Deep Learning Based Lung Region Segmentation with Data Preprocessing by Generative Adversarial Nets

AUTHOR(S):

Nitta, Jumpei; Nakao, Megumi; Imanishi, Keiho; Matsuda, Tetsuya

CITATION:

Nitta, Jumpei ...[et al]. Deep Learning Based Lung Region Segmentation with Data Preprocessing by Generative Adversarial Nets. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) 2020: 1278-1281: 9176214.

ISSUE DATE:

2020

URL:

http://hdl.handle.net/2433/265387

Deep Learning Based Lung Region Segmentation with Data Preprocessing by Generative Adversarial Nets

Jumpei Nitta[1], *Student Member, IEEE*, Megumi Nakao[1], *Member, IEEE*, Keiho Imanishi[2], and Tetsuya Matsuda[1], *Member, IEEE*

***Abstract*— In endoscopic surgery, it is necessary to understand the three-dimensional structure of the target region to improve safety. For organs that do not deform much during surgery, preoperative computed tomography (CT) images can be used to understand their three-dimensional structure, however, deformation estimation is necessary for organs that deform substantially. Even though the intraoperative deformation estimation of organs has been widely studied, two-dimensional organ region segmentations from camera images are necessary to perform this estimation. In this paper, we propose a region segmentation method using U-net for the lung, which is an organ that deforms substantially during surgery. Because the accuracy of the results for smoker lungs is lower than that for non-smoker lungs, we improved the accuracy by translating the texture of the lung surface using a CycleGAN.**

I. INTRODUCTION

Recently, endoscopic surgery has become widely performed thanks to the development of medical technology and instruments. Endoscopic surgery is less painful than conventional laparotomy and the patient recovers more quickly. The surgical wound is also smaller than in open surgery, but the operation is difficult and time-consuming. In endoscopic surgery, because a camera is inserted into the body and the operation is performed through a monitor, it is necessary to understand the three-dimensional structure of the target organ in order to improve safety. However, it is not easy to capture the three-dimensional structure of an organ from the narrow surgical field of the endoscope, as several studies have reported [1][2].

For example, the liver is an organ that does not deform much during surgery. One proposed method supports liver surgery by projecting a virtual image of organ blood vessels and tumors created from preoperative computed tomography (CT) images on the laparoscopic image as a guide [3]. In contrast, in organs with large intraoperative deformations, there is a great difference between the organ shape obtained by preoperative CT images and the organ shape during operation. Thus it is not possible to simply project a virtual image onto a laparoscopic image, as can be done for the liver. In the lung in particular, because the lung parenchyma collapses in response to changes in intrapulmonary pressure during the operation, the amount of deformation is large and can exceed 50% [4]. Therefore, it is necessary to estimate the deformation to identify the tumor position. Studies have found that to estimate the deformation, it is possible to use the shape of the organ in the intraoperative endoscopic image [5], but it is necessary to segment the lung region from the endoscopic image for clinical application. Wu et al. [6] proposed a lung region segmentation method that focuses on the continuity of organ deformation in surgical videos.

In this paper, we propose a lung region segmentation method using U-net, which is a model that was proposed for medical image segmentation [7]. Accuracy varies in lung region segmentation using U-net, and in particular, the results of smoker lung images are less accurate than those of non-smoker lung images, where smoker lung images stand for the images include purple color lung and thoracic cavity. Therefore, to improve the accuracy of lung region segmentation for smoker lung images, we propose a data preprocessing method that performs image style transfer using a CycleGAN-based model [8]. Many studies on image style transfer using deep learning have been reported [9][10]. Among them, CycleGAN is a method that changes only the texture of a local object while maintaining the global features of the image. CycleGAN is an extension of generative adversarial networks (GANs) [11]. It is an unsupervised learning method that learns the features of two image domains and translates image styles belonging to one domain to styles belonging to the other domain. As the data, we prepared frame images extracted from surgical videos of thoracoscopic lung cancer resection. We extracted scenes that do not include any surgical tools; examples are shown in Figure 1. In the experiment, we trained both U-net and CycleGAN using these data. For the U-net training, we also prepared mask images that represent the correct lung region. For the CycleGAN training, we divided the lung images into two domains, non-smoker lung images and smoker lung images, as shown in Figure 1, and used them as the training data.

II. METHODS

A. U-net Lung Region Segmentation Model

We constructed a lung region segmentation model using the U-net structure. The input of the model is a target frame image in which the lung region needs to be predicted, and the output is the lung region in the target frame. The size of the input frame image to be predicted is 256×256, and the number of channels is three (i.e., RGB channels). The output lung region is a single-channel binary image also 256×256 in size. In the output image, the white area represents the lung region and the black area represents the background.

