Chest CT Segmentation and Model Comparative Analysis

Meet Oza (mgo26), Sriya Challa (sc2677), Bushra Tasneem (bt353), Nathaniel Nirmal (na453)

Cornell Tech

ABSTRACT

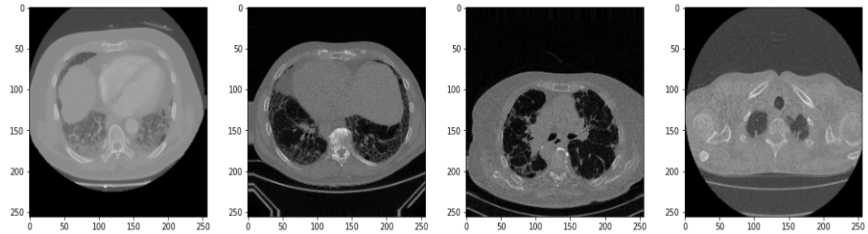
Our objective for this project is to perform lung CT image segmentation using the following architectures: U-Net, U-Net++, and V-Net, and compare the three models to find which model provides more precise region segmentation with the highest score and the lowest loss. We performed a multi-class segmentation since the dataset contains the following three classes: lung mask, heart mask, and trachea mask. The model with the highest performance metrics and the lowest loss score was the U-Net++ model. These discoveries play a crucial role in the progression of medical imaging, particularly in improving the precision of diagnosing lung diseases.

BACKGROUND AND LITERATURE REVIEW

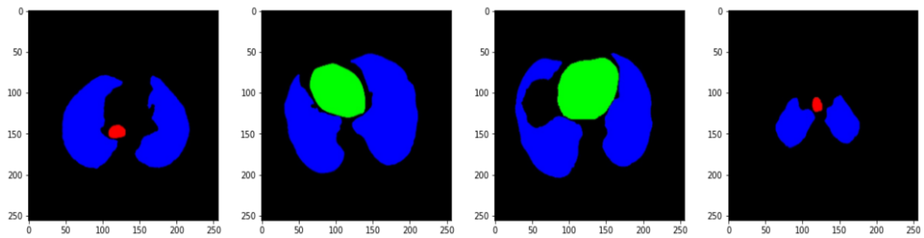
Lung disorders are the third prominent cause of death worldwide. Early diagnosis of lung infections/diseases through AI-enabled advanced clinical diagnosis systems could save people before any severe damage to the lungs. Segmentation is an important step in the overall process. In this paper, we aim to add knowledge to this area by creating lung image CT segmentation models using U-net, U-net++, and V-net architectures. To train our models, we use a dataset of 16708 images. Ideally, we want to use this model for accurate segmentation of lung CT scans to help physicians diagnose patients more efficiently. In the literature, Skourt et al demonstrated lung segmentation using U-net[1]. They were able to achieve an accurate segmentation with a 0.95 Dice-Coefficient index based on a small image set that contains a few hundred of manually segmented lung images. Murugappan et al developed a computationally efficient and robust deep learning model for lung segmentation using CT images with DeepLabV3 + networks for two-class (background and lung field) and four-class (ground-glass opacities, background, consolidation, and lung field) using a publicly available database for COVID-19 that contains 750 chest CT images[2]. The researchers proposed an improved DeepLabV3+ network with ResNet-18 which requires fewer network parameters and achieved higher accuracy in semantic segmentation compared to the original DeepLabV3+ network. In another paper published in 2021 which also focused on image segmentation of COVID-19 lung images, Saood and Hatem attempted to detect and label infected tissues on CT lung images USING SegNet (which is a scene segmentation tool) and U-net[3]. They found that U-net performed better than SegNet as a multi-class segment with a 0.91 mean accuracy.

DATASET

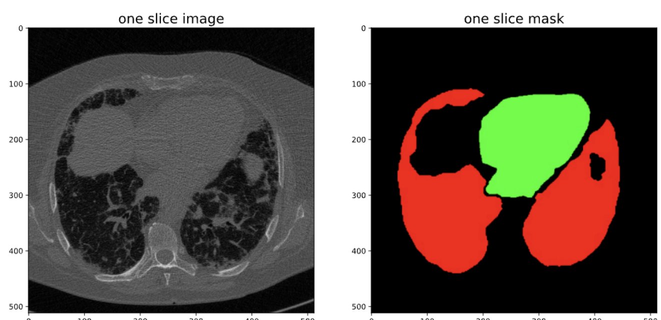
The dataset is adapted from the Lung Segmentation dataset by Kónya et al., 2020[4]. It contains chest CT scan images and their corresponding segmentations - lungs, heart, and trachea. Since there are 3 classes - lung mask, heart mask, and trachea mask, it is a multi-class segmentation.



