SERBIAN JOURNAL OF ELECTRICAL ENGINEERING
Vol. 22, No. 2, June 2025, 281-307
UDC: 004.932:616.314 **DOI: https://doi.org/10.2298/SJEE2502279P**
Original scientific paper

Redefining Dental Image Processing: De-Convolutional Component with Residual Prolonged Bypass for Enhanced Teeth Segmentation

Kumar Prasun[1], Anil Verma[1], Rajiv Mishra[1]

Abstract: Dental diseases have risen in the past few years due to improper hygiene. Early detection and diagnosis can control this rapid growth in dental diseases. Therefore, different traditional techniques are employed for the detection of dental problems. However, these classical techniques such as X-Ray and CT scans are considered to be time-consuming, ineffective, and prone to errors due to human intervention. Hence, AI techniques are used to obtaining precise outcomes for dental-related issues. The conventional ML (Machine Learning) techniques are inefficient for obtaining enhanced outcomes as the efficiency of ML techniques heavily depends on image processing approaches. They are performed and also the quality of the features that have been extracted. Further, ML techniques lack in producing better outcomes while dealing with huge datasets. Therefore, the proposed model employs DL (Deep Learning) techniques due to its capability to learn the features strongly from the data by using a general-purpose learning procedure. So, DL techniques can work efficiently on huge datasets. The proposed DC (De-convolution Component) with RES (Residual Prolonged Bypass) is employed in the present research work as it is responsible to increase the spatial resolution of the feature maps and helps in recovering lost spatial information during the down sampling process. Likewise, the RES model aids in proficiently proliferating both low-level and high-level features to the deep layers, which help in generating better-segmented images. RES model includes prolonged bypass paths that carry feature information across multiple layers. This ensures that features extracted at earlier layers (low-level features) are available at much deeper layers. Implementation of the present research work contributes to enhancing the overall performance and effectiveness in detecting and diagnosing various dental issues and possesses the capability to work on both small and massive datasets effectively. Also, the proposed work contributes to deliver better accuracy, IoU

[1]Deparment of Computer Science and Engineering, Indian Institute of Technology, Patna, India;
kumar_2422cs12@iitp.ac.in, https://orcid.org/0009-0008-1363-4834;
anil_2321cs11@iitp.ac.in, https://orcid.org/0009-0004-0457-8966;
rajivm@iitp.ac.in, https://orcid.org/0000-0002-4910-5749

Colour versions of the one or more of the figures in this paper are available online at https://sjee.ftn.kg.ac.rs

©*Creative Common License CC BY-NC-ND*

(Intersection Over Union) and Dice coefficient, compared to Multi-Headed CNN and Context Encoder-Net, thereby assisting dental professionals in the detection and diagnosis of various dental issues due to the effectiveness of the proposed model.

Keywords: Dental, UNet, Convolutional neural network, Residual Prolonged Bypass, Tufts dataset, Attention mechanism.

1 Introduction

Dental diseases have high risk of occurrence; therefore, detecting them is crucial to avoid further complications. Various techniques, such as X-ray, CT scans and other manual methods, are employed for the detection of dental issues [1, 2]. However, manual detection can be time-consuming and can be prone to errors due to human interpretation. Hence, AI-based techniques such as ML and DL are used. However, the efficiency of ML techniques heavily depends on the image processing techniques that are performed and the quality of the features that have been extracted. Therefore, DL techniques are used due to their capability to learn the features directly from the raw data by employing a general-purpose learning procedure [3].

Segmentation of teeth is considered one of the fundamental and crucial practices in dentistry for the diagnosis and treatment of various teeth-related problems [4]. Therefore, TSAS (two-stage attention segmentation network) was used on dental X-ray images [5] to address concerns related to tooth segmentation, which are usually caused by the low contrast of images. The experimental outcome demonstrated that results attained by the TSANet employed segmentation were much better than the existing models on dental P X-ray images [5]. Segmentation is, therefore, a technique primarily used to find the region of a particular area. Hence, effective medical image segmentation should be employed to enhance the segmentation process. As an alternative, encoder-decoder DCNN model for segmentation has been used [6]. The encoder-decoder DCNN utilized an additional skip connection. When compared to the existing ResNet model, encoder-decoder DCNN delivered better outcomes for segmentation. However, the encoder-decoder DCNN employed fewer parameters than DenseNet. Effectivity of the DCNN model was evaluated by comparing it with the ResNet-based, AttnNet and DenseNet-based models, where it was found that the encoder-decoder DCNN delivered better segmentation results compared to these alternatives [6].

It is quite significant that teeth are segmented from DRR (Digital Restoration Rendering) in restorative dentistry and orthodontics. There are a few limitations in X-ray images, including blurred boundaries. To overcome these inevitable issues, the U-Net model is introduced [7]. U-Net is a convolutional neural network architecture primarily designed for semantic segmentation tasks,

particularly in the field of biomedical image analysis. Its unique structure, characterized by a contracting path (encoder) and an expansive path (decoder) enables it to effectively capture and reconstruct spatial information. Further, the squeeze-excitation module is engaged in the encoder-decoder mechanism. The existing work has possessed drawbacks like blurred, low contrasted images with noisy pixels. So, the proposed model would overcome these limitations. Further, skip connection was employed for a residual semantic gap. Besides, Multi-scale Aggregation attention Block was used in the bottleneck layer given that the image was characterized by low-contrast features and irregular shapes of teeth. The effectiveness and efficiency of the model was evaluated using a clinical dental PX-ray image dataset. The outcome of the recommended model was compared to the existing models, and the U-Net technique utilized various evaluation techniques such as context semantics, contrast enhancement, data augmentation, Multi-Scale Feature Extraction for assessing the efficiency of the model by using Dice, Volumetric overlap error and relative volume difference for dental P X-ray segmentation of teeth [7]. Analyzing dental radiographs is considered one of the most important aspects of clinical practice. Hence, CNN model is used [8] because it aided in easily detecting and numbering teeth. A dataset utilized in the study comprises of 1352 panoramic radiographs of adults to train the model. And, the spatial arrangement of teeth was enhanced by utilizing the VGG16-Net model. Eventually, the model was evaluated using sensitivity, specificity, and precision. From the outcome, it was identified that a computer-aided model improves in detection and numbering of teeth effectively, thereby saving time and benefits in enhancing the effectiveness of the model.

Though existing studies have delivered significant results for effective segmentation, they still lagged in delivering enhanced and accurate images due to the usage of ineffective algorithms. Thus, **Table 1** shows the limitations faced in the previous works.

Table1 depicts some of the pitfalls faced by the state-of-the-art approaches, such as Low IoU, Dice coefficient, overfitting of the model, inability to handle diverse datasets and exhibits low accuracy rate. They have a few training samples that can severely hinder the effectiveness of deep learning models making it crucial to ensure an adequate amount of high-quality training data for optimal result. Therefore, the proposed model utilizes UNet that comprises of DC with RES model. It aids in delivering accurate and clear segmented images to the corresponding input images (radiographs). RES technique utilized in the proposed model aids in minimizing the problems associated with vanishing and exploding gradients by skipping some layers due to the utilization of identity mapping and can efficiently propagate both high as well as low-level features to the DL. Finally, the model's performance is evaluated using different evaluation metrics.

Table 1
Limitations of Existing Work.

References	Model	Limitations
[3]	Context Encoder-Net model	Low IoU score of 0.8164 and dice coefficient of 0.8662 is gained by Context Encoder-Net model
[9]	CNN model	Inability to handle huge datasets
[10]	Multi-Headed CNN model	Low IoU and Dice coefficient of 0.918 is attained by Multi-Headed CNN model
[11]	Sparse voxel octree and 3D-CNN	The number of training data available is insufficient, which makes it difficult to achieve optimal performance with deep learning models.
[12]	CenterNet ResNet, EfficientNet and MobileNet.	Class imbalance problem is faced by the model
[13]	MobileNet V2	Accuracy obtained by the model is 0.87, which can be improved further by employing better models

The objectives of the research are:

– To accomplish accurate and flawless image segmentation from the corresponding radiographs using DC with RES model.
– To evaluate the performance of the proposed model using various evaluation metrics such as IoU and Dice coefficient, accuracy, precision, recall and F1-score.
– To assess the effectively of the proposed model by comparing it with the existing models.

1.1 Paper organization

Section 2 discusses traditional approaches used in a related field alongside various methods. Section 3 outlines the methodology employed in the proposed system. The results and achievements of the proposed method are presented in Section 4. Lastly, the conclusion and future plans for the proposed method are detailed in Section 5.

