

Automatic Lung Nodule Segmentation and Intra-Nodular Heterogeneity Image Generation

Jiangdian Song , Member, IEEE, Shih-Cheng Huang, Brendan Kelly, Guanqun Liao, Jingyun Shi , Ning Wu, Weimin Li, Zaiyi Liu, Lei Cui , Matthew P. Lungre, Michael E. Moseley , Peng Gao, Jie Tian , Fellow, IEEE, and Kristen W. Yeom 

Abstract—Automatic segmentation of lung nodules on computed tomography (CT) images is challenging owing to the variability of morphology, location, and intensity. In addition, few segmentation methods can capture intra-nodular heterogeneity to assist lung nodule diagnosis. In this study, we propose an end-to-end architecture to perform fully automated segmentation of multiple types of lung nodules and generate intra-nodular heterogeneity images for clinical use. To this end, a hybrid loss is considered by introducing a Faster R-CNN model based on generalized intersection over union loss in generative adversarial network. The Lung Image Database Consortium image collection dataset, comprising 2,635 lung nodules, was combined with 3,200 lung nodules from five hospitals for this study. Compared with manual segmentation by radiologists, the proposed model obtained an average dice coefficient (DC) of 82.05% on the test dataset. Compared with *U-net*, *NoduleNet*, *nnU-net*, and other three models, the proposed method achieved comparable performance.

Manuscript received July 2, 2021; revised September 30, 2021 and November 26, 2021; accepted December 12, 2021. Date of publication December 15, 2021; date of current version June 6, 2022. This work was supported in part by the China Scholarship Council and in part by the National Natural Science Foundation of China under Grant 82001904. (*Corresponding authors:* Peng Gao; Jie Tian.)

Jiangdian Song is with the College of Health Management, China Medical University, Shenyang, Liaoning 110122, China, and also with the School of Medicine Stanford University, Stanford, CA 94306 USA (e-mail: song.jd0910@gmail.com).

Lei Cui is with the College of Health Management, China Medical University, Shenyang, Liaoning 110122, China (e-mail: lcui@cmu.edu.cn).

Shih-Cheng Huang, Brendan Kelly, Matthew P. Lungre, Michael E. Moseley, and Kristen W. Yeom are with the School of Medicine Stanford University, Stanford, CA 94306 USA (e-mail: mschuang@stanford.edu; brendanskelly@me.com; mlungren@stanford.edu; moseley@stanford.edu; kyeyom@stanford.edu).

Guanqun Liao is with the Department of Hepatobiliary Surgery Foshan Hospital, Southern Medical University, Foshan 528000, Guangdong, China (e-mail: keepthinking@126.com).

Jingyun Shi is with the Department of Radiology, Shanghai Pulmonary Hospital, Shanghai 200433, China (e-mail: shijingyun89179@126.com).

Ning Wu is with the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences, Beijing 100021, China (e-mail: cjr.wuning@vip.163.com).

Weimin Li is with the Department of Respiratory and Critical Care Medicine, West China Hospital, Chengdu, Sichuan 610044, China (e-mail: weimin003@163.com).

Zaiyi Liu is with the Department of Radiology Guangdong General Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510055, China (e-mail: zyliu@163.com).

Peng Gao is with the First Affiliated Hospital of China Medical University, Shenyang, Liaoning 110122, China (e-mail: gaokaji16@163.com).

Jie Tian is with the Beijing University of Aeronautics and Astronautics, Beijing 100191, China (e-mail: tian@ieee.org).

Digital Object Identifier 10.1109/JBHI.2021.3135647

on lung nodule segmentation and generated more vivid and valid intra-nodular heterogeneity images, which are beneficial in radiological diagnosis. In an external test of 91 patients from another hospital, the proposed model achieved an average DC of 81.61%. The proposed method effectively addresses the challenges of inevitable human interaction and additional pre-processing procedures in the existing solutions for lung nodule segmentation. In addition, the results show that the intra-nodular heterogeneity images generated by the proposed model are suitable to facilitate lung nodule diagnosis in radiology.

Index Terms—Computed tomography, generative adversarial network, generalized intersection over union, intra-nodular heterogeneity, lung nodule segmentation.

I. INTRODUCTION

LUNG CANCER is one of the most common type of cancers according to the latest global cancer statistics, leading to the highest number of cancer-related deaths worldwide [1]. Frequently used medical imaging techniques, such as traditional computer-assisted diagnosis as well as emerging radiomics and deep learning methods, have proven significant in improving the diagnosis and prognosis of lung cancer in current clinical practice [2]. Based on ingeniously designed phenotypic decoders for heterogeneity analysis of lung lesions, new roadmaps for the early detection, diagnosis, and prognosis of lung lesions have been created [3], [4]. The precise segmentation of lung nodules on computed tomography (CT) images and assessment of intra-nodular heterogeneity are prerequisites for these techniques, which not only reduce the bias of subsequent radiological analysis but also improve the work efficiency of radiologists.

Currently, manual delineation of lung nodules in CT scan is most commonly performed slice by slice by a radiologist. Although this manual segmentation method may guarantee accuracy to a certain extent, the procedure involves several shortcomings. First, the radiologist must identify and delineate the boundary of the lung nodule on each CT slice, which is a tedious procedure that relies on subjective experience. Second, the number of patients with lung nodules enrolled in such studies has grown from dozens to hundreds or even thousands in recent years, making manual segmentation less feasible. Finally, for the current big data-driven medical imaging workflow, manual delineation of the region of interest (ROI) contradicts the objective of using artificial intelligence. In summary, end-to-end

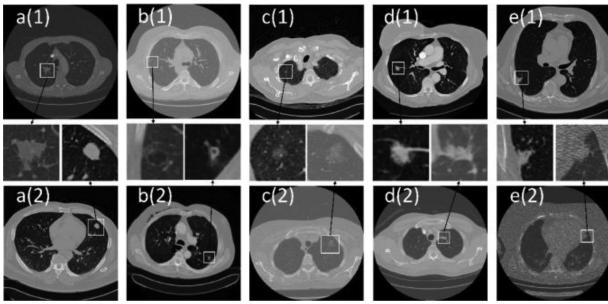


Fig. 1. Different types of lung nodules: (a) solid nodules; (b) solid nodules with cavity; (c) GGNs; (d) and (e) juxta-vascular and juxta-pleural lung nodules, respectively.

algorithms for automatic segmentation of lung nodules on CT images are increasingly necessary to improve medical imaging workflows.

Automating lung nodule segmentation has long been challenging owing to their irregular shapes, the variability of nodule locations, and the extreme variation of intensity for different types of nodules on CT images [5]–[8]. As shown in Fig. 1(a), the solid nodules have various shapes, and the presence of cavities (Fig. 1(b)) further limits the sensitivity of automatic segmentation algorithms for the solid nodules. For ground glass nodules (GGN), as shown in Fig. 1(c), the intensity is similar to that of lung parenchyma on CT images, hindering automatic detection and delineation. For lung nodules adhering to the vessel, chest wall, or mediastinum (Fig. 1(d–e)), the identical visual characteristics and intensity with the surrounding tissues on CT images increase the difficulty of automatic segmentation. Moreover, even for the same type of lung nodule, the pixel-value intensity varies significantly with different CT scanning parameters, as shown in the two rows of Fig. 1. All these factors make precise automatic delineation of lung nodules very challenging.

Although lung nodule segmentation methods have been reported, barriers remain in applying these methods to clinical diagnosis in radiology. One challenge is that current segmentation methods focus only on segmentation accuracy, whereas the clinical evaluation of ROIs, such as texture-based intra-nodular heterogeneity, still relies on hand-crafted features (such as gray-level run length matrix and gray-level co-occurrence matrix) or visual evaluation by radiologists [9], [10]. Moreover, intra-nodular heterogeneity descriptors are susceptible to the differences in scanning conditions, image preprocessing, and feature engineering; therefore, a consensus for determining intra-nodular heterogeneity for clinical diagnosis has not yet been established in the field of radiology [11]. In addition, ROI segmentation and heterogeneity information generation in these methods are always divided into independent steps. However, manual intervention and image processing connecting these steps hinder the reproducibility of intra-nodular heterogeneity descriptors [11]. Therefore, automatic methods for lung nodule segmentation and intra-nodular heterogeneity image generation are needed in current radiological practice.

A. Related Works

Many algorithms have been proposed to overcome the above-mentioned challenges of lung nodule segmentation. These algorithms can be divided into two groups: those based on traditional image recognition methods, and those based on emerging deep learning models.

