SANet: A Slice-Aware Network for Pulmonary Nodule Detection

Jie Mei, Ming-Ming Cheng, Gang Xu, Lan-Ruo Wan, and Huan Zhang

Abstract—Lung cancer is the most common cause of cancer death worldwide. A timely diagnosis of the pulmonary nodules makes it possible to detect lung cancer in the early stage, and thoracic computed tomography (CT) provides a convenient way to diagnose nodules. However, it is hard even for experienced doctors to distinguish them from the massive CT slices. The currently existing nodule datasets are limited in both scale and category, which is insufficient and greatly restricts its applications. In this paper, we collect the largest and most diverse dataset named PN9 for pulmonary nodule detection by far. Specifically, it contains 8,798 CT scans and 40,439 annotated nodules from 9 common classes. We further propose a slice-aware network (SANet) for pulmonary nodule detection. A slice grouped non-local (SGNL) module is developed to capture long-range dependencies among any positions and any channels of one slice group in the feature map. And we introduce a 3D region proposal network to generate pulmonary nodule candidates with high sensitivity, while this detection stage usually comes with many false positives. Subsequently, a false positive reduction module (FPR) is proposed by using the multi-scale feature maps. To verify the performance of SANet and the significance of PN9, we perform extensive experiments compared with several state-of-the-art 2D CNN-based and 3D CNN-based detection methods. Promising evaluation results on PN9 prove the effectiveness of our proposed SANet. The dataset and source code is available at https://mmcheng.net/SANet/.

Index Terms—Pulmonary Nodule Detection, Nodule Dataset, Slice Grouped Non-local, False Positive Reduction.

1 Introduction

Lung cancer has become one of the main causes of cancer death worldwide [1], [2]. Pulmonary nodules are the lesions in the lungs, which have a high probability of evolving into malignant tumors. Diagnosis of the pulmonary nodules at an early stage and timely treatments are the best solutions for lung cancer. The thoracic computed tomography (CT) is an effective tool for the early diagnosis of the pulmonary nodules [3], which plays an important role in reducing the mortality of lung cancer [4]. In the CT images, the absorption levels of X-ray for nodules and other tissues are often the same. However, nodules are usually isolated and spherical, which are quite different from the vessels and bronchus' continuous pipe-like structure. Since interpreting CT data requires analyzing hundreds of images at a time, an experienced doctor often takes about 10 minutes to perform a thorough examination of a patient. Moreover, there are many small nodules, and different types of nodules have different morphology. It is a big challenge for doctors to accurately identify and diagnose the malignancy of nodules [5].

Computer-aided diagnosis (CAD) systems have been developed to assist doctors in interpreting the CT images more effectively and accurately [6], [7]. Traditional CAD systems detect nodule candidates mostly relying on the morphological operations or low-level descriptors [8], [9], [10]. These methods often obtain inferior detection results due to the variety of nodules in size, shape, and types. With the development of deep learning, convolutional neural networks (CNNs) such as Faster R-CNN [11], SSD [12], and YOLO [13] have been proposed and proven to be effective in object detection. CNNs have also been introduced in the field of medical image analysis [14], [15], [16], [17], [18], [19], [20], [21], [22]. For pulmonary nodule detection, CNN-based methods [7], [23], [24], [25], [26], [27]are proven to be much more effective than the traditional methods. Compared to the 2D object detection in natural images, pulmonary nodule detection is much harder since it is a 3D object detection problem using 3D CT data. Some studies such as [26], [28] utilize 2D region proposal networks (RPNs) to obtain proposals in each 2D slice of the 3D CT images, then the 2D proposals are merged across slices to generate 3D proposals. Nowadays, more and more methods [7], [23], [24] adopt 3D CNN based models to deal with the CT data and directly generate 3D proposals. Compared to the 2D CNN, 3D CNN has much more parameters, making it need more time and more GPU memory to train. However, the performance of 3D CNN models is better than 2D ones for CT data, as a comparative study shown in [28].

With the publicly available of several CT datasets such as LIDC-IDRI [29] and LUNA16 [30], CNN-based methods have become the trend for pulmonary nodule detection. These datasets allow researchers to develop and evaluate their algorithms for nodule detection under the same evaluation metrics, further promoting the CAD systems' application in practice. However, dataset like LUNA16 [30], the most widely used dataset nowadays, only contains 888 CT scans with a limited number and categories of labeled pulmonary nodules, which is insufficient for the training of 3D CNNs and restricts its application in the diagnosis of lung cancer. Therefore a more extensive and more diverse dataset is urgently needed.

This paper introduces a new large-scale dataset, namely PN9 (pulmonary nodule with 9 categories), for pulmonary nodule detection. First, we collect CT images from two major hospitals and different scenes such as the clinic, hospitalization, and physical examination. After performing quality assurance, 8,798 CT scans from 8,798 different patients are obtained. Next, all the private health information of the patients is removed through the data

- J. Mei, M.M. Cheng, and G. Xu are with the TKLNDST, College of Computer Science, Nankai University, Tianjin 300350, China.
- L.R. Wan, and H. Zhang are with the InferVision.
- M.M. Cheng is the corresponding author (E-mail: cmm@nankai.edu.cn).

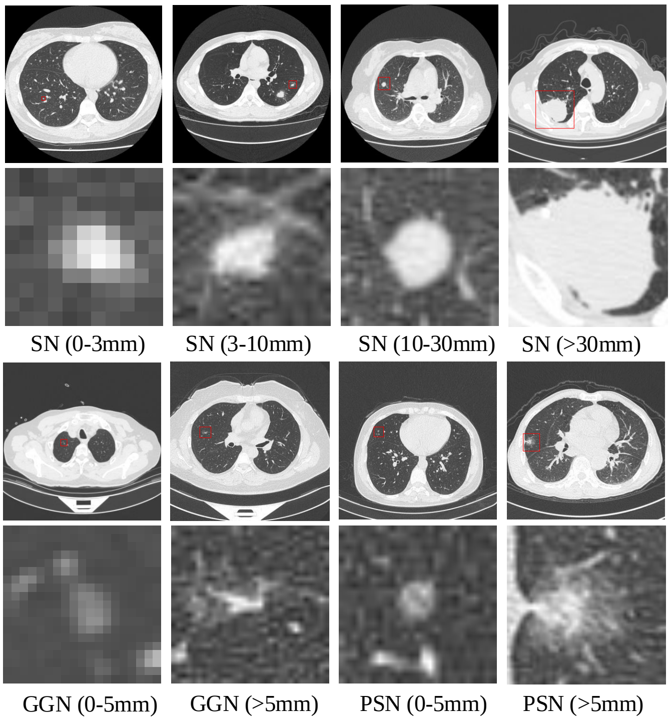

Fig. 1: Examples of the pulmonary nodules in the PN9. Each image belongs to a different class of nodules. SN, GGN, and PSN denote solid, ground-glass, and part-solid nodules, respectively. The size of each nodule is labeled in the parentheses. The 1st and 3rd rows are the complete slices, while the rest two rows are zoomed-in images, respectively.

masking process. Attending physicians then further check and annotate the CT images with the bounding boxes and category labels following a two-phase process. Finally, we obtain the PN9 dataset, including 40,439 annotated nodules, which are divided into 9 common classes. Compared with the current existing pulmonary nodule datasets, PN9 contains a larger number of CT scans and more classes, contributing to the pulmonary nodule detection and allowing researchers to design more effective algorithms based on the rich properties of nodules. Meanwhile, more small-size pulmonary nodules annotated in PN9 help diagnose small nodules more accurately and result in the earlier treatment of patients. Fig. 1 shows some examples of the labeled pulmonary nodules in PN9.

What's more, we propose a slice-aware network (SANet) for pulmonary nodule detection. We first introduce an encoder-decoder architecture network to learn the feature of nodules, since their size is much smaller than the common objects in natural images. According to doctors' diagnosis way, we propose a slice grouped non-local module (SGNL) and add it to the encoder network. SGNL is able to capture long-range dependencies among any positions and any channels of one slice group in the feature map. And 3D region proposal network is introduced to generate pulmonary nodule candidates with high sensitivity, while this detection stage usually comes with many false positives. Subsequently, we develop a false positive reduction module (FPR) by using the multi-scale feature maps. To validate the performance of SANet and the significance of the PN9 dataset, we perform extensive experiments compared with several state-of-the-art 2D CNN-based and 3D CNN-based object detection algorithms. Promising evaluation results on PN9 prove the effectiveness of our proposed SANet.

Our contributions are summarized as follows.

- We construct a new pulmonary nodule dataset, called PN9, which contains 8,798 CT scans and 40,439 annotated nodules of 9 different classes. To the best of our knowledge, PN9 is the largest and most diverse dataset for pulmonary nodule detection by far.
- We propose a slice-aware network (SANet) for pulmonary nodule detection, which mainly contains a slice-grouped non-local module and a false positive reduction module.
- Compared with previous state-of-the-art 2D CNN-based, 3D CNN-based detection methods, and experienced doctors, SANet makes full use of the characteristics in CT images and achieve better performance in several evaluation metrics.

2 Related Work

This section introduces the related works of pulmonary nodule detection and reviews the existing pulmonary nodule datasets.

2.1 Pulmonary Nodule Detection

Unlike general 2D object detection, pulmonary nodule detection is a 3D object detection problem using 3D CT images. It draws more and more attention in recent years because of its great clinical value. Traditional nodule detection methods mostly rely on hand-designed descriptors or morphological operations. Messay *et al.* [8] introduce a fully automated lung segmentation algorithm, which combines morphological processing and intensity thresholding to detect and segment lung nodule candidates simultaneously. Jacobs *et al.* [9] adopt shape, texture, intensity features, and a novel set of context features to detect subsolid pulmonary nodules. A novel work based on global segmentation methods is proposed in [31] for lung nodule candidate detection, and it is combined with simple rule-based filtering and mean curvature minimization. [10] uses manually designed filters to screen the possible pulmonary nodules in CT scans, which highly depends on professional medical knowledge. However, it is difficult for these methods to detect nodules in the complex region, especially nodules that present a high degree of vascular attachment.

