LGAN: Lung Segmentation in CT Scans Using Generative Adversarial Network

Jiaxing Tan*,1, Longlong Jing*,1, Yumei Huo[1,2], Oguz Akin♮,4, Yingli Tian†,1,3
[1]The Graduate Center, The City University of New York, New York, NY, 10016
[2]The College of Staten Island, The City University of New York, New York, NY 10314
[3]The City College, The City University of New York, New York, NY 10031
[4]Memorial Sloan-Kettering Cancer Center, New York, NY, 10065

***Abstract*—Lung segmentation in computerized tomography (CT) images serves as an important procedure in various lung disease diagnosis, where current approaches are mostly designed with a series of manually empirical parameter adjustment involved steps. To achieve a fully-automatic end-to-end segmentation method, in this paper, we propose a novel deep learning Generative Adversarial Network (GAN) based lung segmentation schema, which denoted as LGAN. Based on the evaluation results on a dataset containing 220 individual CT scans, our proposed LGAN outperforms the current state-of-the-art methods by two metrics: segmentation quality and shape similarity. The results obtained with this study demonstrate that the proposed LGAN schema can be used as a promising tool for automatic lung segmentation due to its simplified procedure as well as its good performance.**

***Index Terms*—Deep Learning, Lung Segmentation, Generative Adversarial Network, Medical Imaging, CT Scan**

I. Introduction

Lung segmentation is an initial step in analyzing medical images to assess lung disease. Researchers proposed a number of lung segmentation methods which fall into two categories: hand-crafted feature-based methods and deep learning-based methods. Compared to the hand-crafted feature-based methods, such as region growing [1], active contour model [2], and morphological model [3], deep learning-based methods ([1], [2], [3], [4], [5], [6]) could automatically learn important features [7] without manually empirical parameter adjustments.

Existing hand-crafted feature-based lung segmentation methods are usually performed through a series of procedures with manually empirical parameter adjustments. Various sets of 2D-based [3] and 3D-based methods [8] are developed to achieve a high quality segmented result. However, these traditional segmentation techniques are designed for specific applications, imaging modalities, and even datasets. They are difficult to be generalized for different types of CT images or various datasets since different kinds of features and different parameter/threshold values are extracted from different datasets. Moreover, the feature extraction procedure is monitored by users to manually and interactively adjust the features/parameters for analyzes.

In this paper, to solve the medical image segmentation problem, especially the problem of lung segmentation in CT scan images, we propose LGAN (Generative Adversarial Network-based Lung Segmentation) schema which is an end-to-end deep learning model for lung segmentation from CT images based on a Generative Adversarial Network (GAN) structure combining an EM distance-based loss function. In the proposed schema, a deep Deconvnet Network is trained to generate the lung mask while an Adversarial Network is trained to discriminate segmentation maps from the ground truth and the generator, which, in turn, helps the generator to learn an accurate and realistic lung segmentation of the input CT scans. The performance analysis on a dataset selected from the LIDC-IDRI dataset [9] shows the effectiveness and stability of this new approach. The proposed LGAN schema outperforms the state-of-the-art methods on our dataset and debuts itself as a promising tool for automatic lung segmentation and other medical imaging segmentation tasks.

II. The Proposed Method

A. The Proposed LGAN Schema

The design of the proposed LGAN schema is, via the competition game between the lung mask generator network and mask discrimination network, to force the generated lung segmentation mask to be more consistent and close to the ground truth.

As illustrated in Fig. 1, LGAN consists of two networks: the mask generator network and the discriminator network, and both of them are convolution neural networks. The mask generator network is to generate the lung segmentation masks based on the grayscale input images of CT slices, while the discriminator computes the EM distance between the predicted masks and the ground truth masks.

The LGAN schema takes a slice of the lung CT scan I_i as input, then the generator predicts a mask M_i to illustrate the pixels belong to the lung. The quality of lung segmentation is judged by how well M_i fools the discriminator network. In the rest of this section, we describe the three main components of our LGAN schema: Generator Network, Discriminator Network, and Training Loss.