In the first group, region-growing, level set, graph cut, and other algorithms based on pattern recognition have been developed to estimate the difference between the nodule and its surrounding tissues. Song *et al.* proposed a toboggan-based 3D region-growing algorithm to automatically detect potential nodule candidates [6]. Twenty-six neighborhoods of each confirmed lung nodule voxel were evaluated based on the adaptive intensity range to generate a complete nodule. Using 850 nodules from the Lung Image Database Consortium image collection (LIDC-IDRI) dataset [12] and 121 nodules from an in-house dataset, an accuracy of 96.35% was achieved for lung nodule detection as well as a dice coefficient (DC) of 81.57% compared with manual segmentation by radiologists. However, the potential leakage of region-growing for juxta-pleural nodules limited the utility of this segmentation method for clinical practice. To solve this problem, a modified spherical region-growing method was proposed [13]. Based on the seed point input by a human observer, a voxel would be considered nodular only if it met the spherical growth condition and volume variation limit. This method was validated on 80 patients and achieved a DC of 88% when compared with manual segmentation results. To better identify the region of the lung nodules, some algorithms in this group require extra human interaction or extraction of pulmonary parenchyma in advance. A semi-automatic geodesic active contour model in level set was proposed in [14] using the threshold of mean intensity, which was acquired from the manually selected lung nodule reference slice. Based on the seed point selected by a human observer, convex-hull, geodesic distance, and selective enhancement filtering were sequentially applied in [15] to remove the pleural surface, adhesive blood vessels, and blood vessels of non-solid nodules, respectively. A total of 891 nodules were used and an overlap of 70%–80% was achieved when compared to segmentations performed by radiologists. Based on the lung parenchyma segmentation, a weighted support vector machine algorithm was proposed in [16] to distinguish nodules from adjacent vascular bundles. A dot filter was used to extract the spherical architectures and suppress other shapes in the non-vessel tree group, obtaining 87.81% sensitivity on 204 nodules from the LIDC-IDRI dataset.

For deep learning-based segmentation methods, a convolutional neural network (CNN)-based network structure is most widely used. Wang *et al.* [8] proposed a multi-scale CNN architecture that performed voxel-wise predictions using pre-selected lung nodule image patches on CT images. This method demonstrated a segmentation accuracy of 82% compared with manual segmentation on 893 lung nodules from the LIDC-IDRI dataset. However, the additional human interaction of choosing the lung nodule image cube and the preprocessing step of balancing nodule and non-nodule point proportions limited

its application. Recently, a fully automated lung nodule detection scheme was proposed using faster regional-CNN (Faster R-CNN) [17]. Based on the nodule candidates acquired by a pre-designed detection network, false-positive reduction and nodule extraction were conducted separately using two other specific CNN architectures. An intersection over union (IoU) of 70.24% and DC of 79.30% were obtained for 223 solid lung nodules from the LIDC-IDRI dataset. A new method called *NoduleNet*, comprising a nodule candidate screening module, a decoupled false-positive reduction module, and finally a nodule segmentation module, was recently proposed [18]. A pre-segmented lung mask was used, and a detection sensitivity of 87.27% and a DC of 83.10% was reported on the LUNA16 dataset. As an emerging deep learning architecture, adversarial networks have been used for segmentation tasks. Among these, generative adversarial networks (GANs) have been proven to be a powerful framework for medical image segmentation [19], [20]. For the lung nodule segmentation task, based on the segmented lung parenchyma images, Zhao *et al.* [21] proposed a GAN-based lung nodule detection architecture using 3D U-net as the generator network. The sum of the pixel-wise loss calculated by generator and the dice coefficient loss for evaluating the overlapping of ground truth image and generated image was used as the total network loss for training. A data augmentation method using a style-based GAN to synthesize training data for robust lung nodule segmentation was proposed in [22]. The image was first input into the style encoder for image style conversion to obtain the augmented data, and then a semantic region-adaptive generator was used for mimic image production. A DC of 85.21% was obtained on the augmented LIDC-IDRI datasets.

In summary, traditional lung nodule segmentation methods are limited by the requirement for additional human processing and the challenge of accurately segmenting adhesive nodules, whereas the current deep learning models may be more rigorous for image pre-processing and pre-selected nodule patches. In addition, the extraction of intra-nodular heterogeneity to assist clinical diagnosis has not been considered in segmentation studies. Therefore, the development of end-to-end lung nodule segmentation and intra-nodular heterogeneity map generation methods is required in current clinical practice. In this study, we integrated a Wasserstein generative adversarial network with a gradient penalty (WGAN-GP) and Faster R-CNN with generalized intersection over union (GIoU) loss for both end-to-end segmentation of GGN, solid, juxta-pleural, and juxta-vascular nodules, and corresponding intra-nodular heterogeneity information generation. The proposed approach is called the *LNHG* model, and it was validated on multiple lung nodule CT image datasets. In contrast to current lung nodule segmentation algorithms, the approach proposed herein does not require any image preprocessing procedures. Manual selection of the nodule cube in CT images or balancing of positive and negative samples are also not required. These advantages overcome the challenge of poor segmentation accuracy of adhesive lung nodules, achieving stable lung nodule segmentation performance for CT images of varied quality and different types of lung nodules, and generating intra-nodular heterogeneity images capturing data that are imperceptible to the naked eye.

II. MATERIAL AND METHODS

A. Generative Adversarial Network

GAN models [23] have proven effective for synthesizing images. The generator network of GAN is designed to generate an image mimicking a target image, and the model learns the characteristics of its corresponding target image. Then, both the target and generated images are input into the discriminator network to calculate the difference between the two. However, it has been proven that the original GAN cannot always generate a significant overlap on the ROI between generated images and target images during training. Thus, gradient vanishing occurs on the generator, and the process of updating of the generative network is interrupted.

To avoid this problem, an improved Wasserstein GAN with a gradient penalty (WGAN-GP) [24] for network regularization was recently proposed. Accordingly,

$$\min_G \max_D L(D, G) = -\mathbb{E}_{x \sim P}[D(x)] + \mathbb{E}_{x \sim P_g}[D(G(X))] + \lambda \mathbb{E}_{\hat{x} \sim P\hat{x}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2], \quad (1)$$

where $\mathbb{E}(\cdot)$ denotes the expectation operator, and P_r and P_g are the distributions of the target and generated images, and D and G represent the discriminator and generator network, respectively. The first two terms on the right side of (1) are the WGAN losses, which represent the losses of the target and generated images, respectively, when they pass through the discrimination network. The last term is the gradient penalty, which is proposed to refine the k -Lipschitz restriction of the original WGAN model. The coefficient λ controls the penalty, and $P\hat{x}$ represents the uniform samples along straight lines between pairs of the points sampled from the target and the generated images. In the WGAN-GP model, the logarithm is no longer needed for the loss calculation.

B. Generalized Intersection Over Union

IoU is a widely used similarity measurement for evaluating benchmarks in image detection tasks. The overlap between 2D or 3D objects is defined by IoU in (2). In the case of a bounding box with axis alignment, IoU can be used as the loss function, and its utility for medical imaging detection tasks has been confirmed [17].

$$IoU = \frac{|B_r \cap B_p|}{|B_r \cup B_p|}, \quad (2)$$

where B_r and B_p are the bounding boxes of the ROI on the target and predicted images, respectively. However, there are limitations in using IoU for loss [25]. When the overlap between B_r and B_p approaches zero, the IoU becomes zero and can never be optimized. In addition, IoU fails to reflect how the overlap between the two objects occurs, because the IoUs for different alignments of two shapes are identical if the volumes (areas) of their intersection in each case are equal [26], [27].

Recently, a generalized IoU (GIoU) was proposed to solve the abovementioned problems [26]. By finding the smallest enclosing convex object C , i.e., the minimum bounding box that could cover B_r and B_p , the ratio of the area of C that does not cover B_r and B_p to the total area of C is calculated. Finally, the

GIoU is obtained using the IoU of B_r and B_p minus the ratio, as given by

$$GIoU = IoU - \frac{|C| / (B_r \cup B_p)|}{|C|}. \quad (3)$$

When using GIoU for loss evaluation in a neural network, the following loss function was used to optimize the difference between B_r and B_p .

$$L_{GIoU} = 1 - GIoU \quad (4)$$

Studies have indicated that GIoU could inherit the ability of IoU to encode the shape attributes into area attributes of the targets being compared [26]. When used as the loss function, it is strongly related to IoU for the overlap of the objects. Further, GIoU can also optimize the network when there is no overlap. For image detection and segmentation task, recent studies have demonstrated the potential of GIoU loss for improving the ROI segmentation accuracy [28], [29].

C. Faster R-CNN

Instead of using selective search for region proposal (RP) extraction, a region proposal network (RPN) was designed in Faster R-CNN to improve the efficiency of foreground candidates searching. By using the RPN and ROI pooling layers, Faster R-CNN is more advantageous for detection tasks than traditional R-CNN-based models [30].

For Faster R-CNN, a backbone such as ResNet was first used for feature map extraction. A set of anchors with various sizes and scales were defined for each pixel on the output feature map of the backbone. For all the anchors, two fully connected layers were applied to perform the classification and bounding box regression. The classifier layer was used to determine the probability of an anchor belonging to the target by comparing the overlap between the anchor and target region, as shown in (5), and the bounding box layer was used to detect the location of target.

$$\begin{cases} \text{Anchor} = \text{Positive RP, if Overlap} > 0.7 \\ \text{Anchor} = \text{Negative RP, if Overlap} < 0.3 \end{cases} \quad (5)$$

Based on the probability of all the RPs, the top ranked positive RPs with high overlap were selected as the input for ROI classification in Fast R-CNN. Smooth L1 loss was then used as the loss function for bounding box layer in this method.

$$Smooth_{L1}(x) = \begin{cases} 0.5x^2, \text{if } |x| < 1 \\ |x| - 0.5, \text{otherwise} \end{cases}, \quad (6)$$

where x denotes the deviation between the parameters of predicted and target bounding box. Recent studies have indicated that the detection task will benefit by replacing the smooth L1 loss with state-of-the-art GIoU loss in such detection networks [31].

