A Novel Automatic Method Based on U-Net for Lung Fields Segmentation

Yuqin Li, *Member, IAENG*, Bing Wang, Ke Zhang*, Zhengang Jiang*, Weili Shi, Weiwu Liu

Abstract—Computer-Aided Diagnosis (CAD) has become a requisite and fundamental part in medical detection and diagnosis. As an indispensable component of CAD, the lung fields segmentation is critical for further analysis. Nowadays, it is well known that the methods of deep convolutional neural networks (DCNNs) have achieved outstanding performance in medical image segmentation. Especially the U-Net and its extensions have obtained promising accuracy on medical image segmentation. However, due to the superimposed regions in lung fields and varied shapes among different individuals, it is difficult to detect and segment the boundaries precisely. Besides, insufficient training dataset may results in poor generalization ability of the networks. To address these problems, this paper uses the standard U-Net as its backbone, and optimizes and improves it effectively. The proposed method can segment the lung fields in chest X-ray images automatically, which integrates U-Net, Bi-directional ConvLSTM (BConvLSTM), Squeeze and Excitation (SE), and fully connected CRF into one framework. In the proposed architecture, a single U-Net network is employed as the backbone. Then, the BConvLSTM is employed to concatenate the features extracted from the encoder and the corresponding decoder. In this way, the skip connection of the U-Net structure is replaced by BConvLSTM. Furthermore, the SE modules are embedded in the decoder to recalibrate the channel-wise features. Besides, the fully connected CRF is used to further refine the initial segmentation contours, which can fully consider the mutual information between pixels in the original image. Compared with diverse lung fields segmentation algorithms on JSRT and MC datasets, the superiority, effectiveness and robustness of the proposed method are verified.

Index Terms—Lung fields segmentation, SE module, U-Net, BConvLSTM, Fully connected CRF

I. INTRODUCTION

NOWADAYS, Computer-Aided Diagnosis (CAD) is an essential tool for automatic detection and diagnosis, which can offer doctors more persuasive judgments. The

Y. Li is a PhD candidate in Changchun University of Science and Technology, Changchun 130022, China. (e-mail: lyqcust@outlook.com).

B. Wang is a master student in Changchun University of Science and Technology, Changchun 130022, China. (e-mail: yxxwangb@163.com).

K. Zhang is an associate professor in Changchun University of Science and Technology, Changchun 130022, China. (Corresponding author, Tel: +86-0431-85583582, e-mail: zhangke@cust.edu.cn).

Z. Jiang is a professor in Changchun University of Science and Technology, Changchun 130022, China. (Corresponding author, Tel: +86-0431-85583582, e-mail: jiangzhengang@cust.edu.cn).

W. Shi is an associate professor in Changchun University of Science and Technology, Changchun 130022, China. (e-mail: shiweili@cust.edu.cn).

W. Liu is a doctor of the Second Hospital of Jilin University, Changchun 130022, China. (e-mail: Liuweiwu2001818@sohu.com).

radiologists use the output of CAD as second opinion guidance to make the final decisions [1]. With the development of auxiliary diagnosis technologies, the performance of CAD has been continuously improved. The technologies of medical image analysis are key components of CAD. As we all know, Chest X-Ray images are widely used for detection and diagnosis of lung diseases. However, some challenges remain in correctly and effectively interpreting lung medical images, such as the difference among radiologists in skill, concentration, and experience. Accurate lung fields segmentation in medical lung images is a critical component that facilitates subsequent analysis. However, accurate lung segmentation has become a challenge for some reasons [2]. On the one hand, there are some superimposed regions and strong edge structures, such as ribs and clavicles. These structures will confuse the objective boundaries. Furthermore, due to differences in patient age, gender, and health conditions, the shape of the anatomical structure of the lungs varies greatly. On the other hand, the related DCNNs methods have been widely adopted in image analysis, especially in the task of image segmentation [2]. However, these models need to be trained on plenty of training samples. Besides, a large number of parameters are produced during the training of the networks. These problems lead to low generalization ability and high memory usage. To mitigate these challenges, researchers at home and abroad have proposed various strategies. This paper discusses it in accordance with traditional and deep learning-based methods.

We divide the traditional segmentation methods into four categories: rule-based methods, pixel classification-based methods, registration-based methods and deformable model-based methods [3].

Rule-based methods contain a sequence of heuristic rules and steps according to the position, shape, intensity, texture, and anatomical information of lung fields [4]-[6]. However, the approximately solutions obtained by these methods are not satisfactory compared to the global optimal solution. Up to now, rule-based methods are usually employed as the initial steps [7].

