ED6001 - Term Paper

November 30, 2021

Dental Panoramic X-ray Segmentation

Aqil K H -EE20S049 and Kumari Rashmi -EE20S051

ABSTRACT

Context. Automatic semantic segmentation in one-shot panoramic x-ray image by using conventional image segmentation methods and deep learning method with U-Net Model and binary image analysis in order to provide diagnostic information for the management of dental disorders, diseases, and conditions.

Aims. It is shown that UNet outperforms other conventional segmentation technique

Methods. Global Thresholding,K-Means,Fuzzy C-Means,Gaussian Mixture Model,Watershed Algorithm,Canny Edge Detection,UNet

Results. The UNet segmentation is very fast and accurate when compared with other segmentation methods.

1. Introduction

X-ray images are a tool that is used in dental medicine to check the state of the teeth, gums, jaws and bone structure of a mouth, allowing diagnosis of problems. Particularly, panoramic X-ray is a useful exam to complement the clinical examination in the diagnosis of dental diseases .This type of examination allows the visualization of dental irregularities such as: Teeth included, bone abnormalities, cysts, tumors, cancers, infections, post-accident fractures. Commonly, dentists request panoramic view of the mouth as a preoperative examination of the teeth. Tooth detection has been object of research during at least the last two decades, mainly relying in threshold and region-based methods. Following a different direction, we are exploring a deep learning method for segmentation of the teeth.It is noteworthy that this image type is the most challenging one to isolate teeth, since it shows other parts of patient's body (e.g., chin, spine and jaws).

Deep Learning has enabled the field of Computer Vision to advance rapidly in the last few years. Here we are using one specific task in Computer Vision called as Semantic Segmentation.The goal of semantic image segmentation is to label each pixel of an image with a corresponding class of what is being represented. Because we're predicting for every pixel in the image, this task is commonly referred to as dense prediction.

2. UNET Architecture and Training

The U-shaped architecture shown in Fig.1 consists of a specific encoder-decoder scheme: The encoder reduces the spatial dimensions in every layer and increases the channels. On the other hand, the decoder increases the spatial dims while reducing the channels. The tensor that is passed in the decoder is usually called bottleneck. In the end, the spatial dims are restored to make a prediction for each pixel in the input image. These kinds of models are extremely utilized in real-world applications. The encoder is just a traditional stack of convolutional and max pooling layers. The second path is the symmetric expanding path (also called as the decoder) which is used to enable precise localization using transposed convolutions. Thus it is an end-to-end fully convolutional network (FCN), i.e. it only contains Convo-

