

Lung Infection by Covid-19 Segmentation from Chest X-ray Images Using U-Net

Pham Minh Hieu

University of Science and Technology of Hanoi

Abstract—Chest X-ray (CXR) imaging is widely used for screening and diagnosis of lung diseases due to its low cost and availability. Accurate segmentation of infected regions in lung X-ray images is an important step toward computer-aided diagnosis systems. In this work, I focus on the segmentation of lung infection regions from CXR images using a deep learning-based semantic segmentation approach. A U-Net architecture is employed to learn pixel-wise representations of infection areas. Experimental results demonstrate the effectiveness of the proposed method in identifying infected regions in COVID-19 chest X-ray images.

I. INTRODUCTION

Chest X-ray imaging plays a crucial role in the diagnosis and monitoring of pulmonary diseases, including COVID-19. Manual delineation of infected regions is time-consuming and highly dependent on expert knowledge. Therefore, automatic segmentation of infection regions using deep learning techniques has gained increasing attention.

In this report, I propose a U-Net-based segmentation framework to automatically segment infection regions from chest X-ray images. The main objective of this work is to accurately localize infection areas within the lung region, which can serve as a foundation for further quantitative analysis and clinical decision support.

II. DATASET

The dataset used in this work is organized into two main components: Infection Segmentation Data and Lung Segmentation Data. The overall directory structure is illustrated as follows:

```
Dataset/  
  Infection Segmentation Data/  
    Train/  
    Val/  
    Test/  
  Lung Segmentation Data/  
    Train/  
    Val/  
    Test/
```

A. Infection Segmentation Data

Each split (Train, Val, Test) in the Infection Segmentation Data contains three categories:

- COVID-19: Chest X-ray images of lungs infected by COVID-19.

- Non-COVID: Chest X-ray images with lung infections not caused by COVID-19.
- Normal: Chest X-ray images of healthy lungs.

Inside each category, the following subdirectories are provided:

- images: Original chest X-ray images.
- infection_masks: Binary masks indicating infection regions.
- lung_masks: Binary masks of the whole lung region.

B. Lung Segmentation Data

The Lung Segmentation Data follows a similar Train/Val/Test split. However, each category only contains:

- images: Chest X-ray images.
- lung_masks: Ground truth lung segmentation masks.

Notably, this dataset does not provide infection masks.

C. Data Selection Strategy

The objective of this work is to perform segmentation of infection regions in lung X-ray images. Therefore, only the Infection Segmentation Data is used for training and evaluation. The Lung Segmentation Data is excluded since it does not contain infection masks and thus cannot directly support the target task.

Furthermore, during training and validation, only images from the COVID-19 category are used as positive samples for infection segmentation, along with a limited number of images from the Normal category.

Although the Non-COVID category also provides infection masks, these masks are entirely black and visually indistinguishable from the masks of Normal images. Including Non-COVID samples may cause the model to incorrectly learn that infected lungs not caused by COVID-19 correspond to normal (non-infected) regions. To avoid this ambiguity and ensure that the model focuses on meaningful infection patterns, Non-COVID images are excluded from the training and validation process.

III. METHODOLOGY

This section describes the overall framework of the proposed lung infection segmentation approach. The pipeline consists of data preprocessing, model training using a U-Net architecture, and inference on unseen test images.

A. Data Preprocessing

All chest X-ray images are provided as grayscale images with a fixed resolution of 256×256 . Therefore, no resizing operation is required during preprocessing.

Each image is first normalized by scaling pixel intensities to the range $[0, 1]$, followed by standardization to $[-1, 1]$ using

$$x' = 2(x - 0.5).$$

To match the input requirement of the pretrained encoder, the single-channel image is replicated into three channels, resulting in an input tensor of size $3 \times 256 \times 256$.

For the segmentation masks, pixel values are binarized such that all non-zero values are mapped to 1, representing infection regions, while zero values indicate background. The mask is then reshaped to a single-channel binary tensor of size $1 \times 256 \times 256$.

B. Segmentation Network

A U-Net architecture is employed for pixel-wise segmentation of infection regions. The network follows an encoder-decoder structure with skip connections to preserve spatial information at different scales.

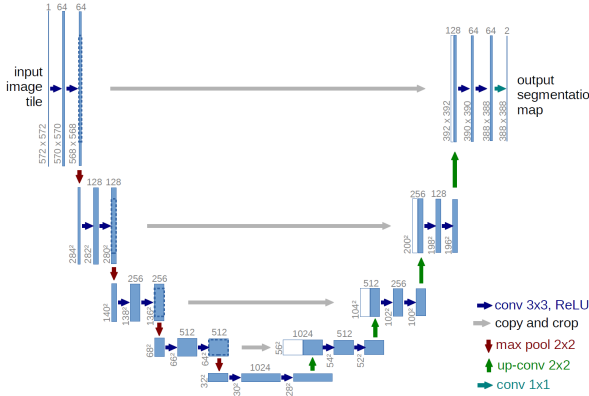


Fig. 1: U-Net architecture.

