Lung infection segmentation

BI12-497 Nguyen Xuan Minh Vu

March 2024

1 Introduction

X-rays are a form of electromagnetic radiation, similar to visible light. Unlike light, x-rays have higher energy and can pass through most objects, including the body. Medical X-rays are used to generate images of tissues and structures inside the body. If X-rays traveling through the body also pass through an X-ray detector on the other side of the patient, an image will be formed that represents the "shadows" formed by the objects inside of the body. The X-ray images that result from this process are called radiographs. The radiograph is crucial, specifically in the COVID-19 period.

During the COVID-19 pandemic, there was an overload in the number of patients infected coming to the hospitals. The most affected organ caused by this disease is the lung. Therefore, there are demands for rapidly detecting and segmenting the infection area inside the lungs. The manual process may take time and cannot respond to the needs of diagnosing various cases. The deep learning method is one of the most suitable solutions for this problem. Currently, post-COVID-19, we can utilize deep learning in the lung COVID-19 segmentation as the data is huge and various. This is not only for treating COVID-19 but for any diseases or pandemics coming in the future.

The segmentation task in medicine is crucial as doctors can accurately diagnose the position of infection and the stage of the disease. The traditional segmentation methods are time-consuming; however, the advancement in machine learning significantly reduces the processing time.

In this report, I try to implement COVID-19 infection segmentation in people's lung X-ray images using a deep learning method.

2 Background

2.1 COVID-19 Lung Damage

COVID-19 can cause lung diseases such as pneumonia and, acute respiratory distress syndrome (ARDS) or sepsis. In pneumonia, the lungs become filled with fluid and inflamed, leading to breathing difficulties. In more severe cases, ARDS is caused when the blood vessels leak fluid, leading to shortness of breath and lastly lung failure. On the other hand, sepsis occurs when an infection reaches, and spreads through the bloodstream, causing tissue damage.

2.2 Convolutional Neural Network (CNN)

CNN excels at processing data like images, speech, and audio signals. This is due to their unique architecture composed of three main layers: The convolutional layer is to extract information. The pooling layer reduces the dimensionality of the data and captures spatial variance. Lastly, the fully connected layer combines the extracted features from earlier layers and makes the final classification or prediction.

3 Dataset

In this project, I will use the COVID-QU-Ex dataset, provided by Qatar University 33,920 chest X-ray images, including 11,956 COVID-19 cases, 11263 Non-COVID but viral or bacterial pneumonia, and 10701 normal cases. A conventional radiography machine captures those images and presents the state of the organs inside the body. Each category includes three parts: X-ray image, lung mask image, and infection mask image.

4 Method

4.1 Data preprocessing

The images are resized into 256x256. The X-ray images are scaled using the MinMaxScaler method. They are fed into the model as the input. The annotation images are the masked images corresponding to the input. They include three areas: Non-lung area, lung area, and infection area. Their pixel values are set to 0, 125, and 255 respectively.

4.2 Network architecture

The network architecture is described in 1. The model is in the "U" shape. The left side will undertake the task of extracting features. It does the 3-by-3 convolutions and uses batch normalization and the ReLU activation function after each convolution layer. For down-sampling, the 2-by-2 MaxPool with stride = 2 will increase the channel size twice. The output of the left side is then fed to the right side to do the predicting process and produce the result of a segmented image. It uses 2-by-2 up-convolutions results concatenate with the corresponding layers on the left side and continue do the 3-by-3 convolution. In the last layer, the 1-by-1 convolution, BatchNorm, and Sigmoid are performed to achieve the segmented prediction with the same height and width as the original resized

