

Automatic Lung Segmentation in Chest X-ray Images Using Improved U-Net

Wufeng Liu (✉ lwf@haut.edu.cn)

Henan University of Technology

Jiixin Luo

Henan University of Technology

Yan Yang

Henan University of Technology

Wenlian Wang

Nanyang Central Hospital

Junkui Deng

Nanyang Central Hospital

Liang Yu

Henan University of Technology

Research Article

Keywords: Deep Learning, Chest X-ray, Lung Segmentation, Semantic Segmentation, U-Net

Posted Date: August 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-842374/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Automatic Lung Segmentation in Chest X-ray Images Using improved U-Net

Wufeng Liu^{1,*}, Jiaxin Luo¹, Yan Yang¹, Wenlian Wang^{2,+}, Junkui Deng^{2,+} and Liang Yu¹

¹Henan University of Technology, Zhengzhou 450001, China

²Nanyang Central Hospital, Nanyang 473009, China

*lwf@haut.edu.cn

+these authors contributed equally to this work

ABSTRACT

Automatic and highly accurate lung segmentation in chest X-ray (CXR) images is the basis of computer-aided diagnosis systems, because the lung is the region of interest of many diseases, and it can show useful information through its contours. However, automatic lung segmentation is immensely challenging due to extreme variations in the shape, obscure lung area, or opacity caused by lung diseases reaches high-intensity values. In the face of these severe situations, the model may segment the lung boundary incorrectly. We designed an improved U-Net network: using the pre-training Efficientnet-b4 as the encoder, and the residual block and LeakyRelu activation function are used in the decoder. The network can not only extract features with high efficiency but also avoid the gradient explosion caused by the multiplication effect in gradient backpropagation. We constructed a CXR lung field segmentation dataset (Haut) based on the NIH CXR dataset. In particular, this lung segmentation dataset contains some serious abnormal cases, such as lung deformation, pleural effusion, covered by foreign matters, or CXR blur caused by severe lung disease. The improved U-Net is evaluated on Haut, JSRT, and Montgomery County (MC) datasets. Experimental results show that our network can achieve high-precision lung segmentation.

Index Terms—Deep Learning, Chest X-ray, Lung Segmentation, Semantic Segmentation, U-Net

Introduction

Among the existing medical imaging methods, X-ray is one of the most commonly used diagnostic technology as it is widely available, low cost, non-invasive, and easy to acquire^{1,2}. Chest radiography is the most popular and important imaging modality used in the diagnosis of various pulmonary diseases, such as tuberculosis, lung cancer, emphysema, atelectasis, pneumothorax, etc. The application of deep learning in the field of medical imaging can help medical experts carry out screening and diagnosis and reduce their burden. Although CXR images are widely used and are a research hotspot in recent years, it still is one of the most challenging medical images to interpret³.

The significance of automatic lung segmentation is that it can assist radiologists in the diagnosis of related diseases. After surgery, physicians can diagnose the recovery status of the patient via the size of lung regions in CXR images to determine the success of the surgery. Segmentation of the lung becomes challenging due to several reasons: 1. Non-pathological changes: the shape and size of the lung vary with age, gender, and heart size; 2. Pathological changes: the opacity caused by severe lung disease reaches a high-intensity value, which may be misinterpreted as the lung boundary by the automatic method⁴; 3. Foreign body coverage, such as the lung area is obscured by the patient's clothes or medical equipment (pacemaker, infusion line, medical catheter)⁵.

According to the investigation, few studies have evaluated the performance of the deep learning model in lung segmentation of CXR images with serious abnormal lesions, obscured by foreign matters, distortion, and deformation. Most of the reported work on lung segmentation is based on mild lesions or healthy CXR. Therefore, it is necessary to verify the ability of the lung segmentation model on complex CXR images. we propose a lung segmentation dataset and semantic segmentation framework. Our deep learning framework is open-source on GitHub (<https://github.com/2112942597/2985>). We randomly screened 2785 CXRs from the NIH (National Institute of Health) Chest X-ray dataset⁶ (<https://www.kaggle.com/nih-chest-xrays/data>) and invited experienced radiologists to manually label the lung area. In particular, these 2785 images contain some severe lung diseases, and foreign bodies obscure the lungs. In addition, we also designed an excellent lung field semantic segmentation model, which is structured by U-net⁷ and uses the Efficientnet-b4 pre-training model as the backbone. We also performed experiments on the well-known benchmark datasets. such as the Japanese Society of Radiological Technology(JSRT)⁸; Montgomery County(MC)⁹, USA.

In related literature, a large number of methods have been proposed for automatic lung segmentation. These methods have a wide application prospect. It can be divided into two categories: 1. Unsupervised classical methods 2. Supervised deep neural network technology. These classical lung segmentation techniques do not rely on the dataset labeled by professional radiologists, so they are easy to implement. but the lung boundaries obtained may not be optimum due to the heterogeneity of lung field shapes. The accuracy of this kind of algorithm is unstable, and its accuracy is far lower than that of neural network modeling^{3,10}.

With the development of computing power and the continuous enrichment of datasets, in the past few years, deep learning

technology has been used in medical image analysis and achieved good results¹¹⁻¹³. In related literature, the deep neural network is widely used in disease detection, target detection, and medical image segmentation, and has achieved good results¹⁴. In semantic segmentation technology, the chest radiograph is used as the input of a deep neural network, which classifies each pixel into lung region or non-lung region. Hwang, S et al¹⁵. proposed network-based atrous convolutional layers for accurate lung segmentation. Their algorithm was tested on JSRT⁸ and Montgomery County (MC) datasets⁹, and the dice coefficients were 0.9800 and 0.9640, respectively. Rahul et al¹⁶. used full convolution neural networks to segment the lung field of JSRT and MC datasets, with an average accuracy of 98.92% and 97.84% respectively. Mittal, A et al.¹⁷ modified the upsampling part of the famous SegNet architecture¹⁸ and trained it on the JSRT dataset. When tested on JSRT and MC datasets, their model achieves 98.73% accuracy. Ngo, T. A et al.¹⁹ propose a new methodology for lung segmentation in CXR using a hybrid method based on a combination of distance regularized level set and deep structured inference. Their methodology produces an average accuracy on JSTR that varies from 94.8% to 98.5%. The above scholars' research focuses on using better neural networks to achieve high-precision lung segmentation.