The encoder is based on a ResNet-34 backbone pretrained on ImageNet, which helps accelerate convergence and improve feature extraction. The decoder progressively upsamples the feature maps while concatenating corresponding encoder features via skip connections.

The network takes a $3 \times 256 \times 256$ image as input and outputs a single-channel probability map indicating the likelihood of infection at each pixel. A sigmoid activation is applied during inference to obtain pixel-wise probabilities.

To optimize segmentation performance, a hybrid loss function is used, combining Dice loss and binary cross-entropy loss:

$$\mathcal{L} = \mathcal{L}_{Dice} + \mathcal{L}_{BCE}.$$

This combination helps address class imbalance while maintaining stable gradient updates.

C. Training Strategy

The dataset is divided into training and validation sets according to the provided data split. The model is trained for 20 epochs using a batch size of 8.

The Adam optimizer is employed with a learning rate of 1×10^{-4} . During training, the model parameters are updated based on the combined Dice and binary cross-entropy loss.

At each epoch, the model is evaluated on the validation set, and the average validation loss is computed. The model achieving the lowest validation loss is saved and used for subsequent evaluation.

IV. EVALUATION METRICS

Model performance is evaluated using several commonly used segmentation metrics, computed at the pixel level. These include precision, recall, Dice coefficient, and Intersection over Union (IoU). Accuracy is not used as a primary evaluation metric in this study. Due to the large proportion of background pixels in chest X-ray images, a model can achieve a high accuracy by simply predicting most pixels as background, even when failing to correctly segment infection regions.

Given true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), the metrics are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN},$$

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}, \quad \text{IoU} = \frac{TP}{TP + FP + FN}.$$

Evaluation is performed separately on COVID-19, Non-COVID, and Normal test sets to analyze the model's robustness and false positive behavior. In addition to quantitative metrics, predicted masks and overlay visualizations are generated for qualitative assessment.

V. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the quantitative and qualitative evaluation of the proposed lung infection segmentation framework. The experiments were conducted using a U-Net-based architecture with identical hyperparameters across all settings. Two models are considered in this study:

- Model 1: a baseline model trained exclusively on COVID-19 images.
- Model 2: a fine-tuned model trained on COVID-19 images with the inclusion of additional Normal images at an approximate ratio of 3:10.

The motivation for developing Model 2 arises from an error analysis of Model 1. While Model 1 performs well on COVID-19 images, qualitative inspection reveals that it frequently produces false positive infection predictions on Normal chest X-ray images, where no pathological findings are present. This behavior suggests that training solely on COVID-19 images encourages the model to associate general lung textures and structural patterns with infection. To address this limitation, Normal images are introduced during training to

provide explicit negative supervision at the pixel level, leading to the development of Model 2.

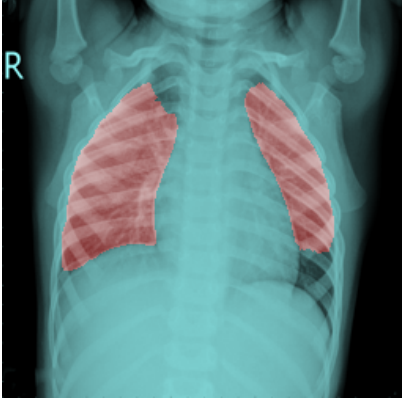


Fig. 2: Example of false positive of Model 1 on Normal image.

A. Comparison at Threshold 0.5

Both models are first evaluated using a fixed threshold of 0.5. The quantitative results on the COVID-19 test set are shown in Table I.

TABLE I: Performance comparison at threshold = 0.5

	Model 1	Model 2
Precision	0.7567	0.7804
Recall	0.8714	0.8242
F1-score	0.7800	0.7699
Dice	0.7800	0.7699
IoU	0.6742	0.6648

Model 1 achieves higher recall, showing strong sensitivity to infected regions in COVID-19 images. However, this also leads to over-segmentation and many false positives on Non-COVID and Normal images. In contrast, Model 2 improves precision by reducing these false positive predictions, although recall decreases by about 5%. This reflects a clear precision–recall trade-off: Model 2 is more conservative but produces cleaner predictions outside the COVID-19 domain.