1) Mask Generator Network: The mask generator network is designed to generate the segmented mask of the input

*Equal contributions.
♮Oguz Akin, MD serves as a scientific advisor for Ezra AI, Inc., which is unrelated to the research being reported.
†Corresponding author. E-mail: ytian@ccny.cuny.edu.

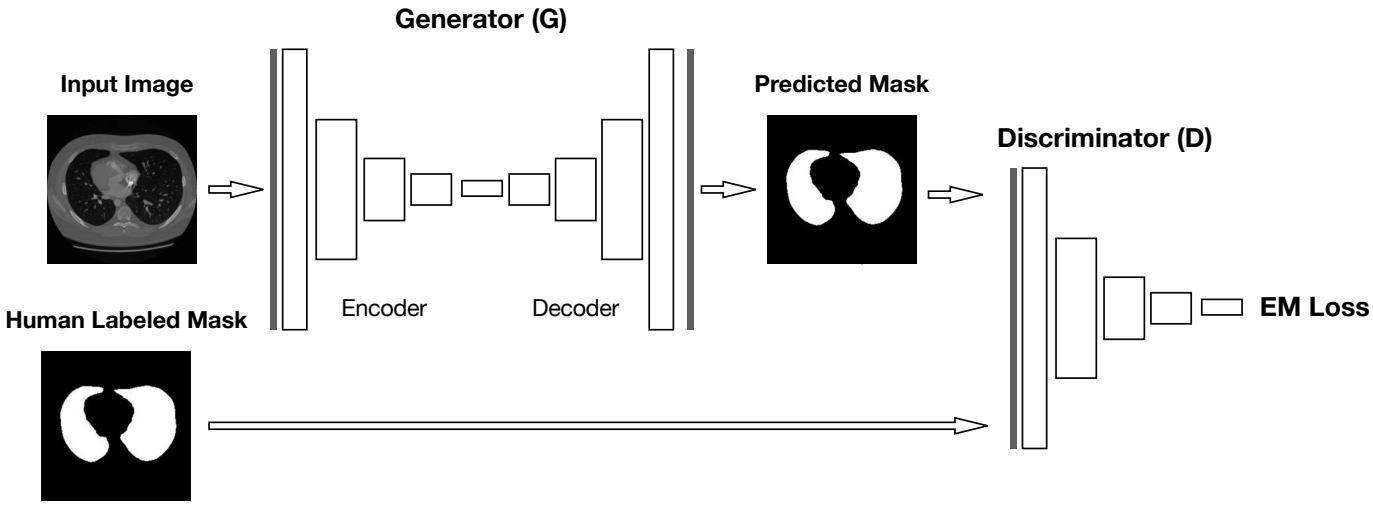

Fig. 1: The pipeline of the proposed LGAN schema which includes a generator network and a discriminator network. A fully convolutional neural network is trained to generate the lung mask while an Adversarial Network is trained to discriminate segmentation masks from the ground truth and the generator, which, in turn, helps the generator to learn an accurate and realistic lung segmentation of the input CT scans.

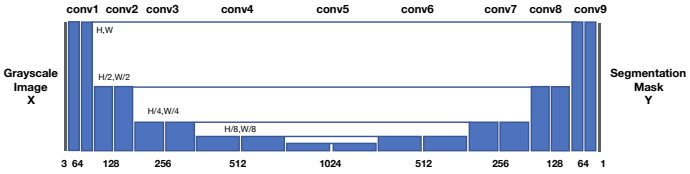

Fig. 2: The architecture of the generator in the proposed LGAN framework. Each blue box represents the feature map generated by convolution block. The number of channels is denoted on the bottom of the box. The lines on the top of the boxes indicate the concatenation operation of the feature map.

