Received: 8 June 2020 | Revised: 27 September 2020 | Accepted: 1 October 2020

DOI: 10.1002/ima.22511

WILEY

RESEARCH ARTICLE

3D conditional generative adversarial network-based synthetic medical image augmentation for lung nodule detection

Tian Bu | Zhiyong Yang | Shan Jiang | Guobin Zhang | Hongyun Zhang | Lin Wei

School of Mechanical Engineering, Tianjin University, Tianjin, China

Correspondence
Shan Jiang, School of Mechanical Engineering, Tianjin University, 135 Yaguan Road, Jinnan District, Tianjin 300350, China.
Email: shanjmri@tju.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 51775368, 51811530310, 81871457

Abstract

A computer-aided detection (CADe) scheme, relying on a large number of high-quality images with annotations, could help radiologists effectively detect lung nodules. However, such medical data are generally difficult to obtain. To address this issue, this paper proposes a novel method based on a conditional generative adversarial network (CGAN) to generate new samples for data augmentation (DA). This method employs a 3D CGAN based on a squeeze-and-excitation mechanism and residual learning to synthesize realistic and diverse lung nodules in chest computed tomography (CT) images to improve the performance of the CADe system. To evaluate the proposed method, this paper uses synthetic samples for DA to train the lung nodule detection network. The results indicate that these synthetic samples, which cover a part of the data distribution unfilled by the original data, can boost the overall performance of the nodule detection network at fixed false positive (FP) rates.

KEYWORDS

3D squeeze-and-excitation network, computer-aided detection, conditional generative adversarial network, data augmentation, lung nodules

1 | INTRODUCTION

Lung cancer is the leading cause of cancer death in America, with a high incidence and low survival rate (24% for men and 23% for women).[1] The 5-year relative survival rate for patients is less than 19%.[2] More than 60% of patients are diagnosed with advanced cancer without early detection.[3,4] Early diagnosis and effective treatment of lung cancer are critical for patients.[5] According to the National Lung Cancer Screening Trial, the early identification of lung cancer by screening with computed tomography (CT) can significantly reduce mortality by approximately 20%.[6] Most lung cancers originate from malignant lung nodules, which can be detected in chest CT scans. The conventional method, detecting lung nodules by reading chest CT scans, is time consuming and can be prone to artificial mistakes.[7] To avoid these problems, computer-aided detection (CADe) systems have been developed, which can help radiologists effectively detect lung nodules and improve their identification accuracy in reading lung CT scans. Conventional CADe systems rely on handcrafted visual features, which contain only part of the lung nodule features. Compared to trained radiologists, the advantages of conventional CADe systems are not obvious.[8] In recent years, with the development of deep learning techniques, various CADe systems based on deep convolutional neural networks (DCNNs) have appeared and achieved great success in the field of medical image analysis.[9] CADe systems based on DCNNs can automatically learn highly discriminative

features from images and are more accurate and efficient than handcrafted feature-based methods.[10] The performance of deep learning-based methods relies on the quality and quantity of the image data. Currently, an important challenge of CADe systems based on DCNNs is the limited quantity and variety of medical images. To solve the issue of shortages of image data, this paper proposes a data augmentation (DA) method that can generate lung nodule images and apply the generated images to enrich the training data of lung nodule images.

Conventional DA techniques mostly adopt geometric or intensity transformation methods, such as random cropping, flipping, rotating, and zooming. Although conventional DA techniques can introduce a certain level of diversity to medical image data, these new images' appearances are essential, similar to the original data, and cannot introduce sufficient fresh features. Therefore, conventional DA methods that improve CADe systems are limited.[11] Recently, generative adversarial networks (GANs),[12] proposed by Goodfellow, have made game-changing strides in the field of image analysis. GANs can learn abundant discriminatory features from the original images and generate various and realistic images.[13] Because of the great performance of GANs in the medical field, an increasing number of researchers have used GAN models to resolve the chronic problem of insufficient labeled data.[14] Maria et al.[15] used a GAN framework (DC-GAN) to generate samples for benign, malignant and a mixture of both tumor types in 2D image slices, aiming to assist the training of radiologists. Mirsky et al.[16] proposed a 3D framework (CT-GAN) based on deep learning to add or remove lung nodules in CT scans and introduce the possibility of an attacker modifying medical images. Jin et al.[17] developed a 3D GAN method to synthesize lung nodules, the results of which indicate that DA with synthetic lung nodules can improve the robustness of the pathological lung segmentation model. There are also many works based on GANs to synthesize lung nodules and improve the performance of DCNN models in nodule classification tasks.[18-21] Liu et al.[22] developed a method for generating lung nodules in chest CT scans with manipulable shapes and textures. Han et al.[23] proposed the 3D multiconditional GAN (MCGAN) to generate realistic and diverse samples to boost accuracy in lung nodule detection. These papers show that mixing real data and generated samples in the training stage can improve lung nodule detection performance. However, to the best of our knowledge, the efficacy of promising methods such as SE-ResNet on GAN-based methods for DA has not been widely explored.

This paper proposes a novel adversarial learning scheme based on a conditional GAN (CGAN)[24] for medical DA. The proposed generator network uses a U-Net structure and can generate 3D lung nodules of multiple scales, with diverse appearances, in different locations in chest CT scans. To enhance the quality of synthetic nodules, this paper introduces a squeeze-and-excitation residual neural network (SE-ResNet) module[25,26] to both the generator network and the discriminator network, which can effectively strengthen the hierarchical feature learning ability and relieve the problem of training difficulties of GANs. By using the proposed framework, we synthesized abundant and various lung nodules based on the Lung Image Database Consortium (LIDC) dataset[27] and used these samples to create a new training dataset. By utilizing the GAN dataset to train a nodule detection network, we show that the proposed methods can be used to generate training samples when medical image data are scarce and achieve acceptable performance. To evaluate the capacity of using the synthetic samples to improve the detection ability of the network, we combined the GAN dataset with the LIDC dataset at a proper ratio as an augmentation dataset to train the nodule detection network. The results imply that the application of the proposed method as a DA technique can effectively improve the performance of the nodule detection network, which achieves a higher sensitivity and competition performance metric (CPM) scores at random FP rates.

2 | METHODS

2.1 | Dataset

This research uses the publicly available chest CT scans dataset, the LUng Nodule Analysis 2016 (LUNA16) dataset,[28] extracted from the LIDC dataset. The LIDC dataset, containing 1018 clinical chest CT scans, is a standard lung nodule database collected by the National Cancer Institute in America. Each CT scan was annotated by four professional thoracic radiologists during a two-phase review process. The LUNA16 dataset includes 888 CT scans with slice thickness ≤ 2.5 mm. In total, there are 1186 nodules with a diameter ≥ 3 mm accepted by at least three out of four radiologists. The LUNA16 dataset is equally divided into 10 subsets for 10-fold cross-validation.

