A FRAMEWORK FOR ONE-SHOT LUNG SEGMENTATION BASED ON LEARNED TRANSFORMATIONS

Qiuli Wang, Zhihuan Li, Wei Chen, Xiaohong Zhang*, Hongqian Wang, Chen Liu†

*School of Big Data & Software Engineering, Chongqing University, Chongqing, 400044, China
†The First Affiliated Hospital of Army Medical University, 400032, Chongqing, China

ABSTRACT

Lung segmentation on CT (Computed Tomography) is a core step for many pieces of research on pulmonary diseases. Supervised methods have attained state-of-the-art accuracy; however, these methods heavily rely on supervised training with large labeled datasets. Labeling CT scans is a time-consuming task and different datasets acquired from different types of equipment may have various data characteristics. In this study, we present a framework for one-shot lung segmentation based on learned spatial and density transformations. Our framework requires only one segmented CT scan (source scan), and transform this scan using spatial and density transformations according to target scans. These target scans can be scans from another dataset or acquired from other types of equipment. Spatial and density transformations learns lung spatial deformation field and structural density using CNNs (Convolutional Neural Networks). Then we will train a supervised U-Net to segment lungs from CT scans using generated scans. As far as we know, we are the first to segment lungs using one-shot method. Our experiments demonstrate that our framework can achieve convincing results even if the source scan with label comes from another dataset.

Index Terms— One-Shot, Lungs, Segmentation, CT (Computed Tomography), U-Net

1. INTRODUCTION

Lung segmentation on CT (Computed Tomography) is a core task for many studies on pulmonary diseases. Supervised learning methods attain state-of-the-art accuracy, but these models have some common drawbacks.

First of all, these supervised methods heavily rely on labeled datasets. Labeling lungs on CT scans requires significant expertise and time, since one case of chest CT contains about 200 slices. As a result, public available lung CT datasets are usually small. For example, VESSEL12 [1] contains 20 cases in training set, Lung CT Segmentation Challenge 2017 [2] contains 60 cases. These datasets can not reflect the clinical diversity of lung structures. Second, CT scans collected from different types of equipment vary in basic settings, which can cause different noise or image quality. Since supervised models' parameters are optimized according to these settings, the performances of the supervised models trained on the partial dataset are not stable on another dataset.

The drawbacks mentioned above have become limitations of clinical application for the supervised lung segmentation model.

In this study, we propose a framework for one-shot lung segmentation. Our method needs only one labeled chest CT scan for VESSEL12, and generate new scans with segmentation map according to data without segmentation map.

The process of our framework is shown in Fig. 1. First, we use two U-Net models to learn the set of spatial and density transformations between the source scan and the target scans separately. Second, we apply spatial and density transformations to the source scan to synthesize new scans. Third, we apply spatial transformations to the segmentation map of the source scan to get a segmentation map for the generated new scans. Finally, we train a supervised U-Net with these generated labeled scans.

The remainder of the manuscript is organized as follows. Section 2 introduces some related work about lung segmentation and transformation networks. Section 3 describes the datasets we use in this study and introduces the architecture and details of our framework. Section 4 reports our experimental results on the public dataset and the clinical dataset. Our conclusions are summarized in section 5.

2. RELATED WORK

2.1. Lung Segmentation

Lung segmentation is a basic task for many studies on pulmonary diseases. Supervised methods attain state-of-the-art accuracy. Diverse works aimed to segment the lung region using different and combined techniques, that goes through region growing, border analysis, shape and probabilistic models, and recently deep learning approaches.