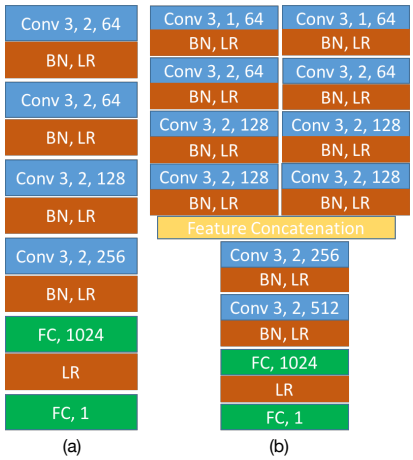

Fig. 3: The architecture of the original Discriminative network (a) and our design (b). *Conv* stands for convolution layer, *FC* stands for fully-connected layer, *BN* stands for Batch Normalization and *LR* stands for LeakyReLU. For each convolution layer, the numbers represent kernel size, (down) pooling stride and number of kernels accordingly. Feature concatenation layer combines feature maps from different branches.

lung CT scan image, which labels all the pixels belonging to the lung. This segmentation task can be addressed as a pixel-wise classification problem to identify whether a pixel belongs to the lung area or not. Given an input CT slice I_i, the generator will predict the category of each pixel and generate a corresponding mask M_i based on the classification result.

The design of the generator is illustrated in Fig. 2, which consists of encoder and decoder parts. The encoder extracts the high-level features from the input CT scan by a bunch of convolution blocks, while the decoder reconstructs the mask from the high-level features. Both encoder and decoder are composed of convolution blocks, which are represented as blue boxes in the figure. The input of the generator network is normalized to 224×224 pixels and the generated mask is the same size as the input.

In the encoder part, each block has two convolution layers, both of which have the same number of filters with filter size 3×3 followed by a max-pooling layer, which performs a 2×2 down-pooling on the feature map. In the decoder part, each block consists of one deconvolution layer and two convolution layers. For the convolution layers, similarly, each has the same number of filters with filter size 3×3. Instead of an up-pooling layer, we use the deconvolution layer with stride 2 as suggested by Tran *et al.* [10] because deconvolution layer can generate better quality images than an up-pooling layer. Skip connections are added between the encoder and decoder to reduce number of parameters as well as reuse extracted feature.

2) Discriminative Network: The task of the discriminative network is to distinguish the ground truth mask from the generated segmentation mask so as to force more consistent and realistic mask to be generated. And EM distance is adapted to measure the difference between the real and the generated mask as it has been proved to be a smooth metric [11].

Given a generator, the discriminator approximates the dis-

TABLE I: Performance comparison of different LGAN structures.

	Mean		Median	
Model	IOU	Hausdorff	IOU	Hausdorff
U-net [12]	0.6248	6.1062	0.7582	5.8310
$LGAN_{Basic}$	0.9018	3.3672	0.9655	3.1622
$LGAN_{Regression}$	**0.9225**	**3.3802**	**0.9715**	3.1622

tance of the distribution of the ground truth (GT) and generated mask (MK) via EM loss:

$$EM(GT, MK) = E[GT] - E[MK]. \tag{1}$$

3) Training Loss: We formulate the task of image segmentation into the framework of GAN by modify the training loss. Specifically, we modify the loss of generator G by adding a Binary Cross Entropy (BCE) loss which calculates the cross-entropy between the generated mask and ground truth mask. Therefore, the loss of the generator network is:

$$\mathbb{BCE}[G(x), Real] - \mathbb{E}_{x\sim P_z}[D(x)], \tag{2}$$

where P_z is the learned distribution from the ground-truth mask by G.

B. Our proposed LGAN structures

We propose five different LGAN structures, each with different Discriminator designs. Due to page limit, we are presenting our basic design and the best design, $LGAN_{Basic}$ and $LGAN_{Regression}$, which can be shown in Fig. 3. For $LGAN_{Basic}$, the discriminator network is to evaluate the generated mask and the ground truth mask separately and minimize the distance between them. For $LGAN_{Regression}$, we design the discriminator as a regression network to approximate the $E[D(G(z))] - E[D(Real)]$ where $D(G(z))$ and $D(Real)$ are evaluated together in the same network setting.

