Adenoid segmentation in X-ray images using U-Net

Ali Abdullah Alshbishiri
College of Computer Science and Engineering
University of Jeddah
Jeddah, Saudi Arabia
1900214@uj.edu.sa

Rani Ghazi Ahmad
Department of Radiology
King Abdulaziz University
Jeddah, Saudi Arabia
rahmad@kau.edu.sa

Muath Abdulrahim Marghalani
Department of Radiology
King Abdulaziz University
Jeddah, Saudi Arabia
m.marghalani.s@gmail.com

Mohammed Ali Alqarni
College of Computer Science and Engineering
University of Jeddah
Jeddah, Saudi Arabia
alqarni@uj.edu.sa

Hassan Aqeel Khan
College of Computer Science and Engineering
University of Jeddah
Jeddah, Saudi Arabia
hakhan@uj.edu.sa

Muhammad Murtaza Khan
College of Computer Science and Engineering
University of Jeddah
Jeddah, Saudi Arabia
mkhan@uj.edu.sa

2021 National Computing Colleges Conference (NCCC) | 978-1-7281-6719-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/NCCC49330.2021.9428866

***Abstract*— Use of machine learning and specifically deep learning-based techniques for medical diagnosis has created a significant impact on early and easier diagnosis in the domain of radiology. Deep learning techniques have demonstrated unprecedented superiority in all facets of medical image analysis ranging from classification to identification to segmentation. The efficacy of deep learning algorithms to process X-ray data and extract meaningful information from it has helped diagnose and provide timely health care to patients. The focus of proposed work is to use DICOM x-ray images for detection and segmentation of adenoid gland using deep learning-based techniques. The distance between the Adenoid gland and soft palate may be used by doctors to identify the severity and type of diseases and hence an automated method for identification of adenoid gland will help in automatic diagnostics. The main challenge is that the size and shape of adenoid gland varies with age and disease and hence because of its deformative property it is difficult to segment it. In this work, we propose to use U-net based technique for segmentation of adenoid gland. To the best of our knowledge this is the first attempt to solve the problem of adenoid detection and segmentation using U-net based deep learning architecture.**

Keywords—deep learning, U-net, Adenoid, segmentation

I. INTRODUCTION

Adenoid hypertrophy may be the result of various negative clinical conditions [1]. Therefore, detection of such conditions by observing adenoid gland is of great interest [2, 3]. One method to observe this condition is the use of endoscopy as discussed by Feres et al. in [1]. Another approach is to make use of x-ray images for determining the size of the adenoid gland as discussed in [4]. The focus of both the approaches is to determine the size of the adenoid, anatomical correlation and its effect on the nasopharynx and airway to assess any restriction in the nasal passage.

Detection of Adenoid is generally done by an expert radiologist. To do this the radiologist may have to process the image so that it becomes clear by applying various intensity-based transformations. Different capturing techniques, position and age of patient causes difficulties in diagnosing. Transformations that would make the Adenoid clearly visible depend on the imaging conditions and hence are subject to the experience of the radiologist. Novice or fatigued users may not be able to apply accurate correction to the image and hence an automated process may make the process of correcting the image or identifying the Adenoid in the image beneficial. Another advantage of automated detection is that it may be used to diagnose any health issues immediately and without increasing the workload of the radiologist.

To analyze the Adenoid gland a radiologist needs to determine its size. This means that they would like to differentiate the Adenoid from its surroundings. This can be referred to as segmentation in computer vision. Therefore, the goal of the automated algorithm should be to identify and segment the Adenoid gland from the x-ray images. It is assumed that the images that will be provided to the algorithm will be from the head and neck region and contain the Adenoid gland in them. The process of automatic segmentation can be divided into pre-processing and then training of deep neural network for segmentation. We propose to the use U-net based architecture for training the network. The input to the network is an image and the output is an image with the Adenoid region highlighted in red color. The algorithm is tested using DICOM data captured at the King Abdulaziz University hospital. The proposed algorithm produces a Dice coefficient score of 74% for the Adenoid region.

