



HAL
open science

LLASN:Lung lobes segmentation using adversarial network

Qingmei Chen, Hui Tang, Jean-Louis Dillenseger

► **To cite this version:**

Qingmei Chen, Hui Tang, Jean-Louis Dillenseger. LLASN:Lung lobes segmentation using adversarial network. 6th International Conference on Signal and Image Processing (ICSIP 2021), Oct 2021, Nanjing, China. pp.240-244, 10.1109/ICSIP52628.2021.9688972 . hal-03403373

HAL Id: hal-03403373

<https://hal.science/hal-03403373v1>

Submitted on 26 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LLASN:Lung lobes segmentation using adversarial network

Qingmei Chen

Laboratory of Image Science and
Technology, School of Computer
Science and Engineering,
Southeast University, Nanjing, China
Centre de Recherche en Information
Biomédicale sino-français,
Laboratoire International Associé,
INSERM, Université Rennes 1, France,
Southeast University, Nanjing, China

Hui Tang

Laboratory of Image Science and
Technology, School of Computer
Science and Engineering,
Southeast University, Nanjing, China
Centre de Recherche en Information
Biomédicale sino-français,
Laboratoire International Associé,
INSERM, Université Rennes 1, France,
Southeast University, Nanjing, China
corinna@seu.edu.cn

Jean-Louis Dillenseger

Univ Rennes, Inserm, LTSI - UMR 1099
F-35000 Rennes, France
Centre de Recherche en Information
Biomédicale sino-français,
Laboratoire International Associé,
INSERM, Université Rennes 1, France,
Southeast University, Nanjing, China
jean-louis.dillenseger@univ-rennes1.fr

Abstract—The segmentation of the lung lobes is generally a prerequisite for some specific diagnostic tasks and even more so for the planning of the therapy of lung diseases. Recently this task has been tackled with the help of deep learning and more particularly by using U-Net networks. However, these networks have great difficulty in segmenting the lobes because the lung is extremely heterogeneous and the fissures between lobes are extremely fine and therefore blurred in the images. This leads to relatively inaccurate segmentation results, especially in terms of spatial continuity over the 2D lung. Inspired by the adversarial networks in medical image segmentation, we propose in this paper a new segmentation scheme: the lung lobes adversarial based segmentation network (LLASN). In this scheme U-Net is used to generate segmentation results and a discriminator network is used to discriminate the generated segmentation results from Ground Truth labels. The proposed method is evaluated on the LUNA16 dataset and the Tianchi Medical AI Competition dataset. Compared to the classical U-Net network, the proposed network improved Dice similarity coefficient from 0.82 to 0.84, 0.83 to 0.84, and 0.59 to 0.64 for left lower lobe, left upper lobe, and right middle lobe in LUNA16 dataset.

Index Terms—lung lobes segmentation, adversarial network, U-Net network

I. INTRODUCTION

Segmentation of the lung lobes is particularly important for the diagnosis, localization or treatment planning of lung diseases such as emphysema or pulmonary nodules. Anatomically, the left lung is divided into the left upper (LUL) and the left lower (LLL) lobes, and the right lung is divided into the right upper (RUL), the right middle (RML), and the right lower (RLL) lobes [1]. Anatomically, these lobes are separated by fine fissures that are blurred or not always visible on CT images. This makes segmentation of lobes particularly difficult because of the incomplete nature of the fissures and the presence of other structures and abnormalities in the lung parenchyma surrounding them. Pulmonary lobes segmentation algorithms can be divided into three categories based on fissure

surface information, lobar boundaries and pulmonary atlases [2]. However the conventional Pulmonary lobes segmentation algorithms are time-consuming, so deep learning networks approaches are now applied for this task. Among all the methods U-Net [3] is widely used in the context of medical image segmentation because of its generalizability and performance. However when applied for lung lobes segmentation, the U-Net has a rather weak performance especially because it has difficulty to treat the spatial continuity of the lobes fissures. In order to improve this spatial continuity of the segmentation results we propose to include U-Net in a new framework inspired by the generative adversarial network(GAN) [4]: the lung lobes adversarial based segmentation network (LLASN). In this framework U-Net is used as a generator network to generate segmentation results. It is followed by a discriminator network that aims to discriminate the generated segmentation results from the Ground Truth labels based on the input CT images. Our expectation is that the loss function provided by the discriminator network encourages the generator network to correct segmentation errors and to produce more accurate and continuous results. To verify this assumption, we will evaluate our presented method on two open access datasets: LUNA16 dataset [5] and the Tianchi Medical AI Competition dataset. The paper is structured as following, the proposed framework with the generator and discriminator networks is presented on section II while section III is focused on the experimental results and analysis.

