# Pulmonary nodule segmentation with CT sample synthesis using adversarial networks

Yulei Qin, Hao Zheng, Xiaolin Huang, Jie Yang, Yue-Min Zhu

# Pulmonary nodule segmentation with CT sample synthesis using adversarial networks

Yulei Qin,[1, a)] Hao Zheng,[1] Xiaolin Huang,[1] Jie Yang,[1, b)] and Yue-Min Zhu[2]

[1)] *Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, 800 Dongchuan RD. Minhang District, Shanghai 200240, China*

[2)] *University Lyon, INSA Lyon, CNRS, INSERM, CREATIS UMR 5220, U1206, F-69621, France*

(Dated: 30 October 2018)

**Purpose:** Segmentation of pulmonary nodules is critical for the analysis of nodules and lung cancer diagnosis. We present a novel framework of segmentation for various types of nodules using convolutional neural networks (CNNs).

**Methods:** The proposed framework is composed of two major parts. The first part is to increase the variety of samples and build a more balanced dataset. A conditional generative adversarial network (cGAN) is employed to produce synthetic CT images. Semantic labels are generated to impart spatial contextual knowledge to the network. Nine attribute scoring labels are combined as well to preserve nodule features. To refine the realism of synthesized samples, reconstruction error loss is introduced into cGAN. The second part is to train a nodule segmentation network on the extended dataset. We build a 3D CNN model that exploits heterogeneous maps including edge maps and local binary pattern maps. The incorporation of these maps informs the model of texture patterns and boundary information of nodules, which assists high-level feature learning for segmentation. Residual unit, which learns to reduce residual error, is adopted to accelerate training and improve accuracy.

**Results:** Validation on LIDC-IDRI dataset demonstrates that the generated samples are realistic. The mean squared error and average cosine similarity between real and synthesized samples are $1.55 \times 10^{-2}$ and 0.9534, respectively. The Dice coefficient, positive predicted value, sensitivity, and accuracy are respectively $0.8483, 0.8895, 0.8511, 0.9904$ for the segmentation results.

**Conclusions:** The proposed 3D CNN segmentation framework, based on the use of synthesized samples and multiple maps with residual learning, achieves more accurate nodule segmentation compared to existing state-of-the-art methods. The proposed CT image synthesis method can not only output samples close to real images but also allow for stochastic variation in image diversity.

Keywords: pulmonary nodule segmentation, computer-aided diagnosis, generative adversarial networks, convolutional neural networks.

## I. INTRODUCTION

Pulmonary cancer has been one of the leading cancers in both men and women and annually causes 1.3 million deaths worldwide[1]. Although the overall 5-year survival rate is only 18%, if early diagnosis and treatment are put into effect timely, the patients' chances of survival can be greatly increased[2]. Pulmonary nodules are small masses in lung and often viewed as an early indication of cancer. The wide-spread use of computer tomography (CT) helps radiologists make accurate diagnosis of nodules. However, due to the high demand for CT scanning and similarity of nodules to lung tissue (e.g., blood vessels and bronchi), it may take radiologists long reading time to analyze suspicious lesions. Therefore, computer-aided diagnosis (CAD) systems are developed to improve doctors' reading efficiency.

Many current CAD systems focus on the detection of pulmonary nodules in CT[3–9]. These CAD systems process CT images and predict the coordinates of bounding boxes that contain suspicious nodules. However, bounding box alone is not sufficient. In clinical practice, radiologists need to measure volumetric changes of nodules to estimate their malignancy likelihood effectively[10–13], which requires manual delineation of nodules' boundaries. The pixel-level manual segmentation by radiologists is time-consuming since nodules differ in size (diameter ranging from 3 to 30 $mm$), shape, brightness, and compactness[6]. Therefore, it is imperative to develop CAD systems for accurate and robust nodule segmentation.

The main difficulty in nodule segmentation is to design an algorithm that adapts to both internal texture and external surroundings of pulmonary nodules. According to the variation in internal texture characteristics, lung nodules can be classified into the categories: solid, part-solid, and ground glass opacity (GGO). The solid nodules exhibit explicit shapes and margins while GGO nodules are of low contrast and have fuzzy boundaries. The part-solid nodules fall in between. Pulmonary nodules can also be classified into the categories: well-circumscribed, juxta-vascular, and juxta-pleural. The well-circumscribed nodules stay inside the lung alone. The juxta-vascular nodules and the juxta-pleural nodules connect vascular structures and pleural surfaces, respectively. Typical cases for each category are shown in Fig. 1.

In the past, several methods have been proposed to mainly segment on solid nodules[14–17]. Dehmeshki *et al.*[14] employed a 3D region growing method for user-interactive segmentation. Their method performs a sphericity-oriented contrast region growing on the fuzzy connec-

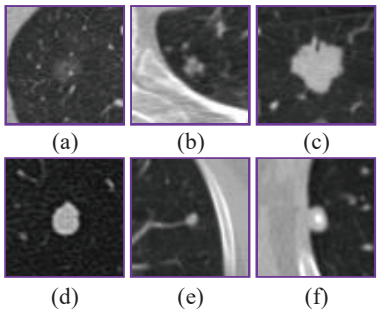


FIG. 1 Typical cases for each nodule type. First row: Nodules are classified by internal texture. (a) GGO; (b) part-solid; (c) solid. Second row: Nodules are classified by external surroundings. (d) well-circumscribed; (e) juxta-vascular; (f) juxta-pleural.

tivity map of the target object. It combines distance and intensity information as growing conditions. Diciotti *et al.*[15] developed an automated method to refine initial rough segmentation results of small juxta-vascular solid nodules. The rough segmentation is corrected by 3D local shape analysis, which removes vessel attachments with nodule boundaries preserved. GGO nodules are not considered in their work. Reeves *et al.*[16] designed an iterative method to separate a nodule from the pleural surface using plane fitting technique. Adaptive thresholding is then applied to adjust segmentation. Wang, Engelmann, and Li[17] proposed a segmentation method that transforms 3D volume of interest (VOI) into 2D images using a spiral-scanning technique. The optimal outlines of nodules in 2D images are delineated by dynamic programming method. Then, they are transformed back to 3D images for surface reconstruction.

Few methods were developed for segmentation of all solid, part-solid, and GGO nodules[18–20]. Kubota *et al.*[18] proposed a general segmentation method. It combines morphological operation and convexity models to segment on juxta-vascular and juxta-pleural nodules without separating lung walls. Qiang *et al.*[19] employed a scheme that utilizes freehand sketch analysis. Nodules are automatically segmented with an improved shape break-and-repair strategy. Mukhopadhyay[20] adopted a two-step segmentation method. It first categorizes nodules by internal texture. Then, vascular structures and pleural surfaces are removed. The method was evaluated on LIDC-IDRI public database[21].

With the development of convolutional neural networks (CNNs)[22–24], researchers tended to employ CNNs for segmentation in an end-to-end manner[25–27]. However, only one method is reported on adopting CNNs for nodule segmentation. Wu *et al.*[28] developed a 3D CNN

model for segmentation of pulmonary nodules from VOI. They evaluated the method on LIDC-IDRI dataset and achieved Dice coefficient of 0.7405.

Despite the fact that there exists an interest in designing CAD systems based on deep learning techniques, the performance of these systems is limited by the availability of large labeled datasets. Medical data are not easy to access due to privacy issues. In addition, it is laboursome for doctors to collect, organize, and annotate them, making the size of dataset restricted. Motivated by recent development of generative adversarial networks (GAN)[29–34], we believe synthetic image generation may be a good choice in the face of the underlying problem of imbalanced and limited data. In order to build a more balanced and diverse dataset, we capitalize on generating nodule CT images through adversarial networks, which is not considered in previously reported works.