Pixel classification-based methods are trained according to the appearance, size, position, and anatomical structures of the lung fields [8]-[10]. Then each pixel is assigned to either the object or background regions. Therefore, it can be regarded as a binary classifier. But the pixel classification-based methods also suffer from classification errors near the boundary.

Registration-based methods use the labeled lung dataset as anatomical atlas to match the target images. Candemir et al. [7] proposed a novel non-rigid registration-driven method to

detect the lung boundaries via an adaptive method based on image retrieval. However, these methods are inefficient and rely on the results of non-rigid registration heavily.

Deformable model-based methods have been extensively studied profit from its shape flexibility and are widely used in medical image segmentation [11], [12]. These methods take the lung boundary as a flexible curve and force it under internal and external potential energy [13]. The internal potential energy ensures that the curve remains bendable and stretchable by calculating the object shape [14]. In addition, by using appearance information, the external pushes the model as close to the boundary as possible. However, these methods tend to produce average shapes. For abnormal cases, it is ineffective due to heavy reliance on the initial model.

The above-mentioned methods need to set some hyperparameters subjectively and extract useful features manually, which relies on manual experience highly. These methods are not automatic and have low generalization. Hence, the results are not reliable enough.

Deep learning-based methods have achieved outstanding performance in many applications [15], [16], especially in medical image segmentation [2], [17], [18]. Convolutional neural networks (CNNs) demonstrated its efficiency due to its prominent achievements in abstract features expression and extraction [19], [20]. However, a plenty of parameters produced during training of the networks causes a large storage overhead. Besides, the repeated operations of convolution during model training process affect computation efficiency. For semantic segmentation tasks, Fully Convolutional Network (FCN) [21] is a basic framework in deep learning, which is trained in an end-to-end manner. In FCN, all fully connected layers in CNN are replaced with convolution operations to maintain spatial resolutions. However, since the deep feature mappings are included in the final results of FCN model, the spatial domain information extracted from the shallow network is ignored. Besides, the FCN model is not sensitive to image details. Drozdzal et al. [22] employed both long and short connections in deep FCN for segmentation, and obtained better results than the original FCN. Ronneberger et al. [23] proposed U-Net network in 2015 with skip architecture. Today, the U-Net network has been regarded as a baseline framework for medical image segmentation. It concatenates the features from shallow and deep layers by skip connections, and can make full use of global position and context information simultaneously. The U-Net network contains two sub-structures: encoding and decoding structures. However, due to the large differences in shape and size of the target tissue or organ, the segmentation performance of a single U-Net model is inefficient.

In this paper, a novel and end-to-end framework is proposed for lung fields segmentation. This paper has the following contributions: Firstly, the BConvLSTM is employed to concatenate the features extracted from the encoder and the corresponding decoder, which replaces the skip connection of U-Net. Secondly, the SE modules are embedded in the decoder to recalibrate the channel-wise features. Then the U-Net with the above structures produces a probability map for each sample. Finally, a fully connected CRF is employed to refine the initial segmentation contours, and the probability map and the original image are used as its input. The rest of this paper is organized as follows: Section 2 shows the related theories and some works in this paper; Section 3 presents the details of the proposed method; Section 4 gives some comparative experiments and discussion; Section 5 is our conclusion.

II. RELATED THEORIES AND WORKS

Today, a lot of DCNNs methods have been widely applied in the task of medical image analysis, especially in organ and tissue segmentation. In most cases, due to the varied shapes and sizes of organs among different individuals, a single U-Net can hardly obtain accurate results. In this research, inspired by the strengths of U-Net, BconvLSTM, SENet, and fully connected CRF structures, a novel and automatic framework is constructed for lung fields segmentation. This section introduces the proposed overall architecture and some related works.

A. U-Net network

Over the past few years, CNN has been introduced to tackle different image processing tasks effectively with its remarkable achievement for feature extraction and expression. However, it causes large storage overhead and lower computation efficiency due to its repeated convolutional operations during training.

The U-Net network is an extension of FCN, which is more compatible with the task of image segmentation. Besides, the skip connection alleviates the loss of feature information and further improves the segmentation performance. It consists of down-sampling and up-sampling structures, and can also be called encoder and decoder, respectively. The down-sampling structure extracts the features information from the input samples, and transmits it through continuous operations to the next layer. The up-sampling structure expands the resolution of the feature maps through deconvolutional operations until the resolution of the images is fully restored.

B. BconvLSTM

In this paper, we use BConvLSTM [24] in our network. The structure of BConvLSTM employs two ConvLSTMs [25], which can transmit input data to forward and backward paths.

As we all know, the architecture of U-Net adopts skip connections to combine the features extracted from the encoder and the corresponding decoder layers. In fact, it is generally considered that the features obtained from the decoder structure involve more semantic information, while the higher resolution exists in the corresponding encoder structure. In this work, we combine these different feature maps with nonlinear functions instead of a simple concatenation.