[1]J. Nitta, M. Nakao, and T. Matsuda are with Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, JAPAN. (email: njum@sys.i.kyoto-u.ac.jp, megumi@i.kyoto-u.ac.jp, tetsu@i.kyoto-u.ac.jp)
[2]K. Imanishi is with e-Growth Co., Ltd., 403, Shimo-Maruya-cho, Nakagyo-ku, Kyoto 604-8006, JAPAN.

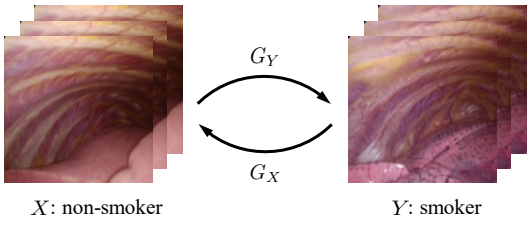

Fig. 1. Generative adversarial nets for lung images, X: non-smoker lung image domain and Y: smoker lung image domain. The dataset for training the adversarial nets was constructed from 1,890 images in total (X: 770 images, Y: 1,120 images).

B. Lung Image Targeted Adversarial Nets

In a CycleGAN, a network is constructed using two GANs. During model training, each GAN tries to generate images belonging to one image domain from images belong to the other image domain. Simultaneously, the image generated by one generator is input to the other generator to generate a reconstructed image that retains the features of the real image. CycleGAN uses the sum of adversarial loss, which is the original loss function of GAN, and cycle consistency loss as the loss function. Let the two image domains be X and Y. Then, the adversarial loss is

$$\begin{aligned}\mathcal{L}_{\text{adv}}(G_X, D_X) = \mathbb{E}_{\boldsymbol{x}}[\log D_X(x)] \\ + \mathbb{E}_{\boldsymbol{y}}[\log(1 - D_X(G_X(y)))],\end{aligned} \tag{1}$$

and the cycle consistency loss is

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G_X, G_Y) = \mathbb{E}_{\boldsymbol{x}}[\|G_X(G_Y(x)) - x\|_1] \\ + \mathbb{E}_{\boldsymbol{y}}[\|G_Y(G_X(y)) - y\|_1],\end{aligned} \tag{2}$$

where G_X tries to generate image $G_X(y)$ that is similar to image x in X from image y in Y, whereas D_X tries to discriminate whether the image is the real image x or generated image $G_X(y)$. The same is true for G_Y and D_Y. By adding the cycle consistency loss to the loss function, the model is trained to generate a reconstructed image that corresponds to the real image. Therefore, we can request the model to generate an image that retains the real image features. To summarize the loss function of CycleGAN, it becomes

$$\begin{aligned}\mathcal{L}_{\text{cgan}} = \mathcal{L}_{\text{adv}}(G_X, D_X) + \mathcal{L}_{\text{adv}}(G_Y, D_Y) \\ + \lambda_{\text{cyc}} \times \mathcal{L}_{\text{cyc}}(G_X, G_Y),\end{aligned} \tag{3}$$

where weight λ_{cyc} controls the relative strength of the adversarial loss and cycle consistency loss.

When segmenting lung regions on translated images, the correct region cannot be obtained if the shape of the lung is deformed. Therefore, when transferring the style of the lung image, we want to translate only the texture of the lung surface from the original lung image to the generated image. However, the original loss function of CycleGAN has insufficient constraints to perform such a translation while maintaining the lung shape. To solve this problem, we modify the feature loss introduced by Nakao et al. [12] and add it to the CycleGAN loss function (3).

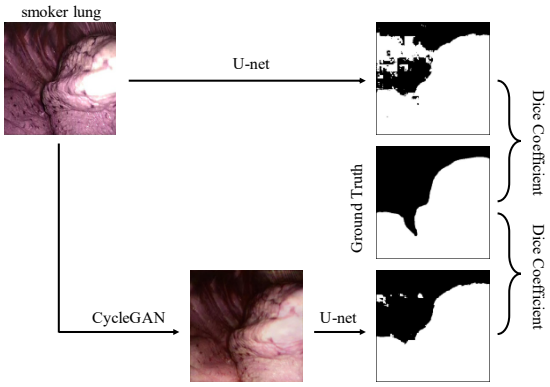

Fig. 2. Experiment architecture for smoker lung image region segmentation with data preprocessing. The U-net performs the lung region segmentation, CycleGAN performs the style transfer, and the ground truth represents correct lung region of the target smoker lung image. The Dice coefficient of the lung region images generated from original smoker lung images and translated smoker lung images were calculated and then compared.

