2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

LUNG SEGMENTATION IN CT IMAGES USING A FULLY CONVOLUTIONAL NEURAL NETWORK WITH MULTI-INSTANCE AND CONDITIONAL ADVERSARY LOSS

Tianyi Zhao[⋆] *Dashan Gao*[†] *Jiao Wang*[†] *Zhaozheng Yin*[⋆]

[⋆] Department of Computer Science, Missouri University of Science and Technology, MO, USA
[†]12 Sigma Technologies, CA, USA

ABSTRACT

Lung segmentation is usually the first step of lung CT image analysis and plays an important role in lung disease diagnosis. We propose an efficient end-to-end fully convolutional neural network to segment lungs with different diseases in CT images. We introduce a multi-instance loss and a conditional adversary loss to the neural network in order to solve the segmentation problem for more severe pathological conditions. Our method is capable of solving the lung segmentation problem under normal, moderate and severe pathological conditions, which is validated on 3 public benchmark datasets with different diseases.

Index Terms— Lung segmentation, CT, Fully convolutional neural network, Multi-instance, Conditional adversarial network

1. INTRODUCTION

Various lung diseases are causing millions of deaths worldwide. Detecting lung diseases at an early stage may dramatically lower the death rate. Screening individuals and applying computer aided diagnosis on the medical input, e.g. Computed Tomography (CT) images, enable the early stage detection of lung diseases. Lung segmentation is usually the first step for lung disease analysis systems. However, most of the existing lung segmentation methods can only work well on lungs with moderate diseases [1].

Traditional lung segmentation methods often deploy shape and region information [1, 2, 3, 4]. Most of the recent state-of-the-art segmentation methods in the medical field are deep-learning based methods. A fully convolutional neural network [5], named Unet, was proposed for cell segmentation [6]. An extension of 3D-Unet was proposed for xenopus kidney segmentation [7]. Several deep learning methods were proposed for brain segmentation problems [8, 9, 10]. Recently a fully convolutional neural network with multi-path enhancement is proposed for lung segmentation on interstitial lung disease databases [11].

Two sample patches of interstitial lung diseases are shown in Fig 1(b). The interstitial lung diseases often have spread textures which are considered moderate in this paper. Patches

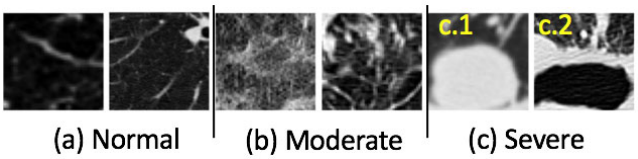

Fig. 1. Patch examples of different lung diseases. (a) Two patches belonging to normal lungs. (b) Two patches with moderate lung diseases. (c) Two patches with severe lung diseases.

from different lungs or the same lung at different locations may have different textures. Two normal lung patches are shown in Fig 1(a). Two severe lung disease patches are shown in Fig 1(c). Fig 1(c.1) has a macro-nodule attached on the lung wall; Fig 1(c.2) has fibro-cavernous. In this paper, we propose an end-to-end fully convolutional neural network to solve lung segmentation problem for various lung diseases, including interstitial lung diseases,cancer, and tuberculosis.

The first contribution of the paper is designing an efficient and effective fully convolutional network for lung segmentation. Efficiency is an important factor for real world diagnosis applications with more than one thousand scans, and a huge network will easily result in over-fitting due to the lack of annotated large-scale lung datasets in the lung segmentation community. The second contribution of the paper is proposing a multi-instance [12] loss to enhance our network on lung segmentation in moderate lung disease cases. Adding multi-instance loss to the intermediate stage of the neural network could enhance the diversity of kernels inside the preceding stage to detect more various low level texture patterns. The third contribution of the paper is that we deploy Conditional Adversarial Network [13, 14] which can learn the global representation of lung segmentation over the data collection to enhance our network on lung segmentation in severe lung disease cases.