The rest of the paper is organized as follows. Section II presents a brief overview of deep leaning methods for segmentation. Section III discusses the proposed methodology by highlighting the pre-processing steps along with data labeling procedure. Section IV covers the results obtained by the proposed network. Section V presents conclusions and opportunities for future work.

II. LITERATURE REVIEW

To the best of our knowledge there is no existing work that focuses on the use of deep learning for automatic extraction of Adenoid gland. Since literature on the exact topic was not available, we focused on identifying deep learning algorithms that were used for segmentation of biomedical images.

Determining the volume and shape of organs and substructures is of importance in biomedical imaging and hence tasks related to segmentation are important. In terms of volumetric structures, it is the set of voxels which are of interest however, for two-dimensional shape simple segmentation results are sufficient. One of the most well-known methods for medical image segmentation was proposed by Ronneberger et al. in 2015 [5]. The novelty of the method revolved around a combination of equal amounts of down sampling and up sampling in encoder and decoder layers. This allowed the proposed U-net to perform segmentation on the whole image instead of the patch-based approaches utilized by ordinary Convolutional Neural Network (CNN) based methods. In 2016, Cicek et al. [6] extended the use of U-net for 3D segmentation by feeding the 2D U-net with slices from the same volume. V-net a variant of U-net was proposed by Milletari et al. in [10] by utilizing the Dice coefficient as the objective function.

Authors in [11] proposed to use U-net for diagnosing osteoporosis to address the problem of diagnostic variability. Dong et al. proposed the use of U-net for detection and segmentation of brain tumors in [12]. In [13], Shaziya et al. proposed to use U-net for segmentation of lungs cancer while in [14] the authors proposed to use U-net for organ segmentation to support radiotherapy. Wu et al. proposed the use of U-net for automatic detection of coronary artery in [15].

Recurrent neural networks (RNN) have also found success for segmentation tasks. Xie et al. proposed to use RNN for segmentation of histopathological images in [16]. To incorporate bidirectional information RNN must be applied four times in different directions. Chen et al. combined U-net and bi-directional LSTM-RNNs to segment structures in 3D electron microscopy images [17]. In [18], authors propose a fully convolutional neural network (fCNN) for segmenting brain MRI images, pectoral muscle in breast and coronary arteries in cardiac CT angiography.

The literature on use of deep neural networks for segmentation or for processing biomedical images is extensively large however, we have tried to briefly present different approaches for the task of segmentation and on the types of problems. Since Adenoid detection using U-net was not found during our literature review and since U-net is a popular segmentation technique for small datasets we propose to use it for segmentation of Adenoid gland.

III. Proposed Methodology

Detection of Adenoid gland from Digital Imaging and Communications in Medicine (DICOM) x-ray images is challenging because of the variability in shape, size and position of the gland. To the best of our knowledge the detection of size and shape of the Adenoid has not been attempted using deep learning techniques. In this work we propose to use U-net for detection and segmentation of adenoid gland from lateral head x-rays.

The challenges associated with this task involve classic image processing problems. The images obtained are at different scales or in other words have varying degree of zoom. The images are not aligned with respect to translation or rotation. The images do not have the same exposure. The view varies between left and right side for acquisition. The region being captured varies significantly. All the highlighted challenges are evident from the sample images presented in Figure 1. It may be noted that all the images presented in Figure 1 contain the adenoid gland and that it is visible to an expert human observer.

To make the processing of images reliable the first step in the process was to pre-process the images so that they have approximately the same dimensions, view, and intensity range. Also, since the images contain information about the patients that information must be removed for the purposes of privacy. After pre-processing all the images were manually labeled. Finally, U-net based deep learning algorithm was trained using the labeled data. The details of individual steps are presented below.