We use ResNet [13] as a feature extractor and extract the features of images x, y, $G_X(y)$, $G_Y(x)$, $G_X(G_Y(x))$, and $G_Y(G_X(y))$. The feature loss is the sum of the differences in the extracted features of each image. Let f be a function that extracts image features using ResNet. Then, the feature loss becomes

$$\begin{aligned}\mathcal{L}_{\text{feat}} = \mathbb{E}_{\boldsymbol{x}}[\|f(G_Y(x)) - f(x)\|_2] \\ + \mathbb{E}_{\boldsymbol{y}}[\|f(G_X(y)) - f(y)\|_2] \\ + \mathbb{E}_{\boldsymbol{x}}[\|f(x) - f(G_X(G_Y(x)))\|_2] \\ + \mathbb{E}_{\boldsymbol{y}}[\|f(y) - f(G_Y(G_X(y)))\|_2] \\ + \mathbb{E}_{\boldsymbol{x}}[\|f(G_X(G_Y(x))) - f(G_Y(x))\|_2] \\ + \mathbb{E}_{\boldsymbol{y}}[\|f(G_Y(G_X(y))) - f(G_X(y))\|_2].\end{aligned} \tag{4}$$

By adding the feature loss to the loss function, the model trained to generate images retains the real image features.

Finally, we use loss function

$$\mathcal{L} = \mathcal{L}_{\text{cgan}} + \lambda_{\text{feat}} \times \mathcal{L}_{\text{feat}}, \tag{5}$$

to train the CycleGAN model, where λ_{feat} is the weight that controls the relative strength of each term. The smoker lung image translation model G_X^* is obtained by solving

$$G_X^*, G_Y^* = \arg \min_{G_X, G_Y} \max_{D_X, D_Y} \mathcal{L}(G_X, G_Y, D_X, D_Y). \tag{6}$$

III. EXPERIMENTS AND RESULTS

A. Lung Dataset and Experiment Architecture

In this paper, we prepared a dataset consisting of 25 sets of images to train the models. These image sets include 20 to 50 frame images of a scene without surgical tools extracted from each of the 25 surgical videos of thoracoscopic lung cancer resection provided by the Department of Thoracic Surgery, Kyoto University. These 25 sets of images include 12 non-smoker lung image sets and 13 smoker lung image sets. Furthermore, we augmented the dataset by performing parallel movement, affine transformation, zooming, and contrast changes on the extracted frame images. In addition,

TABLE I
LUNG REGION SEGMENTATION RESULTS FOR NON-SMOKER, ORIGINAL SMOKER, AND TRANSLATED SMOKER LUNG IMAGES: DICE COEFFICIENT VALUES FOR EACH CASE

Non-smoker			Smoker	
			Original	Translated
Case 1.1	0.9832	Case 4.1	0.9095	0.9738
Case 1.2	0.9937	Case 4.2	0.7515	0.8402
Case 2.1	0.9911	Case 5.1	0.9164	0.8824
Case 2.2	0.9550	Case 5.2	0.9083	0.8470
Case 3.1	0.9741	Case 6.1	0.9800	0.9823
Case 3.2	0.9785	Case 6.2	0.9354	0.9054
Mean	0.9793	Mean	0.9002	0.9052

we manually created images that represent the lung region corresponding to each frame image as the ground truth for lung region segmentation. We used all 25 sets for the U-net model training and used two sets for the CycleGAN model training. To test the models, we used 12 images from six surgical videos (i.e., two images from each video), where five videos were not included in the training dataset and one video was the same video as CycleGAN training dataset video while two images were from different scenes. The 12 images consisted of six non-smoker lung images and six smoker lung images.

Using these data, we performed the experiment shown in Figure 2. First, we trained the lung region segmentation model (U-net) and generated the lung region image of the original smoker lung images. We then evaluated the Dice coefficient of the result using the correct lung region (ground truth). Next, we trained the image style transfer model (CycleGAN) to translate the style of smoker lung images to non-smoker lung images. The lung region image was generated from this translated smoker lung image using the lung region segmentation model and its Dice coefficient was calculated. Finally, we compared the Dice coefficient of these two lung region images to verify the effectiveness of the data preprocessing using CycleGAN.