With the development of deep learning, many CNNs have been proposed for object detection. Some methods have two stages, like [11], [32], [33], [34], while most recent methods [12], [13], [35], [36], [37], [38] have one stage that the bounding boxes and class probabilities are predicted simultaneously. CNN-based methods have also been introduced in field of the pulmonary nodule detection. Ding *et al.* [26] introduce a deconvolutional structure to Faster RCNN for candidate detection on axial slices. Setio *et al.* [24] propose multi-view ConvNets for pulmonary nodule detection. The inputs are a set of 2D patches from differently oriented planes. The outputs from multiple 2D ConvNets are combined using a dedicated fusion method to get the final results. These methods require post-processing to integrate 2D proposals into 3D proposals, which is inefficient and may affect the accuracy of nodule detection.

Recently, more and more studies adopt 3D CNN-based models due to the 3D nature of CT images. Dou *et al.* [27] propose a method employing 3D CNNs for nodule detection from CT scans, and introduce an effective strategy encoding multilevel contextual information to deal with the large variations and hard mimics of lung nodules. In [39], a 3D CNN with an encoder-decoder structure is developed for pulmonary nodule detection.

It also adopts a dynamically scaled cross-entropy to reduce the false positive rate and the squeeze-and-excitation structure to fully utilize channel inter-dependency. Zhu *et al.* [23] propose a 3D Faster R-CNN with 3D dual-path blocks for nodule detection and a U-Net-like [40] architecture to effectively learn nodule features. To promote further researches in this field, Liao et al. [7] adopt a 3D RPN to detect pulmonary nodules. They introduce a leaky noisy-OR gate to evaluate the cancer probabilities by selecting the top five nodules based on the detection confidences. [41] proposes a novel multi-scale gradual integration CNN to learn features of multi-scale inputs with a gradual feature extraction strategy, which reduces many false positives. An end-to-end probabilistic diagnostic system is introduced in [42], which contains a Computer-Aided Detection (CADe) module for detecting suspicious lung nodules and a Computer-Aided Diagnosis (CADx) module for patient-level malignancy classification. Harsono *et al.* [43] propose a lung nodule detection and classification model I3DR-Net, which combines the I3D backbone with RetinaNet and modified FPN framework. Song *et al.* [44] develop a 3D center-points matching detection network (CPM-Net) for pulmonary nodule detection. It automatically predicts the position and aspect ratio of nodules without the manual design of anchor parameters.

There are also some works for pulmonary nodule classification [28], [45], [46]. Before CNN is introduced, features like shape, 3D contour, and texture are widely used for nodule diagnosis [45], [47], [48]. Subsequently, a multi-scale CNN for capturing nodule heterogeneity by extracting features from stacked layers is proposed for nodule classification [49]. Inspired by the deep dual-path network (DPN) [50], Zhu *et al.* [23] propose a 3D DPN to learn the features of nodules and adopt a gradient boosting machine (GBM) for nodule classification. Additionally, Some multi-instance learning and deep transfer learning approaches are employed for nodule classification of patient-level [51], [52], [53]. For example, Hussein *et al.* [51] adopt graph regularized sparse multi-task learning to incorporate the complementary feature from nodule attributes for malignancy evaluation.

2.2 Pulmonary Nodule Datasets

Some of the pulmonary nodule datasets have been released [6], [29], [54], [55], which make it possible for researchers to develop and evaluate their CNN-based methods for pulmonary nodule detection under unified evaluation metrics. In 2010, ANODE09 was proposed by Van *et al.* [6], which only contains 55 CT scans acquired using a single scanner and scan protocol. Besides, it contains a limited number of larger nodules that generally are more likely to be malignant. Subsequently, several datasets with larger pulmonary nodules are introduced. The LIDC-IDRI dataset [29] contains 1,018 CT scans with annotations from four experienced radiologists. The nodules are divided into three categories: nodules $\geqslant 3$ mm, nodules < 3 mm, and non-nodule. Manual 3D segmentation is implemented for nodules categorized as nodules $\geqslant 3$ mm. The CT scans in LIDC-IDRI are collected from seven different academic institutions and a range of scanner models. The LUNA16 dataset [30] is collected from the LIDC-IDRI [29], where CT scans with a slice thickness greater than 3 mm are discarded. In fact, the slice thickness of all CT scans in the LUNA16 dataset [30] is less than 2.5 mm. Besides, scans with missing slices or inconsistent slice spacing are also excluded. This dataset eventually contains 888 CT scans with considering the 1,186 nodules annotated by the majority of the radiologists as positive examples. All nodules in LUNA16 are categorized as nodules $\geqslant 3$ mm. The DSB 2017 [7] contains 2,101 CT scans, while it only includes binary labels indicating whether a scan is diagnosed with lung cancer or not. However, since different types of nodules have different morphology and cancer probabilities, these datasets with few annotations and limited types of nodules are insufficient for practical application in lung cancer diagnosis.

3 Proposed Method

Different from the 2D object detection in general natural images, pulmonary nodule detection is a 3D object detection problem using 3D CT data. In order to make full use of the 3D space information between different slices, we propose a slice-aware network (SANet), as shown in Fig. 2. In SANet, a slice grouped non-local (SGNL) module is proposed to capture long-range dependencies among any positions and any channels of one slice group in the feature map. And we further develop a false positive reduction module to improve the performance of nodule detection, especially nodules of small size.

3.1 Network Structure

Encoder-Decoder Architecture. In our SANet, 3D ResNet50 is adopted as an encoder due to its outstanding performance in feature extraction [56]. However, the size of nodules varies greatly and is much smaller compared with common objects in natural images. 3D Resnet50, which encodes the CT images with five 3D convolutional blocks, can not explicitly describe the features of nodules and lead to poor performance in detecting nodule candidates. To address this problem and enable the network to capture multi-scale information, we employ a U-shaped encoder-decoder architecture [40]. The decoder network consists of two $2 \times 2 \times 2$ deconvolution layers for up-sampling the feature map to an appropriate size. Each output feature map of deconvolutional layers is concatenated with the corresponding output in the encoder network, whose channel is adjusted by a $1 \times 1 \times 1$ convolutional layer. The feature maps produced by our encoder-decoder network are defined as $\{M_{res1}, M_{res2}, M_{res3}, M_{res4}, M_{res5}, M_{de1}, M_{de2}\}$, respectively.

3D Region Proposal Network. To generate pulmonary nodule candidates, a $3 \times 3 \times 3$ convolutional layer is employed over the concatenated feature map M_{de2}. The $3 \times 3 \times 3$ convolution is followed by two parallel $1 \times 1 \times 1$ convolutional layers for regressing the 3D bounding box of each voxel (*i.e.,* Reg Layer in Fig. 2) and predicting classification probability (*i.e.,* Cls Layer in Fig. 2). Based on the distribution of nodule size, we design five anchors with sizes 5, 10, 20, 30, and 50. Each anchor is specified six regression parameters: central z-, y-, x- coordinates, depth, height, and width. The multi-task loss function is defined as:

$$\begin{aligned} L_{RPN} = L\left(\{p_i\}, \{t_i\}\right) = \frac{1}{N_{cls}} \sum_i L_{cls}\left(p_i, p_i^*\right) \\ + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}\left(t_i, t_i^*\right), \end{aligned} \tag{1}$$

where i is the index of $i-th$ anchor in one 3D patch. N_{cls} and N_{reg} are the numbers of anchors considered for computing classification loss and regression loss, respectively. λ is a parameter used to balance the two losses. p_i is the predicted probability of $i-th$ anchor being a nodule, p_i^* is 1 if the anchor is positive and

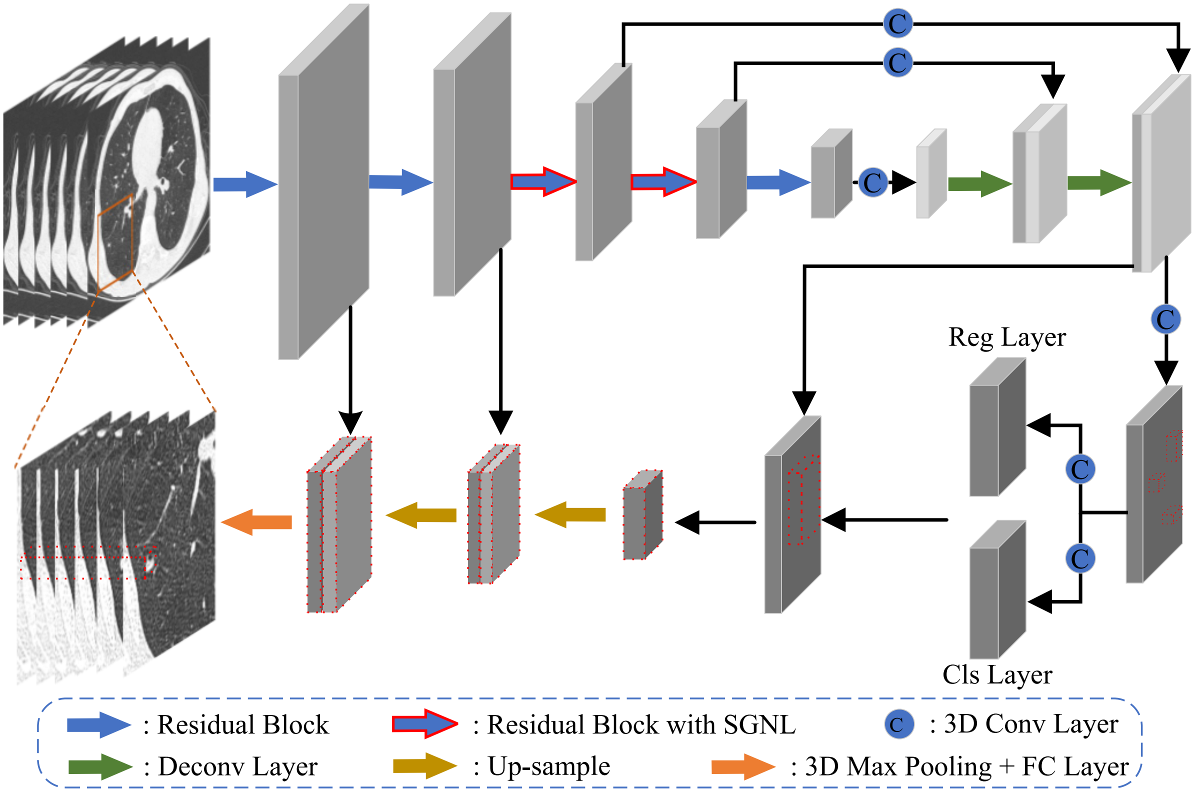

Fig. 2: Overall architecture of the proposed slice-aware network (SANet). The red dashed box represents the nodule candidate. And the CT images below are zoomed-in images of the ones above.

