

LUNG SEGMENTATION WITH NASNET-LARGE-DECODER NET

Youshan Zhang

Computer Science and Engineering,
Lehigh University, Bethlehem, PA, USA
yoz217@lehigh.edu

ABSTRACT

Lung cancer has emerged as a severe disease that threatens human life and health. The precise segmentation of lung regions is a crucial prerequisite for localizing tumors, which can provide accurate information for lung image analysis. In this work, we first propose a lung image segmentation model using the NASNet-Large as an encoder and then followed by a decoder architecture, which is one of the most commonly used architectures in deep learning for image segmentation. The proposed NASNet-Large-decoder architecture can extract high-level information and expand the feature map to recover the segmentation map. To further improve the segmentation results, we propose a post-processing layer to remove the irrelevant portion of the segmentation map. Experimental results show that an accurate segmentation model with 0.92 dice scores outperforms state-of-the-art performance.¹

Index Terms— Lung segmentation, NASNet-Large Net, Encoder, Decoder

1. INTRODUCTION

Lung cancer is one of the malignant tumors with the fastest increase in morbidity and mortality and the greatest threat to human health and life. In the past fifty years, many countries have reported a marked increase in the incidence and mortality of lung cancer [2]. The lung-related disease is kept on the list of top ten causes of death in the country in the United States [3, 4]. Therefore, the prevention of lung cancer is important.

From the definition, lung cancer is a malignant lung tumor that is characterized by uncontrollable growth in the lung tissue. Clinicians find that the cure rate of early lung cancer is over 90%. Hence, the early diagnosis of lung cancer is essential since it could control the disease, reduce the mortality rate, and increase the patient's survival rate when the treatment is more likely curative. Chest radiograph (CXR) is commonly performed as diagnostic imaging for lung cancer and Pneumonia diagnosis. However, there are a number of factors, such as the positioning of the patient and depth of in-

spiration, that can affect the appearance of the CXR, complicating interpretation further. In addition, clinicians are faced with reading high volumes of images every shift. It is difficult for the physician to obtain an accurate diagnosis without the help of an additional tool. Therefore, it is necessary to develop segmentation tools to improve the effectiveness of the treatment.

However, designing an effective lung segmentation method is a challenging problem since the ROIs are often confused with the lung tissue. A large number of medical image analysis techniques have been proposed for lung segmentation, such as threshold method [5, 6, 7], region-based method [8, 9], genetic method [10], level set method [11, 12, 13] and artificial neural network [14, 15, 16], etc.

The threshold method is one of the most common and straightforward segmentation methods in lung segmentation. It is a region segmentation technology, which divides the gray value into two or more gray intervals, chooses one or more appropriate thresholds to judge whether the region meets the threshold requirement according to the difference between the target and the background, and separates the background and the target to produce a binary image. Threshold processing has two forms: global threshold and adaptive threshold. The global threshold only sets one threshold, and the adaptive threshold sets multiple thresholds. The target and background regions are segmented by determining the threshold at the peak and valley of the gray histogram [5, 6]. Level set methods are also widely used in the segmentation task. The basic idea of the level set method for image segmentation is the continuous evolution of curve motion. The boundary of the image is searched until the target contour is found and then the moving curve is stopped. Curves are moved along every three-dimensional section of images to slice different levels of the three-dimensional surface. The level of the obtained closed curves of each layer change over time, and finally get a corresponding shape extraction contour [11].

Deep neural networks have also been applied in segmentation and they supersede many traditional image segmentation approaches. Garcia published a review of deep learning approaches that aimed to present an overview of deep learning-based segmentation [17]. There are several models to address the segmentation task. Fully convolutional network (FCN)

¹This work is written before our previous paper [1].