B. Lung Region Segmentation

We trained the model introduced in Section II-A using the dataset introduced in Section III-A. For the model training and evaluation, we used the Dice coefficient of the output image of the model and correct lung region image. We evaluated this model using 12 lung images. The second and fourth columns in Table I show the results.

These results reveal that the accuracy of lung region segmentation for smoker lung images is inferior to that of non-smoker lung images. However, in case 6.1, the value of the Dice coefficient is higher than some non-smoker lung image values. This is considered to be due to the fact that some smoker lungs have clear contour lines and can be easily distinguished from the thoracic cavity.

C. Style Transfer by Adversarial Nets

We trained CycleGAN using 770 non-smoker lung images as X and 1,120 smoker lung images as Y. Each of these image sets were extracted from one surgical video from the dataset described in Section III-A. As the generator G_X and G_Y, we used U-net. The weights in loss function $\mathcal{L}$ are $\lambda_{\text{cyc}} = 10.0$, $\lambda_{\text{feat}} = 5.0$.

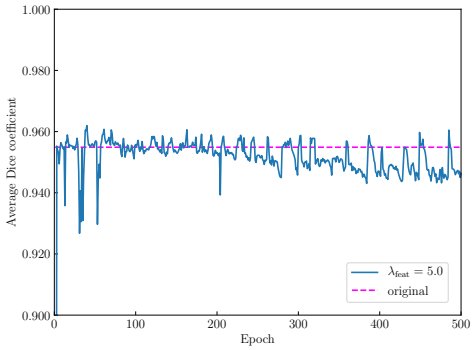

Fig. 3. Average Dice coefficient of the results for each epoch. The blue line represents the translated smoker lung image Dice coefficient and the pink dashed line represents the original smoker lung image Dice coefficient.

To find the model with the best accuracy of lung region segmentation, we performed a style transfer to the smoker lung images using the all models in each epoch. Then, we segmented the lung regions of the translated images and evaluated their Dice coefficient. We trained the model for 500 epochs and we used the training data for evaluation. Figure 3 shows the result. In this figure, the Dice coefficient of the translated image is higher than that of the original image up to about epoch 200. Since the Dice coefficient decreases smoothly as the number of epochs increases, it is considered that the models trained for 100 to 200 epochs are the models with the highest accuracy. Therefore, in later experiments, we used a 163-epoch model, which had the highest Dice coefficient of these models.

Figure 4 shows the results of translated smoker images using the 163-epoch model and lung region images generated by the lung region segmentation model. Columns 4 and 5 of Table I compare the lung region segmentation results for the original and translated smoker lung images. As shown in the table, in three cases, the Dice coefficient value of the translated images decreased. In these cases, the CycleGAN model failed to translate the images properly. However, in case 4.1, 4.2 and 6.1, the Dice coefficient value increased, where case 6.1 was the image from the CycleGAN training dataset video. In these cases, the translation of the image succeeded and lung region segmentation accuracy increased. The reason why translation succeeded in case 4.1 and 4.2 is that the lung color of these case images resembled the lung color of the image in the training data of the model. In contrast, we consider that the translation of cases 5.1 and 5.2 failed because the images differ from the images in the training data. In these failed cases, the contour line of the lung images was blurred by the translation, and this is the reason why the Dice coefficient value decreased. For case 6.2, translation failed since the appearance of the lung differs from the lung in training data.