Due to the difference in feature information between the encoder and the decoder, this paper adds the BConvLSTM module in the skip connection to reduce the semantic gap. BConvLSTM contains two ConvLSTM blocks, the following is a brief description of ConvLSTM.

Standard LSTM fail to consider spatial correction information. To mitigate this problem, ConvLSTM [25] introduces convolution operation into input-to-state and

state-to-state transitions. It models temporal dependence while preserving spatial information. ConvLSTM includes four elements: input gate i_t, output gate o_t, memory cell c_t, and forget gate f_t. During the training process, the memory cell is used to memorize or forget information in the sequence, and accordingly update the information stored in it in time. The input gate and output gate control the input and output messages in the memory cell respectively. Besides, the forget gate is used to discard some irrelevant information in the memory cell. The process of ConvLSTM at time t can be formulated as (1) [25]:

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \\
h_t &= o_t \circ \tanh(c_t).
\end{aligned} \tag{1}
$$

where $*$ denotes the operator of the convolution and $\circ$ is Hadamard product.

Obviously, the ConvLSTM mentioned above accumulates past information in the memory cells, which can be seen as simply "remembering" the past sequences. In this paper, we employed BConvLSTM in the skip connection. The BConvLSTM network is a variant of the ConvLSTM network, which takes into account the information in both the forward path and the backward path. In this way, it can enhance the predict results by considering different information. The process of BConvLSTM can be formulated as (2) [24].

$$
Y_t = \tanh(W_y^{H^f} * H_t^f + W_y^{H^b} * H_{t-1}^b) \tag{2}
$$

where H^f and H^b are the hidden state tensors.

C. SENet

Recently, many mechanisms have been proposed to improve the focusing ability of the network in acquiring spatial information, such as attention and perception. The architecture of Squeeze-and-Excitation networks (SENet) [26] is a kind of attention mechanisms, which can highlight important features and suppresses nonsignificant features. In this paper, SE modules are embedded in the decoder to recalibrate the channel-wise features. Next, we give a simple introduction to the SE modules.

The SE modules are used to capture the relationships between feature channels, and the interdependence relationship can be modeled explicitly. Assigning weights to each channel in the feature maps can enhance useful features and suppress others. Therefore, the adaptive calibration of the feature channel can be realized. As shown in Fig. 1, SE module includes two structures: squeeze and excitation. Squeeze operation uses the global average pooling (GAP) algorithm to aggregate the global spatial information into a channel descriptor, which can be formulated as (3). Excitation operation obtains the interdependence between channels through the fully connected layer, and assigns a corresponding proportion of weights to each channel. This process can be formulated as (4), it can capture channel-wise dependencies. In this way, recalibration of the original features is realized in the channel dimension.

$$
z_c = \mathbf{F}_s(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \tag{3}
$$

$$
s = \mathbf{F}_e(z, \mathbf{W}) = \sigma(g(z, \mathbf{W})) = \sigma(\mathbf{W}_2 ReLU(\mathbf{W}_1 z)) \tag{4}
$$

where $\mathbf{F}_s$, $\mathbf{F}_e$ represent the squeeze and excitation operations, respectively.
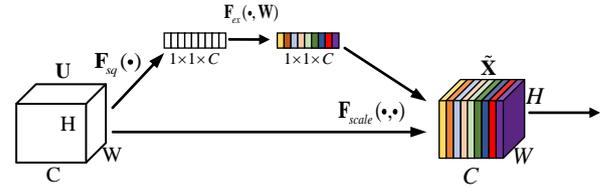

Fig. 1. The Squeeze-and-Excitation module.

D. Fully connected CRF

In 2011, Krähenbühl et al. [27] proposed the Fully connected conditional random field (CRF) algorithm, which can be used as a post-processing step for segmentation method. The CRF method originated from the Markov Random Field (MRF), which consists of two parts: unary and pairwise potentials. Unary potentials consist of a single pixel or image patch, and pairwise potentials include adjacent pixels or image patches. Different from the basic CRF, the fully connected CRF calculates pairwise potentials on all pairs of pixels in an image. Equation (5) is the energy function of fully connected CRF.

$$
E(x) = \sum_i \theta_i(x_i) + \sum_{i<j} \beta_{ij}(x_i, x_j), \quad i, j \in \{1, 2, \ldots, N\} \tag{5}
$$

where N is the number of pixels, x_i denotes the label that assigned to pixel i. The operator θ, β represent unary term and pairwise term respectively. By minimizing (5), a sequence of x_i can be obtained, which can be regarded as the solution of the function. The unary potential is formulated as (6), which is the negative log-likelihood.