Each tensor has the following shape: number of slices, width, height, and number of classes, where the width and height number of slices are individual parameters of each tensor ID, and the number of classes = 3. The dataset has images in RGB format and masks in RGB format along with train csv which contains image names.



The scans are converted to images of each slice and there are a total of 16,708 images and their masks. Sample image representation:



The dataset can be downloaded in .zip format and the size can be accommodated over Google Drive which can be then linked to Google Colab where we will be running our model and tests. In terms of pre-processing, the images were resized to 256,256,3 which is suitable for traditional CNN models.

ANALYSIS

1. Brief:

This research paper presents a comparative analysis of three deep learning models—U-Net, U-Net++, and V-Net—for CT lung segmentation. The objective is to evaluate and compare the performance of these models in accurately delineating lung regions in medical imaging data. The study employs evaluation metrics such as loss function(BCEDiceLoss), Dice score, and Jaccard score to assess the segmentation performance of each model quantitatively. The dataset used in this study comprises a diverse set of CT scans with annotated lung regions. Each model is trained on the dataset using a common preprocessing pipeline, including data augmentation techniques to enhance model generalization.

2. Models:

2.1 U-Net:
U-Net[5] is a widely used architecture for image segmentation tasks, known for its simplicity and effectiveness. The model consists of a contracting path, a bottleneck, and an expansive path. Skip connections between the contracting and expanding paths help preserve spatial information and enhance segmentation accuracy.

2.2 U-Net++:
U-Net++[6] is an extension of the U-Net architecture, designed to address segmentation accuracy and model expressiveness limitations. The model incorporates nested skip pathways, creating a more intricate architecture to capture hierarchical features at multiple scales.

2.3 V-Net:
V-Net[7] is specifically tailored for volumetric image segmentation tasks such as medical imaging. It leverages a 3D convolutional neural network (CNN) architecture, enabling the model to capture spatial dependencies in three dimensions. This makes V-Net well-suited for volumetric CT lung segmentation.

3. Evaluation Metrics:

To quantitatively assess the performance of each model, three key evaluation metrics are employed which are the standard loss function, Dice Score, and Jaccard Score. Below is a brief description and mathematical function for the two scores:

3.1 Dice Score:

The Dice score[8], also known as the F1 score, measures the overlap between predicted and ground truth segmentations. It is computed as twice the intersection of the predicted and ground truth masks divided by the sum of their areas. A higher Dice score indicates better segmentation accuracy.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

3.2 Jaccard Score:

The Jaccard score[9], also known as the Intersection over Union (IoU), quantifies the similarity between the predicted and ground truth segmentations. It is calculated as the intersection of the predicted and ground truth masks divided by the union of their areas. A higher Jaccard score signifies better segmentation performance.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

4. Model Training and Pipeline:

The model training pipeline for U-Net, U-Net++, and V-Net in the context of CT lung segmentation begins with the loading of annotated CT scan data, which undergoes data augmentation to enhance model generalization. Model architectures are configured with specified layers, filter sizes, and activation functions. Loss functions, such as binary cross-entropy or Dice loss, are selected, and optimizers like SGD or Adam are chosen for minimizing these losses during training. The training process involves iterative adjustments of model parameters over multiple epochs, with periodic validation on separate datasets to prevent overfitting. Hyperparameters, including learning rate and batch size, are tuned for optimization. Following training, models are evaluated on a test dataset using metrics like Dice score and Jaccard score. Optional post-processing techniques may refine predicted segmentations, and the results, including quantitative metrics and visualizations, are analyzed to compare the segmentation performance of U-Net, U-Net++, and V-Net in CT lung segmentation.