D. Proposed LNHG Model

Fig. 2 shows the pipeline of the proposed *LNHG* model. $R(x) \in \mathbb{R}^{N \times N}$ is defined as a binarized radiologist segmented lung nodule image, and $O(x) \in \mathbb{R}^{N \times N}$ is the corresponding

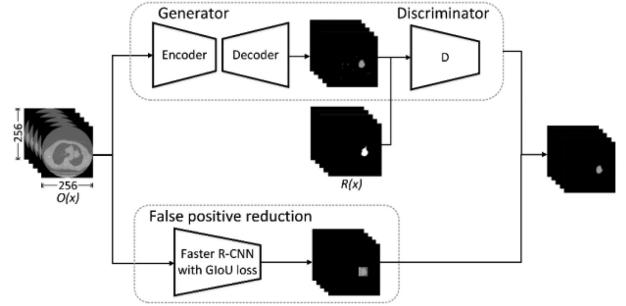


Fig. 2. Flowchart of the proposed *LNHG* model. Two branch networks, a WGAN-GP for segmentation and a Faster R-CNN for false-positive reduction, are simultaneously and separately executed. The WGAN-GP segmentation branch (above) is used to produce lung nodule image and intra-nodular heterogeneity directly from the original CT image, where the Faster R-CNN branch (below) is used to reduce potential false-positive candidates. $R(x)$ denotes the images manually segmented by radiologists, and $O(x)$ denotes the original lung CT images.

original lung CT image. The aim of the generating process is to optimize the *LNHG* model that maps $O(x)$ to $R(x)$.

$$LNHG : O(x) \rightarrow R(x) \quad (7)$$

The proposed *LNHG* model comprises two branches, the first branch is meant for potential lung nodule segmentation and intra-nodular heterogeneity information generation, and the second branch is used to reduce potential false-positive candidates produced by the first branch. In the first branch, a WGAN-GP is used to generate lung nodule images directly from the original CT image. According to the pre-experiment, the first branch network can not only recognize the morphology of potential lung nodules, but also capture subtle intra-nodular heterogeneous activities from original CT images. Meanwhile, based on previous studies, the pipeline of the second branch using Faster R-CNN with GIoU loss helps reduce potential fake nodules resulting from imbalance of pixels between nodule and non-nodule on CT images.

The WGAN-GP model comprises two networks: a generator network G that learns the characteristics of the $R(x)$ image so that the $O(x)$ image continuously fits the $R(x)$ image and a discriminator network D to differentiate between the distributions of the $R(x)$ image P_r and generated image P_g . The encoder and decoder are symmetrically arranged in the generator network G , corresponding to four 3D convolution and deconvolution layers. To speed up the learning process, WGAN-GP is modified by introducing short connections to accumulate the results of each 3D convolution with the corresponding 3D deconvolution results, as shown in **Fig. 3**. The sequence of the number of filters used in the generator network G is 32, 64, 128, 256, 128, 64, 32, and 1. Kernels of $3 \times 3 \times 3$ are used in each layer with stride 1. Each of the encoders successively consists of 3D convolution, 3D batch normalization, and LeakyReLU layers. For the first three decoders, a short connection is performed after 3D deconvolution; then 3D batch normalization and LeakyReLU layers are used. For the last decoder, the ReLU layer is employed after the short connection [32].

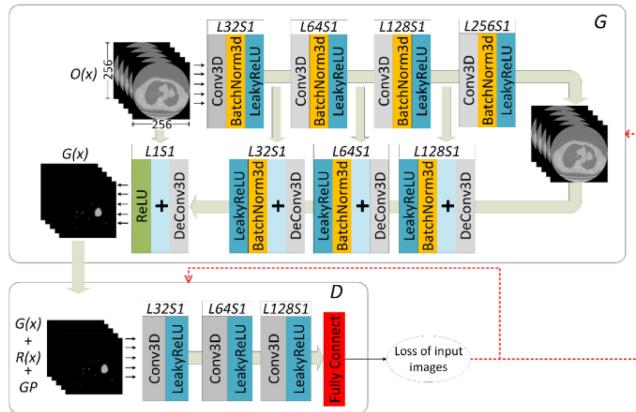


Fig. 3. Flowchart of the WGAN-GP model. L and S in each layer represent the number of feature maps and the stride, respectively. $R(x)$ denotes the images manually segmented by radiologists, and $O(x)$ denotes the input original CT images. $G(x)$ represents the generated image, and GP denotes the image generated by gradient penalty in the WGAN-GP framework. “+” indicates element-wise sum.

The structure of the discriminator network D consists of three convolutional layers of 32, 64, and 128 filters, as shown in Fig. 3. The kernel size is set to $3 \times 3 \times 3$ with stride 1 in all convolution layers. LeakyReLU is used after the 3D convolution layers. A fully connected layer is then employed to produce the discriminant result.

Based on the output of the two branches, a fusion step is then performed to obtain the final lung nodule segmentation image of the proposed *LNHG* model. Based on the lung nodule candidates generated by the second branch, the connected domains of potential lung nodules that were generated by the first branch network with the same position as these candidates are considered as true nodules. Other connection domains on the image generated by the first branch network are correspondingly removed. In addition, if the Faster R-CNN branch network fails to identify lung nodule on CT image (empty image output), the image generated by the first branch network is used as the final lung nodule segmentation image, and vice versa. A detailed explanation of the workflow of the proposed model is presented in https://github.com/JD910/LNHG#supp_s1

E. Loss Function

In this study, a hybrid loss based on multiple networks is proposed to improve network performance. For the WGAN-GP, in addition to the WGAN-GP losses of L_{Gloss} and L_{Dloss} , two other network losses are simultaneously used in this study: mean square error (MSE) loss and perceptual loss. For the Faster R-CNN, GIoU loss was introduced to replace smooth L1 loss to improve the nodule detection efficiency. The calculation of loss function of WGAN-GP is presented in Fig. 4.

1) L_{Gloss} and L_{Dloss} : According to (1), the loss function of the generator network G is formulated as $L_{Gloss} = -\mathbb{E}_{x \sim P_g}[D(G(x))]$. L_{Dloss} is the sum of the losses of the $R(x)$ image, generated image $G(x)$, and GP image when passing through the discriminator network, as shown in (1) and in Fig. 4.

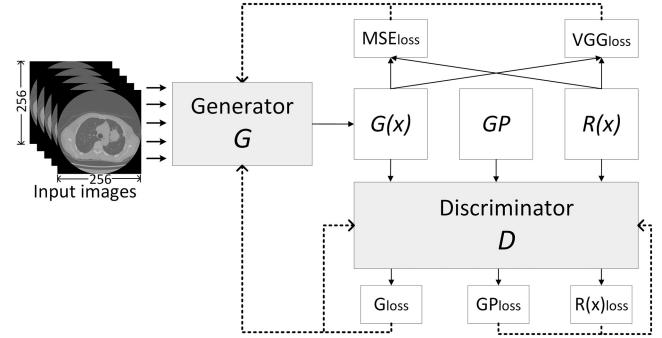


Fig. 4. Loss function of WGAN-GP in the proposed *LNHG* model. $G(x)$ denotes the generated image by generator network, $R(x)$ represents manual segmentation by radiologists, and GP denotes the image obtained using the gradient penalty function in the WGAN-GP framework. Loss backward propagation is represented by dotted line.

2) MSE Loss: The MSE loss function is the most common loss function for pixel-level transform tasks, which minimizes the pixelwise differences between the $R(x)$ image and the generated image $G(x)$. The MSE loss is calculated as follows.

$$L_{MSE} = \frac{1}{whd} \|G(x) - R(x)\|^2 \quad (8)$$

where w , h , and d represent the dimensions of the image. $G(x)$ and $R(x)$ denotes the generated image using the proposed model and the manual segmentation of lung nodule, respectively.

3) VGG Loss: Perceptual loss can be used to evaluate the high-frequency details of an image, which addresses the shortcomings of MSE loss that only concentrates on pixel-level information. The well-known ImageNet pre-trained VGG-19 network [33] was adopted as the perceptual feature extractor. The VGG-19 network comprises 16 convolutional layers, and the subsequent three layers are fully connected. The output of the 16th convolutional layer is the feature extracted by the VGG network and is used in the perceptual loss function. Thus, only the first 16 layers are used as the feature extractor. Because the pretrained VGG network accepts color images as input, the grayscale CT images were duplicated to make RGB channels before they were fed into the VGG network. The specific perceptual loss function based on the VGG network is expressed as follows.

$$L_{VGG} = \frac{1}{whd} \|VGG(G(x)) - VGG(R(x))\|_F^2 \quad (9)$$

where VGG is the feature extractor, $\|\cdot\|_F^2$ denotes the Frobenius norm; and w , h , and d represent the dimensions of the feature maps. For convenience, the perceptual loss computed by VGG-19 is defined as L_{VGG} .

Finally, the loss of the generator of WGAN-GP in the hybrid loss function of the proposed *LNHG* model is expressed as follows.

$$LNHG_{Gloss} = \lambda_1 L_{MSE} + \lambda_2 L_{VGG} + \lambda_3 L_{Dloss} \quad (10)$$

For comparison, six other networks were also trained:

- L-MSE:** *LNHG* model only with MSE loss in WGAN-GP to validate the performance of the hybrid loss design used in this study.