$$
\theta_i(x_i) = -\log P(x_i) \tag{6}
$$

where $P(x_i)$ represents the probability of assigning pixel i to the target. Generally speaking, the calculation of $P(x_i)$ can correspond to the output of the network. Pairwise potentials allow efficient inference by using a fully connected graph. The formula is as follows:

$$
\beta_{ij}(x_i, x_j) = \mu(x_i, x_j) \left[w^{(1)} \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\tau^2} \right) + w^{(2)} \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\zeta^2} \right) \right] \tag{7}
$$

where I represents an image, p_i and p_j represent the positions of pixels I_i, I_j respectively. Equation (7) consists of two Gaussian kernels, and the parameters $w^{(1)}$ and $w^{(2)}$ denote the weights of different Gaussian kernels. Besides, $\sigma_\alpha, \sigma_\tau, \sigma_\xi$ control the "scale" of Gaussian kernels.
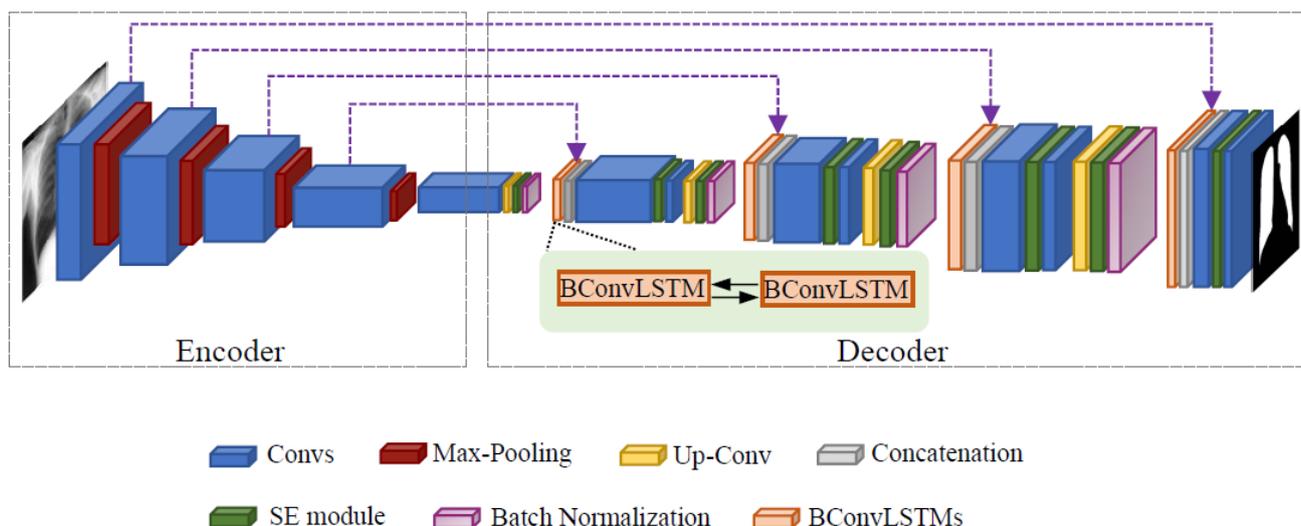
Fig. 2. The network architecture of the proposed method in this paper

III. PROPOSED METHOD

A. Network architecture

The proposed method of lung fields segmentation in this paper is depicted in Fig. 2. In our research, we adopt the U-Net as the backbone. Inspired by the theories of BconvLSTM and SENet, this paper puts forward a new model to obtain a coarse lung fields segmentation result, and then uses fully connected CRF to refine the initial contours. The proposed network includes encoder and decoder as in Fig. 2. We will introduce the encoder and decoder parts in detail as follows.

B. Encoder

The encoder consists of five layers, each of which includes two convolution operations followed by a ReLU activate function. In addition, except for the last layer in the encoder, each layer contains a max-pooling operation of size 2*2. While each layer in the encoder progressively extracts image representations, the dimensions of these representations increase layer by layer. Through a sequence of convolution operations, various features are yielded of the network. In the end, the last layer in the encoder generates images with high semantic information.

C. Decoder

The decoder includes some convolution, up-convolution, and batch normalization operations. The SE and BConvLSTM modules are also embedded in the decoder to enhance the representation of the network. More definitions and information about these modules are described in section II. In a standard U-Net, it uses skip connections to concatenate features extracted from encoder to decoder with the same resolutions, so as to transmit high-resolution information throughout the network. In essence, the feature maps of the encoder include a large amount of local information, and the feature maps of decoder contain more semantic information.

