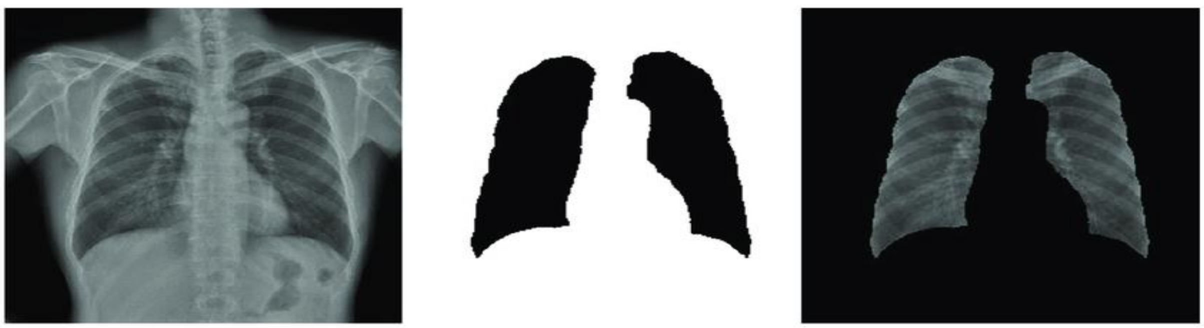
LUNG SEGMENTATION

Kaushik Jagini (MS Computer Science Eng.), Abhishek Srivastava (MS Data Science)

Problem Statement:

Surgery planning, disease diagnosis, and prognosis all depend on being able to accurately identify organ boundaries in medical pictures. The technique of correctly separating the regions and borders of the lung field from surrounding thoracic tissue is known as lung segmentation, and it is a crucial first step in pulmonary image processing for many clinical decision support systems. In this study, we suggest an approach based on deep learning for segmenting the lung regions in chest X-rays.



Motivation:

Additional computer analysis of these anatomical regions is available when lung fields are accurately defined. This enables the extraction of clinically relevant features that can be used to train a machine learning algorithm for the detection of disease and anomalies. In order to improve the caliber of care and patient outcomes, these computational tools can assist physicians in making rapid, precise diagnoses. For chest disorders such lung cancer, pulmonary edema (fluid in the lung), pleural effusion (fluid between the lung and chest cavity), pneumonia (infection by bacteria, viruses, fungus, or parasites), and tuberculosis, chest radiography is the most often used diagnostic imaging test (bacterial infection). Chest disorders cause more than 10 million deaths worldwide each year. According to a survey, 6.3 million people died from ischemic heart disease, 4.3 million from lower respiratory infections, 2.2 million from chronic obstructive pulmonary disease, 2 million from tuberculosis, and 0.9 million from other chest disorders in 1990. (lung cancer). Many treatments are only effective in the early, symptomless stages of most diseases. For accurate lung CT image analysis, such as lung cancer detection, lung image segmentation is a necessary step. The change in the visual extent of the disease over time is an essential measure of response to therapy and a predictor of death, and it can be used to quantify the progression, regression, or stagnation of lung disease. Additionally, segmented lungs can be employed for image registration, particularly in 3-D scans. As a result, the radiologists will be able to assess the volumetric changes in the lungs and determine the percentage of the patients' healthy parenchyma region that has been functionally reduced.

Literature review:

The delineation of anatomical features and other regions of interest is automated or made easier with the help of medical picture segmentation, which is essential in many imaging applications. Since the last ten years, the segmentation of lung fields in CXR has gained a lot of attention in the literature. For this work, a thorough review of lung segmentation methods for chest radiographs was conducted.

The majority of the works mentioned used the JSRT dataset as their image database. The Japanese Society of Radiological Technology (JSRT) created a database of chest radiographs (with and without lung nodules) and made it available to the public along with its ground truth clinical data. Only two studies, as far as we are aware, have used portable radiography of the chest. This demonstrates the lack of research on portable chest radiographs, which are extremely crucial, especially for critically ill patients.

Early segmentation techniques for CXR were divided into four groups: rule-based techniques, pixel classification-based techniques, deformable model-based techniques, and hybrid techniques. A rule-based scheme includes a series of actions, tests, and rules. The techniques include local thresholding, region growth, edge and ridge detection, morphological procedures, fitting of geometrical models, functions, and dynamic programming. Pixel classification-based technique, on the other hand, is more generic and basically models the intensities of the picture and classifies the pixels into lung field or background. The rule-based scheme is presented using Bezier interpolation of prominent control points. used the fuzzy clustering method (FCM)-based scheme, and one of their comparison outcomes was the post-processed pixel classification approach.