In this paper, we propose a CNN-based framework for pulmonary nodule segmentation. By adopting adversarial networks, synthetic samples are generated to achieve a more balanced training dataset. With interpretable feature maps incorporated and residual learning strategy introduced, the segmentation model performs robustly on all kinds of nodules with-out radiologists' manual intervention. The main contributions are as follows: (1) We employ a conditional GAN that generates nodule CT images to extend the LIDC-IDRI dataset. Since original annotation is only the boundary of each nodule, we design a method to obtain ten-channel semantic labels of nodule patches. These labels not only contain contextual information but also represent nodules' semantic attributes. Based on semantic labels, synthetic samples are generated through adversarial networks. The $L2$ reconstruction error loss is introduced into cGAN to increase the realism of generated samples. The imbalanced data problem is alleviated by such expansion of dataset, which prevents overfitting for the training of segmentation model. Hence, the performance of our segmentation method gets improved. (2) We propose a 3D CNN model that accurately segments pulmonary nodules. To generate segmentation masks, a 3D U-Net[35] similar network is exploited. Multiple hetero- geneous maps, including edge maps and texture feature maps, are introduced as inputs and leveraged by the CNN model to learn high-level features. For edge maps, we apply Canny operator[36] and Sobel operator[37] to detect the edges of nodule images, which lay a foundation for the task of segmentation. Local binary patterns (LBP)[38] are chosen to capture spatial structure of nodules' textures. Since there exists a great difference in textures between solid, part-solid, and GGO nodules, these texture feature maps are considered informative

for the network to generate accurate segmentation results for each kind of nodule. The 3D architecture of our model aims at better utilizing volumetric knowledge of 3D CT images. Besides, residual learning is employed to resolve vanishing gradient problem. It promotes effective feature learning and accelerates training process. (3) The proposed CNN-based segmentation framework is evaluated on the public LIDC-IDRI dataset.

## II. MATERIALS

The public LIDC-IDRI dataset is used to generate synthetic nodule images and validate the proposed segmentation method. The dataset contains 1010 patients' CT scans. Each CT scan was reviewed by four experienced radiologists through a two-stage process: blinded and unblinded reading. In the blinded phase, each radiologist reviewed and marked each CT scan independently. In the unblinded phase, with three other radiologists' marks provided, each radiologist modified the original annotations to improve the quality of ground-truth labels. Nodules, of which the diameters are larger than 3 $mm$, are annotated with the boundaries and nine semantic attributes of subtlety, internal structure, margin, calcification, sphericity, lobulation, spiculation, texture, and malignancy. In our experiments, we exclude CT scans whose slice thickness is greater than 2.5 $mm$ in consideration of image quality. Hence, there are 888 CT scans with 1182 nodules in total. The distributions of these nodules are listed in Table I in terms of texture and size. For each nodule, the rating scores of its attributes are computed as the average of ratings from the four radiologists. The definition of attributes' scoring can be found in Table II. The scores of internal structure and calcification reflect corresponding classes while other feature scores represent sequential degrees. First, the internal area inside each nodule's boundary is filled to obtain its ground-truth label. For each CT slice, the Hounsfield unit (HU) is clipped in the range of [-1200 HU, 600 HU]. Then, all slices are normalized to $[0, 255]$ and resampled to the same spacing of $1 \times 1 \times 1$ $mm$. VOI cubes containing nodules are cropped from slices based on their coordinates and the cropped size is $64 \times 64 \times 64$ pixels.

TABLE I Distributions of the 1182 pulmonary nodules for experiments.

| | Category | No. of nodules |
|---|---|---|
| Texture | Solid | 927 |
| | Part-solid | 188 |
| | GGO | 67 |
| Diameter | $< 6\ mm$ | 38 |
| | $6 \sim 10\ mm$ | 424 |
| | $> 10\ mm$ | 720 |
| In total | | 1182 |

TABLE II Definition of scoring for each nodule attribute.

| Attribute | Scoring | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Subtlety | Extreme subtlety | | | | Obvious | \ |
| Internal Structure | Soft | Fluid | Fat | Air | \ | \ |
| Calcification | Popcorn | Luminated | Solid | Non-central | Central | Absent |
| Sphericity | Linear | | Ovoid | | Round | \ |
| Margin | Poorly defined | | | | Sharp | \ |
| Lobulation | None | | | | Marked | \ |
| Spiculation | None | | | | Marked | \ |
| Texture | GGO | | Part-solid | | Solid | \ |
| Malignancy | Highly unlikely | Moderately unlikely | Indeter-minate | Moderately suspicious | Highly suspicious | \ |

## III. METHODS

The developed pulmonary nodule segmentation framework is composed of two parts (Fig. 2): (1) Synthetic image generation and (2) 3D CNN-based segmentation. For the first part, adversarial networks are adopted to enhance the diversity of nodule samples and mitigate the problem of imbalanced and limited data. The second part is designed to segment all kinds of nodules from VOI using a 3D CNN model. The details of the proposed framework is presented as follows.

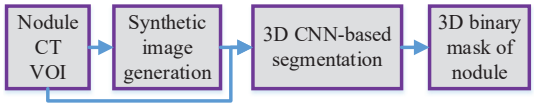


FIG. 2 An overview of the proposed pulmonary nodule segmentation framework. Synthetic nodule images are first generated. Then, both the original and synthesized images are used to train the segmentation model. The segmentation results are 3D binary masks of nodule VOI.

### III.A. Synthetic image generation

As can be observed in Table I, the LIDC dataset is imbalanced in terms of nodule's texture and size. The number of solid nodules is three times as many as that of the rest. Large nodules constitute a great proportion of all nodules. Besides, GGO nodules and small nodules are so limited in quantity that they are easily overwhelmed by other nodules. Con-sequently, the segmentation model may suffer poor performance on the minority categories of nodules if trained on such dataset. To tackle this problem, synthetic image generation then appears as an interesting solution. To do that, slices that contain nodules are first selected from all cropped VOI cubes. A technique of transforming ground-truth labels into ten-channel semantic labels is then designed to introduce abundant contextual information about nodules. Finally, a conditional generative model is employed to translate semantic labels into realistic images.

#### *III.A.1. Semantic label generation*

The ground-truth labels from LIDC-IDRI dataset only describe shape, size, and attributes of nodules. These labels are sufficient for the task of nodule segmentation, but not for sample synthesis. It is difficult for generative adversarial networks to produce authentic images if only information about nodules is provided. The semantic knowledge of nodules' surroundings is of great importance since it depicts external attachments and nodules' relative position in thoracic cavity. For example, two nodules may have similar boundaries but one is attached to pleural surface and the other stays alone. Hence, in order to enable the network to learn from nodules' contextual information, semantic labels are generated as stated in the following. First, slices are thresholded to extract all components with high intensity. The thresholding value is determined by the category of nodule's internal texture. For GGO nodules whose scoring of texture is lower than three, the grayscale value of 60 is chosen. For part-solid and solid nodules, 70 is set as threshold value. These two values are calculated based on the studies of nodule's density distribution in CT[39,40] and our clipping window of Hounsfield unit. Secondly, a disk-shape structuring element with radius of one is adopted for morphological opening. Each slice is opened to remove tiny objects and smooth image since only obvious parenchymal structures, including large vessels and pleural surfaces, are

considered for labeling. Then, connect component analysis is employed to differentiate between vascular structures and pleural surfaces. For each connected component, if its area is larger than 640 or if it intersects at least two borders of image with a minimum area of 32, it is labeled as pleural surface with a value of 3. Other components are labeled as vascular structures with a value of 2. The ground-truth label of nodule is set as 1. This generated label is not accurate enough to directly train a semantic segmentation model, but it provides adequate nodule's surrounding knowledge for image synthesis.

