Chest X-Ray Image Segmentation Using Encoder-Decoder Convolutional Network

Lamin Saidy and Chien-Cheng Lee, *Member, IEEE*
Dept. of Communication Engineering, Yuan Ze University, Taiwan

***Abstract*--This paper presents a deep learning method of segmenting lungs in chest X-Ray image using *Encoder-Decoder Convolutional Network* on the JSRT (Japanese Society of Radiological Technology) lung nodule dataset. The result of the segmentation has proven efficient enough to be applicable in real world medical environments to bring ease in determining the area occupied by the lungs and some other medical diagnosis.**

I. INTRODUCTION

Segmenting medical image is one of the most important task in medical image analysis. Often, it is the first and the most crucial step in clinical applications [1]. Before the intervention of deep learning in computer vision, popular approaches in semantic image segmentation were the likes of TextonForest and Random Forest based classifier [2, 3]. These segmentation approaches were succeeded by deep learning methods, which are more efficient and higher in accuracy [4, 5].

The deep learning methodology used on this paper is an encoder-decoder convolutional neural network called SegNet [6]. It is designed to map low-resolution features to input resolution for pixel-wise classification in order to produce features that are useful for accurate boundary localization. SegNet was built from VGG16 [7], its fully connected layers were removed and replaced with decoder network. These as a result makes SegNet smaller and easier to train. Each layer in the decoder network has a corresponding encoder layer with pixel-wise classifier at the end of decoder network.

II. METHODOLOGY

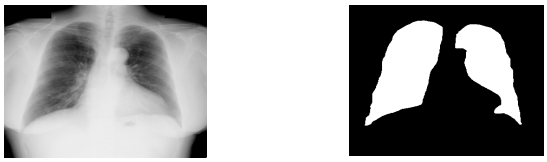
A. Preprocessing

The Japanese JSRT lung nodule dataset used in this experiment consists of 154 images, 119 of which are used for training the model and 35 used for testing. The original images are in 16-bit raw format, which we have mapped to 8-bits PNG file formats. We convert each pixel to three channels and create manual reference segmentations drawn by experts for each mapped image as shown in Fig. 1. Since our goal is to segment the lung regions, we normalize the segmentation mask to 0 and 1 which corresponds to the background and lung regions, respectively.

B. Model Training

This work was supported by the Ministry of Science and Technology, Taiwan. (Grant number: MOST 105-2622-E-155-015-CC3)

Our architecture consists of five encoder layers with their corresponding decoder layers. Each convolutional layer is followed by the batch normalization and ReLU nonlinearity, which in turn is followed by max polling with the stride length of two in the encoder while each convolutional layer in the decoder is preceded by up sampling with the mapped pooling indices from the encoder as shown in Fig. 2.



(a) (b)

Fig. 1. (a) is the original image and (b) is the manual reference segmentation

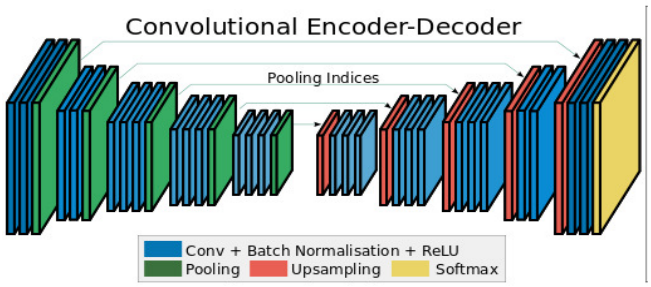


Fig. 2. Architecture of Encoder-Decoder Convolutional Network [6].

The training hyperparameters of the model are as follows:

- Batch size is 4
- Maximum iteration is 40000

The network is trained on GeForce GTX TITAN X equipped with 3072 cuda cores and 12GB memory. Training performance is illustrated on Fig. 3.

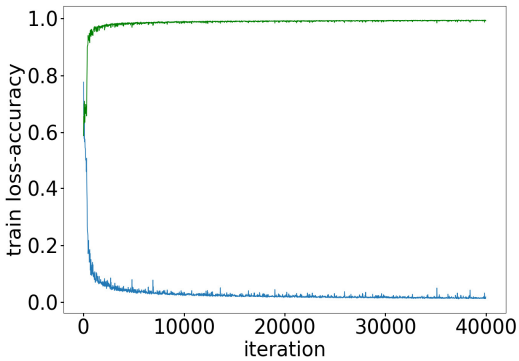
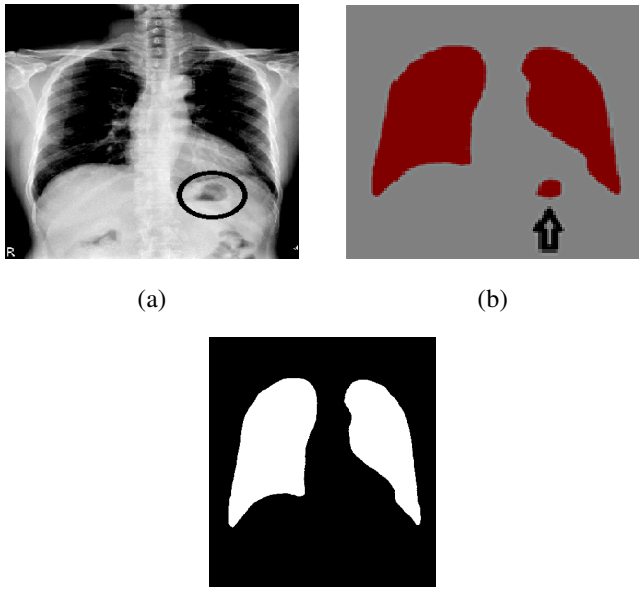


