PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

An automatic approach to lung region segmentation in chest x-ray images using adapted U-Net architecture

Rahman, Md Fashiar, Tseng, Tzu-Liang (Bill), Pokojovy, Michael, Qian, Wei, Totada, Basavarajaiah, et al.

Md Fashiar Rahman, Tzu-Liang (Bill) Tseng, Michael Pokojovy, Wei Qian, Basavarajaiah Totada, Honglun Xu, "An automatic approach to lung region segmentation in chest x-ray images using adapted U-Net architecture," Proc. SPIE 11595, Medical Imaging 2021: Physics of Medical Imaging, 115953I (15 February 2021); doi: 10.1117/12.2581882

SPIE.

Event: SPIE Medical Imaging, 2021, Online Only

An Automatic Approach to Lung Region Segmentation in Chest X-Ray Images Using Adapted U-Net Architecture

Md Fashiar Rahman[a], Tzu-Liang (Bill) Tseng[a,b], Michael Pokojovy[a,c], Wei Qian[d], Basavarajaiah Totada[d], and Honglun Xu[a]

[a]Computational Science Program, UTEP, Texas 79968, El Paso, USA
[b]Department of Industrial, Manufacturing and Systems Engineering, UTEP, Texas 79968, USA
[c]Department of Mathematical Sciences, UTEP, Texas 79968, USA
[d]Department of Electrical and Computer Engineering, UTEP, Texas 79968, USA

ABSTRACT

Segmentation of the lung field is considered as the first and crucial stage in diagnosis of pulmonary diseases. In clinical practice, computer-aided systems are used to segment the lung region from chest X-ray (CXR) or CT images. The task of segmentation is challenging due to the presence of opacities or consolidation in CXR, which are typically produced by overlaps between the lung region and intense abnormalities caused by pulmonary diseases such as pneumonia, tuberculosis, or COVID-19. Recently, Convolution Neural Networks (CNNs) have been shown promising for segmentation and detection in digital images. In this paper, we propose a two-stage framework based on adapted U-Net architecture to leverage automatic lung segmentation. In the first stage, we extract CXR-patches and train a modified U-Net architecture to generate an initial segmentation of lung field. The second stage is the post-processing step, where we deploy image processing techniques to obtain a clear final segmentation. The performance of the proposed method is evaluated on a set of 138 CXR images obtained from Montgomery County's Tuberculosis Control Program, producing an average Dice Coefficient (DC) of 94.21%, and an average Intersection-Over-Union (IoU) of 91.37%.

Keywords: lung segmentation, deep learning, U-Net, chest X-ray

1. INTRODUCTION

Medical imaging is one of the most common and widely used diagnostic techniques employed in modern healthcare. For example, computer tomography (CT) and magnetic resonance imaging (MRI) are successfully used for screening diseases such as lung cancer and many other pulmonary diseases. Advances in medical imaging have greatly improved the accuracy of screening for these diseases at earlier stages. However, 3D medical imaging techniques are still rather expensive and their availability is limited in most parts of the world. Due to this fact, chest X-rays (CXR) remain the most popular and affordable imaging technique worldwide. CXR can provide helpful general insights and serve as an initial diagnostic tool. They are especially commonly used in diagnosing pneumonia, cancer, and other pulmonary diseases including COVID-19.[1,2] This year, researchers have intensively focused on developing new techniques to improve the accuracy of CXR-based detection, especially in connection with COVID-19.[2–4]

An important step in any CXR-based detection system is lung region segmentation, i.e., identification of the lung area in CXR.[5] It provides structural information about shape irregularities and size measurements of the lungs, which can be used to analyze severe clinical conditions. This suggests that good segmentation of the lung region is essential for accurate analysis and further conclusions. For healthy patients, computer-aided systems are typically able to provide reliable segmentation results because the contrast between the lung region and the outer area tends to be strong. However, in patients affected by pulmonary diseases, the lung region appears to be very obscure in CXR images due to the poor contrast between the lung area and the outer region. This primarily happens due to an overlap between the lung region and dense abnormalities caused by pneumonia,

Further author information: (Send correspondence to Md Fashiar Rahman.)
Md Fashiar Rahman: E-mail: mrahman13@miners.utep.edu, Telephone: 1 915 730 1007

Medical Imaging 2021: Physics of Medical Imaging, edited by Hilde Bosmans, Wei Zhao, Lifeng Yu, Proceedings of SPIE Vol. 11595, 115953I · © 2021 SPIE
CCC code: 1605-7422/21/$21 · doi: 10.1117/12.2581882

tuberculosis, COVID-19 or other symptoms, which are known as opacities and consolidations.[6] For this reason, the segmentation task in CXR becomes significantly more complex and challenging. Focusing on such scenarios, the purpose of this work is to present an automatic lung segmentation method that addresses this problem by implementing a two-stage framework based on the U-Net[7] architecture.