Rashid, R et al²⁰. proposed a complete convolution network for lung segmentation. The average accuracy on JSRT, MC, and local data sets are 97.1%, 97.7%, and 94.2% respectively. Ching-Sheng Chang et al's work is based on images as part of the NIH Chest X-ray dataset, randomly select some pictures for manual annotation²¹. They achieved a 94.9 % Jaccard index score. Some scholars combine neural networks with computer graphics morphology to eliminate some fragments produced by automatic lung segmentation. Ching-Sheng Chang et al. used encoder-decoder nets with atrous convolutions. Souza, J. C et al. designed an automatic lung segmentation and reconstruction method based on a depth neural network²². Based on the deep neural network, Lamin Saidy et al. introduce the knowledge of graphic morphology to solve the problem of fragments in lung segmentation²³.

Metrics

Following are the five segmentation performance metrics we use: Accuracy, Sensitivity, Specificity, Dice Coefficient, Jaccard Index. Semantic segmentation can be regarded as pixel-level classification. True Positive (TP): The model prediction is a positive example, which is actually a positive example. False Positive (FP): The model prediction is a positive example, but it is actually a negative example. False Negative (FN): The model prediction is a counterexample, but it is actually a positive example. True Negative (TN): The model prediction is a counterexample, which is actually a counterexample.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1) \quad \text{Specificity} = \frac{TN}{TN+FP} \quad (2) \quad \text{Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{Dice} = \frac{2TP}{2TP+FP+FN} \quad (4) \quad \text{Jaccard Index} = \frac{TP}{TP+FP+FN} \quad (5)$$

Results

Haut dataset contains 1647 normal individuals and 1138 patients CXR's lung field masks, including 193 with Infiltration, 111 with Atelectasis, 78 with Effusion, 65 with Nodule, 54 with Mass, 43 with Pneumothorax, 37 with Cardiomegaly, 37 with Pleural Thickening, 34 with Fibrosis, 25 with Consolidation, 21 with Emphysema, 11 with Edema, 10 with Pneumonia, 2 with Hernia, and 417 with Multiple diseases (including any two or more diseases above). For the JSRT dataset, our model with a pre-trained Efficientnet-b4 base network achieved the accuracy of 98.5%, 98.5% of specificity, 98.4% of sensitivity, 97.9 % of Dice coefficient, and 95.8 % of the Jaccard Index using improved U-Net. For the MC dataset, our model got an accuracy of 98.9%, 99.3% of specificity, 97.5% of sensitivity, 97.7% of dice coefficient, and 95.5% Jaccard index. U-net⁷ with a pre-trained Efficientnet-b4 base network provides advanced performance on the public datasets. Our model with a pre-trained Efficientnet-b4 base network obtained an accuracy of 99.4% on the Haut dataset with 99.5 % of specificity, 99.1% of sensitivity, 98.8% of dice coefficient, and 97.7% of Jaccard index with pre-trained Efficientnet-b4 base network. which is very encouraging and establishes the efficiency of our method. It also proves the effectiveness of our lung segmentation framework.

Comparison with other scholars

We summarized the previous studies of scholars and found that their work needs to be supplemented by later scholars. Firstly, most scholars are based on the JSRT and MC datasets, which do not contain lung segmentation in complex cases (severe pneumonia, foreign body shielding, lung deformation, etc.) Of course, some scholars try to label the NIH Chest X-ray dataset for lung segmentation²¹. But they do not verify the segmentation performance of the model on the benchmark dataset and do not summarize the segmentation scores of different CXR images. Our work complements these defects. To connect with the mainstream research on lung segmentation, we also did a series of experiments on JSRT and MC. The following table is the research on lung segmentation by scholars in recent years and the results of this experiment. In general, the performance of our lung segmentation network is comparable to that of the excellent lung segmentation network proposed in the literature in recent years. That also encourages us to further use the network to evaluate the lung segmentation performance of the Haut dataset.

Literature (Year)	Dataset	Evaluation On Severely Unhealthy CXRs	Method	Performance evaluation				
				Accuracy	Specificity	Sensitivity	Dice Coefficient	Jaccard Index
Ngo et al. ¹⁹ (2015)	JSRT	No	Deep learning + level set method	98.5%	–	–	99.2%	98.5%
Hwang et al. ¹⁵ (2017)	JSRT & MC	No	Atrous Convolutions + Network-wise Training of CNN	–	–	–	98% on JSRT & 96.4% on MC	96.1% on JSRT & 94.1% on MC
W Yang et al. ²⁴ (2017)	JSRT & MC	No	Structured Edge Detector	–	–	–	97.6% on JSRT & 95.6% on MC	95.8% on JSRT & 93.5% on MC
Saidy et al. ²³ (2018)	JSRT	No	Encoder-Decoder Convolutional Network	–	99.25%	95.26%	95.95%	–
Hooda et al. ¹⁶ (2018)	JSRT & MC	No	Improved Fully-Convolutional Network	98.92% on JSRT & 97.84% on MC	–	–	–	95.88% on JSRT & 91.74% on MC
CMittal et al. ¹⁷ (2018)	JSRT & MC	No	Improved SegNet	98.73%	–	–	–	95.10%
Rashid et al. ²⁰ (2018)	JSRT & MC & private	No	U-net	97.1% on JSRT & 97.7% on MC & 94.2% local	98% on JSRT, 98.5% on MC & 97% local	95.1% on JSRT 95.4% on MC & 86.2% local	95.1% on JSRT 95.4% on MC & 88% local	–
Souza et al. ²² (2019)	MC	No	AlexNet for patch classification + reconstruction based on ResNet18	96.97%	96.79%	97.54%	94%	88.07%
CS Chang et al. ²¹ (2020)	NIH	Yes	Encoder-Decoder nets with atrous convolutions.	–	–	–	–	94.9%
Y.M. et al. ²⁵ (2020)	JSRT & MC	No	Dense-Unet	–	98.8% on JSRT & 99.2% on MC	97.9% on JSRT & 98.1% on MC	97.6% on JSRT & 97.9% on MC	95.3% on JSRT & 95.9% on MC
The proposed work	JSRT	Yes	U-net is used as the architecture and the pre-trained Efficient net-b4 is used as the backbone (encoder) , Residual block and LeakyRelu are used for upsampling (decoder)	98.5%	98.5%	98.4%	97.9%	95.8%
	MC			98.9%	99.3%	97.5%	97.7%	95.5%
	Haut			99.4%	99.5%	99.1%	98.8%	97.7%