2.2 | SE-ResNet module

The network degradation in DCNNs restricts the growth of the network depth and the network performance. These problems result in difficulties with deep network optimization. To mitigate the above problems, He et al.[25]

proposed the residual learning concept: a major breakthrough in the deep learning field. The key to residual learning is identity shortcut connections, which is an effective approach to intensify information circulation in deep networks. Applying the residual learning method to the proposed GAN can efficiently solve the instability problem during GAN training.

The SE operation proposed by Hu et al.[26] is different from past works in that it tries to optimize the network spatial dependencies. The SE block focuses on channel-wise information and incorporates a channel attention mechanism into the framework of the network. The SE modules implement channel-wise feature recalibration by specifically modeling the interdependencies between channels. Through this mechanism, the network is able to automatically learn the channel-wise significance level to selectively stress beneficial channel features and suppress less beneficial ones. The SE module improves the performance of the network by enhancing the representational capacity of the channel features. Existing work has shown that integrating the SE modules into DCNNs for image segmentation can produce promising improvements.[29,30]

The SE-ResNet module is a processing unit in a CNN that integrates the advantages of residual learning for feature reuse and the SE operation for feature recalibration. It can be formulated as

$$X^r = F_r(X)$$

$$X^r = \left[X_1^r, X_2^r, ..., X_c^r\right] \tag{1}$$

where X is the input feature. F_r denotes the residual mapping. X_c^r is the cth channel feature map of residual feature X^r.

$$z_c = F_{sq}\left(X_c^r\right) = \frac{1}{L \times H \times W} \sum_{l=1}^{L} \sum_{h=1}^{H} \sum_{w=1}^{W} x_c^r(l,h,w) \tag{2}$$

The function F_{sq} is the squeeze function, producing the z_c statistic of each channel by global average pooling. $L \times H \times W$ are the spatial dimensions of the residual feature.

$$s = F_e(z, W) = \sigma(W_2 \delta(W_1 z)) \tag{3}$$

where z is the set that consists of z_c. F_e represents the excitation function, which calculates the channel-wise weight $s \in R^C$. The excitation function block is composed of two fully connected layers with parameters W_1 and W_2 and two nonlinear activation functions. σ is the sigmoid function, and δ is the rectified linear unit (ReLU) function or LeakyReLU function in different parts.

$$\tilde{X}_c^r = F_{sc}\left(X_c^r, s_c\right) = s_c X_c^r \tag{4}$$

The function F_{sc} refers to channel-wise multiplication between the weight s_c and the feature map $X_c^r \in \mathbb{R}^{L \times H \times W}$, recalibrating the channel response of the residual feature to obtain $\tilde{X}^r$.

$$Y = \delta\left(\tilde{X}^r + X\right) \tag{5}$$

where the output feature Y is obtained after the activation δ. The calculation of $\tilde{X}^r + X$ is accomplished by elementwise addition.

The 3D SE-ResNet block is illustrated in Figure 1.

2.3 | Generative adversarial networks

A GAN[12] is a specific image processing framework based on deep learning, whose target is to synthesize new image samples externally similar to real original data. The most common GAN framework is composed of two neural network models: a generator G and a discriminator D. The generator G aims to generate samples to fool the discriminator D, which is focused on distinguishing real samples from the generated sample. Specifically, generator G samples a random noise vector z from a prior data distribution, such as a Gaussian distribution, and synthesizes a new sample $G(z)$, which is expected to

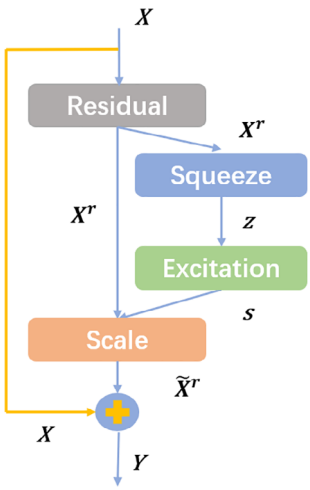

FIGURE 1 3D squeeze-and-excitation residual neural network module [Color figure can be viewed at wileyonlinelibrary.com]

approximate the underlying unknown real data distribution. The input of discriminator D is sample x, and the output obtained by function D is a probability that sample x is real or synthetic. Both the discriminator and generator are adversarially trained against one another in a minimax game while their capabilities are constantly being improved. The loss function for the two-player minimax game is defined as

$$L_{\mathrm{GAN}} = \min_G \max_D \mathbb{E}_x[\log D(x)] + \mathbb{E}_z[\log(1 - D(G(z)))] \quad (6)$$

A GAN model can be extended to a CGAN[24] model with additional information y, such as class labels and attribute information. This conditional information is used in both the generator and discriminator as additional input or total input.[31] The aim of this study is to synthesize pulmonary nodules with realistic and various appearances at random positions, which are harmonious with the surrounding context. To allow lung nodules to naturally blend with other tissues, this work uses a CGAN framework based on the surrounding information. The surrounding information can assist the model in generating and discriminating samples better than random noise z. Similar to the unconditional GAN, the generator and discriminator are playing a zero-sum game where the generator is attempting to generate better images to fool the discriminator, while the discriminator is attempting to distinguish all the spurious images. The only difference between the two models is that the inputs are random noise or surrounding information. The loss function for the CGAN can be defined as

$$L_{\mathrm{CGAN}} = \min_G \max_D \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_y[\log(1 - D(G(y),y))] \quad (7)$$

2.4 | Network architecture

The generator G is a 3D CNN based on the U-Net structure, which can synthesize realistic and various lung nodules at the desired size. Due to the 3D attributes of CT scans, a 3D CNN can better extract dimensional contextual information of lung nodules well and obtain higher quality features than other networks. The U-Net architecture with shortcut connections effectively combines multiple features and increases the training stability. The generator is composed of an encoder and the decoder, which have a spatial mapping relationship. The encoder takes conditional information as input and generates an underlying feature expression, which is transmitted to the decoder and eventually becomes a 3D lung nodule with surrounding content, as shown in Figure 2. The input of the encoder is a small, 3D volume of interest (VOI) cropped from lung CT scans with a missing region in the center, whose size is $32 \times 32 \times 32$ (height $\times$ length $\times$ width). There are five fractionally strided convolutional layers to downsample the image with a $4 \times 4 \times 4$ kernel size in the encoding path, which are interleaved with four 3D SE-ResNet modules. In the decoding path, deconvolutional layers are used to upsample the feature map. After each upsampling operation, the images are incorporated with the matching layers from the encoding path. To prevent overfitting, dropout with a probability of 0.5 is used on the first four deconvolutional layers in the decoding path. To stabilize the training process, we apply batch normalization on each convolutional and deconvolutional layer of the network, except for the input and output layers. In addition to the output layer using a tanh activation function, all layers use ReLU activation functions in the decoding path, and the layers of the encoder use LeakyReLU activation functions.