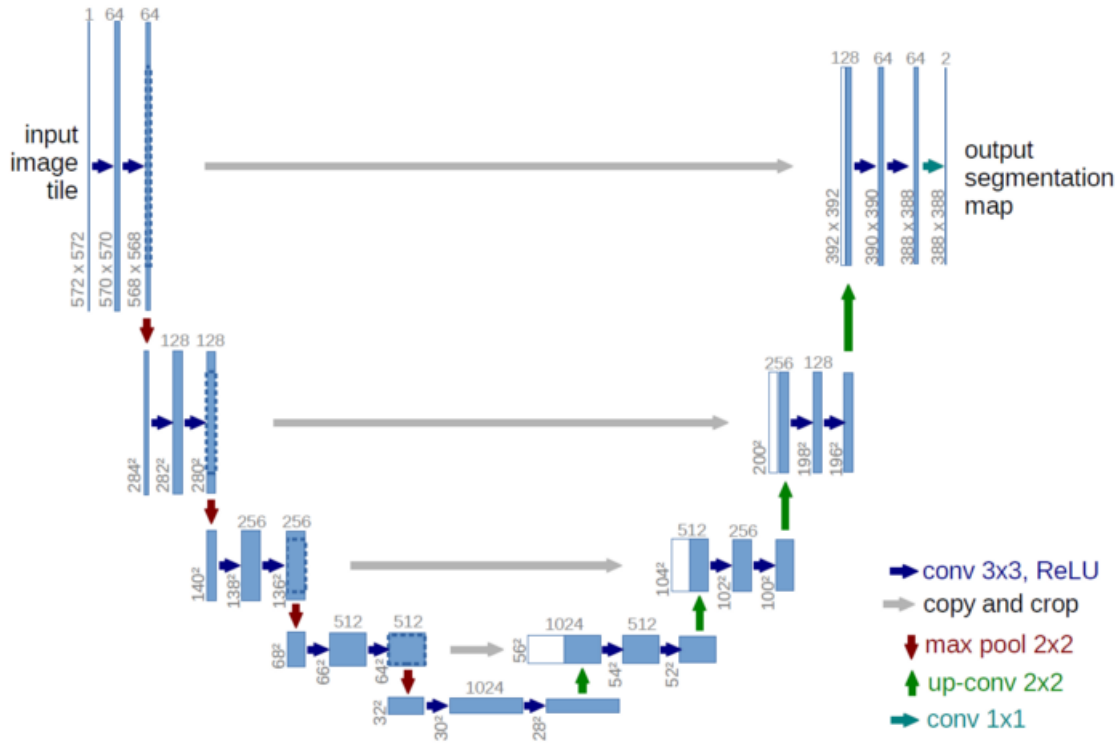


Fig. 1. UNet architecture.

lutional layers and does not contain any Dense layer because of which it can accept image of any size

The architecture shows that an input image is passed through the model and then it is followed by a couple of convolutional layers with the ReLU activation function. We can notice that the image size is reducing from 572X572 to 570X570 and finally to 568X568. The reason for this reduction is because they have made use of unpadded convolutions (defined the convolutions as "valid"), which results in the reduction of the overall dimensionality. Apart from the Convolution blocks, we also notice that we have an encoder block on the left side followed by the decoder block on the right side.

The encoder block has a constant reduction of image size with the help of the max-pooling layers of strides 2. We also have repeated convolutional layers with an increasing number of filters in the encoder architecture. Once we reach the decoder aspect, we notice the number of filters in the convolutional layers start to decrease along with a gradual upsampling in the following layers all the way to the top. We also notice that the use of skip connections that connect the previous outputs with the layers in the decoder blocks.

This skip connection is a vital concept to preserve the loss from the previous layers so that they reflect stronger on the overall values. They are also scientifically proven to produce better results and lead to faster model convergence. In the final convolution block, we have a couple of convolutional layers followed by the final convolution layer. This layer has a filter of 2 with the appropriate function to display the resulting output. This final layer can be changed according to the desired purpose of the project we are trying to perform.

2.1. Training

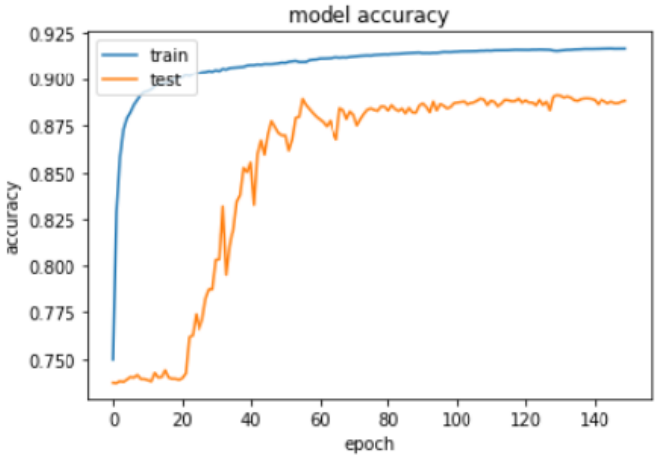
Model is compiled with Adam optimizer and we use binary cross entropy loss function since there are only two classes . We use Keras callbacks to implement:

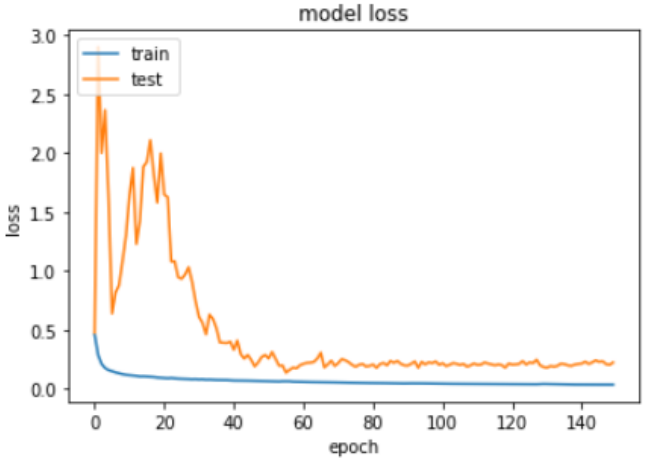
- Learning rate decay if the validation loss does not improve for 5 continues epochs.
- Early stopping if the validation loss does not improve for 10 continues epochs.
- Save the weights only if there is improvement in validation loss.

We use a batch size of 8. Note that there could be a lot of scope to tune these hyper parameters and further improve the model performance.

2.2. The Training and Testing Accuracy and Loss on Different Epoch Number

Figure shows the graph of training and testing accuracy and loss value against change of epoch. The training accuracy graph is shown by blue line, while testing accuracy is shown by orange line.





2.3. Optimizers Used For Training

Adam (Adaptive Moment Estimation) works with momentums of first and second order. The intuition behind the Adam is that we don't want to roll so fast just because we can jump over the minimum, we want to decrease the velocity a little bit for a careful search. In addition to storing an exponentially decaying average of past squared gradients like AdaDelta, Adam also keeps an exponentially decaying average of past gradients Its Advantages are:

- The method is too fast and converges rapidly.
- Rectifies vanishing learning rate, high variance.

2.4. UNet Result

3. Otsu Thresholding

We have used Otsu's method to segment the image in 2 classes i.e., background and foreground. It is a type of Global thresholding method.

3.1. Algorithm

- Compute histogram and probabilities of each intensity level.
- Set up initial class probability and initial class means.
- Step through all possible thresholds maximum intensity.
- Update mean and variance.
- Compute between class variance.
- Desired threshold corresponds to the maximum value of between class variance.

3.2. Otsu Thresholding Result

4. K-Means Clustering

We have used K-Means here to segment the image in 5 classes. Have taken 5 because foreground might have 4 different types of intensity. Have used K-Means here to understand the different clusters of our teeth x-ray. It is a type of hard clustering method. The K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible.

4.1. Algorithm

- Choose the number of clusters k
- Select k random points from the data as centroids
- Assign all the points to the closest cluster centroid
- Recompute the centroids of newly formed clusters
- Repeat steps 3 and 4

4.1.1. Stopping Criteria for K-Means Clustering

There are essentially three stopping criteria that can be adopted to stop the K-means algorithm:

- Centroids of newly formed clusters do not change
- Points remain in the same cluster
- Maximum number of iterations are reached

4.2. K-Means Result

5. Gaussian Mixture Model

A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians.

The GaussianMixture object implements the expectation-maximization (EM) algorithm for fitting mixture-of-Gaussian models. expectation-maximization converges faster than gradient descent algorithm, however expectation minimization can converge at local maxima.

5.1. Gaussian Mixture Model

6. Fuzzy C-Means

Fuzzy C-Means clustering is a soft clustering approach, where each data point is assigned a likelihood or probability score to belong to that cluster. It is generalized form of k-means algorithm

6.0.1. Algorithm

The step-wise approach of the Fuzzy c-means clustering algorithm is:

- Fix the value of c (number of clusters), and select a value of m (generally 1.25<m<2), and initialize partition matrix U
- Calculate cluster centers (centroid).
- Update Partition Matrix
- Repeat the above steps until convergence.

6.1. Fuzzy C-Means Result

7. Watershed segmentation

Watershed segmentation is a region-based technique that utilizes image morphology . It requires selection of at least one marker (“seed” point) interior to each object of the image, including the background as a separate object. The markers are chosen by an operator or are provided by an automatic procedure that takes into account the application-specific knowledge of the objects. Once the objects are marked, they can be grown using a morphological watershed transformation.