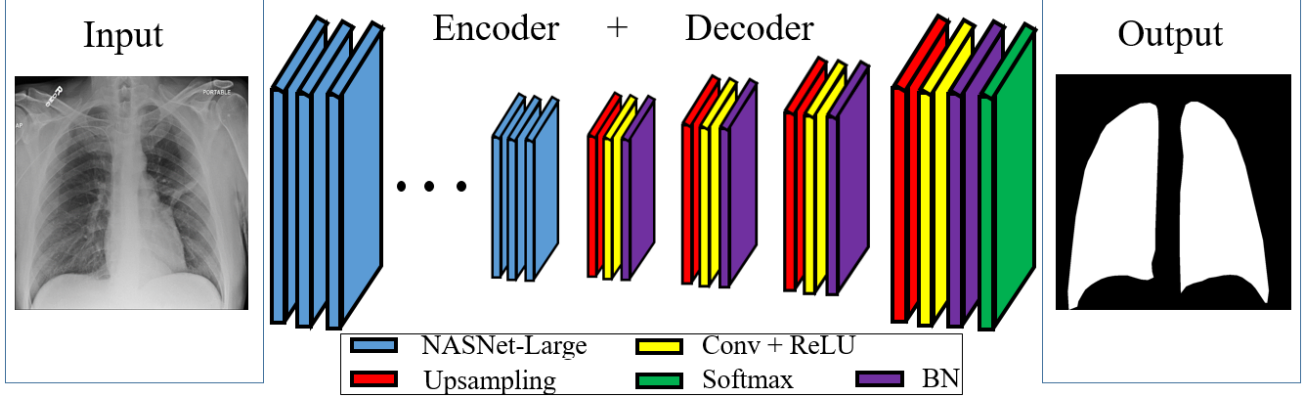


Fig. 1: The architecture of NASNet-Large segmentation net. The encoder consists of the first 414 layers from NASNet-Large model. There are four blocks in the decoder, and each block contains Upsampling, Conv+ReLU, and BN layer. (Convolution (Conv), Batch normalization (BN), Rectified linear units (ReLU)). The final decoder is fed into a softmax layer for pixel-wise lung prediction.

[18] is one of the earliest work for deep learning based image segmentation problem, which performs end to end segmentation. It is a convolution network for the dense prediction that does not need to pass through from the full connection layer. This method leads to the possibility of segmenting images of any size effectively, and it is much faster than the patch classification method. Almost all of other more advanced methods follow this architecture. However, there are several limitations of FCN model, such as the inherent spatial invariance causes the model fails to take into account of useful global context information and its efficiency in high-resolution scenarios is worse and is not available for real-time segmentation. Another problem that causes the difficulty using CNN networks in segmentation is the existence of pooling layers. The pooling layer not only enlarges the sensing field of the upper convolution layer but also aggregates the background and discards part of the location information. However, the semantic segmentation method needs to adjust the category map accurately, so it requires to retain the location information abandoned in the pooling layer. Later encoder-decoder architecture is widely used in the segmentation. Segnet [19] and U-net [20] are representative encoder-decoder architectures. It first selects a classification network such as VGG-16, and then removes its fully connection layer to produce low-resolution image representation or feature mapping. This part of the segmentation network is usually called an encoder. A decoder is part of the network, which can learn how to decode or map these low-resolution images to the prediction at the pixel level. The difference between different encoder-decoder architecture is the design of the decoder. Another type of methods uses dilated convolutions layers and removes the pooling layer structure. Chen et al. proposed Deeplab model, which used the dilated convolutions, and it used the fully connected conditional random field to realize the atrous spatial pyramid pooling (ASPP) in the spatial space [21].

In this paper, we provide a lung segmentation model using one of the common encoder-decoder architectures for image segmentation with a deep learning model called NASNet-Large-decoder. This architecture can erase unnecessary information provided in lung images. Our network achieves an accurate segmentation result with a 0.92 dice score.

Our contributions are in three folds:

1. We are the first to define a NASNet-Large-decoder net to segment the chest radiography images;
2. To remove the irrelevant small segmented parts, we are the first to propose a post-processing layer, which is able to filter out the false segmented section in prediction images;
3. Experiments results demonstrate that the proposed model achieves the highest dice and IoU score over state-of-the-art.

This paper is organized as follows: in section 2, the NASNet-Large-decoder net architecture is summarized; we present the segmentation results in section 3; In section 4, we discuss the advantages and disadvantage of proposed model and conclude in section 5.

2. METHOD

In this section, we first introduce the proposed segmentation net and then describe a post-processing layer for filter out the irrelevant segmented section in the prediction image.

2.1. NASNet-Large segmentation net

The NASNet-Large segmentation net contains an encoder and decoder, which is followed by a classification layer.