- *L-VGG*: *LNHG* model only with VGG loss in *WGANGP* to validate the performance of the hybrid loss design.
- *L-RF*: *LNHG* model removed Faster R-CNN module to show the output image of *LNHG* without false-positive reduction branch.
- *U-net*
- *NoduleNet*[18]
- *nnU-net (3D)* [34]

III. EXPERIMENTS

A. Datasets

A total of 1018 cases from the LIDC-IDRI dataset were used in this study, and the ROI of each lung nodule on CT images was manually delineated by one to four radiologists. Based on the latest LIDC nodule size report [35], all the 2635 lung nodules collected from the LIDC-IDRI dataset were used in this study, including a total of 14266 lung nodule CT images. In addition, a total of 3200 lung nodules from 1775 patients were collected from five first-class hospitals in China. The ROI of each lung nodule (mask = 1, otherwise = 0) on the CT images was delineated slice by slice by at least two experienced radiologists using ITK-Snap (v3.6.1), and a total of 37053 lung nodule CT images were included.

Entire CT images were used as input for the end-to-end study design. Using the original DICOM images, with a size of 512×512 , from the abovementioned datasets would result in time-consuming calculations, even on GPU-based computing hardware. Thus, both the $O(x)$ and $R(x)$ images were downsampled to a size of 256×256 . The minimum diameter of a lung nodule after downsampling was defined as 3 mm.

Data normalization on CT image was performed to enhance image contrast when constructing the dataset files. A total of 38758 CT images (including 5835 lung nodules) were used in this study. Fivefold cross-validation was used to divide all the nodules into training and test datasets in this study, and the model was then validated on an external test dataset. All the nodule images were divided into volumes with a shape of $3 \times 256 \times 256$, i.e., each input data (volume) was consisted of three continuous CT images for 3D convolution. For each nodule, if there were less than three CT images when composing the last volume, the last nodule CT image will be used to fill in. However, the order of the internal CT slices remained unchanged.

To test the performance of the *LNHG* model on an external dataset, 100 lung nodules from 91 patients from Stanford University Hospital, with a total of 430 CT images, were collected. Radiologists used OsiriX to manually delineate the lung nodules one slice at a time. In addition, lung segmentation masks of the external test dataset were manually delineated by our local radiologists for *NoduleNet*.

B. Evaluation Metrics

DC [36] was used to measure the segmentation accuracy of lung nodules.

$$DC_{R,G} = (2 |R \cap G| / (|R| + |G|)) * 100\% \quad (11)$$

where $DC_{R,G}$ is the overlap between R and G , the two clustered nodule volumes (target and generated images). The segmentation results obtained by the *LNHG*, *L-MSE*, *L-VGG*, *L-RF*, *U-net*, *NoduleNet*, and *nnU-net* models were compared with the golden standard, $R(x)$ images provided by radiologists, to calculate the DC.

In addition, a comparison of lung nodule detection accuracy in the above-mentioned methods was conducted. Precision-recall was used to evaluate the performance of the models on the external test dataset. The details of metrics to evaluate lung nodule detection and the corresponding results are presented in https://github.com/JD910/LNHG/Supplementary_S2.

C. Validation of Intra-Nodular Heterogeneity

In addition to the measurements of segmentation accuracy, we further explored the authenticity and credibility of the intra-nodular heterogeneity generated by using the proposed approach and validated its value for clinical diagnosis.

First, based on the intra-nodular heterogeneity information generated by the *WGANGP* branch network in this study, an ad-hoc analysis was conducted to verify its stability and validity. By fixing the output of the second branch network, the following training strategies, with epoch of 80, 100, 140, 160, and 220, were used to train the first branch network. The training strategies were represented as (1) to (5) for clarity, and the corresponding segmentation results were collected for performance comparison.

Second, a reader study was conducted to evaluate the potential of clinical application of the intra-nodular heterogeneity images generated. A total of 100 pathologically confirmed benign and 100 malignant lung nodules were selected for the reader study. The original CT images and corresponding generated intra-nodular heterogeneity images of all the nodules were provided to three local radiologists with more than 5-year radiological experience. The three radiologists classified the nodules and scored whether the intra-nodular heterogeneity images helped them make a correct diagnosis. That is, a score of zero indicates that the images are not useful, whereas a higher score indicates that the radiologist considered that the intra-nodular heterogeneity images were more important for correct diagnosis. A score of 100 indicates the radiologist considered that correct diagnosis could be obtained on the intra-nodular heterogeneity images without the help of the original slices.

D. Training Details

All the models were implemented using PyTorch (version: 1.7.1). A computer with two Intel Core i9 processor and four 12 GB TITAN XP graphic cards was used for training and test in cross-validation. The trained model was then provided to the independent external test dataset. A pre-experiment was conducted using 2000 randomly selected CT images to determine the hyper-parameters under the experimental condition, which are listed as follows. For the *WGANGP* network, a batch size of 12 was used; the number of training iterations was set to 160; the Adam optimizer was used with a variable learning rate between $1e-4$ and $1e-7$; the parameters for the Adam optimizer

TABLE I

AVERAGE DC OF THE MODELS (BASED ON THE NODULES ON THE $R(x)$ IMAGE AND THE CORRESPONDING NODULES ON THE GENERATED IMAGE)

	<i>LNHG</i>	<i>U-net</i>	<i>NoduleNet</i>	<i>nnU-net</i>	<i>L-MSE</i>	<i>L-VGG</i>
Training	87.30%	92.88%	86.39%	76.90%	72.12%	73.81%
Test	82.05%	83.45%	81.22%	73.25%	68.36%	69.65%

Note that nodulenet is on the lung16 (from LIDC-IDRI) dataset

were set to $\alpha = 5e-5$, $\beta_1 = 0.5$, and $\beta_2 = 0.9$; the size of the input CT image was set to 256×256 . For the Faster R-CNN, a batch size of 20 was used; the number of training epochs was set to 100; the Adam optimizer was used, and the learning rate of RPN was set as $1e-5$ with a weight decay of $5e-4$. To take advantage of Faster R-CNN in detecting small objects, anchor scales of [32, 64, 128], which are suitable for RN detection of lung nodules, were used. The parameters of λ_1 , λ_2 , and λ_3 in (10) were 1.0, 1.0, and $1e-1$, respectively; and λ in (1) was set to 10. In addition, cross-validation was used to the other methods for comparison experiment. The setting of batch size, training epoch, and other hyperparameters of the other methods followed the previous works [18], [34], [37]. The code used in this study is published on GitHub and available at the following URL. <https://github.com/JD910/LNHG>, including the WGAN-GP for lung nodule segmentation and intra-nodular heterogeneity image generation, and the Faster R-CNN for false-positive nodule reduction.

IV. RESULTS

Comprehensive experiments were conducted to test the proposed *LNHG* model, and that of the *L-MSE*, *L-VGG*, *L-RF*, *U-net*, *NoduleNet*, and *nnU-net* models.

The results on the entire training dataset indicated that the *LNHG* model obtained an average DC of 87.30% compared to the manual segmentation by radiologists. The average DC of the test dataset was 82.05%. No significant performance reduction was observed between the training and test datasets using the *LNHG* model ($P = 0.352$, *t test*). According to the results of the *U-net* model, an average DC of 92.88% was realized on the training dataset, and an average DC of 83.45% was obtained on the test dataset ($P < 0.05$, *t test*). The DCs of the *NoduleNet* and *nnU-net* models were 86.39% and 76.90% on the training dataset and 81.22% and 73.25% on the test dataset, respectively ($P > 0.05$, *t test*). However, the DCs of other WGAN-GP-based models were significantly lower compared with the models mentioned above. The average DC of the *L-MSE* and *L-VGG* models on the test dataset was 69.07%. The segmentation accuracy increased by 12.98% when the proposed hybrid loss function ($P < 0.01$ between the two segmentation results by *t test*) was used, as shown in Table I.

The segmentation results of lung nodules using *LNHG*, *U-net*, *NoduleNet*, *nnU-net*, and that achieved by the radiologists are presented in Fig. 5. Note that this study was performed in an end-to-end manner; that is, nodules were segmented directly from the original CT images. The intercepted nodule patches are shown in the figures to better display the segmentation results. The pseudo-color images represented in Fig. 5(c)–(f)

TABLE II

SEGMENTATION DC OF SOLID NODULE, JUXTA-PLEURAL NODULE, JUXTA-VASCULAR NODULE, AND GGN BY *LNHG*, *U-NET*, AND *NNU-NET* MODELS ON THE TEST DATASET. OTHER DENOTES THE AVERAGE OF *L-MSE* AND *L-VGG* MODELS

DC	<i>LNHG</i>	<i>U-net</i>	<i>nnU-net</i>	Other (avg.)
<i>Solid</i>	83.96%	84.32%	73.37%	67.02%
<i>Juxta-pleural</i>	82.30%	82.20%	72.20%	69.30%
<i>Juxta-vascular</i>	81.02%	82.52%	72.96%	70.20%
<i>GGN</i>	82.88%	83.33%	73.01%	66.18%

NoduleNet is not included since the different test dataset. (calculated based on the nodules on the $R(x)$ image and the corresponding nodules on the generated image)

were acquired by intra-nodular heterogeneity rendering of the lung nodule segmentation images output by the *LNHG*, *U-net*, *NoduleNet*, and *nnU-net* models, respectively. The change in color intensity indicates the intra-nodular heterogeneity of the lung nodule obtained by each algorithm. For the intra-nodular heterogeneity, the color indicates nodular phenotype heterogeneity from strong (red and yellow) to weak (blue and green) on the generated images. To better illustrate the intra-nodular heterogeneity, a more detailed illustration is presented in Fig. 6. Four lung nodules in Fig. 5 and the corresponding intra-nodular heterogeneity images generated by using the *LNHG* model are separately enlarged to illustrate the subtle heterogeneity in detail. In addition, the results of the intra-nodular heterogeneity rendering directly on the lung nodules on the original CT images are presented for better comparison.