The discriminator D is a typical CNN structure, as shown in Figure 2. The input of the discriminator is a 3D patch of size $32 \times 32 \times 32$, which is the combination of a surrounding context and a patch centering on a fake or real nodule. The discriminator network is composed of four convolution layers and three 3D SE-ResNet modules, which is similar to the encoding path of the generator. We also apply dropout and batch normalization operations on the three middle convolution layers to inject randomness and avoid network collapse to a certain extent. We used strided convolutions instead of pooling operations to decrease the spatial dimensionality. The LeakyReLU activation function is used on the first layers, and the last output layer uses the sigmoid activation function.

3 | EXPERIMENTS AND RESULTS

3.1 | Implementation details

In the stage of model training, we use the Adam optimizer with learning rates of 0.0002 and 0.0001 for the generator and the discriminator, respectively. The model was trained for 100 epochs with a batch size of 32 samples, in which the generator and the discriminator are trained alternately. This model is trained on a computer with four NVIDIA TITAN Xp 12 GB graphics processing units (GPUs). Prior to the training stage of the lung nodule detection model, we perform a preprocessing operation on the CT scans to obtain the lung segmentation

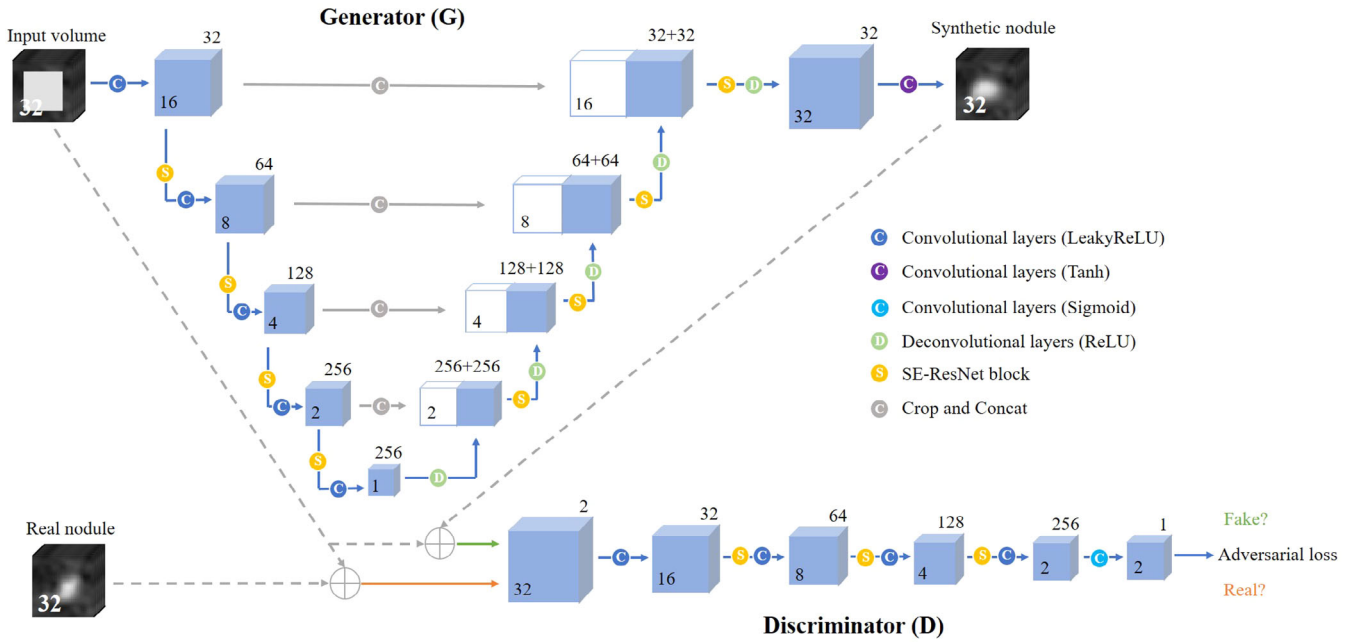

FIGURE 2 3D conditional generative adversarial network architecture for lung nodule generation. The number inside the cube represents the spatial size, and the number outside the cube represents the number of channels [Color figure can be viewed at wileyonlinelibrary.com]

images. During training, we adopt stochastic gradient descent (SGD) with a momentum of 0.9. This model was trained for 100 epochs for each fold with a multitask learning loss function. The initial learning rate was 0.01, which was reduced to 0.001 and 0.0001 during the training process. This model is also trained on a computer with four NVIDIA TITAN Xp 12 GB GPUs.

3.2 | Proposed method-based lung nodule generation

To train the proposed network model, we choose a training set of nodules from the LUNA16 dataset with approximate diameters between 5 mm and 30 mm, which includes 916 lung nodules in total. Based on experience, training GAN on the data obtained by scaling the nodule image with half of the patch side length as the average nodule diameter can produce better results.[17,23] We extracted the nodule patches from the CT scans and scaled them based on the ratio between 16 mm and the diameter of annotation. Then, we resample these patches with $32 \times 32 \times 32$ size with a resolution of $1 \times 1 \times 1$ mm. After that, we apply histogram equalization to the cropped images to increase the contrast, and we apply normalization to improve the regularization ability of the model. The image preprocessing operation can help the model learn the features better and improve the performance of the model. After training, the discriminator is abandoned, and the generator is used to synthesize the expected samples. To further evaluate the performance of the proposed method, we also implement a baseline method without 3D SE-ResNet blocks, which is trained in the same way as above. The generated images of the proposed method and baseline method are illustrated in Figure 3. It can be seen in these figures that the images generated by the proposed network have higher visual quality than the images generated by the baseline method. Some lung nodules in CT images generated by the proposed method are presented in Figure 4. All patch slices from a sample are shown in Figure 5.

