

LGAN: Lung Segmentation in CT scans using Generative Adversarial Network

Jiaxing Tan^{1,a}, Longlong Jing^{1,a}, Yumei Huo^a, Lihong Li^a, Oguz Akin^b, Yingli Tian^{*a}

^a*The City University of New York, New York, 10016, USA*

^b*Memorial Sloan Kettering Cancer Center, New York 10065, USA*

Abstract

Lung segmentation in computerized tomography (CT) images plays an important role in various lung disease diagnosis. Most of the current lung segmentation approaches are performed through a series of procedures with manually empirical parameter adjustments in each step. Pursuing an automatic segmentation method with fewer steps, we propose a novel deep learning Generative Adversarial Network (GAN)-based lung segmentation schema, which we denote as LGAN. The proposed schema can be generalized to different kinds of neural networks for lung segmentation in CT images. We evaluated the proposed LGAN schema on datasets from Lung Image Database Consortium image collection (LIDC-IDRI) and Quantitative Imaging Network (QIN) collection with two metrics: segmentation quality and shape similarity. Also, we compared our work with current state-of-the-art methods. The experimental results demonstrated 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 improved performance and efficiency.

Keywords: Deep Learning, Lung Segmentation, Generative Adversarial Network, Medical Imaging Analysis, Thorax CT Images

*Corresponding author. Email: ytian@ccny.cuny.edu.

¹Equal contributions

1. Introduction

Computerized tomography (CT) is a clinical imaging modality, that is key to, and sometimes the first step in, the diagnosis of various lung diseases, including lung cancer, which has been the leading cause of cancer-related deaths in the United States with the overall five-year survival rate of 17% [1], but the survival rate could be increased to 55% if localized, where usually performed by CT scan analysis. Lung CT scan images allow a physician to confirm the presence of a tumor, measure its size, identify its precise location and determine the extent of its involvement with other nearby tissue. With an increasing use of CT imaging for clinical studies, it has become almost compulsory to use computers to assist radiologists in clinical diagnosis and treatment planning [2], [3], [4], [5], [6], [7].

Among all the Lung CT computer aided detection/diagnosis applications, lung segmentation is an initial step in analyzing medical images obtained 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 [8], active contour model [9], and morphological-based models [10], deep neural network based methods [11], [12], [13] could automatically learn representative features [14] without manually empirical parameter adjustments.

Existing hand-crafted feature-based lung segmentation methods are usually performed through a series of procedures with manual empirical parameter adjustments. 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. Compared to the hand-crafted methods, the deep learning-based methods [10], [15], require less data-specific hyper-parameters and generally perform better than the hand-crafted methods.

In this paper, we propose an end-to-end deep learning Generative Adversarial Network based lung segmentation schema, LGAN, where the input is a slice of lung CT scan and the output is a pixel-wise mask showing the position of the lungs by identifying whether each pixel belongs to lung or not. Furthermore, the proposed schema can be generalized to different kinds

of networks with improved performance.

Recently, several deep learning-based pixel-wise classification methods have been proposed in computer vision area and some of them have been successfully applied in medical imaging. Early deep learning-based methods are based on bounding box [16]. The task is to predict the class label of the central pixel(s) via a patch including its neighbors. Kallenberg *et al.* [17] designed a bounding box based deep learning method to perform breast density segmentation and scoring of mammographic texture. Shin *et al.* [18] compared several networks on the performance of computer-aided detection and proposed a transfer learning method by utilizing models trained in computer vision domain for medical imaging problem. Instead of running a pixel-wise classification with a bounding box, Long *et al.* [19] proposed a fully convolutional network (FCN) for semantic segmentation by replacing the fully connected layers with convolutional layers. An Auto-Encoder alike structure has been used by Noh *et al.* [20] to improve the quality of the segmented objects. Later, Ronneberger *et al.* [21] proposed a U-net model for segmentation, which consists of a contracting part as an encoder to analyze the whole image and an expanding part as a decoder to produce a full-resolution segmentation. The U-net architecture is different from [19] in that, at each level of the decoder, a concatenation is performed with the correspondingly cropped feature maps from the encoder. This design has been widely used and proved to be successful in many medical imaging applications such as Lumbar Surgery [22] and gland segmentation [23]. Most recently, Lalonde *et al.* [12] designed a convolutional-deconvolutional capsule network, called SegCaps, to perform lung segmentation, where they proposed the concept of deconvolutional capsules.

After the emergence of Generative Adversarial Network GAN-based models [24], which have shown a better efficiency in leveraging the inconsistency of the generated image and ground truth in the task of image generation, Luc *et al.* [25] proposed a GAN-based semantic segmentation model. The motivation is to apply GAN to detect and correct the high-order inconsistencies between ground truth segmentation maps and the generated results. The model trains a segmentation network along with an adversarial network that discriminates segmentation maps coming either from the ground truth or from the segmentation network. Following this idea, Zhao *et al.* [13] proposed to use adversary loss to perform lung segmentation, where the segmentor is a fully convolutional neural network. Both of their models employ the original GAN structure, which, however, due to its loss function

design, original GAN suffers from the problem of learning instability such as mode collapse, which means all or most of the generator outputs are identical ([26],[27]).

To avoid this problem, Arjovsky *et al.* [26] proposed an optimized GAN structure which uses a new loss function based on the Earth Mover (EM) distance and in the literature is denoted as WGAN. It should be noted that WGAN is designed to solve the same problem as the original GAN, which is to leverage the inconsistency of the generated image and ground truth in the task of image reconstruction instead of generating an accurate segmentation from a given type of images.