Table 1 . Comparison of results of the proposed method and recently related works.

Lung segmentation in benchmark datasets (JSRT&MC)

The JSRT dataset⁸ is created by the Japanese Society of Radiological Technology in collaboration with the Japanese Radiological Society. It contains 247 PACXRs of 2048×2048 resolution. Among 247 images, 93 are normal and 154 are abnormal having different TB manifestations. These images are stored in PNG format with 2048 × 2048 pixels having 12 bits grayscale. The abnormality of images is graded from extremely subtle to obvious. In our work, we used Segmented Chest Radiograph (SCR) dataset¹⁰ as ground truth corresponds to the JSRT dataset consists of manually delineating boundaries of lungs, heart, and clavicles.

The Montgomery County (MC) dataset is created by the Department of Health and Human Services, Montgomery County, Maryland, USA. The dataset contains 138 CXRs including 80 healthy cases and the remaining 58 are cases of tuberculosis. They can also be made available in Dicom format upon request. The size of the X-rays is either 4,020×4,892 pixels.⁹

The network was trained using two-third of the images, in which 20% of the data were reserved for validating the training process and tuning the models, and the image size is adjusted to 256 * 256.

In general, our model achieves excellent segmentation scores in dealing with two benchmark data sets (mild disease, no foreign body occlusion, high image quality). That shows the reliability of our data set and model. However, since these two public datasets do not contain complex chest radiographs, we also need to verify the ability of the model to process complex chest radiographs on Haut datasets. The Jaccard Index is an extremely important metric to evaluate our method because it represents the rate of lung pixels correctly segmented, which is directly related to the objective of our work. Data enhancement techniques are used to generate new images to compensate for the limited size of the dataset. HorizontalFlip, Cropping, and Rotation are transformations used to generate new images.

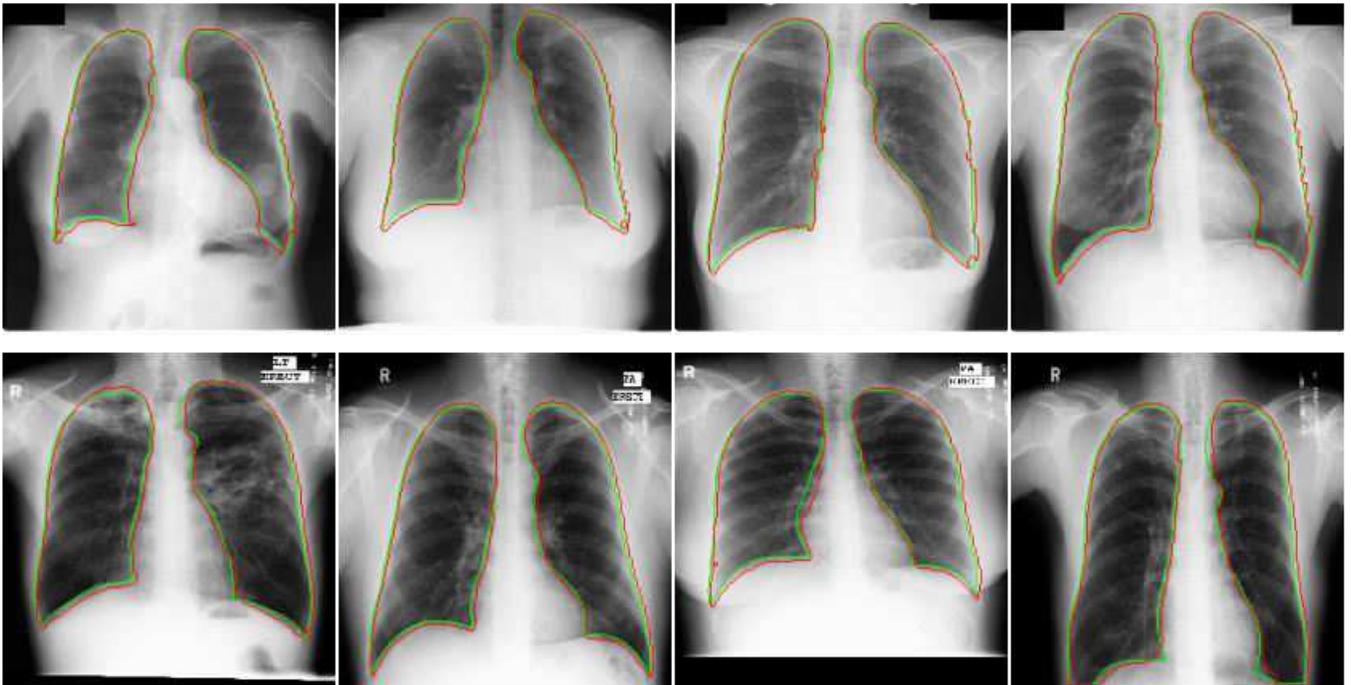


Figure 1. As shown in the picture, the first row is the lung segmentation result in the JSRT dataset, and the second row is the segmentation result in the MC dataset. On these two public datasets, our segmentation model has no obvious errors, and the predicted lung field is the same as the real field. The ground-truth lung boundary is depicted by green, and the automatically segmented lung boundary by our method is presented by red color.