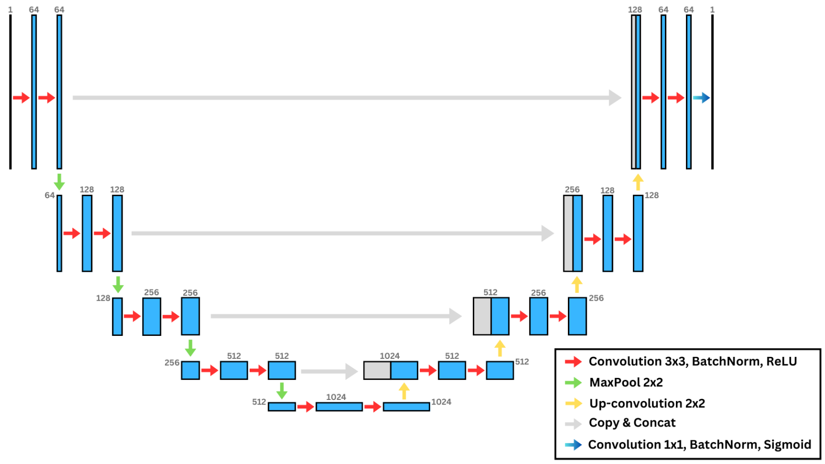


Figure 1: Overall Unet custom model

image.
The special of this architecture is that there is no fully connected layer. Therefore, the input size can be varied and the learning process can be faster.
There are some adjustments in the model compared with the original one to suit the current problems. Firstly, the Unet model uses copying and cropping to concatenate two blocks as the size is lost in the convolution steps. I only do the concatenation but at each convolution layer, I add the padding so that the size will not change in that process and I can get the last output with the same size as the input. Secondly, I use the BatchNorm method between the convolution and the ReLU, which makes the model faster and more stable as the data is continuously re-scaled and re-centered. Lastly, I use the Sigmoid activation function to the last layer to get a matrix with values between 0 and 1 to do the criterion.

4.3 Training

For training, I use the Adam optimizer algorithm with a learning rate = 0.0001, and weight decay = 0.00000001. For the loss function, I use the Binary Cross Entropy Loss. The batch size for training is set as 16 while validating is 8 in this work. The model runs through 100 epochs.

4.4 Data splitting

The training dataset has 3728 samples: 1864 COVID-19 cases, 932 non-COVID, and 932 normal cases. For the validation set, there are 932 samples with 466, 233, and 233 cases respectively. The test set includes 1164 images.

4.5 Data postprocessing

For the same format as the input image, the prediction should have three different pixel values 0, 125, and 255 representing three areas in the X-ray images. However, the prediction has various values of pixels so I will assign each pixel to the nearest one of three values 0, 125, and 255. The result of the postprocessing process is shown in 2.

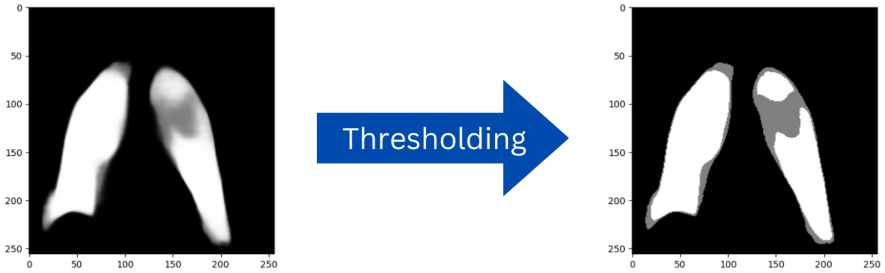


Figure 2: Postprocessing process illustration (Original output (left) and processed output (right))

5 Evaluation

5.1 Loss progression

The losses after 100 epochs are illustrated in 3. For the first 30 epochs, loss in training and validating are significantly decreased, reaching 0.15. After that, the validating loss is between 0.16 and 0.14 while the training loss still decreases.

5.2 Metrics

During this study, the Mean Square Error and the Mean Absolute Error metrics are used to evaluate the model before the post-processing process. To investigate the model after post-processing, we use precision, recall, F1 score, and Intersect over Union (IoU) metrics. These metrics are shown in Table 1.

Table 1: Evaluation metrics

Dataset	**Before post-processing**		**After post_processing**		
	MSE	**MAE**	**Precision**	**Recall**	**IoU**
Train data	0.00288	0.01394	0.9378	0.9361	0.9731
Validation data	0.00830	0.02438	0.8872	0.8863	0.9363
Test data	0.00873	0.02522	0.8856	0.8829	0.93337

