# ONE-SHOT LUNG SEGMENTATION BASED ON LEARNED TRANSFORMATIONS IN A CROSS-DATASET SCENARIO

*Qiuli Wang, Zhihuan Li, Wei Chen, Kun Wang, Jiuquan Zhang*[⋆]*, Chen Liu*[†]*, Xiaohong Zhang*[‡*]

[‡]School of Big Data & Software Engineering, Chongqing University, Chongqing, 400044, China
[†]The First Affiliated Hospital of Army Medical University, Chongqing, 400032, China
[⋆]Chongqing University Cancer Hospital & Chongqing Cancer Institute, Chongqing, 400030, China

## ABSTRACT

Lung segmentation on CT is a core step for many studies on pulmonary diseases. Supervised methods have attained state-of-the-art accuracy. However, these methods heavily rely on supervised training with large labeled datasets. In this study, we present a framework for one-shot lung segmentation based on learned spatial and density transformations in a cross-dataset scenario. Our framework requires only one labeled CT (Computed Tomography) scan: source scan, and the source scan is transformed according to target scans using spatial and density transformations, which learn the lung spatial deformation field and structural density based on CNNs. The target scans are not limited to the same dataset as the source scan, but also come from another dataset or another types of CT equipment. And, using the spatial transformation, our framework can generate labels for each transformed scans. Finally, we will train a supervised U-Net with these generated labeled scans. Moreover, a multi-density MSE loss is proposed to improve transformations since the ribs are natural boundaries for the lungs. As far as we know, our study is the first to segment lungs using the one-shot method. Experiments demonstrate that our framework can achieve convincing results on VESSEL12 and clinical chest CT.

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

## 1. INTRODUCTION

Lung segmentation on CT (Computed Tomography) is a core step for many studies on pulmonary diseases. Supervised learning methods have attained 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 over 200 slices [1]. As a result, public available lung CT datasets, like [1, 2] are usually small. So that 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 lead to different noise or image quality. As a result, the performances of the supervised models trained on the particular dataset are not stable on another dataset. The drawbacks mentioned above have become limitations of clinical application for the supervised lung segmentation models.

In this study, we propose a framework for one-shot lung segmentation. Our method needs only one labeled chest CT scan from VESSEL12 and generates new scans with segmentation maps according to the target scans, even if these target scans come from another dataset. As shown in Fig 1, we first have a source scan with segmentation maps from VESSEL12. Second, two U-Net models are adopted to learn spatial and density transformations between source scan and target scans. Third, spatial and density transformations are applied to the source scan to synthesize new scans. Then, we apply spatial transformations to the segmentation maps of the source scan to generate labels for the new scans. Finally, we train a supervised U-Net with these generated labeled scans.

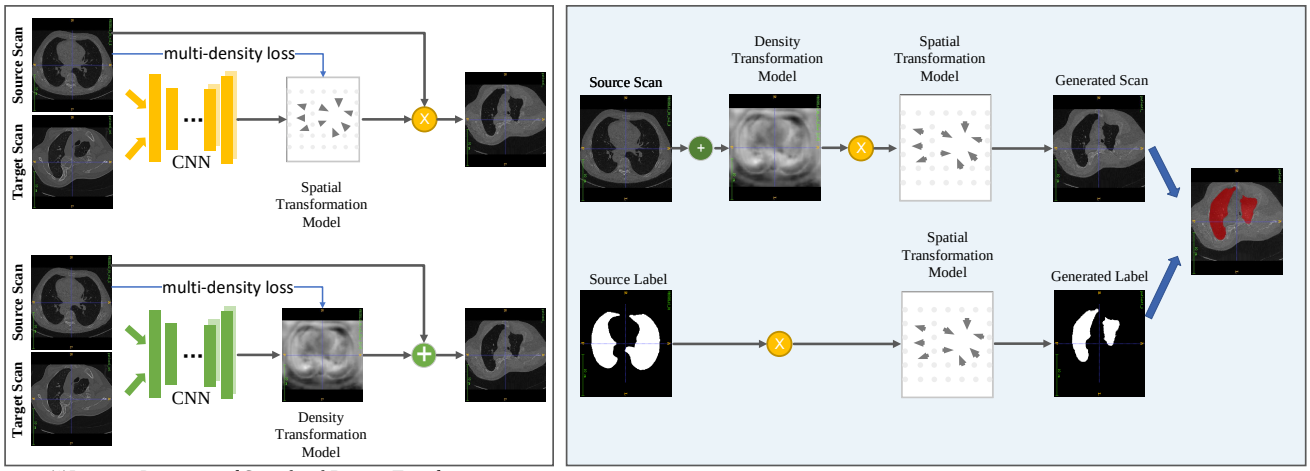
The major contributions of this paper can be summarized as follows: (1)We propose a framework for one-shot lung segmentation based on learned transformations. The whole framework needs only one labeled example; (2)We design a multi-density MSE loss to improve the transformations, so that we can take advantages of natural boundaries of lungs, i.e., ribs; (3)We evaluate our framework on VESSEL12 and clinical CT scans. Our framework achieves a Dice Score of 96.31 on VESSEL12 and shows good generalization ability in a cross-dataset scenario.