In this paper, to solve the medical image segmentation problem, especially the problem of lung segmentation in CT scan images, we propose LGAN schema which is a general deep learning model for segmentation of lungs from CT images based on a Generative Adversarial Network structure combining the 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 LIDC-IDRI and QIN datasets shows the effectiveness and stability of this new approach. A very preliminary version of this work has been reported [28]. The main contributions of this paper include:

1. We propose a novel end-to-end Generative Adversarial Network-based lung segmentation schema with EM distance to perform pixel-wised lung segmentation.
2. We apply the LGAN schema to five different GAN structures for lung segmentation and compare them with different metrics including segmentation quality and shape similarity.
3. We perform experiments and evaluate our five LGAN segmentation algorithms as well as the baseline U-net model using LIDC-IDRI and QIN datasets with ground truth masks generated by transitional lung segmentation method with further corrections from our radiologists.
4. Our experimental results show that the proposed LGAN schema outperforms current state-of-the-art methods and debuts itself as a promising tool for automatic lung segmentation.

2. The Proposed Method

In this section, we first introduce the background knowledge of Generative Adversarial Network and then present the proposed LGAN schema.

2.1. Generative Adversarial Networks

Generative Adversarial Network is a deep generative model initially proposed in [24], and later improved by DCGAN [29] and WGAN [26]. A general GAN model consists of two kinds of networks named as the generator network and the discriminator network. The generator network is trained to generate an image similar to the ground-truth and meanwhile the discriminator network is trained to distinguish the generated image from the ground-truth image. By playing this two-player game, the results from the discriminator network help the generator network to generate more similar images and simultaneously the generated images as the input data help the discriminator network to improve its differentiation ability. Therefore, the generator network and the discriminator network are competing against each other while at the same time make each other stronger.

Mathematically, the goal of the generator network G is to learn a distribution p_z matching the ground-truth data in order to generate the similar data, while the goal of discriminator network D is also to learn the distribution of the ground-truth data but for distinguishing the real data (i.e. from the real distribution p_d) from the generated data from G . The adversarial comes from the min-max game between G and D , and is formulated as:

$$\min_G \max_D \mathbb{E}_{x \sim p_d(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))], \quad (1)$$

where, for a given real data x and the corresponding generated data $G(z)$, the adversarial discriminator is trained to maximize the probability output for the real data x (that is, $\mathbb{E}_{x \sim p_d(x)} [\log D(x)]$) and minimize the probability output for the generated data (that is, $\mathbb{E}_{x \sim p_d(x)} [\log D(G(z))]$) which is equivalent to maximizing $\mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$, and on the other side, the generator network is trained to generate $G(z)$ as similar as possible to x so that the discriminator outputs the bigger probability value for $G(z)$, that is, to maximize $\mathbb{E}_{x \sim p_d(x)} [\log D(G(z))]$ and equivalently to minimize $\mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$.

Luc *et al.* [25] employed GAN model to perform segmentation task, where the role of the generator has been changed from generating synthetic images to generating segmentation masks for the original images, which has been

proved to be effective on the task of lung segmentation by Zhao *et al.* [13]. The details of GAN-based segmentation design will be specified in the next section.

As illustrated in [26], the original GAN structure, which although achieves a great performance in various tasks, including replicating images, human language, and image segmentation, suffers from a mode collapse problem due to its loss design. To make the training process more stable, Arjovsky *et al.* proposed WGAN using Earth Mover (EM) distance to measure the divergence between the real distribution and the learned distribution [26]. Specifically, given the two distributions, p_d and P_z , with samples $x \sim p_d$ and $y \sim P_z$, the EM distance is defined as:

$$W(p_d, P_z) = \inf_{\gamma \in \Pi(p_d, P_z)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|], \quad (2)$$

where $\Pi(p_d, P_z)$ represents the set of all joint distributions $\gamma(x, y)$ whose marginals are respectively p_d and P_z , and the term $\gamma(x, y)$ represents the cost from x to y in order to transform the distributions p_d into the distribution P_z . The EM loss actually indicates optimal transport cost. In this design, the loss for the generator network is:

$$L_G = -\mathbb{E}_{x \sim p_z} [D(x)]. \quad (3)$$

And the loss for the discriminator network is:

$$L_D = \mathbb{E}_{x \sim p_z} [D(G(x))] - \mathbb{E}_{x \sim p_{real}} [D(x)]. \quad (4)$$

With the EM distance-based loss, the GAN model becomes more powerful in generating high-quality realistic images and outperforms other generative models. While the WGAN is designed for image reconstruction, here we take advantage of the basic idea of WGAN, and design an efficient and enhanced deep learning GAN-based lung segmentation schema.

2.2. Our LGAN Schema

Our LGAN schema is designed to force the generated lung segmentation mask to be more consistent and close to the ground truth mask and its architecture is illustrated in Fig. 1.