2. METHODOLOGY

The proposed framework is summarized in Fig. 1. The methodology is comprised of two major steps. In the first stage, we collect, pre-process and extract a number of image "patches" to train the adapted UNet architecture. The output of this stage is the initial segmentation. Subsequently, this initial segmented image is post-processed using image processing techniques to obtain the final segmentation. The details of these steps are discussed in the subsequent subsection.

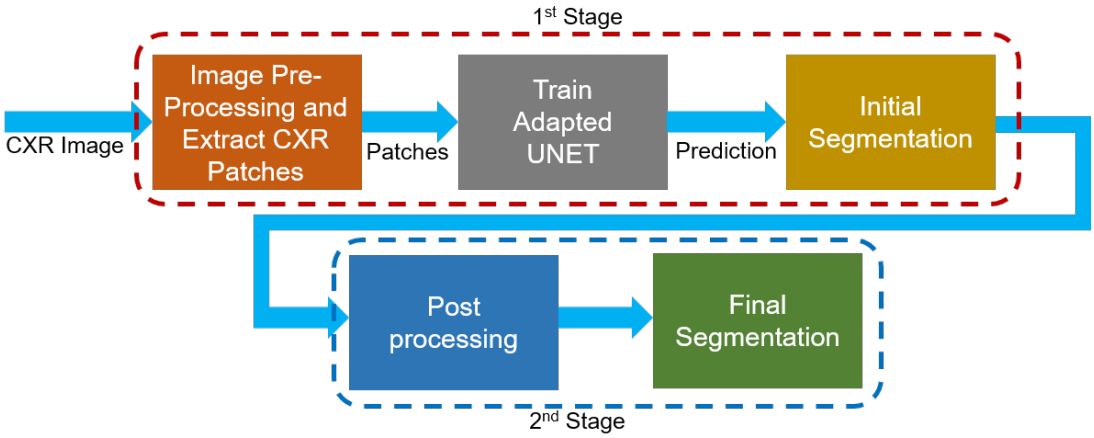

Figure 1. Our proposed two-stage framework for lung segmentation

2.1 Dataset description

The dataset used in this paper has been collected from Tuberculosis Control Program of he Department of Health and Human Services of Montgomery County (MC), Maryland, USA. This is one of the public datasets (https://lhncbc.nlm.nih.gov/publication/pub9931) used to benchmark the task lung segmentation algorithms. It contains 138 posterior-anterior X-rays, of which 80 X-rays are normal and 58 X-rays are abnormal with manifestation of tuberculosis. The X-rays are available in two different sizes: either 4020×4892 or 4892×4020 pixels. The dataset also comes with the corresponding ground truth (masks) which makes it suitable for supervised learning process. The masks are binary images consisting of white pixels for the lung region (left and right lungs) with the black background.

2.2 Pre-processing and Patch Extraction

Since the X-rays in MC dataset are relatively large, we first downscale the images to 1024×1024 to reduce the computational cost of training the model. As mentioned earlier, the images appear in two different sizes, direct resizing of these images could distort the original size proportion. To address this issue, we first place the image at the center of 4892×4890 background image and later downscale it to the final resolution of 1024×1024 pixels. After resizing, the X-ray pixel intensity is normalized[8] i.e., the pixel values are scaled into the range from 0 to 1, where 0's represent dark black and 1's represent white. The values between 0 and 1, represent the shades of gray. This is useful as the MC data may have different ranges of pixel values and normalizing them help to apply the same algorithms across the images.