2 Literature Review

The subsequent sections outline various techniques employed by existing methods for teeth segmentation.

DL techniques have attained a lot of attention in current researches due to the impressive and promising outcomes in terms of prediction, detection, and classification. Likewise, different DL methods have been used for analysis of

panoramic dental radiographies, as these methods aid in delivering various solutions for medical professionals in terms of detection of affected area in the images. However, poor quality of images and doctor's fatigue can lead to various problems and eventually encumber the treatment. Therefore, this paper [14] employed automatic detection of teeth using DL techniques termed as ERFNet (Efficient Residual Factorized). CNN model was trained using annotated data in order to attain the semantic segmentation. Further, various image processing techniques have been employed for segmentation and assist in improving the bounding boxes used for detecting teeth. Finally, Various metrics were employed for semantic segmentation, which includes recall, precision, accuracy, and F1-score [14]. Similarly, a study [15] employed DL method which employed instance segmentation technique of teeth. This method focused on segmenting each tooth in panoramic X-Ray images. The segmentation system was based on the mask region, which relied on CNN model in order to achieve instance segmentation. However, the computer vision tasks have been combined for detecting objects with pixel-level segmentation, which has allowed for identifying and allocating each object instances within an image. Dataset containing 10 different categories of buccal images with total of 1500 images was employed in this paper. The recommended study also utilized various evaluation metrics such as precision, recall, accuracy and specificity in order to assess the performance of the proposed model and was able to delivered 88% of F1-score, 94% of precision, 84% of recall and 99% of specificity over 1224 unseen images for teeth segmentation.

Correspondingly, DL based DeNTNET method [16] has employed panoramic dental radiographs. DeNTNet model was used for segmentation of teeth. Around, 12,179 images were utilized, in which, 11,189 images were trained by DeNTNet model, then 190 images were validated and finally around 800 panoramic dental radiographs were tested. The objective of the study [17] was to assess the usage of CNN system for identifying the VRF (Vertical Root Fracture) on panoramic images. CNN based DL model was employed for detection of VRF by employing DetectNet with the help of SW application DIGITS version 5. The reliability of the model was further increased by implementing 5-fold cross validation (CV) method. The dataset comprised of 300 panoramic images which were downloaded in JPEG format. Though the existing study aided in detecting the VRF panoramic images, the process of preparing the dataset and developing the model was regarded as time-consuming, but this challenge may be addressed in the future [17]. Moreover, dental caries is identified by employing MLP-NN (multi-layer perceptron -NN) [18] as large percentage of adults are affected by dental caries in recent years. Therefore, it is extremely crucial to detect the presence of dental caries as early as possible and it has been considered to be one of the challenging tasks for the dentist to detect the dental caries. Likewise, the BPNN (Back Propagation Neural Network) [18] with 10-fold cross validation has

been used for enhancing oral health management and identification in tooth decay issues. Therefore, reliable and sturdy diagnostic tools and digital radiography must be utilized for diagnosing the dental caries. In most cases, digital radiographies were employed for detecting various dental abnormalities as these images employed for the process had relatively low level of radiation, which aided in better segmentation process. In order to employ better segmentation technique, the Laplacian filtering, window based adaptive threshold, morphological operations, statistical feature extraction and back-propagation neural network model was compared with existing operations likely, SVM, KNN, NB, Bagging, RF, and XGBoost techniques and different metrics were utilized for assessing the performance of the MLP-NN model [18]. The CNN based DL model in DBR (Dental Bitewig Radiographs) was used in the study [19] by employing U-Net and VGG-16 architecture for both identification and segmentation of dental caries. This model aided the clinicians for identifying the tooth caries swiftly and dependably. The radiographic dataset which consisted of 621 ABR (Anonymized Bitewing Radiographs) was employed [19].

Different diagnostic tools such as PD-X ray imaging have been employed for diagnosis of various dental concerns. However, resolutions of PD-X rays were considered to be comparatively low. Hence the U-Net model [20] was implemented with the loss function, which demonstrated an advanced segmentation accuracy of teeth in PD-X ray images than the loss function obtained by the existing U-Net model. The dataset implemented in the recommended paper consisted of 162 PD-X ray images, in which 102 images were used for training and the remaining 60 images were used for testing. Lastly, DI (Dice Index) of 0.894 and JI (Jaccard index) of 0.809 have been obtained - from the experimental outcome, it was identified that the existing U-Net model revealed sophisticated segmentation accuracy of 0.864 and 0.927 [20]. Various segmentation techniques have been incorporated in the recommended study [21] which includes instance segmentation and semantic segmentation techniques along with numbering of teeth (TN). Mask R-CNN, ResNet, PA Net, HTC (Hybrid Task Cascade) were considered, in which the instance segmentation and the TN were found to be feasible by the network. Out of all these methods, PA-Net model was identified to deliver better outcome on SS [22]. The existing study implemented AlexNet, DenseNet169, VGG-19, ResNet50, MobileNetV3 and VGG-16. It was identified that DenseNet 169 was found to deliver much better precision rate than the PA-Net and aided in delivering performance for dental professionals [21].

2.1 Problem identification

From the assessment of the previously discussed work, core concerns of state-of-the art emphasized as below:

– The existing CNN [9] is considered to be incapable of handling huge datasets — it is essential to be able to incorporate larger datasets with a greater volume of diverse data. Utilizing more extensive datasets can provide the model with a richer variety of examples, thereby improving its ability to generalize and perform accurately across different scenarios.
– IoU score obtained by the existing mask RCNN model is low which results in ineffectual outcome for image segmentation [23].
– Reliance on limited datasets can result in generalization issues for diverse or underrepresented populations in clinical practice [17]. AI systems may have difficulty interpreting complex or subtly featured cases [18]. Integrating AI in clinical settings requires efforts to ensure it complements clinician expertise, while a tooth edge-weighted loss function might not address all segmentation variations [20].

3 Proposed Methodology

Dental diseases are inherently unavoidable, necessitating early detection and diagnosis. To achieve this, classical approaches such as CT scan and panoramic dental radiography were applied. But these methods were limited by their susceptibility to human error. To overcome these pitfalls, AI based strategies are opted. However, conventional ML techniques are considered to be ineffective for achieving robust and accurate model for abnormality detection because conventional ML algorithms are not as optimized as DL algorithms. Besides, the geometric variability and unique shapes of teeth further complicate the task making it difficult for traditional ML algorithms to accurately capture and differentiate features. So, we are proposing a model that employs DL techniques for accurate tooth segmentation by utilizing DC with RES model and demonstrate its performance on a Tuffs dental dataset. The overall flow of the proposed model is depicted in Fig. 1.

Initially, Tuft dataset is loaded and then, the data is pre-processed using various pre-processing techniques in order to eliminate noisy data and aid in handling the missing values as the raw data is considered to be unusable. Moreover, pre-processing helps in improving the quality of the input images and improves the shape, and texture of the images. Therefore, process of pre-processing is crucial in order to ensure that the input data is reliable and consistent. Further, the data is split in train and test parts, and the train data is passed to the next step which is segmentation to detect the affected region present in the image by using de-convolutional component with RES (Residual Prolonged Bypass) UNet architecture. The proposed model aids in achieving robust and accurate model for abnormality detection. It also assists in enhancing the dental image radiograph and image segmentation. De-convolution component employed in the proposed algorithm is presented in Fig. 2 that employs up-

convolutions present in UNet architecture to carry out up-sampling. The up-convolutions are employed to increase the spatial resolution of feature maps which allows the decoder to reenact the segmentation output at the original image resolution. In addition, RES is implemented in DC as it helps in minimizing the problems associated with vanishing and exploding gradients by skipping some layers due to utilization of identity mapping. Implementation of RES benefits the convergence of the training process of NN which makes the proposed model efficient for segmentation. Then, the image is passed to attention mechanism, which permits the model to extract the increasingly complex features from the data and helps in identifying the appropriate segmented images for the corresponding input images. Hence, the proposed model overcomes the problem of vanishing gradient issue, overfitting the model and the inability to capture full contextual feature information and handling complex data, thereby producing better segmented images. Finally, the performance metrics such as dice coefficient and IoU value is employed for evaluating the performance of the proposed model.