0 otherwise. In our work, an anchor that has Intersection-over-Union (IoU) overlap equal to or higher than 0.5 with any ground-truth nodule box is considered as positive ($p_i^* = 1$). If there are no anchors that meet the above condition, the anchor with the highest IoU overlap is assigned a positive label. On the other hand, an anchor having IoU less than 0.02 with all ground-truth boxes is considered as negative ($p_i^* = 0$). t_i is a vector denoting the predicted 6 parameterized coordinates for nodule position, and t_i^* is the ground-truth vector. For notational convenience, subscript i is ignored and t_i and t_i^* are defined as:

$$t = \left(\frac{z - z_a}{d_a}, \frac{y - y_a}{h_a}, \frac{x - x_a}{w_a}, \log\frac{d}{d_a}, \log\frac{h}{h_a}, \log\frac{w}{w_a}\right), \tag{2}$$

$$t^* = \left(\frac{z^* - z_a}{d_a}, \frac{y^* - y_a}{h_a}, \frac{x^* - x_a}{w_a}, \log\frac{d^*}{d_a}, \log\frac{h^*}{h_a}, \log\frac{w^*}{w_a}\right), \tag{3}$$

where $x, y, z, w, h,$ and d represent the predicted box's center coordinates, width, height, and depth. $x^*, y^*, z^*, w^*, h^*,$ and d^* are the parameters for the ground-truth box. $x_a, y_a, z_a, w_a, h_a,$ and d_a denote the parameters of the anchor box. What's more, we use weighted binary cross-entropy loss for L_{cls} and smooth L_1 loss [34] for L_{reg}.

3.2 Slice Grouped Non-local Module

In the thoracic CT images, vessels and bronchus are the continuous pipe-like structure, while nodules are usually isolated and spherical. To diagnose nodules from other tissues, doctors need to view multiple consecutive slices to capture the correlation among them. According to the diagnosis way of doctors, we propose a slice grouped non-local module (SGNL, as shown in Fig. 3) based on the non-local module in [57]. The SGNL can learn explicit correlations among any elements across slices.

Review of Non-local Operation. Let $\mathbf{X} \in \mathbb{R}^{D\times H\times W\times C}$ denote the input feature map for the non-local module, Where $D, H, W,$ and C represent depth, height, width, and the number of channels. The original non-local operation in [57] is defined as:

$$\mathbf{Y} = f\left(\theta\left(\mathbf{X}\right), \phi\left(\mathbf{X}\right)\right) g\left(\mathbf{X}\right), \tag{4}$$

where $\mathbf{Y} \in \mathbb{R}^{D\times H\times W\times C}$. $\theta\left(\cdot\right), \phi\left(\cdot\right), g\left(\cdot\right) \in \mathbb{R}^{DHW\times C}$ are implemented by $1\times 1\times 1$ convolution and can be written as:

$$\theta\left(\mathbf{X}\right) = \mathbf{X}\mathbf{W}_\theta,\ \phi\left(\mathbf{X}\right) = \mathbf{X}\mathbf{W}_\phi,\ g\left(\mathbf{X}\right) = \mathbf{X}\mathbf{W}_g, \tag{5}$$

where $\mathbf{W}_\theta, \mathbf{W}_\phi,$ and $\mathbf{W}_g$ are weight matrices to be learned. The function $f\left(\cdot,\cdot\right)$ is used to compute the similarity between all locations in the feature map. In [57], they describe several choices for f, where the dot-produce is probably the simplest one, *i.e.,*

$$f(\theta(\mathbf{X}), \phi(\mathbf{X})) = \theta(\mathbf{X})\phi(\mathbf{X})^\top \tag{6}$$

Slice Grouped Non-local. The origin non-local module can capture long-range dependencies among any positions in the feature map. However, the affinity between any channels is also important for discriminating the fine-grained objects as explored in [58], [59]. We consider cross-channel information in the origin non-local operation to model long-range dependencies among any positions and any channels.

We reshape the output of Eq. 5 by merging channel into position and obtain $\theta\left(\cdot\right), \phi\left(\cdot\right), g\left(\cdot\right) \in \mathbb{R}^{DHWC}$. Our SGNL operation computes the response $\mathbf{Y}$ as:

$$\mathbf{Y} = f\left(vec\left(\theta\left(\mathbf{X}\right)\right), vec\left(\phi\left(\mathbf{X}\right)\right)\right) vec\left(g\left(\mathbf{X}\right)\right), \tag{7}$$

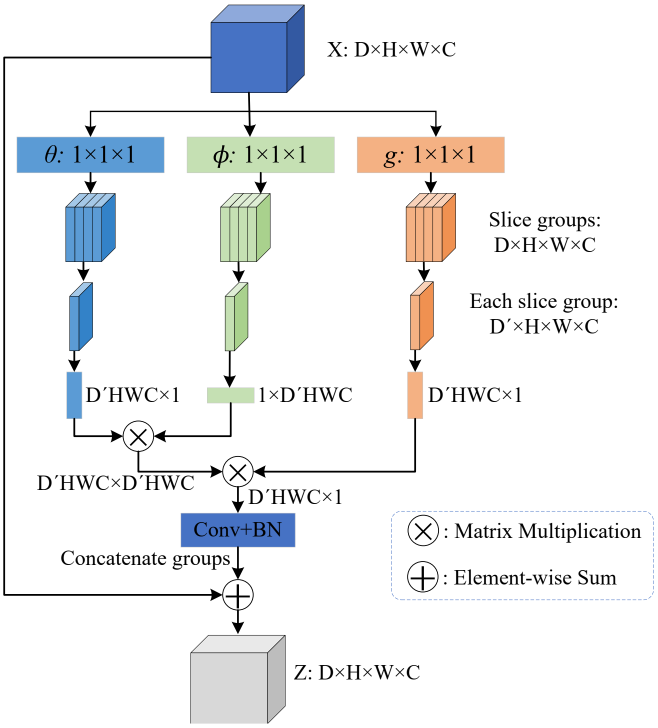

Fig. 3: The slice grouped non-local module (SGNL). After three $1 \times 1 \times 1$ convolutional layers, the feature maps are divided into multiple groups along the depth dimension. The depth dimension is grouped into $D' = D/G$, where G is the group number.

where vec denotes that it is a vector after reshape operation.

Since there is a $DHWC \times DHWC$ pairwise matrix, the computational complexity is much higher than the original non-local module, so directly implementing the SGNL is not feasible. Recently, some studies explore the idea of group convolution, such as Xception [60], MobileNet [61], ResNeXt [62], and Group normalization [63], dividing channels into groups has been proven to be effective in improving the performance of CNN. We introduce the group idea in SGNL and consider the characteristics of nodule detection in CT images, group the depth dimension D into G groups, as shown in Fig. 3, each of which contains $D' = D/G$ depths of the feature map. Each group is executed independently by Eq. 7 to compute $\mathbf{Y}'$, and the results are concatenated along the depth dimension to obtain $\mathbf{Y}$. In CT images, one nodule usually exists in several consecutive slices, and utilizing all depths to detect the nodule is unnecessary. The slice grouping operation can capture the similarity between any positions and any channels in one group, which augments the discrimination of nodules with different sizes correspond to information in one slice group.

Fig. 3 illustrates the workflow of SGNL module for each group. The SGNL operation in Eq. 7 is wrapped into the SGNL block, which is defined as:

$$\mathbf{Z} = concatenate \left(BN \left(\mathbf{Y}'\mathbf{W}_z \right) \right) + \mathbf{X}, \tag{8}$$

where $\mathbf{W}_z$ represents a $1 \times 1 \times 1$ convolutional layer and BN is a Batch Normalization [64]. "*concatenate*" denotes that all groups are concatenated along the depth dimension. The residual connection "$+\mathbf{X}$" makes the SGNL compatible with the existing neural network blocks. For the configuration of SGNL block, we add 5 blocks (2 blocks on the *res*3 and 3 blocks on the *res*4, to every other residual block) into 3D ResNet50 following [57].

TABLE 1: Comparison with the existing datasets of the pulmonary nodule. 'Scans' indicates the number of CT scans. 'Nodules' denotes the number of labeled nodules. 'Class' means the class number. And 'Avail' denotes whether the dataset is available.

Dataset	Year	Scans	Nodules	Class	Avail
ANODE09 [6]	2010	55	710	4	Yes
LIDC-IDRI [29]	2011	1,018	2,562	3	Yes
LUNA16 [30]	2016	888	1,186	2	Yes
DSB 2017 [7]	2017	2,101	N/A	2	No
PN9	2020	8,798	40,439	9	Yes

3.3 False Positive Reduction

The candidate detection stage is introduced to detect nodule candidates with high sensitivity, which usually carries many false positives. Some thoracic tissues, such as nodular-like structures, mediastinal structures, large vessels, and scarring, are often found as false positives. We further propose a false positive reduction module (FPR) to reduce the number of false positives among the nodule candidates and generate the final results.

