Lung Image Segmentation Using Deep Learning Methods and Convolutional Neural Networks

Alexander Kalinovsky, Vassili Kovalev
United Institute of Informatics Problems, Belarus National Academy of Sciences
Surganova St., 6, 220012 Minsk, Belarus
gakarak@gmail.com, vassili.kovalev@gmail.com, http://imlab.grid.by/

Abstract: *This paper presents results of the first, exploratory stage of research and developments on segmentation of lungs in X-Ray chest images (Chest Radiographs) using Deep Learning methods and Encoder-Decoder Convolutional Neural Networks (ED-CNN). Computational experiments were conducted using GPU Nvidia TITAN X equipped with 3072 CUDA Cores and 12Gb of GDDR5 memory. Comparison of resultant segmentation accuracy with manual segmentation using Dice's score has revealed that the average accuracy achieves 0.962 with the minimum and maximum Dice's score values of 0.926, 0.974 respectively, and standard deviation of 0.008. The study was performed in the context of large-scale screening of population for lung and heart diseases as well as development of computational services for international portal on lung tuberculosis. The results obtained with this study allow concluding that ED-CNN networks may be considered as a promising tool for automatic lung segmentation in large-scale projects.*

***Keywords*:** Image segmentation, Deep Learning, Convolutional Neural Networks, Lung.

1. INTRODUCTION

The image segmentation problem. Medical Image segmentation is known to be one of complicated problems in the image processing and image analysis field [1]. Typically, segmentation of target image objects comes before other image analysis stages and therefore any mistakes of incorrect detection of objects' borders affect all the subsequent steps severely. This paper is dealing with chest X-Ray images, which are also known as chest radiographs.

Despite the problem of segmentation of lung component in X-Ray images of chest has been addressed in several studies (see, for example, [2, 3]), the results of fully automatic extraction of lung region remains unsatisfactory in many occasions. This is especially true in case of segmentation of lungs, which are affected by various pathological processes and/or severe changes associated with age. The problem of an automatic and accurate segmentation worsened even further in the scenario of massive screening of population [4] where it moves into the Big Data domain [5].

Deep Learning and Convolutional Neural Networks. Recently, there can be observed an explosion of interest to the Deep Learning methodology. Such a methodology is commonly understood as a branch of machine learning methods, which capitalize on algorithms that attempts to model high-level abstractions in data using multiple processing layers. The corresponding computational architecture and multiple processing/abstraction layers typically represented using Convolutional Neural Networks (CNN). Such a great interest to the deep neural networks in general and CNN in particular can be partly explained by the fact that since 2009, they have won many official international pattern recognition competitions, achieving the first superhuman visual pattern recognition results in limited domains (see [6] for up-to-date review of the field).

Encoder-Decoder Convolutional Neural Networks. The first approaches employing deep learning methods for image segmentation were similar to the ones, which already examined earlier in previous image processing and pattern recognition works. They tried to directly adopt deep learning architectures for categorization small image patches or pixel neighborhoods to certain classes [7]. More recently, Vijay Badrinarayanan and colleagues from University of Cambridge have presented a novel and practical deep fully convolutional neural network architecture for semantic pixel-wise segmentation termed SegNet [8, 9]. This core trainable segmentation engine consists of an encoder network, corresponding decoder network followed by a pixel-wise classification layer. The role of the decoder network is to map the low resolution encoder feature maps to full input resolution feature maps for pixel-wise classification. The resultant Encoder-Decoder Convolutional Neural Network (ED-CNN) can be viewed as a next step of generalization of neural networks.

The purpose of this study was to examine the ability of the deep learning methods and ED-CNN neural networks to segment the lung component in chest X-Ray images. From application point of view, this study was performed in the context of large-scale screening of population for lung and heart diseases, which resulted in X-Ray image databases containing up to millions of items [10] as well as development of computational services [11] for international portal on Lung Tuberculosis hosted by Amazon [12]. Since it is not feasible and impractical to assess the efficiency of ED-CNN networks immediately on the whole image database, this study was subdivided into the following three subsequent stages:

(1) An exploratory trial based on a small set contained few hundred of manually segmented chest images, which used for both training and testing. Drawing conclusions regarding the potential utility of ED-CNN networks.

(2) Modification of ED-CNN networks and extensive testing on separate training and test sets containing thousands of cases each. Adaptation network architecture for lung segmentation in 3D Computed Tomography (CT) images and testing.