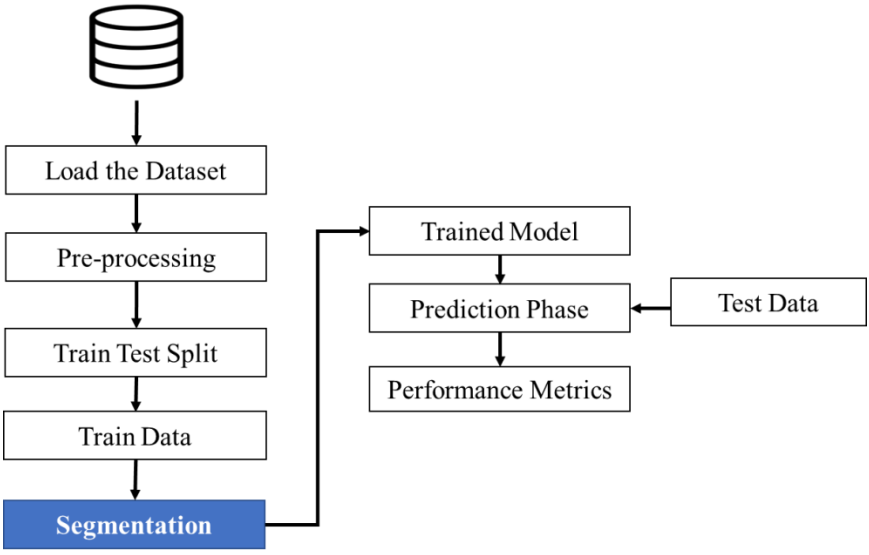


Fig. 1 – *Overall mechanism of the proposed work.*

3.1 Segmentation using proposed UNet model

UNet-model is specially employed for image segmentation tasks due to its immense capability of handling high resolution images and the ability to produce precise segmentation maps. Its graphical representation is depicted in Fig. 2.

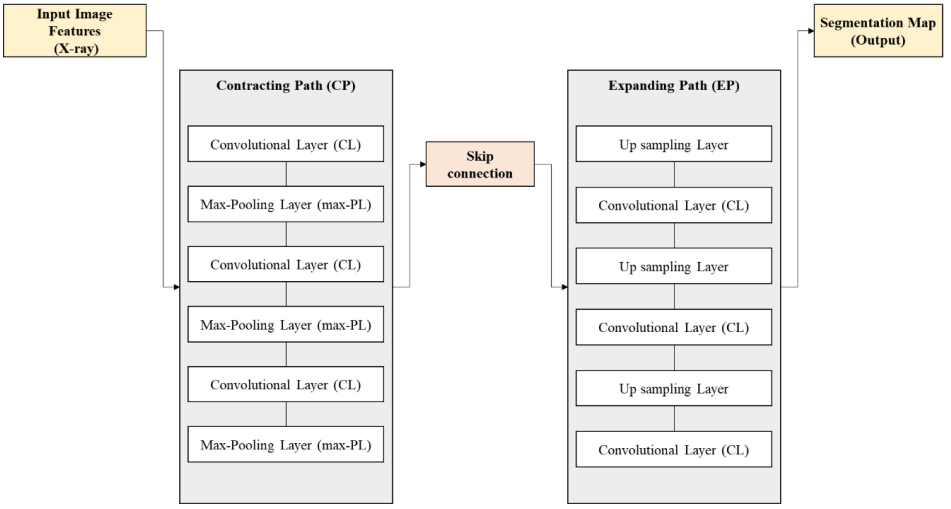


Fig. 2 – *UNet Model.*

The UNet architecture is trained end to end on a huge Tuft dataset of annotated images. The Tuft dataset contains 1000 selected radiographs that were identified and saved in a standard image format using a special identifier. These radiographs were meant to be analysed by a student and an expert from the Tufts University School of Dental Medicine. The dataset consists of six distinct elements, such as radiographs, eye-tracker maps, masks for the maxilla-mandibular region of interest, written descriptions for each radiograph, and labelled masks including masks for each individual tooth, all accompanied by their respective labels with the aim to predict a pixel level segmentation map for each image as shown in Fig. 3. The major components of the UNet architecture include expanding contracting path and skip connections as depicted in Fig. 3. The (CP) Contracting path consists of a sequence of CL and max pooling layer with down samples input images and aids in extracting the features from it. The CL applies a set of filters to input images and produces FM (Feature Maps), whereas the max pooling layer down samples the FM by attainting the highest value with the windows of pixels. The final segmentation is obtained by Upsampling the features through CP and coalescing them with the features that are obtained from the input images in the EP (Electronic Photography). Then CL applies a series of filters to the feature maps which are up-sampled with the purpose to build a final segmentation map. However, in contrast, spatial resolution of the FM is boosted by up-sampling layer by repeating the values within the window of pixels. Additionally, skip connections are used for bypassing one or more levels which expand the path and helps in linking them to the respective layers in CP. The conventional skipping connection assists in high level and low-level information from the input images, which needs to be

incorporated into the model and it benefits in increasing the precision of segmentation map. Further, conventional skipping connection aids in increasing the complexity and the requirement of the memory is humongous. Though the traditional model offers various advantages, it still lacks in different aspects such as potential for overfitting, poor generalization and difficulty in handling complex tooth structure. The proposed model, thus, utilizes DC with RES for resulting better image segmentation process, as goal of the model is to produce more accurate, precise, and reliable segmentation maps. This involves correctly identifying and classifying each pixel in the image to the appropriate segment. Hence, Fig. 3 shows the mechanism involved in proposed UNet Model.

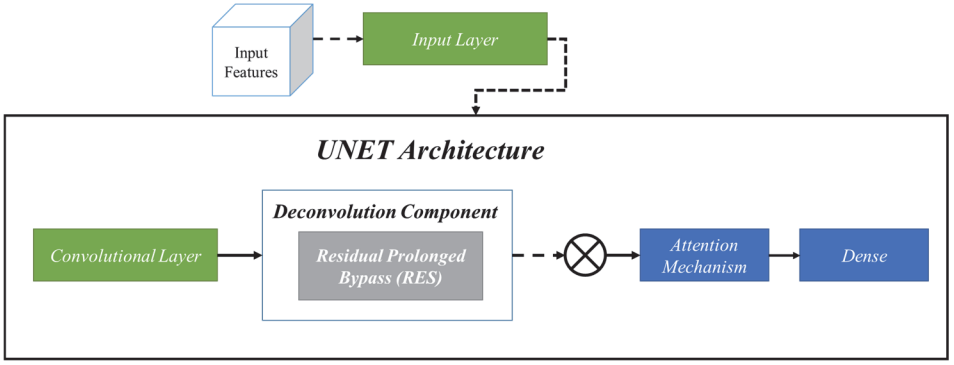


Fig. 3 – *Segmentation – De-convolution Component Utilizing Residual Prolonged Bypass (RES).*

Fig. 3 shows the process of segmentation, in which, the input features are passed to input layer. This input layer takes the input signals and then passes that input to the next layer. And, the output from the input layer is passed to Convolutional Layer (CL). CL aids in transforming input images in order to extract the features from them and provides flexibility in learning. In the next step, the values are entered into Deconvolution Component (DC). DC is a component which usually denotes the use of up-convolutions in UNet architecture with the aim to carry out up-sampling. Up-convolutions are utilized to proliferate the spatial resolution of the FM, which permits the decoder to reconstruct the output obtained during the process of segmentation at the original image resolution. After numerous pooling and convolutions, the FM becomes smaller and smaller which results in lower resolution of images. Therefore, with the aim to recuperate from the low-resolution image, skip connections are utilized for reducing the step size of the Up-sampling at Shallow Layer. Then, RES technique is utilized for minimizing the problems associated to vanishing and exploding gradients by skipping some layers due to utilization of identity mapping and can efficiently propagate both high level as well as the low-level features to the DL. Then, the value obtained from the bottleneck layer is passed

to attention mechanism (AM) section, which aids in selective focus on the most relevant parts of the input. This process of employing DC with RES technique is depicted in Fig. 4.