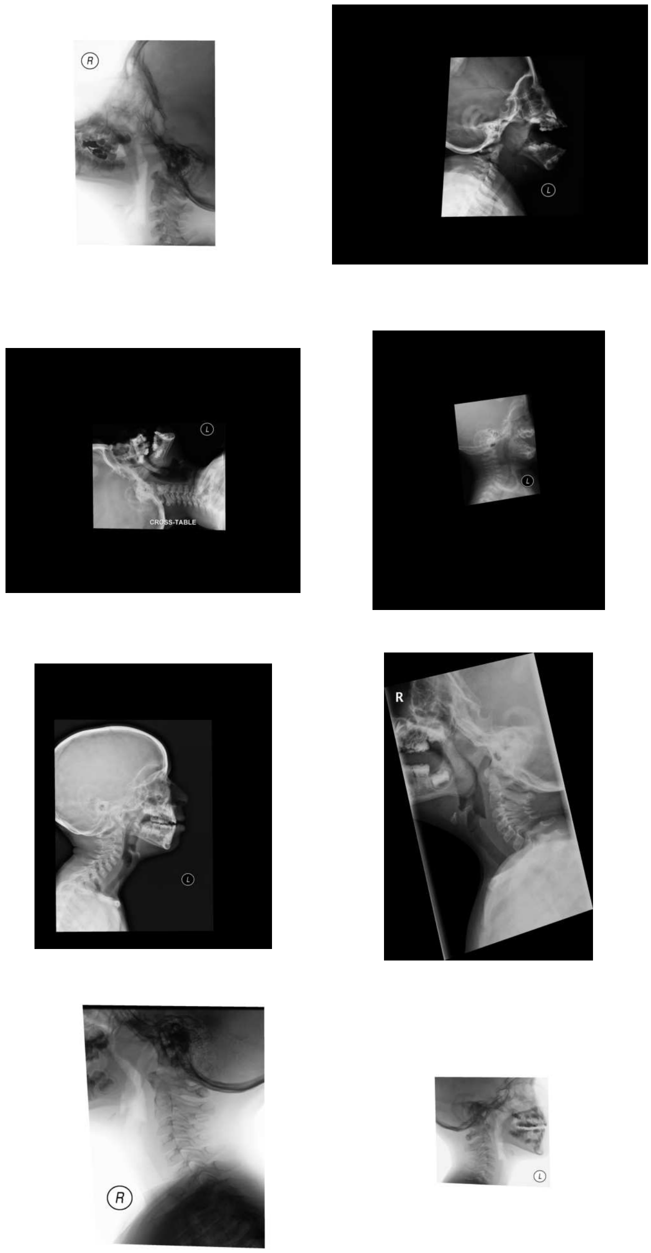


Fig. 1. Sample dicom images of lateral head and neck view

A. Preprocessing of dicom images

The dataset comprised of DICOM images and they contain information patient's personal information along with information about the captured image. As a first step any information related to the name, age, gender of patient or the doctor was removed form the dataset. Next, information related to the lateral view i.e., right, left or both was extracted. Also, information related to the top left corner and bottom right corner was extracted. This information was used to crop to the region of image that contains the x-ray. This information was obtained from the Collimator edge information available in the DICOM images. Any other information related to field of view rotation, distance between pixels and exposure time was extracted. Since the DICOM image comprises of a single monochromatic channel it is converted into a three-channel image by copying the same monochromatic channel in each of the red, green and blue channels.

As shown in Figure 1, some images have black background while others have a white background thus indicating that the images have inverted intensity variation range. This issue was fixed by observing if the Photometric Interpretation field of DICOM image is Monocrhome1 or not. If the field was Monochrome1 the images were inverted by subtracting the image from the maximum value in the image. This resulted in all images have a black background with high density regions i.e., bones appearing white in the image.