In addition, nine semantic attributes are introduced to describe nodules in more details. These features represent nodule's internal variation of intensity and shape, and are closely related to diagnosis. For each nodule, nine original binary ground-truth labels are multiplied with its nine attribute scorings respectively. Each label corresponds to one attribute scoring. Then, all nine attribute labels and one semantic label are concatenated together to form a ten-channel image as the input of cGAN. The input size is $10 \times 64 \times 64$ and the process of label generation is shown in Fig. 3.

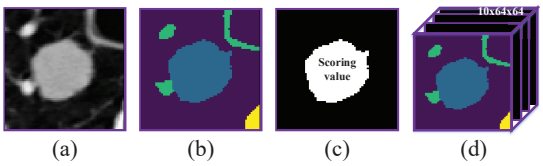


FIG. 3 The process of generating ten-channel labels. (a) CT image of nodule; (b) Generated semantic label containing a nodule, pleural surfaces, and vascular structures; (c) Each attribute's scoring value is multiplied with a binary ground-truth label; (d) The semantic label and nine attribute labels are concatenated as an input image with ten channels.

### *III.A.2. Conditional generative adversarial network*

The initially proposed GAN[29] learns to generate samples from the random noise vector.

The noise $z$ is passed explicitly into the generator as input. Different from the original GAN, the random noise $z$ of cGAN[30] is introduced into the generator during the process of generating samples. The cGAN maps $z$ to the realistic CT image $y$ in the conditional setting of a ten-channel semantic label $x$. The training of cGAN involves gaming between the generator model $G$ and the discriminator model $D$. The objective function of the original cGAN[30] is defined as:

$$\mathcal{L}_{cGAN} = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(y|x)] + \mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[\log(1 - D(G(z|x)|x))],$$
$$G, D = \arg\min_G \max_D \ \mathcal{L}_{cGAN}, \tag{1}$$

where $p_z$ and $p_{data}$ here denote the prior noise distribution and the real nodule data distribution, respectively. $G$ tries to capture the nodule images' distribution with the condition of label $x$ and its generated sample is $G(z|x)$. $D$ estimates the probability that the current pair is real nodule data pair $(x, y)$ rather than synthetic data pair $(x, G(z|x))$. $G$ is trained by minimizing such adversarial loss while $D$ by maximizing it. It is noted that $G$ is optimized to output images that are difficult for $D$ to distinguish from real ones. To directly guide $G$ to produce samples that are similar to realistic images, we introduce $L2$ reconstruction error loss to the training of generator as the following:

$$\mathcal{L}_G = \mathbb{E}_{x,y\sim p_{data}(x,y),z\sim p_z(z)}[\|y - G(z|x)\|_2^2], \tag{2}$$

where the real nodule data $y$ serve as the ground-truth for $G(z|x)$. Such $L2$ loss function penalizes the model to explicitly reduce the difference between real CT images and synthetic images. The modified objective function is given by:

$$G, D = \arg\min_G \max_D \ \mathcal{L}_{cGAN} + \lambda\mathcal{L}_G, \tag{3}$$

where $\lambda$ is a weight balancing these two terms. We set $\lambda = 100$ in the present study. The adversarial training process is illustrated in Fig. 4. Note that the noise here, to a certain degree, can be viewed as an implicit input.

The architecture of our cGAN is depicted in Fig. 5. The 2D U-Net structure is used as a backbone to build a generative model, which generates synthetic images in an encoder-decoder fashion with skip paths. For the contracting path, instead of max-pooling layer used in the original U-Net[26], strided convolution layer is adopted to downsample the image, followed by a batch normalization (BN) layer and a leaky rectified linear unit (ReLU) layer. For the expansive path, we employ transposed convolution to upsample feature maps

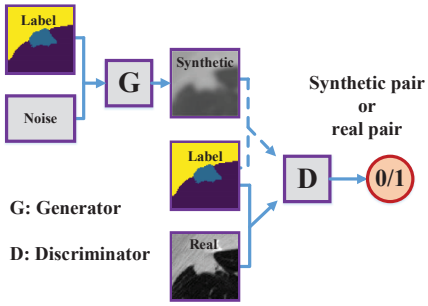


FIG. 4 The training of cGAN proceeds by alternatively training $G$ and $D$. Given a label image and a noise vector, $G$ is trained to obtain a realistic image. The synthetic pair and real pair refer to the ten-channel label concatenated with synthetic image and real image, respectively. $D$ learns to distinguish real pairs from synthetic fake pairs.

to increase resolution and concatenate them with features from skip path. The BN layer, ReLU layer and dropout layer are also introduced. Then, a fully convolutional net (FCN) is designed as a discriminator model. Except the first layer, all strided convolution layers are followed by a BN layer and a leaky ReLU layer. The pooling layer in both generator model and discriminator model is replaced by strided convolution because the latter learns to summarize the pixels within its kernel by a weighted element-wise multiplication. Different from max-pooling or avg-pooling, the way that strided convolution reduces feature dimensionality is not determined in advance but learnable during training.

As shown in Fig. 5, the noise $z$ is implicitly taken as input to the generator. We use dropout layer on the expansive path to introduce noise[33] by randomly deactivating neurons with a probability of 0.5. Previous study on dropout layer[41] proves that such layer adds noise to the output features and thus improves robustness to the variation of input images. Furthermore, the dropout layer provides regularization to prevent over-fitting by reducing co-dependency among neurons. It randomly deactivates neurons during the training process, thereby preventing the model from learning interdependent set of feature weights[42].

### III.B. Pulmonary nodule segmentation

The overall nodule segmentation architecture is given in Fig. 6. As pulmonary nodules have different internal textures and segmentation method should adapt to such variety, we introduce texture maps to implicitly impart to the network the ability of apprehending whether current nodule is GGO, part-solid, or solid. In addition, edge maps are concatenated

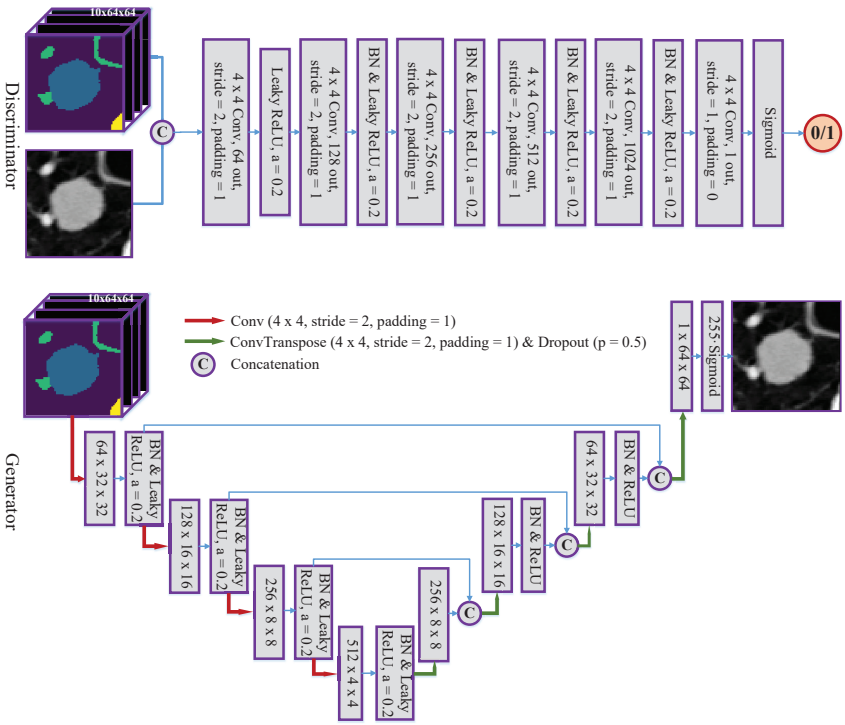


FIG. 5 The network architecture of the proposed cGAN.

as inputs since they provide rich knowledge about margins and boundaries of nodule images, thereby assisting the task of segmentation. The 3D CNN segmentation model is an end-to-end model that exploits a 3D U-Net[35] similar structure. Residual learning is brought into the network to improve the performance of segmentation.

