Segmentation and Localization of Anatomical Structures in Chest X-Ray Images using Advance Deep Learning Architectures

Author

Saad

00000319501

Supervisor

Dr. Sajid Gul Khawaja

Co-Supervisor

Dr. Muhammad Usman Akram

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

JANUARY 2022

Segmentation and Localization of Anatomical Structures in Chest X-Ray Images using Advance Deep Learning Architectures

Author

Saad

00000319501

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Computer Engineering

Thesis Supervisor

Dr. Sajid Gul Khawaja

Thesis Supervisor's Signature: _____

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD

January 2022

Declaration

I certify that this research work titled "*Segmentation and Localization of Anatomical Structures in Chest X-Ray Images using Advance Deep Learning Architectures"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Saad
00000319501

Language Correctness Certificate

This thesis has been read by an English expert and is free of most typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

Saad
00000319501

Signature of Supervisor

Dr. Sajid Gul Khawaja

Copyright Statement

Acknowledgements

By the grace of the lord above, my Master's thesis is finally complete. It was quite an arduous exercise that could not have been completed without the help of God Almighty and the strength that he bestowed upon me.

I am also eternally grateful to all those, especially my friends and family, who helped me with the thesis every step of the way. Their support was what made me accomplish this herculean task. Without them, it would not have been possible.

Lastly, I would like to offer my sincere thanks to my incredible supervisors and the entire committee: Sir Usman Akram, Sir Sajid Gul Khawaja, Sir Wasi Haider and Sir Arsalan Shaukat. Without their support – in every sense of the word – and constant guidance, this thesis would never have been conceived or completed (and definitely not in time). I cannot thank them enough for their role in the completion of this thesis and report.

Dedicated to my exceptional parents: ***Zia Kibria & Sofia Zia****, and adored sister whose tremendous support and cooperation led me to this accomplishment*

Abstract

The COVID 19 pandemic has been a very tough time for people around the globe, be it medical professionals who had to work for long hours or patients who had to deal with shortage of medical professionals or delays in their diagnosis. Chest X-rays have been a valuable tool for identifying COVID in a patient and tracking its progression along with other techniques. However, due to the large number of patients and by extension, chest x-rays, the healthcare professionals are facing a real problem. Therefore, any technique that can help in early diagnosis and reduces effort of medical professionals can essentially be lifesaving. Recently, many researchers have tried to help medical professionals by using advanced deep learning techniques such as Convolutional Neural Network for automatic diagnosis from chest X-rays. In this research, use of multiple advance deep learning like Yolact and Yolact++ to segment and localize anatomical structures in chest x-ray image is explored. This would help medical professionals to look at different anatomical structures independently and this would reduce their effort and time consumption in diagnosis. Furthermore, isolation of these anatomical structures can also help train other specific deep networks for diagnosis of diseases as correct localization will help reduce the noise in the images.

Key Words: *Chest X-Rays, Yolact, Yolact++, Segmentation, Localization*

Table of Contents

Chapter 1 Introduction

According to a report by WHO [1] , Chronic obstructive pulmonary disease (COPD) affects around 65 million people globally, with 3 million dying each year, making it the third greatest cause of mortality. Asthma affects around 334 million people worldwide, and it is the most prevalent chronic illness of childhood, impacting 14% of all children. Pneumonia kills millions of people every year and is the top cause of mortality in children under the age of five. Tuberculosis (TB) affects around 10 million people worldwide each year, with 1.4 million deaths, making it the most prevalent deadly infectious illness. Lung cancer is the most lethal cancer, killing 1.6 million people each year. Chronic respiratory illness claims the lives of 4 million individuals worldwide. Also, according to a report by WHO [2], cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, claiming the lives of an estimated 17.9 million people per year. So in totality, heart and circulatory disorders and respiratory disorders [1] [2] [3] cause the most number of deaths around the globe. Whenever there is any issue in upper thoracic region, be it cardiovascular or respiratory, chest x-rays are first and foremost procedures to be done. Chest x-rays vast usage lies in the fact that it is cheap, non-invasive, easily available and easily doable. Globally, each year, there are 2 billion chest x-rays being done [4]. While x-rays are quite common, trained and qualified practitioners that can understand those x-rays are very scarce.

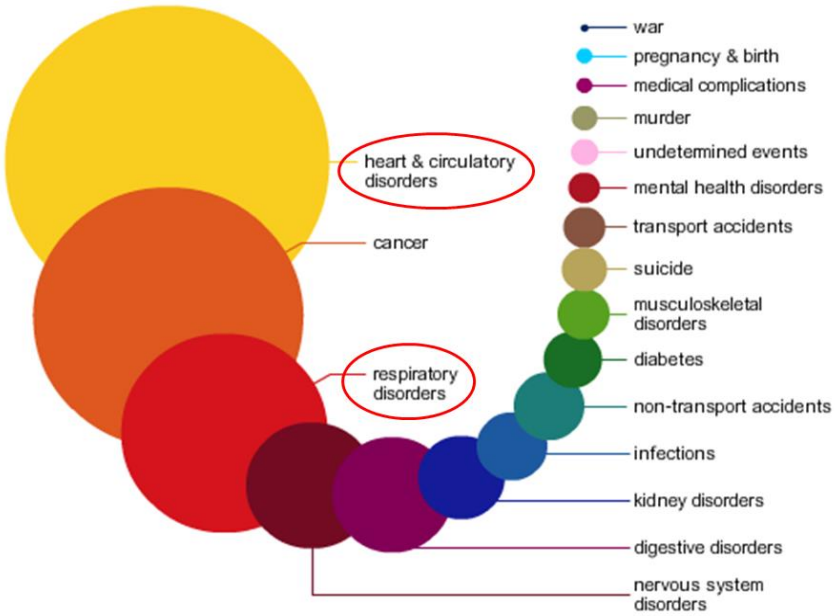

Figure 1-1 WHO Data for Cause of Deaths

The world as we know it is being changed thanks to the advances in the field of deep learning and AI as fictitious things like autonomous cars come into the market. We see a sudden increase in the use of AI in all fields to be a part of these advances and incorporate these recent advances for the betterment of their respective system. In all sectors that Artificial Intelligent Networks are used, you can see a steady boom as they bring results which are achievable by human intervention. Nowadays, health care is one of these sectors where Deep Learning techniques are being used to help medical decision support systems vastly and modern computing technological leaps are being used to help in the advancement of the medical sciences. Hence, computer vision techniques such as segmentation and localization of anatomical structures of chest X-Rays can play pivotal role in implementation of automated and intelligent medical decision support systems. Also, extracted anatomical structures can be used as an input in other deep learning problems involving those structure i.e. that of a heart, a lung, or others. This research compares different modern deep learning techniques such as Yolact and Yolact++ for the segmentation and localization of anatomical structures in the thoracic region using chest x-ray.

1.1 Motivation

In recent times, with the world facing pandemics like COVID-19 and having a significant increase in population, medical professionals and systems have been under immense pressure and it has been a challenge for them to tend to patients around the world. The hospitals have groaned under the immense numbers of patients around the globe. Medical staff have been overworked and there is a shortage of personnel in the staff of hospitals around the globe. This fact is highlighted by the studies carried out in Rosman, D. A. et al. [5] and Ali et al. [6] where it was shown that as of 2015, Rawanda had only 11 radiologists for a population of over 12 million people of Rwanda whereas Liberia having a population of four million had only two practicing radiologists in the same timeline. Similarly, Pakistan, like other third world countries, suffers from shortage of medical professionals. In 2019 it was estimated that approximately 562,000 individuals developed TB in Pakistan but only 360,000 were notified and diagnosed [7] whereas, 202,000 people were not diagnosed due to the low doctor to patient ratio (which was staggering 1:1300 in Pakistan as of 2019). This leads to large number of un-diagnosed cases resulting in general population causing high mortality rate.

So, in this scenario, any automated medical decision support system can be really crucial for timely diagnosis and provide a wider coverage as it is not limited to the knowledge of the

current doctor. In addition, this system can also have worldwide coverage so that the lives of millions of people can be catered to with the help of an automated medical decision support system. The job of the doctor is to interpret the X-rays of patients to understand what type of symptoms or discrepancy has made them unwell. The machine itself is not intuitive enough to understand the symptoms of a patient just by looking at the image of the x-ray, therefore, segmentation and localization of anatomical structures in chest x-rays such as the heart, lungs can help medical decision support systems identify the disease and facilitate it and is the motivation behind this research.

1.2 Problem Statement

This research is based on improving the accuracy of medical decision support systems in the field by making accurate models which will be able to work in a real-time environment. In this paper we look into, improving the segmentation and localization of anatomical structures in chest x-rays can really help medical decision support systems and train other specific deep networks for diagnosis of diseases as correct segmentation and localization will help reduce the noise in the images. The purpose of this research is to explore deep learning techniques such as Yolact and Yolact++ in order to improve the segmentation and localization of anatomical structures in chest x-rays. We believe that applying these models which have shown incredible results in the field of Instance segmentation in the field of computer vision will allow us to perform medical decisions in a real-time environment while also being able to provide accuracy in prediction on par with licensed doctors.

1.3 Aims and Objectives

The major objectives of the research are as follow:

- Segmentation of different anatomical structures present in a CXRs need attention for more understanding of a CXR
- Localization along with segmentation is not used for CXR
- Advance deep learning-based models have not been utilized for localization and segmentation of CXR structures

1.4 Structure of Thesis

This dissertation work is structured as follows:

Chapter 2 covers the importance of chest x-rays in medical science and how fast and accurate diagnosis can be ensured using segmented and localized anatomical structures present in chest cx--rays.

Chapter 3 gives a review of the literature and the significant work done by researchers in the past few years for the segmentation and localization of anatomical structures in chest x-ray images.

Chapter 4 consists of the proposed methodology in detail. It includes two main modules: using Yolact and Yolact++ for segmentation and localization.

Chapter 5 introduces the databases used for evaluation purposes. All the experimental results are discussed in detail with all desired figures and tables.

Chapter 6 concludes the thesis and reveals the future scope of this research

Chapter 2 Importance of Segmentation & Localization of Anatomical Features of Chest X-Rays

It is standard practice in the medical sector to diagnose any problems a person may have in the thoracic region with x-rays. The reason that x-rays are such a standard practice for an initial diagnosis is that it provides a clear pictorial view on the inside of the patient while also being a non-invasive procedure and giving negligible radiation exposure to the patient. Moreover, it is also the preferred method by doctors as it provides the doctor with a detailed look into the patient's symptoms and it also gives a more standardized approach as the tolerance of different people can be different. Afterward, the physicians take a look at each anatomical structure like lungs, heart, ribcage, clavicle in the x-ray for diagnosis/analysis and then look into solving the patient's illness. This chapter will briefly cover the basics of chest x-ray, diagnoses that can be done from segmented anatomical structures present in the chest, and its use cases.

