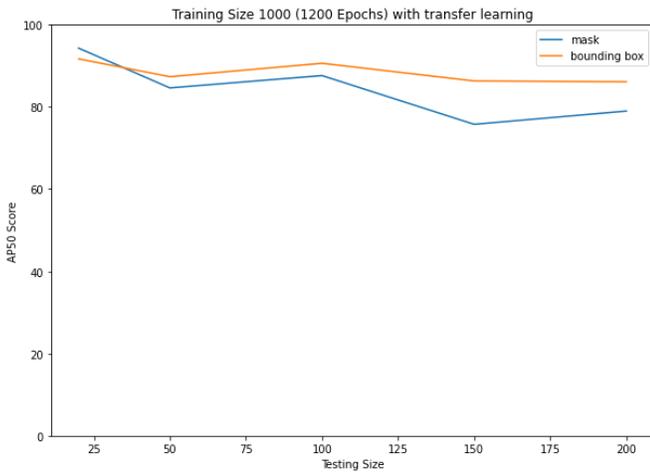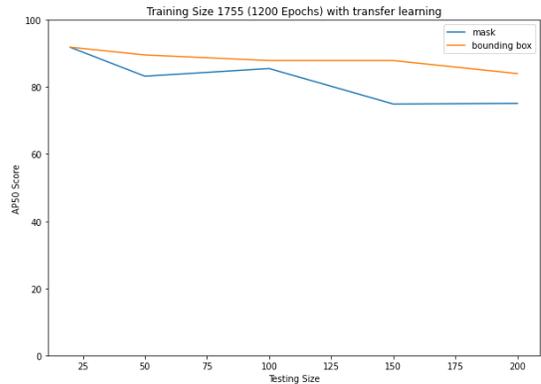Fig. 18. Comparing Results of 1000 Images with and without Transfer Learning.

Fig. 19. AP50 Scores Obtained for Dataset of 1755 Images at 1200 Epochs.

Also, the 1755 images dataset is trained without transfer learning for 1600 epochs and obtained 82.3398 and 62.7846 AP50 scores respectively for bounding boxes and mask segmentation as shown in Fig. 20.

Fig. 16. AP50 Scores Obtained for Dataset of 1000 Images at 1200 Epochs.

Fig. 17. AP50 Scores Obtained for Dataset of 1000 Images at 1600 Epochs without Transfer Learning.
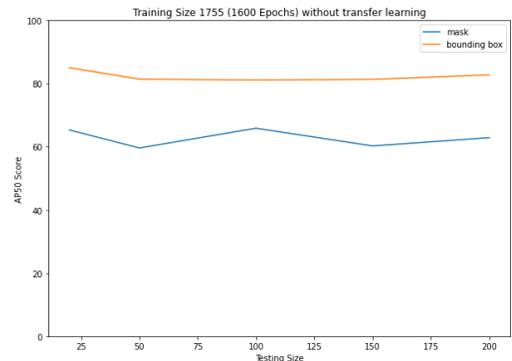

Fig. 20. AP50 Scores Obtained for Dataset of 1755 Images at 1600 Epochs without Transfer Learning.

Although the results obtained seem stable and do not alter much, the size of the dataset showed that the model could be trained further and improved. The 1755 images dataset was trained further for 2400 epochs, and results obtained showed improvement. It obtained AP50 scores of 87.9548 and 76.1272 for bounding box and mask segmentation, respectively as shown in Fig. 21.
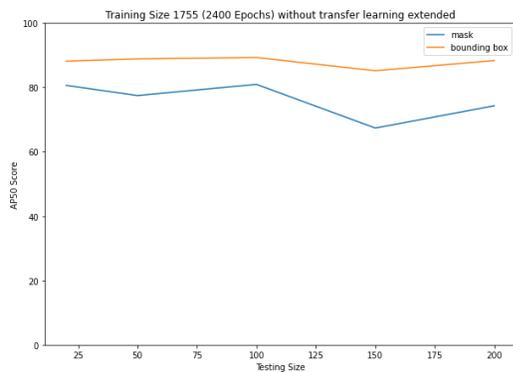
Fig. 21. AP50 Scores Obtained for Dataset of 1755 Images at 2400 Epochs without Transfer Learning.

As observed, the results have improved and altered less, showing that the complexity change in images will not affect the model much. Also, as the 1000 images dataset has obtained better results for mask segmentation, the results may vary depending on the complexity of images but with less intensity.
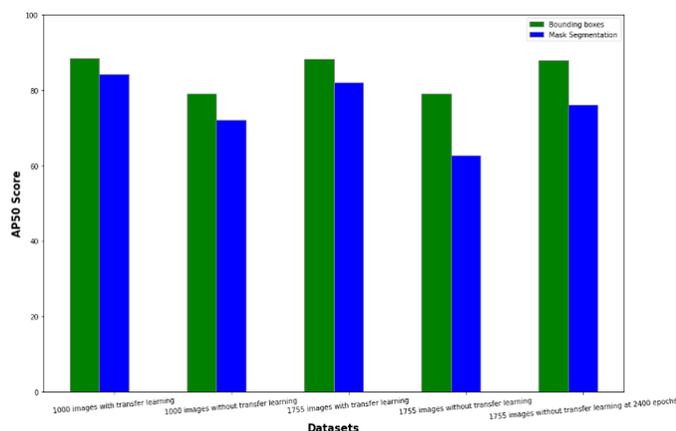

Fig. 22. Comparing AP50 Scores of 1000 and 1755 Images Datasets.