3.3 | Improving lung nodule detection

After the proposed method was trained, we further explored whether our work benefits lung nodule detection. For this purpose, we implement a lung nodule detection model. This network is a 3D region proposal network (RPN) model, which is optimized with 3D SE-ResNet blocks based on the nodule detection network.[32] To improve the precision of the detection results, this model applies positive sample balancing and hard negative mining operations. Prior to training, we need to use the proposed generator to perform DA. We sample the 3D patches centered on annotated nodules from the LUNA16 dataset and erase the central region containing

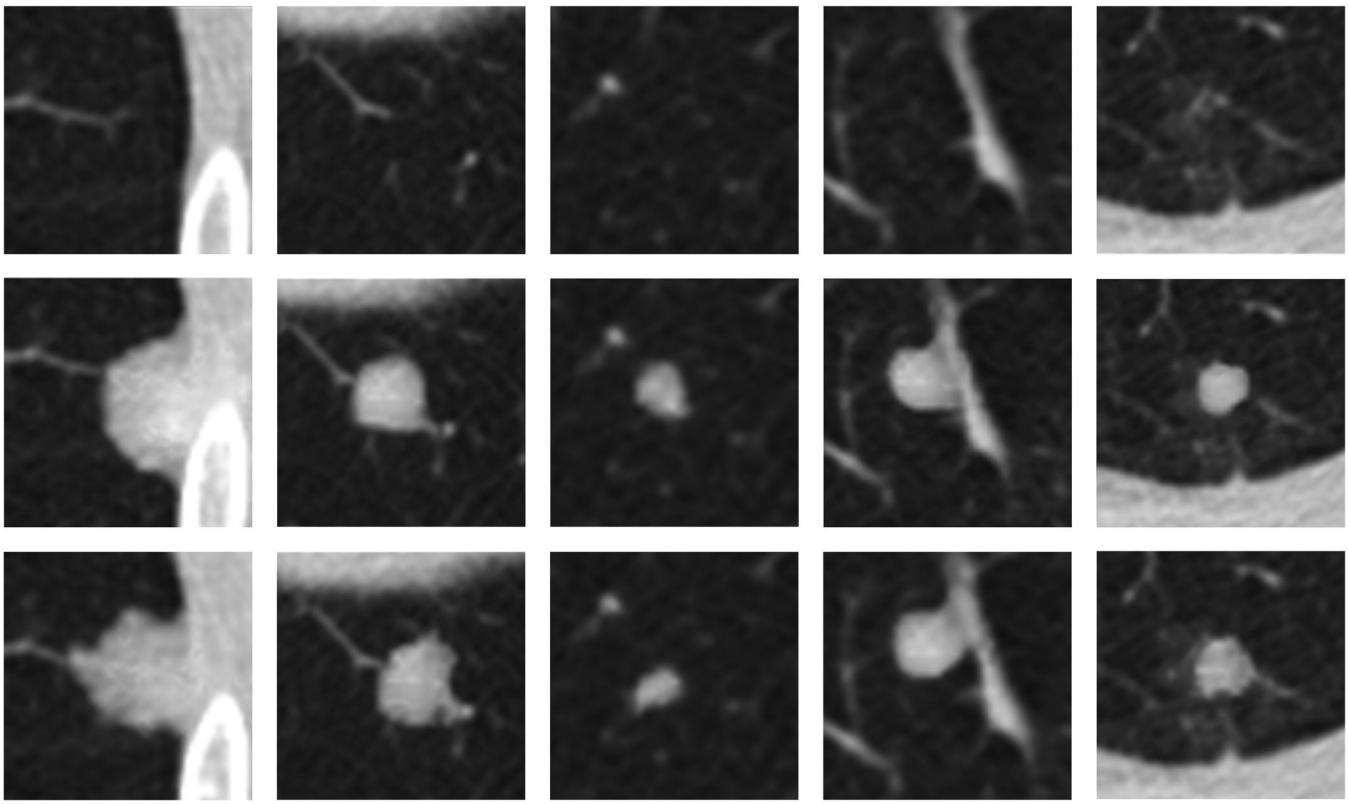

FIGURE 3 2D axial view of some example results. First row: original images, second row: synthetic images generated by the baseline method, and third row: synthetic images generated by the proposed method

the nodule as a mask. To control the size of the generated nodules, we will cut patches of different sizes according to the expected diameter and scale them to a fixed size of $32 \times 32 \times 32$. Then, we use the generator of the proposed method to synthesize new samples based on the inputs with the mask. After that, we perform inverse preprocessing operations including denormalization, de-equalization, and rescaling the images back to the original resolution. The resulting samples are pasted back into the original CT scans, which are employed as the training data for the lung nodule detection model. In the original experiment, the training dataset contains only the 888 CT scans from the LUNA16 dataset, which is randomly and equally divided into 10 subsets. We use the 10-fold cross-validation scheme to train the detection network with a ratio between the training set, the validation set, and the test set of 7:2:1. To confirm the effect of the baseline and proposed DA methods, we compare the detection results on the same test set with the same validation set and different training sets. These training sets select different proportions of the generated samples together with all real samples as follows: "Baseline" with 100% additional synthetic samples based on the baseline method compared to the number of real samples; and "70% DA", "100% DA" and "130% DA" with 70%, 100%, and 130% additional synthetic samples based on the proposed method, respectively.

We evaluate the results based on the free-response receiver operating characteristic (FROC) curves, which show the identification sensitivity on the corresponding FPs per scan (FPs/scan), and the CPM scores, which is the average sensitivity at seven predefined FP rates: 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs/scan according to the LUNA16 challenge. During the lung nodule detection stage, 1171 nodules were detected by the '100% DA' model, which achieved a high sensitivity of 98.74% with 14.85 candidates per scan on average. The model "Baseline" detected 1163 nodules and achieved a sensitivity of 98.06% with 15.64 candidates per scan on average. In addition, 1155 nodules were detected by the model trained by the original data from the LUNA16 dataset, which achieved a sensitivity of 97.39% with 16.73 candidates per scan on average. Compared to the model trained by the original data, the performance of the model "100% DA" is superior in terms of the CPM score and sensitivity. Compared to the model using the baseline method for DA to train the nodule detection network, the model using the proposed method for DA can achieve better performance.