Traditional methods (i.e., region growing, border analysis

(1)Learning Progresses of Spatial and Density Transformations   (2)Transform the Source Scan and Segmentation Map to New Scans with Labels
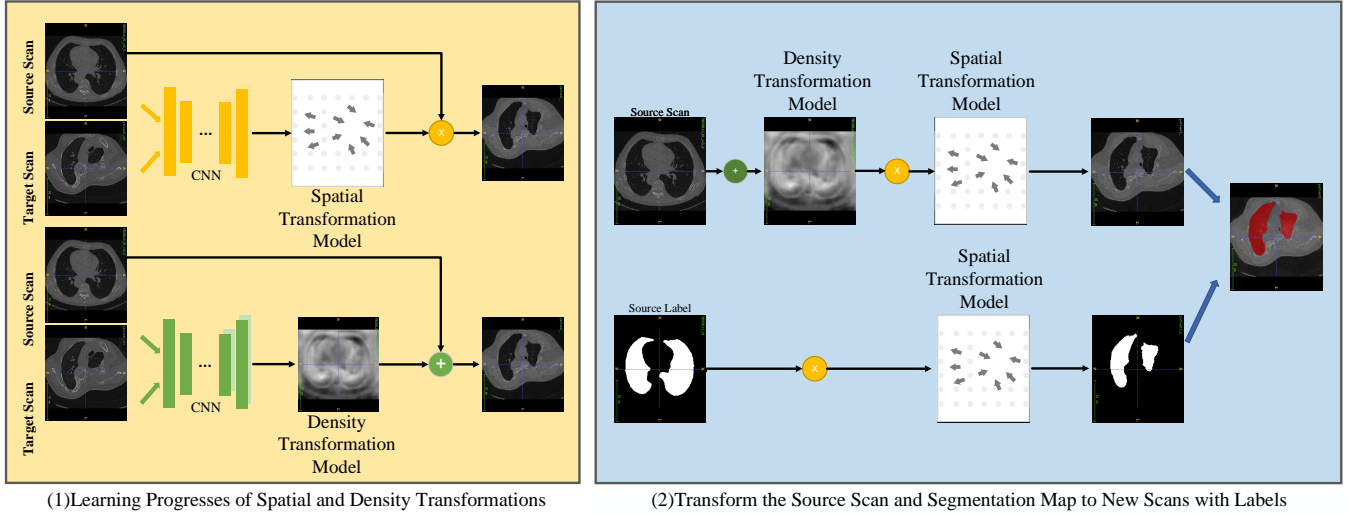
Fig. 1. The overview of our framework. (1) The learning processes of spatial and density transformations. We use two separate U-Net models to learn the transformations. (2) The process of transforming the source scan with segmentation maps to new labeled scans. Different from transforming a case, transforming a label only needs the spatial transformation.

and so on) is mature, but have some disadvantages in medical image segmentation. Methods like region growing [3] relied heavily on the selection of seed points, and the segmentation speed is inefficient. Active contour model [4] segmentation results are greatly influenced by the initial contour setting. Border analysis [5] performs worse when regions have no obvious boundary. These disadvantages limit the clinical application of these methods.

Deep learning methods, especially CNN models (Convolutional Neural Network) have been widely used to segment medical images, like U-Net [6], H-DenseUNet [7], 3D U-Net [8] and so on. Jeovane H. Alves et al. Study in [9] extract lungs from CT images using fully convolutional networks, which achieved 96.89 ± 3.14 in Dice scores on VESSEL12. Ahmed Soliman et al. [10] achieved 99.0 ± 0.5 in Dice scores on VESSEL12 using adaptive appearance-guided shape modeling. Study in [9] achieved 99.19 ± 0.47 in Dice sources, which is so far the state-of-art. Most deep learning methods heavily rely on labeled datasets. However, public available big-scale medical datasets are very rare, which also limits the clinical application of supervised deep learning models.

2.2. Spatial Transformer Networks

Spatial transformer networks [11] and its variations have been used in a variety of image analyses. Parametric spatial transform models have been used to align and classify handwritten digits [12–14]. In medical image registration, a spatial deformation model is used to establish semantic correspondences between images.

Guha Balakrishnan et al. [15] proposed VoxelMorph, a unsupervised learning-based method, to learn spatial trans-

formations. VoxelMorph used CNN based on U-Net to learn registration between brain MRI 3D volumes. Amy Zhao et al. [16] proposed a method based on spatial and appearance transform. This model used two U-Net models to implement data argumentation using learned transformations for one-shot medical image segmentation.

3. MATERIALS AND METHODS

3.1. Datasets

In this study, we use two datasets, which contain 110 cases. The first dataset is VESSEL12 [1], which was proposed for the evaluation of both semi and automatic methods for lungs' blood vessel segmentation in CT scans. A total of 20 scans were available for testing (plus three as examples). Scans have an average of 432 slices, and each scan contains segmentation maps for lungs. This dataset is available online[1]. The second dataset is a clinical private dataset, which is collected from the Radiology Department, The First Affiliated Hospital of Army Medical University. A total of 90 scans are included. For each scan, we only keep the series under the lung window, which is one of the most common CT windows in the clinic. Since it is a private dataset, which is collected directly from the clinic, we don't have segmentation maps for each scan.