7.1. Algorithm

The watershed transform finds "catchment basins" and "watershed ridge lines" in an image by treating it as a surface where light pixels are high and dark pixels are low.

Segmentation using the watershed transform works better if you can identify, or "mark," foreground objects and background locations. Marker-controlled watershed segmentation follows this basic procedure:

- Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment.
- Compute foreground markers. These are connected blobs of pixels within each of the objects.
- Compute background markers. These are pixels that are not part of any object.
- Modify the segmentation function so that it only has minima at the foreground and background marker locations.
- Compute the watershed transform of the modified segmentation function.

7.2. Watershed segmentation Result

8. Canny Edge Detection

We have used Canny Edge detector to understand the edge of teeth. We have tried with different threshold values to detect the edge of the x-ray image The Canny filter is a multi-stage edge detector. It uses a filter based on the derivative of a Gaussian in order to compute the intensity of the gradients.The Gaussian reduces the effect of noise present in the image. Then, potential edges are thinned down to 1-pixel curves by removing non-maximum pixels of the gradient magnitude. Finally, edge pixels are kept or removed using hysteresis thresholding on the gradient magnitude. The Canny has three adjustable parameters: the width of the Gaussian (the noisier the image, the greater the width), and the low and high threshold for the hysteresis thresholding. The general criteria for edge detection include:

- Detection of edge with low error rate, which means that the detection should accurately catch as many edges shown in the image as possible
- The edge point detected from the operator should accurately localize on the center of the edge.
- A given edge in the image should only be marked once, and where possible, image noise should not create false edges.

8.1. Canny Edge Detection Algorithm

- Apply Gaussian filter to smooth the image in order to remove the noise
- Find the intensity gradients of the image
- Apply non-maximum suppression to get rid of spurious response to edge detection
- Apply double threshold to determine potential edge
- Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