Rotation or shearing of images was not addressed and the images with rotation or shear were used without change. From Figure 1, the actual exposed part of the image varies in size. This issue was addressed by cropping the image by making use of Collimator left vertical edge, right vertical edge, upper horizontal edge, and lower horizontal edge fields. Next the size of each image was made the same by scaling. This is a requirement since for a network the size of input layer is fixed and hence images with varying dimensions cannot be used. Therefore, all images were all rescaled to the resolution of 2048x2048 pixels. The steps of pre-processing stage are shown in Figure 2 and sample pre-processed images are shown in Figure 3.

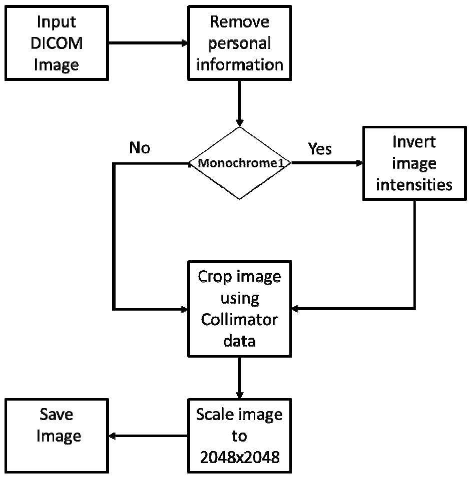


Fig. 2. Pre-processing to obtain images having same size and intensity range

B. Dataset

The dataset used for this research comprised of DICOM images of lateral head and neck region. It was collected at King Abdulaziz University hospital. The dataset comprised of 672 images. The data was split into 416 images for training and 256 test images. Thus, the training to testing split was approximately 60% and 40%, respectively. The dataset collection was performed independent of the study; this means that the data was not acquired specifically for this study and hence was captured in the real-world scenario as much as possible. Hence, it can be considered as a representation of the actual problem solving that will observed when diagnosing patients in daily hospital environment.

C. Data labeling

Once the data was acquired the next phase was labeling of data for training and testing. To achieve this task, we have used the expertise of the radiology department of King Abdulaziz University hospital. Two expert radiologists, part of the research team, marked the ground truth. To simplify the process the procedure for marking the ground truth a document was shared with the radiologists and a small video was shared with them as well. To facilitate the process an open-source image editing tool called Gimp was used. Sample ground truth marked images provided by the experts are shown below in Figure 4. The adenoid is marked in red color so that it visible. From the two samples in Figure 4 it is clear that the shape, size, location and orientation of adenoid gland varies significantly in the captured images.

Fig. 3. Images of Figure 1 after pre-processing

D. Data augmentation

The size of dataset is relatively small for machine learning tasks. Therefore, we have performed data augmentation to increase the number data samples. This is done at run time i.e., when data is read for training and hence does not require any additional storage space. The data is augmented by translating, rotating or flipping the images. This increases the number of collected samples and variability in the dataset.

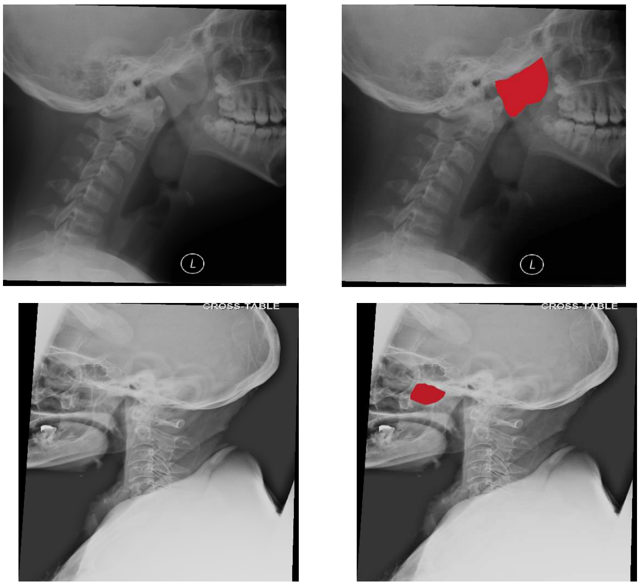


Fig. 4. Results of ground truth (data labeling)