Concatenating these feature maps simply may result in feature redundancy. To solve this problem, instead of the simple skip connection between the encoder and decoder, this research uses BConvLSTM to concatenate the features extracted from the encoder and the decoder. Moreover, SE modules are employed in the decoder to highlight important features and suppress others. In this way, adaptive calibration of the feature channels can be realized. Furthermore, the Batch Normalization (BN) operation is applied to speed up the process of training, which can increase the network stability.

Besides, a ReLU operation is employed as the activate function of the intermediate layers, and sigmoid is applied to the output layers to assign labels. Finally, the Adam optimizer is applied to optimize the loss function.
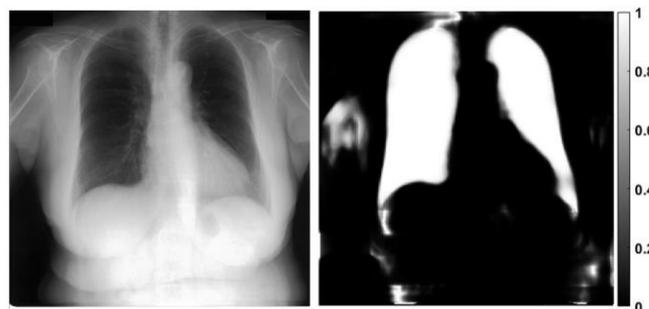

Fig. 3. Left: A sample randomly selected from the JSRT dataset. Right: Probability map of the corresponding segmentation output by the proposed network.

D. Post-processing

In this work, U-Net with BConvLSTM and SE modules produces a probability map for each sample. Specifically, the proposed network assigns a probability to each pixel in the image. Fig. 3 gives a group of randomly selected image from the JSRT dataset and its segmentation probability map. In the probability map, the intensity of each pixel represents the likelihood that the pixel belongs to a part of the lung region. In our study, the algorithm of fully connected CRF is employed as the post-processing method, and the probability map and the original image are taken as its input. This algorithm is

considered to be a highly efficient inference method. The potentials of the pairwise edge is defined by a linear combination of Gaussian kernels in the feature space. In this way, the surrounding pixels are considered in the original images.

IV. EXPERIMENT

A. Dataset information

1) JSRT dataset

In this paper, we use a publicly available dataset of the Japanese Society of Radiological Technology (JSRT) [28] to evaluate the performance of the segmentation models. The JSRT dataset is collected from one American institution and thirteen Japan institutions. The image grayscale in JSRT is 12-bit, and the spatial resolution is 0.175 mm. The lung segmentation labels for the JSRT dataset are manually annotated by radiologists and can be accessed at: http://www.isi.uu.nl/Research/Databases/SCR/. Fig. 4 shows gold standard lung segmentation masks and its original images. The JSRT dataset consists of 247 PA chest radiographs with 2048*2048 image resolution, the dataset was divided to two folds: fold 1 includes 124 images and fold 2 has 123 images.
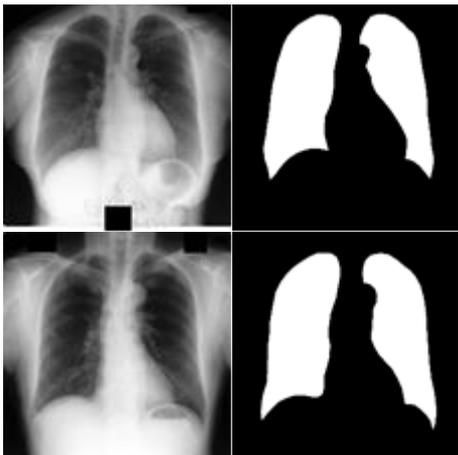

Fig. 4. Chest X-Ray images and corresponding ground truth in JSRT dataset

2) MC dataset

In this research, the Montgomery County Chest X-Ray database (MC) is also employed to evaluate the model performance. The MC dataset is a publicly available dataset created by the Department of Health and Human Services, Montgomery County, Maryland, USA. This dataset contains Chest X-Ray images collected under Montgomery County's Tuberculosis screening program. The MC dataset includes 138 Chest X-Ray images, of which 80 are from normal cases and 58 are abnormal cases with manifestations of tuberculosis. The images of the MC dataset are 12-bit gray-scale and their resolutions are either 4020×4892 or 4892×4020. The pixel spacing in vertical and horizontal directions is 0.0875mm. In addition, the MC database has gold standard segmentations under the supervision of a radiologist [7], [29]. Fig. 5 shows gold standard lung segmentation masks and its original images.
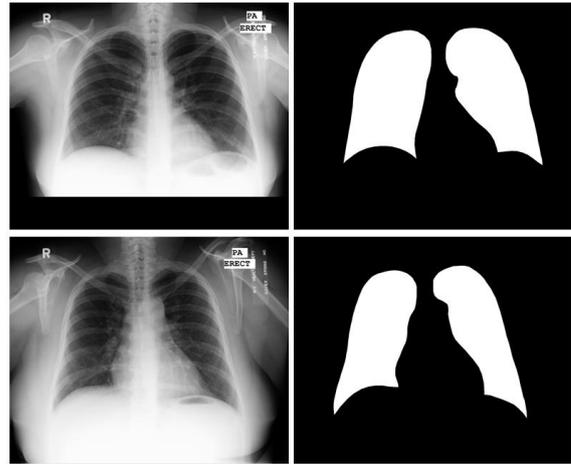

