Segmentation of COVID19 X-ray images using U-net

Nguyen Hoang Tung
Student ID: 22BA13318
University of Science and Technology of Ha Noi (USTH)
Email: tungtimo0808@gmail.com

I. EXPLORATORY DATA ANALYSIS

A. Dataset Overview

1) Training Set: The training set is provided in the `Train` folders of both *Lung Segmentation Data* and *Infection Segmentation Data*. The training data include chest X-ray images together with their corresponding ground-truth masks.

For the **Lung Segmentation Data**, the training set is a subset of the COVID-QU-Ex dataset and contains chest X-ray images with their corresponding lung masks.

For the **Infection Segmentation Data**, the training set is constructed from a combination of COVID-QU-Ex and QaTa-Cov19 datasets. It includes:

- **COVID-19**: chest X-ray images with corresponding lung masks and infection masks.
- **Non-COVID**: chest X-ray images with corresponding lung masks.
- **Normal**: chest X-ray images with corresponding lung masks.

For COVID-19 samples, both lung masks and infection masks are provided, while for Non-COVID and Normal samples, only lung masks are available. All masks are manually annotated by medical experts and are used as ground-truth for training the segmentation models.

2) Test Set: The test set is provided in the `Test` folders of both *Lung Segmentation Data* and *Infection Segmentation Data*. It contains chest X-ray images without using their labels during inference time.

Each test sample includes:

- **Image**: the chest X-ray image.

For evaluation purposes, the corresponding ground-truth lung masks and infection masks are available. The trained models are applied to the test images to predict:

- Lung segmentation masks.
- Infection segmentation masks (for COVID-19 cases).

These predicted masks are then compared with the ground-truth masks to evaluate the segmentation performance using standard metrics such as Dice coefficient and Intersection-over-Union (IoU).

B. Class Distribution

From Table I, it can be observed that the infection segmentation subset includes 2,913 COVID-19 images with corresponding lung and infection masks, 1,457 Non-COVID images, and

TABLE I
DATA DISTRIBUTION OF THE INFECTION SEGMENTATION SUBSET

Split	Class	Images	Lung Masks	Infection Masks
Train	COVID-19	1864	1864	1864
Train	Non-COVID	932	932	932
Train	Normal	932	932	932
Validation	COVID-19	466	466	466
Validation	Non-COVID	233	233	233
Validation	Normal	233	233	233
Test	COVID-19	583	583	583
Test	Non-COVID	292	292	292
Test	Normal	291	291	291
Total	COVID-19	**2913**	**2913**	**2913**
Total	Non-COVID	**1457**	**1457**	**1457**
Total	Normal	**1456**	**1456**	**1456**

1,456 Normal images, result in a total of 5,826 chest X-ray images. The dataset is split into training, validation, and test sets with a consistent class distribution across all splits. This balance partition ensures that the model is trained and evaluated under fair and representative conditions, while also reducing potential bias toward any specific class.

TABLE II
DATA DISTRIBUTION OF THE LUNG SEGMENTATION DATASET

Split	Class	Images	Lung Masks
Train	COVID-19	7658	7658
Train	Non-COVID	7208	7208
Train	Normal	6849	6849
Validation	COVID-19	1903	1903
Validation	Non-COVID	1802	1802
Validation	Normal	1712	1712
Test	COVID-19	2395	2395
Test	Non-COVID	2253	2253
Test	Normal	2140	2140
Total	COVID-19	**11956**	**11956**
Total	Non-COVID	**11263**	**11263**
Total	Normal	**10701**	**10701**
Total	**All**	**33920**	**33920**

From Table II, it can be observed that the lung segmentation

dataset consists of 33,920 chest X-ray images, including 11,956 COVID-19, 11,263 Non-COVID, and 10,701 Normal cases, each with corresponding lung masks. The dataset is split into training, validation, and test sets with a consistent class distribution across all splits. This balanced partition helps ensure a fair and reliable evaluation of the lung segmentation model.

II. MODEL IMPLEMENTATION

In this work, we build a two-stage deep learning system to analyze COVID-19 from chest X-ray images. The first stage is used for lung segmentation. It extracts the lung region from the input image. The second stage is used for infection segmentation. It detects the infected areas inside the lung region. This design helps the model focus only on important areas and reduces the effect of the background.

A. Overall Pipeline

Given an input chest X-ray image, the lung segmentation model is applied first. It produces a binary lung mask. Then, the original image is multiplied by this mask to keep only the lung area. This masked image is passed to the infection segmentation model. The final system produces two outputs: a lung mask and an infection mask.

Both stages use fully convolutional neural networks with an encoder–decoder structure. We follow the U-Net architecture because it is simple and effective for medical image segmentation.

B. Network Architecture

We use the same U-Net structure for both lung and infection segmentation. The network has two main parts: an encoder and a decoder. The encoder extracts features from the input image at different scales. The decoder restores the spatial resolution to produce the final segmentation mask.

Skip connections are used between the encoder and the decoder. They help keep important spatial details. The input of the network is a single-channel chest X-ray image. The output is a pixel-wise probability map. A sigmoid activation function is used in the last layer to produce values between 0 and 1.

