Deep Learning-Based Lung Segmentation using U-Net

Avinash Rajaraman Swaminathan
Yeshiva University
aswamina@mail.yu.edu

Abstract

Lung segmentation is a crucial step in analyzing chest X-ray images for the accurate diagnosis and treatment of lung-related diseases. In this paper, I propose a deep learning-based approach for lung segmentation in chest X-ray images using the U-Net architecture. The method leverages the power of convolutional neural networks (CNN) and a custom decoder to segment lung regions from X-ray images effectively. I demonstrate the effectiveness of the approach on a dataset of lung X-ray images, achieving high Intersection over Union (IoU) and Dice scores, which indicate the quality of the segmentation results.

1. Introduction

Building upon the remarkable success of Residual Networks (ResNets) in various computer vision tasks, including image classification, object detection, and segmentation, the proposed work extends the ResNet architecture to encompass a 101-layer CNN structure comprising multiple residual block layers. The ResNet architecture introduces skip connections between layers, enabling the network to learn the residual mapping between input and output, thus addressing the vanishing gradient problem and facilitating the training of deeper networks.

The proposed U-Net model processes input images through a series of convolutional layers and ReLU activation functions within its encoder, effectively extracting rich feature representations. The network leverages residual blocks to facilitate efficient learning of the residual mapping between input and output.

Trained on a lung dataset [4] containing images from various categories, the model demonstrates high accuracy in the classification task. Data augmentation techniques are employed to enhance the model's performance, with the proposed model achieving a test IOU score of 88.50%.

2. Related Work

In recent years, numerous approaches have been proposed for lung segmentation in chest X-ray images, with deep learning techniques proving to be particularly effective. This section discusses some of the key works in this domain, highlighting the similarities and differences with the proposed method.

2.1. U-Net and its Variants

The U-Net architecture is characterized by its symmetric encoder-decoder structure and skip connections, which help maintain fine-grained spatial information for accurate segmentation. Several variants and adaptations of the U-Net have been proposed, including V-Net [1], and Attention U-Net [2], further improving the original U-Net's performance.

2.2. Combining ResNets with U-Net

Integrating ResNet-based encoders with U-Net architectures [3] has emerged as a promising approach for medical image segmentation. ResNet's skip connections aid in mitigating the vanishing gradient problem and allow for deeper networks to be trained effectively. Several works have combined ResNets with U-Net to improve segmentation performance in various medical imaging tasks.

2.3. Comparison with the Proposed Method

The proposed method extends the U-Net architecture by integrating a pre-trained ResNet-based encoder and a custom decoder. This combination enables the model to effectively learn rich feature representations from the chest X-ray images, resulting in high-quality lung segmentation. Compared to the existing works, the method achieves competitive performance in terms of IoU and Dice scores, demonstrating its effectiveness for lung segmentation in chest X-ray images.

The model is evaluated on the lung images dataset [4], a pre-compiled and acquired dataset containing train and test images and masks of lungs for image segmentation. The

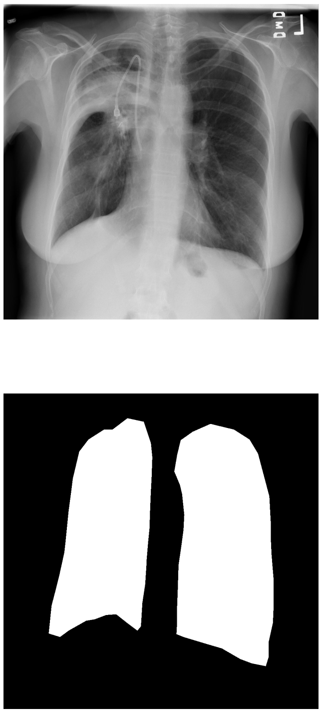

Figure 1: Lung Image and Mask

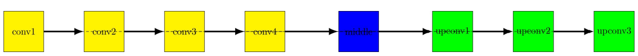

Figure 2: Encoder-Decoder Architecture

experiments show that the proposed model achieved a validation IOU score of 87.94%, and validation Dice score of 93.56%. The results demonstrate that the proposed model is an effective and robust architecture for medical image segmentation tasks.

3. Methods

In this study, I propose a U-Net based model with a ResNet backbone for lung segmentation from chest X-ray images. The U-Net architecture has been widely used for various segmentation tasks, and its success can be attributed to its symmetric encoder-decoder structure, which allows for both high-level and low-level features to be learned and combined effectively. I used the ResNet architecture as the backbone of the model due to its strong ability to learn hierarchical features while mitigating the vanishing gradient problem.

3.1. Model Architecture

Given U-Net model consists of an encoder, a middle layer, and a decoder. The encoder uses the ResNet101 architecture, with the fully connected layers removed. I have replaced the first convolutional layer of ResNet with a custom layer to accommodate different input channels. The middle layer consists of two convolutional layers followed by ReLU activation functions. The decoder has three upconvolutional layers followed by a series of convolutional layers and ReLU activation functions. The final layer in the decoder uses a 1x1 convolution to produce the output segmentation mask.

3.2. Loss Function and Optimizer

I have used the binary cross-entropy with logits loss (BCEWithLogitsLoss) as the loss function. This loss function is suitable for binary segmentation tasks, as it measures the error between the predicted mask and the ground truth mask. The loss function can be defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)) \tag{1}$$

where N is the number of pixels, y_i is the ground truth label for the i-th pixel, and $\hat{y}_i$ is the predicted probability for the i-th pixel.

The model is trained for 200 epochs using the Adam optimizer with an initial learning rate of 0.0001.

3.3. Evaluation of Performance Indicators

To evaluate the model's performance, I have used the Intersection over Union (IoU) and Dice scores as the primary metrics. These metrics are commonly used to measure the overlap between the predicted segmentation mask and the ground truth mask.