As shown in Fig. 2, we take advantage of the multi-scale feature maps to reduce false positives considering that the features produced by the shallow block in ResNet contain rich spatial details with high resolution. By cropping the feature maps $M_{res1}, M_{res2}, M_{de2}$ using nodule candidates, we obtain three regions of interest (RoI) of different scales: $R_{res1}, R_{res2}, R_{de2}$. R_{de2} is up-sampled and concatenated with R_{res2}, then it is concatenated with R_{res1} after up-sampled. The final RoI is converted by 3D max pooling, followed by two Fully connected (FC) layers to obtain classification probability and bounding-box regression offsets. The loss function is the same as the above 3D RPN, which will reduce false positives and further optimize the regression parameters of the box.

4 Proposed Dataset

We collect and annotate a new large-scale pulmonary nodule dataset named PN9, which contains 8,798 thoracic CT scans and a total of 40,439 annotated nodules. In this section, we present the process of data acquisition and the detailed properties of the proposed dataset.

4.1 Data Collection and Annotation

4.1.1 Data Collection

The CT images of PN9 are mainly collected from two major hospitals and different scenes such as the clinic, hospitalization, and physical examination, etc. The year of each CT image being taken varies from 2015 to 2019. For the initial CT images obtained with CT plain scan, quality assurance is performed by confirming of selecting Digital Imaging and Communications in Medicine (DICOM) fields. Besides, the images with object interference and severe respiratory motion artifacts are excluded to ensure quality. When collecting data in the hospitals, all the protected health information of the patients contained in the DICOM headers of the images is removed through the data masking process, including the patient's name, institution name, referring physician's name, and so on.

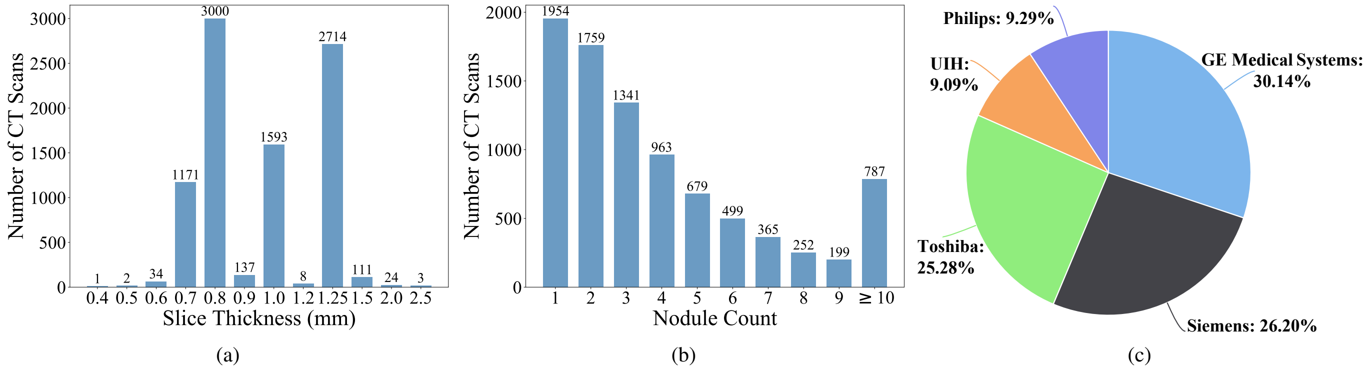

Fig. 4: Statistics of the proposed PN9 dataset. (a) Slice thickness distribution of CT scans. (b) Distribution of nodule count in one patient. (c) Percentage of CT manufacturer.

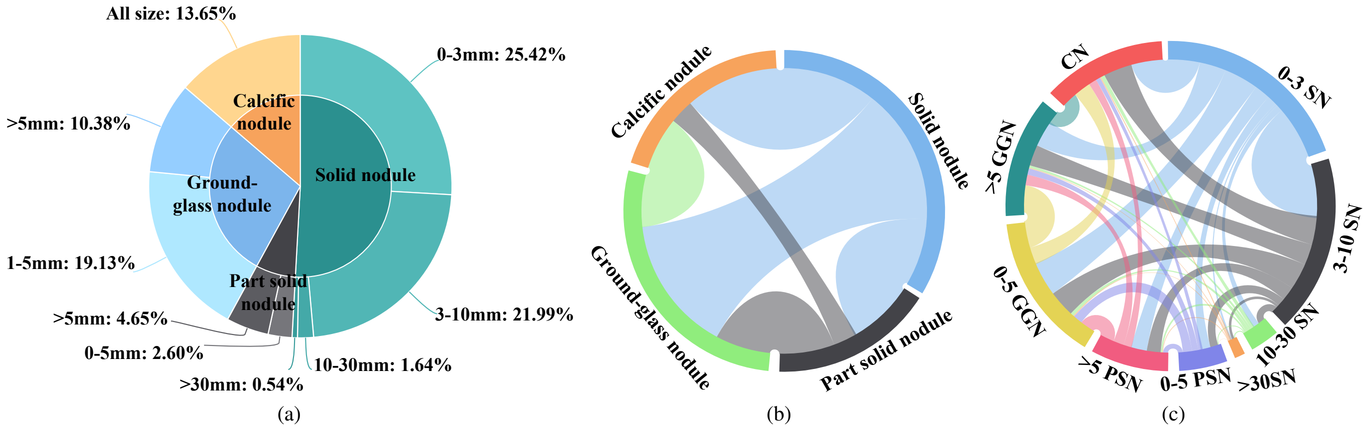

Fig. 5: Statistics of class in PN9. (a)Taxonomy of the PN9 dataset. It contains four super-classes and nine sub-classes. The percentage represents the proportion of a certain class of nodules to all nodules. (b) Mutual dependencies among super-classes. (c) Mutual dependencies among sub-classes.

4.1.2 Data Annotation

In order to annotate the pulmonary nodules as accurately as possible, we adopt a two-phase process for the interpretation of CT scans. Meanwhile, the doctors who annotating the DICOM images are attending physicians from the major hospitals. At the first annotation phase, the DICOM images of one CT scan are interpreted by an attending physician from the hospital and checked by another doctor. Then, the medical report of the patient is generated. The medical report contains information on the type, size, and the approximate location of each pulmonary nodule. Then we obtain medical reports of all CT scans acquired at both hospitals. The second phase is for detailed annotations. Each case is annotated slice by a slice of the pulmonary nodules by an attending physician, who refers to the hospital's medical report at the same time. For each pulmonary nodule at one slice identified by the doctor, the bounding box and category information are stored in a single XML file. By referring to the medical guidelines [65], [66], [67], and satisfying the needs of hospitals, we divide the pulmonary nodule into 9 different categories according to the nodule type and size. Doctors classify pulmonary nodules in terms of the category criteria. Then another doctor reviews and modifies the annotations to form the final annotation. If there is any inconsistency between the two doctors at the second annotation phase, they will discuss to determine the final annotation.

By implementing the above two-phase annotation procedures, we finally obtain 8,798 CT scans with 40,439 annotated nodules. All images are collected from hospitals, and the data distribution is consistent with the clinical circumstance.

4.2 Dataset Properties

CT Manufacturer. The CT scans in PN9 are obtained by a series of CT manufacturers and corresponding models, as shown in Fig. 4 (c). PN9 includes 2,652 scans from ten different GE Medical Systems scanner models, 2,305 scans from eleven different Siemens scanner models, 2,224 scans from three different Toshiba scanner models, 800 scans from two different United Imaging Healthcare (UIH) scanner models, and 817 scans from six different Philips scanner models.

Slice Thickness. Since the images of thick slice are not optimal for CAD analysis [68], [69], we mainly collect the CT scans with thin-slice. As illustrated in Fig. 4 (a), slice thickness ranges from 0.4 mm to 2.5 mm, and most are located at 0.7, 0.8, 1.0, and 1.25 mm. Besides, the pixel spacing ranges from 0.310 mm to 1.091 mm, with a mean of 0.706 mm.

Nodule count. In Fig. 4 (b), we illustrate the distribution of nodule count in one patient. We observe that approximately 68 % of patients have nodules less than 5 in our PN9. However, there are about 9 % of patients with more than 10 nodules, which may be difficult to detect.

Class. Our PN9 has a hierarchical class structure, and its detailed taxonomy is shown in Fig. 5. According to the property of the pulmonary nodules, all nodules in our dataset are first divided into four upper-level classes (denoted as super-class), including solid nodule (SN), part-solid nodule (PSN), ground-glass nodule (GGN), and calcific nodule (CN). Meanwhile, To satisfy the practical demands of doctors and hospitals, we further subdivide the super-class referring to the medical guidelines [65], [66], [67]. Each nodule is assigned with a subordinate class (denoted as sub-class) belonging to a certain super-class based on the nodule size. For example, sub-class 0-3mm solid nodules (denoted as 0-3SN) are defined as any nodules identified to be super-class solid nodules with the most significant in-plane dimension in the range of 0-3 mm. And 9 different sub-classes are finally obtained. The statistics of nodules in each class are shown in Fig. 5 (a). In Fig. 5 (b-c), we show the mutual dependencies among super-classes and sub-classes, respectively. The larger width of a link between two classes indicates a higher probability for the two classes' nodules appearing in one patient simultaneously. For example, a patient diagnosed with ground-glass nodules is also likely to have solid nodules.

There are 9 different categories in PN9 covering the most common pulmonary nodule types. However, since some types of nodules rarely appear in real life, the data distribution in our PN9 is imbalanced. As illustrated in Fig. 5, the number of small size nodules is bigger, and nodules like PSN are relatively fewer. The imbalanced distribution can result in the model learning a biased result to those nodules with relatively more samples. Besides, a large number of small size nodules also bring challenges to the accurate detection of nodules.