LGAN consists of two networks: a generator network (G) and a discriminator network (D), and both of them are convolutional neural networks. The generator network is trained to predict the lung masks based on the grayscale

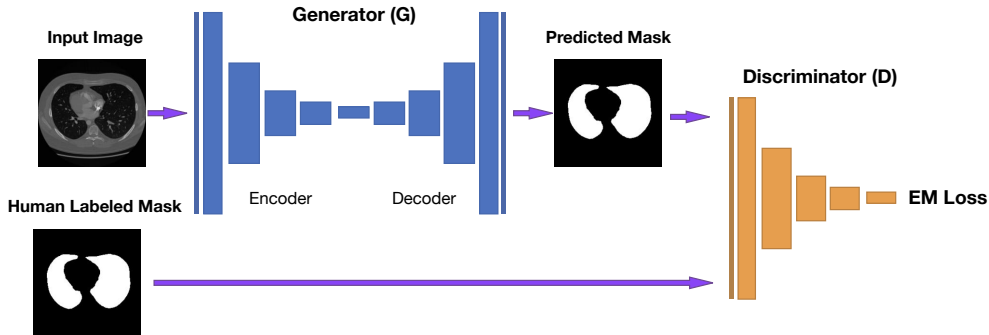


Figure 1: The pipeline of the proposed LGAN schema which includes a generator network (G) and a discriminator network (D). A Deep Fully Convolutional Network is trained to generate the lung mask while an Adversarial Network is trained to discriminate segmentation maps from the ground truth and the generator, which, in turn, helps the generator to learn an accurate and realistic lung segmentation of the input CT scans.

input CT slices, while the discriminator computes the EM distance between the predicted masks and the ground truth masks to help the generator to learn accurate and realistic lung segmentation masks.

During the training, 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 the 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.

2.2.1. Generator Network

The generator network is designed to generate the segmented mask of the input lung CT scan image. The mask 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 architecture of our designed generator is illustrated in Fig. 2. The generator model consists of encoder and decoder parts. The encoder extracts multi-scale features from gray input CT scans by a bunch of convolution blocks, while the decoder predicts masks from the multi-scale features extracted by the encoder. Both encoder and decoder are composed of convolution blocks, and the feature map for each block is represented as blue

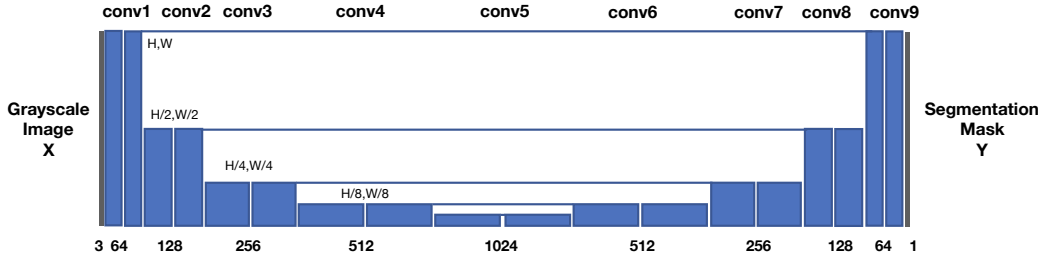


Figure 2: The architecture of the generator network in the proposed framework. Each blue box represents the feature map generated by convolution block. The number of channels is denoted on the bottom of the box. The lines on the top of the boxes indicate the concatenation operation of the feature map.

boxes in the figure. 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.

Following DCGAN [29], we employ LeakRelu as the activation function for the convolution layers which is first proposed in [30]. As shown below, to alleviate potential problems caused by ReLU, which sets 0 to all negative values, LeakyReLU set a small non-zero gradient NegativeSlope, which is user pre-defined, to negative values. In the equation below, we represent this negative slope as α .

$$\text{LeakyReLU}(x, \alpha) = \max(x, 0) + \alpha \times \min(x, 0). \quad (5)$$

At the final layer, a 1×1 convolution is performed to map each component feature vector to the final segmentation mask.

2.2.2. Discriminative Network

The task of the discriminative network is to distinguish the ground truth mask from the generated segmentation mask. The EM distance is employed to measure the difference between the real and the learned distributions as it has been proved to be a smooth metric [26].

Given a generator, the discriminator approximates $E[\cdot]$ function such that the EM loss $E_{x \sim p_z}[x] - E_{x \sim p_{real}}[x]$ is approximated by $D_{x \sim p_z}(G(x)) -$

Table 1: The list of all the proposed five LGAN structures and their corresponding descriptions.

Network	Input of the Discriminator Network
$LGAN_{Basic}$	Generated mask, one at a time.
$LGAN_{Product}$	Segmented original image based on the predicted mask.
$LGAN_{EF}$	Mask and original image are combined as one input with two channels.
$LGAN_{LF}$	Mask and original image as two inputs.
$LGAN_{Regression}$	Approximate EM loss based on prediction and ground truth directly.

$D_{x \sim p_{real}}(x)$. Compared to the discriminator in the vanilla GAN, which performs a classification task, the new discriminator is actually performing a regression task (approximating the function $E(\cdot)$).

Based on the different assumptions that could help improve the performance of the discriminator network, we propose five different designs for the discriminator network, which thus yield five different LGAN structures as listed in Table 1. We discuss these five designs one by one in the next section.

2.2.3. Training Loss

As the original WGAN is designed for image generation tasks, here we modify the training loss to fit for the segmentation task. Specifically, we modify the loss of generator G by adding a Binary Cross Entropy (BCE) loss which calculates the cross-entropy between the generated lung mask and ground truth lung mask. Therefore, the loss of the generator network is:

$$\text{BCE}[G(x), Real] - \mathbb{E}_{x \sim p_z}[D(x)], \quad (6)$$

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

For the training loss of the discriminator D , different designs for the discriminator network may have different training loss functions which are described in the next section.

