

Interpretative Computer-aided Lung Cancer Diagnosis: from Radiology Analysis to Malignancy Evaluation

Shaohua Zheng^{a,1}, Zhiqiang Shen^{a,2}, Chenhao Pei^{a,3}, Wangbin Ding^{a,4}, Haojin Lin^{a,5}, Jiepeng Zheng^{b,6}, Lin Pan^{a,*,7}, Bin Zheng^{b,**,8} and Liqin Huang^{a,9}

^aCollege of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China

^bThoracic Department, Fujian Medical University Union Hospital, Fuzhou 350001, China

ARTICLE INFO

Keywords:

Computer-aided diagnosis
Malignancy evaluation
Pulmonary nodule
Radiology analysis

ABSTRACT

Background and Objective: Computer-aided diagnosis (CAD) systems promote diagnosis effectiveness and alleviate pressure of radiologists. A CAD system for lung cancer diagnosis includes nodule candidate detection and nodule malignancy evaluation. Recently, deep learning-based pulmonary nodule detection has reached satisfactory performance ready for clinical application. However, deep learning-based nodule malignancy evaluation depends on heuristic inference from low-dose computed tomography (LDCT) volume to malignant probability, which lacks clinical cognition.

Methods: In this paper, we propose a joint radiology analysis and malignancy evaluation network (R2MNet) to evaluate the pulmonary nodule malignancy via radiology characteristics analysis. Radiological features are extracted as channel descriptor to highlight specific regions of the input volume that are critical for nodule malignancy evaluation. In addition, for model explanations, we propose channel-dependent activation mapping (CDAM) to visualize the features and shed light on the decision process of deep neural network (DNN).

Results: Experimental results on the LIDC-IDRI dataset demonstrate that the proposed method achieved area under curve (AUC) of 96.27% on nodule radiology analysis and AUC of 97.52% on nodule malignancy evaluation. In addition, explanations of CDAM features proved that the shape and density of nodule regions were two critical factors that influence a nodule to be inferred as malignant, which conforms with the diagnosis cognition of experienced radiologists.

Conclusion: Incorporating radiology analysis with nodule malignant evaluation, the network inference process conforms to the diagnostic procedure of radiologists and increases the confidence of evaluation results. Besides, model interpretation with CDAM features shed light on the regions which DNNs focus on when they estimate nodule malignancy probabilities.

1. Introduction

Lung cancer is the most common cause of cancer death worldwide [1]. Lung cancer screening using low-dose computed tomography (LDCT) scans has been proved as an effective tool to reduce patient mortality [2]. However, A thorough inspection of a CT scan usually takes a radiologist around 10 minutes and diagnosis results are influenced by the doctor's experience and emotion. With the increasing number of CT images, the data volumes to be analyzed overwhelm the capacity of radiologists. Computer-aided diagnosis (CAD) systems have the potential to reduce this burden. In recent years, deep learning-based methods have demonstrated impressive performance in medical image processing, and taken up a dominant position in the design of CAD systems [3, 4, 5, 6, 7, 8].

A general deep learning-based CAD system for lung cancer diagnosis includes 1) a pulmonary nodule detection module that detects candidate pulmonary nodules, and 2) a nodule malignancy evaluation module that diagnoses the suspicious nodules proposed by the previous stage. Deep learning-

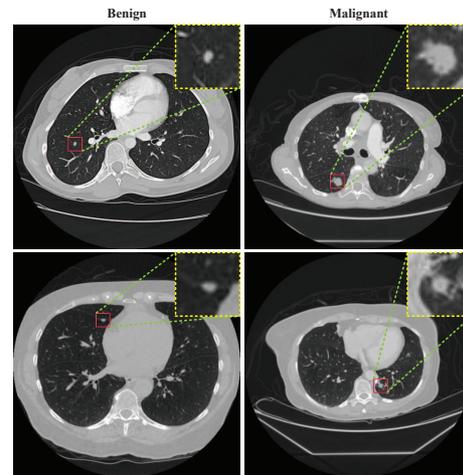


Figure 1: Examples of benign (the left column) and malignant nodules (the right column). The red rectangles emphasize the nodule locations and the yellow dashed rectangles highlight the nodule areas. Figure best viewed in color.

based nodule detection has achieved remarkable results. However, deep learning-based nodule malignancy evaluation models that straightforwardly predict malignant probabilities are short of explanations of which regions deep neural networks (DNNs) focus on [9, 10]. Doctors estimate nodule malignant

*Corresponding author at: College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China E-mail: panlin@fzu.edu.cn).