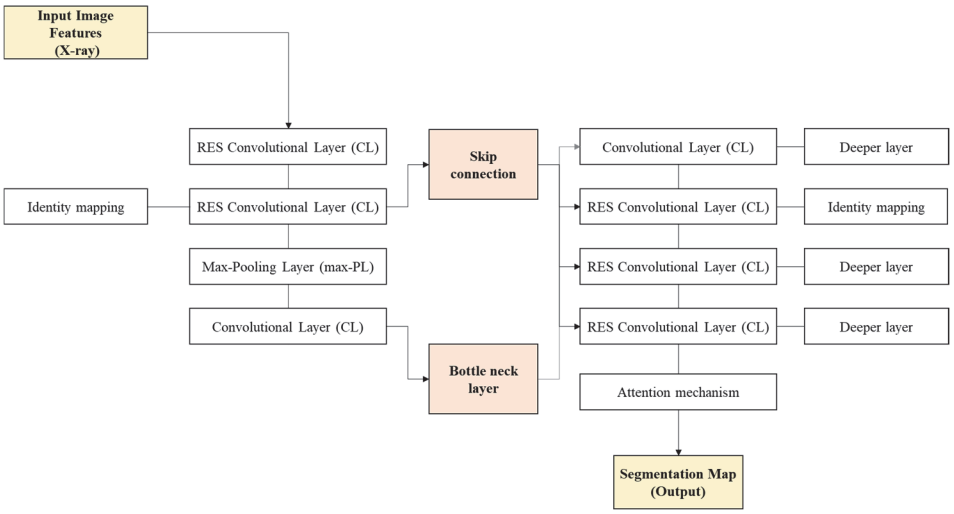


Fig. 4 – *Process involved in DC with RES technique.*

Here in Fig. 4, the skip connection from the initial RES convolutional layer goes to the convolutional layer directly above it. From there, the signal passes through the series of RES convolutional layer with each RES convolutional layer connected to the next one. The signal then reaches the attention mechanism, which processes the features before producing the final segmentation map output. Besides, the deeper layer blocks on the right side of the diagram represent additional convolutional layers that extract more complex features and the identity mapping blocks indicate that the input to these deeper layers is directly added to the output of the deeper layers and it is a common technique in residual networks to aid training and improve performance. Further, Fig. 5 showcases the gradient loss obtained by the proposed method.

In Fig. 5, dental image is fed into the neural network as the input x. In the next step, image goes through the first convolutional layer, which applies a set of filters to extract initial features, followed by ReLU activation function to introduce non-linearity. The output is then passed through second convolutional layer, which further processes the features extracted by the first layer. The result after passing through these layers is the intermediate output $F(x)$ which denotes the features learned by the network up to this point. The input image x is directly added to the output $F(x)$ of the CL through a skip connection. This addition helps preserve the original spatial information of the input image, which is significant for accurate segmentation. The output from the convolutional layers $F(x)$ and the

original iInput x are added together. This results in the final output of the residual block(y), which incorporates both learned features and the original image data. Final output obtained as $y = F(x) + x$ denotes the final feature map, which is employed for segmentation. This output will be further processed by subsequent layers in the network for producing final segmentation map.

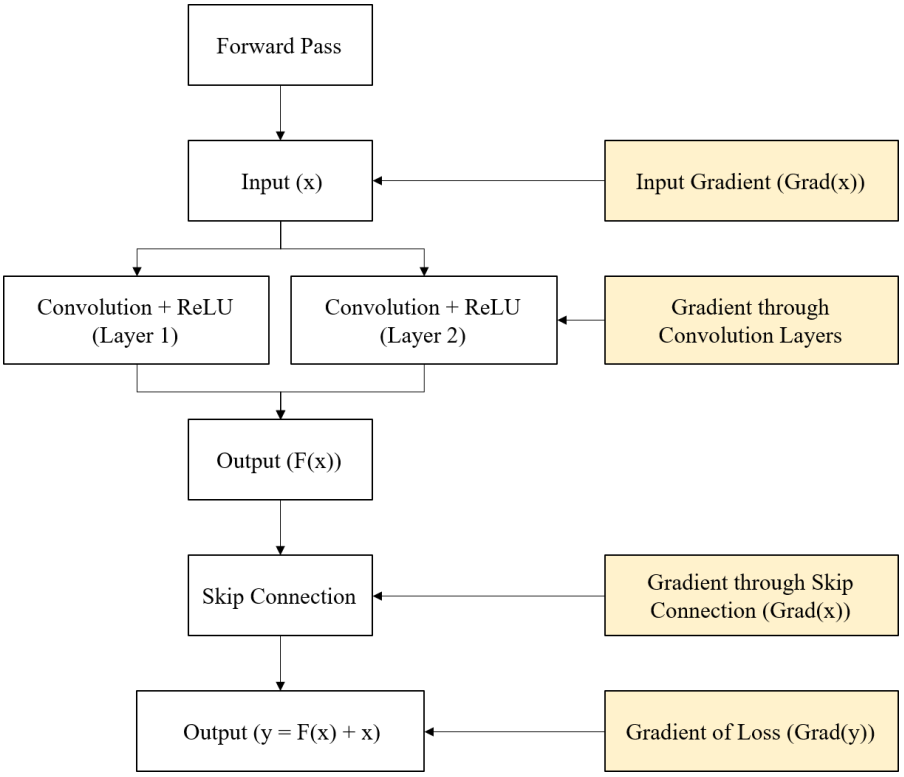


Fig. 5 – *Vanishing Gradient Computation.*

Moreover, on the right side, the yellow blocks represent the gradients being computed and propagated through the network. The "input gradient Grad(x)" indicates the gradient of the loss with respect to the input (x), while the "gradient of loss Grad(y)" represents the gradient of the loss concerning the output (y). Additionally, the "Gradient through Convolution Layers" shows the gradient calculated through the convolutional layers, and the "gradient through skip connection Grad(x)" reflects the gradient derived from the skip connection. This configuration facilitates efficient backpropagation by allowing the input to be directly added to the final output, enhancing learning and performance in deep networks.

Algorithm I showcases the process involved in DC with UNet for carrying out segmentation.

Algorithm I: RES Technique
Input • Itruftdataset: Input image from the truft dataset • u: truft • (a, b): coordinates in the image • W, H: width and height of the image • Q (class probability) **Output** • S: Segmented image • v = (s,t): Coordinates in the output segmented image • Q(c): Class probability for class c **Steps** **1. Input definition:** Itruftdataset = It where: $u = (a, b) \subset Y2$ with constraints: $0 \leq a < W, \quad 0 \leq b < H$ **2. Output definition:** $S\sim(v) = Q(c \mid It)$ where: $v = (s, t) \subset Y2$ with constraints: $0 \leq s < w/8, \quad 0 \leq t < H/8$ **3. Class probability:** Q(c) = class probability for class c **4. Mapped size:** The output segmented image S~ has a size of: Width = W/8, Height = H/8

Algorithm I shows the process involved with RES technique for segmentation process. The process begins with input processing, where an image from the truft dataset is taken and all points within it are ensured to adhere to its spatial bounds. This is followed by a segmentation process that applies techniques based on conditional probabilities derived from a model trained on these images. Each pixel or segment is evaluated using these probabilities, which are then used to assign class probabilities based on learned patterns from the training data. The output generation phase produces a segmented image with reduced resolution compared to the input, typically by down-sampling it. Finally, segments in this lower-resolution image are represented along with their respective class probabilities, providing a concise final representation of the original data at a reduced scale.

3.2 Proposed deconvolutional component with residual prolonged bypass

Residual prolonged bypass is primarily used for connecting the output of one previous CL to the input of another subsequent CL. RES aids in allowing the gradients to flow via network directly without passing through non-linear

activation functions. RES predominantly assists in the gradient vanishing problem. Hence, RES can be described as a technique utilized in the proposed model that helps with minimizing the problems of vanishing and exploding gradients by skipping some layers due to utilization of identity mapping, which can efficiently propagate both high level as well as the low-level features to the DL.

The residual unit generates $G(a)$ by processing input a through two weighted layers. The output $W(a)$ is obtained by adding a to $G(a)$. When considering $W(a)$ as the ideal predicted output that matches the ground truth, achieving this desired output depends on accurately obtaining $G(a)$. This indicates that the two weighted layers within the residual unit must effectively produce the required $G(a)$ which guarantees the ideal output $WG(a) = G(a) + a$.

$G(a)$ is obtained from a as follows,

$$a \rightarrow wt_1 \rightarrow \text{ReLU} \rightarrow wt_2. \tag{1}$$

In (1) the process starts with an input a, then wt_1 is applied to a, which adjusts its importance. Then, ReLU activation function is used, which makes all negative values zero and keeps positive values as they are. This helps simplify complex data. Then, another set of weights wt_2 is applied to further adjust the output. Equation (2) shows the residual connection in neural networks

$$G(a) + a \rightarrow \text{ReLU}. \tag{2}$$

In (2), $G(a)$ is resulted after passing a through layers (based on (1)). Adding a back into $G(a)$ creates a residual connection $G(a) + a$ and this helps in preserving information form earlier stages and improves learning by allowing gradients to flow more easily during training. Again, ReLU is used, which introduces non-linearity.

$$a \rightarrow wt_1 \rightarrow \text{ReLU} \rightarrow wt_2 \rightarrow \text{ReLU} \rightarrow a. \tag{3}$$

Here in (3), wt_1 and wt_2 represent the two weight layers (convolutional layers) within the residual block. Thereby passing through multiple layers, each applying weights and then using ReLU for simplification. Additionally, earlier versions of data $+a$ is used before continuing through more layers.

$$a \rightarrow wt_1 \rightarrow \text{ReLU} \rightarrow wt_2 \rightarrow \text{ReLU} \rightarrow 0. \tag{4}$$

Here in (4), represents a neural network's forward pass where input data a flows through multiple layers, each applying linear transformations via weights (wt_1 wt_2) followed by non-linear ReLU activations until reaching the final output. Therefore, the in simple terms, the equations (1 – 4) shows that neural network process information by adjusting inputs with weights and applying ReLU functions to simplify the data. Besides, residual connection used helps in retaining important details by adding original inputs back into later stages of processing. Algorithm II depicts the process in residual prolonged bypass.