(3) Porting resultant software solutions to a powerful workstation equipped by modern GPUs and incorporation into the target environment.

Thus, this paper dedicated to the first, exploratory stage of the whole bunch of prospective research and developments on lung segmentation in radiological images using deep learning methods and recent neural network approaches.

2. MATERIALS

The image set consisted of 354 X-Ray chest images, each of which accompanied by lung masks resulted from manual segmentation. These images originated from two different sources:

- 107 images from tuberculosis portal [12] (image Source 1),
- 247 images from open Japanese JSRT Database [13] (image Source 2).

Original images from both sources and corresponding masks of the lung component are illustrated in Fig. 1.

Fig.1 – Example of original chest X-Ray images from two image sources and their lung masks obtained manually.

It was hoped that the use of inhomogeneous dataset containing chest images acquired in different countries with the help of different scanners would be helpful for obtaining more objective and conclusive testing results.

3. METHODS

The basic elements of the SegNet neural network architecture (Fig. 2) can be viewed as a stack of convolution layers (Encoder) with their corresponding de-convolution layers (Decoder). The network architecture used in this work had 4 encoding and 4 decoding layers. Every encoder layer reduces the input feature map size by factor of 2. Therefore, the combined sub-sampling rate was equal to 16. It is commonly known that large scaling factors can potentially improve desired properties of displacement, rotation and scale invariance of the convolution network being considered in the spatial domain. Also, in case of chest X-Ray image segmentation, the original input images already partly aligned due to the natural top-bottom orientation of patient's body within the scanner. Consequently, the lung area is typically located near the image center and the top part of lung situated in the upper half of the image. Thus, the relatively large value of scale factor such as 16-fold represents a good spatial tolerance for the problem in hand.

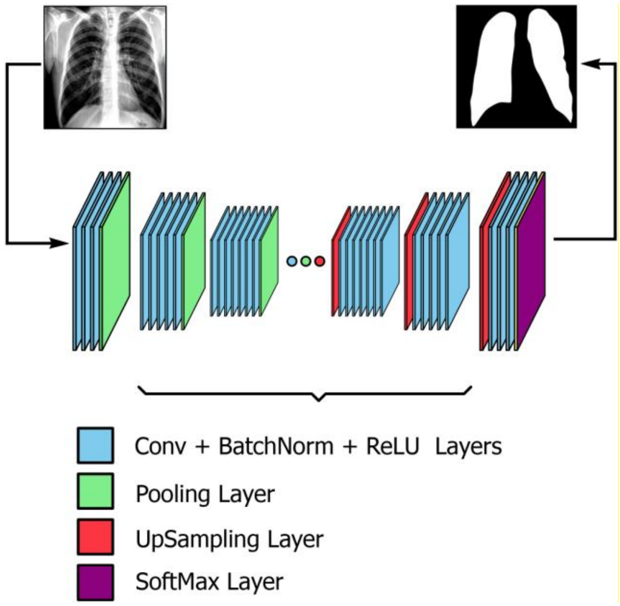

Fig.2 – Architecture of Deep Encoder-Decoder Convolutional Neural Networks

In this work we used ReLU as the nonlinear activation function [13]. The MaxPooling sub-sampling was used on the encoding stage and MaxPooling up-sampling (un-pooling) utilized for the decoding stage. At every stage, the window size was set to a small patch of 2x2 pixels in size, without overlapping.

It is known that the problem of unpooling in decoder layers is not uniquely defined. In order to solve this problem in SegNet, the upsampling of feature map in decoder layer was implemented using max-pool index from corresponding encoder layer (see Fig. 3). Every convolution and deconvolution layer maintains a fixed number of filters (Fig. 3), which was set to 64 filters.

On the final layer of ED-CNN neural network we used SoftMax function of the following type:

$$y_k = \frac{\exp(x_k)}{\sum_i \exp(x_i)}.$$

At the classification stage, the following two techniques have been used for reducing the influence of X-Ray intensity variations in the original images to the neural network being employed:

(a) At a preprocessing stage, we transform the intensity of each input image using the histogram equalization technique [14].

(b) The Local Contrast Normalization (LCN) procedure [15] was applied at the input of encoding layers.

At the experimentation step the ED-CNN neural network was trained on a graphics processor Nvidia TITAN X equipped with 3072 CUDA Cores and 12Gb of GDDR5 memory. The network training parameters were set to:

Batch size: 6 (the minimum batch size to place network into GPU memory),

Type of Solver: SGD Caffe solver,

Number of iterations: 5000,

Number of epochs: 85.