II. METHOD

The classical GAN aims to generate new data from the given dataset using an adversarial process. However, Luc et al. [6] proposed to replace the generator network by a segmentation-dedicated CNN network for natural image semantic segmentation. We will exploit this idea for the segmentation of the lobes of the lung.

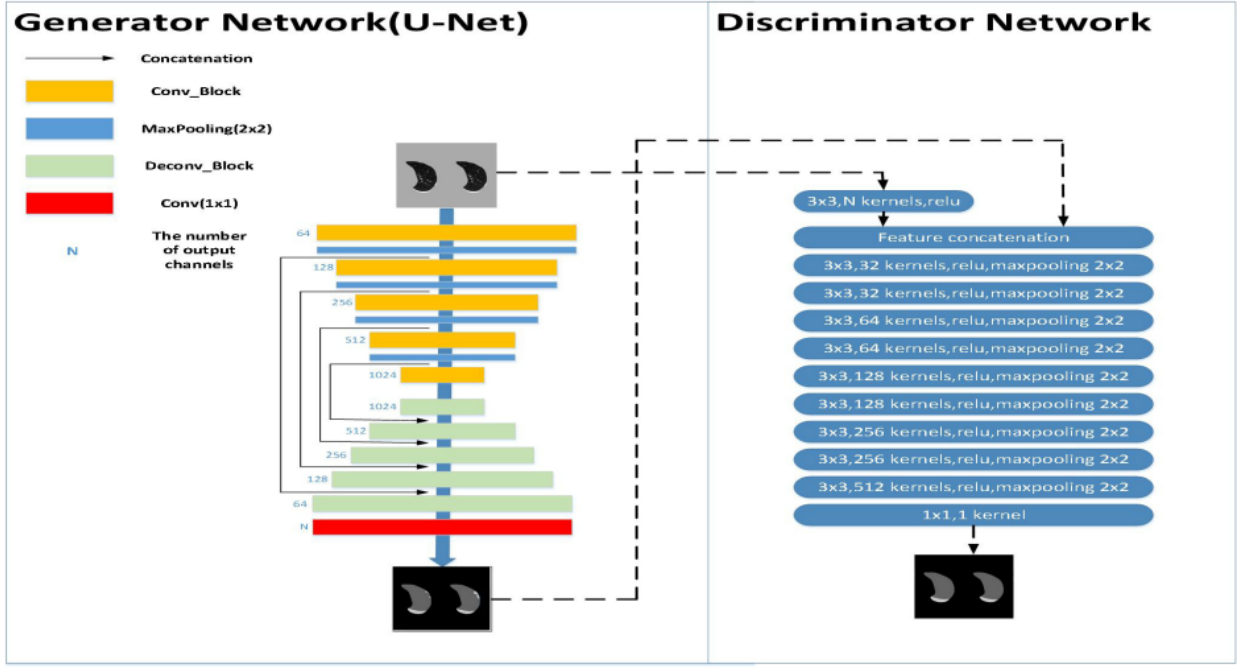


Fig. 1. The structures of proposed LLASN framework. On the left, the generator network which is the conventional U-Net framework and on the right, the discriminator framework

As for the classical generative adversarial network the LLASN includes a regular generator network and a discriminator network which are connected together (Fig. 1).

A. Generator Network

This conventional generator network is here a 2D U-Net network. Sub-volumes of CT images are used as the input of the U-Net network, and Ground Truth labels are used as the target of the segmentation network to help the generative network learn to produce better results.

The structure of the U-Net-based generator network is described as follows (Fig. 1-left):

- The U-Net contains an encoding path and a decoding path.
- In the encoding path, each layer contains a convolution block (in orange) composed by two 3×3 convolutional layers followed by a ReLu layer, and finally there is 2×2 max-pooling layer (in blue) in each direction.
- In the decoding path, each layer contains a deconvolution block (in green) composed by one 2×2 deconvolution layer and two 3×3 convolution layers followed by one ReLu layer.
- After a 2×2 deconvolution, there is a concatenation of the feature maps with the correspondingly layer of the same resolution from the encoding path (gray links). This provides localization information from encoding path to decoding path.
- A final 1×1 convolution layer (in red) is used to map the feature map to the desired number of classes.