2.1 Chest X-Rays

In the modern day medical science, chest x-rays are the most commonly used x-rays around the globe. A chest x-ray produces pictures of the heart, lungs, airways, blood vessels, spine, and chest bones, which is why they are so common. An x-ray examination assists doctors in the diagnosis and treatment of medical illnesses by providing a crisper image and giving the doctor more information than the patient's symptoms. It creates photographs of the inside of your body using a little quantity of ionising radiation. The most popular and oldest form of medical imaging is X-rays.

A chest X-ray may show a lot about what's going on within your body, including:

- The situation of lungs
- The size of the heart.
- Fractures.
- Cancer

Because radiation is absorbed differently in various tissues, a chest X-ray produces a black-and-white picture of the organs in the chest. Radiation-blocking structures look white, whereas radiation-allowing structures seem black. Bones appear in light color as they are dense structures. The heart is also visible as a lighter area. Lungs show as darker spots on the photographs because they contain air and block relatively little radiation. A trained radiologist or a specialist doctor examines the images for diagnosis and makes a conclusion on the patient's

condition based on his medical knowledge as to the discrepancies of the current x-ray with that of a healthy person's x-ray and the disease that caused the discrepancy.

2.2 Lungs

On either side of the chest are two spongy, air-filled organs called lungs (thorax). Through tubular branches known as bronchi, the trachea (windpipe) carries breathed air to the lungs. The bronchi become tiny when they divide into smaller and smaller branches (bronchioles). Bronchioles finally end in alveoli, which are clusters of microscopic air sacs. The alveoli take oxygen from the atmosphere and transport it to the bloodstream. The metabolic waste product carbon dioxide goes from the bloodstream to the alveoli, where it is released. Between the alveoli, there exists a thin layer by the name of interstitium which contains cells and blood vessels that support the alveoli. A thin tissue layer that covers the lungs is known as the pleura. The inside of the chest cavity is lined with the same thin tissue known as pleura. With each breath, a small coating of fluid works as a lubricant, helping the lungs to expand and collapse easily.

As the lungs are such an invaluable organ, we would like them to stay in top-notch condition. In case a patient comes into the doctor with symptoms that affect the lungs (e.g., they are having trouble breathing) the physician would normally advise them to do a chest X-ray can be used to diagnose pulmonary abnormalities such excess fluid (pulmonary edoema), fluid surrounding the lung (pleural effusion), pneumonia, bronchitis, asthma, cysts, and malignancies and the x-ray would then be able to help the doctor identify the source of the patient's discomfort using the discrepancy in the x-ray. Moreover, Chest X-rays are so detailed and medical science has evolved so far that medical professionals can identify and correlate problems such as i.e., alterations or issues in the lungs as a result of cardiac difficulties Congestive heart failure, for example, can cause fluid in the lungs.

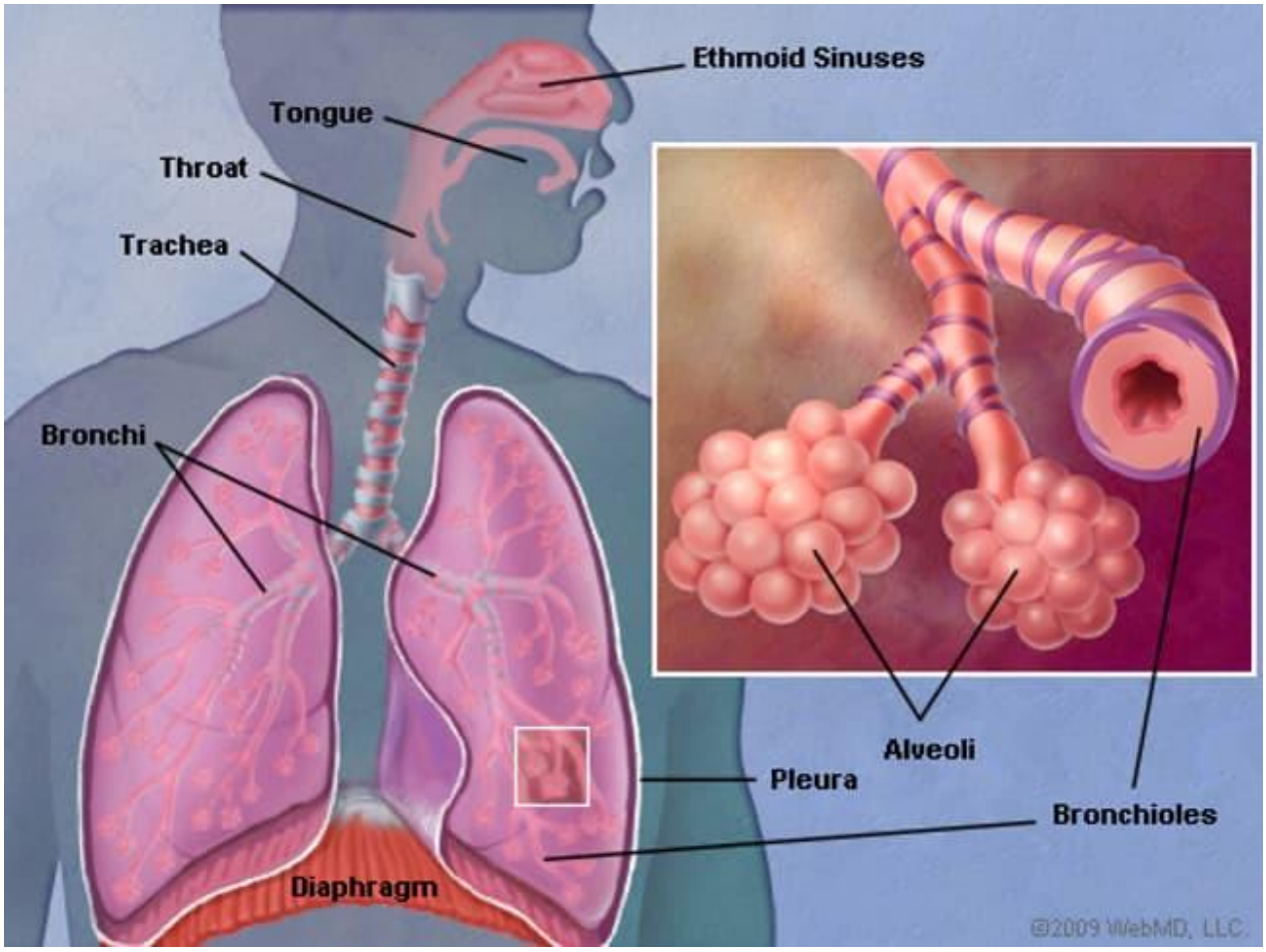

Figure 2-1 Detailed diagram of lungs

2.3 Heart

The heart is a muscular organ about the size of a fist that sits directly behind and to the left of the breastbone. The cardiovascular system, which consists of a network of arteries and veins, circulates blood through heart. It's solely responsible for transporting the blood in the body so proving oxygen or any nutrients that the body might need and also carrying waste like carbon dioxide out of the body.

In matters of the heart, a doctor may prescribe a chest x-ray to identify the discrepancy in the patient's heart that is responsible for causing the problem. For example, heart changes in size and form might suggest heart failure, fluid surrounding the heart, or difficulties with the heart valves. Because the outlines of the big vessels around the heart, such as the aorta, pulmonary arteries, and veins, may be seen on X-rays, it can also reveal aortic aneurysms, other blood vessel issues, or congenital heart disease. Calcium in the heart or blood arteries can also be detected using chest X-rays. Its presence might suggest the presence of lipids and other

substances in your vessels, as well as damage to your heart valves, coronary arteries, heart muscle, and the protective sac that surrounds the heart. So as you can see the chest x-rays allow the doctor to get a clear idea of the discrepancy taking place in the person heart leading to their systems and then asking the patient to go on a treatment plan based on what they seem medically appropriate.

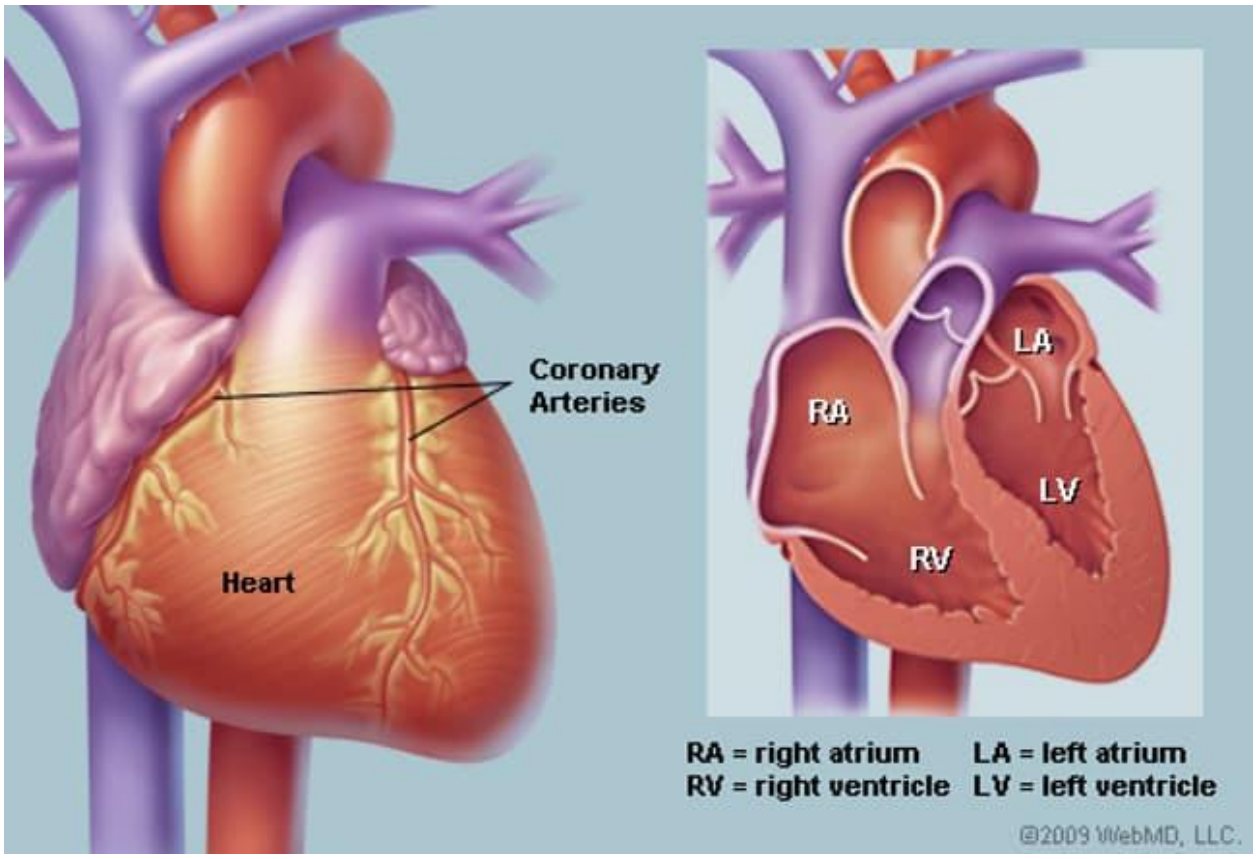

Figure 2-2 Detailed diagram of Heart

2.4 Bones

Chest x-ray contains ribcage and clavicle bones. The rib cage is one of the body's best defenses against impact injury. The ribcage is made up of the ribs and the sternum. The rib cage shields vital organs like the heart, lungs, and liver from injury. There are 24 ribs in the rib cage. Each rib grows out of the spinal cord and wraps around the body. The sternum is a long flat bone in the middle of the chest that is also known as the breastbone. Its purpose is to protect the heart. The first seven ribs, as well as the clavicle, or collarbone, are connected by the sternum. The clavicle, or also known as the collarbone, extends across the front of the shoulder from the sternum to the scapula, or shoulder blade.