3. The Proposed LGAN Structures

In GAN-based image generation tasks, the generated images and the real images are very similar. However, for the lung segmentation task, the pixel intensity in the predicted mask is in $[0, 1]$ while the value in the ground truth

mask is binary, that is, either 0 or 1. This fact may mislead the discriminator to distinguish the generated mask and the ground truth mask by simply detecting if the mask consists of only zeros and ones (one-hot coding of ground truth), or the values between zero and one (output of segmentation network).

With this observation, we explore all possible discriminator designs for lung segmentation task based on various assumptions, and provide five different LGAN structures: LGAN with Basic Network ($LGAN_{Basic}$), LGAN with Product Network ($LGAN_{Product}$), LGAN with Early Fusion Network ($LGAN_{EF}$), LGAN with Late Fusion Network ($LGAN_{LF}$), and LGAN with Regression Network ($LGAN_{Regression}$). Their corresponding architectures are illustrated in Fig. 3, where the structure of each network is optimized based on a structural search. In the rest of this section, we introduce them accordingly.

3.1. LGAN with Basic Discriminator

The basic discriminator is to evaluate the generated mask and the ground truth mask separately and minimize the distance between the two distributions. The architecture of the discriminator network for this design is illustrated as (a) in Fig. 3. We denote the LGAN with this basic discriminator as $LGAN_{Basic}$ and it has a single channel with the network input size of 224×224 .

In $LGAN_{Basic}$, the training loss of G is the same as we described in Section 2.2.3 and the training loss of D is the following:

$$\mathbb{E}_{x \sim p_z}[D(G(x))] - \mathbb{E}_{x \sim p_{real}}[D(x)]. \quad (7)$$

Based on $LGAN_{Basic}$, we conjecture that the discriminator network may have a more precise evaluation if the original image is also provided as additional information. Under this assumption, we examine three strategies and design the $LGAN_{Product}$, $LGAN_{EF}$, and $LGAN_{LF}$ structures.

3.2. LGAN with Product Network

Different from the basic discriminator where the inputs are the segmented mask and the ground-truth mask with only binary value, the product network takes, as the inputs, the lung volume images which are mapped out from the original CT scan image by the segmented mask and the ground-truth mask respectively. That is, given the segmentation mask, we obtain the lung

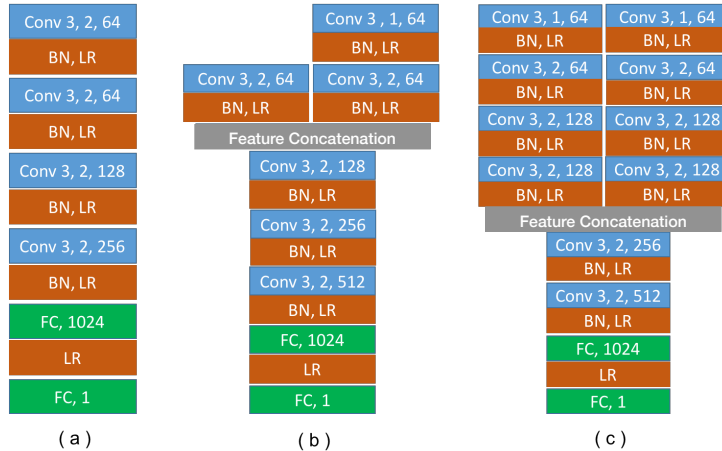


Figure 3: The architectures of three different discriminator networks, where *Conv* stands for convolution layer, *FC* stands for fully-connected layer, *BN* stands for Batch Normalization and *LR* stands for LeakyReLU. For each convolution layer, the numbers represent kernel size, (down) pooling stride and number of kernels accordingly. Feature concatenation layer concatenates feature maps from different branches and feed the concatenated features to its next layer.

volume image by modifying the original image such that the values within the segmented lung area are kept as they are but the values in the rest of the area are set to be 0. This design is motivated by the work of Luc *et al.* [25].

With this input, the discriminator network might be biased by the value distribution. Although in [25] the deep learning model with the product network is not designed based on WGAN, we observe that the product network could still be used in our LGAN model and we define this LGAN structure as $LGAN_{Product}$. The discriminator in $LGAN_{Product}$ differs from $LGAN_{Basic}$ only in the number of channels in inputs, so the discriminator in $LGAN_{Product}$ shares the same architecture as $LGAN_{Basic}$ shown in (a) of Fig. 3.

In $LGAN_{Product}$, the loss of G is the same as we described in Section 2.2.3 and the training loss of D , which is product network, is the following:

$$\mathbb{E}_{x \sim p_z} [D(G(x) \circ I_i)] - \mathbb{E}_{x \sim p_{real}} [D(x)], \quad (8)$$

where \circ is an operation such that $x \circ y$ represents pixel-wise multiplication of matrices x and y .

3.3. LGAN with Early Fusion Network

Instead of taking only the mask as inputs, early fusion network takes both the whole original CT scan image and the segmentation/ground-truth

mask as an input. To keep the design of single input, we concatenate the original image and the mask as one single image with two channels, where one channel is the original CT scan and the other is the mask. We denote *LGAN* with this early fusion discriminator network as $LGAN_{EF}$.