First, when the intra-nodular heterogeneity images were observed by the naked eye, the results indicated that for the boundary region where the CT intensity changed sharply on the four nodules (i.e., the regions where the heterogeneity changed significantly), the intra-nodular heterogeneity was more vivid (yellow to red) by using the *LNHG* model, as shown in Fig. 6b. Although the subtle intra-nodular phenotypes in Fig. 6a(1)–a(3) are difficult to judge by the naked eye, it could be found that the grayscale changes on the original CT image in Fig. 6b(4) were consistent with the phenotypic changes on the intra-nodular heterogeneity images. Compared with other areas of the nodule, the two regions indicated by arrows exhibited a lower gray level and a smaller change in CT intensity, as shown in Fig. 6a(4). The weak intra-nodular heterogeneity expression (green to blue) in Fig. 6b(4) in the two nodular regions is consistent with the actual CT intensity distribution visible to the naked eye in Fig. 6a(4). Further, when the same intra-nodular heterogeneity rendering was directly performed on the lung nodules on the original CT images, the comparison between Fig. 6b and (c) demonstrated that the intra-nodular heterogeneity captured by the *LNHG* model could significantly reflect much more heterogeneous activities than the information conveyed by the original CT image.

Then, the stability of the intra-nodular heterogeneity images generated by the *LNHG* model was investigated, as shown in Fig. 7. The results show that when using training strategies (1) and (2), i.e., when the training epochs were lower than 100, the segmentation for adhesion-type nodules was poor, as shown in Figs. 7(5)–(9). When using the training strategies (3), (4), and

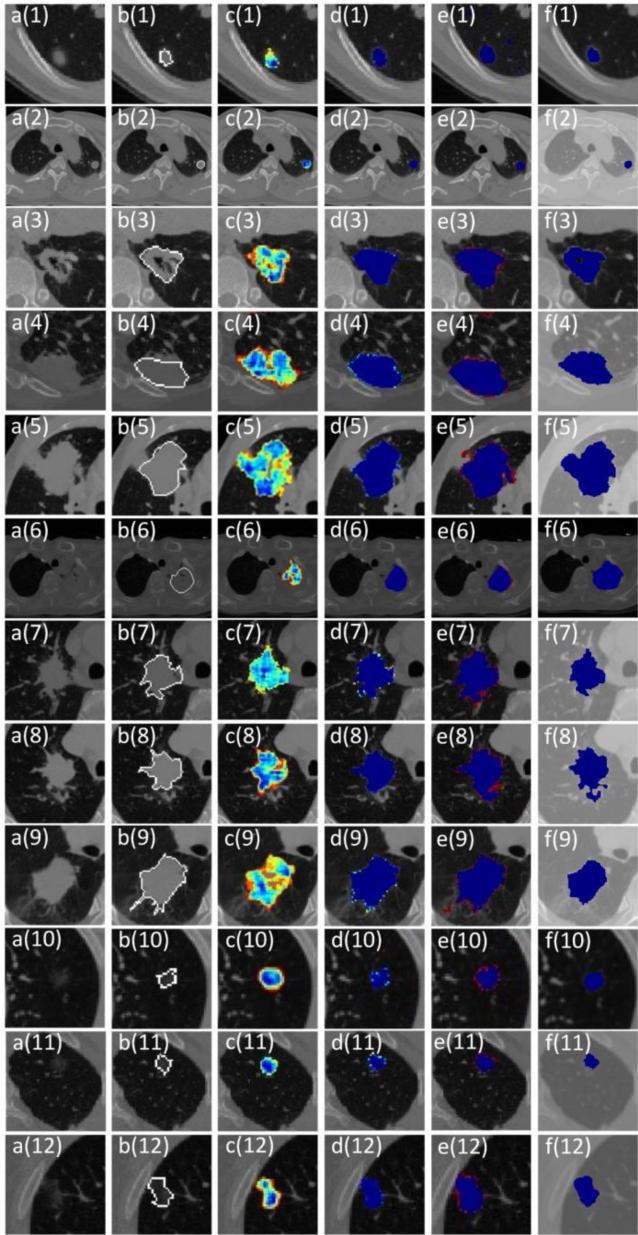


Fig. 5. Segmentation results of solid nodules (1)–(3), juxta-pleural nodules (4)–(6), juxta-vascular nodules (7)–(9), and GGNs (10)–(12) by radiologists (**b**), *LNHG* model (**c**), *U-net* (**d**), *NoduleNet* (**e**), and *nnU-net* (**f**). Lung mask of the CT images in this figure was delineated by our radiologists for *NoduleNet*. **c**, **d**, **e**, and **f** denote the intra-nodular heterogeneity images derived from the lung nodule images output by each model. **(a)** represents the original CT image. For the images generated by the models, the color indicates nodular phenotype heterogeneity on the generated images from strong (red and yellow) to weak (blue and green).

(5), the proposed *LNHG* model was able to generate accurate lung nodules, and the intra-nodular heterogeneity generated by the model tended to be stable and consistent. Furthermore, the results of the reader study indicated that the average score (with standard deviation) of the three radiologists in the reader study were 78.4 (15.63), 66.2 (22.07), and 69.0 (10.30), respectively. Results indicated that all the three radiologists considered the intra-nodular heterogeneity images generated by the proposed

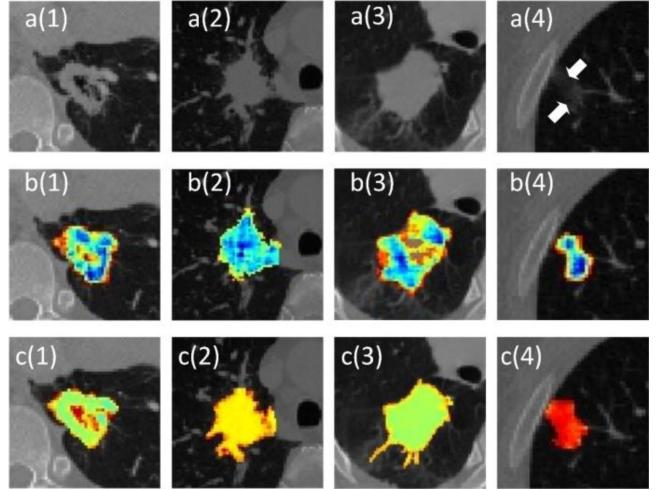


Fig. 6. Illustration of the original lung nodule CT images (in the first row) and the corresponding intra-nodular heterogeneity generated by using the proposed model (in the second row), and the result of the intra-nodular heterogeneity on the lung nodules on the original CT images (in the third row). The four lung nodules are selected from Fig. 5.

method can promote better clinical diagnosis of lung nodules. Examples of the intra-nodular heterogeneity image of benign and malignancy nodules are presented in https://github.com/JD910/LNHG#supp_s3.

To further investigate the performance of the models on each subtype of lung nodules, the segmentation of the GGN, solid, juxta-pleural, and juxta-vascular nodules in the test dataset (one of the test sets in the fivefold cross-validation was used, and the types of lung nodules were distinguished one by one by our radiologists) by the above models were analyzed individually, as shown in Table II. For each subtype, the results indicated that the performance of *LNHG* and *U-net* in this study was better than *nnU-net* in terms of segmentation accuracy, and there was no significant difference in the segmentation accuracy between *U-net* and *LNHG* on the test dataset (all $P > 0.05$ for the four subtypes of nodules). In addition, the results on the external test dataset indicated that the average DC of *LNHG*, *U-net*, *NoduleNet*, and *nnU-net* was 81.61%, 82.80%, 80.53%, and 72.17%, respectively. The results based on the training, internal test, and external test datasets showed that *U-net* achieved better accuracy than other models for lung nodule segmentation. The segmentation of GGN, solid, juxta-pleural, and juxta-vascular nodules from the external test dataset is presented in Figs. 8–11.

The measurement of P , R , and $f1$ achieved by the *LNHG*, *U-net*, *NoduleNet*, *nnU-net*, and *L-RF* models indicate that the proposed model achieved significant improvement for nodule detection than *U-net*, *nnU-net*, and *L-RF* models ($P < 0.01$), as shown in https://github.com/JD910/LNHG/Supplementary_S2. In addition, similar accuracy for lung nodule detection was found between *LNHG* and *NoduleNet* models, which the module to reduce false-positive candidates is designed in the two models. Examples of lung nodule detection and segmentation on the original CT images by *LNHG*, *L-RF*, *U-net*, *NoduleNet*, and *nnU-net* are presented in Fig. 12.