3.2. Methods

We propose to improve one-shot lung CT segmentation by generating new training chest CT with segmentation maps in

---

[1]https://vessel12.grand-challenge.org/

a semi-supervised learning framework.

Following the study [16], we use two U-Net models to learn the spatial and density transformation models. We first have a source scan (atlas) from VESSEL12, which has segmentation maps. Second, we use two U-Net models to learn spatial and density transformations between source scan and target scans. Third, we apply spatial and density transformations to the source scan to synthesize new scans. Third, we apply spatial transformations to the segmentation map of the source scan to get segmentation maps for the generated new scans. Finally, we train a supervised U-Net with these generated labeled scans.

We describe the differences between scans using a combination of spatial and density transformations. Let $T(\cdot)$ denote a transformation from one CT scan to another as a combination of a spatial transformation $t_s(\cdot)$ and a density transformation $t_d(\cdot)$. Then we have $T(\cdot) = t_s(t_d(\cdot))$. We model our spatial and density transformations as follows:

$$t_s(x) = x \circ \mu, \qquad \mu = g\theta_s(x, y) \qquad (1)$$

$$t_d(x) = x + \rho, \quad \rho = h\theta_d(x, y \circ \mu^{-1}) \qquad (2)$$

We define the deformation function $\mu = id + u$, where id is the identity function, u is a smooth voxel-wise displacement field. We use $x \circ \mu$ to denote the application of the deformation μ to x. In our study, x is the source scan.

Then we model our appearance transform model loss as:

$$L(x, y^{(i)}, \mu^{(i)}, \mu^{-1(i)}, \rho^{(i)}, c_x) =$$
$$L_{sim}((x + \mu^{(i)}) \circ \mu^{(i)}, y^{(i)}) + \lambda L_{smooth}(c_x, \rho^{(i)})$$

$$L_{sim}((x + \mu^{(i)}) \circ \mu^{(i)}, y^{(i)}) =$$
$$\sigma_1 L_{sim}((x_{bone} + \mu^{(i)}) \circ \mu^{(i)}, y^{(i)}_{bone}) +$$
$$\sigma_2 L_{sim}((x_{lung} + \mu^{(i)}) \circ \mu^{(i)}, y^{(i)}_{lung})$$

where i indicates the i-th target scan. L_{smooth} is a smoothness regularization function based on the segmentation map of the source scan:

$$L_{smooth} = (c_x, \rho)\nabla\rho \qquad (3)$$

In [16], L_{sim} only calculated single loss of image similarity loss. However, their study was about brain segmentation. The density of the brain has a much smaller range than that of the lungs, so the single loss of image similarity cannot meet the requirements of lung segmentation. In our study, L_{sim} is a multi-density loss, which contains loss for both CT images and bone structure. We use bone structure to improve the transformation since bones like ribs are the natural boundary for lungs. λ, σ_1 and σ_2 are parameters, which can be optimized by back-propagation algorithm. Both density and spatial transformations are learned by separate U-Nets, as shown in Fig. 2.

In Fig. 3, we show two generated cases with its target cases. As shown in Fig. 3, using labeled sources scan from VESSEL12, we can generate new labeled scans by learning spatial and density transformations. In Fig. 3, we can see that the generated scans are very similar to target scans, and the generated segmentation maps can cover the lung areas well. We will release some generated cases on our website.

4. EXPERIMENTS

In this section, we have conducted several experiments to analyze our model in detail. In the first part of the experiments, we quantitatively analyze the performance of our framework. This part of the experiments is carried on with VESSEL12 since VESSEL12 provides segmentation maps for each scan. In the second part of the experiments, we verify the performance of our framework on clinical data. Due to the shortage of labeled data, we cannot quantitatively analyze segmentation results in this part, so that we invite 2 radiologists to give subjective judgments on segmentation results. All experiments are carried on with a GPU NVIDIA Tesla V100.

4.1. Results on VESSEL12

In this section, we experimentally quantitative analyze the performances of our framework on VESSEL12. Since our method requires only one segmented scan, and generate new scans which are different from VESSEL12, 20 scans are all used to verify the performances.

In this part of the experiments, we conduct two experiments. First, we pick one case and its segmentation maps as source case and generate 20 new cases as our training set, then we train a U-Net(Ours) with these cases. We test the performance of this U-Net(Ours) on VESSEL12. Second, we train a U-Net(One Scan) with only one segmented case (i.e., the source case) and test the performance of this U-Net(One Scan) in the same way.