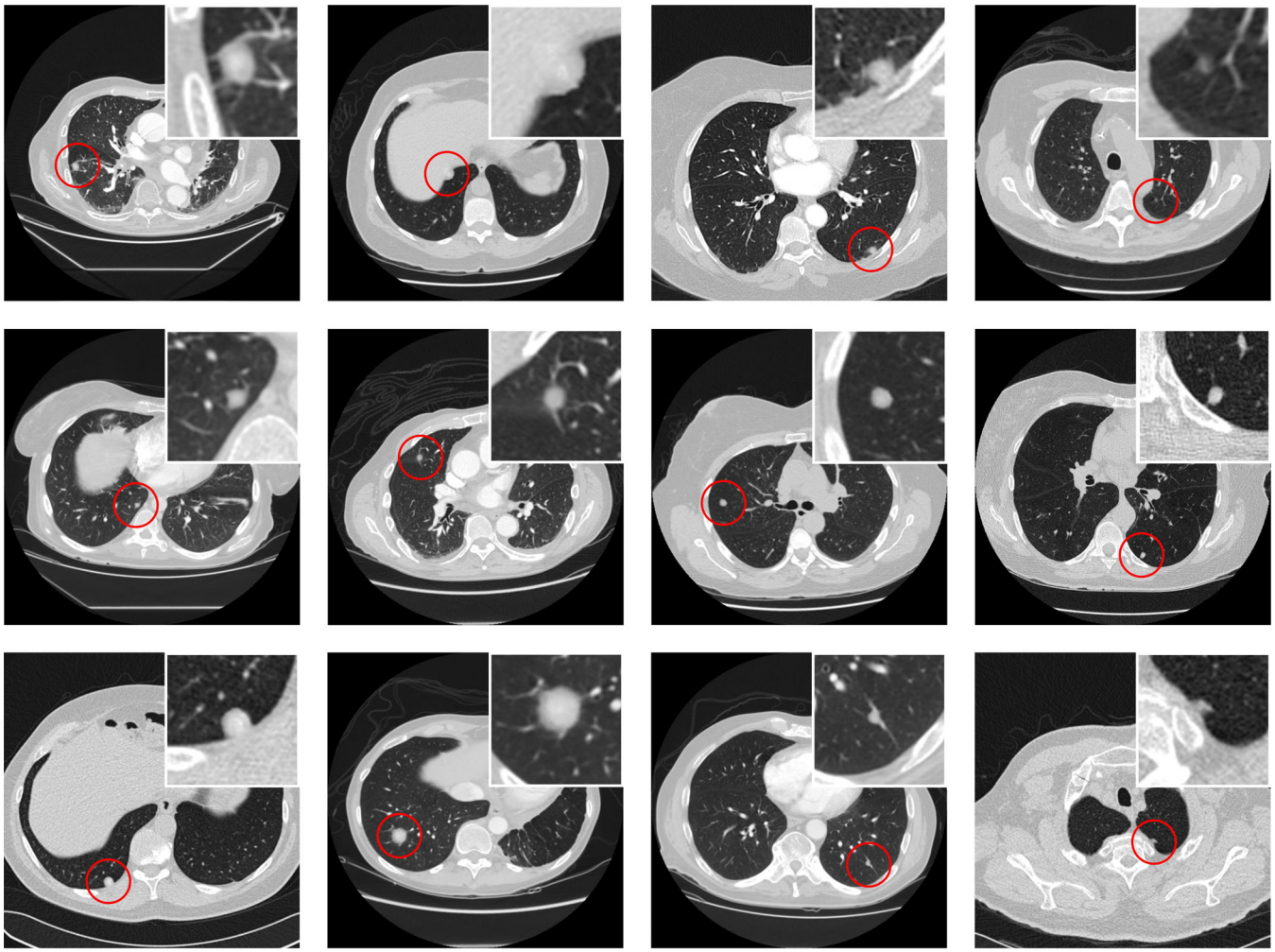

FIGURE 4 Examples of synthetic lung nodules in computed tomography images [Color figure can be viewed at wileyonlinelibrary.com]

The FROC curves are shown in Figure 6, and the quantitative nodule detection results are listed in Table 1.

3.4 | Visual turing test

We recruited three radiologists to evaluate the quality of the synthetic samples. We supplied a dataset composed of 50 real nodules, 50 synthetic nodules generated by the baseline model, and 50 synthetic nodules generated by the proposed model in a random order. All the samples are randomly selected, and their diameters are greater than 3 mm. The lung nodule image is a $32 \times 32 \times 32$ 3D cube with surrounding tissues with a lung nodule in the center. In this trial, the radiologists need to identify real and synthetic nodules to evaluate the visual quality of the synthetic medical images. Table 2 shows that the nodules generated by the proposed network are more visually realistic than the nodules generated by the baseline network, and the probability of the radiologists correctly distinguishing their categories is lower.

3.5 | Visualization using t-distributed stochastic neighbor embedding (t-SNE)

To further analyze the image synthesis capabilities of the proposed network, we use the t-SNE[33] algorithm to visualize the distribution of real and synthetic data. Through using the t-SNE algorithm for dimensionality reduction, we obtain the 2D representation of randomly sampled images, including 500 real images, 500 baseline method-based images, and 500 proposed method-based images. The experimental results are shown in Figure 7. According to the results, the data synthesized by GAN exhibit a similar distribution to the real data and can fill the part not covered by the real distribution to a certain extent. Thus, it is feasible to use synthetic images for data enhancement.

