Optimizing Dental Care: Segmentation and CNN Models for Accurate Panoramic X-ray Diagnostic Classification

Laith Adi | 20711298
Keying Xu | 20783941
Yichen Jiang | 21103079
Raymond Chen | 20790117

University of Waterloo
Winter 2024
STAT940 - Deep Learning

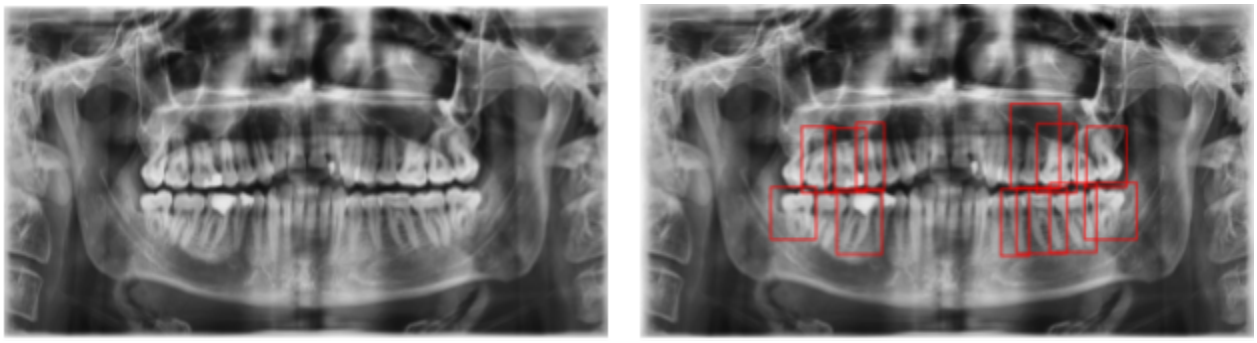
1. Introduction

The project aims to utilize dental panoramic X-rays for diagnostic classification. The data consists of labeled panoramic X-rays, and was introduced as part of the Dental Enumeration and Diagnosis on Panoramic X-rays Challenge (DENTEX). The primary objective is to develop a segmentation model capable of accurately depicting each tooth within the panoramic X-ray images. These segmented tooth images will then be extracted and utilized as inputs for a Convolutional Neural Network (CNN). CNN's role is to predict the diagnostic class for each tooth, thereby facilitating an efficient and precise dental diagnosis.

2. Dataset

The Dental Enumeration and Diagnosis on Panoramic X-rays Challenge (DENTEX) [1] held in conjunction with MICCAI in 2023, sought to propel advancements by facilitating the development of algorithms geared towards accurately identifying abnormal teeth and associated diagnoses on dental radiography. The challenge's dataset, structured using the FDI system, is hierarchically annotated, offering three distinct types of annotated data to participants.

1. Partially labeled dataset with quadrant information: This dataset consists of 693 X-rays where each quadrant is numerically designated.
2. Partially labeled dataset with quadrant and tooth enumeration information: This dataset consists of 634 X-rays where each tooth within a quadrant is numbered according to the FDI numbering system.
3. Fully labeled dataset: This dataset consists of 1005 X-rays featuring quadrant, tooth enumeration, and diagnosis information.



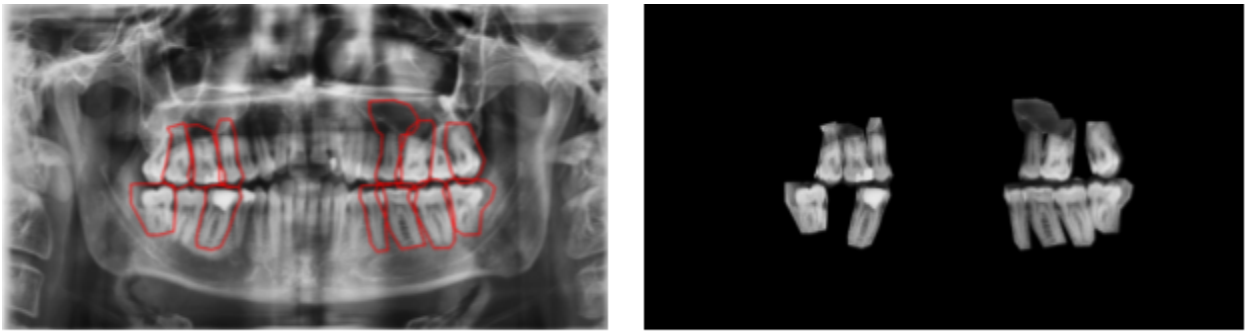
(a) (b)