8.2. Canny Edge Detection Result

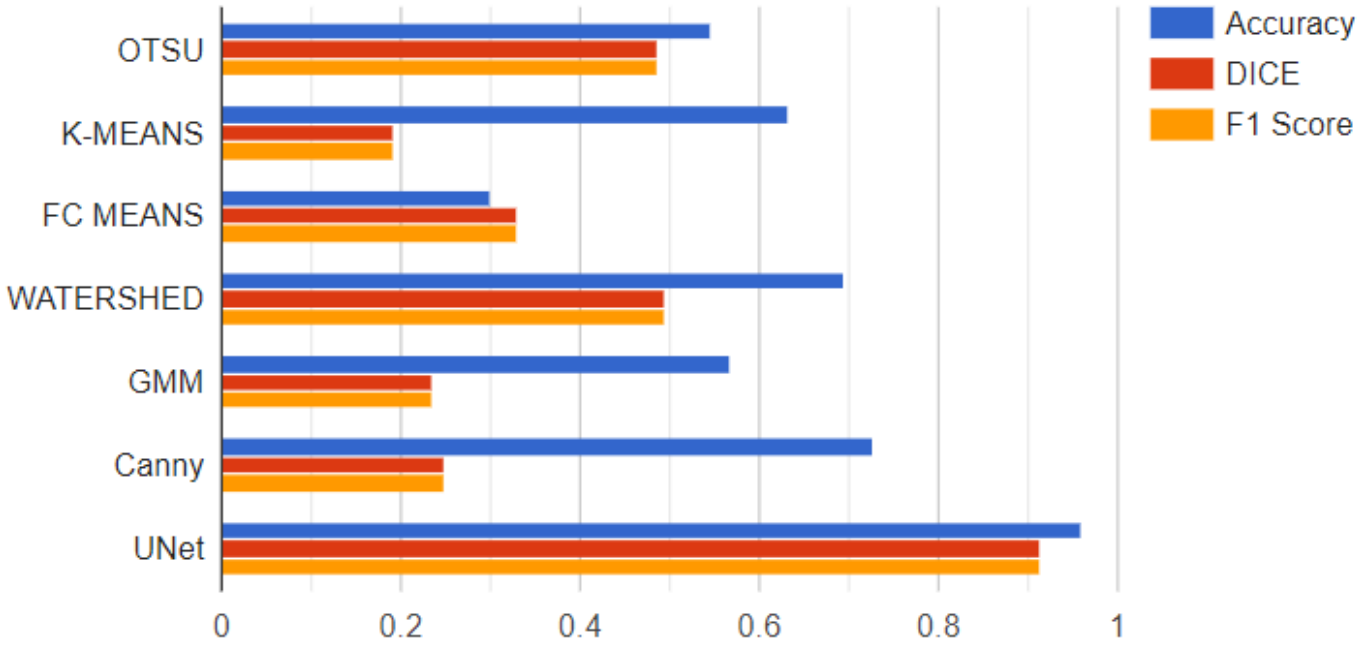
9. EXPERIMENTAL ANALYSIS

To assess the performance of the segmentation, the following metrics were used: Dice, Accuracy,Sensitivity,Specificity,Jaccard similarity coefficient,F1 Score. These metrics were used in a pixel-wise fashion. Table I summarizes the quantitative results found by our system. All standard deviations was found very low, indicating that all the individual results were all near the mean. This fact demonstrates that, although the data set was challenging, the proposed system achieves a good generalization and consistency in the results. In a nutshell, results indicated a good balance between true negative/false negative (specificity, accuracy) and true positive/false positive (accuracy, precision and recall) rates, which are ultimately attested with the F1-score, computing the harmonic mean between recall and precision. UNet demonstrated highly superior results in comparison to unsupervised methods evaluated

In semantic segmentation tasks, we predict a mask, i.e. where the object of interest is present. We only have one type of object to predict thus it is a binary task.We can thus assign the following mapping:

0 for the background class. 1 for the object of interest. Now that we know what we are predicting, we can move to the metrics. Pixel accuracy is perhaps the easiest to understand conceptually. It is the percent of pixels in your image that are classified correctly. While it is easy to understand, it is in no way the best metric. At first glance, it might be difficult to see the issue

with this metric. High pixel accuracy doesn't always imply superior segmentation ability. When our classes are extremely imbalanced, it means that a class or some classes dominate the image, while some other classes make up only a small portion of the image. Unfortunately, class imbalance is prevalent in many real world data sets, so it can't be ignored. The Intersection-Over-Union (IoU), also known as the Jaccard Index, is one of the most commonly used metrics in semantic segmentation… and for good reason. The IoU is a very straightforward metric that's extremely effective.