In case of damage of bones (fractures), doctors prescribe the patient an x-ray to get an idea of the damage the magnitude of the damage the person has suffered. The way this Is done is that normally x-ray radiations are not able to pass through the bones as they are so din caseut in case of a fracture of the bone the racked so we are able to the breaking point as the radiation will pass through that point and we will have a black point rather then the expected white point of a healthy bone.

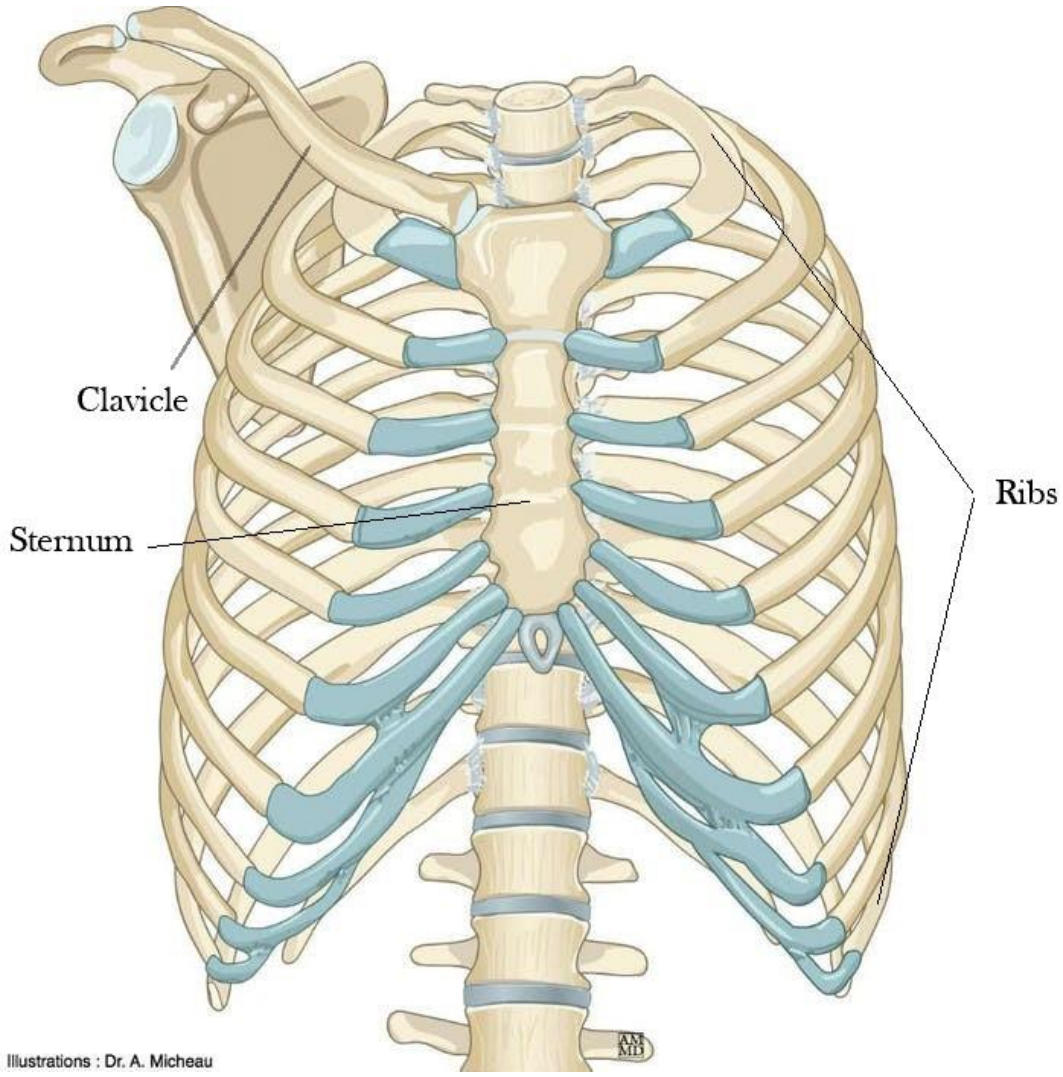

Figure 2-3 Detailed diagram of bones in CXR

2.5 Use-Cases

In modern times, with advancements in deep learning, automated medical decision systems are being developed and deployed to help in the advancements of the medical sciences. There are numerous researches being carried out the increase the efficiency and time complexity of health

systems throughout the world. The reason for this solution is to aid in the development and improvement of the medical world and to be one of the driving forces on the brink of changing and advancing this field.

One of the most fundamental reasons for the success of these applications is segmentation, which is the process of splitting a digital image into several parts (sets of pixels, also known as image objects). The purpose of picture segmentation is to make an image's representation more meaningful and simpler to examine by simplifying and/or changing it. Image segmentation is commonly used to find objects and boundaries (lines, curves, and so on) in images. Image segmentation is the process of labeling every pixel in an image so that pixels with the same label share certain characteristics. Whereas localization is another deep learning method that gives us a bounding box around objects in an image. Its output is x coordinate, y coordinate, height, and width of the bounding box of the object. So, segmentation and localization can help us identify objects in images. Now in our thesis, we are using both of these algorithms to extract lungs, heart, and clavicle from chest x-ray images.

Chest x-rays are one of the most abundantly used diagnosis methods or rather the most initial diagnosis method for any issues in the thoracic region by the physicians. With the emergence of COVID-19, chest x-rays were commonly used for checking conditions of lungs for COVID-19 to get an in-depth analysis of positive patients as well as people who were considered false negatives and had uneasiness with their lungs. CT Scans and COVID-19 tests had time constraints which caused problems as there was a delay in providing patients with the right care and also the classification of whether the person had tested positive for COVID-19 also took time, while x-rays can be obtained in real-time, so this increase in usage increased the demand for more professionals to assess these x-rays. The recent pandemic has shown us that the medical sector is already short-staffed as during COVID-19 the world’s top and foremost countries in the field of medical science were also seen having issues with catering to patients during the pandemic due to the large numbers. The idea behind this paper being that since automation has had such a major role in so many industries bringing it to the medical sector will allow us to get increased coverage while also providing patients with timely care as a fully automated such as this will be able to cut the turnaround time from a few hours to a few seconds. So, any system that could automate this assessment of x-ray automatically can really be beneficial for health care systems as well as patients as it can decrease turnaround reporting time to a few seconds from a few hours. This can in a result be very vital for the greater good

of people and robust working systems can increase the lifesaving capacity of the health care system while also alleviating the doctors from tedious activities to more important ones. Also, these extracted anatomical structures can be given to other deep learning-based automated medical decision support systems that can do the diagnosis like if extracted lungs can be given to any deep learning algorithm to extract the damage done to them, its workload becomes very less and it can work very robustly, same goes for heart etc.

So, this robust method of applying computer vision techniques like image segmentation and image localization of anatomical structures in chest x-rays can indeed help the medical sector. The deep learning field can automate the tedious and repetitive tasks for medical professionals freeing up more time for them to devote to the betterment of their patients and the development of the medical sciences, this is very important as it will intern will lead to the preservation of the human species

Chapter 3 Literature Review

Chest x-rays have been very integral part of medical imagining and there has been a lot of research being done on them. In recent times, various imaging solutions are being utilized but chest x-rays are still the first and basic step in getting any diagnosis in the thoracic region. In the past two years, many researchers have worked on segmentation and localization of anatomical structures in chest x-rays. Segmentation and localization of anatomical structures in chest x-rays is difficult because of ribs being super imposed on lungs and low intensity of heart etc, so it has been quiet a trivial research topic in recent years. This chapter summarizes all those valuable researches.

3.1 Segmentation

Joseph et al. [8] depicted an entirely automatic way of segmentation of X-ray images on a limited data set. It showed an architecture of 2 encoder decoder modules instead of a single deep architecture in order to image pixilation and combat overfitting. Data set used for training and scoring the model was of IBEX Innovations Ltd. The algorithm produced results of 92% accuracy, 92% f1-score and 98% AUC which exceeded the already existing traditional implementations.

Souza et al. [9] tries to automatically segment lungs including the opacities caused by pneumonia and pulmonary tuberculosis etc. The opacities reach high intensity making it difficult to segment automatically in a single region. The algorithm involves using 2 CNNs, one for initial segmentation of lungs and second for reconstruction of missing lungs portions and final output is the combination of both CNNs. For initial segmentation, AlexNet was used while for reconstruction ResNet18 was used. Montgomery County Chest X-ray dataset was used, which contained 138 CXR exams from normal patients but also from several patients with manifestations of tuberculosis, so it was quiet heterogeneous dataset. Results of this research showed promise in the quality of the segmentation, but the dataset was limited and there is still a chance of getting false positive high intensity regions as a result.

In this paper Zhang et al. [10], task-driven generative adversarial network model was used to investigate the unsupervised multi-organ segmentation on X-ray images. The proposed model framework accepts synthetic labelled DRR images produced from 3d CT as input and produces meaningful segmentation results on real Xray images without the use of ground truth annotations. It uses a cycle-GAN substructure to achieve image style transfer and carefully

designed add-on modules to segment organs of interest at the same time. 153 topograms were used for training and it was tested on 500 random annotated images from NIH dataset qualitatively and on 300 random annotated images from JSRT dataset quantitatively. It has been able provide 86% dice with unsupervised learning which was pretty similar to 89% dice score of supervised learning.

Maga et al. [11] showed a combination of CNN and Adversarial Critic Model for segmenting lungs in chest X-rays. The CNN identifies different masks among the X-ray image while the Critic Model determines whether the X-ray contains Lungs or not? The data sets used for training and scoring the model contained Japan Society of Radiological Technology (JSRT), Montgomery and Shenzhen data sets. And the final Dice Score of 0.975 was achieved.