2. METHODOLOGY

In this section, we introduce the three key technical components of our method: (1) the fully convolutional neural net-

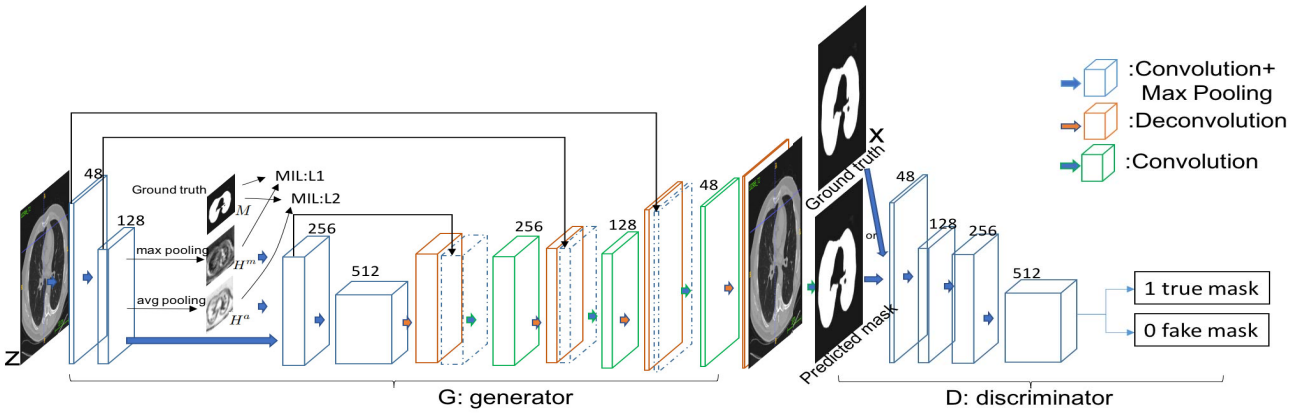

Fig. 2. The overview of our method. One the left side is the detailed implementation of multi-instance loss, as described in Section 2.2. The right side contains the generator and discriminator. The details of generator are described in Section 2.1. It contains 4 convolution layers followed by max-pooling layers and 4 deconvolution layers followed by concatenation layers and convolution layers. The filter size is 8x8 for the 1st convolutional layer, and 2x2 for others. The number of filters for each layer is 48,128,256,512,256,128,48,2. The discriminator contains 4 convolution layers followed by max-pooling layers. The filter size is 8x8 for the 1st convolutional layer, and 2x2 for others. The number of filters for each layer is 48,128,256,512,512,1.

work; (2) a multi-instance loss (MIL) added to the intermediate stage of the network; (3) a conditional adversarial loss to further enhance the network. Each of the three components focuses on the three kinds of diseases as shown in Fig 1(a)(b)(c): normal, moderate and severe.

These three components are illustrated in Fig 2. The fully convolutional neural network is denoted by "G:generator" in Fig 2. The details about multi-instance loss that are added to the intermediate stages of the network are shown in the left box of Fig 2. The "D:discriminator" on the right of the figure is a conditional adversarial network.

2.1. Fully Convolutional Neural Network

In this section, we first briefly review the Unet framework for biomedical image segmentation [6]. Unet first downsizes the input by 4 convolution blocks, then upscales by 4 up-sampling blocks to generate the segmentation map. Each down-sampling convolution block includes two 3x3 convolution layers, each of which is followed by a rectified linear unit and a 2x2 max pooling operation with stride 2. Each up-sampling block consists of one up-convolution layer, a concatenation with the output of corresponding down-sampling convolution blocks, and two 3x3 convolution layers. Unet achieves start-of-the-art performance for many biomedical image segmentation applications. However, since the disease image patches in lung CT images are much less than other patches in normal conditions, the large amount of parameters in the Unet (33M) tends to over-fit the network even after the data augmentation. So we design a more efficient fully convolutional neural network framework. We reduce the convolution and deconvolution block size and kernal size. Our network contains 4 down-sampling convolution layers and 4 up-sampling deconvolution layers. The 4 down-sampling convolution layers are followed by rectified linear units and max-pooling layers. Each of the 4 up-sampling deconvolution layer is followed by a concatenation layer, a convolution layer and rectified linear units. The framework is shown in Fig 2("G:generator"), the parameter size of our network is 4M ,which is much less than the Unet [6], 33M.

2.2. Multi-instance Loss

At pixel level, lung segmentation is a binary classification problem. However, different lung diseases have different textures. Some lung textures are visualized in Fig 1, which could be receptive fields for pixels in the intermediate stages of the network. Applying our multi-instance loss during stochastic gradient descent process lets each iteration update the most input-related kernels, in order to enhance the diversity of kernels. In this paper, we deploy the multi-instance loss to the second and third layers to learn more kinds of low-level texture patterns. Our multi-instance loss could be applied to any intermediate stages of any convolutional neural network, however for this binary classification problem, neurons in the layer abstract layers may not need more difference.