Lung segmentation in complex case (Haut)

Compared with the two benchmark datasets, our Haut dataset contains more complex and diverse CXR images. Through the above comparison, our dataset segmentation model has achieved excellent results on two benchmark datasets. The following figure shows the performance of our lung segmentation model in CXR images under different conditions, including clear lung field, fuzzy lung field, lung field blocked by foreign bodies, and lung field with segmentation failure.

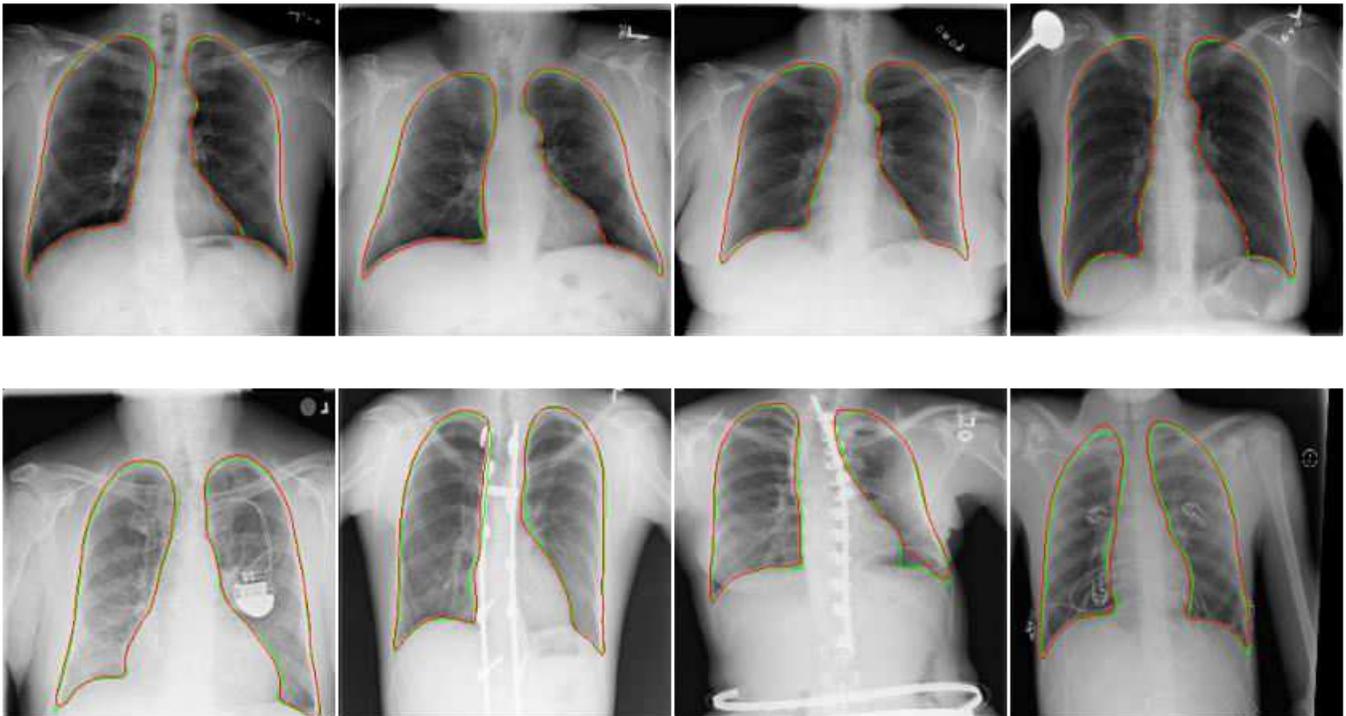


Figure 2. As shown in the figure above, we comprehensively evaluated the Haut dataset. Green represents the real lung field and red represents the lung field predicted by the model. The first line belongs to healthy or mild symptoms, and the effect of lung segmentation is very good. The second line is that there are foreign bodies (various medical devices) blocking the lung area, and the segmentation effect is relatively poor.

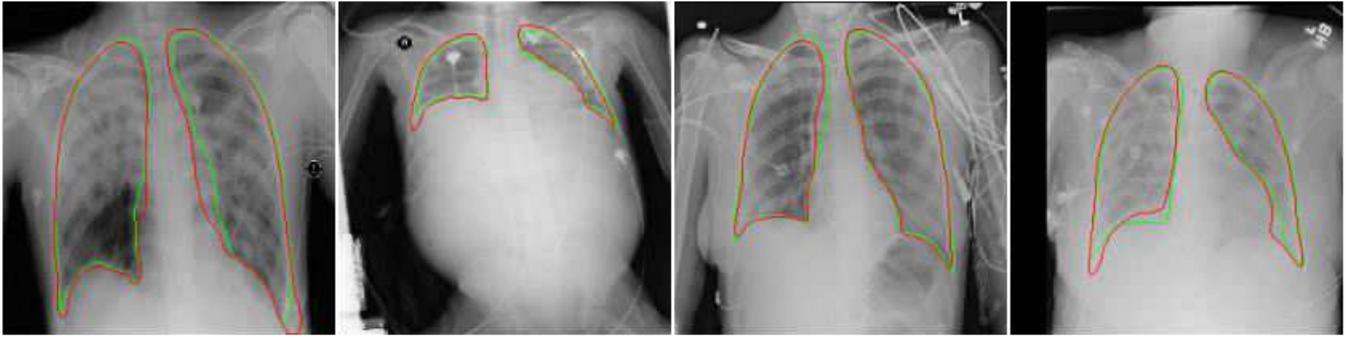


Figure 3. The above is the result of lung segmentation in severe disease (blurred lung area caused by disease) and distorted lung. In those cases, the lung segmentation score is the lowest. The ground-truth lung boundary is depicted by green, and the automatically segmented lung boundary by our method is presented by red color.