Algorithm II: Residual Prolonged Bypass

Input parameters:

- a: Input feature map
- G(a): Output of the residual function
- wt_1, wt_2: Weight matrices for the layers
- ReLU: Activation function (Rectified Linear Unit)

Output parameters:

- W(a): Ideal predicted output

Steps:

1. **Initial processing:** W(a) = G(a) + a
2. **Ideal prediction:** W(a) = ideal predicted output
3. **Processing through Layers:** a → wt_1 ReLU → wt_2
   This indicates that the input a is processed bu weight wt_1, followed by the ReLU activation, and then by weight wt_2.
4. **Residual connection:** W(a) = G(a) + a (where G(a) = 0)
   This means that if G(a) is zero, the output simplifies to: W(a) = 0.
5. **Processing with Residual Function:**
   a → wt_1 ReLU → wt_2 ReLU… → aW(a)
   If G(a) is zero:
   W(a) = G(a) + a (with G(a) = 0)
   Thus:
   W(a) = a

This neural network algorithm processes input data through a series of transformations and residual connections. It starts by defining the output of a residual block, $W(a)$ as the sum of the input feature map a and the output of a residual function $G(a)$, i.e., $W(a) = G(a) + a$. The goal is to achieve an ideal predicted output $w(a)$. The data flows through multiple layers: first, it is transformed by weight matrix wt_1, then activated by ReLU, followed by another transformation using wt_2, resulting in $a \rightarrow wt_1 \rightarrow \text{ReLU} \rightarrow wt_2 \rightarrow W(a)$. If the residual function's output $G(a)$ is zero, the equation simplifies to $W(a) = G(a) + a = 0 + a = a$, indicating no net change from the original input. This setup allows for efficient information flow even in deeper networks where some transformations may not significantly alter the input. Thus, graphical representation of Algorithm II is depicted in Fig. 6.

Fig. 6 shows the pictorial illustration of Algorithm I, where the block labelled Input denotes the image (a), the input is passed to process block, where the process block is denoted by the function $G(a)$ which performs processing on the input. Combine block represents the step where $G(a)$ and a are combined for creating refined image of $W(a)$ Moreover, Bypass path permits the initial processing $G(a)$ to reach the final stage directly and eventually output block

delivers the final segmented image by combining the desired outcome from both processing paths and bypass connection.

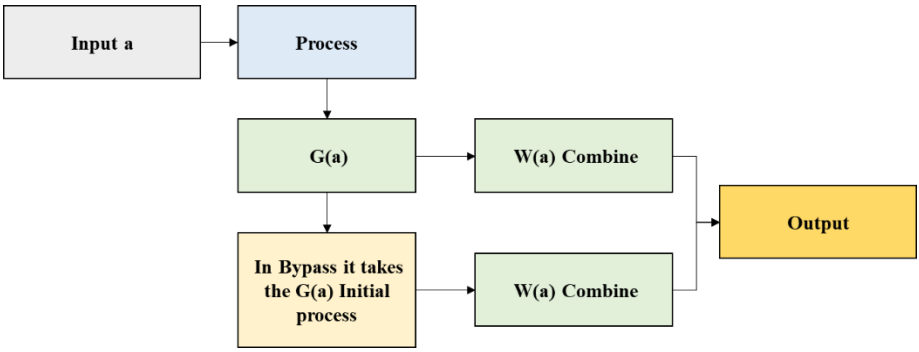


Fig. 6 – *Vanishing Gradient Computation.*

3.3 Attention mechanism

Attention mechanism is predominantly used for permitting the features in the model and focusing on the most relevant parts of input, which aids in improving the robustness of the model. In addition, it also aids in reducing the usage of memory and computational cost by detecting and processing only vital parts of inputs. The proposed model implemented attention mechanism in order to increase the weight of effective features and attention mechanism aids in the process of enhancing the accuracy of segmentation.

In the proposed model, attention mechanism is classified into 2 branches, which includes mask branch and trunk branch. Trunk branch performs feature processing and mask branch aids in preventing wrong gradients with the aim to update the parameters of trunk. The attention mechanism employed in the proposed model comprises of p, t and r, where

$$\text{Specified:} \quad p = 1, \quad t = 2, \quad r = 1. \tag{5}$$

Here p represents the number of pre-processing residual unit before splitting into mask and trunk branches, t is the number of RU in the trunk branch and r denotes the number of Residual Unit which lie between the pooling layer in the mask branch.

The Residual Unit is passed as input to attention mechanism layer and then enhanced images are fetched as output. The trunk branch output is out(a) with the input a, then the mask branch employs top down (up-sampling) and bottom up (down-sampling) in order to learn the size of the mask which softly weight output features out(x). Top-down involves transposed convolutions (deconvolutions) that restore the spatial dimensions of the feature maps, allowing for detailed output images and bottom up approach facilitates the extraction of critical

features and patterns from the data, enabling the model to discern the underlying semantics present within the teeth image. Additionally, these approaches aid in mimicking the fast FF and feedback attention process. Eventually, the output of the attention module AMH is defined as represented in (6)

$$\mathrm{AMH}_{s,ch}(a) = \mathrm{M}_{s,ch}(a) \otimes \mathrm{out}_{i,c}(a) \,, \tag{6}$$

in which *i* ranges over the spatial locations and ch is the channel index from 1 to Ch, pictorial representation is projected in Fig. 8. Where, the process starts with an input *a*, which is passed through a transformation layer that generates $G(a) + a$ harnessing the residual connection to preserve original information while incorporating learned transformations. This output, referred to as the residual unit output, is subsequently fed into a mask branch. Within the mask branch, a combination of top-down and bottom-up processing is employed to effectively integrate multi-scale contextual information. The refined output from this stage is then passed through an attention module, which selectively emphasizes salient features and suppresses irrelevant ones, ultimately producing enhanced feature representations for downstream tasks.

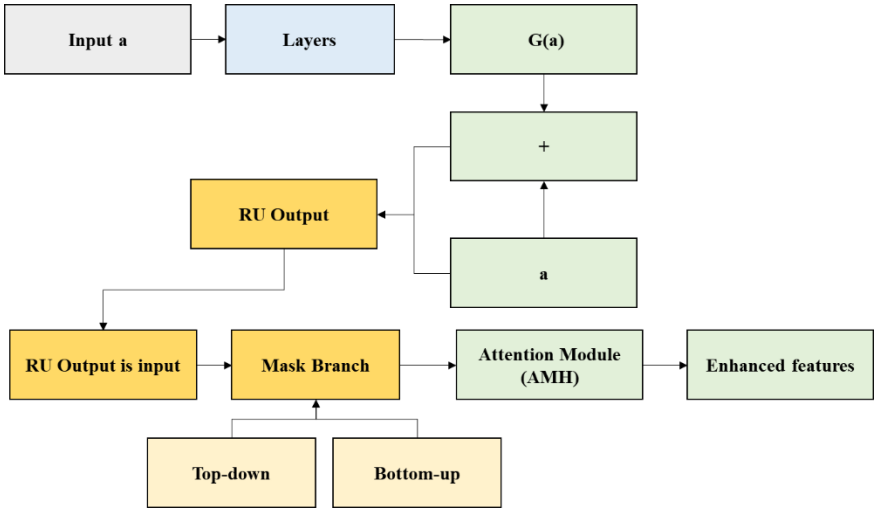


Fig. 7 – *Attention mechanism model.*

Fig. 7 depicts the process involved in attention mechanism, where the output from residual units (enhanced features) is sent as input *a* to attention mechanism layer and this attention mechanism layer focuses on significant parts of the images by assigning weights to different features. Thus, 2 substantial components used in attention mechanism is mask branch and attention module, where mask branch uses top down and bottom up information. Top down information refers to features learned in higher layers of the network and bottom up information

denotes low level features from input layers. Further, attention module aids in calculating attention weights depending on the mask branch output and refines the features from RU.