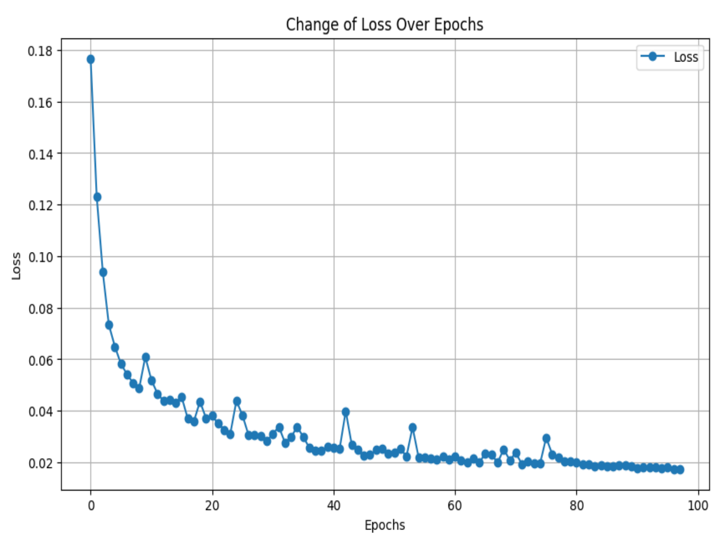
RESULTS

The performance of U-Net, U-Net++, and V-Net in CT lung segmentation was comprehensively evaluated using a diverse dataset of annotated CT scans. Table 1 summarizes the quantitative results of the test set, showcasing the performance metrics for each model.

Metric	Unet	Unet++	Vnet
Loss	0.04	**0.02**	0.04
Dice Score	0.92	**0.94**	0.93
Jaccard Score	0.90	**0.93**	0.92

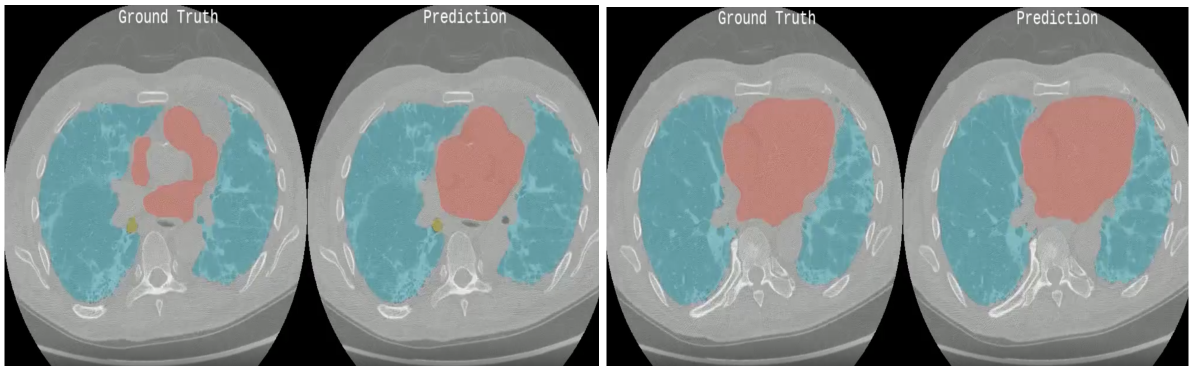
As expected and observed from other literature reviews performed, Unet++ remains relatively better at image segmentation tasks. Vnet is a close cousin of Unet and thus the similar model results. Overall, every model gave a Dice Score of over 0.9 which indicates that the 3 classes were segmented decently overall.

Let us see some more results from our best performing model - Unet++. The below graph shows the loss per epoch.



Below are some of the visualization results of the best performed model.

Ground Truth vs Predictions of image slices:



As you can see, the lungs are segmented out very precisely but there are some boundary issues with segmenting the heart properly. The very small yellow dot on the left image is the trachea and it is segmented well by the model.

INTERPRETATION OF RESULTS

Essentially, models like Unet have 2 parts - a contracting network and an expanding network. This makes them ideal for tasks like image segmentation and thus we can see a Dice score of over 0.9 for such models. In terms of model performances, U-Net++ has consistently demonstrated enhanced performance in image segmentation tasks, a characteristic attributed to its improved skip connections and nested architecture. V-Net, being conceptually akin to U-Net, exhibits comparable results, reaffirming the efficacy of the underlying convolutional neural network (CNN) architecture. Finally, even though we saw Unet++ outperforming others, it is interesting to note the faster convergence of the Vnet model.

CONCLUSION

In conclusion to this project we have evaluated U-Net, U-Net++, and V-Net models for CT lung segmentation, gaining valuable insights into their respective strengths and weaknesses. The analysis revealed the segmentation model’s overall robustness and efficiency, U-Net++'s superior boundary delineation capabilities, and V-Net's effectiveness in processing volumetric data. These findings are pivotal for advancing medical imaging, particularly in enhancing the accuracy of lung disease diagnosis.

Future work can focus on integrating these models into clinical practices and exploring their potential in various medical imaging applications. This will not only improve diagnostic precision but also pave the way for more personalized and efficient healthcare solutions. The evolution of AI in healthcare, as demonstrated by these models, underscores the transformative impact of technology in improving patient outcomes and streamlining medical processes.