### *III.B.1. Heterogeneous maps*

Local Binary Pattern (LBP)[38] characterizes the spatial structure of local image texture by encoding the difference between a center pixel and its neighboring pixels. We use LBP maps as the representation of nodule's texture to describe different types of nodules for the network. For each pixel in the original image, its LBP encoding is computed by thresholding neighboring pixels with its intensity:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^P, s(x) = \begin{cases} 1, \text{ if } x \geq 0 \\ 0, \text{ if } x < 0 \end{cases}, \tag{4}$$

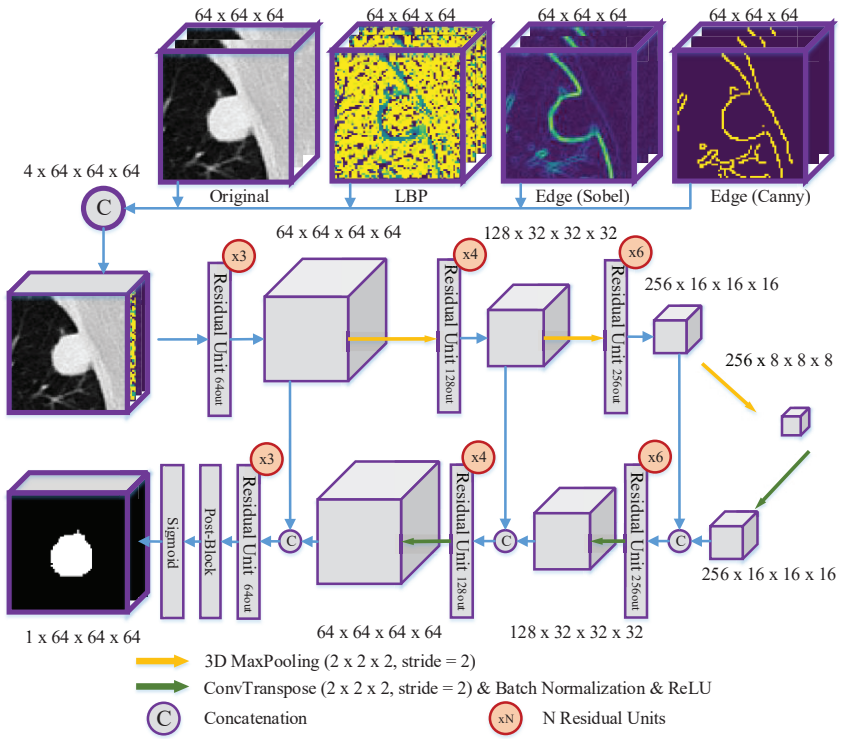


FIG. 6 The network architecture of the proposed segmentation framework.

where $g_c$ and $g_p$ are grayscale values of the center pixel and its surrounding pixels inside a circle with radius of $R$, respectively. The total number of neighboring pixels is $P$.

The LBP operator only considers the relative intensity of neighboring pixels with respect to the center pixel. Its value changes if rotation operation is implemented on the image. Since rotation is used for data augmentation, rotation-invariant LBP is preferred in order to extract essential characteristics of nodule's texture. Hence, we use the new type of LBP:

$$LBP^{ri}_{P,R} = \min\{ROR(LBP_{P,R}, i)|i = 0, 1, .., P-1\}, \tag{5}$$

where $ROR(x, i)$ performs a circular bit-wise right shift on the encoded value $x$, $i$ times. It can be viewed as texture feature detector to capture micro-features that are invariant to rotation. Furthermore, with rotation-invariant LBP texture maps fed into the network, the learned high-level CNN features are rotation-invariant as well. In the experiments, we set $P = 24$ and $R = 3$ and compute LBP maps slice by slice.

In the segmentation task, accurate detection of meaningful edges is fundamental. The edge map reflects the discontinuity of an image. Especially for the solid nodule that has a

clear margin, there exists an abrupt change in intensity around nodule's border. Edge maps are used to filter out useless information and only preserve structure properties of images. For well-circumscribed nodules, the edge maps directly detect their boundaries. For nodules that connect pleural surface or vascular structures, their edge maps also provide the outlines of their attachment. These maps can be viewed as initial segmentation results, which are then polished up by our network for final precise results.

There are many methods for edge detection. In our experiments, two most widely used methods, Sobel[37] and Canny[36] edge detectors, are employed together for each slice since their performance varies depending on the categories of nodules and the integrated use of both the methods is better than using one. For Sobel edge detection, two $3 \times 3$ kernels are convoluted with images to estimate the gradient in $x$ and $y$ directions. After convolution with horizontal and vertical kernels, two images of the approximated gradient of intensity are obtained as $G_x$ and $G_y$. Then, the magnitude of gradient is computed as edge map.

For Canny edge detection, we first smooth the image using a $3 \times 3$ Gaussian filter to reduce noise. Gaussian filter is adopted because it is faster than other non-linear filters such as Median filter. Then, horizontal and vertical Sobel operators are applied to compute the magnitude and orientation of the gradient. After that, non-maximum suppression is performed on the magnitude map to suppress all gradient values except local maxima. Finally, two thresholding values $t_1$ and $t_2$, determined respectively as 10% and 20% of the maximum magnitude's value, are applied to threshold the edge map. All pixels with magnitude value higher than $t_2$ are labeled as edges. Pixels with value higher than $t_1$, which are also 8-connected to the labeled edge pixels, are recursively labeled as edges.

### *III.B.2. Segmentation network*

The input of the network is a four-channel 3D cube, which consists of four different cubes: cropped CT volume cubes, LBP maps, and two edge maps. The full 3D CNN architecture is developed to exploit spatial contextual knowledge for high-level feature extraction. Residual units, which consist of a few stacked layers, are introduced into the network. Given the input $x$ of the residual unit, the underlying mapping to be fit by the layers is denoted as $H(x)$. Rather than directly approximating $H(x)$, these layers approximate a residual function $F(x) = H(x) - x$. By reducing such residual, it is easier to learn the underlying mapping.

This learning strategy is known as residual learning. Specifically, we define a residual unit as:

$$y = \mathcal{F}(x, \{W_i\}) + x, \tag{6}$$

where $x$ and $y$ are the input and output of residual unit, respectively. $\mathcal{F}(x, \{W_i\})$ is a 3D mapping to high-level features, which includes two convolution layers, two BN layers, and one ReLU layer. $\{W_i\}$ contains all learned parameters. Such residual unit allows gradient to propagate directly through a shortcut and thus avoids vanishing gradient problem. The introduction of residual learning benefits optimization process of deep network and improves the accuracy of segmentation. The schematic representation of residual unit is illustrated in Fig. 7.

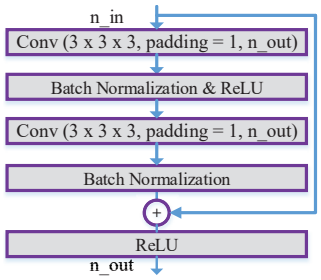


FIG. 7 Residual unit. n_in and n_out denote the number of channels of input cube and output cube, respectively.