Further, in attention mechanism, the attention mask does not function as feature selector during the process of forward inference, but it functions as GUF (Gradient Update Filter) during the processes of back propagation. It plays a crucial role during training (back propagation) by affecting how gradients are updated. This highlights the dual functionality of the attention mask: it helps in processing information during inference and influences learning during training. The attention mask is obtained as follows

$$\frac{\mathrm{dM}\left(\mathrm{a},\mathrm{Grountruth}\right)\mathrm{T}\left(a,\varnothing\right)}{\mathrm{d}\varnothing}=\mathrm{M}\left(a,\mathrm{groundtruth}\right)\frac{\mathrm{dT}\left(x,\varnothing\right)}{\mathrm{d}\varnothing}, \tag{7}$$

in which (7), $\varnothing$ represents trunk branch parameter and Groundtruth denotes mask branch parameter. To these variables, attention mechanism is considered to be more robust and efficient to noisy label and aids in delivering effective model for obtaining enhanced segmented images. Subsequent section deals with results obtained using proposed framework.

4 Results and Discussion

The proposed design has been executed with Python. The first subsection discusses description of the dataset. The second subsection describes the performance metrics. The third subsection presents experimental outcome of the proposed model and the fourth subsection describes the performance analysis of the proposed model. Finally, the Fifth subsection presents comparative analysis to determine the efficacy of the proposed approach over conventional methods.

4.1 Dataset description

Tuft dataset consists of 1000 chosen radiographs which were detected and saved in a generic image format by using a unique identifier. These radiographs were set to be interpreted by a student and an expert from Tufts university school of DM. the dataset comprises of six different components, which includes radiographs, eye tracker generated maps, maxilla-mandibular ROI mask, textual information which describes each and every radiograph, labelled masks, including individual tooth masks, are provided for each radiograph along with their corresponding labels.

Link: http://tdd.ece.tufts.edu/

4.2 Performance metrics

Performance of the model is evaluted in the subsequent section with accordance to various performance metrics such as accuracy, precision, F-measure, recall, IoU and Dice-coefficient.

A Accuracy

Accuracy is the measure of number of total accurate classification and accuracy range is calculated with the following (8)

$$Acc = \frac{\text{TRN} + \text{TRP}}{\text{TRN} + \text{FLN} + \text{TRP} + \text{FLP}}, \tag{8}$$

where TRN represents the number of True negatives, TRP are True positives, FLN are false negatives, and FLP are false positives.

B Precision

Precision is calculated by measuring the precise classification count. It is measured over the total positive count. The precision is calculated with the following (9)

$$\text{precision} = \frac{\text{TRP}}{\text{FLP} + \text{TRP}}. \tag{9}$$

C Recall

Recall calculates the sum of precise positive types conceived of all the optimistic groups and it is evaluated by the ensuing (10),

$$R_c = \frac{\text{TRP}}{\text{FLN} + \text{TRP}} \tag{10}$$

In (11) FLN denotes the number of false negatives.

D F-measure

Another type of F-measure is known as the F1 score. The F1 score represents the weighted harmonic-mean value of recall and precision, the F1 score is estimated with the following (11),

$$\text{F1-score} = 2 \times \frac{Rc \times Pc}{Rc + Pc}, \tag{11}$$

where, P denotes precision and R represents recall.

E Dice-coefficient

The Dice Coefficient DSC deals with the spatial overlap between 2 segmentations, A and B target regions, and is defined in (12),

$$\text{DSC}(\text{A}, \text{B}) = \frac{2(\text{A} \cap \text{B})}{(\text{A} + \text{B})}, \tag{12}$$

where A refers to ground truth Segmentation and B represents the predicted segmentation.

F IoU Value

IoU value measures how well the predicted region overlaps with the ground truth region, with a value ranging from 0 (no overlap) to 1 (perfect overlap),

$$\text{IoU} = \frac{\text{Area of Union}}{\text{Area of Overlap}}. \tag{13}$$

4.3 Experimental results

Experimental outcome of Input images and the respective segmented images are depicted in this subsequent section. **Table 2** shows the radiography images and the segmented images of the teeth in tabulated form.

Table 2 shows the different input images, which are known as radiography images and their corresponding segmented images which are known as masking images. From the table, it can be identified that, proposed model aided in delivering precise and clear images **required** for segmentation. Though the proposed model has produced accurate teeth segmented images, proposed model has also generated misclassified images which are illustrated in **Table 3.**

Table 2
Input Images and Segmented Images.

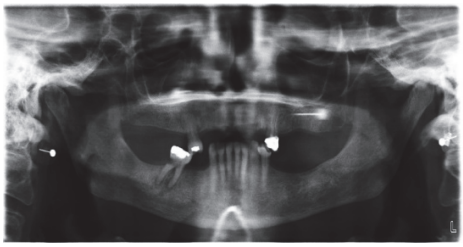
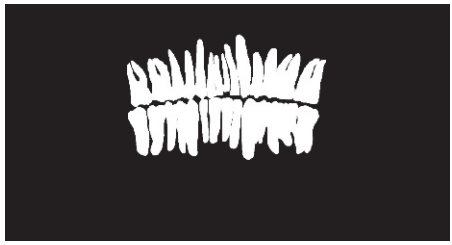
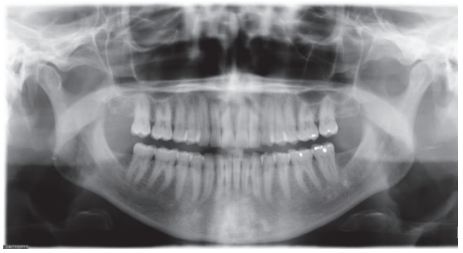
Radiography images	***Mask images***
	
	
	

Table 2–continued

Input Images and Segmented Images.

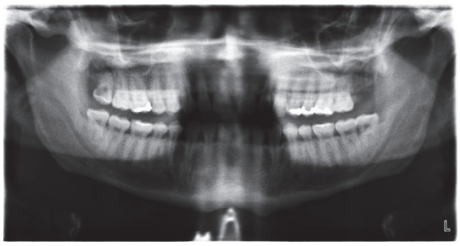
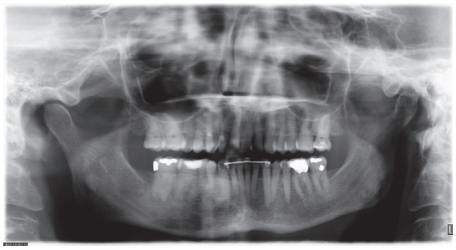
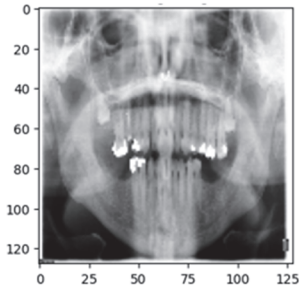
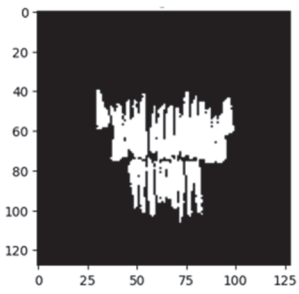
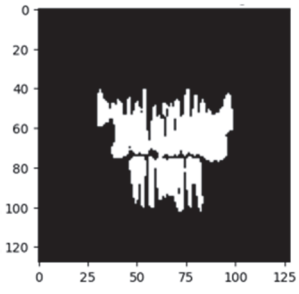
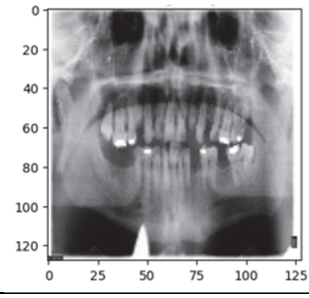
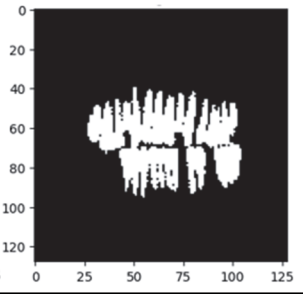
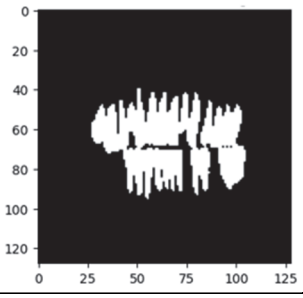
Radiography images	***Mask images***
	
	

Table 3

Under-segmented Images

Testing Image	Testing label	Prediction on Test Image
		
		

Despite proposed model has delivered better Dice coefficient and Mean IoU such as 0.93 and 0.94, Less Mean IoU of 0.9223079 and Dice coefficient of 0.93124 is also attained by proposed model and it is depicted in **Table 3** (2nd row). Likewise, Mean IoU of 0.9189391 and Dice coefficient of 0.920541 is attained by proposed DC with RES model, **Table 3** (3rd row). Thus, the present section shows diverse IoU and Dice Coefficient attained by proposed work for different Teeth images.