Rehman et al. [12] described a conjunction of U-net and 9 different Deep CNNs including ResNet(18,50,101), MobileNet,ChexNet,InceptionV3,SqueezeNEt,DenseNet201 and Vgg19 for initially segmenting the lungs and then classifying whether it's a positive/negative case for TB. The datasets were gathered from various online sources including Kaggle, A. Rosenthal et al. The TB Portals and B. P. Health. (2020). Belarus Tuberculosis Portal. For training the Deep CNNs, transfer learning was used to initialize weights from pre-trained models for classification. After experimentation, the algorithm with optimal scores for classification (without segmentation) was found to be ChexNet having accuracy, sensitivity, precision, specificity and f1-score of 0.965,0.965,0.966,0.965 and 0.965 respectively. The algorithm for classification (with segmentation) produced best results with DenseNet201 of 0.986, 0.986, 0.986, 0.985 & 0.986 for accuracy, sensitivity, precision, specificity and f1-score respectively. In addition, the paper used Score-CAM heat map visualization to confirm the detection of TB.

In Abdel-Basset et al. [13], they tried extracting the similar small regions of the chest image in attempt to extract the important region called image segmentation problem (ISP), which is highly useful for detecting COVID-19 by using x-ray images. The authors discovered a new hybrid multi thresholding approach based on e SMA behavior with WOA for overcoming ISP. Its effectiveness is better than five state of the art algorithms like WOA, SSA, Lshade, HHA, and FFA. The authors used number of images from The Berkeley Segmentation Dataset and Benchmark as dataset.

In Pereira et al. [14], lungs segmentations was used on chest X-ray images which was further effectively used to detect COIVD-19. They used U-Net CNN architecture, and classified using three CNN architectures (VGG, ResNet, and Inception). They composed three class dataset naming Pneumonia, COVID and Normal human. They also used dataset from different sources for their experimentation. With a Jaccard distance of 0.034 and a Dice value of 0.982, the

segmentation was successful. For the multi-class configuration, the classification using segmented pictures achieved an F1-Score of 0.88, and for COVID-19 identification, an F1-Score of 0.83. They got an F1-Score of 0.74 and an area under the ROC curve of 0.9 for COVID-19 identification utilizing segmented pictures in the cross-dataset scenario. The models were trained using proprietary databases including photos from four Chinese hospitals and then evaluated on the open-access CC-CCII dataset, which is publicly available.

3.2 Localization

Wu et al. [15] uses an end to end automated approach of firstly segmenting the 6 zones of lungs (upper, lower and middle of left/right lung) using U-Net with the help of bounding boxes. The anomalous region was labelled with a different color. Subsequently, these labelled images are passed into RetinaNet for classification using lung opacity. The dataset of ChexRay-14 NIH was used for training and validation purposes and the average recall and precision of 81.8% and 89.6% were achieved respectively.

Kholiavchenko et al. [16] discusses an automated method for detecting pneumonia regions in the chest using an ensemble of RetinaNet & Mask-RNN. Both models are trained separately on training data set and weighted average is calculated for the final predictions. The data set used for training and validation were obtained from Kaggle Pneumonia Detection Challenge and the best score of 79.3% recall was achieved. In addition, the paper claims to have achieved a top 3% ranking in the competition.

Singh et al. [17] tried to predict Thorax disease by applying deep convolutional neural network using ImageNet on multiple X-ray images. The dataset used in this research was provided by NIH, it included 25,603 gray scale x-ray images of 1024x1024 pixels from 30,000 unique patients. They used two different models to analyze the x-ray. Firstly, the SoftMax regression and secondly DCNN architecture with account of metadata such as age, gender etc The solution got mean Accuracy of 69.75%.

Zebin et al. [18] tried to diagnose COVID-19 disease by chest X-ray images. For identifying COVID-19 chest X-ray pictures, they used a transfer learning pipeline. The classifier distinguished the images of lungs infected by COVID-19 and Pneumonia form the normal ones. They used multiple pre-trained convolutional backbones for feature extraction and achieved an overall detection accuracy of 90%, 94.3%, and 96.8% for the VGG16, ResNet50, and EfficientNetB0 backbones respectively. For visual explanation they used gradient class activation mapping technique to highlight the regions that are important for predictions. For

dataset they created three class datasets with titles normal patient, COVID-19 patients and patients with bacterial Pneumonia, the images were taken from "COVID-19 Image Data Collection" publicly available on GitHub.

Table 3-1Literature Review Summary

Year	Authors	Dataset	Methodology	Results
2019	Joseph et al. [8]	IBEX Innovations Ltd	Automatic segmentation using 2 encoder decoder modules containing multiple layers of Pooling, Upsampling, ReLU, Softmax Activation and Convolution	92% Accuracy, 92% F1-score, 98% AUC
2019	Souza et al. [9]	Montgomery County Chest X-ray	2 CNN, one for lung segmentation and other for reconstruction of opaque lung portions caused by TB and pneumonia	97% Accuracy, 97% Sensitivity, 96% Specificity, 93% Dice Index, 88% Jaccard score
2020	Zhang et al. [10]	JSRT & NIH	Accepts synthetic labelled DRR images produced from 3d CT as input and segments real X-ray images by using a cycle-GAN substructure	86% Dice Index
2020	Maga et al. [11]	JSRT, Montgomery and Shenzhen	CNN identifies different lungs masks among the X-ray image while the Critic Model determined whether the X-ray contained Lungs	97.5% Dice Index

2020	Rehman et al. [12]	Kaggle Chest X-Ray, NLM, Belarus, NIAID TB and RSNA	Segmentation of lungs and then detection of TB using U-Net and Deep CNN	98.6% Accuracy, 98.6% Sensitivity, 98.6% Precision, 98.5 % Specificity, 98.6% F1-Score
2020	Abdel-Basset et al. [13]	The Berkeley Segmentation Dataset and Benchmark	Hybrid multi thresholding approach based on e SMA behavior with WOA for overcoming Segmentation of lungs.	Its effectiveness is better than WOA, SSA, Lshade, HHA, and FFA
2021	Pereira et al. [14]	Private	U-Net for Lung Segmentation & 3 CNN for classification as COVID-19 effectee	74% F1-Score, Area under the ROC curve of 0.9
2020	Wu et al. [15]	ChexRay-14 NIH	Segmentation of lungs using U-Net then classification using RetinaNet	81.8% Average Recall, 89.6% Precision
2019	Kholiavchenko et al. [16]	Kaggle Pneumonia Detection Challenge	RetinaNet and Mask R-CNN were trained separately and weighted average was calculated for final predictions	75.8% Precision, Recall 79.3%, 77.5% F1-score
2020	Singh et al. [17]	NIH	Deep Convolutional Neural Network unified framework and disease localization using ImageNet	69.75% Accuracy
2021	Zebin et al. [18]	Kaggle CXR Pneumonia dataset, Github	Convolution network with 3 backbones i.e. VGG-16,	Accuracy of 90%, 94.3%, and 96.8% for the

		COVID-19 Image Data Collection dataset	ResNet50 and EfficientNetB0	VGG16, ResNet50, and EfficientNetB0

As we conducted the above literature review, it dawned upon us that segmentation and mask prediction at pixel level of different anatomical structures present in Chest X-Rays need attention and more work to generate better models that understand Chest X-Ray images and anatomical structures present in it. Also, no research has ever used localization with segmentation for Chest X-Rays, so it's either segmentation or localization, but combination of both have never happened as previously there were no sophisticated advanced deep learning models present that had both of these techniques. But, nowadays, there are certain advance deep learning based models which are using localization and segmentation simultaneously but those algorithms are not being used for segmentation and localization of anatomical structures in Chest X-Rays till date, so that posed as a research gap for our thesis research.

Chapter 4 Proposed Framework

This chapter will explain the framework which has been proposed for the segmentation and localization of anatomical structures in chest x-rays. First we, will cover the dataset used, then an explanation of Yolact followed by an explanation of Yolact++ and eventually our own implementation.

4.1 Dataset

We used a publicly available dataset, the Japanese Society of Radiological Technology (JSRT), for our research. JSRT dataset contains 247 CXRs gathered from 14 medical centers. There are 93 normal images and 154 abnormal images among the 247 images. These images are saved in PNG format with 2048x2048 pixels and a grayscale of 12 bits. Image abnormality is graded from extremely subtle to obvious.

Table 4-1 Dataset Summary

Name	Japanese Society of Radiological Technology (JSRT)
Number of Images	247
Resolution	2048x2048
Anatomical Structures	Heart, Left Lung, Right Lung, Left Clavicle, Right Clavicle

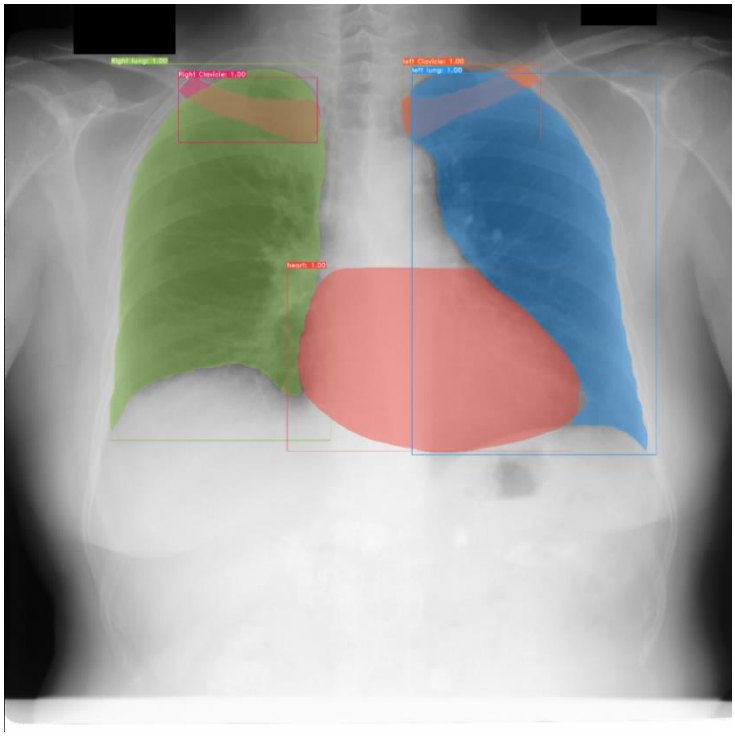

Figure 4-1 Annotated image of JSRT

4.2 Methodology

The methodology we followed has been shown in Figure 4.2. We took the annotated JSRT chest x-ray, parsed the JSRT annotations to COCO style JSON. We then passed the chest x-rays, as well as COCO style annotations JSON file to our deep learning models which were Yolact and Yolact++ to train. Our models then gave us the localized and segmented the chest x-rays from test split of dataset and gave us masks and bounding boxes for all the classes i.e. Heart, Right Lung, Left Lung, Right Clavicle, Left Clavicle.