4.3 Comparison with Other Datasets

In Table 1, we compare the PN9 with several existing pulmonary nodule datasets. Compared to the widely used dataset LUNA16 [30], PN9 contains over 10 times more CT scans and over 30 times more annotated nodules. As for the class diversity, other datasets only have three categories: nodule $\geqslant$ 3 mm, nodule $<$ 3 mm, and non-nodule [29], [30]. Due to these limitations, it is difficult for most of the existing nodule datasets to apply to the practice. However, our PN9 contains many CT scans and 9 classes, which will contribute to the detection and classification tasks of the pulmonary nodules, allowing researchers to design more effective algorithms based on different types of nodules. Besides, there are more pulmonary nodules of small size, like 0-3mm solid nodules and 0-5mm ground-glass nodules. It helps identify small nodules more accurately, then the doctors can diagnose and treat patients earlier. In summary, our dataset not only is larger than the previous datasets, but also has superior diversity and performance.

5 Experiments

5.1 Evaluation Metrics

The Free-Response Receiver Operating Characteristic (FROC) is the official evaluation metric of the LUNA16 dataset [30], which is defined as the average recall rate at 0.125, 0.25, 0.5, 1, 2, 4, and 8 false positives per scan. And a nodule candidate is considered as a true positive when it is located within a distance R from the center of any nodules in the reference standard, where R denotes the radius of the reference nodule. Nodule candidates not located in the range of any reference nodules are considered as false positives. We use this evaluation metric in the experiments and further generalize it as FROC$_{IoU}$, which defines the true positives if the 3D Intersection over Union (IoU) of nodule candidates and any reference nodules is higher than one threshold (3D IoU threshold is defined as 0.25 in experiments).

Besides, we also adopt the 3D mean Average Precision (mAP) as the detection evaluation metric. According to the characteristics of 3D object detection and the dataset PN9, we define the following five metrics: AP@0.25 (AP at 3D IoU = 0.25), AP@0.35 (AP at 3D IoU = 0.35), AP$_s$ (AP for small nodules that correspond size 0-5 mm: volume $<$ 512), AP$_m$ (AP for medium nodules that correspond size 5-10 mm: 512 $<$ volume $<$ 4096), and AP$_l$ (AP for large nodules that correspond size $>$ 10 mm: volume $>$ 4096). Since pulmonary nodule detection is a 3D object detection task, for several 2D detection methods in the comparison experiments, the 2D proposals need to be merged to generate 3D proposals using a method similar to [70].

5.2 Experimental Settings

Data Preprocessing. For the dataset PN9, we split the 8,798 CT scans into 6,707 scans for training and 2,091 scans for testing. During training, we separate 670 CT scans from the training set as the validation set to monitor the convergence of the model. There are three preprocessing steps for the raw CT images. First, all raw data are converted into the Hounsfield Unit (HU) since HU is a standard quantitative value describing radiodensity. Then, the data is clipped into $[-1200, 600]$. Finally, we transform the data range linearly into $[0, 255]$.

Patch-Based Input. For 3D CNN, due to the GPU memory constraint, using the entire CT images as input during training is infeasible. We extract small 3D patches from the CT images and individually input them into the network. The size of the input 3D patch is 128 × 128 × 128 × 1 (Depth × Height × Width × Channel). If a patch exceeds the range of CT images, it is padded with a value of 170, which is the luminance of common tissues and can be distinguished from pulmonary nodules. During the test phase, we take the entire images of one CT scan as input, and do not crop the 3D patches. In order to avoid an odd size of the entire 3D images, they are padded with a value of 170 before being input into the model.

Implementation Details. In our SANet, we use the Stochastic Gradient Descent (SGD) optimizer with a batch size of 16. The initialization learning rate is set to 0.01, the momentum and weight decay coefficients are set to 0.9 and 1×10^{-4}, respectively. The SANet is trained with 200 epochs, and the learning rate decreases to 0.001 after 100 epochs and 0.0001 after another 60 epochs. Besides, our method is implemented using PyTorch. The experiments are performed on 4 NVIDIA RTX TITAN GPUs with 24GB memory.

5.3 Comparison with Other Detection Methods

In this section, we compare the detection performance of our SANet with several state-of-the-art detection methods on the PN9, including 2D CNN-based methods Faster R-CNN [34], RetinaNet [38], SSD512 [12] and 3D CNN-based methods leaky noisy-OR [7], 3D Faster R-CNN [23], DeepLung [23], NoduleNet (N$_2$) [71], I3DR-Net [43], DeepSEED [39].

Comparison based on FROC. We first evaluate the FROC score defined in the LUNA16 dataset [30]. The experiment results are listed in Table 2, and the FROC curves are illustrated in Fig. 6. It is noted that our SANet achieves the best results over other methods, which obtains an improvement of 3.04 % on

TABLE 2: Comparison of our SANet and other methods in terms of FROC on dataset PN9. The values are pulmonary nodule detection sensitivities (unit: %) with each column representing the average number of false positives per CT scan.

Method	0.125	0.25	0.5	1.0	2.0	4.0	8.0	Average
2D CNN-Based Methods:								
Faster R-CNN [34]	10.79	15.78	23.22	32.88	46.57	61.94	75.52	38.10
RetinaNet [38]	8.42	13.01	20.13	29.06	40.41	52.52	65.42	32.71
SSD512 [12]	12.26	18.78	28.00	40.32	56.89	73.18	86.48	45.13
3D CNN-Based Methods:								
Leaky Noisy-OR [7]	28.08	36.42	46.99	56.72	66.08	73.77	81.71	55.68
3D Faster R-CNN [23]	27.57	36.59	46.76	58.00	70.00	80.02	88.32	58.18
DeepLung [23]	28.59	39.08	50.17	62.28	72.60	82.00	88.64	60.48
NoduleNet (N_2) [71]	27.33	38.25	49.40	61.09	73.11	83.28	89.83	60.33
I3DR-Net [43]	23.99	34.37	46.80	60.04	72.88	83.60	89.57	58.75
DeepSEED [39]	29.21	40.64	51.15	62.20	73.82	83.24	89.70	61.42
SANet	**38.08**	**45.05**	**54.46**	**64.50**	**75.33**	**83.86**	**89.96**	**64.46**

TABLE 3: Comparison of our SANet and other methods in terms of FROC$_{IoU}$ (%) on dataset PN9.

Method	0.125	0.25	0.5	1.0	2.0	4.0	8.0	Average
2D CNN-Based Methods:								
Faster R-CNN [34]	3.41	6.97	12.26	20.58	33.05	46.41	57.90	25.80
RetinaNet [38]	2.60	5.56	10.95	19.25	29.29	40.49	51.05	22.74
SSD512 [12]	4.62	8.48	14.76	25.06	40.32	57.27	70.80	31.61
3D CNN-Based Methods:								
NoduleNet (N_2) [71]	21.17	30.23	40.38	51.02	61.26	70.70	76.93	50.24
I3DR-Net [43]	15.64	23.13	37.00	51.54	64.54	72.91	77.53	48.90
SANet	**26.72**	**36.03**	**47.46**	**56.99**	**66.35**	**73.52**	**78.32**	**55.06**

TABLE 4: Comparison of our SANet and other methods in terms of FROC (%) on the dataset LUNA16 [30] .

Method	0.125	0.25	0.5	1.0	2.0	4.0	8.0	Average
Leaky Noisy-OR [7]	59.38	72.66	78.13	84.38	87.50	89.06	89.84	80.13
3D Faster R-CNN [23]	66.20	74.60	81.50	86.40	90.20	91.80	93.20	83.40
DeepLung [23]	69.20	76.90	82.40	86.50	89.30	91.70	93.30	84.20
NoduleNet (N_2) [71]	65.18	76.79	83.93	87.50	91.07	92.86	93.75	84.43
I3DR-Net [43]	63.56	71.31	79.84	85.27	87.60	89.92	91.47	81.28
DeepSEED [39]	**73.90**	**80.30**	85.80	88.80	90.70	91.60	92.00	86.20
SANet	71.17	80.18	**86.49**	**90.09**	**93.69**	**94.59**	**95.50**	**87.39**

TABLE 5: Comparison of SANet and NoduleNet based on AP.

Method	AP@0.25	AP@0.35	AP$_s$	AP$_m$	AP$_l$
NoduleNet (N_2) [71]	46.7	30.2	12.8	45.3	46.4
SANet	**52.2**	**36.6**	**14.1**	**47.6**	**48.6**

TABLE 6: Ablation study for the proposed SGNL and FPR (%). The baseline is the detection model based on 3D ResNet50 and 3D RPN (No. 1). We add the SGNL module and FPR module to show their effectiveness(No. 2 and No. 3). The No. 4 is the complete version of our proposed SANet. The p-value denotes a statistical significance test for FROC (vs. proposed SANet).

No.	SGNL	FPR	FROC	FROC$_{IoU}$	AP@0.25	p-value
1			61.29	51.96	49.0	0.009
2	✔		64.34	53.44	51.3	0.081
3		✔	62.69	52.51	50.2	0.015
4	✔	✔	**64.46**	**55.06**	**52.2**	–

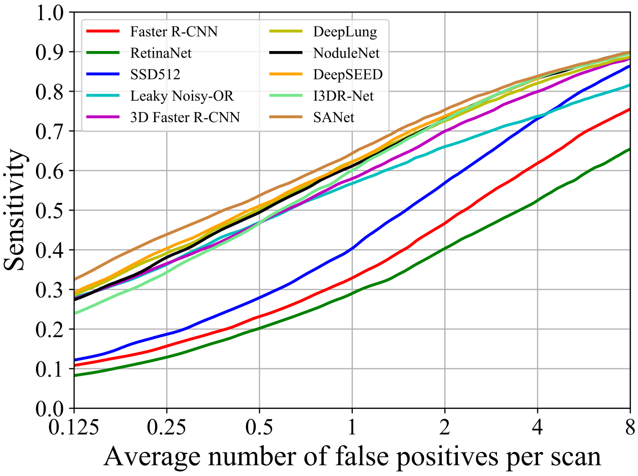

Fig. 6: FROC curves of compared methods and our SANet.

average FROC score over the second-best DeepSEED [39]. And our method especially outperforms the other detection methods by a large margin for the average number of false positives per CT scan smaller than 2. Besides, other 3D CNN-based methods, like NoduleNet (N_2) [71] and DeepLung [23], obtain comparable results. It can be seen that the FROC scores of 3D CNN-based methods are significantly better than 2D CNN-based methods. For example, our proposed SANet improve the SSD512 [12] and Faster R-CNN [34] in terms of average FROC score by 19.33 %, and 26.36 %, respectively. Since the 2D CNN-based methods only utilize the input images of three channels and learn insufficient spatial information, they obtain weak performance for 3D pulmonary nodule detection.