4.4 Performance analysis

Performance of the proposed model is analyzed in the performance analysis using various metrics, which includes Mean IoU, Mean Dice co-efficient, accuracy, F1-score, Recall and precision. **Table 4** shows the values attained by the proposed model using these metrics.

Table 4

Performance Analysis.

Tuft Dental Dataset	
Mean IoU	0.9374645
Mean Dice Coefficient	0.943075021
Accuracy	0.9869
F1-score	0.98
Recall	0.9793
Precision	0.98

IoU value obtained by the proposed model using Tuft dental dataset is 0.9374, similarly, dice co-efficient obtained by the proposed model is 0.943075, accuracy obtained is 0.9869, F1-score is 0.98, Recall attained is 0.9793 and the precision value attained by the proposed model is 0.98. Fig. 8 depicts the graphical representation of the **Table 4**. In which, values attained by employing different evaluation metrics are illustrated.

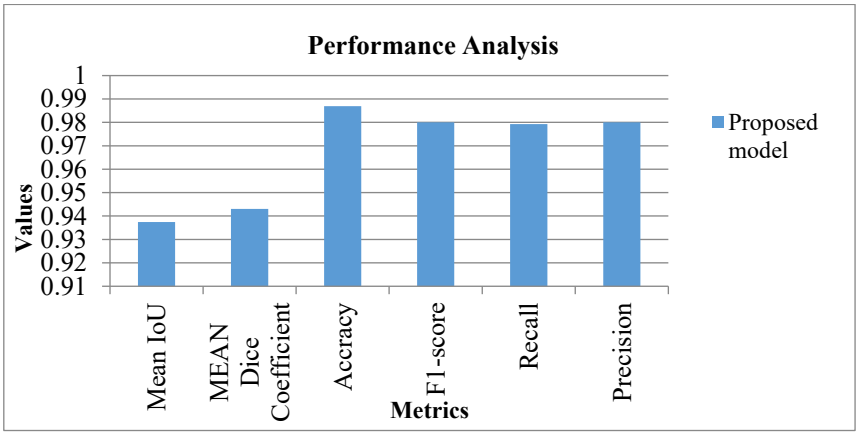


Fig. 8 – *Performance analysis of the proposed and existing model.*

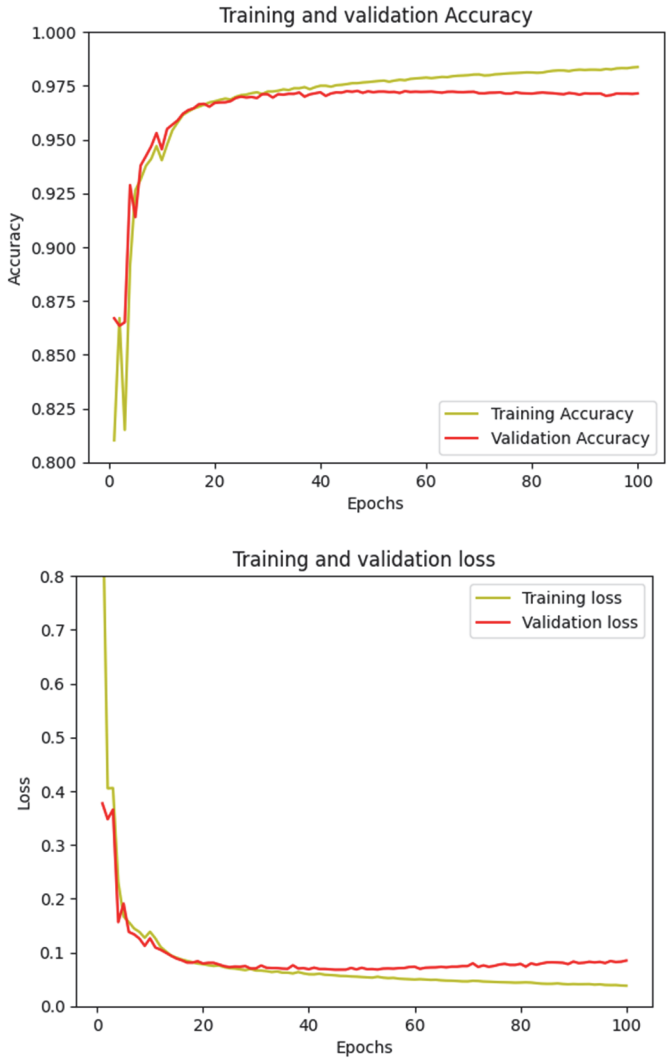


Fig. 9 – *Validation Accuracy and loss of the proposed model.*

Fig. 9 shows the training and validation accuracy and training and validation loss of the proposed model. In the graphs the green line represents training and orange line represents validation.

4.5 Comparative analysis

Comparative analysis aids in showcasing the efficiency of the proposed model by distinguishing the Mean IoU and mean dice co-efficient values of the existing and the proposed model. **Table 5** shows the mean IoU value and mean

dice co-efficient value obtained by both existing Multi-Headed CNN, existing context encoder-net and proposed model.

Table5

Comparative Analysis of the Existing and Proposed Model [10].

Model	Mean IoU	Mean Dice coefficient
Multi-Headed CNN (Existing model) [10]	0.918	0.918
Context Encoder-Net (Existing Model) [3]	0.8164	0.8662
Proposed Model	0.937465	0.943075

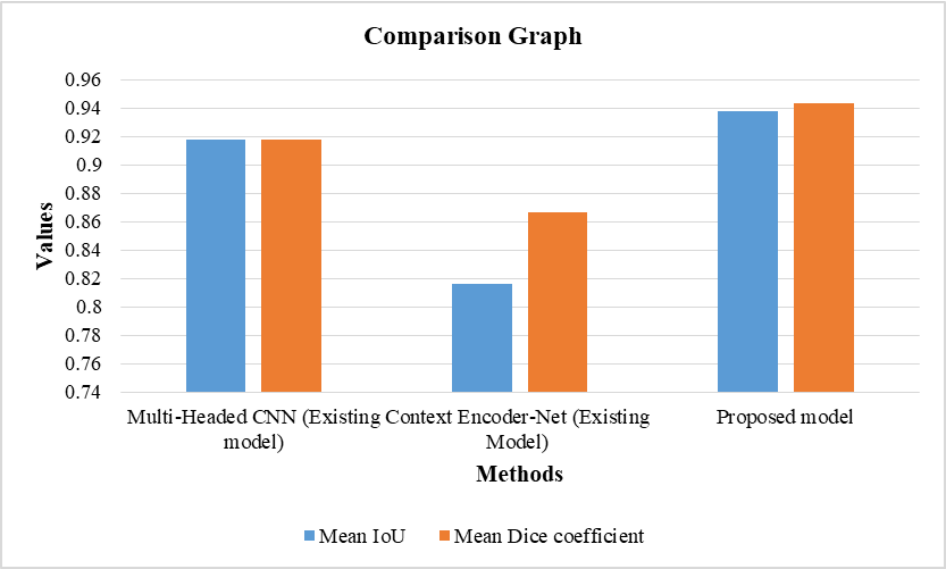


Fig. 10 – *Comparison Graph.*

Fig. 10 shows the graphical representation of the IoU value and Mean Dice co-efficient of the proposed model along with the existing Multi-Headed CNN model and context encoder-net model. Multi-headed CNN is opted for comparison because model used Uses standard architectures like ResNet with fixed resolution, leading to less efficient feature recovery, struggle with large datasets and may not fully leverage data complexity and Provides baseline results with standard accuracy, IoU, and Dice coefficient, but may lack precision. These limitations are overcome by using proposed method. The existing Multi-Headed CNN model achieved an IoU and Dice coefficient of 0.918, while the Context Encoder-Net recorded values of 0.8164 for IoU and 0.8662 for the Dice coefficient. In contrast, the proposed model demonstrated superior performance

with an IoU of 0.937464 and a Dice coefficient of 0.943075. The proposed model enhances tooth segmentation performance by utilizing a Dual Convolution approach combined with RES. This strategy increases the spatial resolution of the feature maps, allowing for better recovery of lost spatial information during the downsampling process. As a result, this improvement is reflected in the model's higher IoU and Dice Coefficient values indicating more accurate segmentation outcomes.