Although it has numerous limitations when used to apply towards discovering appropriate groups in data analysis tasks, the FCM method is the most well-known. To have a more reliable FCM, several researchers have attempted to change the fundamental goal function. FCM has been utilized with some success in the soft or fuzzy segmentation in medical imaging of chest CT, chest MRI, and brain MRI, even though the perfect segmentation of an image is often application dependent. The authors' motivation to use the FCM in their study came from the ease with which the borders of the lung or brain could be seen in CT and MRI images thanks to the unique bone and cell tissue. In their study, Rastgarpour et al. merged a local region-based level set approach with a version of fuzzy clustering in order to segment a few modalities and body parts, including the lung. Shi et al. applied FCM with spatial constraints to segment the lung for CXR. The literature also contains a few publications on additional FCM-based CXR applications. In their implementation of the algorithm for pneumonia identification, Parveen et al. also employed feature vectors to cluster feature data for atypicality detection by segmenting cardiac information (size, contour, and shape). The inhomogeneities of X-Ray imaging are mostly to blame for the paucity of work on lung segmentation for CXR utilizing FCM. FCM is difficult to utilize as a segmentation technique because of the sharp edges at the clavicle and rib cage regions as well as the intensity fluctuation around the lung area.

The deformable model-based approaches represent a recently developed system that has been thoroughly investigated and applied in medical picture segmentation. Lung area segmentation has been effectively used to this shape-flexibility model, Active Shape Model (ASM) and Active Appearance Model (AAM). But both have several drawbacks and restrictions, such as the need for manual initialization and supervision when changing parameters, which results in very variable solutions. The creation of the hybrid scheme aims to improve segmentation outcomes by combining the strategies mentioned previously. Many hybrid algorithms discovered in the literature mix rule- and shape-based techniques, which is quite intriguing to observe. However, the ASM-fused approaches have limitations, whereas the form schemes need computing complexity, learning, and training as well as the optimization process.

Scope of the project and tangible goals –

All of us are aware COVID-19 recently affected everyone worldwide with ~95,700,347 incident cases with ~5,327,014 hospitalization in united states alone. Tuberculosis and pneumonia are also two highly prevalent disease around the world and lung segmentation could help physicians in early and accurate diagnosis of such diseases and hence could save many lives or reduce the overall healthcare resources by significant margin.

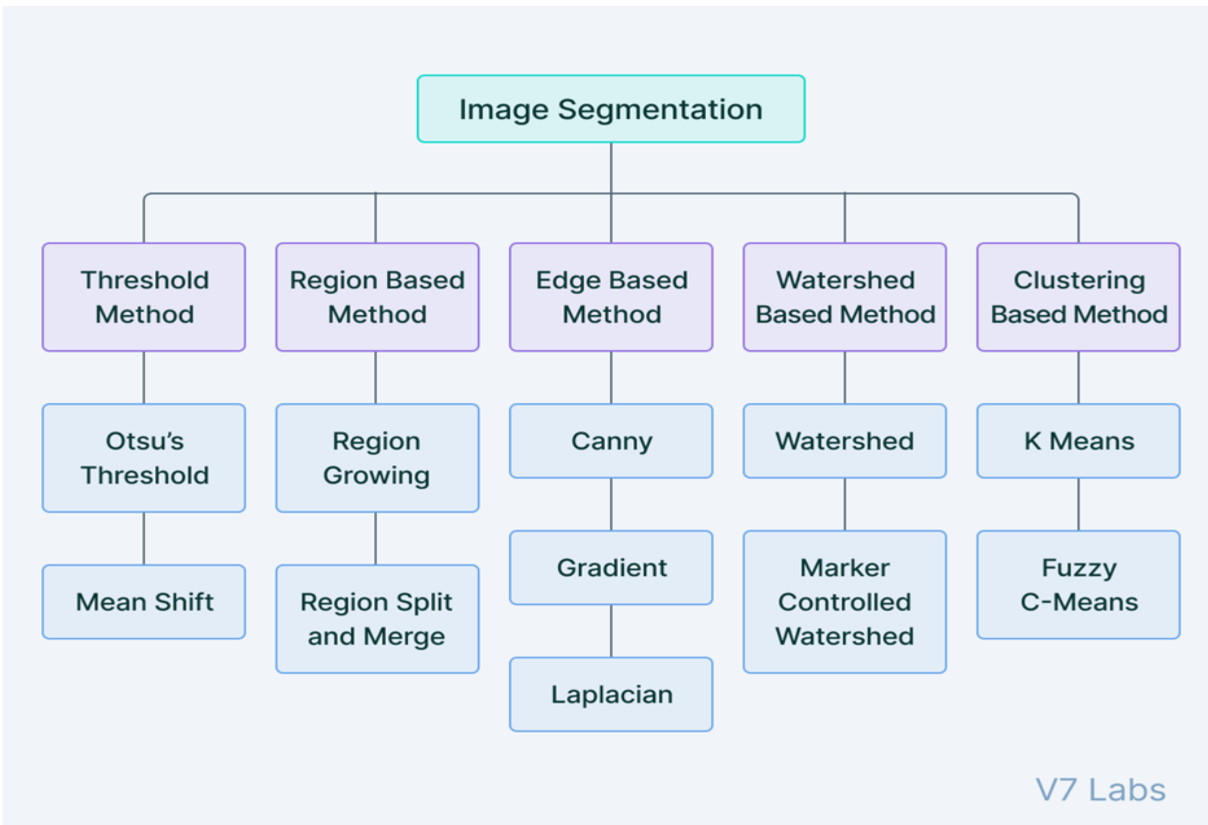
By accurately segmenting lungs and identifying boundaries with 100% precision, we could feed this data into classification models which can classify patients of distinct types of pneumonia with significantly higher accuracy. Although, the scope of this exercise is limited to deriving an approach to automate segmentation of lung given chest X-rays, but this could well be extended as an input to other classification model in downstream.