For the contracting (forward) path, three blocks of residual units are adopted and each block is followed by a max-pooling layer to reduce the dimension of cube. The residual block contains multiple residual units and only the first unit increases the feature channel to the desired size. For the expansive (backward) path, we first use transposed convolution, BN, and ReLU to upsample cube. Secondly, we concatenate it with the corresponding shallow features that propagate via the skip path. Then, the concatenated features are fed into a residual block. At the end of the last residual block, a post block is attached in order to map the 64-channel feature cube to the size of $1 \times 64 \times 64 \times 64$. It is composed of two convolution layers and two BN layers as shown in Fig. 8. In total, the network has 57 convolution layers. For each voxel in the final cube, the probability of being nodule is calculated via a sigmoid function.

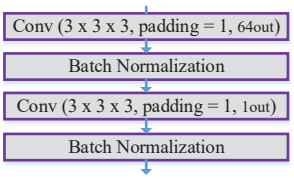


FIG. 8 Post-block.

The loss function of our segmentation network is based on Dice coefficient, which measures the similarity between segmentation results and ground-truth labels. Given two binary volumes $P$ and $T$, the Dice similarity coefficient (DSC) is defined as:

$$DSC = \frac{2\sum_i^N p_i t_i}{\sum_i^N p_i^2 + \sum_i^N t_i^2}, \quad p_i \in P, t_i \in T, \tag{7}$$

where $p_i$ and $t_i$ are voxels in the predicted segmentation result and ground-truth target, respectively. $N$ is the total number of voxels. The value of DSC ranges from 0 to 1 and if $P$ is exactly equivalent to $T$, the DSC achieves the maximum value of 1. In our implementation, the goal being to minimize the loss function, we define the Dice loss as:

$$\mathcal{L}_{seg} = 1 - \frac{2\sum_i^N p_i t_i + \epsilon}{\sum_i^N p_i^2 + \sum_i^N t_i^2 + \epsilon}, \quad p_i \in P, t_i \in T, \tag{8}$$

where $\epsilon$ is a smoothing coefficient that not only prevents division by zero but also avoids overfitting. We set $\epsilon = 1$ here in consideration of Laplace's rule of succession[43,44].

## IV. EXPERIMENTS AND RESULTS

### IV.A. Synthetic image generation

#### *IV.A.1. Experimental settings*

We use the cropped VOI cubes from LIDC dataset to train our cGAN. All slices containing nodules are chosen to generate their semantic labels and the total number of real CT pairs is 4694. Then, we split the dataset into ten subsets and perform ten-fold cross-validation. Each time, nine subsets are used for training. The remaining subset is left for validation, which

generates new synthetic images. Thus, a new synthetic dataset of 4694 slices is obtained.

The cGAN model is initialized from a Gaussian distribution $\mathcal{N}(0, 0.02)$ and optimized using Adam[45] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The initial learning rate for the first 200 epochs is set to 0.0002 and then decreases to 0 linearly after 200 epochs. The model is implemented in PyTorch[46] using 4 NVIDIA GTX 1080Ti GPUs.

### *IV.A.2. Evaluation metrics and results*

It is an open and difficult problem to find suitable metrics for evaluating the quality of synthesized images[33]. In the loss function of our cGAN model, $\mathcal{L}_G$ is explicitly optimized. Hence, it is reasonable and natural to use mean squared error (MSE) and cosine similarity ($S_C$) to evaluate our model. The two metrics are aimed at measuring the similarity between real nodule images and synthetic images. Given a trained generator $G$ and a set of ten-channel semantic labels $\{x^i | i = 1, 2, ..., m\}$, the metrics are defined as:

$$\begin{aligned} MSE &= \frac{1}{m}\sum_{i=1}^{m}\|y^i - G(z|x^i)\|_2^2, \\ S_C &= \frac{1}{m}\sum_{i=1}^{m}\frac{y^i \cdot G(z|x^i)}{\|y^i\|_2\|G(z|x^i)\|_2}, \end{aligned} \tag{9}$$

where $y^i$ and $G(z|x^i)$ here are the vectorized real nodule image and synthetic sample, respectively. The evaluation results inside different categories are given in Table III. The MSE and cosine similarity for all nodules are $1.55 \times 10^{-2}$ and 0.9534, respectively. The MSE of GGO nodules is $1.70 \times 10^{-2}$, which exceeds solid and part-solid nodules. The cosine similarity of solid nodules is higher than that of part-solid and GGO nodules. In terms of nodule's size, small nodules achieve the highest cosine similarity of 0.9556 and medium-sized nodules have the lowest MSE of $1.52 \times 10^{-2}$. All MSE and cosine similarity results are computed on 4694 nodule images of size $64 \times 64$.

Besides, visual examination of generated images is also employed to evaluate our cGAN model. Such evaluation metric is one of the most simple, intuitive yet effective methods to estimate sample's quality. Fig. 9 offers qualitative results of some generated samples. It shows that nodules and their surroundings are well reconstructed through our cGAN model.

TABLE III Quantitative results of synthetic image generation for different nodule categories.

| | Category | MSE ($\times 10^{-2}$) | $S_C$ |
|---|---|---|---|
| Texture | Solid | 1.55 | 0.9538 |
| | Part-solid | 1.47 | 0.9529 |
| | GGO | 1.70 | 0.9491 |
| Diameter | $<6\ mm$ | 1.65 | 0.9556 |
| | $6\sim10\ mm$ | 1.52 | 0.9524 |
| | $>10\ mm$ | 1.55 | 0.9539 |
| All nodules | | 1.55 | 0.9534 |

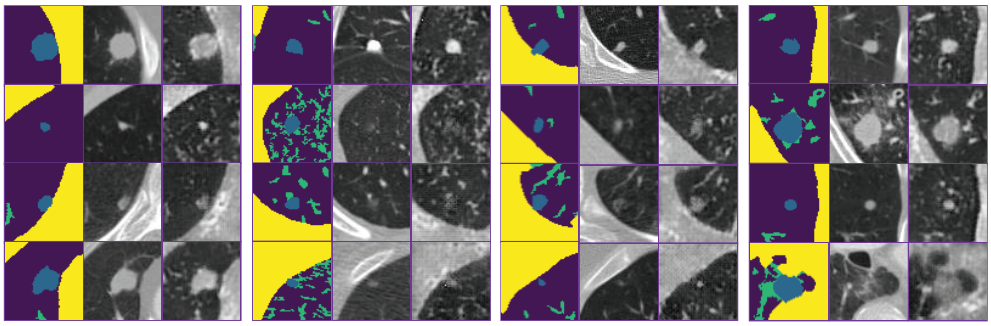


FIG. 9 Examples of generated synthetic images. (a) Input labels; (b) Real images; (c) Generated images. Out of simplicity, ten-channel inputs are briefly displayed as semantic labels.

To corroborate the effectiveness of nine attributes' labels in sample synthesis, we provide a comparison of samples generated with and without the nine attributes in Fig. 10. It shows that if only semantic labels are provided, the synthetic samples resemble real CT images with limited variety. In contrast, with additional nine attributes' labels incorporated as inputs, the cGAN can produce various images according to different configurations of attributes' scorings. By setting the value of texture as 1, 3, and 5, the output nodules indeed exhibit the characteristics of GGO, part-solid, and solid nodules, respectively. The 10-channel inputs (see Fig. 3) allow the cGAN to generate a large variety of nodule images that do not exist in the original LIDC-IDRI dataset, thereby enriching the training data greatly.