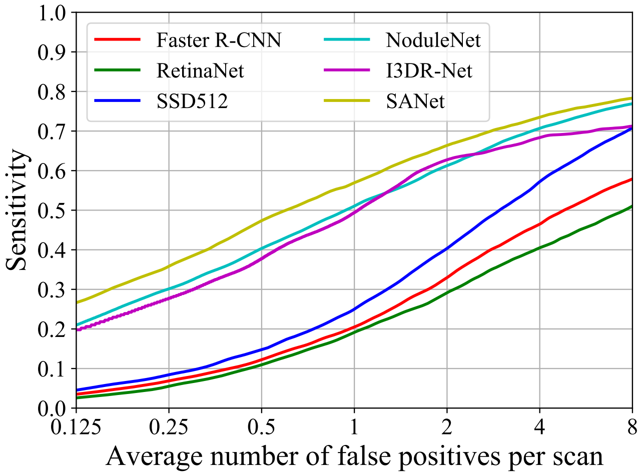

Fig. 7: FROC$_{IoU}$ curves of compared methods and our SANet.

TABLE 7: Comparison of other methods equipped with the proposed SGNL and FPR modules (%).

Model	SGNL	FPR	FROC	FROC$_{IoU}$
DeepLung [23]			60.48	–
DeepLung [23]	✔		61.67	–
DeepLung [23]		✔	61.15	–
DeepLung [23]	✔	✔	62.06	–
NoduleNet (N$_2$) [71]			60.33	50.24
NoduleNet (N$_2$) [71]	✔		62.35	51.13
NoduleNet (N$_2$) [71]		✔	61.18	50.99
NoduleNet (N$_2$) [71]	✔	✔	**62.69**	**52.37**

Comparison based on FROC$_{IoU}$. In the origin FROC, a nodule candidate is defined as true positive if it is located in a distance from the center of any reference nodules. We further generalize the FROC as FROC$_{IoU}$, which defines the true positives based on 3D IoU of nodule candidates and reference nodules. The experiment results are shown in Table 3 and Fig. 7. We do not list the results of leaky noisy-OR [7], 3D Faster R-CNN [23], DeepLung [23], DeepSEED [39] because that the central coordinates and diameters they predict are not matching with the 3D cube of ground truth nodules. Our SANet achieves an FROC$_{IoU}$ score of 55.06 %, which is better than other methods and outperforms the NoduleNet (N$_2$) [71] by 4.82 %. And the FROC$_{IoU}$ scores of 3D CNN-based methods are also better than 2D CNN-based methods.

Comparison based on AP. Since larger nodules usually have a higher suspicion of malignancy, the size of nodules is important for diagnosing lung cancer. We also adopt the 3D AP and define several evaluation metrics based on our dataset PN9. Table 5 lists the results of NoduleNet (N$_2$) [71] and our proposed method SANet. Our SANet improves the NoduleNet in terms of AP@0.25 by 5.5. Considering the size of nodules, we define three metrics AP$_s$, AP$_m$, and AP$_l$ to evaluate the detection performance of small, medium, and large nodules, respectively. It can be seen that our SANet obtains better results on all three metrics, which prove its effectiveness.

Visualization. The visualization of central slices for nodule ground truths and different methods' detection results is shown in Fig. 8. For the nodules of five types, the detected nodule positions of our SANet are consistent with those of ground truth. However, the detection results obtained by other methods are usually offset or larger than the ground truth, especially the 2D CNN-based method. These experiment results verify the superiority of our SANet in the task of nodule detection.

TABLE 8: Ablation study for the proposed SGNL module with different configurations of the SGNL block (%). The p-value denotes a statistical significance test for FROC (vs. SANet with 5-block).

No.	SGNL Block	FROC	FROC$_{IoU}$	AP@0.25	p-value
1	4-block	62.97	53.45	50.8	0.023
2	5-block	**64.46**	**55.06**	**52.2**	–
3	10-block	64.20	53.92	51.1	0.079

TABLE 9: Ablation study for the proposed SGNL module with different numbers of groups G (%).

Groups G	FROC	FROC$_{IoU}$	AP@0.25	AP$_s$	AP$_m$	AP$_l$
1	62.88	53.73	50.6	10.5	**48.2**	48.5
4	**64.46**	**55.06**	**52.2**	14.1	47.6	**48.6**
8	63.80	54.62	51.2	**15.6**	46.6	47.8

Comparison on the LUNA16 dataset. To further validate the performance of the proposed SANet, we conduct experiments on the widely used dataset LUNA16 [30] with 10-fold cross-validation. As shown in Table 4, our SANet achieves the best results for pulmonary nodule detection. For example, it obtains an average FROC score of 87.39 %, which improves the second-best method DeepSEED [39] by 1.19 %. Besides, our SANet outperforms the state-of-the-art nodule detection methods by a large margin for the settings of average number of false positives per CT scan larger than 1.

5.4 Ablation Studies

Effectiveness of Our Proposed Modules. In the SANet model, there are two essential modules: SGNL and FPR. To verify the performance of two modules, we conduct experiments with different settings, as shown in Table 6. The No.1 is the baseline performance without SGNL and FPR. After applying our proposed SGNL and FPR to the baseline, the results obtain improvements in terms of the FROC score by 3.05 % and 1.40 %. These results validate that both modules are helpful for pulmonary nodule detection. Besides, we achieve a 3.17 % improvement in terms of the FROC score if combining SGNL and FPR. The FROC$_{IoU}$ score and AP@0.25 have also been improved by applying the two modules. We also validate the effect of DeepLung [23] and NoduleNet (N$_2$) [71] equipped with our proposed two modules. As listed in Table 7, the FROC scores of NoduleNet (N$_2$) [71] are improved by 2.02 % and 0.85 % with applying our proposed SGNL and FPR modules, respectively. The performance of DeepLung [23] is also improved by adding the two modules. These results verify that our proposed SGNL and FPR modules are beneficial for nodule detection.

Effect of Different Configurations in SGNL. Table 8 shows the results of the proposed SGNL module with different configurations of the SGNL block. We add 4 blocks (to right before the last residual block of stages *res*2 to *res*5), 5 blocks (2 to *res*3 and 3 to *res*4, to every other residual block), and 10 blocks (to every residual block in *res*3 and *res*4) into 3D ResNet50. Compared with the No.3 in Table 6, adding different configurations of the SGNL block brings improvements in terms of the three evaluation metrics. The results of 5 SGNL blocks are best, which provides 1.77 % improvement on FROC and 2.55 % improvement on FROC$_{IoU}$.

We also analyze the influence of the different numbers of groups G in the SGNL module, as listed in Table 9. We achieve

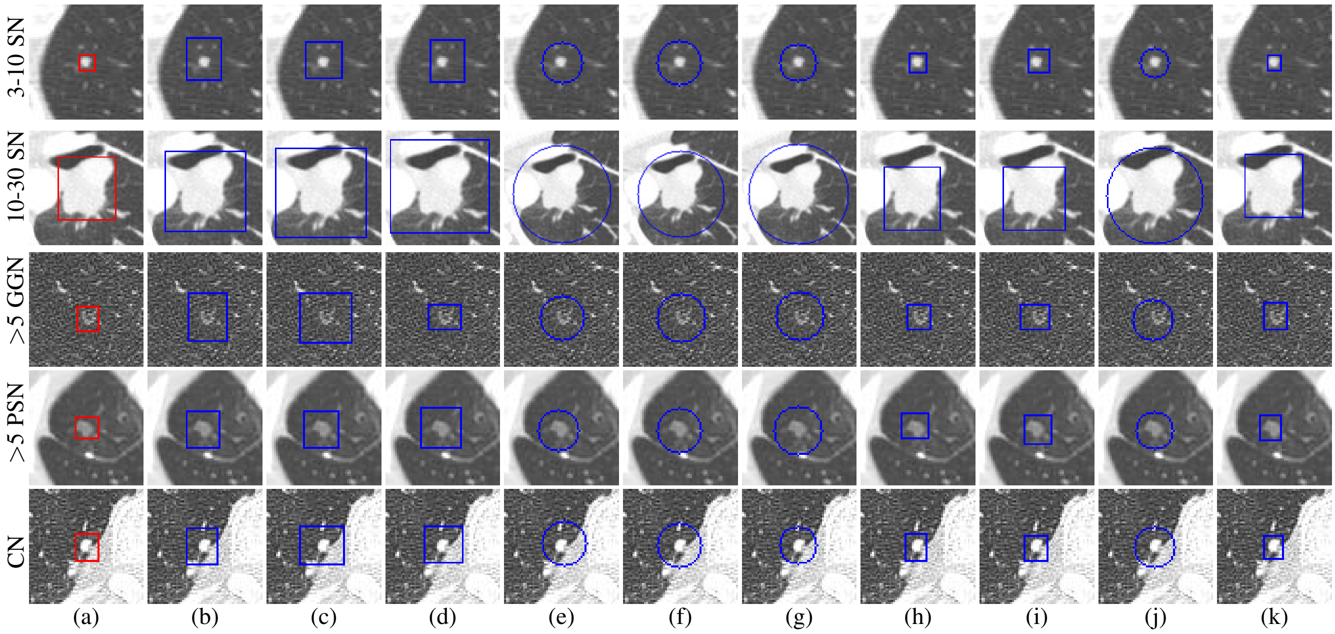

Fig. 8: Qualitative comparison of central slices for our SANet and other methods. The first row to the fifth row show the comparison results with different nodule classes: 3-10 SN, 10-30 SN, >5 GGN, $>$ 5 PSN, and CN, respectively. (a) Ground truth. (b)-(i) Detection results of Faster R-CNN [34], RetinaNet [38], SSD512 [12], Leaky Noisy-OR [7], 3D Faster R-CNN [23], DeepLung [23], NoduleNet (N_2) [71], I3DR-Net [43], DeepSEED [39], and our SANet.