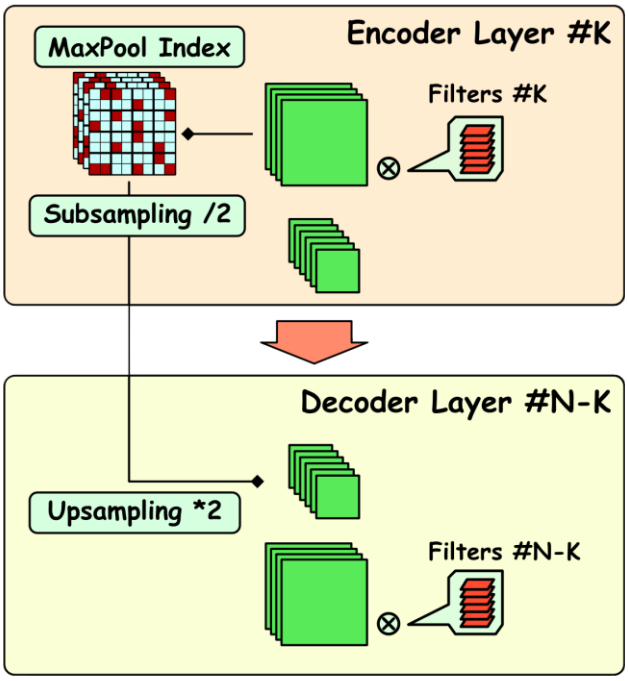

Fig.3 – Interaction scheme between the encoding and decoding neural network layers.

The network training required 11 Gigabytes of GPU memory while the full training time was approximately 3 hours. The resultant automatic segmentation accuracy score assessed by way of comparison with the results of manual segmentation using well-known Dice's score, which calculated as:

$$D_{SCORE} = \frac{S \cap T}{S \cup T},$$

where T is the "true" lung area resulted from manual segmentation, which was treated here as ground truth, and S is the lung area obtained with the automatic segmentation using ED-CNN neural network. In all the occasions, the lung area was measured as the number of image pixels constituting the lung image component.

4. RESULTS

On testing stage, the average accuracy was estimated as 0.962 with the minimum and maximum Dice's score values of 0.926 and 0.974 respectively, and standard deviation of 0.008.

Typical examples of automatic segmentation results obtained using the ED-CNN neural network with the best and worst scores are show in Fig. 4 and Fig. 5 respectively.

5. CONCLUSION

Results reported with this study allow drawing the following conclusions.

(1) The Encoder-Decoder Deep Convolutional Neural Networks may be considered as a promising tool in large-scale projects for automatic lung segmentation in chest X-Ray images. The segmentation accuracy obtained was well comparable with the accuracy provided by a specialized segmentation methods, which are based on the known "segmentation by registration" technique [4]. This technique was implemented by authors earlier and made public on a dedicated web site [11].

(2) The main advantage of the method considered in this work is the fact that the Deep Learning approach followed here is uniform enough and therefore can be applied to a wide range of different medical image segmentation tasks with minimum modifications. Furthermore, the method can be generalized for segmentation of 3D tomography images and solving other medical image analysis problems such as detection of "atypical" image regions, which often associated with lesions and other kinds of abnormalities.

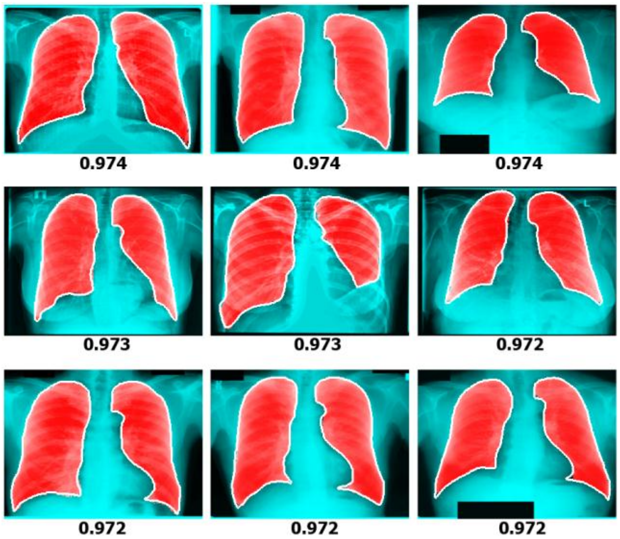

Fig.4 – Example of segmentation results with maximum that is the best Dice's score.