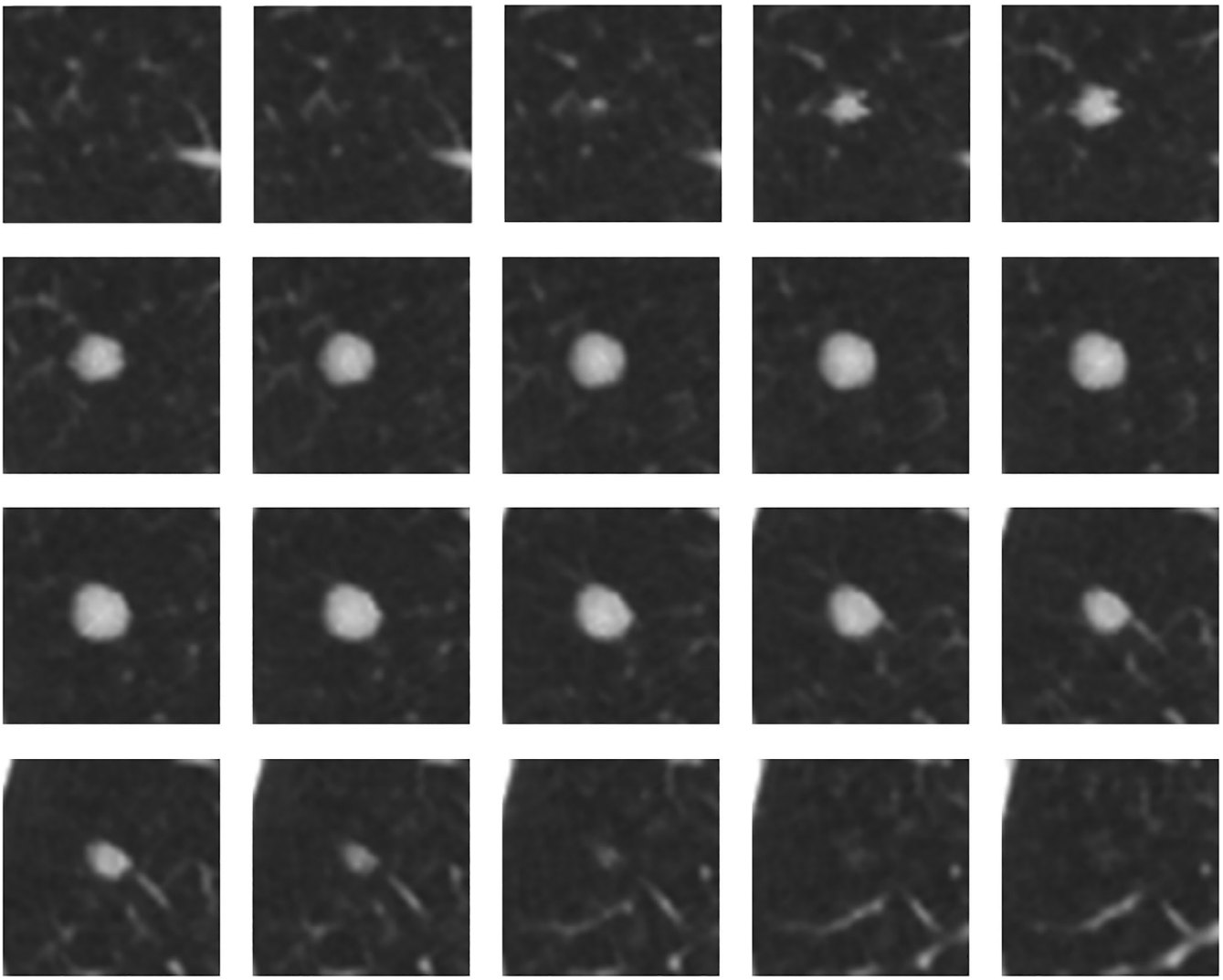

FIGURE 5 All patch slices from a sample

4 | DISCUSSION

In this paper, a novel CGAN-based method for lung nodule generation in CT scans is presented, which achieves promising performance for DA using synthetic samples. There are three primary aspects of the advantages of the proposed method. First, the generator and discriminator in the network are all based on 3D DCNNs. This structure can effectively extract information in space, which is more suitable for 3D medical image processing. Second, the generator based on U-Net structure adopts a multi-resolution decomposition and composition method by combining encoding and decoding paths with skip connection. Third, both networks introduce SE-ResNet blocks to reduce the difficulty of network training and improve the feature extraction capabilities of the network. The experimental results based on the lung nodule detection network show that using the proposed method to generate lung nodules for DA can effectively improve the accuracy of the detection network.

For the CGAN-based experiment, Figure 3 shows the resulting images from the proposed model and the baseline model. In terms of visual quality, the synthetic samples generated by the proposed model appear to be much more authentic than those generated by the baseline model. The same situation also occurred in the visual Turing test. As Table 2 shows, the radiologists encounter difficulty in classifying real nodules and synthetic nodules, especially those generated by the proposed model. The average probability that the radiologists correctly distinguished the synthetic nodules decreased by 10.67% and eventually reached 54.67%. This result indicates that the proposed model can generate sufficiently realistic lung nodules. To intuitively analyze the results, we visualized the distribution of real data and synthetic data using the t-SNE

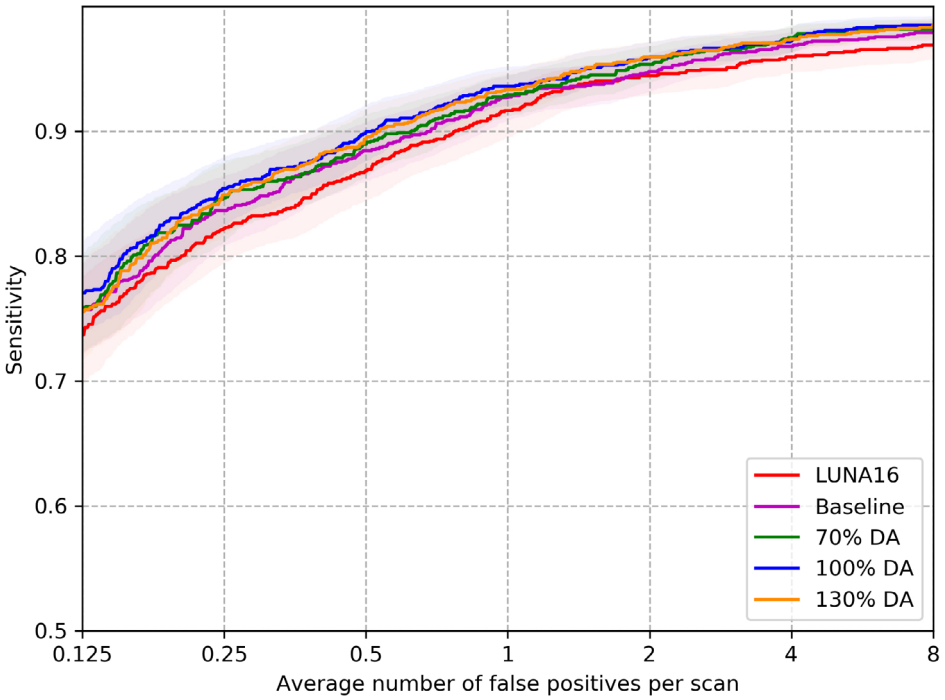

FIGURE 6 Comparison of the free-response receiver operating characteristic curves using the different data augmentation setups [Color figure can be viewed at wileyonlinelibrary.com]

algorithm. As shown in Figure 7, the real images and synthetic images have a similar distribution, indicating that the GAN can learn the distribution of real images and generate reasonable images. Moreover, the synthetic data distribution can partly fill the areas not covered by the real distribution, especially the proposed method-based images. This phenomenon shows the effectiveness of using data generated by GAN for DA. The realistic nodules generated by the proposed method that do not exist in the original data set can provide more plentiful and diverse features for the nodule detection network, resulting in an improvement in the performance of the lung nodule detection network. By introducing the SE-ResNet module, the performance of the model can be further improved. This mechanism can improve the feature extraction capability of the model by adaptively recalibrating the channel-wise feature maps. Additionally, the training difficulty of the

TABLE 1 Lung nodule detection results

Dataset	CPM score	Sensitivity (%)/8 FPs/scan
LUNA16	0.8870	0.9688
Baseline	0.9001	0.9781
70% DA	0.9054	0.9823
100% DA	**0.9123**	**0.9857**
130% DA	0.9072	0.9848

Note: The CPM has an average sensitivity of 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs/scan.

TABLE 2 Visual turing test results by three radiologists

		Accuracy (%)	True	False
Real images	Radiologist 1	46	23	27
	Radiologist 2	72	36	14
	Radiologist 3	64	32	18
Baseline method-based images	Radiologist 1	64	32	18
	Radiologist 2	62	31	19
	Radiologist 3	70	35	15
Proposed method-based images	Radiologist 1	56	28	22
	Radiologist 2	48	24	26
	Radiologist 3	60	30	20

Note: True and false indicate the number of correct and incorrect classifications, respectively, by the radiologists. Accuracy denotes the radiologists' successful classification ratio.