Fig. 5. Chest X-Ray images and corresponding segmentation labeled in the MC dataset.

B. Evaluation metrics

The definitions of TP, FP, TN, and FN are included in Table I. These definitions can be used in the computation of segmentation performance.

TABLE I
THE CONCEPT OF CONFUSION MATRIX

Predict	Ground-Truth	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

1) *Jaccard index (Ω):* is the agreement between the lung area in the real segmentation (GT) and the estimated results (S). It can be formulated as (8):

$$\Omega = \frac{|S \cap GT|}{|S \cup GT|} = \frac{|TP|}{|FP| + |TP| + |FN|} \tag{8}$$

2) *Dice similarity coefficient (DSC):* represents the overlap between the real segmentation and the estimated results. It can be defined as follows:

$$DSC = \frac{|S \cap GT|}{|S| + |GT|} = \frac{2|TP|}{2|TP| + |FN| + |FP|} \tag{9}$$

3) *Sensitive:* denotes the ability that the ground truth can be predicted as the correct objective region. It can be computed as (10):

$$Sensitivity = \frac{T_P}{T_P + F_N} \tag{10}$$

4) *Accuracy:* is the proportion of pixels that are correctly assigned to the objective region or background in the whole image. Its formula as shown in (11):

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \tag{11}$$

C. Data augmentation

Labeling large-scale data by radiologists is both time-consuming and costly. Therefore, it is difficult to obtain a complete medical image dataset. As we all know, deep learning methods, especially convolutional neural networks, require abundant samples for training. Based on this, a potential solution to the problem of insufficient data is augmentation technologies during training. In this paper, we

employ augmentation methods including random rotation, center cropping, scaling, and offset changes to generate new images.

D. Training

In this work, we used two-fold cross-validation method to evaluate the performance of the proposed network. The JSRT and MC datasets were split into 3 subsets: training, validation, and testing. During the training stage, the validation subset is employed to tune the trained model until the validation subset reaches prominent results, at which point the trained model is considered to have the best parameters. Based on this, 20% of the training samples were reserved as validation subset. Besides, to accelerate the network learning process, each up-convolutional layer in the decoder was followed by a BN operation. The epoch is set to 100.

V. RESULTS AND DISCUSSION

In this paper, our experiments were implemented in the Python programming language with the Keras framework. The size of the images was adjusted to 256*256 without any compression.

The proposed network was an extension of U-Net. As we all know, U-Net has been considered as the baseline framework for the task of image segmentation. Hence, this paper compared the proposed network with a single U-Net with the same data augmentation techniques to prove its effectiveness on JSRT and MC Datasets.

To validate the contribution of BConvLSTM to lung fields segmentation, we conducted an ablation experiment without BConvLSTM. The features extracted from the encoder and decoder is concatenated by the BConvLSTM structure in our work, rather than the skip connection of U-Net. In the network, the features in the encoder included a lot of local information, while decoder contained more semantic information. Combing these two features may result in feature redundancy in both local information and semantic information. This paper used BConvLSTM to replace skip connection to mitigate this problem. For a clear comparison, we set the epoch to 50, and used the model trained in step 50 for verification. The segmentation results were shown in Fig. 6.

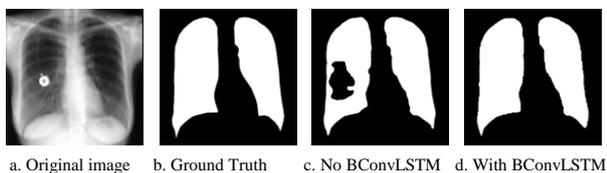a. Original image   b. Ground Truth   c. No BConvLSTM   d. With BConvLSTM

Fig. 6. Segmentation results with or without BConvLSTM

In this work, the SE modules were embedded in the decoder. To prove its effectiveness, an ablation experiment without the SE modules on JSRT was implemented. We set the epoch to 50, and used the model trained in step 50 for verification. From Table II, it is clear that the network with the SE modules produced more accurate results. The reason was that the adaptive feature recalibration of the SE modules can boost the representational ability by enhancing useful features and suppressing weak features.