The input to the generator network is a 512×512 pixel CT image. The generator loss function mainly consists of

two parts, called generator loss and adversary loss. The loss function is as follows:

$$Loss_G = Loss_{cce}(G(I), GT) - \lambda Loss_D^G \quad (1)$$

The first item, $Loss_{cce}(G(I), GT)$, calculates the categorical cross-entropy loss between the Ground Truth labels and the predicted labels. I represents the input CT image in the generating network. GT is the Ground Truth label corresponding to I . The second element, $Loss_D^G$, is called the adversarial network loss and will be described later. λ is a balancing term between the two loss terms and is set to 2 in our case.

B. Generator Network

The discriminator network (Fig. 1-right) is a kind of coding network. The labels generated by the U-Net network and the Ground Truth labels are fed into the discriminator network in one hot coding. At the same time, the input CT images are also introduced as part of the loss function. The discriminator network contains four layers, each layer containing two 3×3 convolutional layers followed by a ReLu layer and a 2×2 max-pooling layer. Finally, a 1×1 convolutional layer is used to discriminate whether the input label is a generated label or a Ground Truth label. The loss function is calculated as follows:

$$Loss_D = Loss_{be}(I, G(I), 0) + Loss_{be}(I, GT, 1) \quad (2)$$

$Loss_{be}(I, G(I), 0)$ is the binary cross-entropy loss between an input CT image I , the predicted label by the generator

network, and label 0, or the input CT image, the Ground Truth label, and label 1. The adversarial network can so encourage the generator network to correct errors in the segmentation results.

C. Adversarial loss

The purpose of the adversarial loss generated by the discriminator network is to increase the probability that the prediction result of the generator network is similar to the Ground Truth label. The adversarial loss is calculated as follows:

$$Loss_D^G = Loss_{be}(I, G(I), 0) \quad (3)$$

The adversarial loss is generated by the coupled training of the adversarial network and the generator network together during the learning process. The adversarial loss is fed back to the generator network, prompting the generator network to better evaluate the loss between the generated and the Ground Truth labels.

But, during the learning process, the GAN network has some difficulties to train and converge. To circumvent this difficulty, the generator network should be trained alone for a while in order to produce some pre-trained weights. Then the generator network and the discriminator network are trained together until the loss function no longer decreases.

III. EXPERIMENTAL MATERIALS AND RESULTS ANALYSIS

A. Experimental materials

The dataset selected for our lung lobes segmentation experiment came initially from the LUNA16 challenge and Tianchi Medical AI Competition. The LUNA16 is composed of 3500 images. We used 2000 images for network training and 1500 for network testing. As Ground Truth we used the manual lobe labeling on LUNA16 provided by Tang et al. [7]. 260 CT images were selected to test the robustness of the network in the Tianchi Medical AI Competition. As Ground Truth, we manually labeled ourselves these 260 CT images using Itk-Snap.

The original CT images were preprocessed before using them in the network. First, the left and right lungs were separated by half. Then a simple threshold followed by a connected regions search were used to extract the lungs and remove the background. Because, the two largest connected regions in 2D are the left and right lung it was easy to adjust a 2D bounding box to obtain the lung region. Then, the processed image was normalized to 0-1.

As the network is designed to work in 2D, we wanted to add some 3D spatial information. To do this we combined two successive slices from the CT volume to form a $512 \times 512 \times 2$ vector image. It is this image that is provided as input to the network.

TABLE I
THE MEAN OF THE DICE COEFFICIENTS OBTAINED USING THE LLASN FRAMEWORK ON BOTH DATASETS ARE HIGHER THAN THESE QUANTITATIVE COMPARISON BETWEEN U-NET AND LLASN FOR LOBES SEGMENTATION ON THE LEFT LUNG. MEAN DICE, MEAN PRECISION AND MEAN RECALL OBTAINED ON THE LEFT LOWER LOBE (LLL) AND THE LEFT UPPER LOBE (LUL) ON OUR TWO DATASETS. IN BOLD THE BEST SCORES.

| | Method | Dice | | Precision | | Recall | |
|---------|--------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | LLL | LUL | LLL | LUL | LLL | LUL |
| LUNA16 | U-Net | 0.82 | 0.83 | 0.91 | 0.89 | 0.90 | 0.88 |
| | LLASN | 0.84 | 0.84 | 0.93 | 0.88 | 0.89 | 0.93 |
| Tianchi | U-Net | 0.82 | 0.80 | 0.91 | 0.84 | 0.88 | 0.89 |
| | LLASN | 0.85 | 0.82 | 0.93 | 0.82 | 0.85 | 0.92 |

B. Evaluation Metrics

In order to evaluate the accuracy of our lobes segmentation, we used the following metrics to compare the results obtained by our methods with the Ground Truth.