MC dataset consists of only 138 X-rays which is insufficient to adequately train the deep learning model. Hence, we extract the patches from the X-ray images to increase the number of training samples. The patches are

small images, which are cropped from the original images. In this paper, we used the patch size of 64×64. The reason behind choosing this patch size is to fit the adapted UNet architecture while preserving the necessary levels of details of lung regions. It should be emphasized that, small patch sizes are insufficient for down-sampling in U-Net architecture and could potentially miss some useful information about the lung region.At the same time, larger patch sizes reduce the training samples. While cropping the patches from the X-ray images, we also cropped the patches from the corresponding mask images. This is done to preserve the ground truth for each of the patches in supervised training process.

2.3 Image Segmentation

In this work, complete X-ray images are segmented based on the patch segmentation technique. The patches are segmented using the U-Net architecture, which is widely applied in segmenting medical images.[9–11] The original U-Net architecture is utilized for comparatively larger images having the dimension of 512×512, 256×256 or 128×128. Usually, in most applications, the U-Net architecture consists of three to four downsample blocks to contract the input images to the size of 32×32 or 16×16 pixels. The contracted image is then up-sampled in the same manner. In this paper, our input image size is 64×64 (cropped patches) which is insufficient for contraction and expansion according to the original U-Net architecture. Hence, we truncated the original U-Net architecture to maintain a reasonable image size while downsampling the patches. The modified network consists of two down-sampling blocks. The patches are contracted to 8×8 pixels and later expanded through two up-sampling blocks as shown in Fig. 2.

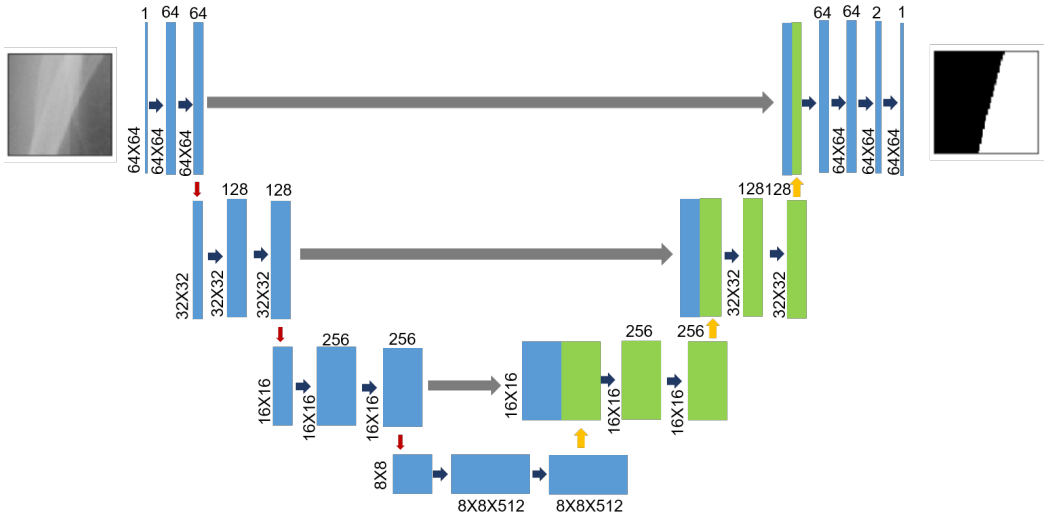

Figure 2. The adapted U-Net architecture for patch segmentation

The network is trained to segment individual patches, which are subsequently merged to obtain the initial segmentation for the whole X-ray image. This is accomplished by plotting the segmented patches on a 512×512 background image. More specifically, for any new image, we crop the patches and feed them as an input into the modified U-Net architecture. This generates segmented patches as the outputs, which are then placed back at the same location of the cropped region in the background image, and thus, produces the whole segmented image.

2.4 Post-processing

Since the whole X-ray image is segmented by merging individually segmented patches, the initial segmentation will invariably contained some noises and inaccuracies. These inaccuracies arise due to the prediction error for individual X-ray patches. For example, a background patch or some portion thereof could be erroneously segmented as lung region, as highlighted in Fig. 3 in red mark. Alternatively, a lung region could be segmented as a background patch, as shown in Fig. 3 in yellow. This issues are sometime inevitable, as some of the background

and lung region patches are difficult to distinguish when a whole X-ray image is subdivided into small patches. This problem is addressed in the second step referred to as the post-processing step. This step confines of several image processing techniques such as image thresholding, erosion, dilation, connected component labeling, and region filling algorithms. The reader is referred to Refs. 12 and 13 for a detailed description of these morphological processes.