TABLE II
QUANTITATIVE COMPARISON WITH SE MODULE IN THE
PROPOSED NETWORK.

Method	Ω(%)	DSC(%)
Without SE module	93.9	95.4
With SE module	95.0	96.6

TABLE III
QUANTITATIVE COMPARISON OF THE DIFFERENT LUNG FIELDS
SEGMENTATION METHODS

Method	Ω (%)	DSC(%)	Sensitivity(%)	Accuracy(%)
SEDUCM [30] (JSRT)	95.2	97.5	-----	-----
Inverted with ELU [17] (JSRT)	95.0	97.4	-----	-----
Rigid [7] (JSRT)	95.4	96.7	-----	-----
Rigid [7] (MC)	94.1	96	-----	-----
Single U-Net (MC)	92.5	94.0	95.6	96.1
Single U-Net (JSRT)	93.8	96.2	95.9	96.8
The proposed method (JSRT)	**95.48**	**97.5**	**96.3**	**97.7**
The proposed method (MC)	**95.1**	**97.3**	**95.9**	**97.0**

To evaluate the performance of the model, this work used four different indicators: Jaccard Index, DSC, Sensitivity, and Accuracy. In this paper, the probability images from the output of the proposed method were converted into binary images by setting the threshold to 0.5. Afterwards, the segmentation performance of the binary image was calculated. Table III presented some quantitative comparison results of different models. From Table III, the SEDUCM achieved 97.5% of DSC on the JSRT dataset, but it needed to predefine the number of features corresponding to various segmentation tasks. Furthermore, the parameters in the method of SEDUCM are adjusted manually, and it took more time to train the model. Therefore, its generalization and automation capabilities were poor. The performance of the Inverted network with ELU [17] achieved a Jaccard index of 95.0% and a DSC of 97.4%. It was obvious that our method achieved higher performance than the recently popular InvertedNet network. The method of [7] was widely applied in medical image segmentation. It was included in the registration-based methods that used a labeled lung database as an anatomical atlas to match the target images. Nevertheless, this method relied on the results of non-rigid registration heavily. Finally, in conclusion, the experiments and discussion demonstrated the effectiveness and reliability of the proposed method.

VI. CONCLUSION

As we all know, CAD can be used for early diagnosis and accurate treatment of lung diseases. Furthermore, CAD is a requisite tool for automatic diagnosis, which can provide radiologists with more persuasive and reliable judgments. A critical component in CAD of lung disease is automatic and accurate lung fields region detection. In this paper, a novel framework for lung fields segmentation is proposed. By using

the U-Net as the backbone, and adding BconvLSTM and SE modules to the network. Then the fully connected CRF is employed to refine the coarse segmentation results. The proposed framework is capable of learning more discriminative features, thereby producing more effective segmentation results. Finally, to prove the effectiveness of the proposed method, four indicators are adopted to evaluate the model performance. Compared with some algorithms, the proposed method shows higher performance on the JSRT and MC datasets.

REFERENCES

[1] N. Saad Mohd, S. A. R. Abu-Bakar, S. Muda, M. Mokji, and A. R. Abdullah. "Fully automated region growing segmentation of brain lesion in diffusion-weighted MRI," *IAENG International Journal of Computer Science*, vol. 39, no. 2, pp155–164, 2012.

[2] M. R. Arbabshirani, A. H. Dallal, C. Agarwal, A. Patel and G. Moore, "Accurate Segmentation of Lung Fields on Chest Radiographs using Deep Convolutional Networks," *In Medical Imaging 2017: Image Processing*, Florida, United States, vol. 10133, pp1013305, 2017.

[3] J. C. Souza, J. O. B. Diniz, J. L. Ferreira, G. L. F. da Silva, A. C. Silva, and A. C. de Paiva, "An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks," *Computer Methods and Programs in Biomedicine*, vol. 177, pp285–296, 2019.

[4] J. Duryea, and J. M. Boone, "A fully automated algorithm for the segmentation of lung fields on digital chest radiographic images," *Medical Physics*, vol. 22, no. 2, pp183–191, 1995.

[5] B. V. Ginneken, and B. M. Romeny, "Automatic segmentation of lung fields in chest radiographs," *Medical physics*, vol. 27, no. 10, pp2445–2455, 2000.

[6] Q. Gao, S. Wang, D. Zhao, and J. Liu, "Accurate lung segmentation for X-ray CT images," *Third International Conference on Natural Computation (ICNC 2007)*, Haikou, China, pp275–279, 2007.

[7] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp577–590, 2013.

[8] B. V. Ginneken, M. B. Stegmann and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp19–40, 2006.