Multi-instance learning classifies bags instead of individual instances. Each bag has multiple instances, but only bag-level label could be given. Here we deploy the standard multi-instance learning assumption, which states that, if at least one instance in the bag is positive then the bag is positive, otherwise if and only if all instances in the bag are negative then the bag is negative. Each intermediate stage in the neural network can be regarded as a multi-instance problem. For one location (i,j) in the feature map $H \in R^{I \times J \times K}$, we have a bag $H_{i,j} \in R^K$. The bag contains K instances. Each instance

$H_{i,j,k}$ represents the activation score from the kth kernel in the preceding stage at location (i, j). If one kernel gives a positive activation at location (i,j), then the pixel at (i,j) is positive.

The multi-instance loss contains two parts, as shown in Fig 2, on the left. For positive pixels (inside lung regions), a bag-level predictive map H^m is calculated by max pooling from feature maps H along the k-axis (note, this is different from the spatial max pooling along the i- or j- axis). $H^m \in R^{I\times J}$. M is the downsized ground truth, $M \in R^{I\times J}$. $M_{i,j} = 1$, if pixel (i,j) is positive (lung regions), otherwise $M_{i,j} = -1$. The loss function for positive pixels is:

$$L1 = \sum_{i,j,M_{i,j}=1} log\left(1 + exp(-M_{i,j} \times H^m_{i,j})\right). \quad (1)$$

By minimizing the loss, we maximize the activation score H^m from the kernel that best matches the texture pattern in the pixel's receptive field.

For negative pixels (outside lung regions), a bag-level predictive map H^a is calculated by average pooling from feature maps H along the k-axis. $H^a \in R^{I\times J}$. The loss function for negative pixels is:

$$L2 = \sum_{i,j,M_{i,j}=-1} log\left(1 + exp(-M_{i,j} \times H^a_{i,j})\right). \quad (2)$$

By minimizing the loss, we minimize the activation score H^a from all kernels so that kernels only give positive scores for lung patterns.

The overall multi-instancce loss function is:

$$L_{MIL} = L1 + \lambda_1 L2 + \lambda_2 \|W\|_2, \quad (3)$$

where W represents weights in the neural network. λ_1, λ_2 are trade-off parameters (both are set as 10 in this paper). To train the whole network, we use stochastic gradient descent method (SGD) to minimize the sum of (1) cross-entropy loss at the final stage and (2) multi-instance loss at intermediate stages. We forward two predictive maps H^m and H^a to the next convolution layer to provide more segmentation information. The whole network is trained in an end-to-end scheme. Learning rate starts from 10^{-4} and decreases when converged.

2.3. Conditional Adversarial Networks

Our generator with the multi-instance loss works pretty well on most of lung CT images, but may fail in cases with severe pathologic conditions (Fig 3), which have Fibro-cavernous disease or Macro-nodule attached on the lung wall or severe interstitial lung disease. In those cases, it's too difficult to recognize those pathologic areas as part of lung regions by the sample-specific loss function, i.e. cross-entropy loss computes segmentation error for each individual image sample. We propose to use the Conditional Adversarial Networks to overcome this problem, which use global information over the entire data collection to help correct the lung segmentation mask on the pathologic areas.

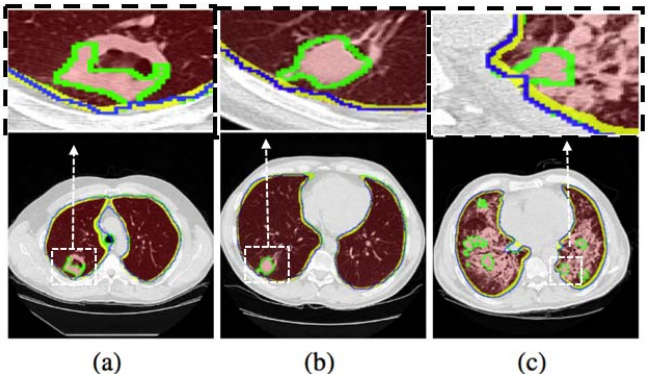

Fig. 3. Example of results. Yellow lines mean the boundary of ground truth; Green lines mean the mask boundary of Unet method; Blue lines mean the mask boundary of gNet+MIL+CAN method; Red masks represents mask generated by our Fusion method. (a) Fibro-cavernous; (b) Macro-nodule attached on the lung wall; (c) severe interstitial lung disease.

Conditional Adversarial Networks (CAN) [13] is a variation of Generative Adversarial Networks (GAN). GAN contains two parts: generator G and discriminator D. The Generator generates fake images from random noises. The Discriminator distinguishes fake images and true images. Generator G and discriminator D are trained interactively until they reach a balance.