- Dice similarity coefficient: The Dice coefficient is generally used to calculate the similarity of two samples, and the range is $[0, 1]$. The calculation formula is as follows:

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|}$$

where X represents the segmentation results estimated by the generator network, and Y represents the Ground Truth labels. The higher the coefficient of Dice, the better the performance of the network.

- The accuracy is the ratio of the number of correct positive predictions to the total number of positive prediction. The formula is as follows:

$$Precision = \frac{TP}{TP + FP}$$

Where TP (True Positives) indicates that the sample is positive and that the prediction result is positive also. FP (False Positives) means that the sample is negative but the prediction result is positive.

- Recall: Recall indicates the ratio of the number of correct positive predictions to the total number of positive samples. The formula is as follows:

$$Recall = \frac{TP}{TP + FN}$$

Where TP (True Positives) indicates that the sample is positive and that the prediction result is positive also. FN (False negatives) means that the sample is positive but the prediction result is negative.

C. Experimental Results

All experiments were tested on the keras framework. Both the generator network and the discriminator network were trained using the Adam optimizer with default parameters $\beta_1=0.9$, $\beta_2=0.999$, and the learning rate set to $1e-5$. The input size of the generator network and discriminator network are $512 \times 512 \times 2$ and $512 \times 512 \times N$, respectively

TABLE II

QUANTITATIVE COMPARISON BETWEEN U-NET AND LLASN USING FOR LOBES SEGMENTATION ON THE RIGHT LUNG. MEAN DICE, MEAN PRECISION AND MEAN RECALL OBTAINED ON THE RIGHT UPPER LOBE (RUL), THE RIGHT MIDDLE LOBE (RML) AND THE RIGHT LOWER LOBE (RLL) ON OUR TWO DATASETS. IN BOLD THE BEST SCORES.

| | Method | Dice | | | Precision | | | Recall | | |
|---------|--------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | RUL | RML | RLL | RUL | RML | RLL | RUL | RML | RLL |
| LUNA16 | U-Net | 0.82 | 0.52 | 0.83 | 0.84 | 0.77 | 0.88 | 0.83 | 0.66 | 0.89 |
| | LLASN | 0.82 | 0.64 | 0.82 | 0.89 | 0.76 | 0.87 | 0.87 | 0.74 | 0.92 |
| Tianchi | U-Net | 0.71 | 0.56 | 0.83 | 0.79 | 0.69 | 0.86 | 0.83 | 0.54 | 0.89 |
| | LLASN | 0.71 | 0.63 | 0.83 | 0.86 | 0.73 | 0.84 | 0.83 | 0.64 | 0.91 |

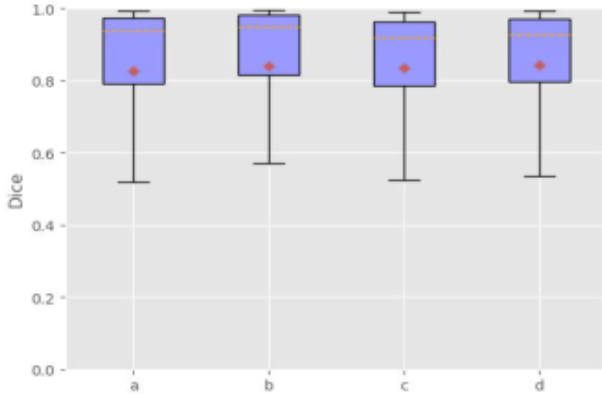


Fig. 2. Left lung lobes from the LUNA16 dataset labeled by Tang [7]: boxplots of the Dice similarity coefficients measured for (a) LLL results obtained by U-Net, (b) LLL results obtained by LLASN, (c) LUL result obtained by U-Net and (d) LUL results obtained by LLASN

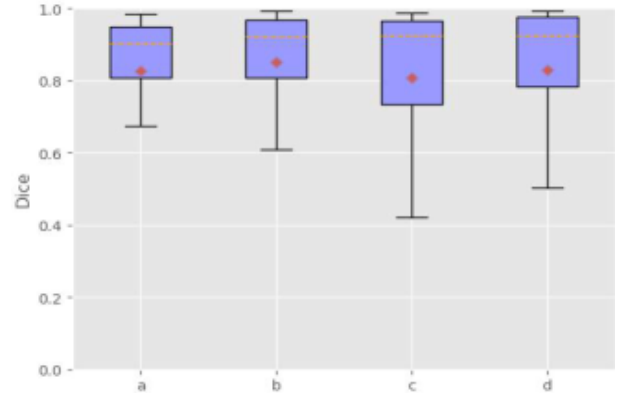


Fig. 3. Left lung lobes from Tianchi Medical AI Competition dataset labeled by Itk-Snap: boxplots of the Dice similarity coefficients measured for (a) LLL results obtained by U-Net, (b) LLL results obtained by LLASN, (c) LUL results obtained by U-Net and (d) LUL results obtained by LLASN