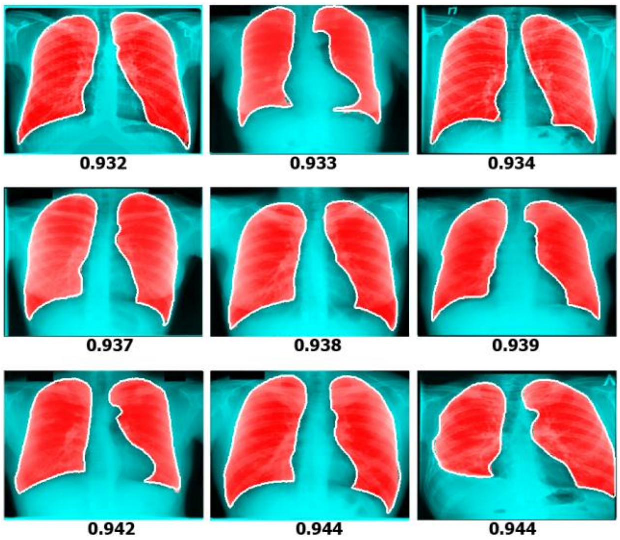

Fig.5 – Example of segmentation results with minimum that is with worst Dice's score.

Acknowledgements. This work was partly funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, U.S. Department of Health and Human Services, USA through the CRDF project OISE-15-61772-1.

6. REFERENCES

[1] Handbook of Medical Image Processing and Analysis, 2nd Edition, I.H.Bankman (Ed.), Academic Press, ISBN 978-0-12-373904-9, San Diego, USA, 2009, 985 p.

[2] S. Candemir et al. Lung Segmentation in Chest Radiographs Using Anatomical Atlases With Nonrigid Registration, *IEEE Transactions on Medical Imaging,* 33 (2), 2014, p. 577-590.

[3] A. Prus, V. Kovalev, P. Vankevich. A method for lung segmentation in massive X-ray screening of the population. *International Journal of Computer Assisted Radiology and Surgery*, 3(1), 2009, p. 367-368.

[4] S. Jaeger et al. Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging,* 33 (2), 2014, p. 233-245.

[5] V.A. Kovalev, A.A. Kalinovsky. Big Medical Data: Image mining, retrieval and analytics, *Proceedings of the International Conference on Big Data and Predictive Analytics,* Belarus State University of Informatics and Radioelectronics, ISBN 978-985-543-146-7, Minsk, Belarus, June 2015, pp. 33-46.

[6] J. Schmidhuber. Deep Learning in Neural Networks: An Overview, *Neural Networks*, vol. 61, 2015, pp. 85–117.

[7] C.Farabet, C.Couprie, L.Najman, and Y.LeCun, Learning hierarchical features for scene labeling, *IEEE PAMI,* vol. 35 (8), 2013, pp. 1915–1929.

[8] V.Badrinarayanan A.Kendall, R.Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling, arXiv preprint, 2015, arXiv:1505.07293

[9] V.Badrinarayanan A.Kendall, R.Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, *arXiv preprint*, 2015, arXiv: 1511.00561.

[10]V.A.Kovalev, V.A.Lapizky, A.A.Dmitruk, A.A Kalinovsky. Big Data in Medicine: A database of chest radiographs for diagnosis, treatment, and scientific research goals, *Proceedings of the International Conference on Big Data and Predictive Analytics*, Belarus State University of Informatics and Radioelectronics, ISBN 978-985-543-146-7, Minsk, June 2015, pp. 66-71 (in Russian).

[11] http://imlab.grid.by/ Last visited 11.04.2016.

[12] http://tuberculosis.by/ Last visited 11.04.2016. Under construction.

[13] X.Glorot, A.Bordes, Y.Bengio. Deep sparse rectifier neural networks. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS),* 2011, Fort Lauderdale, FL, USA, vol 15 of JMLR, 2011, pp. 315-323.

[14] J.A. Stark. Adaptive image contrast enhancement using generalizations of histogram equalization, *IEEE Transactions on Image Processing,* vol. 9 (5), 2000, pp. 889-896.

[15] S.Lyu, E.P. Simoncelli. Nonlinear image represent-tation using divisive normalization. *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR-2008),* Anchorage, Alaska, USA, 23-28 June 2008, pp. 1-8.

[16] http://imlab.grid.by/appsegmxr/ Last visited 11.04.2016.