The difference between CAN and GAN is that the input to the generator in CAN is an image instead of random noise. In this project, our generator loads a lung CT image as the input, and generates its segmentation mask. The generator framework is the same as the neural network with multi-instance loss, as described in Section 2.1 and 2.2. The enhanced network is shown in Fig 2. Generator G and discriminator D are trained by minmax the loss function:

$$\min_G \max_D L_{can} = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))], \quad (4)$$

where z presents the lung CT images given to the generator as input. x presents ground truth segmentation masks. Training discriminator D by maximizing the first term lets discriminator D give a high score for true labels. Training discriminator D by maximizing the second term lets discriminator D give a low score for fake segmentation masks generated by G. Training generator G by minimizing the second term lets generator G generate fake images to trick the discriminator. Discriminator is a global classifier among all samples. Due to this property, generator is trained to learn a more global segmentation mask over data collection. Thus, for cases in a severe pathologic condition which is difficult to segment by

Table 1. **The performance comparison between multiple approaches over three datasets: Dice-coefficient mean $\pm$ var$\times 10^{-1}$**

	LIDC	CLEF	Hug
Morph	0.862 ± 2.93	0.959 ± 0.04	0.769 ± 0.20
Unet	0.970 ± 0.59	0.985 ± 0.05	0.975 ± 0.09
gNet	0.962 ± 0.09	0.961 ± 0.06	0.938 ± 0.53
gNet+MIL	0.982 ± 0.05	0.984 ± 0.02	0.973 ± 0.08
gNet+MIL +CAN	0.949 ± 0.09	0.954 ± 0.06	0.883 ± 0.11
Fusion	**0.983 $\pm$0.05**	**0.985$\pm$ 0.02**	**0.976 $\pm$ 0.07**

itself, the pathologic area could be successfully segmented by CAN method. As show in Fig 3, the blue lines represent the mask boundary generated by CAN, which successfully cover the disease areas.

3. EXPERIMENT

3.1. Dataset

We evaluate the proposed method on three CT datasets. The first dataset is the Lung Image Database Consortium image collection (LIDC) [2, 15], LIDC dataset contains 1018 patients, each of which belongs to one of the three categories: "nodule $\geq$ 3 mm", "nodule $<$ 3 mm", and "non-nodule $\geq$ 3 mm". The second dataset is from ImageCLEFtuberculosis 2017 task (CLEF) [16, 17]. CLEF dataset contains 500 patients, each of which belongs to one of five tuberculosis types: Infiltrative, Focal, Tuberculoma, Miliary, Fibro-cavernous. The last dataset is a multimedia collection of cases with interstitial lung diseases from the University Hospitals of Geneva (HUG) [18]. HUG dataset contains 128 patients, each of which belongs to one of the 13 histological diagnoses of interstitial lung diseases.

3.2. Evaluation of Segmentation

The evaluation metric is 3D Dice-coefficient: $\frac{2|M' \cap M|}{|M'|+|M|}$. M denotes the ground truth segmentation mask, and M' denotes the generated segmentation mask. The methods to be compared include: (1) Morph. A benchmark image processing method based on thresholding, morphological operations and component analysis [1]. (2) Unet [6]. (3) gNet. This is the fully convolutional neural network we described in Section 2.1. (4) gNet+MIL. We add the multi-instance loss with cross-entropy loss to train the fully convolutional neural network, as described in Section 2.2. (5) gNet+MIL+CAN. We further add the conditional Adversarial loss to the fully convolutional neural network, as described in Section 2.3. (6) Fusion. We fuse the generated masks from two networks gNet+MIL and gNet+MIL+CAN together.

Table 2. **The performance comparison between multiple approaches over three datasets on the severe disease area. The values in the 2nd-4th column are the mean dice-coefficient. The values in the last column are the parameter sizes.**

	LIDC	CLEF	Hug
Unet	0.8925	0.9319	0.9074
gNet+MIL	0.8862	0.8808	0.9396
gNet+MIL+CAN	**0.9176**	**0.9613**	**0.9793**

The quantitative results and some qualitative results are shown in Table 1 and Fig 3, respectively. From Fig 3, we can see that although gNet+MIL+CAN method works very well on the severe area, its dice-coefficient doesn't outperform the gNet+MIL method. The reason is that CAN method has a trade-off between local and global results. CAN gets the benefit of better performance at the severe area, but it loses some precision on the boundary. Thus, to achieve the best result, we fuse these two methods together to get the best result. Using M^s denotes the segmentation mask generated from gNet+MIL, M^{sc} denotes the segmentation mask generated from gNet+MIL+CAN, and M^F denotes the segmentation mask generated from the Fusion method, the fusion steps include: getting the difference $M^d = M^{sc} \setminus M^s$; opening on M^d; $M^F = M^d \cap M^s$. As we can see from Table 1, our fusion method achieves the best performance. It's worth noting that we are using High-resolution CT images, which means the volume of each lung could contain more than 5M pixels. 0.001 better dice-coefficient means correctly covering 5k more pixels which are larger than an obvious nodule, as shown in Fig 3.