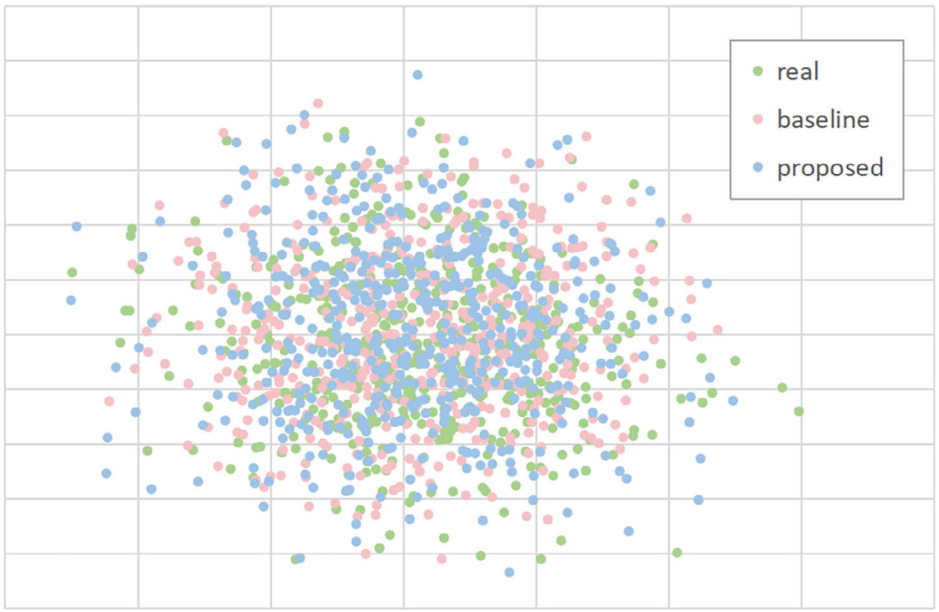

FIGURE 7 t-distributed stochastic neighbor embedding of real and synthetic images [Color figure can be viewed at wileyonlinelibrary.com]

model is further reduced, and the model is more suitable to process highly complex medical images.

After training the proposed model, we evaluate whether our work benefits lung nodule detection. Figure 6 and Table 1 present the quantitative results for the experiment aimed at improving the performance of the lung nodule detection model. We used an original training dataset, including only the LUNA16 data and four mixed training datasets, adding the synthetic images of the baseline or proposed model in different proportions. All test datasets are the same and only contain real data in LUNA16. The original results after training only with the real images are shown in red, and the results of the CGAN-based DA steps are shown in other colors. According to Figure 6 and Table 1, the results of CGAN-based DA can consistently boost the sensitivity at fixed values of FPs/scans, especially at a ratio of 100%. In terms of the CPM, the nodules generated through the baseline and proposed model can improve the detection score by 1.31% and 2.53%, respectively. Compared with the baseline method, using the proposed method to generate images for DA can increase the CPM score of the lung nodule detection network by 1.22%. At 8 FPs/scans, the sensitivity increased from 96.88% to 97.81% (Baseline) and 98.57% (100% DA). The 95% confidence intervals were estimated with bootstrapping. These results of t-SNE visualization indicate that synthetic nodules can partly cover the real data distribution unfilled by the original data. Using the proposed method to synthesize meaningful nodules that do not exist in the real data set can provide more abundant and diverse features for the nodule detection network and effectively improve the generalization ability of the lung nodule detection network. By introducing the SE-ResNet modules, the performance of GAN can be further improved, and the nodule detection network that uses the proposed method for DA can produce higher accuracy and CPM scores. Unexpectedly, adding more generated samples slightly reduce the detection performance, probably because of the imbalance between the real and synthetic nodules. These results validate the effect of using synthetic data to augment training datasets.

Generally, there are two DA methods: conventional methods and GAN-based methods. The conventional DA methods adopting geometric or intensity transformation methods generally cannot introduce sufficient new features, which limits the improvement in the detection methods.[11] Additionally, GAN-based methods have undergone rapid development and can generate enough new samples with diverse characteristics. Maria et al.[15] used a 2D GAN framework to generate lung nodules, which do not contain enough 3D spatial information. Our proposed method can extract more valuable spatial information and more expressive features of lung nodules than 2D GAN-based methods. Currently, some works have used 3D DCNNs without U-Net structures.[18-21] In light of the extensive region of grayscale and blurred borders of the CT scans, a CNN with a U-Net structure is more suitable for dealing with highly complex medical images than a CNN without a U-Net structure; the U-Net structure strengthens the feature extraction capability and accuracy of localization of the network by combining feature maps from corresponding paths. Some works used the U-Net structure,[16,17,22,23] but they considered only the relationships between convolutional layers. Compared with these works, the 3D SE-ResNet block applied to the proposed model focuses on the connections between the feature channels of the convolutional layers,

which can adaptively recalibrate channel-wise feature maps to efficiently extract more representative features. In summary, the network structure used by most researchers for GAN is relatively simple. Some researchers change the network structure, such as U-Net, or design a new loss function to improve the performance of GAN.[17] However, the efficacy of multiple promising methods on GAN-based methods for DA has not been widely explored. In our study, we introduced the 3D SE-ResNet blocks into CGAN to explore the possibility and feasible directions to further improve the performance of GAN. The 3D SE-ResNet block integrates the advantages of residual learning for feature reuse and the SE operation for feature recalibration. Through the adaptive mechanism, the network encourages the model to learn more meaningful features by emphasizing useful features and suppressing weak ones. The SE-ResNet module not only is beneficial for DCNNs to learn the representative nodule features but also allows the GAN to be optimized easily and decreases the difficulty of model training.

Since most of the current research uses the artificial visual evaluation method to evaluate the performance of GAN to generate medical images, we only compare our results with two methods that also use synthetic lung nodules as DA for the lung nodule detection network. The highest CPM score obtained by Han et al.[23] was 0.550, an improvement of 0.032 and the highest CPM score obtained by Liu et al.[22] was 0.8440, an improvement of 0.0844. The highest CPM score obtained in the present study was 0.9123, an improvement of 0.0253. Although the improvement of our CPM score is less than that of the other two networks, the original results of lung nodule detection network we used has the highest CPM score among the three networks. Thus, we cannot simply evaluate network performance by the improvement. Compared with the improvement space of the lung nodule detection network we have adopted, the improvement we have achieved is considerable. And compared with these two networks, the proposed network can generate more realistic nodules, which have slighter shading difference at the edge of the masks. In addition, compared to the baseline method, the higher CPM score also reflects the advantages of the proposed method. In our future work, we will investigate the method of generating medical images with more complex structures and more manipulable attributes.

5 | CONCLUSION

In this paper, we proposed a CGAN-based method to generate manipulable lung nodule samples in CT scans. This method is composed of a generator, which is based on the U-Net structure, and a discriminator; both introduce SE-ResNet blocks to achieve superior performance. Through exhibiting the synthetic images, we indicate that the proposed method can generate diverse and realistic lung nodules at desired positions with the desired size. To evaluate the performance of the proposed method, we used synthetic samples for DA to train a lung nodule detection network. The results indicate that these generated samples, which cover a part of the data distribution unfilled by the original data, can boost the overall performance of the lung nodule detection network by 2.53% in terms of the CPM. Overall, the proposed method shows promising performance for the augmentation of medical images, which could solve the problem of medical image shortages and is not limited to lung nodules. Moreover, these synthetic medical images could help train medical students.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (grant number 81871457); the National Natural Science Foundation of China (grant number 51775368); and the National Natural Science Foundation of China (grant number 51811530310). We are grateful to the LUNA16 challenge organizers for their efforts in collecting and sharing chest CT scan data for comparing pulmonary nodule detection algorithms.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

T.B. made the conception and design of the study, drafted the article, and revised it critically for important intellectual content. Z.Y. and S.J. have provided guidance for experiment and revised the manuscript critically. G.Z., H.Z., and L.W. acquired and analyzed the data.