The architecture of the discriminator network in $LGAN_{EF}$ is shown in (a) of Fig. 3. Different from $LGAN_{Basic}$ and $LGAN_{Product}$, $LGAN_{EF}$ has the input size of 224×224 with 2 channels which are the concatenation of the original CT scan and its mask. In $LGAN_{EF}$, the training loss of G is the same as we described in Section 2.2.3 and the loss of the discriminator network is:

$$\mathbb{E}_{x \sim p_z} [D(G(x) \oplus I_i)] - \mathbb{E}_{x \sim p_{real}} [D(x) \oplus I_i], \quad (9)$$

where \oplus is an operation such that $x \oplus y$ represents concatenation of matrices x and y into a single matrix with 2 channels.

3.4. *LGAN with Late Fusion Network*

Another way of taking both the original image and the mask as an input in the discriminator network is to employ the late fusion technique. Specifically, the input of the discriminator is the concatenation of the high-level feature of the CT scan and the mask. We denote *LGAN* with this type of discriminator as $LGAN_{LF}$.

The corresponding architecture of the discriminator network in $LGAN_{LF}$ is shown in (b) of Fig. 3. There are two branches of convolution layers in the discriminator network while one branch is for the CT slices and the other branch is for lung masks. The two inputs first pass the two parallel branches separately, and then their features are fused by a concatenate layer and pass through several convolution layers and down-sampling layers before they pass through the fully-connected layers to reach the final result. As the CT scan is more complicated and provides more information than masks, we let the CT scan pass through more convolution layers before the feature concatenation layer.

In $LGAN_{LF}$, the training loss of G is the same as we described in Section 2.2.3 and the loss of the discriminator network is:

$$\mathbb{E}_{x \sim p_z} [D(G(x), I_i)] - \mathbb{E}_{x \sim p_{real}} [D(x, I_i)]. \quad (10)$$

3.5. *LGAN with Regression Network*

As we addressed in Section 2.2.2, in WGAN, the EM loss $E[X] - E[Y]$ is approximated by $D(G(z)) - D(Real)$ where $D(G(z))$ and $D(Real)$ are evaluated separately and independently. Differently, we design the discriminator

network 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. The regression discriminator network takes two inputs, the ground truth lung mask and the mask generated by the generator network. The output of the network is the approximated EM distance, and the network is optimized by minimizing the distance. We denote LGAN with this regression discriminator network as $LGAN_{Regression}$.

The architecture of the discriminative network in $LGAN_{Regression}$ is shown in (c) of Figure 3. Similar to the previous networks, the inputs in the regression discriminator network first separately pass through their own convolution branches and down-sampling layers before their features are concatenated together. And then the concatenated features pass through more convolution layers before getting into the fully-connected layer. In this discriminator network, the convolution branch consists of a long set of individual convolution layers.

In $LGAN_{Regression}$, the loss of G is the same as we described in Section 2.2.3 and the loss of the discriminator network is:

$$\mathbb{E}_{x_1 \sim p_z, x_2 \sim p_{real}} [D(G(x_1), x_2)]. \quad (11)$$

By playing the min-max game, the generator prevents the distance computed by discriminator from going to positive infinity while the discriminator network prevents it from going to negative infinity. The generator and the discriminator networks play this min-max game for several rounds until a tie is reached.

4. Experiments

4.1. Datasets

We evaluated our proposed methods on two datasets, LIDC-IDRI [31] and QIN Lung CT datasets [32]. The LIDC-IDRI dataset are selected Lung CT scans from the public database founded by the Lung Image Database Consortium and Image Database Resource Initiative, which contains 220 patients with more than 130 slices per scan. Each CT slice has 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 mask for each CT scans is generated by vector quantization-based lung segmentation method [33] and then corrected by radiologists.

The QIN Lung CT dataset contains 47 CT scans obtained on patients diagnosed with Non-Small Cell Lung Cancer (NSCLC) with mixed stage and histology from the H. Lee Moffitt Cancer Center and Research Institute. Each CT slice has a size of 512×512 pixels. The ground-truth lung masks are generated by first applying a watershed-based lung segmentation [34] and then cleaned by radiologists. The CT slices that have low quality of masks are removed from the dataset. Around 30% of the slices from removed from the original QIN dataset. We use 35 patients’ data for training and the rest 12 patients for testing.

4.2. Experiment Design

Our proposed methods of five different LGAN structures are validated and compared on both LIDC-IDRI and QIN datasets. The comparison on structures are described in Section 3 and are listed in TABLE 1. Furthermore, our best model is compared with the state-of-the-arts for lung segmentation task on LIDC-IDRI dataset following the same settings and evaluation metrics. At last, as our method could serve as a pre-processing step for nodule detection. Therefore, three cases with lung nodules located close to the lung boundary are investigated to understand if our generated masks could include those nodules.

All the models are trained from scratch with Adam [35] optimizer. 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 Segmentor 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.

4.3. Evaluation Metrics

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

4.3.1. 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:

$$IOU = \frac{X \cap Y}{X \cup Y}, \quad (12)$$

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

4.3.2. Shape Similarity

To evaluate the similarity between shapes, the commonly used Hausdorff distance [36] 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 [37] as the shape similarity evaluation metric.