Figure 1

2.1 Data Analysis

In this section, we will observe a panoramic X-ray, along with its bounding boxes and segmentation. Figure 1 (a), provides a clear view of the unaltered panoramic X-ray, serving as our baseline reference. Transitioning to Figure 1 (b), we augment the panoramic X-ray from Figure 1 (a) with bounding boxes, each depicting a tooth alongside its associated diagnostic label. These bounding boxes offer initial insights into the spatial distribution of abnormalities within the dental panorama. Moving forward to Figure 2 (a), we introduce the segmentation coordinates overlaid onto Figure 1 (a). These coordinates, resembling polygons, help facilitate the extraction of cropped tooth images for further analysis. Finally, in Figure 2 (b), we see the

segmentation mask derived from Figure 1 (a). This mask precisely outlines the regions of interest—namely, the individual teeth—crucial for our diagnostic classification objectives.



(a) (b)

Figure 2

3. Methodology

When examining panoramic X-rays, imagine looking at someone directly from the front, capturing the entirety of their mouth in a single image. The initial step in our process involves precisely and efficiently isolating each tooth as a separate image. This critical task is accomplished through the utilization of segmentation models. Once every individual tooth is extracted as a separate image, we can proceed to use them alongside their corresponding diagnosis labels as inputs for training and fine-tuning our Convolutional Neural Network (CNN) model. Training a CNN using cropped segmented images rather than the whole panoramic X-ray enhances model performance by focusing exclusively on the relevant tooth regions, thereby reducing noise and irrelevant information, leading to more precise and efficient diagnostic classification.

3.1 Teeth Segmentation - Pre-processing

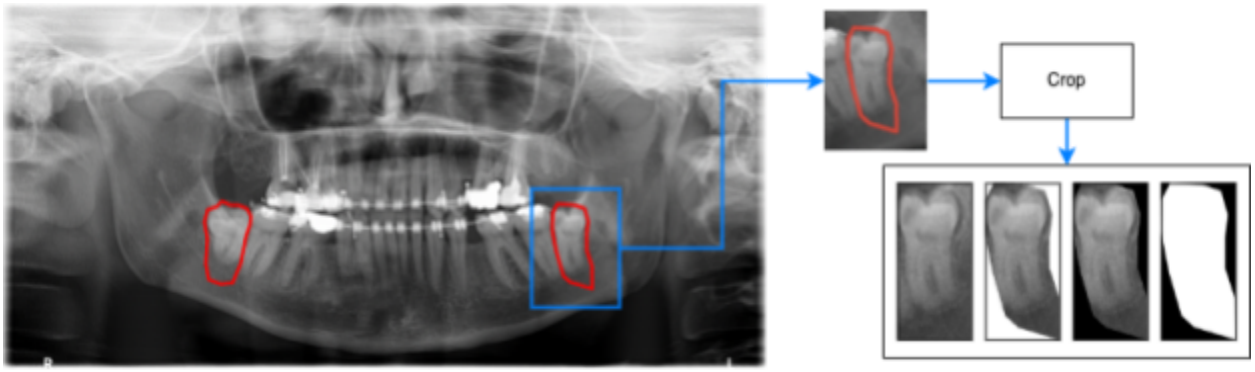


Figure 3: Overview of Teeth Segmentation

Panoramic X-rays are wide images that are about 3000 pixels by 1500 pixels in size, which makes it difficult to conduct training in large batches, as these images do not fit into computer memory. To ensure these images can fit into memory, image cropping was performed to extract the abnormal teeth and their corresponding labels from the dataset. Using the

segmentation masks provided by the dataset, abnormal teeth are extracted using the polygon bounding boxes defined in the annotation file. As shown by figure 3, each abnormal tooth in the panoramic X-rays generates three cropped images and a segmentation mask. The difference in the three cropped images is the color of the background (pixels outside of the bounding box).