3.3. Evaluation of Segmentation on Severe Area

To better show the advantage of our algorithm, we use 3D-Dice-coefficient on the severe disease areas to evaluate the proposed method. Table 2 shows the comparison result from three methods: Unet; gNet+MIL; gNet+MIL+CAN, from which we can observe that our method (gNet+MIL+CAN) outperforms Unet and gNet+MIL in severe areas.

4. CONCLUSION

We proposed an efficient fully convolutional neural network, which is enhanced by adding a multi-instance loss and a conditional adversary loss. Our segmentation method is suitable for lung CT images with various diseases, which is validated on three public benchmark datasets.

5. REFERENCES

[1] Awais Mansoor, Ulas Bagci, Brent Foster, Ziyue Xu, Georgios Z Papadakis, Les R Folio, Jayaram K Udupa, and Daniel J Mollura, "Segmentation and image analysis of abnormal lungs at ct: current approaches, challenges, and future trends," *RadioGraphics*, vol. 35, no. 4, pp. 1056–1076, 2015.

[2] Eva M. van Rikxoort, Bartjan de Hoop, Max A. Viergever, Mathias Prokop, and Bram van Ginneken, "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection," *Medical Physics*, vol. 36, no. 7, pp. 2934–2947, 2009.

[3] Shuangfeng Dai, Ke Lu, Jiyang Dong, Yifei Zhang, and Yong Chen, "A novel approach of lung segmentation on chest {CT} images using graph cuts," *Neurocomputing*, vol. 168, pp. 799 – 807, 2015.

[4] Neil Birkbeck, Timo Kohlberger, Jingdan Zhang, Michal Sofka, Jens Kaftan, Dorin Comaniciu, and S. Kevin Zhou, *Lung Segmentation from CT with Severe Pathologies Using Anatomical Constraints*, pp. 804–811, 2014.

[5] Evan Shelhamer, Jonathan Long, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.

[6] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, vol. 9351 of *LNCS*, pp. 234–241.

[7] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 424–432.

[8] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 2016, pp. 565–571.

[9] Hao Chen, Qi Dou, Lequan Yu, and Pheng-Ann Heng, "Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation," *CoRR*, vol. abs/1608.05895, 2016.

[10] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L. Rubin, and Bradley J. Erickson, "Deep learning for brain mri segmentation: State of the art and future directions," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 449–459, Aug 2017.

[11] Adam P. Harrison, Ziyue Xu, Kevin George, Le Lu, Ronald M. Summers, and Daniel J. Mollura, "Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2017.

[12] Jiajun Wu, Yinan Yu, Chang Huang, and Kai Yu, "Deep multiple instance learning for image classification and auto-annotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3460–3469.

[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros, "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016.

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, pp. 2672–2680. 2014.

[15] Bidaut L McNitt-Gray MF Meyer CR Reeves AP Zhao B Aberle DR Henschke CI Hoffman EA Kazerooni EA MacMahon H van Beek EJR Yankelevitz D et al Armato SG III, McLennan G, "The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans," *Medical Physics*, vol. 38, pp. 915–931.

[16] Yashin Dicente Cid, Alexander Kalinovsky, Vitali Liauchuk, Vassili Kovalev, , and Henning Müller, "Overview of ImageCLEFtuberculosis 2017 - predicting tuberculosis type and drug resistances," in *CLEF2017 Working Notes*, Dublin, Ireland, September 11-14 2017, CEUR Workshop Proceedings.

[17] Yashin Dicente Cid, Oscar Alfonso Jiménez del Toro, Adrien Depeursinge, and Henning Müller, "Efficient and fully automatic segmentation of the lungs in ct volumes," in *Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE ISBI*, May 2015, CEUR Workshop Proceedings, pp. 31–35.

[18] Platon A Geissbuhler A-Poletti PA Mller H. Depeursinge A, Vargas A, "Building a reference multimedia database for interstitial lung diseases," *Computerized Medical Imaging and Graphics*, vol. 36, no. 3, pp. 227–238.