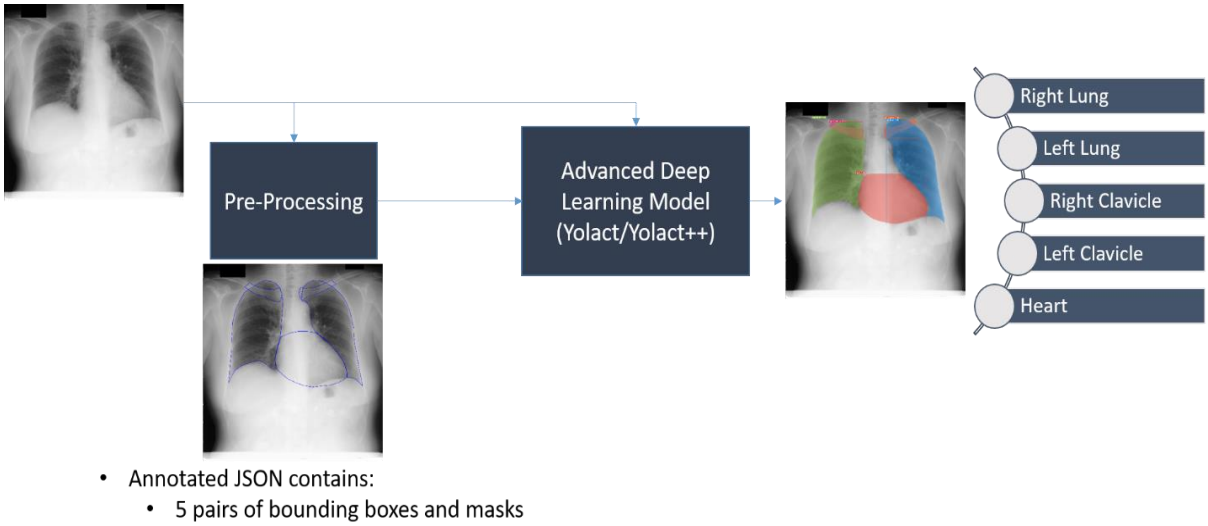

Figure 4-2 Proposed Methodology Summary

4.2.1 Preprocessing

We took the JSRT public dataset which came with chest x-rays and their respective annotation files. These JSRT annotations weren't COCO style JSON annotation files. So we wrote a parser which took JSRT annotation files and returned us a JSON which was similar to JSON's being given by Labelme utility as an output. Labelme utility was advised by the GitHub repository of Yolact and Yolact++ for annotation. After parsing each JSRT annotation we got a JSON file that contained all the relevant information regarding annotations and the image. Afterwards we created a text file that contained all the class labels i.e. left lung, right lung, heart, left clavicle, right clavicle etc. Finally we used instance segmentation utility by the name of labelme2coco to convert labelme styled annotated dataset to COCO style annotated dataset. Output of this utility was our prepared dataset.

We used the dataset prepared above to train the Yolact and Yolact++ models and evaluated them to get mAP scores for both the models.

4.2.2 Advanced Deep Learning Models

Image segmentation has always been seen as a computationally expensive step that cannot be achieved in a real-time environment. The problem is a mixture of both object detection and semantic segmentation as the problem tackles the prediction of object instances and their per-pixel segmentation masks, the final goal being to assign a class label to each pixel like that of semantic segmentation, the way it differs from semantic segmentation being that it treats multiple objects belonging to the same class as individual objects. In the same way that real-time object detectors (e.g., SSD and YOLO) proposed a one-stage architecture so that they could be used in a real-time environment unlike two-stage. So that's what make Yolact and Yolact++ so effective.

4.2.2.1 Yolact

YOLACT proposes to be a solution for this problem as it provides a solution similar to the way real-time object detectors like SSD and YOLO after faster than the pro R-CNN and R-FCN, through the removal of one of their stages of production so that they can work in real-time, and by making up for accuracy with other means. The YOLACT is similarly an alternative to the two-stage instance segmentation models that are readily available (e.g., Mask-RCNN) with a one-stage approach so as to allow us to deploy in a real-time environment. This solution for achieving image segmentation is not as simple as object detection, as we only had other models trying to attain object segmentation in the past (Straight to Shapes and Box2Pix) perform in a real-time environment and their accuracies were far from the desired percentage for modern-day models.

To address this problem, we have the YOLACT model which works by breaking up the image segmentation into two parts. It will generate a dictionary of prototype masks over the image. Secondly, it works by predicting a set of linear combination coefficients per instance. The above steps are the core process as all that remains is to perform the image segmentation and we do that by linearly combining the prototypes using the corresponding predicted coefficients from the model and then crop them using a bounding box towards the end to get the image segmentation. The Bounding Box in itself is a rough mask, which approximates the contour of the mask with the minimum bounding rectangle. This is very different from the other image segmentation model as the conventional way to perform segmentation is repooling or other

methods (e.g., mask voting), these steps put huge time constraints on the algorithm and the model is then not able to perform in a real-time environment. The YOLACT model because of its aforementioned structure is able to forgo these steps as it simply relies on linear combinations. Moreover, the YOLACT model is able to do the generation of prototype masks and predict per-instance masks coefficients in parallel which allows it to do the tasks almost seamlessly.

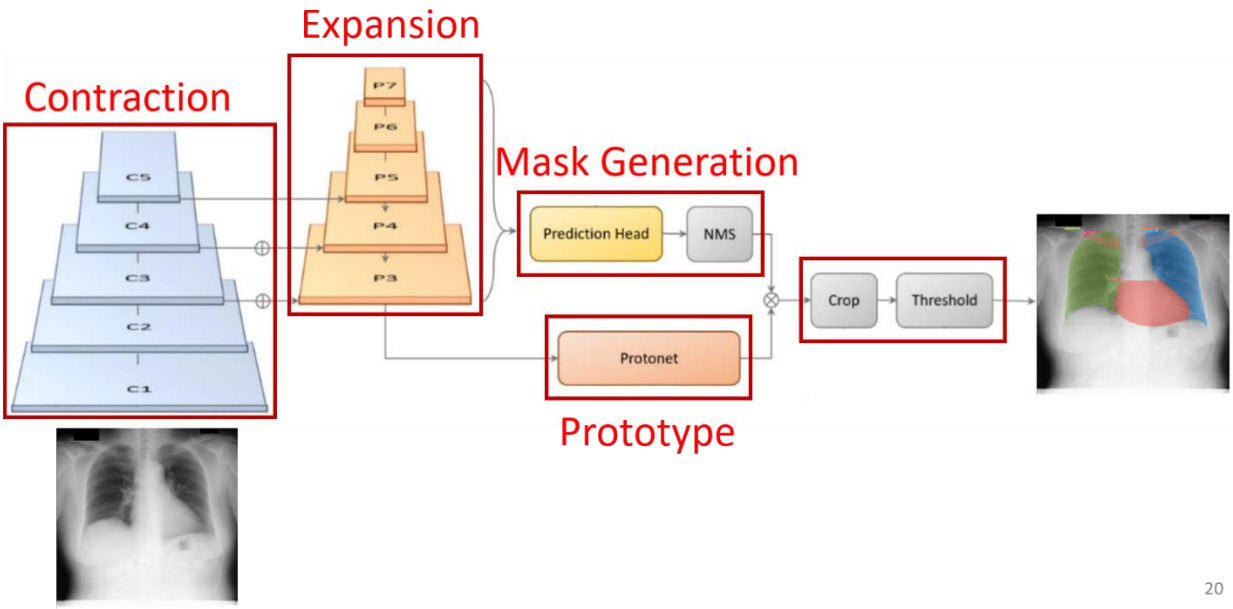

Figure 4-3 Architectural View o Yolact

As shown in the architecture diagram Figure 4.3. The YOLACT model we have used a RESNET-101 along with the FCN as the default feature backbone. The next step is done as a parallel process so as to increase the efficiency of the algorithm by decreasing the time latency. The first branch involves that the prototype masks are generated with the help of the FCN and in the second branch it performs detection and predicts class detection and mask coefficient for each of the instances.

According to the following design decisions, the prototype generating branch (protonet) forecasts a set of k prototype masks for the full image:

1. Using protonet from deeper backbone characteristics that provide strong and high-quality masks, the final layer with k channels from FPN P3 is evaluated, and it is up sampled to one-fourth the input image's size to improve performance on small objects.

2. Individual prototype losses are not expressly considered; instead, the total mask loss after assembly is evaluated.

3. A relu (non linearity) operation is done on the protonet's output to maintain it unbound, allowing the network to generate huge, overwhelming activations on prototypes it is highly certain about for the background.

Following Figure 4.4 shows a depiction of Protonet architecture. It shows simple upsampling followed by 3x3 convolution until the last stage where the convolution is 1x1.

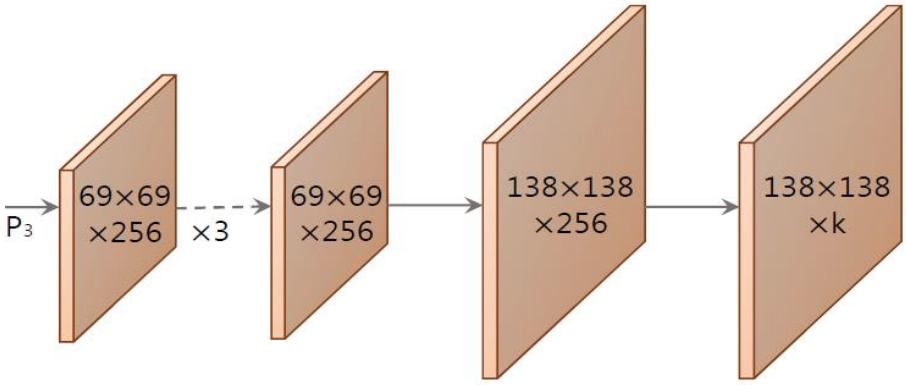

Figure 4-4 Protonet incorporated in Yolact

In the mask generating branch, we got 2 basic components working, one being prediction head and other being fast NMS.