**Corresponding author at: Thoracic Department, Fujian Medical University Union Hospital, Fuzhou 350001, China E-mail: lacustrian@163.com).

ORCID(s):

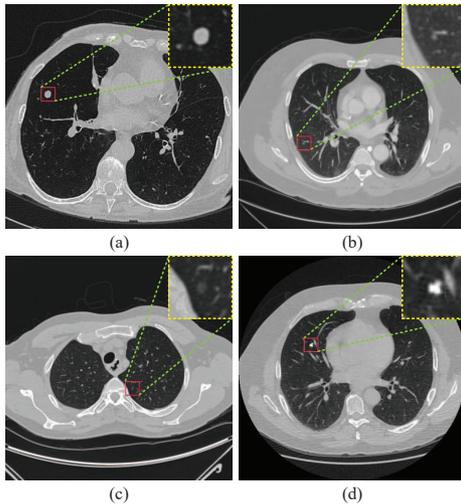


Figure 2: Examples of nodules labeled as solid nodule (a), mix ground-glass opacity nodule (b), ground-glass opacity nodule (c), and calcified nodule (d). The red rectangles emphasize the nodule locations and the yellow dashed rectangles highlight the nodule areas. Figure best viewed in color.

risk mainly according to the shape and density of the nodules as well as other pathology information. Qualitatively, compared to the benign nodules, the malignant ones often have larger volumes, varied density, and irregular shapes. Examples of benign and malignant nodules are illustrated in Fig.1. The inference results of the DNNs, therefore, lack confidence and interpretation.

To overcome the problems mentioned above, we propose a joint radiology analysis and malignancy evaluation network (R2MNet) that evaluates nodule malignancy according to radiology analysis. Specifically, radiology analysis aims to classify nodules as solid nodules (SN), ground-glass opacity nodules (GGO), mix GGO nodules (MGGO), and calcified nodules (CN) as shown in Fig.2. The purpose of malignancy evaluation is to estimate malignant risk of nodule. R2MNet consists of two sub-networks, the radiology analysis network (RNet) and the malignancy evaluation network (MNet) to implemented these two task, respectively. To consolidate the two sub-networks, we design assisted gating units (AGUs) embedded in the MNet to transform the feature maps extracted by RNet as a channel descriptor to capture channel dependencies of that by MNet. Moreover, model interpretability is crucial in CAD. To enable our model explainable, we propose channel-dependent activation mapping (CDAM) that adopts channel dependencies of activation maps themselves for features interpretation. Extensive experiments on LIDC-IDRI [11] via five-fold cross-validation demonstrate that the proposed R2MNet achieves satisfactory performance on nodule malignancy evaluation. Moreover, its inference process conforms to clinical diagnosis procedure which increases the confidence level of evaluation results. Our contributions can be summarized as follows:

- We propose R2MNet that integrates two sub-networks

(RNet and MNet) to inference malignant risk via radiology analysis. The RNet extracts radiological feature using new labeled data. MNet evaluates nodule malignancy.

- To conjoin the two sub-networks of R2MNet, we design the AGUs embedded in MNet to transform the feature maps extracted by RNet as a channel descriptor to capture channel dependencies of that by MNet.
- To enable our model interpretable, we propose CDAM that exploits channel dependencies of the activation maps for visualizing explanation.
- Extensive experiments on the LIDC-IDRI dataset indicate that our method achieves promising accuracy for nodule malignancy evaluation. Remarkably, the inference process conforms to clinical diagnosis procedure.

The rest of this paper is organized as follows. In Section 2, we review the relevant literature. Datasets and their corresponding preprocessing are specified in Section 3. Section 4 elaborates on the proposed methods. Experiments setting and results are shown in Section 5. In Section 6, we discuss the experiment results and analyze the superiority and limitations of our approach. Section 7 concludes this paper.