E. Segmentation Architecture and Training

The main task of this work was to identify the adenoid gland in x-ray images. This process may be referred to as segmentation and in certain cases semantic segmentation. For this task we explored different segmentation techniques and our research highlighted that encoder-decoder based architecture is generally successful in such tasks. One such network called U-net has been very successful for tasks related to biomedical image segmentation [5]. U-net was inspired from a fully convolutional neural network proposed by Long et al. in [7]. Also, U-net has been found to be the most efficient algorithm for semantic segmentation of biomedical images especially in case of a limited dataset [8]. In [9], a number of architectures (U-net, X-net & SegNet) were tested for segmentation of three features on dental radiographs; results demonstrated that U-net significantly outperformed other architectures. This robust performance on a number of existing tasks was the primary reason why we decided to employ U-net for our task.

The main idea behind U-net is to use a contracting path and an expansive path while making use of convolutions. On the encoder side i.e., the contracting path, the convolutions are followed by max pooling operation with a stride of 2 pixels for down sampling the images. While, on the decoder side i.e., the expansive path, the features pass through an upscaling operation which is followed by convolution operation. It should be noted that the number of feature maps in the contracting path are doubled after down sampling while on the expansive path the number of feature maps is halved after each up sampling. The structure of the U-net used in this research is shown below in Figure 5.

U-net can be trained in a supervised manner and hence the training data set along with its marked ground truth or labels is provided to the algorithm as input. The algorithm randomly samples the input data and performs training and validation using the training data set. On each iteration the network parameters are updated, and validation accuracy is calculated. The network was trained employing different number of iterations in multiples of 20 and it was observed that the network presented the best results after 120 iterations.

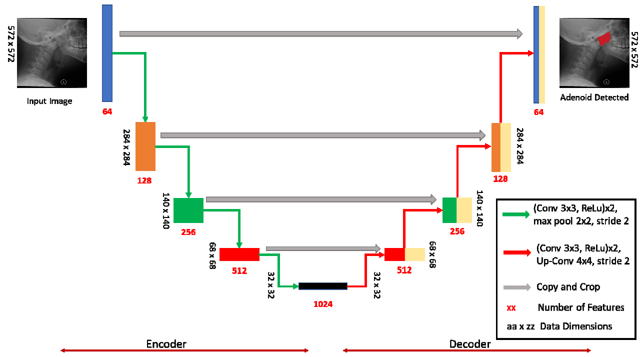


Fig. 5. Unet architecture used for Adneoid segmentation

IV. EXPERIMENTATION AND RESULTS

To test the performance of the algorithm the data was split up in to training and test data. The test data set comprised of 256 images. It is assumed that all the images provided to the U-net have the Adenoid gland present in them. However, the orientation, location and shape of the gland is different in them.

To assess the quality of semantic segmentation we employed Intersection over Union (IoU) based quantitative quality assessment measure. IoU is the ratio between number of overlapping pixels between reference and predicted label (mask) and total number of pixels in both labels (masks). Mathematically, it can be represented as follows:

$$\text{IoU} = \text{Ground Truth} \cap \text{Prediction} \,/\, \text{Ground Truth} \cup \text{Prediction}$$
$$= TP / (TP + FN + FP) \qquad (1)$$

True positive (TP) represents the number of pixels that are correctly predicted as belonging to target class. False Positive (FP) represents number of class pixels which are incorrectly marked as belonging to target class. False Negative (FN) represents number of pixels which are missed by the algorithm and are not marked as belonging to the target class. The advantage of using IoU measure is that it gives a good idea of how accurately the shape of the Adenoid gland was detected.