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

$$HDist(\mathfrak{M}, \mathfrak{G}) = \max \left(\sup_{x \in \mathfrak{M}} \inf_{y \in \mathfrak{G}} \|x - y\|, \sup_{x \in \mathfrak{G}} \inf_{y \in \mathfrak{M}} \|x - y\| \right). \quad (13)$$

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

5. RESULTS

5.1. Comparison results of our proposed different structures

First we compare the performance of the proposed different structures on LIDC-IDRI dataset, and the experimental results are shown in TABLE 2. The LGAN models achieve a significant improvement compared with the baseline U-net, which also serves as the segmentation network. The performance of LGAN is more than 20% higher than the baseline U-net, which demonstrates the effectiveness of the LGAN. All the LGAN designs achieve better performance than the generator alone, among which the $LGAN_{Regression}$ obtains the best performance.

The performance comparison results on QIN dataset is shown in TABLE 3. Each of our designs has achieved better performance with $LGAN_{Regression}$ obtains the best performance. Since all the slices that have noisy masks are removed from this dataset and the size of the dataset are very small, therefore, the baseline model obtain relative better performance. Our proposed method is very robust on different CT dataset.

To qualitatively study the performance of the proposed architectures and demonstrate the strength of our proposed LGAN framework, we compare the performance of all the models on three CT slices. As shown in Fig. 4, the significant improvement in predicted lung masks using LGAN structures can be observed. The regions with red circles indicate where the network fails. Among all the methods, the baseline method performs worst on all the three slices, while our proposed $LGAN_{Regression}$ performs best and obtains masks that are highly similar to the ground truth.

Table 2: Performance comparison of different LGAN structures on LIDC-IDRI dataset. The numbers in bold indicate the best results.

Model	Mean		Median	
	IOU	Hausdorff	IOU	Hausdorff
<i>Baseline</i>	0.625	6.106	0.758	5.831
<i>LGAN_{Basic}</i>	0.902	3.367	0.966	3.162
<i>LGAN_{Product}</i>	0.886	3.410	0.964	3.317
<i>LGAN_{EF}</i>	0.902	3.284	0.966	3.162
<i>LGAN_{LF}</i>	0.791	3.532	0.944	3.0
<i>LGAN_{Regression}</i>	0.923	3.380	0.972	3.162

Table 3: Performance comparison of different LGAN structures on QIN dataset. The numbers in bold indicate the best results.

Model	Mean		Median	
	IOU	Hausdorff	IOU	Hausdorff
<i>Baseline</i>	0.893	2.787	0.976	2.449
<i>LGAN_{Basic}</i>	0.917	2.758	0.976	2.236
<i>LGAN_{Product}</i>	0.919	2.714	0.977	2.449
<i>LGAN_{EF}</i>	0.913	2.679	0.978	2.236
<i>LGAN_{LF}</i>	0.920	2.687	0.978	2.459
<i>LGAN_{Regression}</i>	0.938	2.812	0.978	2.449

One problem of the neural network-based methods is that it is hard to know what really happened inside the network. Therefore, besides segmentation quality, we would like to attain further insight into the learned convolution models. We select the most representative feature maps obtained by each layer of the generator. The feature maps of LGANs are illustrated in Fig. 5. The feature maps show that the network can extract the major information about lung boundary through the contracting part, and then gradually expanding the extracted highly compressed features into a clear mask. Obviously, the lung area tends to have a brighter color, which means higher activation, than the rest parts of the image.

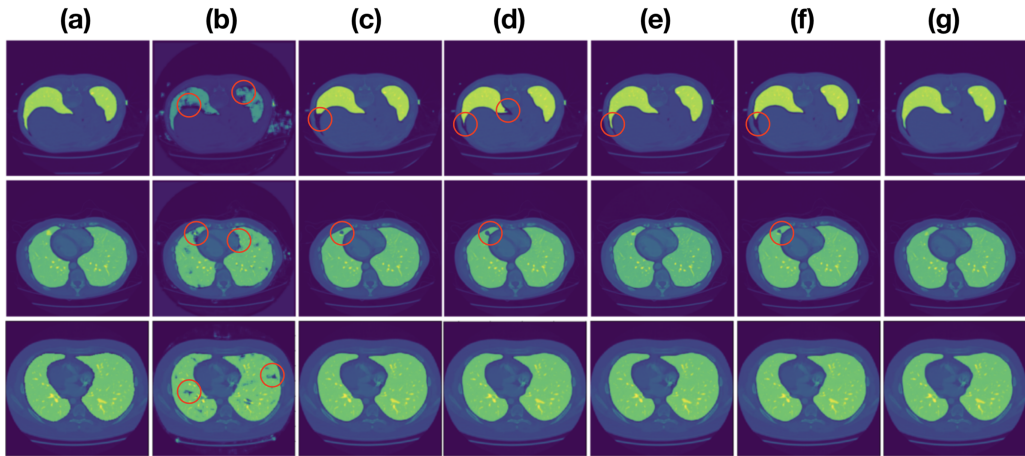


Figure 4: Segmentation results of different LGAN structures on 3 lung CT slices. Column (a) is the overlap of the ground truth lung masks and the original CT slices. The pictures from (b) to (g) correspond to the results of networks of baseline, $LGAN_{Basic}$, $LGAN_{Product}$, $LGAN_{EF}$, $LGAN_{LF}$, and $LGAN_{Regression}$. The region with green color represents lung areas, while the rest regions are other tissue structures, such as the surrounding soft tissue, muscle, bone, and lung blood vessels.