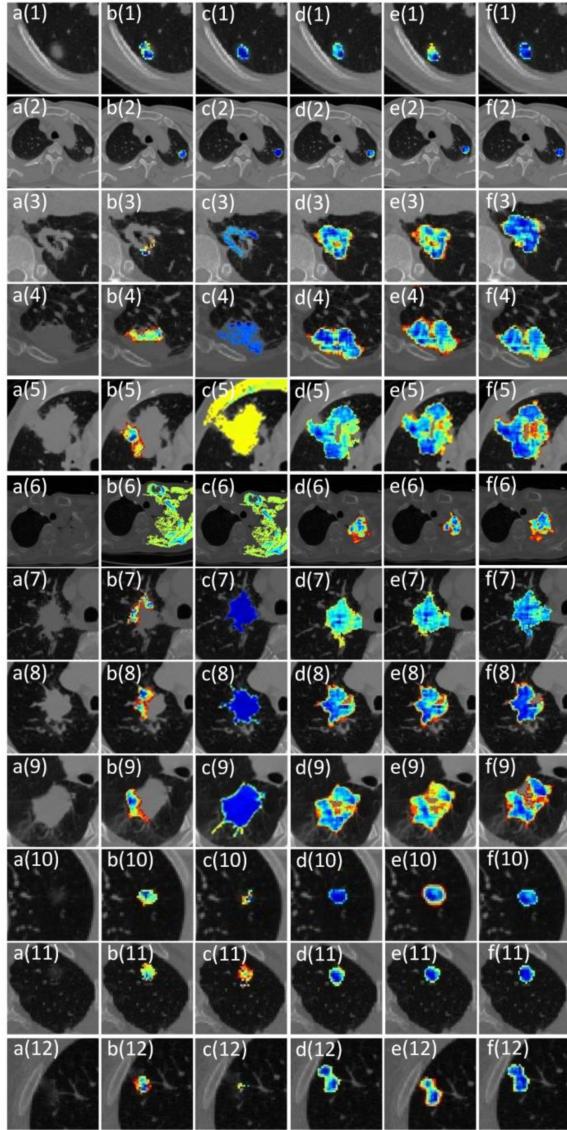


Fig. 7. Segmentation results of lung nodules (in Fig. 5) using different training strategies. (a) represents the original lung nodule CT images; (b), (c), (d), (e), and (f) represent the lung nodule segmentation results and the generated heterogeneity using the above-mentioned training strategies (1) to (5).

V. DISCUSSION

In this paper, an end-to-end network for both accurate lung nodule segmentation on CT images and intra-nodular heterogeneity generation was proposed to promote the application of lung nodule segmentation methods in radiological diagnosis. The experimental results, based on a total of 2884 cases acquired by different scanning machines from multiple centers, demonstrate that the algorithm achieved credible segmentation accuracy for solid, GGN, and adhesive lung nodules on CT images. In addition, it held promise to improve the current lung nodule segmentation paradigm by generating vivid and valid intra-nodular heterogeneity information for clinical use. Moreover, the reader study demonstrated that the intra-nodular

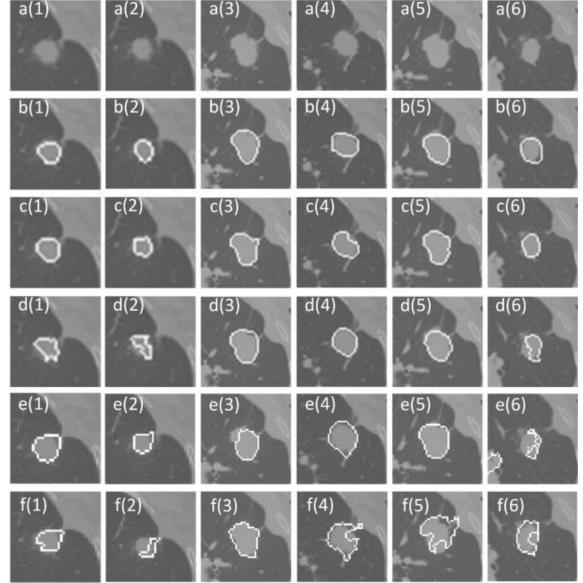


Fig. 8. Solid nodule from the external test dataset. (a) represents the original CT images. ROIs of segmentation in (b)–(f) are from radiologists and *LNHG*, *U-net*, *NoduleNet*, and *nnU-net* models, respectively.

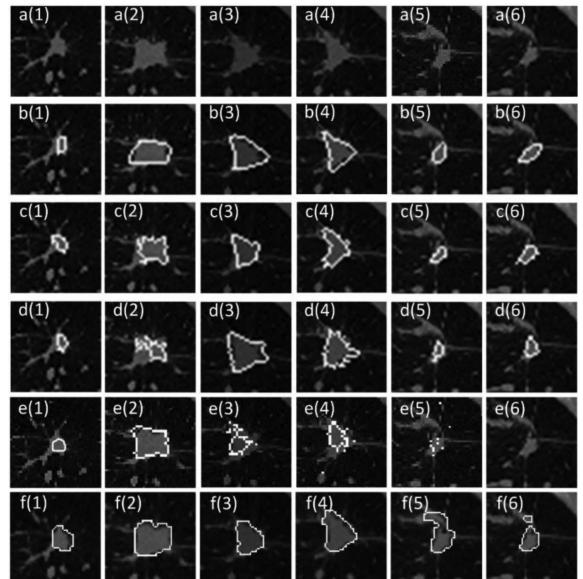


Fig. 9. Juxta-vascular nodule from the external test dataset. (a) represents original CT images. ROIs of segmentation in (b)–(f) are from radiologists and *LNHG*, *U-net*, *NoduleNet*, and *nnU-net* models, respectively.

heterogeneity images generated by the proposed approach have the potential to facilitate radiological diagnosis of lung nodules.

This study has indicated that a hybrid loss design based on the integration of WGAN-GP and Faster R-CNN can contribute to accurate lung nodule segmentation with fewer false-positive nodules. For lung nodule segmentation, the results of this study revealed that WGAN-GP is potentially an effective network as it obtained a competitive DC compared with U-net-based methods and *NoduleNet*. Compared with previous studies [6], [8], [21], [34], our study involved considerably more data from both

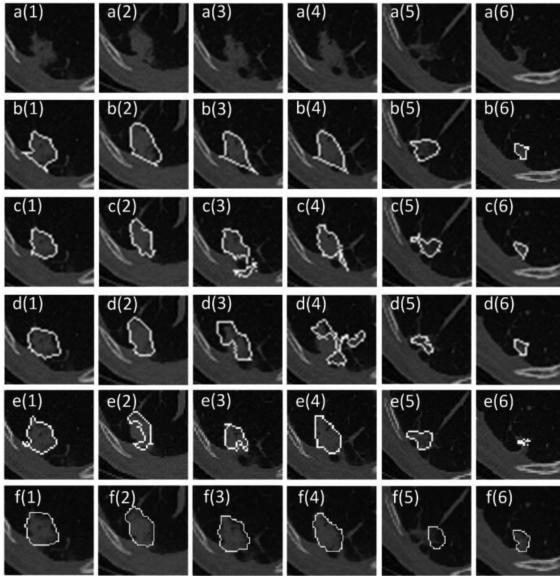


Fig. 10. Juxta-pleural nodule from the external test dataset. (a) represents original CT images. ROIs of segmentation in (b)–(f) are from radiologists and *LNHG*, *U-net*, *NoduleNet*, and *nnU-net* models, respectively.

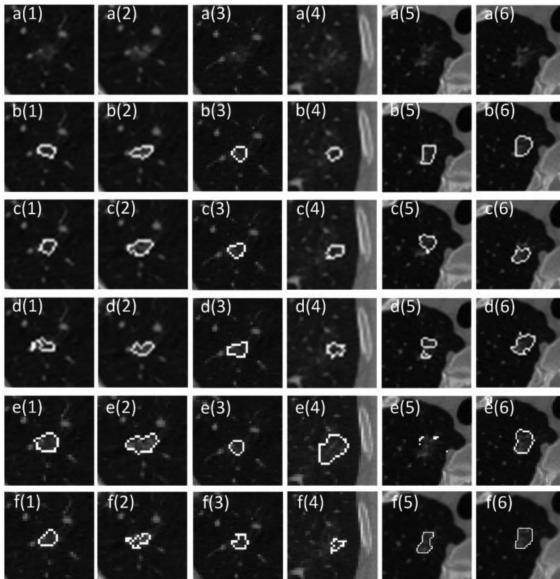


Fig. 11. GGNs from the external test dataset. (1)–(3) and (4) denote two GGNs from one patient, and (5)–(6) denote a GGN from another patient. (a) represents the original CT images. ROIs of segmentation in (b)–(f) are from radiologists and *LNHG*, *U-net*, *NoduleNet*, and *nnU-net* models, respectively.

the LIDC-IDRI dataset and five participating hospitals, and all the models achieved a DC of >73% on the test datasets. *U-net* obtained the best model generalizability on the datasets from different sources, with a DC reduction of 9.43% on the test dataset. However, based on the automatic network topology optimization, our results indicated that the *nnU-net* variant of *U-net* reduced the overfitting. These above findings show that *U-net* is a malleable baseline architecture for lung nodule segmentation.