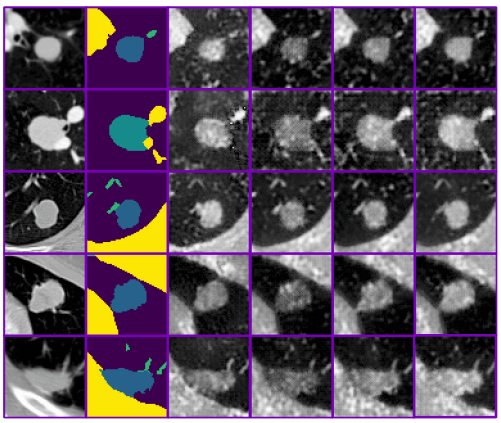


FIG. 10 Comparison of five synthetic samples generated with and without the nine attributes labels. (a) Real CT images; (b) Semantic labels; (c) Images generated without nine attributes.(d), (e), and (f) stand for the images generated with nine attributes and their texture scores are set to 1, 3, and 5, respectively.

## IV.B. Pulmonary nodule segmentation

### *IV.B.1. Experimental settings*

In segmentation experiments, we first replace the original slices in the 1182 nodule cubes with the generated slices to form new VOI. In total, 2364 nodule CT cubes are used, with half from the LIDC-IDRI dataset and half from our generated images. All cubes are of the same size: $64 \times 64 \times 64$. Ten-fold cross validation is adopted to evaluate our model. It should be noted that each time we use nine subsets of LIDC-IDRI dataset and their corresponding synthesized samples to train our model. Then we evaluate the model on the remaining one LIDC-IDRI subset. The validation set has no overlap with the training set.

The segmentation model is initialized from a Gaussian distribution $\mathcal{N}(0, 0.01)$ and trained using Adam[45] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for 150 epochs. The data augmentation method includes random rotation between $[0°, 180°]$, random flipping, and random axis swapping. The initial learning rate is set to 0.05 and decreases by half after every 20 epochs. The validation time for each nodule VOI is within 0.1 second and the segmentation model is also implemented in PyTorch.

### *IV.B.2. Evaluation metrics and results*

The performance of the proposed segmentation model is measured by four metrics: DSC, positive predicted value (PPV), sensitivity and accuracy. The DSC, defined in Eq. 7, is one of the most commonly used evaluation criteria. The PPV and sensitivity are respectively defined by:

$$\begin{aligned} PPV &= \frac{\sum_i^N p_i t_i}{\sum_i^N p_i^2}, \quad p_i \in P, t_i \in T, \\ Sensitivity &= \frac{\sum_i^N p_i t_i}{\sum_i^N t_i^2}, \quad p_i \in P, t_i \in T, \end{aligned} \tag{10}$$

where $P$ is the predicted result and $T$ is the ground-truth label. $N$ is the total number of voxels of VOI cubes. All numerators of DSC, PPV and sensitivity are the intersection voxels between $P$ and $T$. For DSC, its denominator is the average union voxels of $P$ and $T$ while for PPV and sensitivity, their denominators are the voxels predicted as positive for nodule region and true lesion voxels, respectively.

In addition, hard thresholding is applied on the probability map to obtain the binary segmentation result. Voxels having probability higher than 0.5 are considered as foreground objects. Then, accuracy is computed as:

$$\begin{aligned} Accuracy &= \frac{\sum_i^N \mathbb{1}(p_i == t_i)}{N}, \quad p_i \in P, t_i \in T, \\ \mathbb{1}(\text{statement}) &= \begin{cases} 1, \text{ if statement is True} \\ 0, \text{ otherwise} \end{cases}, \end{aligned} \tag{11}$$

where $\mathbb{1}(\cdot)$ is an indicator function.

Table IV summarizes the segmentation results in terms of four metrics. The average DSC, PPV, sensitivity and accuracy of all nodules are 0.8483, 0.8895, 0.8511, 0.9904, respectively. The performance of the proposed method on GGO nodules is the worst in terms of DSC, sensitivity, and accuracy. It is noted that nodules with larger diameter or solid texture have the highest segmentation scores in any evaluation metric.

The comparison of segmentation results with state-of-the-art methods[20,28,35] is given in

TABLE IV The segmentation results of the proposed model for different nodule categories.

| | Category | DSC | PPV | Sensitivity | Accuracy |
|---|---|---|---|---|---|
| Texture | Solid | 0.8605 | 0.8927 | 0.8681 | 0.9909 |
| | Part-solid | 0.8096 | 0.8755 | 0.8023 | 0.9891 |
| | GGO | 0.7865 | 0.8850 | 0.7515 | 0.9871 |
| Diameter | $< 6mm$ | 0.7776 | 0.8748 | 0.7719 | 0.9849 |
| | $6 \sim 10mm$ | 0.8382 | 0.8788 | 0.8494 | 0.9897 |
| | $> 10mm$ | 0.8578 | 0.8966 | 0.8560 | 0.9911 |
| All nodules | | 0.8483 | 0.8895 | 0.8511 | 0.9904 |

Table V. All the methods are evaluated on LIDC-IDRI dataset and the commonly used metric is Dice coefficient. Our model achieves the highest DSC score of **0.8483** and it outperforms existing methods. The traditional segmentation techniques by Mukhopadhyay[20] can not adapt to large variation of nodules such as size, shape and texture. Although both methods by Çiçek *et al.*[35] and Wu *et al.*[28] adopt 3D CNNs for segmentation task, our method surpasses them over 10% on average.

TABLE V Comparison of segmentation results in DSC.

| Approach | DSC |
|---|---|
| Mukhopadhyay[20] | 0.3900 |
| Çiçek *et al.*[35] | 0.7197 |
| Wu *et al.*[28] | 0.7405 |
| Proposed method | **0.8483** |

Table VI summarizes the results of quantitative comparison between different configurations having different inputs. A control group of four methods is constituted to evaluate our proposed method. Seg-NMaps refers to the proposed method without taking any map as input to the segmentation network. Seg-NEdge and Seg-NLBP denote the proposed method that does not use edge maps and LBP maps, respectively. For Seg-NSynthetic, generated synthetic samples are not added into the dataset to train our model. The Seg-NMaps method has the lowest DSC of 0.7993. Both LBP and edge maps contribute to better results, increasing DSC to 0.8176 and 0.8101, respectively. Without the extension of dataset, the Seg-NSynthetic method achieves the lowest accuracy of 0.9876. Except sensitivity, the proposed method enjoys the highest scores on other three metrics, which demonstrates the pertinence of each component of the proposed method. The accuracy of all methods is over 0.98.

More visually, qualitative results of different validation samples are shown in Fig. 11.

TABLE VI Quantitative comparison results of the control group.

| Approach | DSC | PPV | Sensitivity | Accuracy |
|---|---|---|---|---|
| Seg-NMaps | 0.7993 | 0.8523 | 0.8121 | 0.9881 |
| Seg-NEdge | 0.8176 | 0.8233 | **0.8610** | 0.9891 |
| Seg-NLBP | 0.8101 | 0.8559 | 0.8261 | 0.9890 |
| Seg-NSynthetic | 0.8104 | 0.8431 | 0.8596 | 0.9876 |
| Proposed method | **0.8483** | **0.8895** | 0.8511 | **0.9904** |