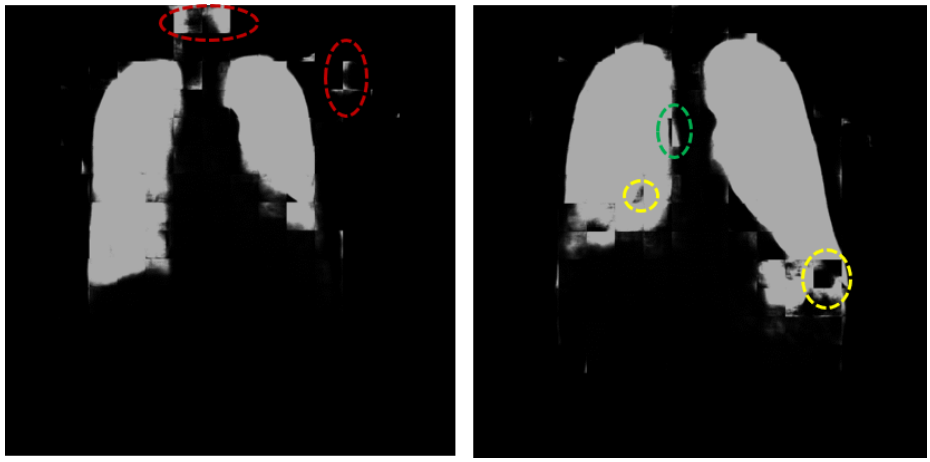

Figure 3. Illustration of noises and inaccuracies in initial segmentation

To make the post-processing uniform, we first use image thresholding to linearly transform all black pixels to 0 and white pixels to 255. Then the left and the right lung regions are separated by grouping them into two large connected regions. This is done to avoid any subsequent complication while deploying the erosion and dilation operation. After separating the lung region, all scattered small noisy "stains" are removed. However, there could remain some smalls region connected to the lung contour through tiny edges, as circled in Fig. 3 in green. To remove such kind of noises, the erosion operation is used to separate this small region and subsequently erase these regions. The size of the regions is determined by using the connected component labeling algorithm.[13] Deploying the erosion operation removes some pixels along the contour of the lung region which shrinks the size of the actual lung region. To reconstruct the contour, we apply the dilation operation. We set the kernel size of 7×7 both for erosion and dilation operations based on several experiments with the segmented images. At this point, the segmented images looks regular and clear without having any major noises. However, there still may exists some gaps (marked in yellow in Fig. 3) inside the segmented lung region. These gaps are closed by the morphological closing operation. The kernel size used in the closing operation was set to 23×23. At the end of post-processing step, we combine the two separate lung regions (left and right) to obtain the final segmentation.

3. EXPERIMENTAL DESIGN AND TRAINING

As described in Subsec. 2.1, we used 138 X-ray images from the MC dataset in our experiment. The dataset was divided as 70%-15%-15% for the training, validation and test purposes respectively. Thus, we have only 97, 21 and 20 images for training, validation and test respectively, which is insufficient to train a deep learning model, especially convolution neural networks (CNNs). Usually, CNNs require a significant number of samples in the training phase.[14] To increase the training sample, we use small patches from the original X-ray instead of whole images. Patches of size 64×64 are cropped from the images with the stride of 32 as described in Subsec. 2.2. The purpose of using the stride of 32 is to maintain overlapping in any two successive patches. Based on this patch extraction strategy, we finally get a total of 141,312 patches, among which 99,328 are training patches, 21,504 are validation patches and 20,480 are testing patches. We also have the same number of mask patches as reported in Tab. 1.

The model is trained for 55 time epochs. We use the adaptive learning rate optimization algorithm (Adam[15]) and binary cross-entropy loss function to train the network. To avoid any possible overfitting issues, the data

Table 1. Split the dataset for training, validation and testing.