Experiments' results are listed in Table. 1. We use Dice Score (DSC) [17] to evaluate the performances of segmentation results. According to Table. 1, supervised methods have achieved very convincing results. The state-of-art has achieved a Dice Score of 99.19. The U-Net trained with only labeled scan form VESSEL12 achieves a Dice Score of 91.69. The U-Net trained with generated scans achieves 96.31, 2.88 lower than the state-of-art. Considering our method requires only one segmented scan, our results are quite convincing and there is a lot of room for improvement if we have more labeled scans.

Moreover, we observe that some particular scans with severe diseases are difficult to segment. For example, our method only achieves a Dice Score of 91.44 on VESSEL12_17 in the testing set, which lower the average score. That would be very helpful if we can get more labeled scans that have severe diseases.
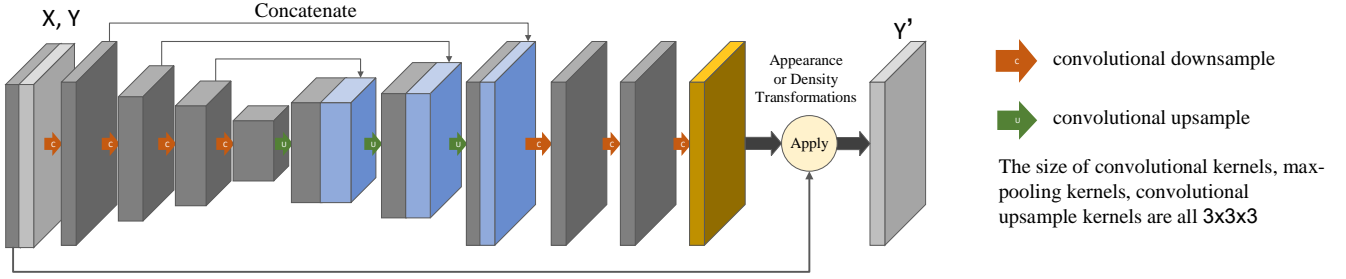
Fig. 2. The structure of U-Net. X and Y indicates source scans and target scans. Y' indicates the generated results. U-Net models for spatial and density transformations are exactly the same.
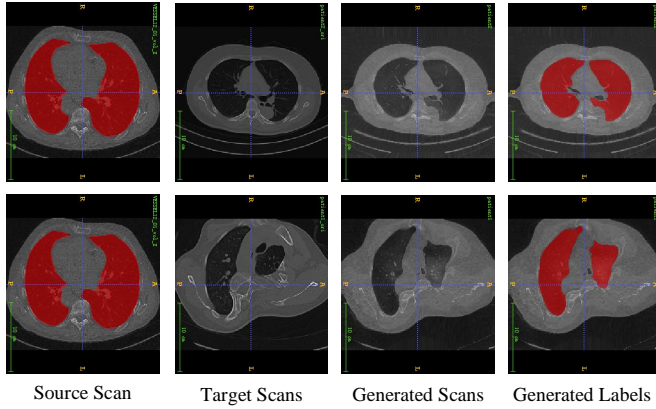

Source Scan   Target Scans   Generated Scans   Generated Labels

Fig. 3. Comparisons between the source scan, target scans, generated scans and generated labels. The generated scans are similar to target scans, but still keep some features from the source scan. The generated segmentation maps can cover the lung areas well.

Table 1. Comparison among Different Methods

Methods	Dataset	DSC
Soliman et al. [10]	VESSEL12	99.00
Alves et al. [9]	VESSEL12	99.19
U-Net(One Scan)	VESSEL12	91.69
U-Net(Ours)	VESSEL12	96.31

4.2. Results on Clinical Data

In this section, we analyze the performances of our model on clinical data. We got 90 cases of chest CT, 60 cases are used to generate new data with labels, 30 cases are used to verify the segmentation results. Due to the shortage of labels of the private dataset, we invite two radiologists to compare the segmentation results manually.

We conduct two experiments in this part. First, we train a U-Net(VESSEL12) with all VESSEL12 data, then test this model on the testing set (30 cases). Second, we use the source case with segmentation maps (the same as section 4.1) and generate 60 new cases with segmentation maps. We train a U-Net(ours) and test this model one the 30 testing cases.