III. Experimental Results

A. Experiment Design

Our proposed LGAN structures are validated and compared on a dataset from Dr. Lihong Li's previous work [13] containing 220 patients selected from the publicly available LIDC-IDRI dataset [9]. Each scan contains more than 130 slices with a size of 512×512 pixels. We randomly select 180 patients' scans as the training data and the other 40 patients' scans as the testing data for experiments. The ground-truth for lung segmentation masks are from [13], which were obtained by first using Vector Quantization-based Lung segmentation to filter out major lung parts, then applying region growing to smooth the result, and finally corrected by radiologists.

We use Adam [14] optimizer with a batch size of 32. All models are trained from scratch without using pretrained weights. The learning rate is set to 10^{-5}, momentum to 0.9, and weight decay to 0.0005. The network is initialized with a Gaussian distribution. During testing, only the mask generator network is employed to generate the final mask. The source code will be made publicly available on the project website following the acceptance of the paper.

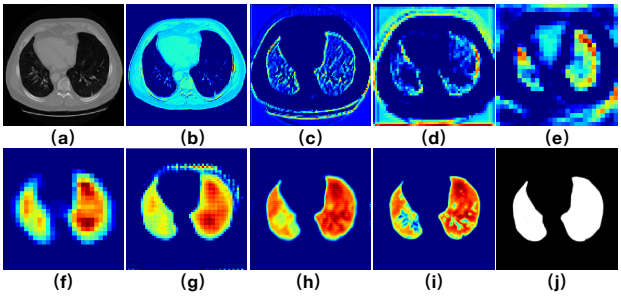

Fig. 4: Visualization of the activation maps of the generator network. The activation maps from (b) to (i) correspond to the output maps from lower to higher layers in the generator. We select the most representative activation in each layer for effective visualization. The image (a) is the input image and image (k) is the predicted mask. The finer details of the lung are revealed, as the features are forward-propagated through the layers in the generator. It shows that the learned filters tend to capture the boundary of the lung.

B. Evaluation Metrics

We take two metrics to evaluate the performance of the networks: segmentation quality and shape similarity.

1) Segmentation Quality: Intersection over Union (IOU) score is a commonly used for semantic segmentation. Given two images X and Y, where X is the predicted mask and Y is the ground truth. The IOU score is calculated as:

$$IOU = \frac{X \cap Y}{X \cup Y}, \tag{3}$$

which is the proportion of the overlapped area to the combined area.

2) Shape Similarity: To evaluate the similarity between shapes, the commonly used Hausdorff distance [15] is employed to measure the similarity between the segmented lung and the ground truth. In this paper, we use the symmetrical Hausdorff distance mentioned in [16] as the shape similarity evaluation metric.

Given generated mask $\mathfrak{M}$ and groundtruth $\mathfrak{G}$, the symmetrical Hausdorff distance is calculated as:

$$HausDist(\mathfrak{M}, \mathfrak{G}) = \max\left\{\begin{matrix}\sup_{x\in\mathfrak{M}} \inf_{y\in\mathfrak{G}} \|x-y\|, \\ \sup_{x\in\mathfrak{G}} \inf_{y\in\mathfrak{M}} \|x-y\|\end{matrix}\right\}. \tag{4}$$

For all three evaluation metrics, we compute and compare their mean values as well as their median values.

C. Effectiveness of the proposed Discriminator Structure

The goal of our first experiment has two folds. The first is to demonstrate the improvement of our proposed LGAN structure compared with U-net alone. The other is to compare our two LGAN structure designs, $LGAN_{Basic}$ and $LGAN_{Regression}$. As shown in Table I, both LGAN structures achieve a significant improvement compared with U-Net, with more than 20% higher performance, which demonstrates the effectiveness of the LGAN. Furthermore, $LGAN_{Regression}$ significantly outperforms the $LGAN_{Basic}$ by a large margin.

TABLE II: Performance Comparison with the state-of-the- arts (3D Dice-score).