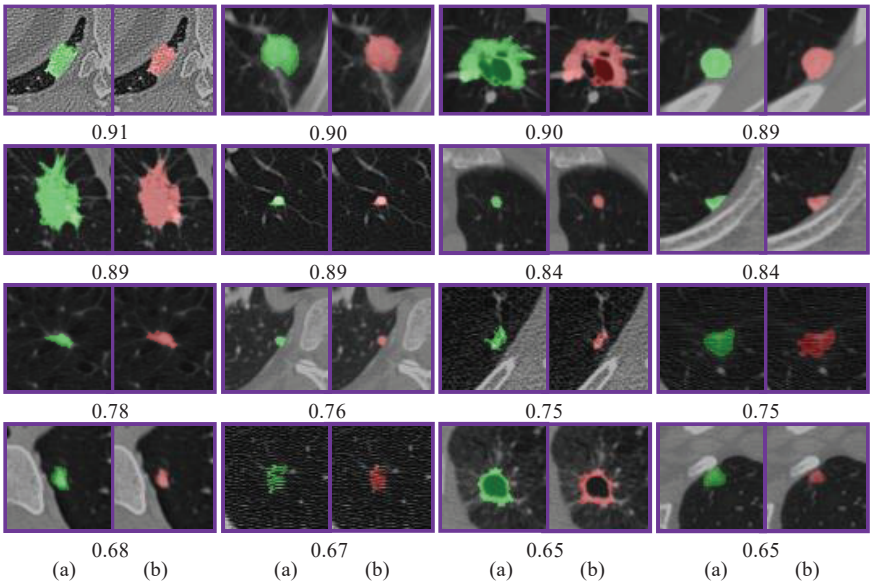


FIG. 11 Qualitative segmentation results of validation samples. (a) Ground-truth labels are in green; (b) Predicted nodules are in red. The score beneath each pair is Dice coefficient of the result. Central slice of each VOI cube is displayed for simplicity.

## V. DISCUSSION

We have shown that the proposed method can achieve accurate segmentation on pulmonary nodules. Its most distinctive characteristics include (1) the adoption of adversarial networks to promote samples' diversity for a more balanced training dataset and (2) the 3D segmentation network that takes advantage of interpretable heterogenous maps and residual learning. The results on synthetic image generation show that the cGAN simulates well the real nodule images to generate satisfactory samples and that each component of the segmentation network is instrumental in improving accuracy.

There are mainly two elements contributing to the realistic synthetic image generation. One is the preprocessing technique designed to obtain the ten-channel label that is rich in semantic information about nodule's attributes and surroundings. In conventional synthetic image generation[32–34], only ordinary geometry images are viewed as conditions of adversarial networks. Such type of images is in lack of depiction of context and characteristics of nodules. The other is the modification on the objective function of the original cGAN. The objective function defined in the original cGAN[30] only considers producing images that can deceive the discriminator, which is not sufficient in our medical setting. In contrast, the introduced $L2$ reconstruction error loss [Eq. (2)] explicitly minimizes the difference between generated images and real nodule images.

During the semantic label generation process, all nine channels were employed to represent nodules attributes. No selection or weighted combination of the nine attributes was conducted in advance because all these attributes are important for describing nodules. Each attribute is annotated and corrected meticulously by radiologists. If without nine channels, the diversity of synthetic images is substantially reduced. Given a one-channel label of a nodule such as a disk mask, the cGAN will produce a geometrically similar sample that only differs from the original data in grayscale value. While with the attributes provided, various nodule images can be generated by changing the rating scores of each attribute.

Although the patterns or styles of synthetic samples are kept similar to real ones, the generated images differ from the existing dataset in specific details. Firstly, in fact, given any artificial 10-channel semantic label, the cGAN can generate realistic CT samples that are missing in the original dataset. In accordance with different scorings of attribute labels, diverse types of synthetic images can be generated to improve the variety of training samples. Secondly, random noise is introduced into the generating process by dropout layer. Even with the same 10-channel semantic labels as inputs, the generated samples are different from their corresponding real CT images. Thirdly, tiny objects, such as small vessels and parenchymal structures, are removed in the generation process of semantic labels. Hence, conditioned on the resulting coarse-grained semantic labels, the synthetic images do possess a high level of variety.

The MSEs of solid and part-solid nodules are smaller than those of GGO nodules (Table III). This can be explained as follows. First, the boundaries of solid and part-solid nodules are clearer and their intensity varies abruptly, which is easier for cGAN to learn the dis-

crepancy between nodules and their surroundings. The second reason is that the internal distribution of GGO nodules is comparatively complex and scattered. The intensity inside GGO is relatively low and not as constant as that of solid nodules. As shown in Fig. 9, compared to real images, nodules on synthetic images tend to be more distinct because they are generated from labels which have sharp margins and specific borders. Due to the introduction of stochastic noise, the background of generated samples has more vascular-like structures than that of real nodule images.

Concerning the segmentation (Table IV), our method performs better on solid and part-solid nodules than on GGO nodules. This is because the boundaries of GGO nodules are fuzzier than other nodules, especially if there exist vascular structures in their vicinity. Furthermore, the number of GGO nodules in LIDC-IDRI dataset is far smaller than that of solid and part-solid nodules and thus the diversity and quality of generated samples are limited. Training on such dataset, the proposed model is difficult to capture strong feature representations for segmentation of GGO nodules. In terms of nodule's size, the larger the nodule is, the better the result is due to the fact that for larger nodules, it is easier to detect their position inside VOI and determine accurate margins.

Table V provided comparison with state-of-the-art methods[20,28,35]. The performance of the traditional method by Mukhopadhyay[20] is poor because it requires careful tuning of hyper-parameters (e.g., thresholding value of density for different nodules), which triggers off weak generalization ability on large dataset. Although Çiçek *et al.*[35] and Wu *et al.*[28] employed deep learning techniques as well, they did not regard the effect of multiple interpretable maps on conveying useful information (e.g. portrayal of nodule's texture by LBP maps and emphasis on nodule's border by edge maps) to the network. Besides, they did not take residual learning into consideration, which is crucial for developing a deep model. Examples in Fig. 11 demonstrate the performance on different kinds of nodules. Compared with well-circumscribed and solid nodules, juxta-vascular and GGO nodules are relatively harder to segment accurately due to their complex outer attachments and internal texture patterns, respectively. The results predicted by our model tend to provide conservative boundaries if the intensity drops sharply at margins. It may be because the inclusion of edge maps makes the model sensitive to borders. In Table VI, a possible reason that Seg-NEdge has the highest sensitivity is that without edge maps, the segmentation is not sensitive to the contours of nodules. It may tend to predict more pixels outside the contour as nodules

than true nodule pixels. According to Eq. (10), the sensitivity becomes high when the numerator increases. If neither the maps nor the synthetic data are used, the proposed framework degenerates back to a normal 3D CNN-based segmentation model, which differs from the existing 3D U-Net in two aspects: (1) the number of feature channels and (2) the introduction of residual learning strategy. Since in this case only real VOIs are fed into the 3D CNN without their features included, the segmentation performance is worse than the proposed framework.

There exist some limitations associated with the proposed framework. First, for the semantic labels in synthetic image generation, we only consider vessels and pleural surfaces and omit other structures such as bones and bronchi. To further improve the realism of generated samples, all structures would need to be labeled, which requires more complicated preprocessing techniques and parameter tuning. Since the quality of semantic labeling is heavily dependent on prior knowledge, it is challenging to develop a method of automatic labeling at an expert level. Second, it is difficult to find the optimal form of introducing random noises into cGAN. In the present study, dropout layer is applied as noise $z$ to generate stochastic output, which is consistent with Isola *et al.*[33]. It needs a different study to determine the impact of noise on the generated samples. Third, the distribution of training dataset is still not even. Although we extend LIDC-IDRI dataset via the cGAN model, the quantity and diversity of some nodules (e.g., GGO and juxta-vascular nodules) are still in shortage. Future work may include designing new schemes to solve imbalanced dataset problem. Finally, it is noted that the segmentation labels of nodules are obtained from radiologists in LIDC. However, the annotation process is decided by each radiologist's subjective judgment[19–21], leading to different ground-truth labels. Hence, the performance of the proposed method may be affected by such variation.