Experiments demonstrate that our method can improve the performance of U-Net. Fig. 4 clearly shows that U-Net trained with our scans has a better performance. Our model can segment lungs better, especially these narrow areas (areas in rectangles). However, both models perform worse when the lungs have severe lung diseases (the last column).

According to two radiologists, in 30 testing cases, our framework has a better performance in 21 cases. Then both methods perform worse on 6 cases, and both segment lungs well on the left 3 cases. We further investigate 6 cases which are difficult to segment. We find that these 6 cases all have severe lung diseases, which lead to great changes in lung density or lung structure.

In a word, using scans generated can improve the performances of segmenting narrow lung areas and healthy cases. However, since we don't have enough labeled scans which have severe diseases, both models have difficulty in segmenting lungs with diseases.

5. DISCUSSION AND CONCLUSION

In this work, we generate labeled lung segmentation data using density and spatial transformations. Our method can help U-Net have a better performance using only one labeled data in the scenario of crossing datasets. Moreover, our method
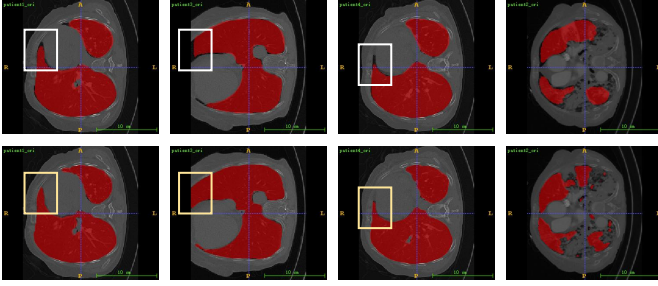
Fig. 4. The top row shows the segmentation results generated by a U-Net trained by VESSEL12. The bottom row shows the segmentation results generated by a U-Net trained by generated scans.

provides a solution when we don have enough labeled data.

However, we only use one labeled source scan in our study, it is very difficult to generate scans which can cover lungs with large changes, i.e., density changes, shape changes. This problem limits the effect of our method. Our future work will focus on adding diversify source scans and improve the quality of generated scans.

6. REFERENCES

[1] Rina D Rudyanto, Kerkstra, et al., "Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the vessel12 study," *Medical image analysis*, vol. 18, no. 7, pp. 1217–1232, 2014.

[2] JS Yang et al., "Data from lung ct segmentation challenge. the cancer imaging archive," 2017.

[3] Rolf Adams and Leanne Bischof, "Seeded region growing," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 16, no. 6, pp. 641–647, 1994.

[4] Michael Kass, Andrew Witkin, and Demetri Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[5] John Canny, "A computational approach to edge detection," in *Readings in computer vision*, pp. 184–203. Elsevier, 1987.

[6] 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.

[7] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng, "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[8] 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*. Springer, 2016, pp. 424–432.

[9] Jeovane H Alves, Pedro M Moreira Neto, and Lucas F Oliveira, "Extracting lungs from ct images using fully convolutional networks," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[10] Ahmed Soliman, Fahmi Khalifa, Ahmed Elnakib, Mohamed Abou El-Ghar, Neal Dunlap, Brian Wang, Georgy Gimel'farb, Robert Keynton, and Ayman El-Baz, "Accurate lungs segmentation on ct chest images by adaptive appearance-guided shape modeling," *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 263–276, 2016.

[11] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu, "Spatial transformer networks," *arXiv: Computer Vision and Pattern Recognition*, 2015.

[12] Søren Hauberg, Oren Freifeld, Anders Boesen Lindbo Larsen, John Fisher, and Lars Hansen, "Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation," in *Artificial Intelligence and Statistics*, 2016, pp. 342–350.

[13] Erik G Learned-Miller, "Data driven image models through continuous joint alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 236–250, 2005.

[14] Erik G Miller, Nicholas E Matsakis, and Paul A Viola, "Learning from one example through shared densities on transforms," vol. 1, pp. 464–471, 2000.

[15] Guha Balakrishnan, Amy Zhao, Mert Sabuncu, John Guttag, and **Adrian V. Dalca***, "Voxelmorph: A learning framework for deformable medical image registration," *IEEE TMI: Transactions on Medical Imaging*, 2019.

[16] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V Guttag, and Adrian V Dalca, "Data augmentation using learned transformations for one-shot medical image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8543–8553.

[17] Lee R Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.