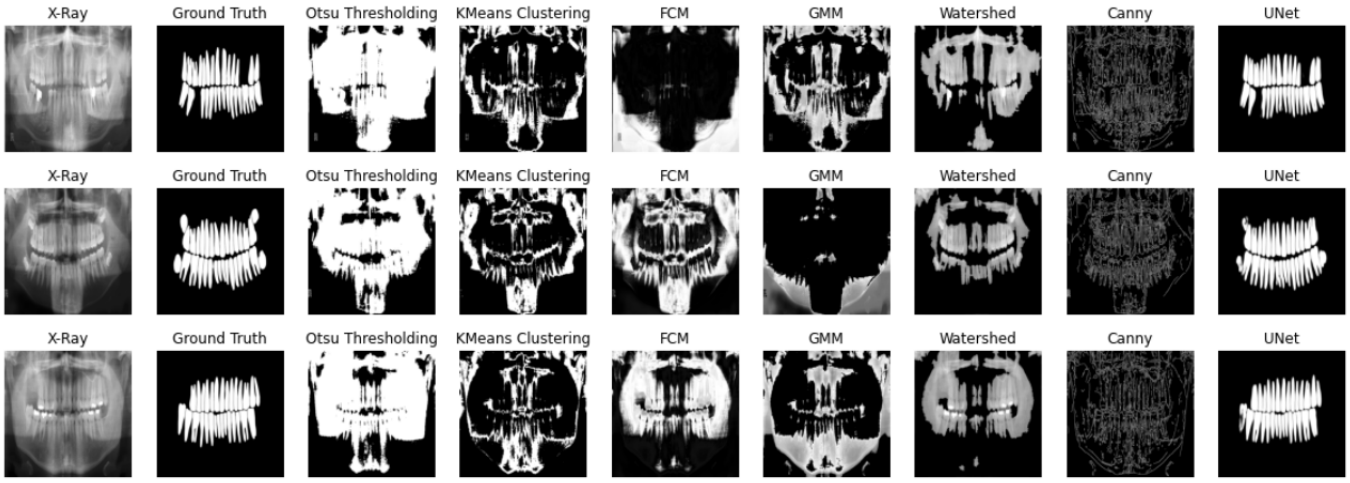


Fig. 2. Experimental Analysis of Various Segmentation Methods

Table 1. COMPARISON OF THE UNSUPERVISED METHODS STUDIED AND UNet

	OTSU	K-MEANS	FC-MEANS	WATERSHED	GMM	Canny	UNet
DICE	0.485346	0.193223	0.329743	0.494625	0.235662	0.249755	0.912742
ACCURACY	0.545353	0.631813	0.300869	0.693508	0.568325	0.725922	0.958622
SENSITIVITY	0.325888	0.191038	0.210365	0.393779	0.196192	0.326454	0.870428
SPECIFICITY	0.967258	0.764056	0.706262	0.877897	0.759335	0.790806	0.987795
JC	0.320434	0.106943	0.197421	0.328573	0.133570	0.142697	0.839489
F1 Score	0.485344	0.193221	0.329742	0.494622	0.235660	0.249752	0.912734

10. Individual Contribution

Methods	Individuals Contributed
OTSU	Kumari Rashmi
K-MEANS	Kumari Rashmi
FC-MEANS	Aqil K H
WATERSHED	Aqil K H
GMM	Aqil K H
Canny	Kumari Rashmi
UNet	Aqil and Rashmi

11. Conclusion

The study was developed using 105 images for training and 27 images for testing purposes. In this study, the UNet based teeth detection model was used to detect teeth in the Panoramic images. This model scores 96 % accuracy, 99% specificity and 87% sensitivity (Table 1). Considering that the UNet system demonstrated promising results on a challenging data set, future work resides on the instance segmentation of each component part of the mouth and teeth, as well as detection of missing teeth.

Model used	Sequential parameters
Activation function (Input)	ReLU
Activation function (Output)	SoftMax
Optimizer	Adam
Loss function	Binary Cross entropy
Number of epochs	150
Batch size	8
Validation split	0.1

The output of semantic segmentation is not just a class label or some bounding box parameters. In-fact the output is a complete high resolution image in which all the pixels are classified. Thus if we use a regular convolutional network with pooling layers and dense layers, we will lose the “WHERE” information and only retain the “WHAT” information which is not what we want. In case of segmentation we need both “WHAT” as well as “WHERE” information. Hence there is a need to up sample the image, i.e. convert a low resolution image to a high resolution image to recover the “WHERE” information. In the literature, there are many techniques to up sample an image. Some of them are bi-linear interpolation, cubic interpolation, nearest neighbor interpolation, unpooling, transposed convolution, etc. However in most state of the art networks, transposed convolution is the preferred choice for up sampling an image. Each pixel we get a value between 0 to 1. 0 represents no salt and 1 represents salt. We take 0.5 as the threshold to decide whether to classify a pixel as 0 or 1. However deciding threshold is tricky and can be treated as another hyper parameter.

Optimization algorithms or strategies are responsible for reducing the losses and to provide the most accurate results possible.We tried different types of optimizers. For Gradient Descent the model got trap at local minima. Required large memory to calculate gradient on the whole dataset. For this data set Adam was the best optimizers. If one wants to train the neural network in less time and more efficiently than Adam is the optimizer.
Oral cancer is a significant health problem throughout the world. It is very important to detect such types of cancer at an earlier stage than the later stage where the treatment becomes unsuccessful. Early detection helps surgeons to provide necessary therapeutic measures which also benefit the patients. This segmentation method can be further developed to preserve the edge details as well as prominent ones to identify tumors in dental radiography.