Project plan with Task-Based deadline –

1. Data Acquisition –
    a. We plan to use chest X-ray data (Montgomery County X-ray Set), which is publicly available. The dataset contains x-rays and corresponding masks. Some masks are missing so it is advised to cross-reference the images and masks. X-ray images in this data set have been acquired from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA. This set contains 138 posterior-anterior x-rays, of which 80 x-rays are normal and 58 x-rays are abnormal with manifestations of tuberculosis. All images are de-identified and available in DICOM format. The set covers a wide range of abnormalities, including effusions and miliary patterns. The data set includes radiology readings available as a text file. However, we intend to explore other publicly available datasets which could better suit our objectives.
2. Data Parsing and data cleansing –

a. Because the data of interest contains images in DICOM format, they need to be appropriately handled while importing them to python and conversion to numpy arrays. We will use pydicom for the same.

3. Labelling –
   a. There are many existing tools to create labelled masks for images, including Labelme, ImageJ, and even the graphics editor GIMP. While these are all great tools, they cannot be integrated within a Jupyter notebook, making them harder to use with many existing workflows. Fortunately, Jupyter Widgets make it easy for us to make interactive components and connect them with the rest of our Python code.
4. Develop Model on train data–
   a. Traditional Methods –



   b. Deep learning-based methods –
      i. U-NET was originally developed for medical image understanding and segmentation. It has vast applications in the domain and has been a key architecture in the medical imaging automation society. The architecture of this network includes two main parts: contractive and expansive. The contracting path consists of several patches of convolutions with filters of size 3×3 and unity strides in both directions, followed by ReLU layers. This path extracts the key features of the input and results in a feature vector of

a specific length. The second path pulls information from the contractive path via copying and cropping, and from the feature vector via up-convolutions, and generates, by a successive operation, an output segmentation map. The key component of this architecture is the operation linking the first and second paths together. This linkage allows the network to attain highly accurate information from the contractive path, thus generating the segmentation mask as close as possible to the intended output.

*We intend to explore more potential methods for image segmentation which could work better, specifically for medical images

5. Parameter hyper tuning on Validation data
6. Model evaluation on test data
7. Highlight limitations and capabilities of the model

Distribution of work

Abhishek Srivastava

- Data Acquisition
- Labeling
- Model development
  - A set of traditional methods – Region based; Clustering based methods
  - Learning based methods
- Model Evaluation and parameter tuning of respective models
- Report and summarize the model results

Kaushik Jagini

- Data Parsing
- Data cleaning
- Model development
  - Remaining set of traditional methods – Watershed method; Edge based method
  - Learning based methods
- Model Evaluation and parameter tuning of respective models
- Report and summarize the model results

Both candidates would attempt to research across the task assigned to deliver insights into each other's work. Model development and evaluation are the tasks where a higher level of collaboration is required and intended.

Reference Links:

1. https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-020-00514-y#:~:text=Lung%20segmentation%2C%20the%20process%20of,many%20clinical%20decision%20support%20systems.
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4521615/
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8720554/
4. https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-015-0014-8#ref-CR1
5. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(96)07493-4/fulltext
6. https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-13-41
7. https://link.springer.com/article/10.1007/s10916-014-0068-3
8. https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-ipr.2011.0128
9. https://link.springer.com/article/10.1007/s10916-009-9255-z
10. https://ieeexplore.ieee.org/document/5740617
11. https://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2011.0263
12. https://ieeexplore.ieee.org/document/974918
13. https://link.springer.com/article/10.1007/s10115-009-0225-z
14. https://link.springer.com/article/10.1007/s10916-014-0021-5
15. https://ieeexplore.ieee.org/document/802752
16. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html
17. https://towardsdatascience.com/how-we-built-an-easy-to-use-image-segmentation-tool-with-transfer-learning-546efb6ae98
18. https://www.v7labs.com/blog/image-segmentation-guide
19. https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-020-00529-5