B. False Positive Analysis on Non-COVID and Normal Images

Since Non-COVID and Normal images have empty ground truth masks, overlap-based metrics are not suitable for evaluating model behavior. Therefore, I measure the Predicted Positive Area Ratio (PPAR), defined as the fraction of pixels predicted as infection in each image.

Figure 3 shows the PPAR distributions for both models on COVID-19, Non-COVID, and Normal test sets.

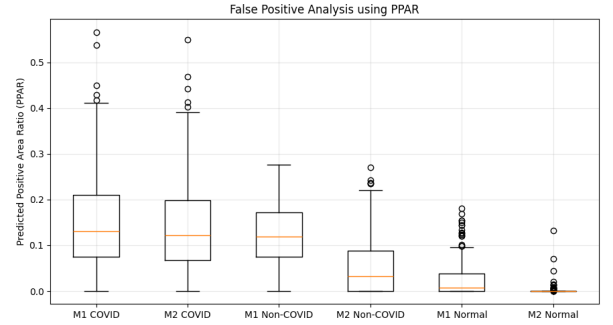
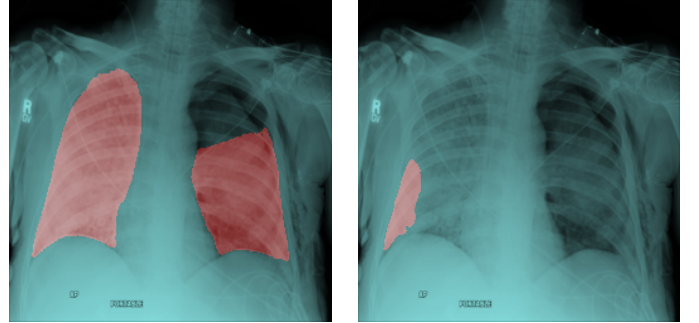


Fig. 3: PPAR distributions for Model 1 and Model 2 on different test sets.

On COVID-19 images, both models produce similar PPAR values, indicating comparable segmentation extent. However, on Non-COVID and Normal images, Model 1 shows much higher PPAR, meaning that it frequently predicts infection where none exists. Model 2 significantly reduces this behavior. On Non-COVID images, some regions are still predicted due to visual similarities with COVID-19 pneumonia, but these activations are sparse and limited. On Normal images, Model 2 produces almost no predicted infection regions.



(a) Model 1 prediction

(b) Model 2 prediction

Fig. 4: Comparison of predictions on Non-COVID images.

These results confirm that adding Normal images during training effectively suppresses false positive activations and improves model robustness.

C. Threshold Tuning for Model 2

Because Model 2 is conservative at the default threshold of 0.5, I further evaluate its performance at a lower threshold of 0.25. The results are shown in Table II.

TABLE II: Model 2 performance at threshold = 0.25

Metric	Value
Precision	0.7587
Recall	0.8525
F1-score	0.7698
Dice	0.7698
IoU	0.6642

Lowering the threshold increases recall by about 3% while keeping precision at a level comparable to Model 1. Importantly, this does not introduce a noticeable increase in false

positives on Non-COVID and Normal images. This shows that Model 2 remains stable even when the decision threshold is relaxed.

Overall, introducing Normal images does not teach the model what Non-COVID pneumonia looks like, but rather what COVID-19 infection does not look like. This leads to better control of false positives and more reliable segmentation results in practical settings. sectionIntroduction

D. Discussion

From a practical perspective, reducing false positive predictions on Normal chest X-ray images is critical, as incorrect infection segmentation may lead to misleading clinical interpretations. Although Model 1 achieves slightly higher recall and overlap-based metrics, its tendency to over-segment healthy lungs limits its reliability in real-world scenarios.

Model 2 offers a better balance between sensitivity and specificity. At the default threshold of 0.5, it produces more conservative and reliable predictions on Normal and Non-COVID images. Furthermore, recall can be effectively improved through threshold tuning without substantially increasing false positives. Therefore, Model 2 is considered the preferred model for lung infection segmentation under realistic data distributions.

The choice of operating threshold allows flexible control over the precision–recall trade-off, enabling the model to be adapted to different application requirements.

VI. CONCLUSION

In this work, I presented a U-Net-based framework for automatic segmentation of lung infection regions from chest X-ray images. By carefully selecting appropriate training data and excluding ambiguous samples, the proposed approach focuses on learning meaningful infection patterns associated with COVID-19. Experimental results demonstrate the potential of deep learning methods in supporting automated analysis of medical imaging data.

REFERENCES

REFERENCES

- [1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *MICCAI*, 2015.