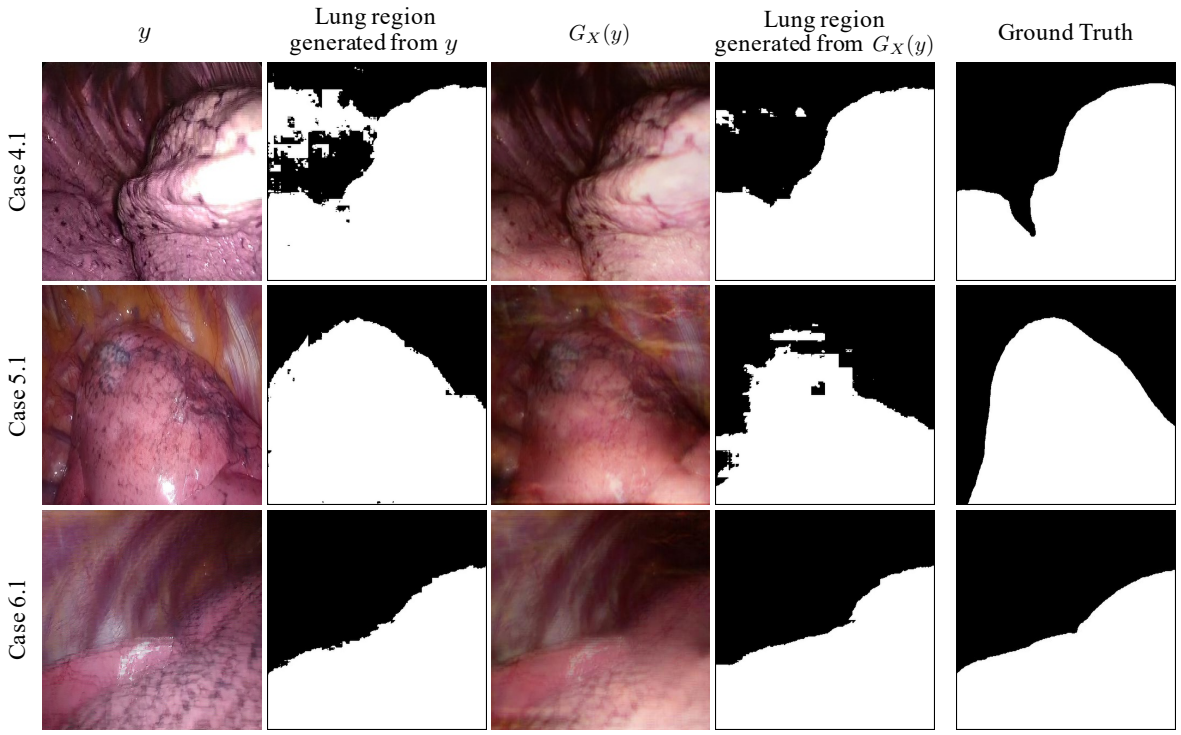

Fig. 4. Translated smoker lung images using the 163-epoch model and generated lung region images using lung region segmentation model. The first and second columns show the original lung images and generated lung region images, respectively, the third and fourth columns show the translated lung images and generated lung region images, respectively, and the fifth column shows the correct lung region images.

IV. CONCLUSIONS

This paper introduced a lung region segmentation method using deep learning and a data preprocessing method using adversarial nets. The results of the lung region segmentation reveal that the proposed method could obtain lung regions with high accuracy. Because the accuracy of the smoker lung images was low, we performed a data preprocessing method that attempted to translate the texture of a smoker lung image to a non-smoker lung image. Using this preprocessing method, we obtained a highly accurate lung region segmentation in one test case. Future challenges include improving the lung region segmentation and the adversarial training to translate image styles accurately in more general cases.

ACKNOWLEDGMENT

We thank Kimberly Moravec, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

REFERENCES

[1] M. Sato, M. Omasa, F. Chen, T. Sato, M. Sonobe, T. Bando, and H. Date, "Use of virtual assisted lung mapping (val-map), a bronchoscopic multispot dye-marking technique using virtual images, for precise navigation of thoracoscopic sublobar lung resection," *The Journal of thoracic and cardiovascular surgery*, vol. 147, no. 6, pp. 1813–1819, 2014.

[2] B. Koo, E. Özgür, B. Le Roy, E. Buc, and A. Bartoli, "Deformable registration of a preoperative 3D liver volume to a laparoscopy image using contour and shading cues," International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017, pp. 326–334.

[3] N. Haouchine, S. Cotin, I. Peterlik, J. Dequidt, M. S. Lopez, E. Kerrien, and M.-O. Berger, "Impact of soft tissue heterogeneity on augmented reality for liver surgery," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 5, pp. 584–597, 2014.

[4] M. Nakao, J. Tokuno, T. Chen-Yoshikawa, H. Date, and T. Matsuda, "Surface deformation analysis of collapsed lungs using model-based shape matching," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–12, 2019.

[5] M. Nakao, A. Saito, and T. Matsuda, "A simulation study on deformation estimation of elastic materials using monocular images," *Int. J. Computer Assisted Radiology and Surgery*, 12(1), S257–258, 2017.

[6] S. Wu, M. Nakao, and T. Matsuda, "Continuous lung region segmentation from endoscopic images for intra-operative navigation," *Computers in biology and medicine*, vol. 87, pp. 200–210, 2017.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234–241.

[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[9] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2414–2423.

[10] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," European conference on computer vision, 2016, pp. 694–711.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Advances in neural information processing systems, 2014, pp. 2672–2680.

[12] M. Nakao, K. Imanishi, N. Ueda, Y. Imai, T. Kirita, and T. Matsuda, "Three-dimensional generative adversarial nets for unsupervised metal artifact reduction," *arXiv preprint arXiv:1911.08105*, 2019.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.