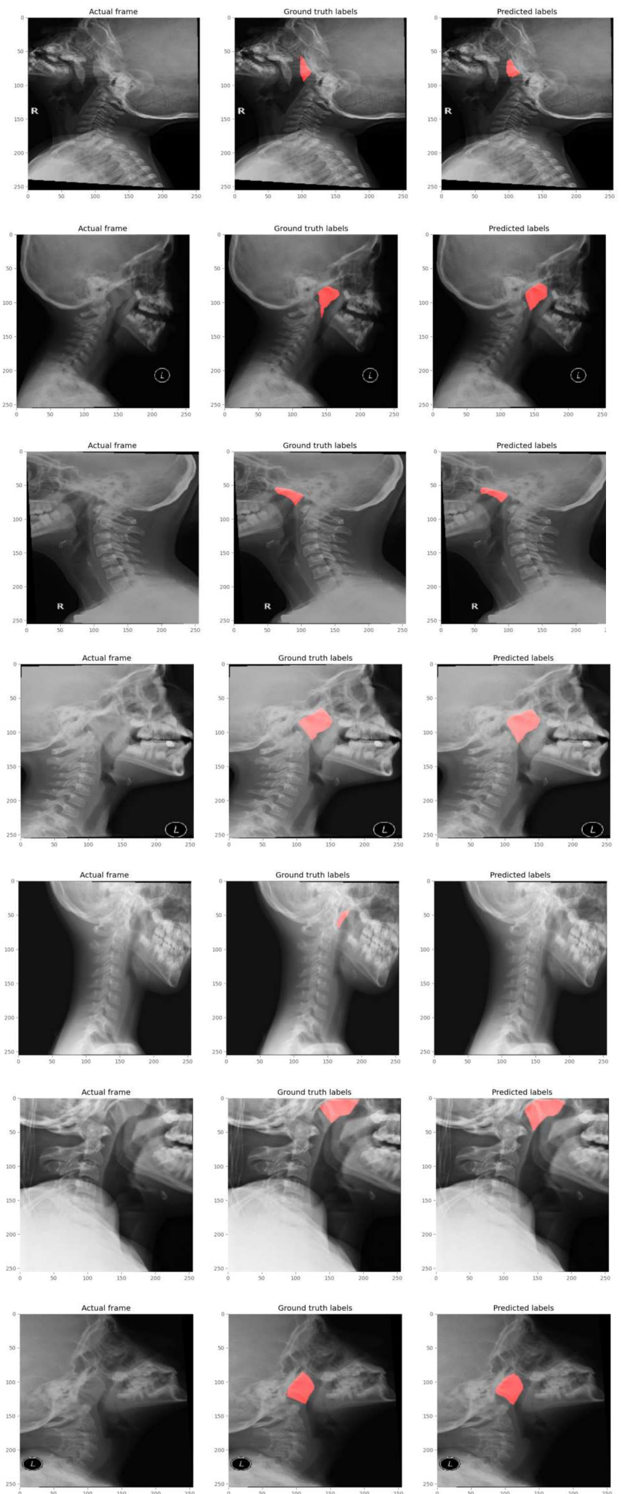
To further evaluate the quality of results we have employed a second quantitative quality assessment measure titled 'Dice coefficient'. It is defined as

$$\text{Dice} = 2\,TP / (2TP + FN + FP) \qquad (2)$$

The difference between IoU and Dice is that IoU penalizes single instance of incorrect segmentation. This means that an algorithm which is incorrect only a few times may result in an IoU score that is much less than that of dice coefficient. Dice coefficient is better at determining average performance as compared to each individually.