The prediction head in depicted in Figure 4.5, here, the input is passed through a 3x3 FCN twice. Afterwards, data is passed through a class coefficient branch, a box coefficient branch and a mask coefficient branch simultaneously. This is for c classes, a anchors for feature layer Pi and k prototypes. Its output is fed to Fast NMS. These 3 heads have separate loses:

- Classification loss: Log based as shown in figure [19]

$$L_{conf}(x,c) = -\sum_{i\in Pos}^{N} x_{ij}^{p} log(\hat{c}_i^p) - \sum_{i\in Neg} log(\hat{c}_i^0) \quad \text{where} \quad \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$$

- Box regression loss: Combination of different factors such as Intersection over Union (IoU) and Generalized IoU (GIoU) as shown in figure [19]

$$L_{loc}(x,l,g) = \sum_{i\in Pos}^{N} \sum_{m\in\{cx,cy,w,h\}} x_{ij}^{k} \text{smooth}_{\text{L1}}(l_i^m - \hat{g}_j^m)$$

$$\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx})/d_i^w \qquad \hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy})/d_i^h$$

$$\hat{g}_j^{w} = \log\left(\frac{g_j^w}{d_i^w}\right) \qquad \hat{g}_j^{h} = \log\left(\frac{g_j^h}{d_i^h}\right)$$

- Mask loss L mask: Binary Cross Entropy (BCE) which means how far are we from actual mask as shown in figure

$$M_{gt}\colon L_{mask} = \mathrm{BCE}(M, M_{gt}).$$

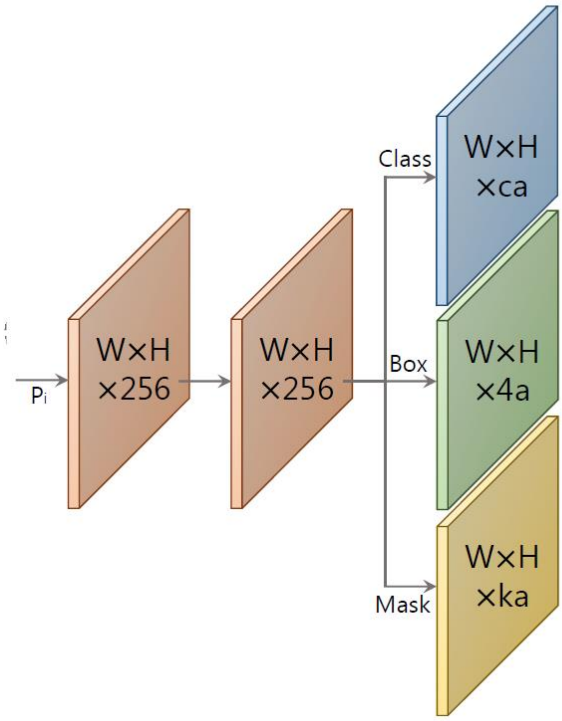

Figure 4-5 Different output heads in Yolact

NMS is used to reduce duplicate detections in most object detectors. The NMS procedure is carried out in the following order: sort the detected boxes descending by confidence for each of the c classes in the dataset, and then delete all those with lower confidence than it that have an IoU overlap higher than a threshold for each detection. Even though it is quick, it is a significant impediment to achieving 30 frames per second. To overcome the sequential nature of conventional NMS, Yolact uses the Fast NMS, in which each instance may be determined to be maintained or discarded in concurrently; to do so, we leverage previously deleted detections to suppress additional detections, which is not feasible in classic NMS.

These two parallel steps are combined by a simple linear combination of the generated prototypes and the mask coefficients. After which we apply sigmoid linearity to produce the final mask. We have 3 losses to train our model the classification loss, box regression loss, and mask loss. During the training period, we use do not crop the bounding box in the evaluation period to still get the small objects in the prototypes but in the final evaluation, we crop the

final masks with the bounding box and after thresholding, we are able to arrive at the high-quality masks for each object instances.

This gives us the benefit of having market standard performance from the model as its masks allow us to get the full extent of the image without the loss we normally get when repooling, coupled with the easy integration of generating prototypes and coefficients which is very helpful. In addition, the time constraint for this model is also less as it is comparable to a one-stage backbone detector because of its ability to be run in parallel threads and its simplistic assembly process. Moreover, because of its simple assembly structure can also be GPU-accelerated matrix-matrix multiplication making the model stand out in performing the image segmentation in a real-time environment.

4.2.2.2 Yolact++

The Yolact++ proposes an improved version of the YOLACT architecture. The change in the architecture helps us not only increase in accuracy but the compact size of the new architecture along with the new compact size allows it to run in low-compacity embedded systems while also performing in performance as the Frame Rate per second is also increased. The First increase in efficiency is caused by using Deformable Convolution Networks which cause an increase in the accuracy because of the DCN network makes up for the lack of not having a re-sampling process the use of DCN will align the sub-optimal sampling which have been obtained to a canonical reference region, also the DCN also gives us the added advantage of allowing the network to handle target instances with respect to their scale, rotation and aspect ratios. Secondly, the model employs a Fast Mask re-scoring network which consists of a 6-layer FCN with ReLU non-linearity per conv layer and a final global pooling layer. Finally, the Yolact++ boats an anchor-based backbone detector and tune the parameters to set the optimum value of aspects ratios and scales.

Chapter 5 EXPERIMENTATION AND RESULTS

This chapter will explain the results we got for our research. First, we will explain the metrics used to judge the output of the research like Intersection over Union, Average Precision, mean Average Precision and the then we will present these metrics for Yolact and Yolact++ for our own annotated JSRT dataset.

5.1 Performance Metrics:

Every deep learning algorithm's performance is evaluated by some fix metrics. To judge our algorithm, following are the performance metrics we would be using.

5.1.1 Intersection over Union (IOU):

IoU is a number between 0-1 that is defined by area of overlap between the predicted and ground truth bounding box divided by union of region between the predicted and ground truth bounding box. If IoU id 1, it indicates that there is a full overlap while lower values mean the predicted bounding box is away from the ground truth bounding box. It is used to calculate mean average precision.

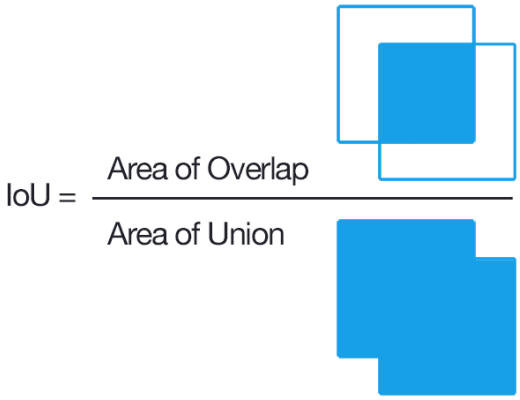

Figure 5-1 Intersection over union pictorial representation

5.1.2 Average Precision (AP):

AP is a metric used to measure accuracy of object detectors. Mathematically it is area under the curve for precision-recall curve plotted from 0 to 1.

Now, precision is a measure of how accurate our predictions are.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Or in simple terms,

$$Precision = \frac{True\ Positive}{Total\ Positive\ results}$$

Similarly, recall means how accurate did we predict all positives.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Or in simple terms,

$$Recall = \frac{True\ Positive}{Total\ Positive\ inputs}$$

So, AP is area under the curve of precision-recall graph.

$$AP = \int_0^1 p(r)dr$$

5.1.3 Mean Average Precision (mAP):

While evaluating COCO style datasets, mAP is used for measuring correctness of multi class detection. mAP is calculated over different IoU's, starting from 0.5 to 0.95. So, for each IoU, mean of AP for all classes with same IoU is calculated and final mAP contains percentage means of AP for all classes over mentioned above range of IoUs.

5.2 Experimentation

So, for experimentation, we first transformed our annotated JSRT dataset to COCO style annotation as Yolact and Yolact++ only take COCO styled annotations to train. So we parsed the annotations and created COCO style annotated JSON files with 90-10 split for training and validation. Afterwards, we trained Yolact and Yolact++ on these custom parsed JSON and got performance metrics explained above.

5.2.1 Yolact

So, we tested Yolact on our JSRT dataset which had 5 classes i.e. heart, left lung, right lung, left clavicle, right clavicle with a 90-10 training validation split. So we got average precisions for each class at different IoUs. Table 5.1 contains the AP's on different IoU's ranging from 0.5 to 0.95 with an interval of 0.05 for each class.

Table 5-1 APs on different IoU's of all classes for Yolact

Class	Box/Mask	0.5	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Heart	Box	1.0	1.0	1.0	1.0	1.0	1.0	0.94	0.63	0.065	0.01
	Mask	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.93	0.39	0.02
Left Lung	Box	1.0	1.0	1.0	1.0	1.0	1.0	0.90	0.81	0.41	0.0
	Mask	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.49
Right Lung	Box	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.92	0.31	0.0
	Mask	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.55
Left Clavicle	Box	1.0	1.0	1.0	1.0	0.83	0.70	0.62	0.34	0.13	0.0
	Mask	1.0	1.0	1.0	1.0	1.0	0.72	0.61	0.41	0.02	0.0
Right Clavicle	Box	1.0	1.0	1.0	1.0	1.0	1.0	0.73	0.47	0.05	0.01
	Mask	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.66	0.06	0.0

Also, the results gave us mean average precision (mAP) for IoU's ranging from 0.5 to 0.95 with an interval of 0.05 and a final average on all IoU's. Table 5.2 shows these mAP's.

Table 5-2 mAP for Yolact across different IoUs

	All	0.5	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Box	75.78	100	100	100	100	96.53	93.94	84.42	63.22	19.38	0.32
Mask	83.73	100	100	100	100	100	94.30	92.18	79.98	49.52	21.29

5.2.2 Yolact++

So, we tested Yolact++ on our JSRT dataset which had 5 classes i.e. heart, left lung, right lung, left clavicle, right clavicle with a 90-10 training validation split. So we got average precisions for each class at different IoUs. Table 5.3 contains the AP's on different IoU's ranging from 0.5 to 0.95 with an interval of 0.05 for each class.

Table 5-3 APs on different IoUs of all classes for Yolact++

Class	Box/Mask	0.5	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Heart	Box	1.0	1.0	1.0	1.0	1.0	1.0	0.87	0.38	0.05	0
	Mask	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.62	0.43	0
Left Lung	Box	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.92	0.29	0.04
	Mask	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.31
Right Lung	Box	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.31	0.03
	Mask	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.6
Left Clavicle	Box	1.0	1.0	1.0	1.0	0.82	0.69	0.69	0.61	0.23	0
	Mask	1.0	1.0	1.0	1.0	1.0	0.93	0.54	0.54	0.09	0
Right Clavicle	Box	1.0	1.0	1.0	1.0	1.0	0.94	0.94	0.31	0.06	0
	Mask	1.0	1.0	1.0	1.0	1.0	1.0	0.94	0.56	0.02	0

Also, the results gave us mean average precision (mAP) for IoU's ranging from 0.5 to 0.95 with an interval of 0.05 and a final average on all IoU's. Table 5.4 shows these mAP's.