($N = 4$ for the left lobe segmentation, $N = 5$ for the right lobe segmentation).

Table I and Table II shows the mean scores (Dice coefficients, Precision and Recall) obtained using our LLASN framework and the classical U-Net to segment the lobes on the two testing datasets: LUNA16 labeled by Tang [7] and Tianchi Medical AI Competition labeled using Itk-Snap. Table I shows the mean score for the left lobe segmentation and Table II for the right lobe segmentation.

However, mean scores are not always representative of the quality of the segmentation on a huge number of data. The distributions of these scores is also a good quality indicator. Fig. 2 to 5 shows the Dice similarity coefficients distributions obtained for each of the five lobes on the images of the two testing datasets: LUNA16 labeled by Tang [7] and Tianchi Medical AI Competition labeled using Itk-Snap. We always compared the results obtained by LLASN to these obtained by the classical U-Net alone.

The mean of the Dice coefficients obtained using the LLASN framework on both datasets are higher than these using the U-Net, and the variances of the Dice coefficients obtained using the LLASN framework are smaller than these using U-Net, indicating that LLASN can indeed improve the accuracy of lung lobes segmentation and has better robustness. This is especially true for the segmentation of the right lung. In fact, the right lung presents 3 classes (instead of 2 classes for the left lung).

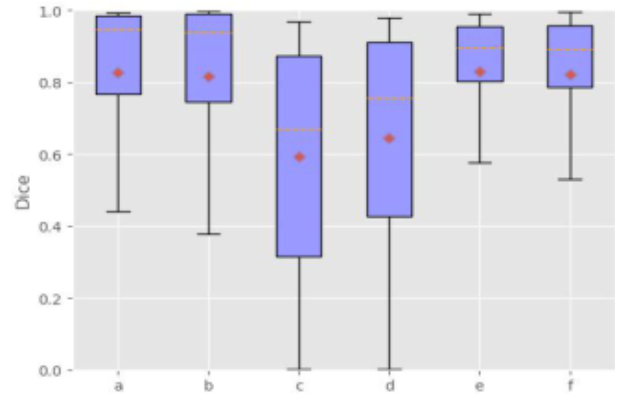


Fig. 4. Right lung lobes from the LUNA16 dataset labeled by Tang [7]: boxplots of the Dice similarity coefficients measured for (a) RUL results obtained by U-Net, (b) RUL results obtained by LLASN, (c) RML results obtained by U-Net and (d) RML results obtained by LLASN, (e) RLL results obtained by U-Net and (f) RLL results obtained by LLASN

The Right Middle lung is rather difficult to segmented because it is smaller and moreover embedded and twisted between the two other one. So two interfaces, one between the upper and the middle lobes and one between the middle and the lower lobes, must be estimated. In some slices all the 3 lobes can be seen. This Right Middle lobe segmentation difficulty can be noticed on the scores which are always lower than this of the other lobes. Also the distribution of

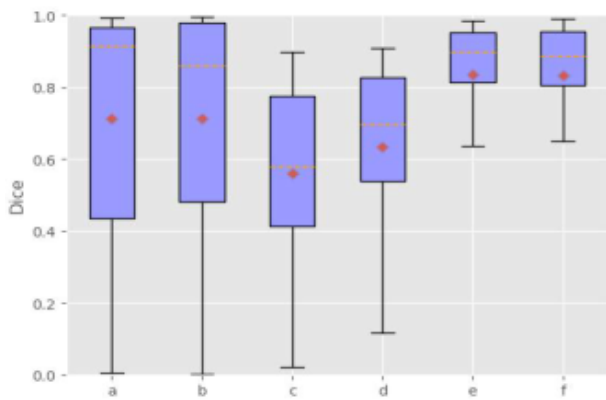


Fig. 5. Right lung lobes from the Tianchi Medical AI Competition dataset labeled by Itk-Snap: boxplots of the Dice similarity coefficients measured for (a) RUL results obtained by U-Net, (b) RUL results obtained by LLASN, (c) RML results obtained by U-Net and (d) RML results obtained by LLASN, (e) RLL results obtained by U-Net and (f) RLL results obtained by LLASN

the Dice coefficient is also larger for the Right Middle lobe than for the other lobes. LLASN always shows better scores and smaller distributions than U-Net in this difficult task. The effectiveness of LLASN compared to U-Net is so verified by the experimental results.