The architecture is shown in Fig.1. There are two significant differences in our model comparing with Segnet, which employs the pre-trained vgg16 network for encoder. Our NasnetLarge-decoder net uses the first 414 layers of NasnetLarge net (which is a well-trained classification net on ImageNet) as the encoder to decompose images [22]. We do not use the pre-trained weighted but retrain the layers using the new data to fit the NasnetLarge net in our experiment since our dataset is significantly different from the ImageNet. Another one is that the decoder is different and there are no pooling indices in our model since the NasnetLarge net can produce detailed information for the decoder.

An appropriate decoder in the decoder network can up-sample its input feature map using the max-pooling layer. The decoding technique is illustrated in Fig 1. There are four blocks in the decoder. Each block first begins with an up-sampling layer, which can expand the feature map, and then feature maps are followed by convolution and rectified linear units. A batch normalization layer is then applied to each of these maps. The first decoder, which is closest to the last encoder, can produce a multi-channel feature map. This is similar to the Segnet, which can generate a different number of sizes and channels as their encoder inputs. The final output of the last decoder is fed to a trainable soft-max classifier. And the output of this softmax layer is a K channel image of probabilities where K is the number of classes (two in our problem). The predicted segmentation corresponds to the class with maximum probability at each pixel.

2.2. Post-processing layer

However, there are some small parts that are not the true lung in the prediction result. We then propose a post-processing layer, which can filter the irrelevant part in the image. The first step of the post-processing layer is to classify the predicted image into several parts², and we then select the first two largest areas (the lung area) as the final segmented lung. As shown in the left image in Fig.2, the red box is the wrong prediction of the image (false negative). After the post-processing, the red box is removed. Therefore, the prediction result will be improved if we filter out these irrelevant parts in images.

3. RESULTS

3.1. Datasets and parameters

Our lung segmentation data is from RSNA pneumonia detection dataset. The whole dataset can be downloaded from <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>. To remove the unrelated features, we focus on the lung segmentation, the lung is manually

²The major step of the post-processing step is using the connectedComponentsWithStats function in OpenCV.

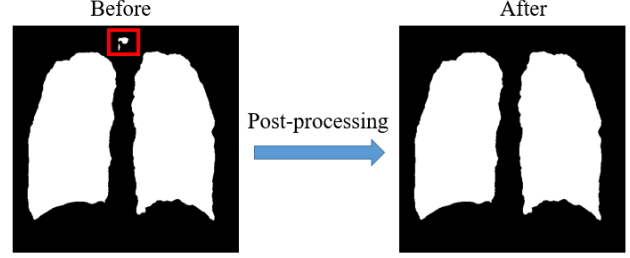


Fig. 2: The post-processing of prediction result. The left one is the prediction result from the proposed net, and the right one is the post-processing result using the proposed post-processing layer. The red box is an irrelevant feature.



Fig. 3: An example lung image in the training dataset (left: raw image; middle: ground truth mask image; right: mask overlay with raw image).

segmented³. There are 800 training images and 200 test images in this dataset. Fig. 3 shows an example of lung image in the training dataset.

Our experimentation is based on Keras, which is a high-level neural networks API, written in python, and is able to run on top of either TensorFlow or Theano, and runs seamlessly on CPU and GPU. In addition, our network was trained on a graphics processor NVIDIA TITAN Xp equipped with 12Gb of memory in order to exploit its computational speed. Hence, we run our code on a GPU in order to greatly accelerate the execution with the Tensorflow backend. The network parameters are set to:

1. Batch size: 4
2. Step size: 5
3. Number of epochs: 100

3.2. Metrics

To evaluate the performance of our NASNet-Large segmentation net, we use the dice coefficient index as a similarity metrics to indicate the goodness of the segmentation results since the dice score is currently widely used in the segmentation task. Furthermore, we also report the IoU score. The two metrics are defined in the following formula:

³The dataset is available at: https://github.com/YoushanZhang/Lung_Segmentation.