Upon completion of cropping all panoramic X-rays, 4 datasets with cropped images were created. The first, second, and third datasets have backgrounds that are original, black, and white, respectively. The last dataset contains masks of the cropped teeth. The labels of each cropped tooth were extracted into a csv file with columns “Image ID”, “Image Label”, and “Image Source”. The cropped datasets and labels are then used to train CNNs for classification.

3.2 Teeth Segmentation - Modeling

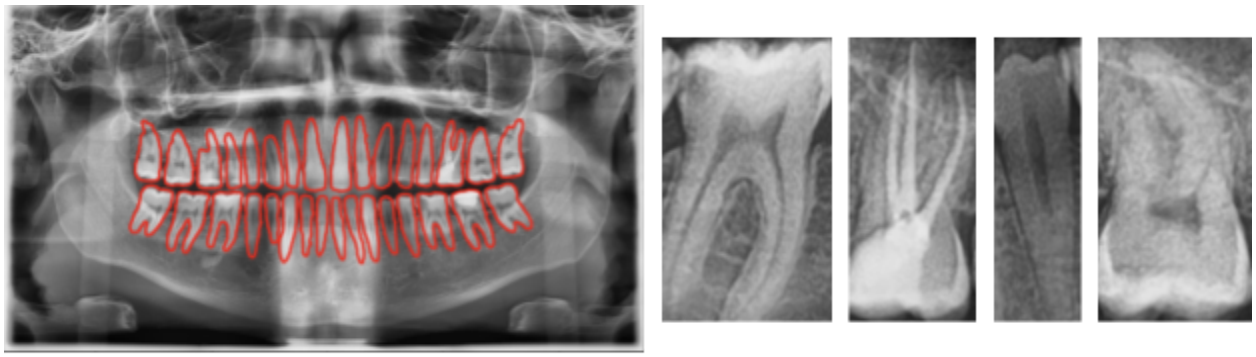
The datasets provided contained segmentation information of abnormal teeth and as a result, we were able to crop out each abnormal tooth from the images. However, segmentation information may not always be available. Therefore, we explored segmentation models that output a mask of teeth for panoramic X-rays. Using the masks, we would then be able to crop out individual teeth. These cropped images were used strictly for analytic purposes and not for training CNNs, because the performance of the CNNs is heavily bottlenecked by the performance of the segmentation model. The objective is to determine the feasibility of removing the need for segmentation information in the annotation file provided in the dataset.

Our first approach was to train a segmentation model from scratch, but quickly, we came to the realization that the dataset was not suited for doing so. More specifically, the segmentation information creates masks of abnormal teeth only. As a result, the model learns contradicting patterns due to the inconsistency of the masks. For instance, consider the molars. In a training instance where the molars are masked, the model learns that the molars are foreground objects. On the other hand, the model learns that the molars are background objects in a training instance where the molars are not masked. This inconsistency confuses the model, causing it to be uncertain about whether it should produce masks for the molars. Consequently, the model's performance was unsatisfactory. Additionally, because of the sensitivity of panoramic X-rays, it was challenging to find a suitable dataset to train our own segmentation model.

Opening Erosion

Figure 4: Opening is used to remove noise and Erosion is used to erode away the boundaries of foreground object.