Dataset	Training	Validation	Test	Total
MC Dataset	97	21	20	138
X-ray patches	99,328	21,504	20,480	141,312
Mask patches	99,328	21,504	20,480	141,312

were augmented using the Keras data generation function. We apply 5% shearing and zooming, and apply the horizontal flipping to each batch of patches. Here, the batch size was set to 64. The model is trained on a 3.6 GHz Intel(R) Core (TM) i7-7700 CPU, 16GB RAM and a single NVIDIA GTX 1080 Ti Graphics Processing Unit (GPU). It took 8 hr 20 min to complete the entire training. After training over 55 epochs, we obtained the training accuracy of 98.65% and validation accuracy of 97.89%. Notice that, here, the accuracy refers to the segmentation of small patches, instead of the whole X-ray images. The training accuracy and loss are shown in Fig. 4

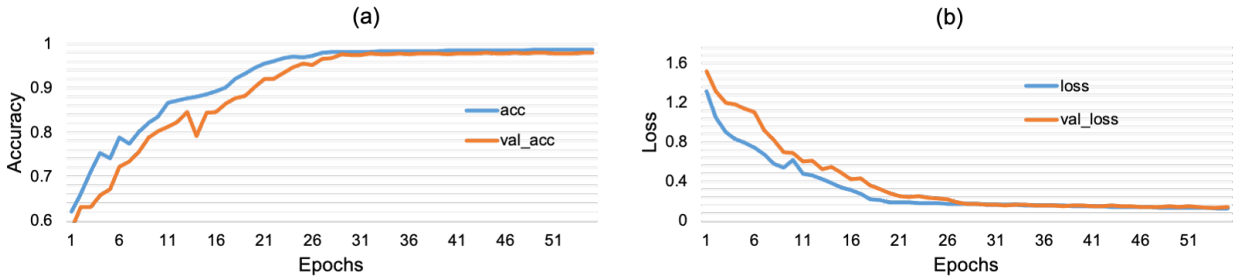

Figure 4. Training history (a) training accuracy (b) training loss

After completing the training process, we tested using individual patches with the trained model. Some segmentation samples are shown is shown in Fig. 5.

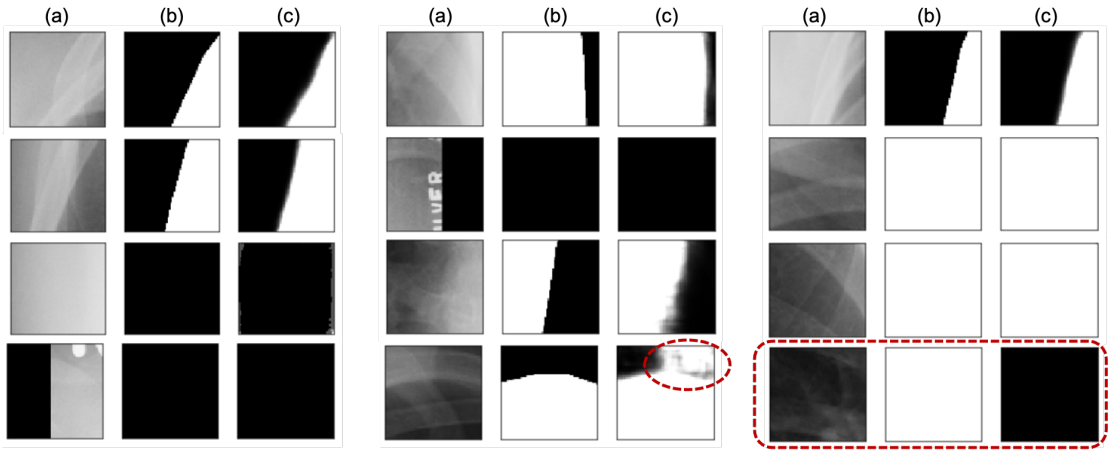

Figure 5. Segmentation result of x-ray patches (a) original x-ray patches (b) ground truth mask (c) predicted mask

Obviously, the trained model can segment the patches with moderate accuracy. However, some segmented masks appeared to be incorrect, as circled in Fig. 5. We use this trained model to predict the "labels" patches for any particular test image and later merge them together to obtain the initial segmentation for the whole X-ray image.

4. RESULT AND DISCUSSION

This section describes the segmentation result for the whole X-ray images. As mentioned earlier, to obtain the complete segmentation, we first crop the patches (64×64) from any particular image. The patches are labelled using the model trained in Sec. 3. Later, the predicted masks are plotted and merged together to generate the initial segmentation. Some segmentation results are shown in Fig. 6 (column (b)).