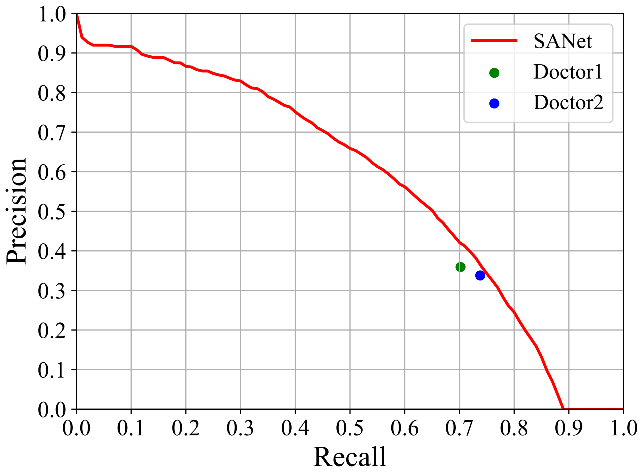

Fig. 9: The precision-recall curve for the nodule detection on the small-scale pulmonary nodule testing dataset. The 'Doctor1' and 'Doctor2' denote the detection results of two experienced doctors.

TABLE 10: Analysis of how different CT manufacturers affect the performance. 'GE' denotes GE Medical Systems.

CT Manufacturer	GE	Philips	Siemens	Toshiba	UIH
Number in Test Set	657	174	523	557	180
FROC	66.63	65.07	65.73	60.65	65.50

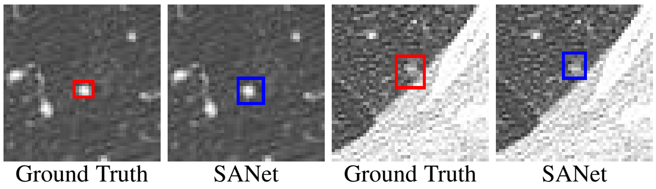

Fig. 10: Visualization of some cases that our SANet fails. The first two columns are nodules of 0-3 SN, and the last two columns are nodules of 0-5 PSN.

an FROC score of 64.46 % when $G = 4$, which improves two other settings $G = 1$ and $G = 8$ by 1.58 % and 0.66 %, respectively. Besides, the AP_s score is best when $G = 8$ and the AP_m score is best when $G = 1$. These experimental results are in line with our expectations. The slice grouping operation is proposed to capture the relationship between any positions and any channels in one group, which augments the discrimination of 3D pulmonary nodules in different sizes. If there are fewer groups, each group contains more consecutive slices, it is beneficial to detect large nodules but limits the detection of small nodules. Each group contains few slices when too many groups are split, restricting the detection of nodules with large size. Since the overall performance is best and AP of nodules with different sizes is comparable, we set the number of groups $G = 4$.

5.5 Comparison with Experienced Doctors

Furthermore, We compare the detection performance of our SANet and two experienced doctors with at least 10 years' clinical experience. We collect an additional small-scale pulmonary nodule testing dataset that contains 120 CT scans. After annotated accurately by several attending physicians from major hospitals, this testing dataset contains 2,137 annotated nodules with the golden standard. The other two experienced doctors, who never diagnose the small-scale testing dataset, are invited to individually identify lung nodules. Each doctor label one 3D bounding box and category for each nodule. If the 3D IoU of one nodule candidate and the golden standard is higher than one threshold, this candidate will be considered as true positive. Then the detection results of two experienced doctors can be obtained: precision 35.92% and recall 70.20% for doctor1, precision 33.78% and recall 73.80% for doctor2. And the detection performance of the two doctors is not high, which is one of the major challenges in nodule diagnosis and treatment. As for our SANet, it is trained on dataset PN9 and tested on this small-scale testing dataset, which is evaluated using AP@0.25.

The PR curve of our SANet and the detection results of two experienced doctors are shown in Fig. 9. It is noted that the performance of our model is better than two doctors on their individually diagnosed nodules, which validates that SANet surpasses the human-level performance and is suited for pulmonary nodule detection. Some pulmonary nodules are small, while different nodules have different morphology. Thus, doctors usually recognize the obvious nodules in CT images and fail to identify each nodule, especially a large number of small nodules. For a nodule containing multiple consecutive slices, they often miss some slices and affect the performance of nodule detection. Our SANet takes much less time to identify nodules than doctors. Therefore, doctors will diagnose pulmonary nodules with higher efficiency and accuracy by taking advantage of our SANet, which will further help the early diagnosis and treatment of lung cancer. Meanwhile, we hope future researches on our PN9 will further promote the detection results for pulmonary nodules and help its application in clinical circumstances.

5.6 Discussion

As shown in the above experiments, our proposed SANet obtains the best performance on our PN9 and public dataset LUNA16 [30] compared with other state-of-the-art detection methods. Besides, the performance of SANet is better than two experienced doctors. However, there are still some failure cases of our method. As illustrated in Fig. 10, SANet may generate larger bounding boxes than the ground truth when identifying nodules with class 0-3 SN. Since the nodules of PSN usually have a fuzzy border, SANet may not identify the entire nodules of PSN and produce smaller bounding boxes than the ground truth. In the future, we will consider the attributes of different categories to detect pulmonary nodules better.

We also analyze the influence of different CT manufacturers. As listed in Table 10, we report the FROC score of our SANet for CT scans in the test set with different manufacturers. The difference between the results is small except CT manufacturer of Toshiba. The reason is that the CT scans in the test set obtained by Toshiba happen to contain more small nodules than other manufacturers, which affects its performance. In general, the influence of different CT manufacturers is small since our PN9 contains sufficient CT scans from different manufacturers.

6 Conclusion

In this paper, we propose a new large-scale dataset named PN9 for pulmonary nodule detection. Specifically, it contains 8,798 CT scans, 9 classes of common pulmonary nodules, and 40,439 annotated nodules. Compared with the currently existing pulmonary nodule datasets, the PN9 has a much larger scale and more categories, which is beneficial for the CNN-based methods and practical application. What's more, we develop a slice-aware network (SANet) for nodule detection. The SGNL module is introduced to learn explicit correlations among any position and any channels of one slice group in the feature map. And we propose a false positive reduction (FPR) module to reduce the false positives generated in the detection stage of 3D RPN. Extensive experiment results on dataset PN9 demonstrate the superior performance of the SANet. We hope our dataset PN9 and method SANet will promote future researches on pulmonary nodule detection and further help the application of deep learning in clinical circumstances.

Acknowledgment

This research was supported by the Major Project for New Generation of AI under Grant No. 2018AAA0100400, NSFC (61922046), S&T innovation project from Chinese Ministry of Education, and Tianjin Natural Science Foundation (18ZXZNGX00110).

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: a Cancer Journal for Clinicians*, vol. 66, no. 1, pp. 7–30, 2016.

[2] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[3] M. Infante, S. Cavuto, F. R. Lutman, G. Brambilla, G. Chiesa, G. Ceresoli, E. Passera, E. Angeli, M. Chiarenza, G. Aranzulla *et al.*, "A randomized study of lung cancer screening with spiral computed tomography: three-year results from the dante trial," *American Journal of Respiratory and Critical Care Medicine*, vol. 180, no. 5, pp. 445–453, 2009.

[4] N. L. S. T. R. Team, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395–409, 2011.

[5] S. Singh, D. S. Gierada, P. Pinsky, C. Sanders, N. Fineberg, Y. Sun, D. Lynch, and H. Nath, "Reader variability in identifying pulmonary nodules on chest radiographs from the national lung screening trial," *J. Thoracic Imag.*, vol. 27, no. 4, p. 249, 2012.

[6] B. Van Ginneken, S. G. Armato III, B. de Hoop, S. van Amelsvoort-van de Vorst, T. Duindam, M. Niemeijer, K. Murphy, A. Schilham, A. Retico, M. E. Fantacci *et al.*, "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the anode09 study," *Med. Image Anal.*, vol. 14, no. 6, pp. 707–722, 2010.

[7] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network," *IEEE Tran. on Neural Networks and Learning Systems*, 2019.

[8] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient cad system for pulmonary nodule detection in ct imagery," *Med. Image Anal.*, vol. 14, no. 3, pp. 390–406, 2010.

[9] C. Jacobs, E. M. van Rikxoort, T. Twellmann, E. T. Scholten, P. A. de Jong, J.-M. Kuhnigk, M. Oudkerk, H. J. de Koning, M. Prokop, C. Schaefer-Prokop *et al.*, "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Med. Image Anal.*, vol. 18, no. 2, pp. 374–384, 2014.

[10] E. Lopez Torres, E. Fiorina, F. Pennazio, C. Peroni, M. Saletta, N. Camarlinghi, M. E. Fantacci, and P. Cerello, "Large scale validation of the m5l lung cad on heterogeneous ct datasets," *Med. Phys.*, vol. 42, no. 4, pp. 1477–1489, 2015.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Adv. Neural Inform. Process. Syst.*, 2015, pp. 91–99.

[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Eur. Conf. Comput. Vis.* Springer, 2016, pp. 21–37.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 779–788.