TABLE I. IoU and Dice Scores for Background and Adenoid

	IoU Score	Dice Score
Background	0.99	0.99
Adenoid	0.62	0.74



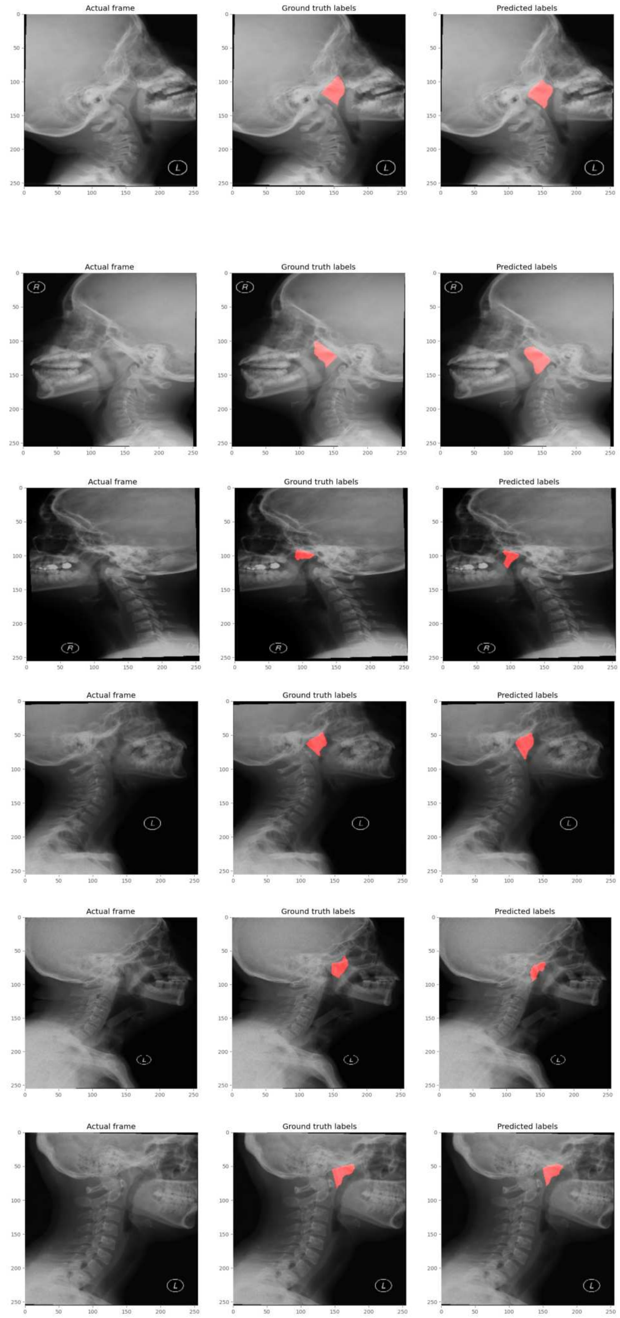


Fig. 6. Visual comparison between ground truth and predicted results

Qualitative assessment of some of the results, presented in Figure 6, suggest that the algorithm is performing adequately at identifying the Adenoid region. However, quantitative analysis demonstrates that the algorithm has a Dice score of 0.74 for the Adenoid while 0.99 for the background. Thus, the performance for Adenoid segmentation task is being affected by the imbalance in the number of pixels between the background and Adenoid classes.

V. CONLCUSION

In this work we have presented the results of segmentation of Adenoid gland using U-net based deep learning method. The results were calculated using true data acquired at the King Abdulaziz University hospital and thus are a good representation of real-world problem. To prepare the data for deep learning pre-processing had to be performed so that the images followed the same monochrome format and were of the same size. The training and test data comprised of 416 and 256 images respectively and the algorithm returned an IoU score of 0.62 and a Dice score of 0.74 for the 256 test images.

Qualitative analysis of the results suggest that the algorithm is performing significantly well however, to improve the qualitative and quantitative scores it may be better to select a region of interest in the image so that there is less bias towards the background class pixels which are far more in number as compared to the target class pixels. To further improve the results Generative Adversarial Networks (GANs) may be explored to increase the overall size of the training data. Finally, variants of U-net with different encoders may be explored to increase the IoU and Dice scores.

Acknowledgment

The authors would like to thank King Abdulaziz University hospital for providing the data required for this research.

References

[1] M. Feres, J. Hermann, A. Sallum, and S. Pignatari, "Endoscopic evaluation of adenoids: Reproducibility analysis of current methods", Clincial and Experimental Otorhinolaryngology, vol. 6, no. 1, pp. 36–40, March 2013.