Despite the difficulties of finding a good dataset, we came across a publicly available pre-trained teeth segmentation model [4] with its training data unavailable. The model outputs a mask for all teeth, which then can be used to crop each individual tooth. However, the masks were not optimal as the model wasn't trained on our dataset. Thus, standard computer vision techniques [5] like those shown in Figure 4 are applied to the masks for better alignment with teeth. The contours (i.e. bounding box) of the masked teeth are then constructed and used in the cropping algorithm.



Polygon bounding boxes Cropped images

Figure 5: Example of a panoramic X-ray that the pre-trained model generated good results on.

The performance of the model wasn't ideal, performing reasonably in some cases and poorly in others. In Figure 5, the model was able to segment teeth very well, allowing the cropping algorithm to generate cropped images similar to those in Figure 3. On the other hand, Figure 6 shows that the model sometimes can segment multiple teeth together as one. As a result, the cropped image would contain multiple teeth as well. Nevertheless, a segmentation model can replace the segmentation annotations provided in the original dataset, because results demonstrate that segmentation models can correctly mask teeth from panoramic X-rays.

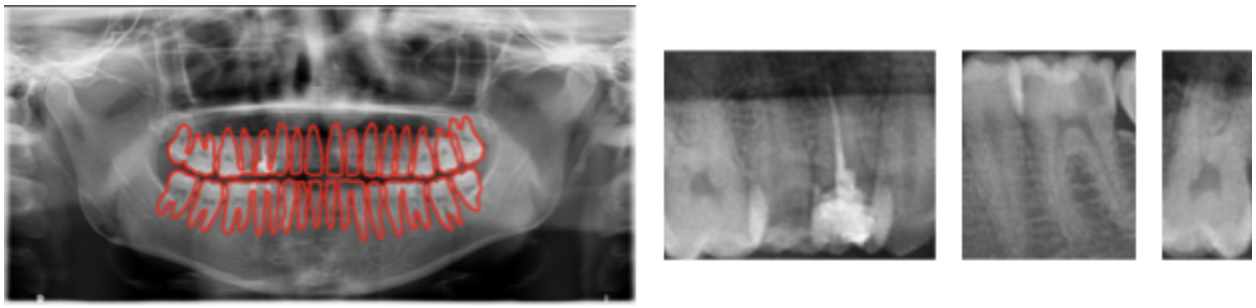


Figure 6: Example of a panoramic X-ray that the pre-trained model generated poor results on.

3.3 Prediction of Diagnostic Classes Using Convolutional Neural Network

There are four possible diagnostic classes, which are carries, deep carries, periapical lesions, and impacted teeth (Figure 7). Using segmented images, we applied two CNN structures to predict the diagnostic class for a given tooth. The two CNN structures are Residual Neural

Network (ResNet), and Visual Geometry Group (VGG)[2]. For both methods, we used cropped images as shown below. Data was divided into a training set (90%) and a validation set (10%).