The better behavior of LLASN compared to U-Net for lobes segmentation can be observed on some visual segmentation examples (Fig. 6). We can see that often the label map segmented by U-Net lacks spatial continuity, while the LLASN scheme we proposed, is more accurate and continuous. This is particularly the case on the fourth row of Fig. 6 which shows the segmentation of the right lobe at the level of the 3 lobes. The LLASN shows results much closer to the Ground Truth than U-Net.

IV. CONCLUSION

In this study, we proposed a segmentation of the lung lobes based on an adversarial network. In this framework, the U-Net, classically used to do segmentation, is put into the GAN as a label generator. The discriminator network of the GAN is then used to encourage the generator to obtain better results and especially a better spatial continuity. The results obtained on two datasets, LUNA 16 and Tianchi Medical AI Competition, demonstrate that the LLASN scheme can enhance spatial continuity and improve accuracy especially for the segmentation of the right middle lobe which is considered as rather difficult. In addition, the robustness of the LLASN scheme is also demonstrated in this study. The adversarial loss obviously encouraged the generative network to segment the lobes to produce more accurate results. In the future, the LLASN scheme will be applied to 3D lung lobes segmentation for further improvement.

ACKNOWLEDGMENTS

This work was supported by the French National Research Agency (ANR) through the VATSop project (ANR-20-CE19-0015).

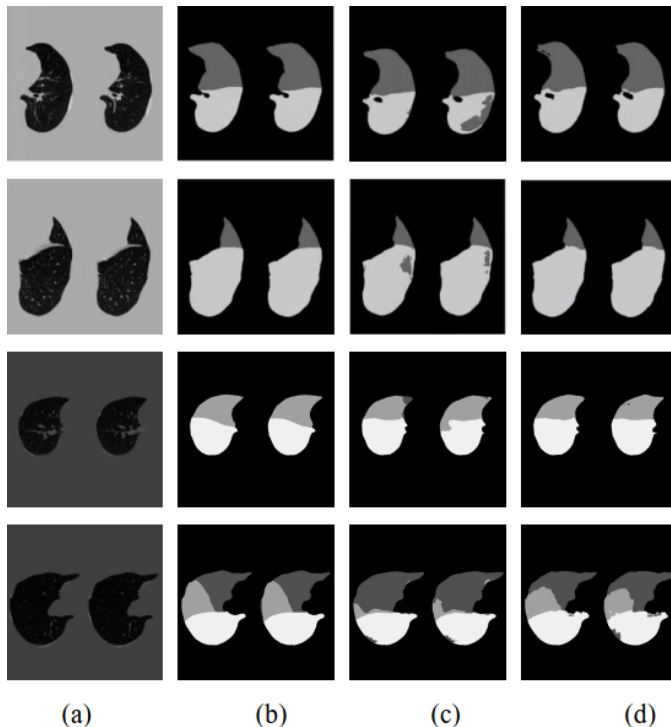


Fig. 6. 4 visual segmentation examples on left lungs (the 2 upper rows) and right lungs (the 2 lower rows). On each row we have (a) the preprocessed input CT image, (b) the Ground Truth, (c) the labels estimated by U-Net and (d) the labels estimated by the proposed LLASN

REFERENCES

- [1] Y. Peng, H. Zhong, Z. Xu, et al., "Pulmonary lobe segmentation in CT images based on lung anatomy knowledge," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–15, 2021.
- [2] T. Doel, D. J. Gavaghan, and V. Grau, "Review of automatic pulmonary lobe segmentation methods from CT," *Computerized Medical Imaging and Graphics*, vol. 40, pp. 13–29, 2015.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *MICCAI 2015*. 2015, vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241, Springer.
- [4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., "Generative Adversarial Nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [5] A. Setio, A. Traverso, T. de Bel, et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, 2017.
- [6] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using Adversarial Networks," in *NIPS Workshop on Adversarial Training*, Barcelona, Spain, Dec. 2016.
- [7] H. Tang, C. Zhang, and X. Xie, "Automatic pulmonary lobe segmentation using Deep Learning," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019.