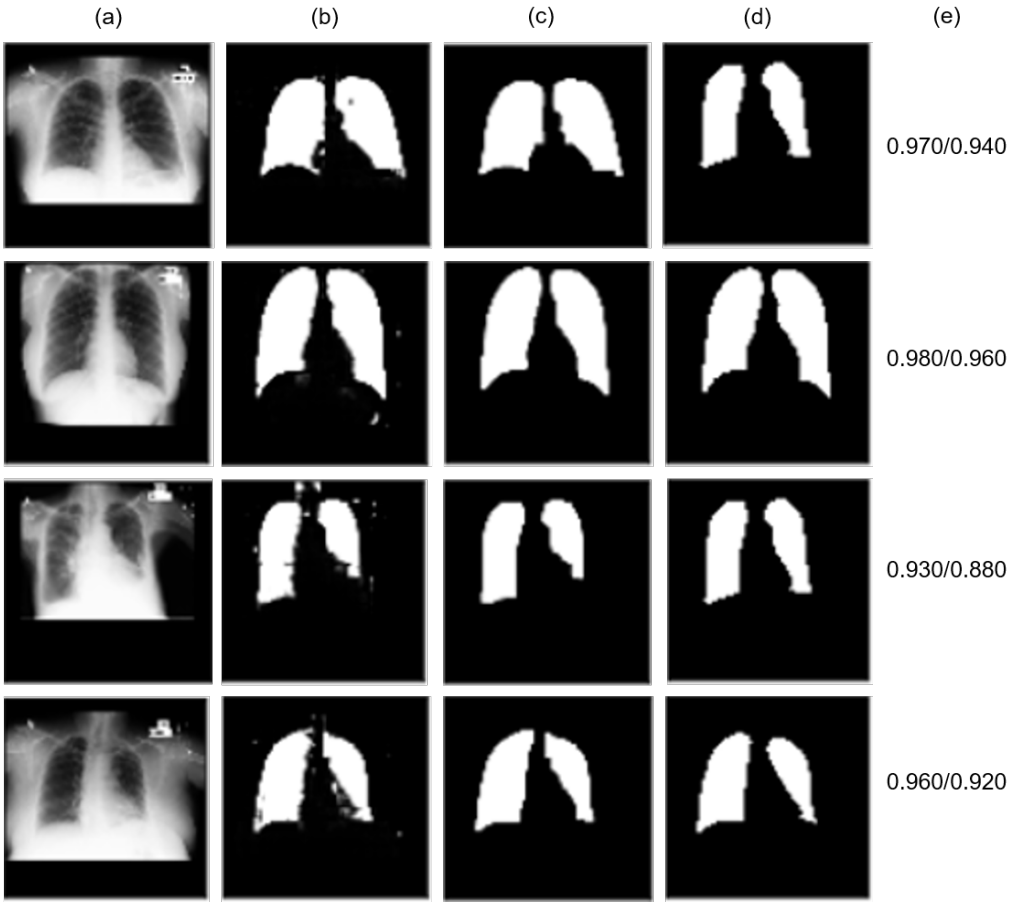

Figure 6. Segmentation result for the whole X-ray (a) original X-ray (b) segmentation before post-processing (c) segmentation after post-processing (d) ground truth mask (e) DC/IoU value

It is observed that, integrating the segmented patches can generate excellent segmentation results with minimal error in good scenarios, where the lung region is visually apparent (1st and 2nd row in Fig. 6). In problematic cases as shown in the 3rd and 4th rows in Fig. 6, the segmentation results are suboptimal due to the presence of unavoidable prediction errors, such as tiny scattered white pixels, extraneous non-lung region connected with the lung boundary or even some little gaps in the lung segmented region. These issues are resolved using the post-processing step as described in Subsec. 2.4. The segmentation results after the post-processing are shown in Fig. 6 (column (c)).