C. Data Preprocessing

All images are resized to a fixed size before training. Pixel values are normalized to the range $[0, 1]$. This helps the training process become more stable.

During training, we apply simple data augmentation. This includes random horizontal flipping and small rotations. These methods increase the diversity of the training data and help the model generalize better.

For the infection segmentation stage, the input image is first masked by the predicted lung region. This ensures that the second model only looks at the lung area.

D. Training Strategy

The two models are trained in two separate steps. First, the lung segmentation model is trained using the full lung segmentation dataset. After training is finished, this model is used to generate lung masks for the infection segmentation dataset.

Then, the infection segmentation model is trained using the masked lung images and the corresponding infection masks. In this way, each model learns its own task independently, which makes the training process simpler and more stable.

III. EXPERIMENTAL RESULTS

TABLE III
RESULTS OF THE INFECTION SEGMENTATION ON THE TEST SET

Metric	Value
Dice coefficient	0.7957
IoU	0.7449
Loss	0.0649

TABLE IV
RESULTS OF THE LUNG SEGMENTATION ON THE TEST SET

Metric	Value
Dice coefficient	0.9768
Loss	0.0559

Table IV and Table III summarize the quantitative results of the proposed two-stage segmentation system on the test set.

For the lung segmentation task, the model achieves a Dice coefficient of 0.9768 with a loss value of 0.0559. This result shows that the model can segment the lung regions very accurately and produce high-quality lung masks.

For the infection segmentation task, the model obtains a Dice coefficient of 0.7957 and an IoU of 0.7449, with a loss value of 0.0649. This task is more difficult because infection regions often have unclear boundaries and large shape variations. However, the results indicate that the model can still localize and segment most infection areas correctly.

Overall, these results confirm that the proposed two-stage pipeline works effectively. The accurate lung segmentation in the first stage provides reliable input for the second stage, which helps improve the stability and performance of the infection segmentation model.

IV. COMPARISON WITH PUBLISHED RESULTS

We compare our experimental results with the findings reported in the COVID-QU-Ex study by Tahir et al., which also uses the COVID-QU-Ex dataset and state-of-the-art segmentation models for both lung and infection segmentation tasks. In their work, the authors achieved a lung segmentation performance with Dice coefficient of 97.99% and an IoU of 96.11% on the test set. For infection segmentation, their model

TABLE V
COMPARISON WITH THE COVID-QU-EX BASELINE METHOD

Method	Task	Dice	IoU
Tahir et al. (2021)	Lung segmentation	0.9799	0.9611
Ours	Lung segmentation	**0.9768**	–
Tahir et al. (2021)	Infection segmentation	0.8821	0.8305
Ours	Infection segmentation	**0.7957**	**0.7449**

obtained an infection localization Dice score of 88.21% and an IoU of 83.05%. :contentReferenceindex=0

Compared to these results, our lung segmentation model achieved a Dice score of 97.68%, which is close to the published result and demonstrates comparable segmentation quality. Although our infection segmentation result (Dice = 79.57%, IoU = 74.49%) is lower than that reported by Tahir et al., this difference is expected due to the inherent difficulty of infection segmentation and may also be influenced by differences in model design, training strategy, or data preprocessing.

Overall, our results are consistent with the trends observed in existing literature, and they confirm the effectiveness of using deep learning–based segmentation models on the COVID-QU-Ex dataset. The comparison also highlights that infection segmentation remains a challenging task and there is room for future improvement in this area.

V. CONCLUSION

In this work, we proposed a two-stage deep learning system for automatic COVID-19 analysis from chest X-ray images. The first stage focuses on lung segmentation to extract the lung region, while the second stage performs infection segmentation inside the predicted lung area. This design helps the model focus on relevant regions and reduces the influence of background noise.

Experiments were conducted on the COVID-QU-Ex dataset. The lung segmentation model achieved a very high Dice coefficient of 0.9768 on the test set, which shows that the lung regions can be segmented very accurately. The infection segmentation model achieved a Dice score of 0.7957 and an IoU of 0.7449. Although this task is more difficult due to unclear boundaries and large variations of infection regions, the results are still reliable and clinically meaningful.

The comparison with the baseline method reported in the COVID-QU-Ex study shows that our lung segmentation performance is comparable to state-of-the-art results, while the infection segmentation performance is reasonable given the simplicity of the proposed pipeline. These results confirm that accurate lung segmentation is an important factor for improving infection segmentation.

In the future, the performance of the infection segmentation stage can be further improved by using more advanced network architectures, stronger data augmentation, and better post-processing techniques. In addition, the proposed system can be extended to support severity grading and other clinical applications.

REFERENCES

[1] A. M. Tahir, M. E. H. Chowdhury, A. Khandakar, Y. Qiblawey, U. Khurshid, S. Kiranyaz, N. Ibtehaz, M. S. Rahman, S. Al-Madeed, S. Mahmud, M. Ezeddin, K. Hameed, and T. Hamid, "COVID-19 Infection Localization and Severity Grading from Chest X-ray Images," *Computers in Biology and Medicine*, vol. 139, p. 105002, 2021.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of MICCAI*, 2015, pp. 234–241.