[2] F. Pagella, A. Colombo, O. Gatti, G. Giourgos, and E. Matti, "Rhinosinusitis and otitis media: the link with adenoids", International Journal of Immunopathology and Pharmacology. no. 23, pp.38-40, 2010.

[3] S. Izu, C. Itamoto, M. Hallinan, G. Pizarro, S. Tufik, S. Pignatari, R. Fujita, "Obtrusive sleep apnea syndrome (OSAS) in mouth breating children", Brazilian Journal of Otohinolaryngology, vol.76, no.5, pp. 552-556, 2010.

[4] A. Mlynarek, M. Tewfik, A. Hagr, J. Manoukian, M. Schloss, T. Tewfik, and J. Choi-Rosen, "Lateral neck radiography versus direct video rhinoscopy in assessing adenoid size", The Journal of Otolarngology, vol. 33, no.6, pp.360-365, 2004.

[5] O. Ronneberger, P. Fischer, T. Brox, "Dental X-ray image segmentation using a U-shaped deep convolutional network", International Symposium on Biomedical Imaging, pp.1-13, 2015.

[6] O. Cicek, A. Abdulkadir, S. Lienkamp, T. Brox, O. Ronnenberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation", International Conference on Medical Image Computing and Computer-Assisted Intervetion, pp.424-432, 2016.

[7] J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440, 2015.

[8] G. Litjens, T. Kooi, B. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J. Laak, B. Ginneken, C. Sanchez, "A survey of deep learning in medical image analysis", Medical Image Analysis, no. 42, pp.60-88, 2017.

[9] H. Khan, M. Haider, H. Ansari, H. Ishaq, A. Kiyani, K. Sohail, M. Muhammad, S. Khurram, "Automated feature detection in dental periapical radiographs by using deep learning", Oral surgery, oral medicine, oral pathology and oral radiology, In press. 2020.

[10] F. Milletari, N. Navab, S. Ahmadi, "V-Net : Fully convolutional neural networks for volumetric medical image segmentation", 4th IEEE International Confernece on 3D Vision, pp. 565-571, 2016.

[11] K. Hatano, S. Murakami, H. Lu, J. Tan, H. Kim, T. Aoki, "Detection of phalange region based on U-Net", International Conference on Control, Automation and Systems, pp.1338-1342, 2018.

[12] H. Dong, G. Yang, F. Liu, Y. Mo, Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks", Communications in Computer and Information Science, pp. 506–517, 2017.

[13] H. Shaziya, K. Shyamala, R. Zaheer, "Automatic Lung Segmentation on Thoracic CT Scans Using U-Net Convolutional Network", Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing (ICCSP), pp.643–647, 2018.

[14] N. Yagi, M. Nii, S. Kobashi, "Abdominal organ area segmentation using u-net for cancer radiotherapy support", IEEE International Conference on Systems, Man and Cybernetics, pp.1210–1214, 2019.

[15] W. Wu, J. Zhang, H. Xie, Y. Zhao, S. Zhang, L. Gu, "Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint". Computers in Biology and Medicine, vol. 118, 103657, 2020.

[16] D. Xue, R. Zhang, H. Feng, Y. Wang, "CNN-SVM for microvascular morphological type recognition with data augmentation", Journal of Medical and Biological Engineering, no.36, vol.6, pp.755–764, 2016.

[17] J. Chen, L. Yang, Y. Zhang, M. Alber, D. Chen, "Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation", In proceedings of Advances in Neural Information Processing Systems (NIPS), pp. 3036–3044, 2016.

[18] P. Moeskops, J. Wolterink, B. Velden, K. Gilhuijs, T. Leiner, M. Viergever, I. Isgum, "Deep learning for multi-task medical image segmentation in multiple modalities". In proceedings of Medical Image Computing and Computer-Assisted Intervention, vol. 9901 of Lect Notes Comput Sci. pp. 478–486, 2016.