[9] M. F. McNitt-Gray, H. K. Huang and J. W. Sayre, "Feature selection in the pattern classification problem of digital chest radiograph segmentation," *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp537–547, 1995.

[10] Z. Shi, P. Zhou, L. He, T. Nakamura, Q. Z. Yao and H. Itoh, "Lung segmentation in chest radiographs by means of gaussian kernel-based FCM with spatial constraints," *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, Tianjin, China, pp428–432, 2009.

[11] Y. Shi, F. Qi, Z. Xue, L. Chen, K. Ito, H. Matsuo and D. Shen, "Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics," *IEEE Transactions on Medical Imaging*. vol. 27, no. 4, pp481–494, 2008.

[12] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. M. Romeny and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp924–933, 2002.

[13] T. Mclnerney, and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. 1, no. 2, pp91–108, 1996.

[14] A. Mittal, R. Hooda, and S. Sofat, "LF-SegNet: A Fully Convolutional Encoder–Decoder Network for Segmenting Lung Fields from Chest Radiographs," *Wireless Personal Communications*, vol. 101, pp511–529, 2018.

[15] W. Sun, J. Wang, B. Zheng, and Z. Li. "A Novel Convolutional Neural Network Voiceprint Recognition Method Based on Improved Pooling Method and Dropout Idea." *IAENG International Journal of Computer Science*, vol. 48, no. 1, pp202–212, 2021.

[16] Q. Zheng, M. Yang, X. Tian, X. Wang, and D. Wang. "Rethinking the Role of Activation Functions in Deep Convolutional Neural Networks for Image Classification," *Engineering Letters*, vol. 28, no. 1, pp80–92, 2020.

[17] A. A. Novikov, D. lenis, D. Major, J. Hladůvka, M. Wimmer, and K. Bühler, "Fully convolutional architectures for multi-class segmentation in chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp1865–1876, 2018.

[18] A. Kalinovsky and V. Kovalev, "Lung image Segmentation using deep learning methods and convolutional neural networks," *2016 Pattern Recognition and Information Processing*, Minsk, Belarus, pp21–24, 2016.

[19] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. M. Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp924–933, 2002.

[20] D. Li, D. Yang, J. Zhang, and X. Zhang. "Ar-ann: Incorporating association rule mining in artificial neural network for thyroid disease knowledge discovery and diagnosis," *IAENG International Journal of Computer Science*, vol. 47, no. 1, pp25–36, 2020.

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Boston, MA, USA, pp3431–3440, 2015.

[22] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," *In Deep Learning and Data Labeling for Medical Applications*, Athens, Greece, pp179–187, 2016.

[23] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, Munich, Germany, pp234–241, 2015.

[24] H. Song, W. Wang, S. Zhao, J. Shen, and K. M. Lam, "Pyramid dilated deeper ConvLSTM for video salient object detection," *In Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, pp715–731, 2018.

[25] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *29th Annual Conference on Neural Information Processing Systems*, Montreal, Canada, pp802–810, 2015.

[26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake, Utah, pp7132–7141, 2018.

[27] P. Krähenbühl, and V. Koltun, "Efficient inference in fully connected CRFs with gaussian edge potentials," *Advances in Neural Information Processing Systems (NIPS)*, Granada, Spain, pp109–117, 2011.

[28] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "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*, vol. 174, no. 1, pp71–74, 2000.

[29] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y. Wang, P. Lu, and C. J. McDonald, "Automatic tuberculosis screening using chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp233–245, 2013.

[30] W. Yang, Y. Liu, L. Lin, Z. Yun, Z. Lu, Q. Feng and W. Chen, "Lung Field Segmentation in Chest Radiographs from Boundary Maps by a Structured Edge Detector," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp842–851, 2018.

Yuqin Li is currently pursuing her Ph.D. in the School of Computer Science and Technology at Changchun University of Science and Technology, Changchun, China. Her research interests include medical image process and computer-aided diagnosis.

Bing Wang is currently a master student in Computer Science and Technology at Changchun University of Science and Technology, Changchun, China.. His major research is medical image segmentation.

Ke Zhang is currently an Associate Professor of the School of Computer Science and Technology, Changchun University of Science and Technology. Her research interests include feature learning, saliency detection, image retrieval, and action recognition.

Zhengang Jiang is currently a Professor of the School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, China. His major research areas include virtual reality, augmented reality, image processing, and computer-aided diagnosis and surger**ies.**

Weili Shi is an Associate Professor of the School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, China. His major research includes augmented reality, computer graphics, and image processing.

Weiwu Liu is currently a doctor of the Second Hospital of Jilin University. He received his Ph.D. degrees in 2013. His major research is medical imaging technology.