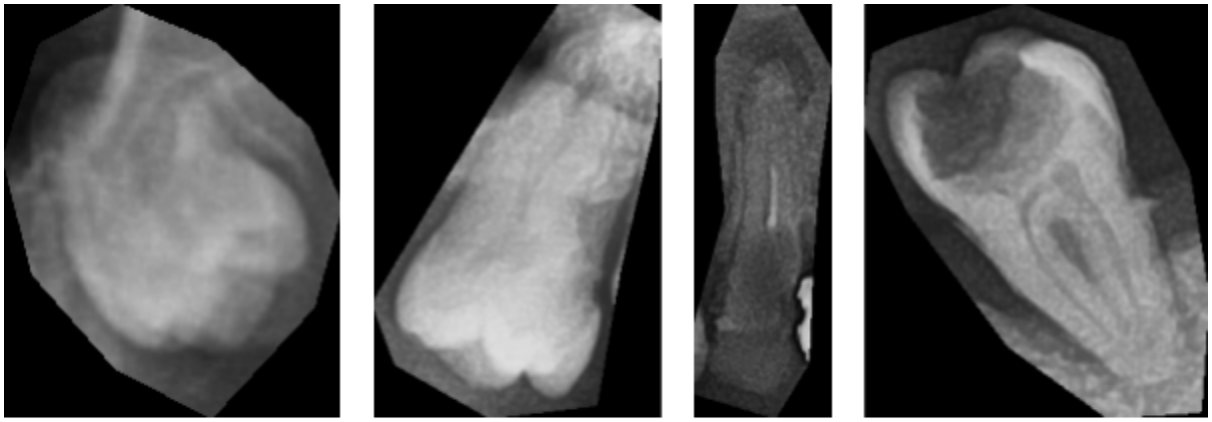


Figure 7: Four diagnostic classes of teeth

3.3.1 Residual Neural Network (ResNet) with 50 layers

ResNet50 [3] is an innovative neural network for image classification. ResNet50 contains skip connections within its architecture. Skip connections help to mitigate the vanishing gradient problem during training. It does this by skipping over some layers when performing forward and backward propagation. Here, we adopted a ResNet-50, which is trained on more than a million images from the ImageNet database.

As shown in Figure 7, cropped images have different sizes. We unify sizes of all images to 224 by 224 pixels. The batch size was set to be 32. Below is a list of additional layers that are added to the ResNet50 model.

(1) Global Average Pooling 2D layer;
(2) 512 Dense ReLU layer;
(3) Dropout layer with 0.5 dropout probability;
(4) 256 Dense ReLU layer;
(5) Dropout layer with 0.5 dropout probability;
(6) 128 Dense ReLU layer;
(7) Dropout layer with 0.5 dropout probability;
(8) 64 Dense ReLU layers;
(9) Dropout layer with 0.5 dropout probability;
(10) 5 Dense softmax layer.

Figure 8 shows the loss and accuracy calculated on the training and validation sets, through 10 epochs. At epoch 10, the validation loss was reduced from 1.1229 to 0.7098, while the validation accuracy was improved from 0.6554 to 0.7542.

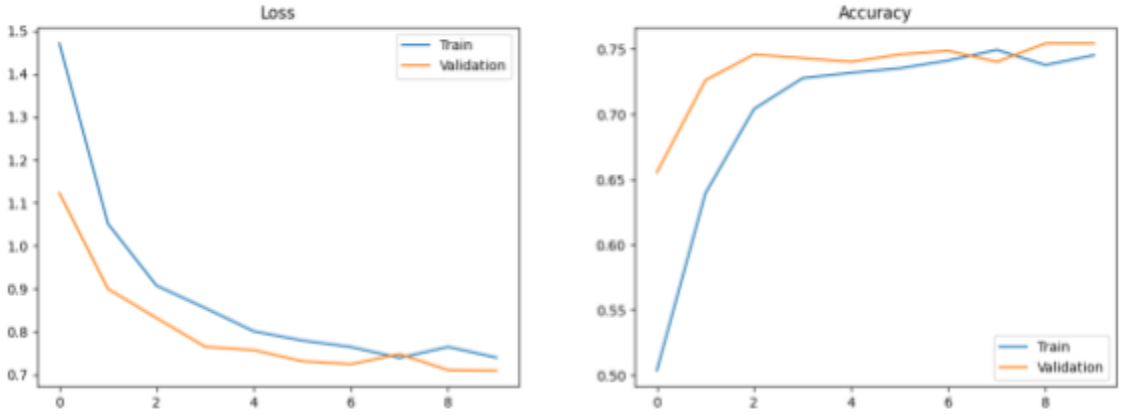


Figure 8: (left) Training and validation loss for ResNet-50. (right) Training and validation accuracy for ResNet-50.

3.3.2 Visual Geometry Group model (VGG) with 16 layers

VGG16[2] is a deep CNN developed by Visual Geometry Group at University of Oxford. The VGG model achieves high performance with small convolution filters and 16-19 weight layers. In the original paper of the VGG model, there are 6 different configurations proposed, and in our project, we trained and tested on VGG16.