Table 5-4 mAP for Yolact++ across different IoUs

	All	0.5	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Box	76.48	100	100	100	100	96.4	92.6	90.08	64.88	19.48	1.4
Mask	83.23	100	100	100	100	100	98.6	89.69	74.76	50.17	18.53

5.3 Validation of framework

We tested the Yolact++ models trained on JSRT dataset on 2 unseen datasets. The first dataset it was tested was called Montgomery County X-Ray set. It contained 138 chest x-ray images with a resolution of 4020x4892. Figure 5.2 contains inputs and results for this dataset.

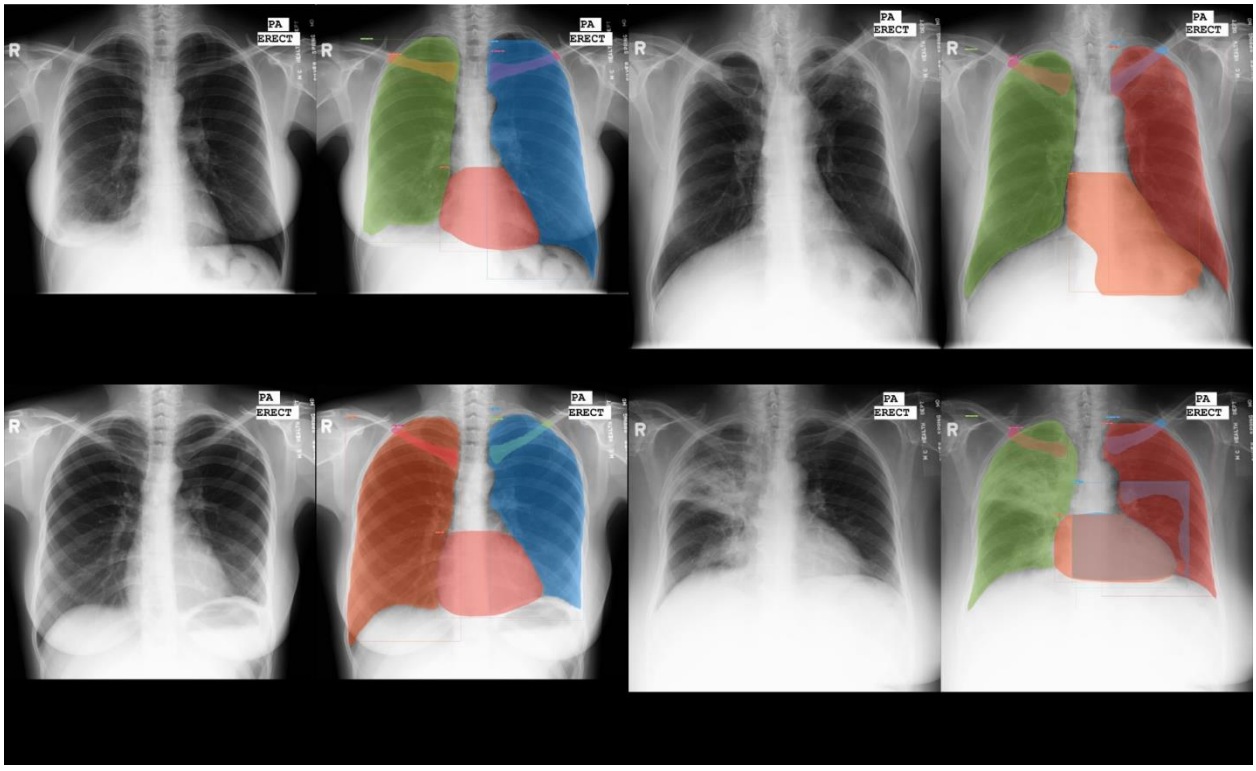

Figure 5-2 Input and predicted results on Montgomery dataset

Second dataset it was tested against was called Shenzhen Hospital X-ray Set. It contained 662 images with a resolution of 2987x2945. Figure 5.3 contains inputs and results for this dataset.

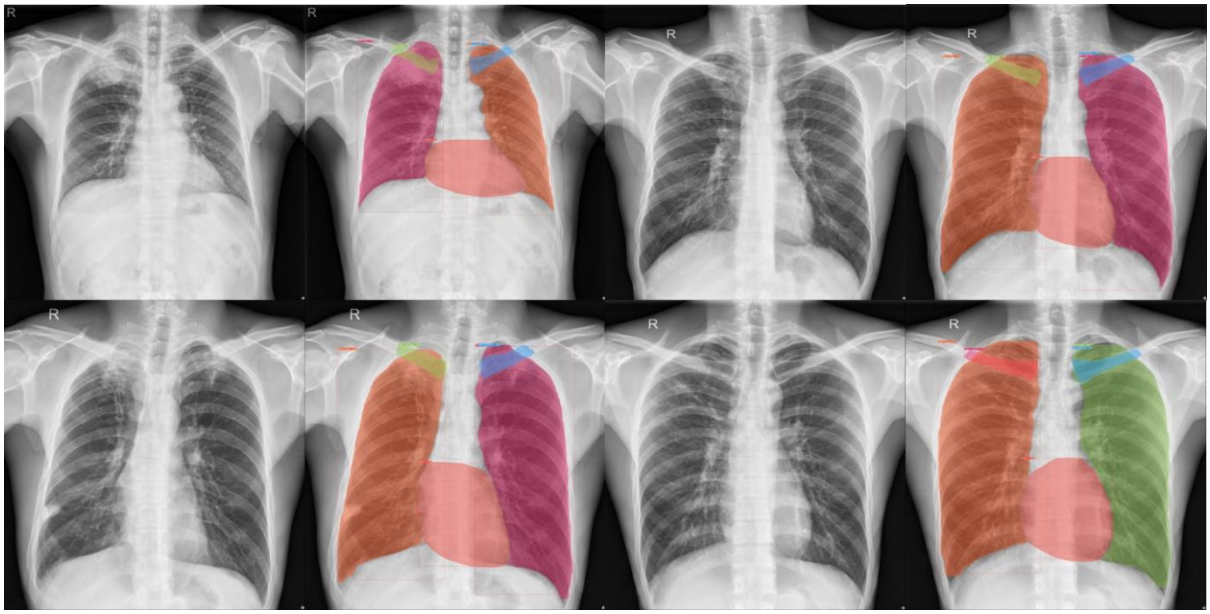

Figure 5-3 Input and predicted results on Shenzhen dataset

As we can see, the model was predicting all 5 classes pretty accurately. We do not have any numbers though as no other datasets except JSRT had all these 5 classes annotated.

5.4 Comparative Analysis

For comparative analysis, we trained and validated with same dataset on Yolo v5 and U-Net. For Yolo v5, we got the mAP's as well as predicted outputs. Table 5.5 contains mAP's of bounding boxes by Yolo v5 versus mAPs of masks and bounding boxes by Yolact++. Figure 5.4 shows labelled and predicted images by Yolo v5. As we can see in Table 5.5, Yolact++ is giving masks at a pretty similar average precision till 70 percent IoU which is a pretty high standard for any deep learning model while Yolo v5 is giving bounding boxes which are easier to find.

Table 5-5 mAP for Yolact++ vs Yolo v5

	IoU	50	55	60	65	70	75	80	85	90	95
Yolact++	Box	100	100	100	100	96.4	92.6	90.08	64.88	19.48	1.4
	Mask	100	100	100	100	100	98.6	89.69	74.76	50.71	18.53
Yolo v5	Box	100	100	100	100	100	100	100	99.8	96.02	77.31

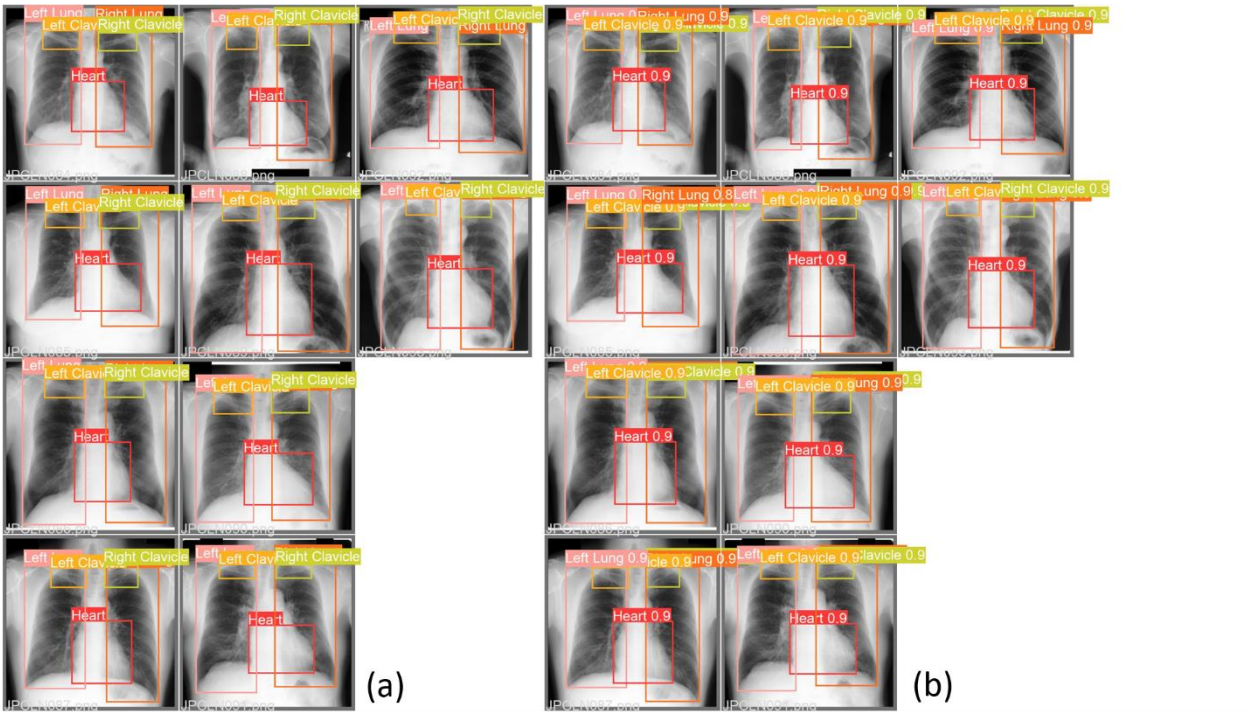

Figure 5-4 (a) Labelled Images (b) Predicted Labels by Yolo v5

For U-Net, we got average IoU's for each class but we have to run 2 models as Clavicles were intermingled and overlapping on Lungs so there was one model having Heart, Left Lung and Right Lung while the other model had Left Clavicle and Right Clavicle. Table 5.6 contains the mean IoU of each class while Figure 5.5 shows us the predicted masks on some samples. Now Table 5.6 shows a similar mean IoUs to Yolact++.