From Fig. 22, one can observe that the bounding box AP50 was highest for the 1755 images dataset trained without transfer learning for 2400 epochs but the mask segmentation AP50 score was highest for the 1000 images dataset with transfer learning at 1200 epochs. One can still observe that results are stable and vary less when there is an increase in dataset size with increased epochs.

VI. Conclusion

The need for expertise in dental clinics is a must. Making the process of taking and reading radiograph must be improved. Also, robotic treatment in dental science needs a system to understand and identify dental objects at the pixel level. This work provides a model that can identify the dental objects at pixel level and has obtained an Average Precision-50 score of 84 and 88 for mask segmentation and bounding boxes respectively for 1000 images dataset at 1200 epochs. The model used in this work is Mask RCNN with RESNET50 architecture. The training was done in two variants of with transfer learning and without transfer learning and with increasing dataset sizes and epochs.

The work demonstrates the mask generated around a tooth with pixel level accuracy which can be further developed for robotic treatment in the future. The model showcases results which are considerable to further explore with more resources for better results that are satisfactory for medical treatment. The AP50 scores of 1000 images dataset prove to be good when trained for 1200 epochs. While the 1755 dataset at 1200 epochs give AP50 scores of 88 and 82 for bounding boxes and mask segmentation. Increasing the number of epochs without transfer learning give similar results for bounding boxes but mask segmentation has reduced by a lot. The results of 1755 dataset trained without transfer learning for 1600 and 2400 epochs give 82 and 87 AP50 scores respectively but the mask segmentation scores have declined to 62 and 76. Further training of model may improve the scores but for more better and practical results it is suggested to increase the dataset size.

The work can be further improved by taking a larger dataset, as observed there is an improvement in results with an increase in dataset size. Also, in this work, the classes of objects have been generalized to three. More specific classes can also be taken for object detection with the help of a dentist or a person with good knowledge in dental science, and work can be improved further.

REFERENCES

[1] Abdel-Mottaleb, J. Z. (2005). A content-based system for human identification based on bitewing dental X-ray images. Pattern Recognition, 2132-2142.

[2] Abdel-Mottaleb, M. H. (2005). Classification and numbering of teeth in dental bitewing images. Pattern Recognition, 577-586.

[3] Abdel-Mottaleb, O. N. (2005). A system for human identification from X-ray dental radiographs. Pattern Recognition, 1295-1305.

[4] Anil K. Jain, H. C. (2004). Matching of dental X-ray images for human identification. Pattern Recognition, vol. 37, 1519-1532.

[5] Anny Yuniarti, A. S. (2012). Classification and Numbering of Dental Radiographs for an Automated Human Identification System. TELKOMNIKA, 137-146.

[6] Chen H, Z. K. (2019). A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Sci Rep 9.

[7] Chen, A. K. (2004). Matching of dental X-ray images for human identification. Pattern Recognition, 1519-1532.

[8] Chen, H., Zhang, K., Lyu, P. et al. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Sci Rep 9, 3840 (2019). https://doi.org/10.1038/s41598-019-40414-y

[9] Everingham, M. V. (2010). The PASCAL Visual Object Classes (VOC) Challenge. Int J Comput Vis 88, 303–338.

[10] Geert Litjens, T. K. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 60-88.

[11] G. Zhu, Z. Piao and S. C. Kim, "Tooth Detection and Segmentation with Mask R-CNN," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 070-072, doi: 10.1109/ICAIIC48513.2020.9065216.

[12] He, K. a. (2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1904-1916.

[13] Huang, P. L. (2010). An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information. Pattern Recognition, 1380-1392.

[14] Jindan Zhou, M. A.-M. (2005). A content-based system for human identification based on bitewing dental X-ray images. Pattern Recognition, vol. 38, 2132-2142.

[15] Kaiming He, G. G. (2017). Mask R-CNN. Clinical Orthopaedics and Related Research.

[16] Kandelman, D. a. (2012). Oral health care systems in developing and developed countries. Periodontology 2000, 98-109.

[17] Khalifa, M. L. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable Cities and Society.

[18] Kim, C.; Kim, D.; Jeong, H.; Yoon, S.-J.; Youm, S. Automatic Tooth Detection and Numbering Using a Combination of a CNN and Heuristic Algorithm. Appl. Sci. 2020, 10, 5624. https://doi.org/10.3390/app10165624

[19] Kim J, L. H. (2019). DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. Sci Rep 9.

[20] Kumar, S. R. (2021). Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images. Chaos, Solitons & Fractals.

[21] Nomir Omaima, A.-M. M. (2007). Human Identification From Dental X-Ray Images Based on the Shape and Appearance of the Teeth. IEEE Transactions on Information Forensics and Security, 188-197.

[22] Omaima Nomir, M. A.-M. (2005). A system for human identification from X-ray dental radiographs. Pattern Recognition, 1295-1305.

[23] S. Tohnak, A. M. (2007). Synthesizing Dental Radiographs for Human Identification. Journal of Dental Research, 1057-1062.

[24] Said, E. a. (2006). Teeth segmentation in digitized dental X-ray films using mathematical morphology. IEEE Transactions on Information Forensics and Security, 178-189.

[25] Shashikant Patil, V. K. (2019). Algorithmic analysis for dental caries detection using an adaptive neural network architecture. Heliyon, vol. 5.

[26] Tsung Yi, e. a. (2016). Feature Pyramid Networks for Object Detection. Clinical Orthopaedics and Related Research.

[27] Yang Jie, X. Y. (2018). Automated Dental Image Analysis by Deep Learning on Small Dataset. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 492-497.

[28] You W, H. A. (2020). Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. BMC Oral Health 20.

[29] Zhang, Z. L. (2018). Large-scale retrieval for medical image analytics: A comprehensive review. Medical Image Analysis, 66-84.