## 2. RELATED WORKS

### 2.1. Lung Segmentation

Lung segmentation is a basic task for many studies on pulmonary diseases. Diverse studies have been proposed to segment the lung region using different and combined techniques, like region-growing, border analysis, shape and prob-

*This work was partially supported by the National Natural Science Foundation of China (Grant No. 61772093), the Chongqing Major Theme Projects (Grant No. cstc2018jszx-cyztzxX0017).




(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. Two U-Net models are adopted 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 spatial transformation.

abilistic models, and recently deep learning approaches.

Traditional methods (i.e., region-growing, border analysis and so on) are mature but have some disadvantages in medical image segmentation [3–5]. Although, a study in [6] achieved $99.00 \pm 0.5$ in Dice Score using adaptive appearance-guided shape modeling.

Deep learning methods, especially CNN models (Convolutional Neural Network) have been widely adopted to segment medical images, like U-Net [7], H-DenseUNet [8], 3D U-Net [9] and so on. Study in [10] extracted lungs from CT images using fully convolutional networks. Their method achieved $99.19 \pm 0.47$ in Dice Scores on VESSEL12, which is the state-of-the-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, such as handwritten digits classification [12, 13], medical image registration, or data argumentation. Guha Balakrishnan et al. [14] proposed VoxelMorph, an unsupervised learning-based method, to learn spatial transformations and finish the task of registration between brain MRI 3D volumes. Amy Zhao et al. [15] proposed a method based on spatial and appearance transform. This model adopted two U-Net models to implement data argumentation using learned transformations for one-shot brain MRI image segmentation.

## 3. MATERIALS AND METHODS

### 3.1. Datasets

In this study, we use two datasets, which contain a total of 110 cases. The first dataset is VESSEL12 [1], which was proposed for the evaluation of both semi and automatic methods for lung blood vessel segmentation in CT scans. A total of 20 scans are available for training or testing. 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. Since it is a private dataset collected directly from the clinic, we do not have segmentation maps for each scan. The data collected had been reviewed and desensitized.

### 3.2. Methods

We propose to improve one-shot lung CT segmentation by generating new training chest CT with segmentation maps in a semi-supervised learning framework. The key point of this framework is the learning of spatial and density transformations.

Following the study [15], 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 $T(\cdot) = t_s(t_d(\cdot))$. Our framework formulates spatial and density transformations as follows:

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

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

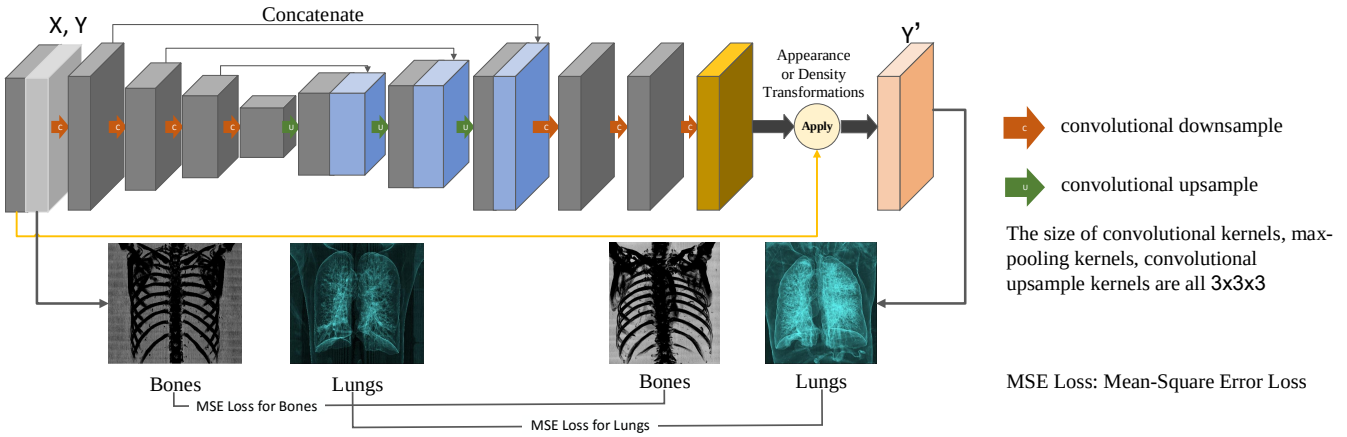



**Fig. 2**. The structure of U-Net. X and Y indicate source scans and target scans. Y' indicates the generated results. U-Net models for spatial and density transformations are the same. The loss function contains two parts: MSE for bones, and MSE for lungs. Bones like ribs are natural boundaries for lungs, which can help to improve the transformations.

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

We define the deformation function $\mu = id + u$, where id is the identity function, and $u$ is a smooth voxel-wise displacement field. $x \circ \mu$ denotes the application of the deformation $\mu$ to $x$, and $x + \rho$ denotes the application of the deformation $\rho$ to $x$. In our study, $x$ is the source scan. $g\theta_s$ and $h\theta_d$ are trainable parameters for two transformation models. $\mu^{-1}$ indicates the reverse deformation.

The parameters of framework are optimized as follows:

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

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

where $i$ indicates the $i$-th target scan. $L_{smooth}$, which is calculated as $(c_x, \rho)\nabla\rho$, is a smoothness regularization function based on the segmentation map of the source scan. $L_{sim}$ is mean-squared error for the image similarity loss. In [15], $L_{sim}$ only calculated single loss of image similarity loss. However, their study was about brain segmentation. The spatial structures of the brain have a much smaller range than that of the lungs, so the single loss of image similarity cannot meet the requirements of lung transformations. In our study, $L_{sim}$ is a multi-density loss, which contains losses for both CT images and bone structures. We use bone structures to improve the transformation since bones like ribs are the natural boundaries for lungs. Parameters like $\lambda$, $\sigma_1$ and $\sigma_2$, can be optimized by back-propagation algorithm. Both density and spatial transformations are learned by separate U-Nets, as can be seen in Fig 2.

We show in Fig 3 two generated cases with its target cases. It can be seen that the generated scans are very similar to target scans, and the generated segmentation maps can cover the lung regions well.

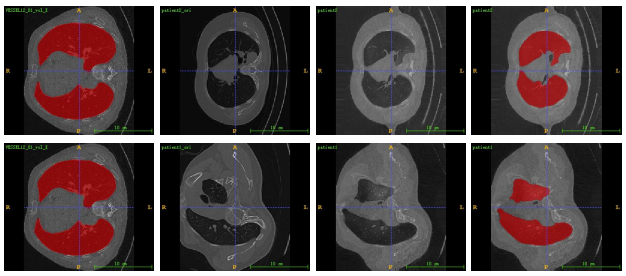



**Fig. 3**. Comparisons between the source scan, target scans, generated scans and generated labels. The generated scans are similar to target scans, but still keep some features from the source scan. The generated segmentation maps can cover the lung regions well.

## 4. EXPERIMENTS

In this section, we have conducted several experiments to analyze our framework in detail. In the first subsection, we quantitatively analyzed the performance of our framework on with VESSEL12. Since VESSEL12 provided segmentation maps for each scan, we were able to calculate Dice Scores [16]. In the second subsection, we verified the performance of our framework on clinical data. Due to the shortage of labeled data, we could not quantitatively analyze segmentation results, so that we invited 2 radiologists to give subjective judgments on segmentation results.

The source scan was a scan from VESSEL12, and the 60 target scans were collected from the clinic. There are not

many requirements for the selection of the source scan, except that the selected source scan needs a clear view of lungs and healthy bone structures.