## VI. CONCLUSION

We have proposed a two-part CNN-based framework for pulmonary nodule segmentation. In the first part, adversarial networks are employed to synthesize nodule samples. It targets at building a more diverse and balanced dataset for the subsequent model training. Semantic labels, together with nine attribute scoring labels, are exploited to provide semantic and contextual knowledge. Reconstruction error loss is introduced to improve re-

alism. Such method of extending dataset presents several advantages. The boundaries and semantic attributes of nodules are preserved during generation process. Moreover, the random noise produced by dropout layer allows for the variation of spatial surroundings and thus boosts image diversity. In the second part, multiple feature maps are incorporated as inputs into the 3D CNN model. With residual learning strategy, the segmentation model trained on the extended dataset enjoys a high level of generality. The results on LIDC-IDRI dataset demonstrate that our 3D CNN model achieves more accurate nodule segmentation compared to existing state-of-the-art methods, which suggests its potential value for clinical applications.

## ACKNOWLEDGMENTS

This study was partly supported by National Natural Science Foundation of China (NSFC, No.61603248, No.6151101179, and No.61572315), 973 Plan of China (No.2015CB856004), Committee of Science and Technology, Shanghai, China (No.17JC1403000), and the Region Auvergne Rhne-Alpes of France under the project CMIRA COOPERA/EXPLORA PRO 2016.

The authors are grateful to the anonymous reviewers for their helpful comments.

## REFERENCES

[a)]Electronic mail: qinyulei@sjtu.edu.cn

[b)]Author to whom correspondence should be addressed. Electronic mail: jieyang@sjtu.edu.cn

[1]L. A. Torre, R. L. Siegel, and A. Jemal, "Lung cancer statistics," in *Lung Cancer and Personalized Medicine* (Springer, 2016) pp. 1–19.

[2]R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," CA: A Cancer Journal for Clinicians **66**, 7–30 (2016).

[3]T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," Medical Image Analysis **14**, 390–406 (2010).

[4]E. Lopez Torres, E. Fiorina, F. Pennazio, C. Peroni, M. Saletta, N. Camarlinghi, M. Fantacci, and P. Cerello, "Large scale validation of the M5L lung CAD on heterogeneous CT

datasets," Medical Physics **42**, 1477–1489 (2015).

[5]C. Jacobs, E. M. van Rikxoort, T. Twellmann, E. T. Scholten, P. A. de Jong, J.-M. Kuhnigk, M. Oudkerk, H. J. de Koning, M. Prokop, C. Schaefer-Prokop, *et al.*, "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," Medical Image Analysis **18**, 374–384 (2014).

[6]A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken, "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," IEEE Transactions on Medical Imaging **35**, 1160–1169 (2016).

[7]M. Sakamoto and H. Nakano, "Cascaded neural networks with selective classifiers and its evaluation using lung X-ray CT images," arXiv preprint arXiv:1611.07136 (2016).

[8]Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-d CNNs for false positive reduction in pulmonary nodule detection," IEEE Transactions on Biomedical Engineering **64**, 1558–1567 (2017).

[9]X. Huang, J. Shan, and V. Vaidya, "Lung nodule detection in ct using 3d convolutional neural networks," in *Proceedings of the Biomedical Imaging, 2017 IEEE 14th International Symposium on* (2017) pp. 379–383.

[10]D. F. Yankelevitz, A. P. Reeves, W. J. Kostis, B. Zhao, and C. I. Henschke, "Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation," Radiology **217**, 251–256 (2000).

[11]B. de Hoop, B. van Ginneken, H. Gietema, and M. Prokop, "Pulmonary perifissural nodules on CT scans: rapid growth is not a predictor of malignancy," Radiology **265**, 611–616 (2012).

[12]D. O. Wilson, A. Ryan, C. Fuhrman, M. Schuchert, S. Shapiro, J. M. Siegfried, and J. Weissfeld, "Doubling times and CT screen–detected lung cancers in the pittsburgh lung screening study," American Journal of Respiratory and Critical Care Medicine **185**, 85–89 (2012).

[13]L. R. Goodman, M. Gulsun, L. Washington, P. G. Nagy, and K. L. Piacsek, "Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements," American Journal of Roentgenology **186**, 989–994 (2006).

[14]J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach," IEEE Transactions on Medical Imaging

**27**, 467–480 (2008).

[15]S. Diciotti, S. Lombardo, M. Falchini, G. Picozzi, and M. Mascalchi, "Automated segmentation refinement of small lung nodules in CT scans by local shape analysis," IEEE Transactions on Biomedical Engineering **58**, 3418–3428 (2011).

[16]A. P. Reeves, A. B. Chan, D. F. Yankelevitz, C. I. Henschke, B. Kressler, and W. J. Kostis, "On measuring the change in size of pulmonary nodules," IEEE Transactions on Medical Imaging **25**, 435–450 (2006).

[17]J. Wang, R. Engelmann, and Q. Li, "Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique," Medical Physics **34**, 4678–4689 (2007).

[18]T. Kubota, A. K. Jerebko, M. Dewan, M. Salganicoff, and A. Krishnan, "Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models," Medical Image Analysis **15**, 133–154 (2011).

[19]Y. Qiang, Q. Wang, G. Xu, H. Ma, L. Deng, L. Zhang, J. Pu, and Y. Guo, "Computerized segmentation of pulmonary nodules depicted in CT examinations using freehand sketches," Medical Physics **41**, 041917 (2014).

[20]S. Mukhopadhyay, "A segmentation framework of pulmonary nodules in lung CT images," Journal of Digital Imaging **29**, 86–103 (2016).

[21]S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, *et al.*, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," Medical Physics **38**, 915–931 (2011).

[22]Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE **86**, 2278–2324 (1998).

[23]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM **60**, 84–90 (2017).

[24]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.

[25]J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015) pp. 3431–3440.

[26]O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (2015) pp. 234–241.

[27]F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings of the 2016 Fourth International Conference on 3D Vision* (2016) pp. 565–571.

[28]B. Wu, Z. Zhou, J. Wang, and Y. Wang, "Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction," arXiv preprint arXiv:1802.03584 (2018).

[29]I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the Advances in Neural Information Processing Systems* (2014) pp. 2672–2680.

[30]M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784 (2014).

[31]A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434 (2015).

[32]A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 3 (2017) p. 6.

[33]P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," arXiv preprint arXiv: 1611.07004 (2017).

[34]J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," arXiv preprint arXiv:1709.01872 (2017).

[35]Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (2016) pp. 424–432.

[36]J. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence **8**, 679–698 (1986).

[37]I. Sobel, "An isotropic 3× 3 image gradient operator," Machine Vision for Three-Dimensional Scenes , 376–379 (1990).

[38]T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern

Analysis and Machine Intelligence **24**, 971–987 (2002).

[39]H.-U. Kauczor, K. Heitmann, C. P. Heussel, D. Marwede, T. Uthmann, and M. Thelen, "Automatic detection and quantification of ground-glass opacities on high-resolution CT using multiple neural networks: comparison with a density mask," American Journal of Roentgenology **175**, 1329–1334 (2000).

[40]B. Zhao, G. Gamsu, M. S. Ginsberg, L. Jiang, and L. H. Schwartz, "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm," Journal of Applied Clinical Medical Physics **4**, 248–260 (2003).

[41]S. Park and N. Kwak, "Analysis on the dropout effect in convolutional neural networks," in *Proceedings of the Asian Conference on Computer Vision* (2016) pp. 189–204.

[42]I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, Vol. 1 (Cambridge: MIT press, 2016).

[43]D. Jurafsky and J. H. Martin, *Speech and language processing*, Vol. 3 (London: Pearson, 2014).

[44]S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach* (Malaysia: Pearson Education Limited, 2016).

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

[46]A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proceedings of the Advances in Neural Information Processing Systems-Workshop* (2017).