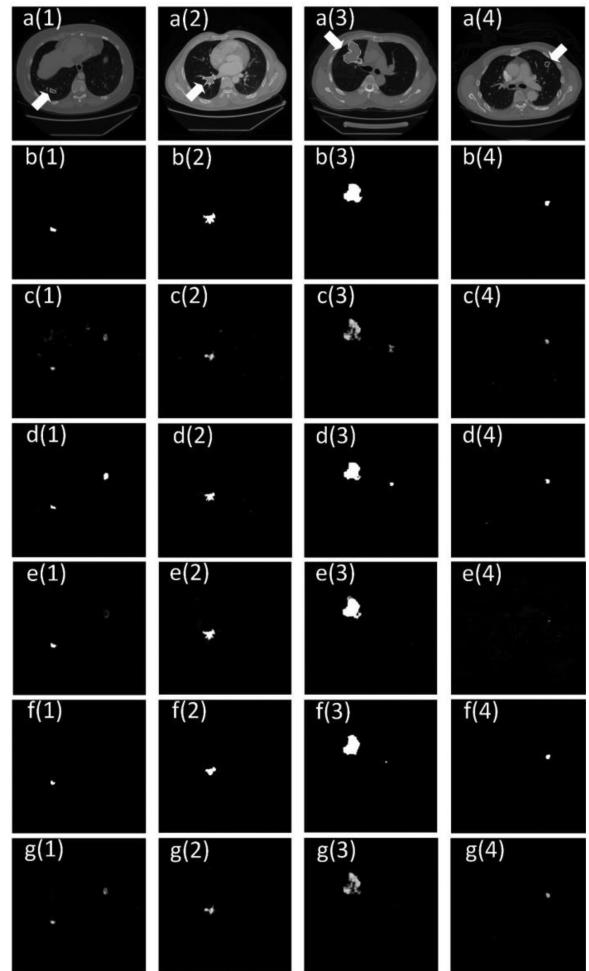


Fig. 12. Segmentation images of solid (1), juxta-vascular (2), juxta-pleural (3), and GGO (4) nodule segmented by the radiologist (b), *L-RF* model (c), *U-net* model (d), *NoduleNet* model (e), *nnU-net* model (f), and *LNHG* model (g). The manual delineation by radiologist in the original CT image are highlighted in white and arrow in (a).

However, overfitting has been proven a challenge that affects the performance of neural networks from both perspectives of methodological studies and application studies on lung nodule segmentation [38], [39]. Therefore, for the *U-net*-based models, *NoduleNet*, and the proposed *LNHG* model, consideration should be given to reducing the potential risk of overfitting in future segmentation studies.

Previous studies have proven that adding GIoU loss to R-CNN-based architecture is effective in eliminating fake candidates [28], [29]. Faster R-CNN with GIoU loss along with the segmentation network was used in this study to reduce false positives that may be learned by WGAN-GP, thereby avoiding the mis-segmentation caused by the imbalance of pixels between target and non-nodular samples on CT images. Based on the comparison between our method and the two *U-net*-base methods, it could be concluded that the proposed approach is superior in reducing false-positive candidates of lung nodule on the output images, which also supports its clinical use on lung nodule diagnosis. By adding a specific lung nodule detection

module, *NoduleNet* achieved a competitive detection accuracy to our method. However, advance segmentation of lung masks in *NoduleNet* indicates additional manual interaction. Therefore, the end-to-end design of the proposed approach has the advantage of simplifying the segmentation process, which makes it easier to apply to clinical practice.

To reduce the potential bias introduced by human intervention, image processing, and feature engineering in the workflow of intra-nodular heterogeneity assessment, we demonstrated that stable and clinically available intra-nodular heterogeneity could be provided by the *LNHG* model, compared with that of the *U-net*, *NoduleNet*, and *nnU-net* models. In particular, we proved the correlation between the subtle heterogeneity inside the lung nodules captured by the *LNHG* model and the CT phenotype, and demonstrated the stability and validity of the generation of intra-nodular heterogeneity based on the different training strategies. As shown in Fig. 6, in the generated intra-nodular heterogeneity images, a stronger heterogeneity expression was presented where the nodular phenotype intensely changed such as the boundary and the intra-nodular regions with non-uniform grey levels. This finding indicated that the subtle changes in the intensity inside the nodules, which are difficult to recognize by the human eye, were captured by the *LNHG* model. In addition, the comparison of the intra-nodular heterogeneity from the images generated by *LNHG* and that of from the original CT images in Fig. 6 indicated that the proposed model could provide more heterogeneous information for clinical diagnosis. As there are no standard metrics for evaluating the intra-nodular heterogeneity images, a clinical diagnosis experiment was designed by providing the images to radiologists to verify the clinical value of the intra-nodular heterogeneity images in assisting lung nodule diagnosis. Based on the evaluation from three radiologists, an average score of 71.2 was obtained by the radiologists. This finding indicated that the radiologists believed that using the intra-nodular heterogeneity images for radiological diagnosis may play a positive role in radiologists' diagnosis of lung nodules.

Automatic detection and generation of intra-tumoral heterogeneous activity is of significance for clinical practice, especially in lung cancer. Existing studies have shown its potential for diagnosis and prognosis by evaluating intra-tumoral heterogeneity based on the pseudo-color images of gray level distribution or phenotypic texture features on radiographic images [40], [41]. However, the radiological image-based intra-tumoral heterogeneity is susceptible to the differences in scanning conditions, image preprocessing, and feature engineering [11]. Therefore, it is difficult to establish a consensus for evaluating intra-tumoral heterogeneity in the field of radiology [42]. The results of this study are expected to overcome the above challenges. Automatic lung nodule segmentation will avoid the inconsistencies in delineating the ROI by different radiologists, and the authentic intra-nodular heterogeneity holds potential to inform a unified heterogeneous knowledge in this field.

The end-to-end design for automatic lung nodule segmentation as well as intra-nodular heterogeneity generation on CT images is an immediate need in this field of study. Although end-to-end segmentation models have been proposed in a previous

work [17], only the segmentation result of solid nodules was reported, and other types of nodules such as juxta-pleural, juxta-vascular, and GGN were not analyzed. The recent *nnU-net* is promising to be a general method for end-to-end segmentation, but only 63 lung nodules were reported for lung nodule segmentation in [34]. This study used approximately 40000 CT images to achieve the automatic segmentation of four existing lung nodule types, with an average DC of >82% on the test dataset. Previous studies indicated that the WGAN network may generate images that look natural, yet cause a severe distortion for medical diagnostics [43]. Thus, a hybrid loss function including MSE, VGG, and GIoU with multi-network was used to ensure high quality and precise location of the generated lung nodule.

There are limitations in this study. The datasets used in this study may not cover all lung nodule types; thus, the segmentation accuracy still needs to be improved. In the future, this study will be extended to include more external datasets for the automatic segmentation of lung nodules. In addition, the mechanism for generating intra-nodular heterogeneity images and the standard metrics of image evaluation are not yet clear. Therefore, the network topology of WGAN-GP and the quantitative evaluation of intra-nodular heterogeneity images should be further explored. Finally, the result indicates that certain small nodules (<3 mm) cannot be accurately detected because the proportion of small lung nodules on the CT image is slight relative to the entire image. Therefore, in addition to the existing four sub-types (i.e., GGN, solid, juxta-pleural, and juxta-vascular nodules), more detailed sub-categories of lung nodules by type, size, and scale should be created in the future, and a specific segmentation algorithm for each sub-category should be designed to improve accuracy.

In conclusion, an end-to-end model was presented for the automatic lung nodule segmentation as well as intra-nodular heterogeneity image generation from CT images, which is a new paradigm to effectively utilize the WGAN-GP network and Faster R-CNN with GIoU loss. This approach effectively addressed the challenges of inevitable human interaction and pre-processing procedures that were necessary for the existing methods in the segmentation task of lung nodules. The proposed method exceeded the existing methods in terms of vivid and valid intra-nodular heterogeneity information generation and achieved competitive results for segmentation of lung nodules on multiple datasets. This study confirmed the effectiveness of the WGAN-GP network for automatic lung nodule segmentation, addressed the challenge of poor segmentation accuracy of adhesive types of lung nodules, and achieved stable performance for CT images of varied quality and different types of lung nodules. The generation of the delicate intra-nodular heterogeneity information on CT images paves the way for the future application of automatic lung nodules segmentation methods in clinical diagnosis.

REFERENCES

- [1] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021.