Here, we adopted the pretrained VGG16 model from pytorch, using the batch size of 32 and 10 epochs to fine-tune the model. Using the same set of cropped images as shown in section 3.3.1, the training loss is reduced from 0.9023 to 0.3660, and the validation loss is reduced from 0.7378 to 0.5597. The validation accuracy is increased from 72.24% to 76.49%, with a peak accuracy of 77.05% at the seventh epoch, as shown in Figure 9.

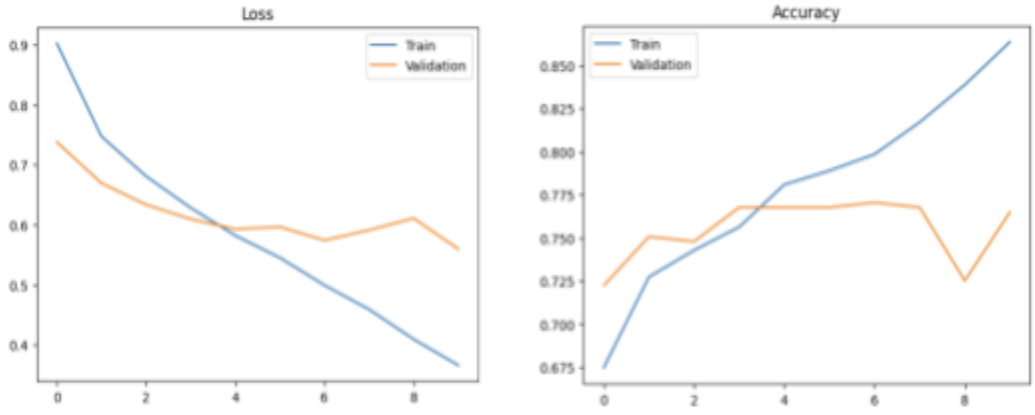


Figure 9: (left) Training and validation loss for VGG16. (right) Training and validation accuracy for VGG16.

4. Conclusion

In conclusion, we have leveraged segmentation and CNN models to extract individual teeth from panoramic X-rays to conduct diagnostic classification. Utilizing the DENTEX fully annotated dataset, the project developed methodologies for teeth segmentation and diagnostic class prediction. Despite facing challenges from the dataset's novelty and complexities, promising results were achieved through the application of ResNet50 and VGG16 CNN architectures. The ResNet50 model exhibited an improvement in validation accuracy from 65.54% to 75.42% after 10 epochs, while the VGG16 model showcased a peak accuracy of 77.05% at the seventh epoch, with a validation accuracy of 76.49% by the end of training. These outcomes, along with reductions in validation loss over epochs, highlight the effects of the developed methodologies. Source code is available at https://github.com/laithadi/dental-panoramic-xray-segmentation-and-classification.

5. Limitations and Challenges

Navigating the landscape of our project, it becomes evident that while promising, there exist notable limitations and challenges that underscore the non-trivial nature of our endeavor. Firstly, the novelty of the data presents a significant hurdle; being only a year old, there is a scarcity of existing research or established methodologies tailored to this specific dataset. Compounding this issue is the intricate nature of the data itself, characterized by three separate datasets, each endowed with distinct labeling schemas. These variances pose a formidable challenge, necessitating the development of sophisticated strategies to harmonize and integrate disparate labeling structures effectively.

References

[1] Er, S. (2023, June 21). Dentex Challenge 2023. Zenodo. https://zenodo.org/records/7812323#.ZDQE1uxBwUG

[2] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. https://arxiv.org/pdf/1409.1556.pdf

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep residual learning for image recognition. arXiv.org. https://arxiv.org/abs/1512.03385

[4] RobertSmithers. (2022, May 17). Teeth Segmentation. GitHub. https://github.com/RobertSmithers/TeethSegmentation/blob/main/models/best_unet_051722_v1.pth

[5] OpenCV. (n.d.). Morphological Transformations. OpenCV https://docs.opencv.org/4.x/d9/d61/tutorial_py_morphological_ops.html