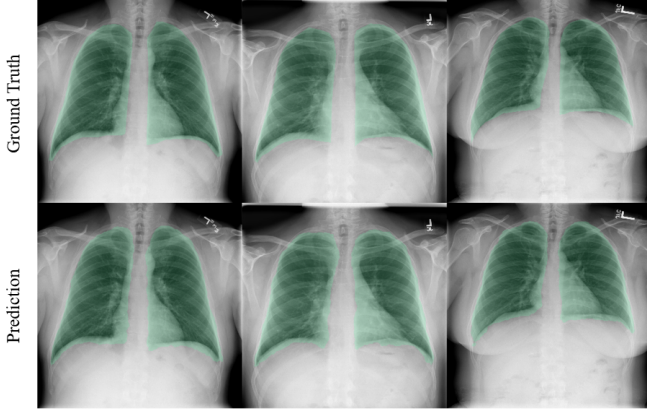


Fig. 4: Ground truth VS. prediction results. The first row is the ground truth mask overlay with the raw image, and the second row is the prediction result overlay with the raw image.

$$Dice = 2 \times \frac{|A \cap B|}{|A| + |B|}, \quad IoU = \frac{|A \cap B|}{|A \cup B|},$$

where A is ground truth mask, and B is the prediction mask.

Table 1: Segmentation results comparison

Methods	IoU	Dice score
Segnet [19]	0.82	0.87
U-Net [20]	0.84	0.88
Deep-Lab [21]	0.85	0.89
Nastnetlarge-net	0.86	0.91
Nastnetlarge-net-Post	0.87	0.92

3.3. Segmentation results

As shown in Fig.4, it compares the predicted segmented lung image with the ground truth image in the test dataset. The prediction image is close to the real mask, which demonstrates the high performance of our model. We also compare prediction results with state-of-the-art methods. Tab. 1 lists the comparison results of different models (Notice that the metrics are only reported on the lung area and exclude the background area). Our Nastnetlarge-net has a higher IoU and dice score than other models, and we also observe that the post-processing layer outputs the highest scores, which illustrates that the post-processing layer is useful in the segmentation tasks.

4. DISCUSSION

One of the distinct advantages of our model is that it achieves a higher dice and IoU score. And there are two reasons: the

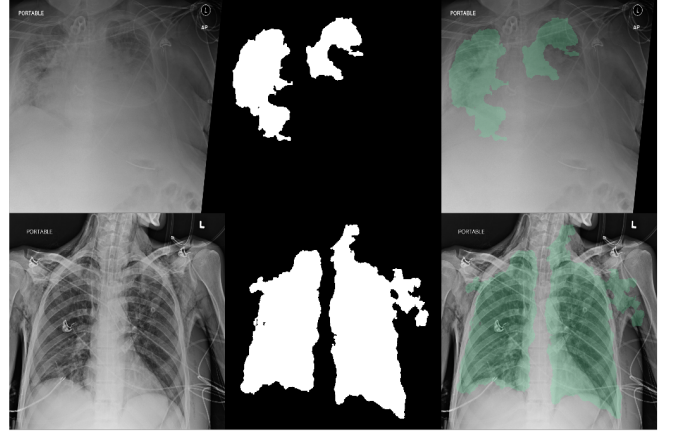


Fig. 5: Two failures in our model. The first row is caused by the light parts and the second row is caused by the dark parts.

designed NASNet-Large segmentation net is suitable for image segmentation, and the post-processing layer filters out the unnecessary parts in the image, which improves the segmented results. Although our model achieves a 0.92 dice score, it still fails in some cases. As shown in Fig. 5, we observe that the segmentation results are worse in these two situations. There are also two reasons, and one is that there is no similar image in the training dataset, which leads to low performance. Second is that our model is not robust enough to deal with some unusual images. Compared with the normal training images as shown in Fig. 3, two worse raw images (in Fig. 5) are either too light (first row) or too dark (second row), and it is even difficult for a human to distinguish the lung area and other tissues. A straightforward solution is to include more similar images in the training images.

5. CONCLUSION

In this paper, we are the first to present a lung segmentation using the NASNet-Large-decoder architecture, and we get an accurate segmentation with a 0.92 dice score. A post-processing layer is employed to remove the unnecessary part of the prediction map. The proposed model can also be applied to a wide area of different medical image segmentation tasks. Our objective in the next stage is to design a more robust encoder and decoder model for application in all different cases.