ORCID

Shan Jiang https://orcid.org/0000-0002-5518-5335

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69:7-34. https://doi.org/10.3322/caac.21551.
2. Torre LA, Siegel RL, Jemal A. Lung cancer statistics. *Adv Exp Med Biol*. 2016;893:1-19. https://doi.org/10.1007/978-3-319-24223-1_1.
3. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early lung cancer action project: overall design and findings from baseline screening. *Lancet*. 1999;354:99-105. https://doi.org/10.1016/S0140-6736(99)06093-6.
4. Zhang G, Yang Z, Gong L, Jiang S, Wang L. Classification of benign and malignant lung nodules from CT images based on hybrid features. *Phys Med Biol*. 2019;64:125011. https://doi.org/10.1088/1361-6560/ab2544.

5. Jemal A, Siegel R, Ward E, et al. Cancer statistics. *CA Cancer J Clin*. 2008;58:71-96. https://doi.org/10.3322/CA.2007.0010.
6. Wu GX, Raz DJ. Lung cancer screening. *Cancer Treat Res*. 2016;170:1-23. https://doi.org/10.1007/978-3-319-40389-2_1.
7. Zhang G, Jiang S, Yang Z, et al. Automatic nodule detection for lung cancer in CT images: a review. *Comput Biol Med*. 2018;103:287-300. https://doi.org/10.1016/j.compbiomed.2018.10.033.
8. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph*. 2007;31:198-211. https://doi.org/10.1016/j.compmedimag.2007.02.002.
9. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. https://doi.org/10.1016/j.media.2017.07.005.
10. Gong L, Jiang S, Yang Z, Zhang G, Wang L. Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks. *Int J Comput Assist Radiol Surg*. 2019;14:1969-1979. https://doi.org/10.1007/s11548-019-01979-1.
11. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*. 2018;321:321-331. https://doi.org/10.1016/j.neucom.2018.09.013.
12. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets; 2014. https://arxiv.org/abs/1406.2661. (Accessed April 15, 2020).
13. Kazeminia S, Baur C, Kuijper A, et al. GANs for medical image analysis; 2018. https://arxiv.org/abs/1809.06222v3.
14. Wolterink JM, Kamnitsas K, Ledig C, et al. Generative adversarial networks and adversarial methods in biomedical image analysis; 2018. https://arxiv.org/abs/1810.10352.
15. Chuquicusma MJM, Hussein S, Burt J, et al. How to fool radiologists with generative adversarial networks? A visual turing test for lung cancer diagnosis. *Proc Int Symp Biomed Imaging*. 2018;2018:240-244. https://doi.org/10.1109/ISBI.2018.8363564.
16. Mirsky Y, Mahler T, Shelef I, et al. CT-GAN: malicious tampering of 3D medical imagery using deep learning. In: Proceedings of the USENIX Security Symposium; 2019, 2019: 461–478. https://www.usenix.org/conference/usenixsecurity19/presentation/mirsky.
17. Jin D, Xu Z, Tang Y, et al. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. *Proc Med Image Comput Comput Assist Interv*. 2018;2018:732-740. https://doi.org/10.1007/978-3-030-00934-2_81.
18. Yang J, Liu S, Grbic S, et al. Class-aware adversarial lung nodule synthesis in CT images; 2018. https://arxiv.org/abs/1812.11204.
19. Zhao D, Zhu D, Lu J, Luo Y, Zhang G. Synthetic medical images using F&BGAN for improved lung nodules classification by multi-scale VGG16. *Symmetry*. 2018;10:519. https://doi.org/10.3390/sym10100519.
20. Kommrusch S, Pouchet L. Synthetic lung nodule 3D image generation using autoencoders. *Proc Int Jt Conf Neural Netw*. 2019;2019:1-9. https://doi.org/10.1109/IJCNN.2019.8852224.
21. Wang Q, Zhou X, Wang C, et al. WGAN-based synthetic minority over-sampling technique: improving semantic fine-grained classification for lung nodules in CT images. *IEEE Access*. 2019;7:18450-18463. https://doi.org/10.1109/ACCESS.2019.2896409.
22. Liu S, Gibson E, Grbic S, et al. Decompose to manipulate: manipulable object synthesis in 3D medical images with structured image decomposition; 2018. https://arxiv.org/abs/1812.01737.
23. Han C, Kitamura Y, Kudo A, et al. Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection. *Proc Int Conf 3D Vision*. 2019;2019:729-737. https://doi.org/10.1109/3DV.2019.00085.
24. Mirza M, Osindero S. Conditional Generative Adversarial Nets; 2014. https://arxiv.org/abs/1411.1784.
25. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2016;2016:770-778. https://doi.org/10.1109/CVPR.2016.90.
26. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2018;2018:7132-7141. https://doi.org/10.1109/TPAMI.2019.2913372.
27. Armato IIISG, McLennan G, Bidaut L, 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*. 2011;38:915-931. https://doi.org/10.1118/1.3528204.
28. Setio AAA, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal*. 2017;42:1-13. https://doi.org/10.1016/j.media.2017.06.015.
29. Rundo L, Han C, Nagano Y, et al. USE-net: incorporating squeeze-and-excitation blocks into U-net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing*. 2019;365:31-43. https://doi.org/10.1016/j.neucom.2019.07.006.
30. Roy AG, Nassir N, Christian W, et al. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. *Proc Medi Image Comput Comput Assist Interv*. 2018;2018:421-429. https://doi.org/10.1007/978-3-030-00928-1_48.
31. Isola P, Zhu J, Zhou T, et al. Image-to-image translation with conditional adversarial networks. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2017;2017:5967-5976. https://doi.org/10.1109/CVPR.2017.632.
32. Liao F, Liang M, Li Z, Hu X, Song S. Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky Noisy-OR network. *IEEE Trans Neural Netw Learn Syst*. 2019;30(11):3484-3495. https://doi.org/10.1109/TNNLS.2019.2892409.
33. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579-2605. https://jmlr.org/papers/v9/vandermaaten08a.html.

How to cite this article: Bu T, Yang Z, Jiang S, Zhang G, Zhang H, Wei L. 3D conditional generative adversarial network-based synthetic medical image augmentation for lung nodule detection. *Int J Imaging Syst Technol*. 2020;1–12. https://doi.org/10.1002/ima.22511