Disease (case number)	Multiple diseases (417)	Infiltration (193)	Atelectasis (111)	Effusion (78)	Nodule (65)	Mass (54)	Pneumothorax (43)	Cardiomegaly (37)	Pleural Thickening (37)	Fibrosis (34)
Jaccard Index	96.7%	97.1%	96.9%	96.4%	97.8%	97.4%	97.6%	97.7%	97.8%	97.6%
Disease (number)	Consolidation (25)	Emphysema (21)	Edema (11)	Pneumonia (10)	Hernia (2)	NO Finding (1647)	Sum (2785)			
Jaccard Index	96.3%	97.0	96.9	97.5	98.5	97.7%	97.4%			

Table 2. CXR images with different cases have different segmentation scores. Multiple diseases mean that a CXR image with two or more diseases.

Discussion

In this study, we evaluated the efficacy of our model for lung segmentation on the JSRT, MC, and Haut datasets. Five segmentation performance indexes: Accuracy, Sensitivity, Specificity, Dice coefficient, and Jaccard index are used to evaluate the model, we achieved excellent lung segmentation results. The excellent segmentation score shows the reliability of our segmentation model.

It is found that the transparency of the lung region, whether there is occlusion, and the shape of the lung will affect the results of lung segmentation to varying degrees. As can be seen in Figure 2, it is difficult for the model to distinguish the lung region and lung boundary under the turbidity of the lung region caused by serious lung diseases. In addition, abnormal lung morphology is also difficult to segment. This is consistent with the results of other scholars.

The automatic lung segmentation model performs poorly in processing images of some diseases, such as pulmonary consolidation, lung effect, lung edema, and atelectasis. These diseases will make a large number of exudates (tissue fluid, fibrin, etc.) fill the alveolar cavity and pleural cavity, resulting in lung densification and turbidity. It seriously affects the texture of the lung region in CXR images, so that the automatic lung segmentation model may misinterpret these textures.

In addition, the automatic lung segmentation model is poor in dealing with severe lung deformation caused by congenital or acquired factors.

A Singh et al.²⁶ recently published their lung segmentation study. Their scores far exceed those of previous scholars. But their data is absurd. Generally speaking, the Jaccard index is smaller than the Dice coefficient. But their result is just the opposite, which is very suspicious. So we didn't compare their experimental data.

Methods

In image segmentation tasks, especially medical image segmentation, U-Net⁷ is undoubtedly one of the most successful methods. Compared with FCN²⁷, SegNet¹⁸, and DeeplabV3+²⁸, U-Net uses skip connection in the same stage instead of direct supervision and loss back transmission on high-level semantic features, which ensures that the finally recovered feature map integrates more low-level features, and also enables the fusion of features of different scales, Thus, multi-scale prediction and deep supervision can be carried out. upsampling also makes the information such as the restored edge of the segmented image finer. A challenge of deep learning for medical image processing is that it often provides few samples, and U-Net still performs well under this limitation. Based on these advantages, we choose U-Net as the framework of the automatic lung segmentation model. Our further experiments show that U-net with Efficientnet-b4 always achieves the best results. The input size of the model is 256 * 256 * 3 and the output size is 256 * 256 * 1. Our experiment with Imagenet's pre-trained base networks. The network architecture used in this work has five coding layers and five decoding layers. The encoder is composed of efficientnet-b4 (We try different pre training networks as the backbone, and the effect of Efficientnet-b4 is the best).

The decoder consists of five blocks, Each decoding layer includes a dropout layer, two-dimensional convolution and padding

layer, and finally two residual blocks and a LeakyReLU. The function of the dropout layer is to improve the generalization ability of the model and prevent the model from overfitting. The two-dimensional convolution layer continues to extract image information. Two residual blocks²⁹ can prevent gradient disappears and make information spread better.

Residual block is the most important module in Resnet³⁰. The method of the residual block is to add a quick connection between the input and output of some network layers. The quick connection here defaults to identity mapping. In other words, it directly adds the original input and output without any change. The deeper the network is, the more obvious the gradient disappears, and the training effect of the network will not be very good. But now the shallow network can not significantly improve the network performance. This is a contradictory problem, but the residual block effectively solves the contradiction of how to avoid the disappearance of the gradient when deepening the network.

LeakyReLU was used as the ³¹activation function. The function of LeakyReLU is very similar to that of ReLU. There is the only difference in the part where the input is less than 0. The value of the part where the input of ReLU is less than 0 is 0, while the value of the part where the input of LeakyReLU is less than 0 is negative and has a slight gradient. If ReLU is used as the activation function of the middle layer, when the gradient of the backpropagation process is 0, the corresponding weight and bias parameters cannot be updated this time. Then the neuron can no longer learn. This phenomenon is called “neuron death”. So we use LeakyReLU as the activation function of the middle layer to avoid this problem. Finally, we apply a 1×1 convolution layer and then use the “Sigmoid” activation function to output the mask. The initial learning rate of the model is set to 0.0002. Every three epochs, the model is not improved and the learning rate is automatically reduced by half.

The data enhancement tool we used is the “albumentations” (<https://github.com/albumentations-team/albumentations>). It is a fast training data enhancement library for OpenCV. It has a very simple and powerful interface that can be used for a variety of tasks (segmentation and detection). It is easy to customize and it is very convenient to add other frameworks. It can convert the data set pixel by pixel, such as blur, downsampling, Gaussian point making, Gaussian blur, dynamic blur, RGB conversion, random atomization, etc; In this work. we use random gamma, random blur, horizontal flip, normalization, and other data enhancement methods. The specific model code and data enhancement code have been open-source on GitHub. Our model is trained using the Tensorflow-2.40 platform on NVIDIA GeForce RTX 3060 GPU with Intel CPU Core i5-11600K@ 3.9GHz, 32GB RAM.