Segmentation results are quantitatively evaluated with two performance metrics - Dice Coefficient (DC) and Intersection-Over-Union (IoU). These two metrics are mostly used to evaluate the performance of semantic segmentation.[16] IoU is the area of the overlap between the predicted segmentation (seg_p) and the ground truth (seg_{gt}) divided by the area of union between the predicted segmentation and the ground truth. It is defined as

$$IoU = \frac{\text{area}(\text{seg}_\text{p} \cap \text{seg}_\text{gt})}{\text{area}(\text{seg}_\text{p} \cup \text{seg}_\text{gt})}, \tag{1}$$

The dice Coefficient is two times the area of the overlap between the predicted segmentation and the ground truth divided by the sum of the area of the predicted segmentation and the ground truth. It is expressed as

$$DC = \frac{2 \cdot \text{area}(\text{seg}_\text{p} \cap \text{seg}_\text{gt})}{\text{area}(\text{seg}_\text{p}) + \text{area}(\text{seg}_\text{gt})}\,, \tag{2}$$

The value of both metrics ranges from 0 to 1 (0-100%) with 0 representing no overlap and 1 indicating perfect overlapping segmentation. We calculated the DC and IoU values for all of test X-ray images. Then, we averaged them and obtained the overall DC value of 94.21% and IoU value of 91.37%. Fig. 6 (column (e)) shows the DC and IoU values for corresponding segmented images.

5. CONCLUSION

In this work, we proposed a two-stage framework based on adapted U-Net architecture and image processing techniques. In the first stage, we extracted small patches from the original images and used them to train the adapted U-Net architecture. The U-Net architecture was modified to utilize small cropped patches. The second stage is mainly represented by a post-processing step, which includes some traditional image processing techniques such as erosion, dilation, connected component labelling and region-fill algorithms. The combination of two stages was shown to generate promising lung segmentation results. It is worth to emphasize that, this framework can efficiently utilize small datasets for training convolution neural networks. In our experiment, we achieved the value of 94.21% and 91.37% for Dice coefficient and intersection-over-union, respectively. In the future, we intend to extend and improve our methodology to utilize limited number of x-ray for the COVID-19 cases and thus, contribute to improved X-ray based diagnosis.

REFERENCES

[1] Jaeger, S., Candemir, S., Antani, S., Wáng, Y.-X. J., Lu, P.-X., and Thoma, G., "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery* **4**(6), 475 (2014).

[2] Mahmud, T., Rahman, M. A., and Fattah, S. A., "Covxnet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization," *Computers in Biology and Medicine* **122**, 103869 (2020).

[3] Wang, L., Lin, Z. Q., and Wong, A., "Covid-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images," *Scientific Reports* **10**(1), 1–12 (2020).

[4] Zhang, J., Xie, Y., Li, Y., Shen, C., and Xia, Y., "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," *arXiv preprint arXiv:2003.12338* (2020).

[5] Qin, C., Yao, D., Shi, Y., and Song, Z., "Computer-aided detection in chest radiography based on artificial intelligence: a survey," *Biomedical Engineering Online* **17**(1), 113 (2018).

[6] Skoura, E., Zumla, A., and Bomanji, J., "Imaging in tuberculosis," *International Journal of Infectious Diseases* **32**, 87–93 (2015).

[7] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 234–241, Springer (2015).

[8] Pei, S.-C. and Lin, C.-N., "Image normalization for pattern recognition," *Image and Vision computing* **13**(10), 711–723 (1995).

[9] De Moor, T., Rodriguez-Ruiz, A., Mérida, A. G., Mann, R., and Teuwen, J., "Automated lesion detection and segmentation in digital mammography using a u-net deep learning network," in [*14th International Workshop on Breast Imaging (IWBI 2018)*], **10718**, 1071805, International Society for Optics and Photonics (2018).

[10] Shaziya, H., Shyamala, K., and Zaheer, R., "Automatic lung segmentation on thoracic ct scans using u-net convolutional network," in [*2018 International Conference on Communication and Signal Processing (ICCSP)*], 0643–0647, IEEE (2018).

[11] Le, H. T. and Pham, H. T.-T., "Brain tumour segmentation using u-net based fully convolutional networks and extremely randomized trees," *Vietnam Journal of Science, Technology and Engineering* **60**(3), 19–25 (2018).

[12] Gil, J. Y. and Kimmel, R., "Efficient dilation, erosion, opening, and closing algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(12), 1606–1617 (2002).

[13] Park, J.-M., Looney, C. G., and Chen, H.-C., "Fast connected component labeling algorithm using a divide and conquer technique.," *Computers and Their Applications* **4**, 4–7 (2000).

[14] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *nature* **521**(7553), 436–444 (2015).

[15] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization in: Proceedings of the 3rd international conference for learning representations (iclr'15)," *San Diego* (2015).

[16] Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., and Blaschko, M. B., "Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 92–100, Springer (2019).