Model	Mean	Median
Morph [3]	0.862±2.93	-
U-net [12]	0.970±0.59	0.98449
SegCaps [5]	-	0.9847
$\text{LGAN}_{\text{Regression}}$	**0.985±0.03**	**0.9864**

D. Comparison with the State-of-the-arts

Furthermore, we compare the performance of our LGAN model with the current state-of-the-arts of lung segmentation on LIDC-IDRI dataset, including the traditional benchmark method [3], U-net model [12], and SegCaps [5]. The commonly used 3D Dice-score metrics and the mean as well as median values are calculated following the same settings. As shown in TABLE II, our model achieves the highest score comparing to current state-of-the-arts with an average Dice-score of 0.985 and a median Dice-score of 0.9864.

E. Discussion

The experimental results demonstrate that our proposed LGAN structures significantly outperform the U-Net structure, which shows the effectiveness of the LGAN schema. Our model outperforms current state-of-the-arts of segmentation task on a subset of LIDC-IDRI dataset with higher Dice-score.

The generator network in our LGAN model is designed based on the currently most widely used benchmark method, U-Net. As the task of finding an optimal network structure is still ongoing, our LGAN schema could also be optimized correspondingly. A deeper network design would extract higher-level features but requires more data as well as more parameters and higher computation cost. Patch normalization and random initialization in our model training show a significant effect and the optimization method by Adam [14] is also used in our work.

IV. CONCLUSIONS

Lung segmentation is usually performed by methods such as thresholding and region growing. Such methods, on the one hand, require dataset-specific parameters, and on the other hand, require a series of pre- and post-processing to improve the segmentation quality. To reduce the processing steps for lung segmentation and eliminate the empirical based parameters adjustments, we have proposed a Generative Adversarial Network based lung segmentation schema (LGAN) by redesigning the discriminator with EM loss. The lung segmentation is achieved by the adversarial between the mask generator network and the discriminator network which can differentiate the real mask from the generated mask. Such adversarial makes the generated mask more realistic and accurate than a single network for image segmentation. Moreover, our schema can be applied to different kinds of segmentation networks.

V. ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under award number IIS-1400802 and Memorial Sloan-Kettering Cancer Center Support Grant/Core Grant P30 CA008748. The authors would appreciate Dr. Lihong Li for managing data and providing groundtruth, which were used in [13].

REFERENCES

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

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

[3] A. Mansoor, U. Bagci, B. Foster, Z. Xu, G. Z. Papadakis, L. R. Folio, J. K. Udupa, and D. 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.

[4] A. P. Harrison, Z. Xu, K. George, L. Lu, R. M. Summers, and D. J. Mollura, "Progressive and multi-path holistically nested neural networks for pathological lung segmentation from ct images," *arXiv preprint arXiv:1706.03702*, 2017.

[5] R. LaLonde and U. Bagci, "Capsules for object segmentation," *arXiv preprint arXiv:1804.04241*, 2018.

[6] T. Zhao, D. Gao, J. Wang, and Z. Tin, "Lung segmentation in ct images using a fully convolutional neural network with multi-instance and conditional adversary loss," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 505–509.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] S. Sun, C. Bauer, and R. Beichel, "Automated 3-d segmentation of lungs with lung cancer in ct data using a novel robust active shape model approach," *IEEE transactions on medical imaging*, vol. 31, no. 2, pp. 449–460, 2012.

[9] S. Armato and et al, "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Med Phys.*, vol. 38, no. 2, pp. 915–931, 2011.

[10] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Deep end2end voxel2voxel prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 17–24.

[11] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[12] O. Ronneberger, P. Fischer, and T. 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.

[13] H. Han, L. Li, F. Han, B. Song, W. Moore, and Z. Liang, "Fast and adaptive detection of pulmonary nodules in thoracic ct images using a hierarchical vector quantization scheme," *IEEE journal of biomedical and health informatics*, vol. 19, no. 2, pp. 648–659, 2015.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[15] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. Springer Science & Business Media, 2009, vol. 317.

[16] S. Nutanong, E. H. Jacox, and H. Samet, "An incremental hausdorff distance calculation algorithm," *Proceedings of the VLDB Endowment*, vol. 4, no. 8, pp. 506–517, 2011.