Backbone	Accuracy	Specificity	Sensitivity	Dice Coefficient	Jaccard Index
VGG16	99.0%	99.7%	96.1%	97.5%	95.1%
DenseNet121	99.4%	99.4%	99.2%	98.7%	97.5%
Inceptionv3	99.0%	99.6%	97.5%	98.3%	96.7%
ResNet101	99.2%	99.6%	98.1%	98.4%	97.0%
Efficientnet-b4	99.4%	99.5%	99.1%	98.8%	97.7%

Table3. Test accuracy of different backbones(pre-trained on ImageNet dataset).

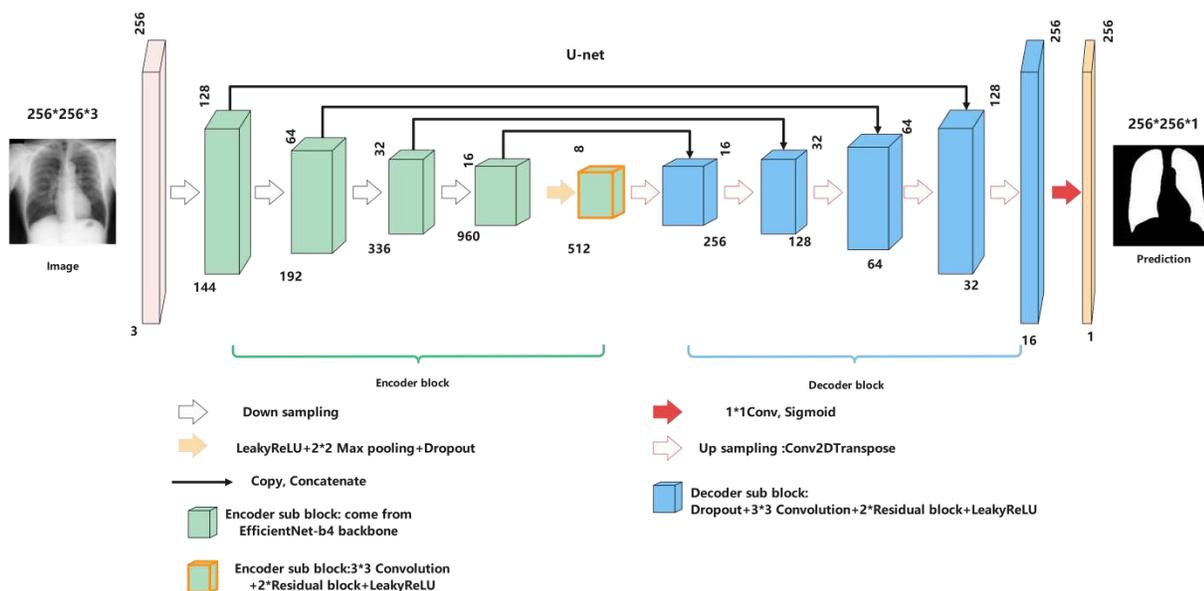


Figure 4. Architecture of U-net with EfficientNet-b4 Encoder.

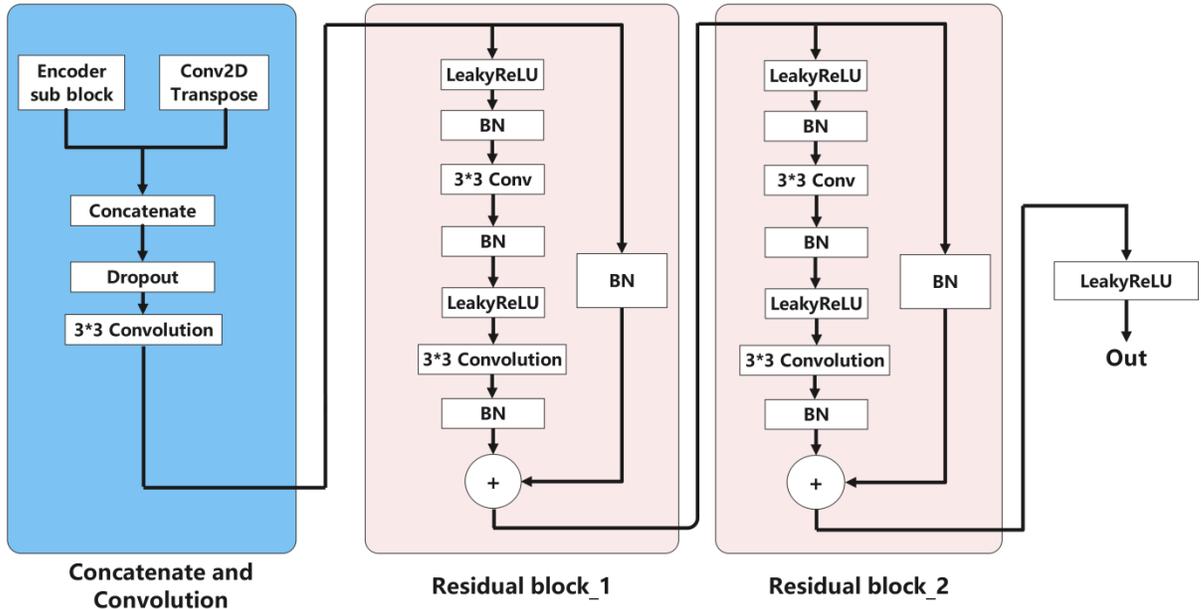


Figure 5. Encoder sub-block as shown. BN refers to Batch Normalization.

Loss function Utilizing the Dice score as the loss function results in the best overlap rates in three datasets.

$$DSC_{(A,B)} = \frac{2|A \cap B|}{|A| + |B|} = \frac{TP}{2TP + FP + FN} \quad L_{dsc} = 1 - dsc \quad (7)$$

Computer graphics morphological repair Considering that fragments (FP) and holes (FN) will appear in lung segmentation of some CXRs, we used two optimization methods to eliminate false positives and false negatives in segmentation. For fragment (FP), we use a connected domain filtering algorithm. Only the two largest connected regions in the image (corresponding to the left and right lungs of the human body) are retained, and small fragments are filtered out. For holes (FN), we use the flood filling algorithm to repair them.