Intersection over Union (IoU) is defined as the ratio of the intersection between the predicted mask and the ground truth mask to their union. The IoU score can be calculated as:

$$\text{IoU} = \frac{A \cap B}{A \cup B} = \frac{\sum_{i=1}^{N}(y_i \cdot \hat{y}i)}{\sum i = 1^N (y_i + \hat{y}_i - y_i \cdot \hat{y}_i)} \tag{2}$$

where A and B represent the ground truth mask and the predicted mask, respectively.

The Dice score, also known as the F1 score or Sørensen-Dice coefficient, is defined as the ratio of twice the intersection between the predicted mask and the ground truth mask to the sum of their areas. The Dice score can be calculated as:

$$\text{Dice} = \frac{2 \cdot A \cap B}{A + B} = \frac{2 \cdot \sum_{i=1}^{N}(y_i \cdot \hat{y}i)}{\sum i = 1^N (y_i + \hat{y}_i)} \tag{3}$$

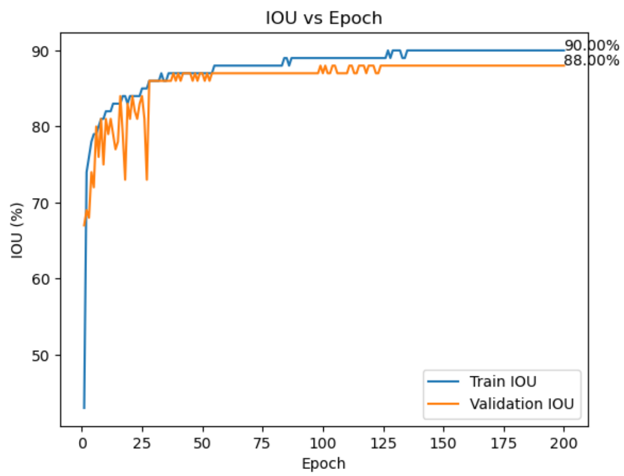

Figure 3: Comparison of IOU Score vs. Epoch during training and validation

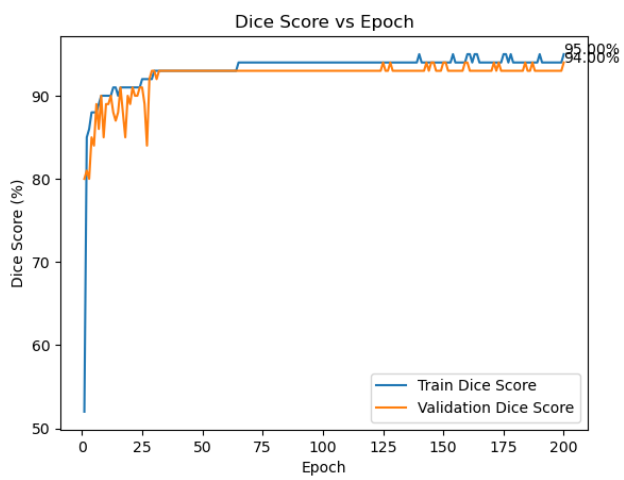

Figure 4: Comparison of Dice score vs. Epoch during training and validation

Based on the plots presented in the section, it can be inferred that the proposed U-net model performed well during training and validation, achieving a training IOU of 89.62% and a validation IOU of 87.94%. Furthermore, the plot of IOU vs. epoch shows a similar change in IOU for both training and validation sets as the number of epochs increases, with the validation IOU staying close to the training accuracy. With respect to loss, both training and validation is at baseline level consistently after certain epochs and almost converges

3.4. Results

The model achieved a final validation IoU of 87.94% and a validation Dice score of 93.56%. The test set evaluation resulted in an IoU of 88.50% and a Dice score of 93.88%,

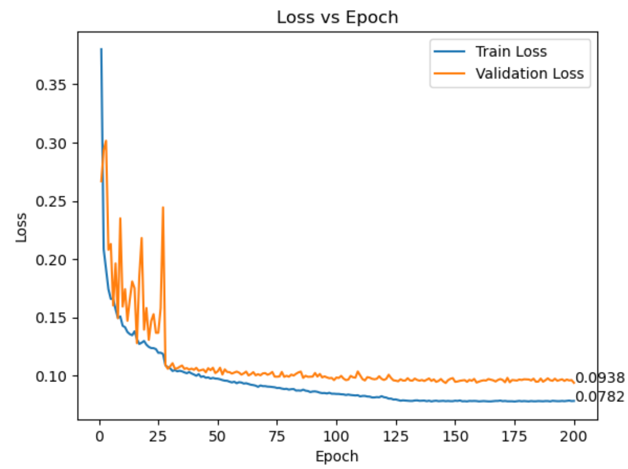

Figure 5: Comparison of Loss vs. Epoch during training and validation

indicating that the model generalizes well and meets the performance criteria of an IoU score higher than 0.8.

4. Conclusion

In this study, I have developed a U-Net based model with a ResNet backbone for lung segmentation using chest X-ray images. The model demonstrates promising performance, achieving a test IoU of 88.50% and a test Dice score of 93.88%. These results indicate that the proposed model can effectively segment lungs in chest X-ray images and could potentially serve as a valuable tool for assisting in diagnosing lung-related diseases. Future model improvements can include exploring other advanced architectures and backbone networks to further enhance the model's segmentation performance.

References

[1] Abolfazl Abdollahi, Biswajeet Pradhan, and Abdullah Alamri. Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. *IEEE Access*, 8:179424–179436, 2020. 1

[2] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 1

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 1

[4] Youshan Zhang. Lung segmentation with nasnet-large-decoder net. *arXiv preprint arXiv:2303.10315*, 2023. 1