From the experimental result it can be identified that, the proposed model aids in delivering better outcome for teeth segmentation due to the implementation of DC along with RES. The proposed DC with RES model aids in producing an enhanced segmented image to the corresponding radiographs as the DC employs transposed convolutions in the UNet architecture which aids in delivering better outcome than the existing ones. Further, incorporation of RES model efficiently propagates both high-level and low-level features to the deeper layers. This integration helps in preserving fine-grained details and improving the network's ability to accurately segment dental structures, such as teeth or dental caries. Due to these reasons, the proposed model delivers better outcome for image segmentation than the existing models.

6 Conclusion

Usage of AI in dental field provides various beneficial outcomes due to the plenty imagery and non-imagery based data. Similarly, the dental radiographies deliver vital information for the treatment and diagnosis of the disease. However, the existing methods are not effective enough to accomplish a robust and effective model for teeth segmentation. Hence, the proposed model employed DC with RES model for teeth segmentation. DC, also known as the transpose convolution or Upsampling layer, was responsible for increasing the spatial resolution of the feature maps. It helped to recover lost spatial information during the down sampling process. Similarly, RES model aided in proficiently proliferated both low-level and high level features to the deep layers. Further, the proposed model was assessed using various evaluation metrics such as IoU and Dice co-efficient, accuracy, recall, precision, F1 score. The proposed model achieved an IoU value of 0.9375, a Dice coefficient of 0.9431, an accuracy of 0.9869, an F1 score of 0.98, a recall rate of 0.9793, and a precision of 0.98. In future, various DL based algorithms can be incorporated for enhancing the radiographs which aids the dental professionals for providing suitable treatments.

7 References

[1] Diagnostics: A Systematic Comparison of Techniques for Accurate Prediction of Dental Disease Through X-Ray Imaging, International Journal of Intelligent Computing and Cybernetics, Vol. 17, No. 1, February 2024, pp. 161 – 180.

[2] K. Orhan, G. Ünsal: Artificial Intelligence in Dentistry, Ch. 18, Digital Dentistry: An Overview and Future Prospects, Springer, Cham, 2024.

[3] K. Panetta, R. Rajendran, A. Ramesh, S. Rao, S. Agaian: Tufts Dental Database: A Multimodal Panoramic X-Ray Dataset for Benchmarking Diagnostic Systems, IEEE Journal of Biomedical and Health Informatics, Vol. 26, No. 4, April 2022, pp. 1650 – 1659.

[4] M. Xu, Y. Wu, Z. Xu, P. Ding, H. Bai, X. Deng: Robust Automated Teeth Identification from Dental Radiographs Using Deep Learning, Journal of Dentistry, Vol. 136, September 2023, p. 104607.

[5] Y. Zhao, P. Li, C. Gao, Y. Liu, Q. Chen, F. Yang, D. Meng: TSASNet: Tooth Segmentation on Dental Panoramic X-Ray Images by Two-Stage Attention Segmentation Network, Knowledge-Based Systems, Vol. 206, October 2020, p. 106338.

[6] M. Jafari, D. Auer, S. Francis, J. Garibaldi, X. Chen: DRU-Net: An Efficient Deep Convolutional Neural Network for Medical Image Segmentation, arXiv:2004.13453v1 [eess.IV], April 2020, pp. 1 – 5.

[7] S. Hou, T. Zhou, Y. Liu, P. Dang, H. Lu, H. Shi: Teeth U-Net: A Segmentation Model of Dental Panoramic X-Ray Images for Context Semantics and Contrast Enhancement, Computers in Biology and Medicine, Vol. 152, January 2023, p. 106296.

[8] D. V. Tuzoff, L. N. Tuzova, M. M. Bornstein, A. S. Krasnov, M. A. Kharchenko, S. I. Nikolenko, M. M. Sveshnikov, G. B. Bednenko: Tooth Detection and Numbering in Panoramic Radiographs Using Convolutional Neural Networks, Dentomaxillofacial Radiology, Vol. 48, No. 4, May 2019, p. 20180051.

[9] A. S. AL-Malaise AL-Ghamdi, M. Ragab, S. A. AlGhamdi, A. H. Asseri, R. F. Mansour, D. Koundal: Detection of Dental Diseases through X-Ray Images Using Neural Search Architecture Network, Computational Intelligence and Neuroscience, Vol. 2022, No. 1, January 2022, p. 3500552.

[10] L. Nashold, P. Pandya, T. Lin: Multi-Objective Processing of Dental Panoramic Radiographs, Available at: http://cs231n.stanford.edu/reports/2022/pdfs/118.pdf

[11] S. Tian, N. Dai, B. Zhang, F. Yuan, Q. Yu, X. Cheng: Automatic Classification and Segmentation of Teeth on 3D Dental Model Using Hierarchical Deep Learning Networks, IEEE Access, Vol. 7, June 2019, pp. 84817 – 84828.

[12] A. Hossam, K. Mohamed, R. Tarek, A. Elsayed, H. Mostafa, S. Selim: Automated Dental Diagnosis Using Deep Learning, Proceedings of the 16th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, December 2021, pp. 1 – 5.

[13] S. Vinayahalingam, S. Kempers, L. Limon, D. Deibel, T. Maal, M. Hanisch, S. Bergé, T. Xi: Classification of Caries in Third Molars on Panoramic Radiographs Using Deep Learning, Scientific Reports, Vol. 11, June 2021, p. 12609.

[14] M. P. Muresan, A. R. Barbura, S. Nedevschi: Teeth Detection and Dental Problem Classification in Panoramic X-Ray Images using Deep Learning and Image Processing Techniques, Proceedings of the IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, September 2020, pp. 457 – 463.

[15] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, L. Oliveira: Deep Instance Segmentation of Teeth in Panoramic X-Ray Images, Proceedings of the 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Parana, Brazil, October 2018, pp. 400 – 407.

[16] J. Kim, H.- S. Lee, I.- S. Song, K.- H. Jung: DeNTNet: Deep Neural Transfer Network for the Detection of Periodontal Bone Loss Using Panoramic Dental Radiographs, Scientific Reports, Vol. 9, November 2019, p. 17615.

[17] M. Fukuda, K. Inamoto, N. Shibata, Y. Ariji, Y. Yanashita, S. Kutsuna, K. Nakata, A. Katsumata, H. Fujita, E. Ariji: Evaluation of an Artificial Intelligence System for Detecting Vertical Root Fracture on Panoramic Radiography, Oral Radiology, Vol. 36, No. 4, October 2020, pp. 337 – 343.

[18] V. Geetha, K. S. Aprameya, D. M. Hinduja: Dental Caries Diagnosis in Digital Radiographs Using Back-Propagation Neural Network, Health Information Science and Systems, Vol. 8, No. 1, December 2020, p. 8.

[19] I. S. Bayrakdar, K. Orhan, S. Akarsu, Ö. Çelik, S. Atasoy, A. Pekince, Y. Yasa, E. Bilgir, H. Sağlam, A. F. Aslan, A. Odabaş: Deep-Learning Approach for Caries Detection and Segmentation on Dental Bitewing Radiographs, Oral Radiology, Vol. 38, No. 4, October 2022, pp. 468 – 479.

[20] Y. Nishitani, R. Nakayama, D. Hayashi, A. Hizukuri, K. Murata: Segmentation of Teeth in Panoramic Dental X-Ray Images Using U-Net with a Loss Function Weighted on the Tooth Edge, Radiological Physics and Technology, Vol. 14, No. 1, March 2021, pp. 64 – 69.

[21] M. Al-Sarem, M. Al-Asali, A. Y. Alqutaibi, F. Saeed: Enhanced Tooth Region Detection Using Pretrained Deep Learning Models, International Journal of Environmental Research and Public Health, Vol. 19, No. 22, November 2022, p. 15414.

[22] C. Muramatsu, T. Morishita, R. Takahashi, T. Hayashi, W. Nishiyama, Y. Ariji, X. Zhou, T. Hara, A. Katsumata, E. Ariji, H. Fujita: Tooth Detection and Classification on Panoramic Radiographs for Automatic Dental Chart Filing: Improved Classification by Multi-Sized Input Data, Oral Radiology, Vol. 37, No. 1, January 2021, pp. 13 – 19.

[23] J.- H. Lee, S.- S. Han, Y. H. Kim, C. Lee, I. Kim: Application of a Fully Deep Convolutional Neural Network to the Automation of Tooth Segmentation on Panoramic Radiographs, Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, Vol. 129, No. 6, June 2020, pp. 635 – 642.