6. REFERENCES

- [1] Youshan Zhang, Brian D Davison, Vivien W Talghader, Zhiyu Chen, Zhiyong Xiao, and Gary J Kunkel, "Automatic head overcoat thickness measure with nasnet-large-decoder net," in *Proceedings of the Future Technologies Conference*. Springer, 2021, pp. 159–176.
- [2] Anthony J Alberg, Malcolm V Brock, Jean G Ford, Jonathan M

- Samet, and Simon D Spivack, "Epidemiology of lung cancer," *CHEST*, vol. 143, no. 5, pp. e1S–e29S, 2013.
- [3] Rui, P and Kang, K, "National ambulatory medical care survey: 2015 emergency department summary tables. table 27," https://www.cdc.gov/nchs/data/nhamcs/web_tables/2015_ed_web_tables.pdf, 2015.
- [4] Sherry L Murphy, Jiaquan Xu, Kenneth D Kochanek, Sally C Curtin, and Elizabeth Arias, "Deaths: final data for 2015," 2017.
- [5] Shiyang Hu, Eric A Hoffman, and Joseph M Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images," *IEEE transactions on medical imaging*, vol. 20, no. 6, pp. 490–498, 2001.
- [6] Jiantao Pu, Justus Roos, A Yi Chin, Sandy Napel, Geoffrey D Rubin, and David S Paik, "Adaptive border marching algorithm: automatic lung segmentation on chest ct images," *Computerized Medical Imaging and Graphics*, vol. 32, no. 6, pp. 452–462, 2008.
- [7] Joseph K Leader, Bin Zheng, Robert M Rogers, Frank C Sciarba, Andrew Perez, Brian E Chapman, Sanjay Patel, Carl R Fuhrman, and David Gur, "Automated lung segmentation in x-ray computed tomography: development and evaluation of a heuristic threshold-based scheme1," *Academic radiology*, vol. 10, no. 11, pp. 1224–1236, 2003.
- [8] Nilanjan Ray, Scott T Acton, Talissa Altes, Eduard E De Lange, and James R Brookeman, "Merging parametric active contours within homogeneous image regions for mri-based lung segmentation," *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 189–199, 2003.
- [9] Gao Guorong, Xu Luping, and Feng Dongzhu, "Multi-focus image fusion based on non-subsampled shearlet transform," *IET Image Processing*, vol. 7, no. 6, pp. 633–639, 2013.
- [10] Serhat Özkes, "Rule-based lung region segmentation and nodule detection via genetic algorithm trained template matching," *İstanbul Tıcaret Üniversitesi Fen Bilimleri Dergisi*, vol. 6, no. 11, pp. 17–30, 2007.
- [11] Luminita A Vese and Tony F Chan, "A multiphase level set framework for image segmentation using the mumford and shah model," *International journal of computer vision*, vol. 50, no. 3, pp. 271–293, 2002.
- [12] Margarida Silveira, Jacinto Nascimento, and Jorge Marques, "Automatic segmentation of the lungs using robust level sets," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 4414–4417.
- [13] Amal A Farag, Hossam E Abd El Munim, James H Graham, and Aly A Farag, "A novel approach for lung nodules segmentation in chest ct using level sets," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5202–5213, 2013.
- [14] Daw-Tung Lin, Chung-Ren Yan, and Wen-Tai Chen, "Autonomous detection of pulmonary nodules on ct images with a neural network-based fuzzy system," *Computerized Medical Imaging and Graphics*, vol. 29, no. 6, pp. 447–458, 2005.
- [15] Murat Ceylan, YÜKSEL ÖZBAY, OSMAN NURİ UÇAN, and Erkan Yildirim, "A novel method for lung segmentation on chest ct images: complex-valued artificial neural network with complex wavelet transform," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 18, no. 4, pp. 613–624, 2010.
- [16] Jinsa Kuruvilla and K Gunavathi, "Lung cancer classification using neural networks for ct images," *Computer methods and programs in biomedicine*, vol. 113, no. 1, pp. 202–209, 2014.
- [17] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.
- [18] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [19] Vijay Badrinarayanan, Alex Kendall, and Roberto 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.
- [20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [21] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.
- [22] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.