- [2] A. Kumar, M. Fulham, D. Feng, and J. Kim, "Co-learning feature fusion maps from PET-CT images of lung cancer," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 204–217, Jun. 2019, doi: [10.1109/TMI.2019.2923601](https://doi.org/10.1109/TMI.2019.2923601).
- [3] G. Wu *et al.*, "Sparse representation-based radiomics for the diagnosis of brain tumors," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 893–905, Apr. 2018, doi: [10.1109/TMI.2017.2776967](https://doi.org/10.1109/TMI.2017.2776967).
- [4] J. Song *et al.*, "A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy," *Clin. Cancer Res.*, vol. 24, no. 15, pp. 3583–3592, Aug. 2018, doi: [10.1158/1078-0432.CCR-17-2507](https://doi.org/10.1158/1078-0432.CCR-17-2507).
- [5] T. Messay, R. C. Hardie, and T. R. Tuinstra, "Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset," *Med. Image Anal.*, vol. 22, no. 1, pp. 48–62, May 2015, doi: [10.1016/j.media.2015.02.002](https://doi.org/10.1016/j.media.2015.02.002).
- [6] J. Song *et al.*, "Lung lesion extraction using a toboggan based growing automatic segmentation approach," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 337–353, Jan. 2016, doi: [10.1109/TMI.2015.2474119](https://doi.org/10.1109/TMI.2015.2474119).
- [7] M. M. Farhangi, H. Frigui, A. Seow, and A. A. Amini, "3-D Active contour segmentation based on sparse linear combination of training shapes (SCoTS)," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2239–2249, Nov. 2017, doi: [10.1109/TMI.2017.2720119](https://doi.org/10.1109/TMI.2017.2720119).
- [8] S. Wang *et al.*, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Med. Image Anal.*, vol. 40, pp. 172–183, Aug. 2017, doi: [10.1016/j.media.2017.06.014](https://doi.org/10.1016/j.media.2017.06.014).
- [9] X. Xu *et al.*, "MSCS-DeepLN: Evaluating lung nodule malignancy using multi-scale cost-sensitive neural networks," *Med. Image Anal.*, vol. 65, 2020, Art. no. 101772.
- [10] S. Napel, W. Mu, B. V. Jardim-Perassi, H. J. Aerts, and R. J. Gillies, "Quantitative imaging of cancer in the postgenomic era: Radio (geno) mics, deep learning, and habitats," *Cancer*, vol. 124, no. 24, pp. 4633–4649, 2018.
- [11] A. Traverso, L. Wee, A. Dekker, and R. Gillies, "Repeatability and reproducibility of radiomic features: A systematic review," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 102, no. 4, pp. 1143–1158, Nov. 2018, doi: [10.1016/j.ijrobp.2018.05.053](https://doi.org/10.1016/j.ijrobp.2018.05.053).
- [12] S. G. Armato, III *et al.*, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Feb. 2011, doi: [10.1118/1.3528204](https://doi.org/10.1118/1.3528204).
- [13] Y. J. Kim, S. H. Lee, K. Y. Lim, and K. G. Kim, "Development and validation of segmentation method for lung cancer volumetry on chest CT," *J. Digit. Imag.*, vol. 31, no. 4, pp. 505–512, Aug. 2018, doi: [10.1007/s10278-018-0051-5](https://doi.org/10.1007/s10278-018-0051-5).
- [14] H. Shakir, T. M. Rasool Khan, and H. Rasheed, "3-D segmentation of lung nodules using hybrid level sets," *Comput. Biol. Med.*, vol. 96, pp. 214–226, May 2018, doi: [10.1016/j.combiomed.2018.03.015](https://doi.org/10.1016/j.combiomed.2018.03.015).
- [15] S. Mukhopadhyay, "A segmentation framework of pulmonary nodules in lung CT images," *J. Digit. Imag.*, vol. 29, no. 1, pp. 86–103, Feb. 2016, doi: [10.1007/s10278-015-9801-9](https://doi.org/10.1007/s10278-015-9801-9).
- [16] Y. Gu *et al.*, "Automatic lung nodule detection using multi-scale dot nodule-enhancement filter and weighted support vector machines in chest computed tomography," *PLoS One*, vol. 14, no. 1, 2019, Art. no. e0210551.
- [17] X. Huang, W. Sun, T. B. Tseng, C. Li, and W. Qian, "Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks," *Comput. Med. Imag. Graph.*, vol. 74, pp. 25–36, Jun. 2019, doi: [10.1016/j.compmedimag.2019.02.003](https://doi.org/10.1016/j.compmedimag.2019.02.003).
- [18] H. Tang, C. Zhang, and X. Xie, "Nodulenet: Decoupled false positive reduction for pulmonary nodule detection and segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2019, pp. 266–274.
- [19] Y. Li and L. Shen, "cC-GAN: A robust transfer-learning framework for HEp-2 specimen image segmentation," *IEEE Access*, vol. 6, pp. 14048–14058, 2018, doi: [10.1109/ACCESS.2018.2808938](https://doi.org/10.1109/ACCESS.2018.2808938).
- [20] X. Dong *et al.*, "Automatic multiorgan segmentation in thorax CT images using U-net-GAN," *Med. Phys.*, vol. 46, no. 5, pp. 2157–2168, 2019, doi: [10.1002/mp.13458](https://doi.org/10.1002/mp.13458).
- [21] C. Zhao, J. Han, Y. Jia, and F. Gou, "Lung nodule detection via 3D U-Net and contextual convolutional neural network," in *Proc. Int. Conf. Netw. Netw. Appl.*, 2018, pp. 356–361, doi: [10.1109/NANA.2018.8648753](https://doi.org/10.1109/NANA.2018.8648753).
- [22] H. Shi and J. Lu, and Q. Zhou, "A novel data augmentation method using style-based GAN for robust pulmonary nodule segmentation," in *Proc. Chin. Control Decis. Conf.*, 2020, pp. 2486–2491, doi: [10.1109/CCDC49329.2020.9164303](https://doi.org/10.1109/CCDC49329.2020.9164303).
- [23] J. P.-A. Ian J. Goodfellow *et al.*, "Generative adversarial nets," 2014, *arXiv:1406.2661*.
- [24] F. A. Ishaan Gulrajani, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
- [25] D. Zhou *et al.*, "Tou loss for 2d/3d object detection," in *Proc. IEEE Int. Conf. 3D Vis.*, 2019, pp. 85–94, doi: [10.1109/3DV.2019.00019](https://doi.org/10.1109/3DV.2019.00019).
- [26] N. T. Hamid Rezatofighi, J. Y. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 658–666, doi: [10.1109/CVPR.2019.00075](https://doi.org/10.1109/CVPR.2019.00075).
- [27] Y. Wang, X. Ning, B. Leng, and H. Fu, "Ship detection based on deep learning," in *Proc. IEEE Int. Conf. Mechatronics Automat.*, 2019, pp. 275–279, doi: [10.1109/ICMA.2019.8816265](https://doi.org/10.1109/ICMA.2019.8816265).
- [28] K. Lin *et al.*, "Face detection and segmentation based on improved mask R-CNN," *Discrete Dyn. Nat. Soc.*, vol. 2020, 2020, Art. no. 9242917, doi: [10.1155/2020/9242917](https://doi.org/10.1155/2020/9242917).
- [29] M. Hassanin, S. Khan, and M. Tahtali, "A new localization objective for accurate fine-grained affordance segmentation under high-scale variations," *IEEE Access*, vol. 8, pp. 28123–28132, 2019, doi: [10.1109/ACCESS.2019.2958608](https://doi.org/10.1109/ACCESS.2019.2958608).
- [30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2016.
- [31] R. Wang, W. Zhang, W. Nie, and Y. Yu, "Gastric polyps detection by improved faster R-CNN," in *Proc. 8th Int. Conf. Comput. Pattern Recognit.*, 2019, pp. 128–133.
- [32] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6021013/pdf/nihms973044.pdf>
- [33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [34] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nat. Methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [35] A. P. Reeves and A. Biancardi, "The lung image database consortium (LIDC) nodule size report," Cornell Univ. Vis. Image Anal. Group, Ithaca, NY, USA, Tech. Rep. 2011, pp. 10–27. [Online]. Available: <http://www.via.cornell.edu/lidc/>
- [36] R. Zhang *et al.*, "Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional densenets," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2149–2160, Sep. 2018, doi: [10.1109/TMI.2018.2821244](https://doi.org/10.1109/TMI.2018.2821244).
- [37] Z. Hao, "Implementation of deep learning framework – Unet, using Keras," Accessed: Feb. 21, 2019, [Online]. Available: <https://github.com/zhihuohao/unet>
- [38] S. H. Khan, M. Hayat, and F. Porikli, "Regularization of deep neural networks with spectral dropout," *Neural Netw.*, vol. 110, pp. 82–90, 2019, doi: [10.1016/j.neunet.2018.09.009](https://doi.org/10.1016/j.neunet.2018.09.009).
- [39] X. Huang, J. Shan, and V. Vaidya, "Lung nodule detection in CT using 3D convolutional neural networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, 2017, pp. 379–383, doi: [10.1109/ISBI.2017.7950542](https://doi.org/10.1109/ISBI.2017.7950542).
- [40] E. Sala *et al.*, "Unravelling tumour heterogeneity using next-generation imaging: Radiomics, radiogenomics, and habitat imaging," *Clin. Radiol.*, vol. 72, no. 1, pp. 3–10, Jan. 2017, doi: [10.1016/j.crad.2016.09.013](https://doi.org/10.1016/j.crad.2016.09.013).
- [41] D. Fehr *et al.*, "Automatic classification of prostate cancer gleason scores from multiparametric magnetic resonance images," *Proc. Nat. Acad. Sci.*, vol. 112, no. 46, pp. E6265–E6273, 2015, doi: [10.1073/pnas.1505935112](https://doi.org/10.1073/pnas.1505935112).
- [42] M.-C. Desserot *et al.*, "Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non-small cell lung cancer tumors: A repeatability analysis in a prospective multicenter cohort," *J. Nucl. Med.*, vol. 58, no. 3, pp. 406–411, 2017.
- [43] Q. Yang *et al.*, "Generative low-dose CT image denoising," in *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*. Berlin, Germany: Springer, 2019, pp. 277–297.