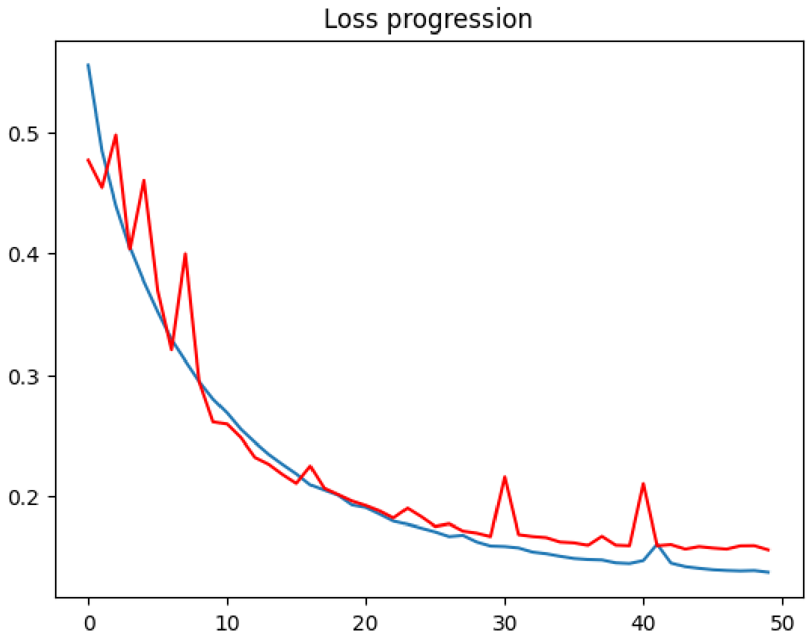
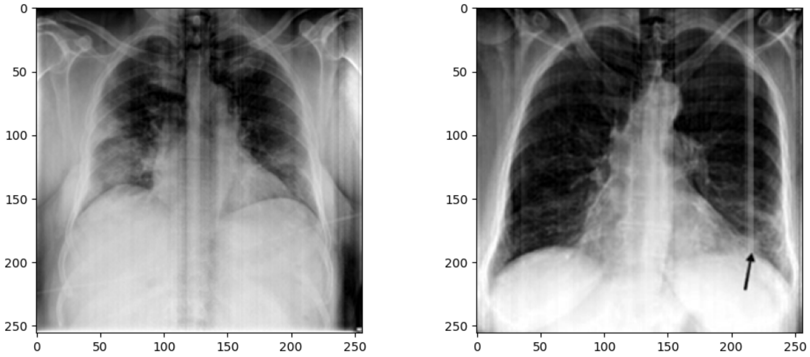


Figure 3: Losses in 100 epochs (blue: training loss, red: validating loss)

5.3 Result

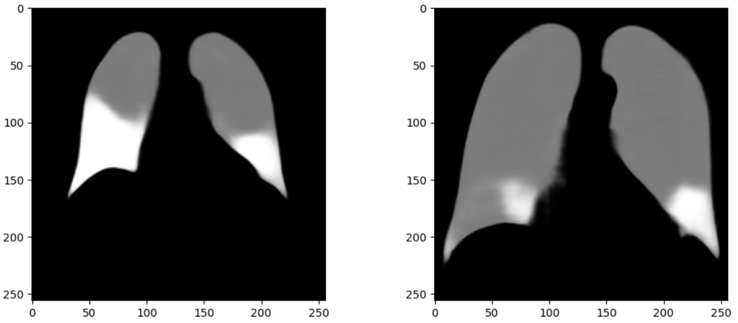
The results are computed using the testing set. Two images are shown in Fig.4. Fig.5, Fig.6, and Fig.7 visually illustrate the prediction and the annotation.



(a) Image 1

(b) Image 2

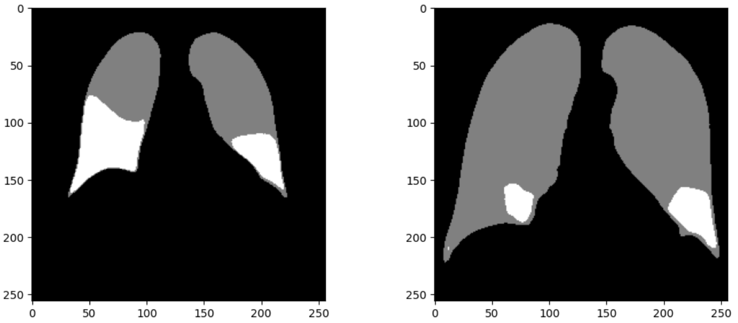
Figure 4: 2 random images from test dataset



(a) Image 1

(b) Image 2

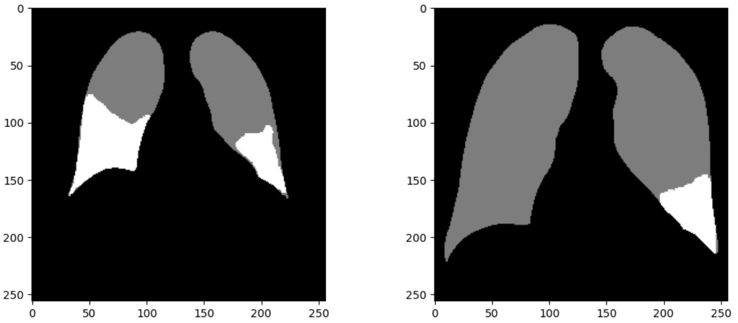
Figure 5: Segmentation through model



(a) Image 1

(b) Image 2

Figure 6: Postprocessing the previous segmentation



(a) Image 1

(b) Image 2

Figure 7: The respective annotation

6 Conclusion

In general, the model performs well on most of the data in both training and validating datasets. There are some cases that which the model cannot fully

detect the segment area as in the real image. However, the post-processing process can handle them and return a good 3-mask image.
I will continue this work in the future to improve the model, do with more datasets, and implement this model in other fields of medicine.