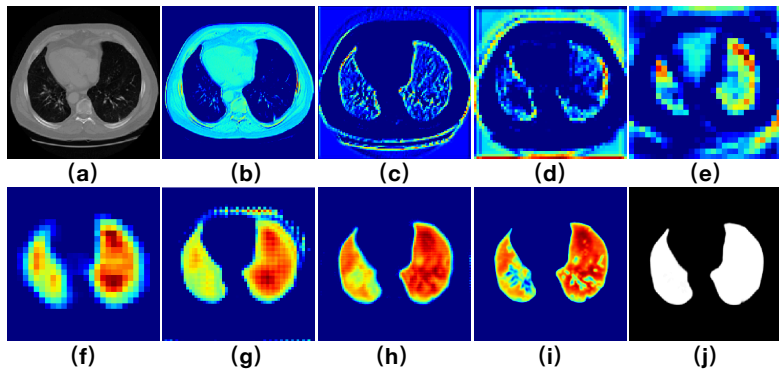


Figure 5: Visualization of activations of the generator network. The activation maps from (b) to (i) correspond to the output maps from lower to higher layers in the generator. We select the most representative activation in each layer for effective visualization. The image (a) is the input image and image (j) 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.

5.2. Comparison with State-of-the-Arts

We compare the performance of our $LGAN_{Regression}$ model with the current state-of-the-arts methods for lung segmentation task on LIDC-IDRI

Table 4: Performance Comparison with the state-of-the-arts methods for lung segmentation task (3D Dice-score) on LIDC-IDRI dataset.

Model	Mean	Median
Morph [10]	0.862±2.93	-
U-net [21]	0.970±0.59	0.9845
Fusion gNet [13]	0.983±0.05	-
Tiramisu [38]	-	0.9841
SegCaps [12]	-	0.9847
LGAN _{Regression}	0.985±0.03	0.9864

dataset, including the traditional method [10], U-net model [21], which serves as our baseline, Tiramisu Network [38], and SegCaps [12]. For fair comparison with others, 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 4, 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. Although SegCaps [12] claims to have fewer parameters, the designed capsule is very memory consuming. Also, our model is much shallower than the 100-layer Tiramisu model [38] and achieves better performance. Meanwhile, our model outperforms the gNet [13], which utilizes the original GAN loss. Comparing to the traditional methods, such as Morph, which requires a series of thresholding, morphological operations, and component analysis, our end-to-end model provides a one-step solution. Moreover, as mentioned in gNet [13], due to the high-resolution of the LIDC-IDRI CT volumes, a better Dice-score with 0.001 improvement indicates that more than $5k$ pixels are correctly predicted.

5.3. Case Study

As lung segmentation usually serves as a pre-processing step for many tasks such as lung nodule detection, we investigate whether the segmented lung areas by our model include all nodules even when the nodules are very close to lung boundary. As shown in Fig. 6, our method can include all the nodules inside the lung area besides achieving high-quality lung mask.

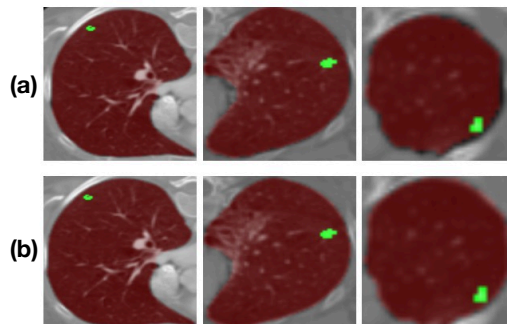


Figure 6: Examples demonstrate that our models are able to segment nodules which are close to lung boundary. The images in rows (a) and (b) are the ground truth lung masks and predicted lung masks by our network. The areas in red represent lung regions and the green areas are the lung nodules annotated by radiologists.

5.4. Discussion and Future Work

For the proposed LGAN schema, we have designed five different discriminative networks, and evaluated these five structures on lung segmentation tasks on two public datasets. The experimental results demonstrate that our proposed LGAN structures significantly outperform current state-of-the-arts of lung segmentation task on LIDC-IDRI dataset with higher Dice-score. Furthermore, our work in this paper makes an important step for lung nodule detection task, especially for detecting the nodules on lung boundary. The generator network in our LGAN model is designed based on the currently most widely used baseline method, U-Net. As the task of finding an optimal network structure is still ongoing, our LGAN schema could also be optimized correspondingly. For the future work, it would be interesting to evaluate the quality of our model by other methods such as nonlinear fitting methods, to employ our method as a pre-processing step for lung nodule detection task, and to extend the proposed schema to the segmentation for other organs such as brain.

6. CONCLUSIONS

Lung segmentation is usually performed by methods such as thresholding and region growing. Such methods, on one hand, require dataset-specific parameters and require a series of pre- and post-processing to improve the segmentation quality, and on the other hand, have low generalization ability to be applied to large-scale diverse datasets. To reduce the processing steps

for lung segmentation and eliminate the adjustments of empirical-based parameters, we have proposed LGAN by redesigning the discriminator with EM loss. The lung segmentation is achieved by the adversarial between the segmentation 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.

Acknowledgments

This work was supported in part by Memorial Sloan Kettering Cancer Center Support Grant/Core Grant P30 CA008748 and National Science Foundation under award number IIS-2041307.