Fig. 3. Training lost and global accuracy per iteration.

III. TESTING

We tested the network on 35 unseen images. There is a softmax classifier at the end of the decoder that does pixel-wise classification and in our case 0 for the background and 1 for the foreground. Since medical images are not as clear as real world images, some of the dark spots on the image different from the lungs whose pixels are similar to the lungs are classified as lungs. Therefore, we applied a contour operation on the output to maintain only two big contours. These corresponds to the left and the right lungs as portrayed in Fig. 4.



(a) (b)

(c)

Fig. 4. (a) is the original image, (b) is the model output and (c) is the post-processed result

We observe the performance of the test result by drawing the contour of the model output and the manual reference segmentation on their corresponding original image as depicted in Fig. 5.

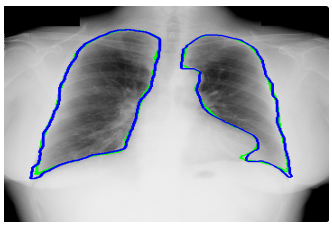


Fig. 5. Green is the model output and blue is the ground Truth

IV. RESULTS

We measure our model accuracy using four measurement criteria [8], dice similarly coefficient, sensitivity, specificity and Hausdorff distance. Each lung is measured separately and the results of first five test images is illustrated in Table I. The average measurements for the overall test images of both left and right lungs is outlined in Table II.

TABLE I
RESULT FOR THE FIRST FIVE TEST IMAGES

File	DSC	Sensitivity	Specificity	H.Distance
JPCLN117.png	L:0. 958	L: 0.960	L: 0.994	L: 1.00
	R: 0.974	R: 0.974	R: 0.991	R: 2.828
JPCLN118.png	L:0. 958	L: 0.954	L: 0.991	L:1.000
	R:0. 975	R:0.980	R: 0.990	R: 2.000
JPCLN119.png	L: 0.963	L: 0.974	L:0.993	L: 1. 000
	R:0.961	R: 0.961	R: 0.993	R: 2.000
JPCLN120.png	L: 0.921	L: 0.870	L: 0.998	L: 5.000
	R: 0.956	R: 0.948	R: 0.995	R:2.000
JPCLN121.png	L: 0.969	L: 0.936	L: 0.997	L: 2.000
	R: 0.970	R: 0.966	R: 0.989	R: 4.000

L is the measurement value for left lung, R is the measurement value for right lung and H is Hausdorff.

TABLE II
AVERAGE MEASUREMENTS FOR THE TEST IMAGES

Measurement	Left lungs	Right lungs
DSC	**0.957**	**0.962**
Sensitivity	**0.952**	**0.960**
Specificity	**0.993**	**0.992**
H.Distance	**2.011**	**3.889**

H means Hausdorff.

V. CONCLUSION

The deep learning segmentation method proposed in this paper includes sequence of operations. These include manual reference segmentation, training and testing the model. Contour operation is applied on each test result to maintain only two big contours then compute the measurements. The results illustrated in Table I and II have proven the model is applicable in real world medical environments and is susceptible to improvement.

REFERENCES

[1] I. Despotovic, B. Goossens, and W. Philips, "MRI segmentation of the human brain: challenges, methods, and applications," *Comput Math Methods Med,* vol. 2015, p. 450341, 2015.

[2] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.

[3] J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," *Commun. ACM,* vol. 56, no. 1, pp. 116-124, 2013.

[4] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843-2851.

[5] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 4, pp. 640-651, 2017.

[6] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 12, pp. 2481-2495, 2017.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[8] C. K. Chama, S. Mukhopadhyay, P. K. Biswas, A. K. Dhara, M. K. Madaiah, and N. Khandelwal, "Automated lung field segmentation in CT images using mean shift clustering and geometrical features," in *Medical Imaging 2013: Computer-Aided Diagnosis*, 2013, vol. 8670, p. 867032.