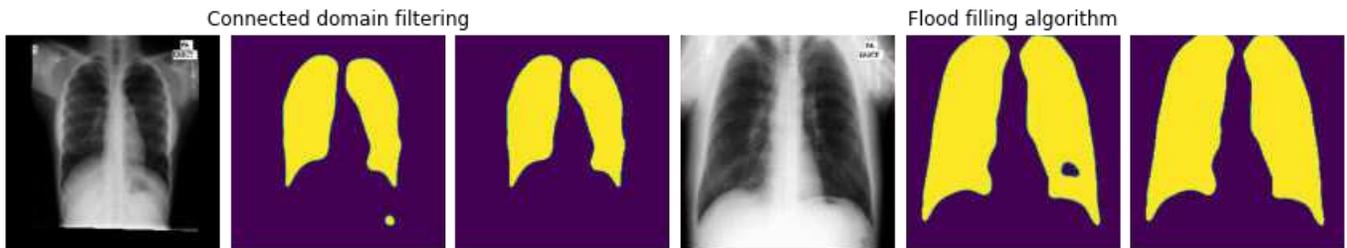


Figure 6. The use of two computer graphics morphology algorithms is shown in the figure above.

Datasets used in the experiment

The Haut dataset is a lung field mask dataset based on part of the NIH Chest X-ray dataset. NIH Chest X-ray Dataset is comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. There are 15 classes (14 diseases, and "No findings"). Images can be classified as "No findings" or one or more disease classes, showing 14 common thoracic pathologies. NIH Chest X-ray dataset itself does not contain lung field labels. We randomly selected 2785 samples and invited doctors (Wenlian Wang and Junkui Deng from Nanyang Central Hospital) to label their lung fields. The Haut dataset contains some chest radiographs that are seriously blurred, obscured, and deformed. To use Efficientnet-b4, the images were downsized to 256x256 pixels as a pre-processing step.

Datasets	JSRT	MC	Haut(private dataset)
Healthy cases	93	80	1647
Unhealthy cases	Lung nodules: 154	Tuberculosis: 58	Multiple diseases: 417 Infiltration: 193 Atelectasis: 111 Effusion: 78 Nodule: 65 Mass: 54 Pneumothorax: 43 Cardiomegaly: 37 Pleural Thickening: 37 Fibrosis: 34 Consolidation : 25 Emphysema: 21 Edema: 11 Pneumonia: 10 Hernia: 2
Total	247	138	2785

Table 4. Three Chest X-Ray datasets were used in the proposed work.

Conclusion

In this paper, an accurate and robust automatic lung segmentation method based on U-Net architecture is proposed. This method uses the pre-trained Efficientnet-b4 as the encoder and uses the residual block and LeakyRelu to optimize the decoder. Our method achieves 95.8% and 95.5% Jaccard Index on JSRT and MC datasets respectively. The accuracy is comparable to that obtained in the advanced literature in recent years. Based on the NIH Chest X-ray dataset, we randomly choose 2785 CXR images from it and invited experienced radiologists to manually mark their lung areas. These 2785 CXR images can be divided into 16 kinds of different situations. We use the above model to evaluate the segmentation performance in the Haut dataset. Achieved 97.4 % of the overall Jaccard Index. However, the lung segmentation scores of different diseases are different. We found that chest radiograph segmentation scores were higher in healthy or mild diseases. The accuracy of lung segmentation is relatively low when the lung area is blurred, blocked by medical equipment, and severely deformed due to serious diseases. We also evaluated lung segmentation for specific diseases.

References

1. Jaeger, S. *et al.* Automatic Tuberculosis Screening Using Chest Radiographs. *IEEE T Med Imaging* **33**, 233-245, doi:10.1109/TMI.2013.2284099 (2014).
2. Mittal, A., Hooda, R. & Sofat, S. J. I. P. Lung Field Segmentation in Chest Radiographs: a historical review, current status, and expectations from deep learning. *IET Image Processing* **11**, 937-952, doi:10.1049/iet-ipr.2016.0526 (2017).
3. Luo, L., Yu, L., Chen, H., Liu, Q. & Heng, P. A. J. I. T. o. M. I. Deep Mining External Imperfect Data for Chest X-ray Disease Screening. *IEEE T Med Imaging* **PP**, 1-1, doi:10.1109/TMI.2020.3000949 (2020).
4. Skoura, E., Zumla, A. & Bomanji, J. J. I. J. o. I. D. Imaging in tuberculosis. *International Journal of Infectious Diseases* **32**, 87-93, doi:10.1016/j.ijid.2014.12.007 (2015).
5. Sema, Candemir, Sameer, Radiology, A. J. I. J. o. C. A. & Surgery. A review on lung boundary detection in chest X-rays. *International Journal of Computer Assisted Radiology and Surgery*, doi:10.1007/s11548-019-01917-1 (2019).
6. Wang, X. *et al.* ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *IEEE Conference on Computer Vision and Pattern Recognition (2017)*, doi:10.1109/CVPR.2017.369 (2017).
7. Ronneberger, O., Fischer, P. & Brox, T. J. S. I. P. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Springer International Publishing*, doi:10.1007/978-3-319-24574-4_28 (2015).
8. Shiraishi, J. *et al.* Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology* **174**, 71-74, doi:10.2214/ajr.174.1.1740071 (2000).
9. Sema Candemir*, S. J., Kannappan Palaniappan, Jonathan P. Musco, Rahul K. Singh, Zhiyun Xue. Lung

- segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE T Med Imaging* **33**, 577-590, doi:10.1109/TMI.2013.2290491 (2014).
10. Ginneken, B. V., Stegmann, M. B. & Loog, M. J. M. I. A. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical Image Analysis* **10**, 19-40, doi:10.1016/j.media.2005.02.002 (2006).
 11. Qin, C., Yao, D., Shi, Y. & Song, Z. J. B. E. O. Computer-aided detection in chest radiography based on artificial intelligence: A survey. *BioMedical Engineering OnLine* **17**, 113-, doi:10.1186/s12938-018-0544-y (2018).
 12. Diniz, J. *et al.* Spinal cord detection in planning CT for radiotherapy through adaptive template matching, IMSLIC and convolutional neural networks. *Computer Methods and Programs in Biomedicine* **170**, 53-67, doi:10.1016/j.cmpb.2019.01.005 (2019).
 13. Park, B., Cho, Y., Lee, G., Sang, M. L. & Kim, N. J. S. R. A Curriculum Learning Strategy to Enhance the Accuracy of Classification of Various Lesions in Chest-PA X-ray Screening for Pulmonary Abnormalities. *Scientific Reports* **9**, doi:10.1038/s41598-019-51832-3 (2019).
 14. Greenspan, H., Ginneken, B. V. & Summers, R. M. J. I. T. o. M. I. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE T Med Imaging* **35**, 1153-1159, doi:10.1109/TMI.2016.2553401 (2016).
 15. Hwang, S. & Park, S. Accurate Lung Segmentation via Network-Wise Training of Convolutional Networks. *Springer* doi:10.1007/978-3-319-67558-9_11 (2017).
 16. Rahul *et al.* An Efficient Variant of Fully-Convolutional Network for Segmenting Lung Fields from Chest Radiographs. *Wireless personal communications*, doi:10.1007/s11277-018-5777-3 (2018).
 17. Mittal, A., Hooda, R. & Sofat, S. J. W. P. C. LF-SegNet: A Fully Convolutional Encoder-Decoder Network for Segmenting Lung Fields from Chest Radiographs. *Wireless Personal Communications*, doi:10.1007/s11277-018-5702-9 (2018).
 18. Badrinarayanan, V., Kendall, A. & Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**, 2481-2495, doi:10.1109/TPAMI.2016.2644615 (2017).
 19. Ngo, T. A. & Carneiro, G. Lung segmentation in chest radiographs using distance regularized level set and deep-structured learning and inference. *IEEE International Conference on Image Processing*, 2140-2143, doi:10.1109/ICIP.2015.7351179 (2015).
 20. Rashid, R., Akram, M. U., Hassan, T. J. I. C. I. A. & Recognition. Fully Convolutional Neural Network for Lungs Segmentation from Chest X-Rays. *International Conference Image Analysis and Recognition*, doi:10.1007/978-3-319-93000-8_9 (2018).
 21. Chang, C. S., Lin, J. F., Lee, M. C. & Palm, C. Semantic Lung Segmentation Using Convolutional Neural Networks. *Bildverarbeitung für die Medizin 2020*, doi:10.1007/978-3-658-29267-6_17 (2020).
 22. Souza, J. C. *et al.* An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Computer methods and programs in biomedicine* **177**, 285-296, doi:10.1016/j.cmpb.2019.06.005 (2019).
 23. Saidy, L. & Lee, C. C. Chest X-Ray Image Segmentation Using Encoder-Decoder Convolutional Network. *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, doi:10.1109/ICCE-China.2018.8448537 (2018).
 24. Yang, W. *et al.* Lung Field Segmentation in Chest Radiographs from Boundary Maps by a Structured Edge Detector. *IEEE Journal of Biomedical and Health Informatics*, 1-1, doi:10.1109/JBHI.2017.2687939 (2017).
 25. Yahyatabar, M., Jouvét, P. & Cheriet, F. Dense-Unet: a light model for lung fields segmentation in Chest X-Ray images. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, doi:DOI:10.1109/EMBC44109.2020.9176033 (2020).
 26. Singh, A. *et al.* Deep LF-Net: Semantic lung segmentation from Indian chest radiographs including severely unhealthy images. *Biomedical Signal Processing and Control* **68**, 102666, doi:10.1016/j.bspc.2021.102666 (2021).
 27. Long, J., Shelhamer, E., Darrell, T. J. I. T. o. P. A. & Intelligence, M. Fully Convolutional Networks for Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (2015)* **39**, 640-651, doi:10.1109/CVPR.2015.7298965 (2015).
 28. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-Decoder with Atrous Separable

- Convolution for Semantic Image Segmentation. *European Conference on Computer Vision* 833-851, doi:10.1007/978-3-030-01234-2_49 (2018).
29. Zhang, J., Xie, Y., Xia, Y. & Shen, C. J. I. T. o. M. I. Attention Residual Learning for Skin Lesion Classification. *IEEE T Med Imaging*, 1-1, doi:10.1109/TMI.2019.2893944 (2019).
 30. He, K., Zhang, X., Ren, S., Sun, J. J. I. C. o. C. V. & Recognition, P. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, doi:10.1109/CVPR.2016.90 (2016).
 31. Liu, Y., Wang, X., Wang, L., Liu, D. J. A. M. & Computation. A modified leaky ReLU scheme (MLRS) for topology optimization with multiple materials. *Applied Mathematics and Computation* **352**, 188-204, doi:10.1016/j.amc.2019.01.038 (2019).

Acknowledgements

This work was supported by the Henan province programs for science and technology development (172102210028) funded by the Department of Science and Technology of Henan Province. We thank Wenlian Wang and Junkui Deng from Nanyang central hospital for labeling CXR images and checking the lung segmentation.

Author contributions

Wufeng Liu and Jiaxin Luo wrote the main manuscript text. Wufeng Liu and Jiaxin Luo performed experiments and prepared figures. Liang Yu and Yan Yang cleansed the dataset. Jiaxin Luo and Liang Yu prepared the dataset and confirmed abnormalities. Wenlian Wang and Junkui Deng labeled CXR images and checked the lung segmentation. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.