Table 5-6 Mean IoU for all classes by U-Net

Class	Heart	Left Lung	Right Lung	Left Clavicle	Right Clavicle
Mean IoU	0.81912156	0.939344	0.94960601	0.74123646	0.80185122

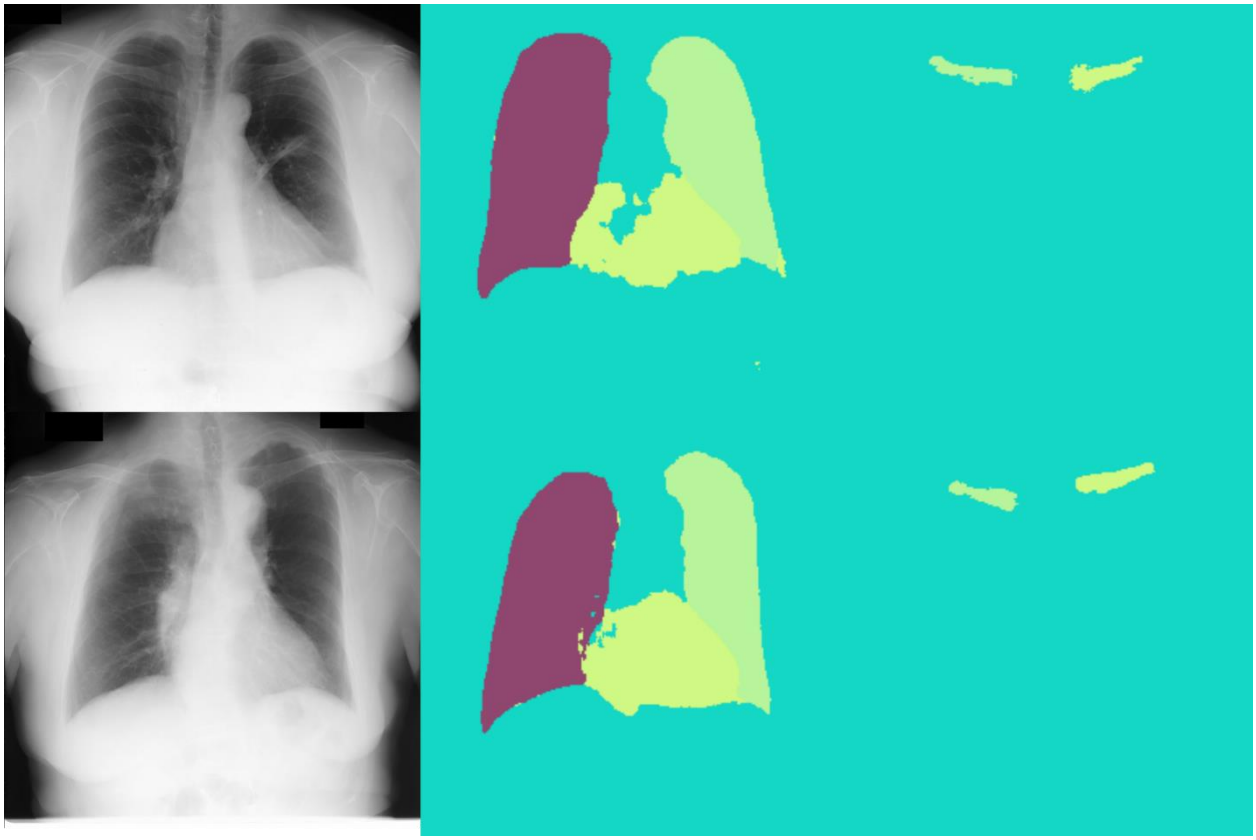

Figure 5-5 (a) Labelled Images (b) Predicted Labels by U-Net

5.5 Discussion

So, we tested Yolact and Yolact++ against our defined performance metrics and against most sophisticated localization and segmentation algorithms separately and one thing that stood out was both of these advanced deep learning algorithms were providing segmentation and localization with similar performance to those of specialized algorithms specified for singular task in less time too. So both of these became natural selection for segmentation and localization simultaneously. Among Yolact and Yolact++, Yolact++ had better results as mentioned above and it trained around 20-25 percent faster than Yolact. So, Yolact++ was doing better segmentation and localization faster than Yolact. Next chapter would conclude with the analysis and future work required to maximize the output of this effort.

Chapter 6 Conclusion & Future Work

6.1 Conclusion

Chest X-Rays are most common diagnostic step in any diagnosis of any ailments in upper thoracic regions as its non-invasive, cost-effective, feasible and accessible to most of the population around the globe. According to a WHO report, nearly 2 billion Chest X-Rays are conducted annually. So Chest X-Rays are accessible but trained radiologists and physicians having specific specializations to understand Chest X-Rays are scarce. According to Rosman, D. A. et al. [5], in 2005, 11 radiologists were serving 12 million of population of Rwanda and according to Ali et al. [6], 4 million people of Liberia had access to only 2 radiologists. So, there was an scarcity of skilled workforce to understand the Chest X-Rays before COVID-19 but with emergence of COVID-19, the situation has gone from bad to worse. So any automated diagnostic and/or decision support system which helps in understanding Chest X-Rays can be really beneficial for current medical systems and mankind. Having said this, all the initial upper thoracic diagnosis requires the medical practitioners to glance and study each anatomical structure present in Chest X-Rays which is a time consuming task. Current CAD and decision support systems have localization and segmentation done separately which is time and compute consuming. Yolact++ on the contrary can do both, localization and segmentation at a very low time and compute cost because it's a very advance deep learning architecture and it is pretty accurate with the masks as shown above. Although the model was trained with just 247 images, it was pretty effective when we tested it on unseen validation datasets such as Montgomery and Shenzhen. The reason for this was there were no class imbalance in the training dataset and Chest X-Rays always had a similar reference style i.e. minor tilts and zooming issues. Also, Yolact++ was compared with advanced localization technique such as Yolo v5 and segmentation technique such as U-Net and results were pretty similar to both. So it was doing what both of them were doing in less time and with more accuracy. Hence, Yolact++ is a compact solution that can be used for segmentation and localization of anatomical structures of Chest X-Rays.

6.2 Contribution

This research contributed towards the following aspects in advanced deep learning algorithms involving Chest X-Rays:

- Utilized one of the most recent architectures YOLACT++ for Chest X-Ray images

- The trained model is tested across different Chest X-Ray datasets to see the robustness of trained model
- Utilizes recent localization and segmentation architecture for anatomical structures present in Chest X-Rays

6.3 Future Work

The current works lays foundation for diagnostics systems which are dependent on segmentation and localization of anatomical structures in Chest X-Rays. In order for this research to yield an appropriate/useable system there are number of future directions which can be followed. Some of such directions are listed below:

- We can collect locally gathered Chest X-Ray dataset to train our model for an enhanced accurate result.
- Fused features from entire image and segmented lung region can enhance classification results.
- Further medical corpus can be incorporated in the current framework in order to improve report generation results.
- The framework can be adopted by various Decision Support Systems such as early detection of COVID19.
- The prototype framework can be extended to a commercial product if added to a medical decision support system or CAD systems.

References

[1] WHO, "The Global Impact of Respiratory Diseases," WHO, 2021.

[2] WHO, "Cardiovascular diseases," WHO, 21 01 2022. [Online]. Available: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1. [Accessed 21 01 2022].

[3] WHO, "WHO reveals leading causes of death and disability worldwide: 2000-2019," WHO, 9 December 2020. [Online]. Available: https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019. [Accessed 21 01 2022].

[4] S. F. D. S. A. R. S. I. L. a. R. E. C. Raoof, "Interpretation of plain chest roentgenogram.," *CHEST Journal, ,* vol. 141, pp. 545-558, 2012.

[5] D. A. Rosman, "Imaging in the land of 1000 hills: Rwanda radiology country report. ," *J Glob Radiol ,* vol. 1, p. 5, 2015.

[6] F. S. H. S. G. K. S. B. &. H. S. Ali, " Diagnostic radiology in Liberia: a country report.," *J Glob Radiol,* vol. 1, p. 6, 2015.

[7] J. M. K. F. N. E. A. R. F. P. M. N. K. e. a. Chakaya, "Global Tuberculosis Report 2020–Reflections on the Global TB burden, treatment and prevention efforts.," *International Journal of Infectious Diseases ,* 2021.

[8] J. C. C.-L. a. A. Q.-B. Bullock, ""XNet: A convolutional neural network (CNN) implementation for medical X-ray image segmentation suitable for small datasets."," *In Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 10953, p. 109531Z. International Society for Optics and Photonics, 2019.,* p. 109531Z, 2019.

[9] J. D. J. F. J. d. S. G. S. A. a. d. P. A. Souza, "An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks.," *Computer methods and programs in biomedicine, 177.,* pp. 285-296, 2019.

[10] Y. M. S. M. T. a. L. R. Zhang, "Unsupervised x-ray image segmentation with task driven generative adversarial networks.," *Medical image analysis,* vol. 62, p. 101664, 2020.

[11] G. M. B. a. L. A. Gaál, "Attention u-net based adversarial architectures for chest x-ray lung segmentation.," *arXiv preprint,* vol. arXiv:2003.10304., 2020.

[12] T. K. A. K. M. I. K. I. K. M. R. H. T. I. M. K. S. M. Z. a. A. M. Rahman, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization.," *IEEE Access,* vol. 8, pp. 191586-191601, 2020.

[13] M. C. V. a. M. R. Abdel-Basset, "HSMA_WOA: A hybrid novel Slime mould algorithm with whale optimization algorithm for tackling the image segmentation problem of chest X-ray images.," *Applied soft computing,* vol. 95, p. 106642, 2020.

[14] L. P. R. B. D. O. L. N. L. C. G. a. C. Y. Teixeira, "Impact of lung segmentation on the diagnosis and explanation of COVID-19 in chest X-ray images.," *Sensors,* vol. 21, p. 7116, 2021.

[15] J. G. Y. K. A. S. A. B. O. M. M. a. S.-M. T. Wu, " Automatic bounding box annotation of chest x-ray data for localization of abnormalities.," *IEEE 17th International Symposium on Biomedical Imaging (ISBI),* pp. 799-803, 2020.

[16] I. K. M. M. T. Y. Y. K. R. a. I. B. Sirazitdinov, "Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database.," *Computers & electrical engineering,* vol. 78, pp. 388-399, 2019.

[17] P. S. P. a. V. A. Singhal, "Interpretation and localization of Thorax diseases using DCNN in Chest X-Ray," *Journal of Informatics Electrical and Electronics Engineering,* vol. 1, p. 1, 2020.

[18] T. a. R. S. Zebin, "COVID-19 detection and disease progression visualization: Deep learning on chest X-rays for classification and coarse localization.," *Applied Intelligence.,* vol. 51, pp. 1010-1021, 2021.

[19] W. D. A. D. E. C. S. S. R. C.-Y. F. a. A. C. B. Liu, "Ssd: Single shot multibox detector.," *European conference on computer vision,* pp. 21-27, 2016.