### 4.1. Results on VESSEL12

In this section, we experimentally quantitative analyzed the performances of our framework on VESSEL12 in the scenario of cross-dataset. Due to the shortage of labeled data, this study only used VESSEL12 for evaluation, except the source scan. The target scans were collected from the clinic.

We conducted two experiments. In the first experiment, we picked one case with its segmentation maps as the source case and generated 60 new cases using 60 clinical cases as our training set. Then we trained a U-Net with these generated cases and tested the performance of this U-Net on VESSEL12. In the second experiment, we trained a U-Net with only one segmented case (i.e., the source case) and tested the performance of this U-Net in the same way.

Experiments' results are listed in Table 1. As can be seen from Table 1, supervised methods have achieved very convincing results. The state-of-the-art has achieved a Dice Score of 99.19. The U-Net trained with one labeled scan form VESSEL12 achieves an average Dice Score of 91.69. The U-Net trained with generated scans achieves 96.31, which is 2.88 lower than the state-of-the-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 can improve the quality of generated scans.

**Table 1**. Comparison among Different Methods

| ***Methods*** | ***Training Data*** | ***DSC*** |
|---|---|---|
| Soliman et al. [6] | VESSEL12 | 99.00 |
| Alves et al. [10] | VESSEL12 | 99.19 |
| U-Net | One Scan from VESSEL12 | 91.69 |
| U-Net | Generated Scans | 96.31 |

Moreover, we observed that some particular scans with severe diseases are difficult to segment. For example, our method only achieved a Dice Score of 91.44 on the 17-th scan of VESSEL12, which lowered the average score. It was because the 17-th scan of VESSEL12 had lung diseases and caused density changes in the lungs. That would be very helpful if we can get more labeled scans that have great changes of densities or structures.

### 4.2. Results on Clinical Data

In this section, we analyzed the performances of our framework on clinical data. The dataset contained 90 cases of chest CT, 60 cases were used to generate new data with labels, 30 cases were used to verify the segmentation results. Due to the shortage of labels of the private dataset, we invited two radiologists to give subjective judgments on segmentation results.

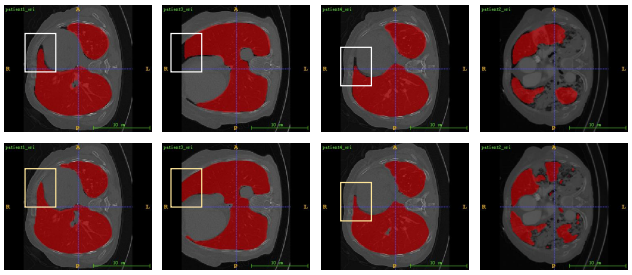


**Fig. 4**. The top row shows the segmentation results generated by a U-Net trained by VESSEL12. The bottom row shows the segmentation results generated by a U-Net trained by generated scans.

We conducted two experiments in this section. First, we trained a U-Net with all VESSEL12 data, then tested this model on the testing set. Second, we generated 60 new cases with segmentation maps. Then we trained a U-Net with generated data and test this model on the 30 testing cases. We chose the same source scan as section 4.1.

The experiment results demonstrate that our method improves the performance of U-Net. Fig 4 clearly shows that U-Net trained with our scans has better performance, especially in these narrow regions (areas in blocks).

However, both models performed worse when the lungs had severe lung diseases (the last column). According to two radiologists, in 30 testing cases, our framework had a better performance in 21 cases. Then both methods performed worse on 6 cases, and both segmented lungs well on the left 3 cases. We further investigated 6 cases which were difficult to segment and found that these 6 cases all had severe lung diseases, which led to great changes in lung densities or lung structures.

In a word, using generated scans can improve the performances of segmenting narrow lung regions and healthy cases on clinical data. Since we don't have enough labeled scans which contain great changes in densities and structures, both models have difficulty in segmenting lungs with severe diseases.

## 5. CONCLUSION

In this work, we proposed a framework for one-shot lung segmentation based on density and spatial transformations in a cross-dataset scenario. Our framework allows the supervised model to have a better performance using only one labeled case. Moreover, our framework provides a solution in a cross-dataset scenario.

However, we only used one labeled source scan in our study, and it was very difficult to generate scans that can cover the lungs with large changes like density changes or shape changes. This problem limited the effect of our framework. Our future work will focus on using diverse source scans and improving 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] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[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] 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.

[15] 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.

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