References

- [1] American Cancer Society, Key statistics for lung cancer, 2016.
- [2] X. Huang, W. Sun, T.-L. B. Tseng, C. Li, W. Qian, Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks, *Computerized Medical Imaging and Graphics* 74 (2019) 25–36.
- [3] W. Zhao, H. Liu, J. K. Leader, D. Wilson, X. Meng, L. Wang, L.-A. Chen, J. Pu, Computerized identification of the vasculature surrounding a pulmonary nodule, *Computerized Medical Imaging and Graphics* 74 (2019) 1–9.
- [4] Z. Xue, K. Wong, S. T. Wong, Joint registration and segmentation of serial lung CT images for image-guided lung cancer diagnosis and therapy, *Computerized Medical Imaging and Graphics* 34 (1) (2010) 55–60.
- [5] W. Yang, e. a. Liu, Lung field segmentation in chest radiographs from boundary maps by a structured edge detector, *JBHI* 22 (3) (2017) 842–851.
- [6] G. Aresta, C. Jacobs, T. Araújo, A. Cunha, I. Ramos, B. van Ginneken, A. Campilho, iW-Net: an automatic and minimalistic interactive lung nodule segmentation deep network, *Scientific reports* 9 (1) (2019) 1–9.

- [7] Z. Peng, X. Fang, P. Yan, H. Shan, T. Liu, X. Pei, G. Wang, B. Liu, M. K. Kalra, X. G. Xu, A method of rapid quantification of patient-specific organ doses for CT using deep-learning-based multi-organ segmentation and GPU-accelerated Monte Carlo dose computing, *Medical Physics* .
- [8] R. Adams, L. Bischof, Seeded region growing, *TPAMI* 16 (6) (1994) 641–647.
- [9] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, *IJCV* 1 (4) (1988) 321–331.
- [10] A. Mansoor, U. Bagci, B. Foster, Z. Xu, G. Z. Papadakis, L. R. Folio, J. K. Udupa, D. J. Mollura, Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends, *RadioGraphics* 35 (4) (2015) 1056–1076.
- [11] A. P. Harrison, Z. Xu, K. George, L. Lu, R. M. Summers, D. J. Mollura, Progressive and Multi-Path Holistically Nested Neural Networks for Pathological Lung Segmentation from CT Images, arXiv preprint arXiv:1706.03702 .
- [12] R. LaLonde, U. Bagci, Capsules for Object Segmentation, arXiv preprint arXiv:1804.04241 .
- [13] T. Zhao, D. Gao, J. Wang, Z. Tin, Lung segmentation in CT images using a fully convolutional neural network with multi-instance and conditional adversary loss, in: *ISBI, IEEE*, 2018.
- [14] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [15] S. Sun, C. Bauer, R. Beichel, Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach, *TMI* 31 (2) (2012) 449–460.
- [16] J. Tan, Y. Huo, Z. Liang, L. Li, Apply Convolutional Neural Network to Lung Nodule Detection: Recent Progress and Challenges, in: *ICSH*, Springer, 214–222, 2017.

- [17] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer, et al., Un-supervised deep learning applied to breast density segmentation and mammographic risk scoring, *TMI* 35 (5) (2016) 1322–1331.
- [18] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, *TMI* 35 (5) (2016) 1285–1298.
- [19] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: *CVPR*, 3431–3440, 2015.
- [20] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: *ICCV*, 1520–1528, 2015.
- [21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *MICCAI*, Springer, 234–241, 2015.
- [22] N. Baka, S. Leenstra, T. van Walsum, Ultrasound Aided Vertebral Level Localization for Lumbar Surgery, *TMI* 36 (10) (2017) 2138–2147.
- [23] S. Manivannan, W. Li, J. Zhang, E. Trucco, S. McKenna, Structure Prediction for Gland Segmentation with Hand-Crafted and Deep Convolutional Features, *TMI* .
- [24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: *NIPS*, 2672–2680, 2014.
- [25] P. Luc, C. Couprie, S. Chintala, J. Verbeek, Semantic segmentation using adversarial networks, *arXiv preprint arXiv:1611.08408* .
- [26] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, *arXiv preprint arXiv:1701.07875* .
- [27] M. Arjovsky, L. Bottou, Towards principled methods for training generative adversarial networks, *arXiv preprint arXiv:1701.04862* .
- [28] J. Tan, L. Jing, Y. Huo, O. Akin, Y. Tian, LGAN: Lung Segmentation in CT Scans Using Generative Adversarial Network, in: *INDIN*, IEEE, 2019.

- [29] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434 .
- [30] A. L. Maas, A. Y. Hannun, A. Y. Ng, Rectifier nonlinearities improve neural network acoustic models, in: ICML, vol. 30, 3, 2013.
- [31] S. Armato, 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.* 38 (2) (2011) 915–931.
- [32] D. Goldgof, et al., Data From QIN LUNG CT, TCIA .
- [33] H. Han, L. Li, F. Han, B. Song, W. Moore, Z. Liang, Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme, *JBHI* 19 (2) (2015) 648–659.
- [34] R. Shojaii, J. Alirezaie, P. Babyn, Automatic lung segmentation in CT images using watershed transform, in: *ICIP*, vol. 2, IEEE, II–1270, 2005.
- [35] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 .
- [36] R. T. Rockafellar, R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.
- [37] S. Nutanong, E. H. Jacox, H. Samet, An incremental Hausdorff distance calculation algorithm, *VLDB* 4 (8) (2011) 506–517.
- [38] S. Jégou, M. Drozdal, D. Vazquez, A. Romero, Y. Bengio, The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation, in: *CVPRW*, IEEE, 1175–1183, 2017.