[14] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1930–1943, 2012.

[15] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[16] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[17] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[18] J. S. Duncan and N. Ayache, "Medical image analysis: Progress over two decades and the challenges ahead," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 85–106, 2000.

[19] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2013, pp. 246–253.

[20] T. Brosch, Y. Yoo, D. K. Li, A. Traboulsee, and R. Tam, "Modeling the variability in brain morphology and lesion distribution in multiple sclerosis by deep learning," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2014, pp. 462–469.

[21] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, 2021.

[22] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, C.-W. Zhao, and M.-M. Cheng, "Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation," *arXiv preprint arXiv:2004.07054*, 2020.

[23] W. Zhu, C. Liu, W. Fan, and X. Xie, "Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification," in *IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2018, pp. 673–681.

[24] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. Van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken, "Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks," *IEEE Trans. Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[25] X. Peng and C. Schmid, "Multi-region two-stream r-cnn for action detection," in *Eur. Conf. Comput. Vis.* Springer, 2016, pp. 744–759.

[26] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2017, pp. 559–567.

[27] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection," *IEEE Trans. Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2016.

[28] X. Yan, J. Pang, H. Qi, Y. Zhu, C. Bai, X. Geng, M. Liu, D. Terzopoulos, and X. Ding, "Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2d and 3d strategies," in *Asian Conf. Comput. Vis.* Springer, 2016, pp. 91–101.

[29] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman *et al.*, "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, 2011.

[30] A. A. A. Setio, A. Traverso, T. De Bel, M. S. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, 2017.

[31] N. Duggan, E. Bae, S. Shen, W. Hsu, A. Bui, E. Jones, M. Glavin, and L. Vese, "A technique for lung nodule candidate detection in ct using global minimization methods," in *Int. Worksh. Energy Minimization Methods in Comput. Vis. Pattern Recog.* Springer, 2015, pp. 478–491.

[32] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 580–587.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015.

[34] R. Girshick, "Fast r-cnn," in *Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

[35] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 7263–7271.

[36] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "Dssd: Deconvolutional single shot detector," *arXiv preprint arXiv:1701.06659*, 2017.

[37] B. Wu, F. Iandola, P. H. Jin, and K. Keutzer, "Squeezedet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving," in *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, 2017, pp. 129–137.

[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[39] Y. Li and Y. Fan, "Deepseed: 3d squeeze-and-excitation encoder-decoder convolutional neural networks for pulmonary nodule detection," in *Int. Symposium on Biomedical Imaging*. IEEE, 2020, pp. 1866–1869.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2015, pp. 234–241.

[41] B.-C. Kim, J. S. Yoon, J.-S. Choi, and H.-I. Suk, "Multi-scale gradual integration cnn for false positive reduction in pulmonary nodule detection," *Neural Networks*, vol. 115, pp. 1–10, 2019.

[42] O. Ozdemir, R. L. Russell, and A. A. Berlin, "A 3d probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans," *IEEE Trans. Med. Image.*, vol. 39, no. 5, pp. 1419–1429, 2019.

[43] I. W. Harsono, S. Liawatimena, and T. W. Cenggoro, "Lung nodule detection and classification from thorax ct-scan using retinanet with transfer learning," *Journal of King Saud University-Computer and Information Sciences*, 2020.

[44] T. Song, J. Chen, X. Luo, Y. Huang, X. Liu, N. Huang, Y. Chen, Z. Ye, H. Sheng, S. Zhang *et al.*, "Cpm-net: A 3d center-points matching network for pulmonary nodule detection in ct scans," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2020, pp. 550–559.

[45] A. El-Baz, M. Nitzken, F. Khalifa, A. Elnakib, G. Gimel'farb, R. Falk, and M. A. El-Ghar, "3d shape analysis for early diagnosis of malignant lung nodules," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2011, pp. 772–783.

[46] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, p. 4006, 2014.

[47] T. W. Way, L. M. Hadjiiski, B. Sahiner, H.-P. Chan, P. N. Cascade, E. A. Kazerooni, N. Bogot, and C. Zhou, "Computer-aided diagnosis of pulmonary nodules on ct scans: Segmentation and classification using 3d active contours," *Med. Phys.*, vol. 33, no. 7Part1, pp. 2323–2337, 2006.

[48] F. Han, G. Zhang, H. Wang, B. Song, H. Lu, D. Zhao, H. Zhao, and Z. Liang, "A texture feature analysis for diagnosis of pulmonary nodules using lidc-idri database," in *IEEE International Conference on Medical Imaging Physics and Engineering*. IEEE, 2013, pp. 14–18.

[49] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *Int. Conf. Inform. Process. Med. Image.* Springer, 2015, pp. 588–599.

[50] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," in *Adv. Neural Inform. Process. Syst.*, 2017, pp. 4467–4475.

[51] S. Hussein, K. Cao, Q. Song, and U. Bagci, "Risk stratification of lung nodules using 3d cnn-based multi-task learning," in *Int. Conf. Inform. Process. Med. Image.* Springer, 2017, pp. 249–260.

[52] W. Shen, M. Zhou, F. Yang, D. Dong, C. Yang, Y. Zang, and J. Tian, "Learning from experts: Developing transferable deep features for patient-level lung cancer prediction," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2016, pp. 124–131.

[53] W. Zhu, Q. Lou, Y. S. Vang, and X. Xie, "Deep multi-instance networks with sparse label assignment for whole mammogram classification," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2017, pp. 603–611.

[54] M. F. McNitt-Gray, S. G. Armato III, C. R. Meyer, A. P. Reeves, G. McLennan, R. C. Pais, J. Freymann, M. S. Brown, R. M. Engelmann, P. H. Bland *et al.*, "The lung image database consortium (lidc) data collection process for nodule detection and annotation," *Academic Radiology*, vol. 14, no. 12, pp. 1464–1474, 2007.

[55] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 770–778.

[57] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 7794–7803.

[58] K. Yue, M. Sun, Y. Yuan, F. Zhou, E. Ding, and F. Xu, "Compact generalized non-local network," in *Adv. Neural Inform. Process. Syst.*, 2018, pp. 6510–6519.

[59] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear cnn models for fine-grained visual recognition," in *Int. Conf. Comput. Vis.*, 2015, pp. 1449–1457.

[60] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 1251–1258.

[61] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.

[62] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 1492–1500.

[63] Y. Wu and K. He, “Group normalization,” in *Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[64] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.

[65] H. MacMahon, D. P. Naidich, J. M. Goo, K. S. Lee, A. N. Leung, J. R. Mayo, A. C. Mehta, Y. Ohno, C. A. Powell, M. Prokop *et al.*, “Guidelines for management of incidental pulmonary nodules detected on ct images: from the fleischner society 2017,” *Radiology*, vol. 284, no. 1, pp. 228–243, 2017.

[66] D. S. Ettinger, D. E. Wood, W. Akerley, L. A. Bazhenova, H. Borghaei, D. R. Camidge, R. T. Cheney, L. R. Chirieac, T. A. D’Amico, T. J. Dilling *et al.*, “Nccn guidelines insights: non–small cell lung cancer, version 4.2016,” *Journal of the National Comprehensive Cancer Network*, vol. 14, no. 3, pp. 255–264, 2016.

[67] F. C. Detterbeck, P. J. Mazzone, D. P. Naidich, and P. B. Bach, “Screening for lung cancer: diagnosis and management of lung cancer: American college of chest physicians evidence-based clinical practice guidelines,” *Chest*, vol. 143, no. 5, pp. e78S–e92S, 2013.

[68] D. Manos, J. M. Seely, J. Taylor, J. Borgaonkar, H. C. Roberts, and J. R. Mayo, “The lung reporting and data system (lu-rads): a proposal for computed tomography screening,” *Canadian Association of Radiologists Journal*, vol. 65, no. 2, pp. 121–134, 2014.

[69] E. A. Kazerooni, J. H. Austin, W. C. Black, D. S. Dyer, T. R. Hazelton, A. N. Leung, M. F. McNitt-Gray, R. F. Munden, and S. Pipavath, “Acr–str practice parameter for the performance and reporting of lung cancer screening thoracic computed tomography (ct): 2014 (resolution 4),” *J. Thoracic Imag.*, vol. 29, no. 5, pp. 310–316, 2014.

[70] F. R. Pereira, D. Menotti, and L. F. de Oliveira, “A 3d lung nodule candidate detection by grouping dcnn 2d candidates.” in *VISIGRAPP*, 2019, pp. 537–544.

[71] H. Tang, C. Zhang, and X. Xie, “Nodulenet: Decoupled false positive reduction for pulmonary nodule detection and segmentation,” in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2019, pp. 266–274.

Jie Mei is a Ph.D. student in College of Computer Science, Nankai University. He is supervised via Prof. Ming-Ming Cheng. His research interests include computer vision, machine learning, and remote sensing image processing.

Ming-Ming Cheng received his Ph.D. degree from Tsinghua University in 2012. Then he did two years research fellow with Prof. Philip Torr in Oxford. He is now a professor at Nankai University, leading the Media Computing Lab. His research interests include computer graphics, computer vision, and image processing. He received research awards, including ACM China Rising Star Award, IBM Global SUR Award, and CCF-Intel Young Faculty Researcher Program. He is on the editorial boards of IEEE TIP.

Gang Xu is a Ph.D. student in Media Computing Lab at Nankai University. He is supervised by Prof. Ming-Ming Cheng. He received his bachelor’s degree from Xidian University in 2018. His research interests include computer vision and machine learning.

Lan-Ruo Wan received her master degree majoring control systems in Imperial College London. Now she serves as Machine Learning Engineer in Infervision Technology Co., Ltd. Her research interests include medical image algorithm and computer vision.

Huan Zhang received his master degree in computer science and technology from Fudan University in 2018. Now he is in charge of algorithm research at InferVision Medical Technology Co., Ltd. His research interests include medical image algorithm and computer vision.