2. Related work

In the following, we review the works related to pulmonary nodule classification, long-range dependencies, and Class Activation Map (CAM)-based explanation.

2.1. Pulmonary nodule classification

In a deep learning-based CAD system, nodule classifiers either reduce false-positive nodules following nodule detectors or evaluate nodule malignancy in the back of the CAD systems. Setio et al. extracted 2D patches from nine symmetrical planes of a cube for false positive reduction [12]. Dou et al. encoded multi-level context information with 3D Convolutional Neural Network (CNN) to reduce false positives [13]. MD-NDNet integrated nodule volumetric information and spatial nodule correlation features from sagittal, coronal, and axial planes to decrease false positive rate [14]. Winkels et al. developed a 3D version of group equivariant convolutional networks that generalizes automatically over discrete rotations and reflection for false-positive reduction [15]. False-positive reduction using CNNs that identifies input CT volumes whether have nodules or not conforms to the clinical basis. However, nodule benign/malignant evaluation directly from CT to malignant probability lacks interpretation of features extracted by CNN [9, 10]. To improve model interpretability, Hussein et al. adopted multiple CNNs based on graph regularized sparse multi-task learning for malignant risk stratification [16]. Similarly, Wu et al. integrated the tasks including classification and segmentation in a multi-task learning manner [17]. In this work, we exploit radiological features as a channel descriptor for nodule

malignancy evaluation. Besides, we employ the proposed CDAM for model explanation. Overview of the proposed model is introduced in Section 4.1.

2.2. Long-range dependencies

Learning long-range dependencies is of great importance in deep neural networks. Long-range dependencies enable networks to capture large receptive field and learn global features. Convolutions are local operations in which long-range dependencies can only be captured when these operations are applied repeatedly. The transformer was one of the first attempts to apply a self-attention mechanism to model long-range dependencies in machine translation [18]. Non-local operation captured the pixel-level pairwise relations for solving computer vision [19]. GCNet improved the Non-local network with less computation while maintained the effectiveness of long-range dependencies capturing [20]. To learn channel-wise dependencies of feature maps, SENet recalibrated the channel dependency with global context features as each channel of feature maps corresponding to the specific region of the input image [21]. Motivated by the superiority of SENet, we propose a AGU for recalibrating channel relationships using specific features as a channel descriptor. Details of the AGU are presented in Section 4.2.

2.3. CAM-based Explanation

Activation maps visualization has been the most mainstream method for CNN interpretation. Specifically, the Class Activation Map (CAM) is one of the widely adopted methods [22]. CAM-based explanations provide feature visualization for explanations with a weighted combination of activation maps learned from CNN [22, 23, 24, 25].

CAM identified discriminative regions by a linear weighted combination of activation maps of the last convolutional layer before the global pooling layer [22]. However, it is only appropriate for a restricted class of CNNs that contain global average pooling layers and fully connected layers. To extend the range of application of CAM, Grad-CAM generalized the definition of the weighting coefficients as the gradient of class confidence concerning the activation map and applies to a significantly broader range of CNN model families [23]. The variation of Grad-CAM, Grad-CAM++ aimed to provide better localization of objects as well as explaining occurrences of multiple objects of a class in a single image [24]. Using gradient to incorporate the importance of each channel towards the class confidence is a natural choice. The gradient information for a deep neural network can be noisy and also tends to vanish due to saturation in sigmoid or the flat zero-gradient region in Rectified Linear Unit (ReLU). Instead of using the gradient information flowing into the last convolutional layer to represent the importance of each activation map, Score-CAM exploited the importance as the linear combination of score-based weights and activation maps [25]. However, the aforementioned methods adopted weighting coefficients derived from external data, which may introduce noise and bias. Therefore, we propose the CDAM for activation maps visualization where the weighting coefficients are calculated from the activation maps

themselves. Details of the CDAM are presented in Section 4.3.

3. Materials

In this section, we introduce the database used in our experiments. Data annotation and preprocessing methods are also specified.

3.1. Dataset

In this study, we use a selected version of the LIDC database [11] provided in the LUNA16 challenge [26] which consists of 888 CT scans comprising a total of 1186 nodules. We have obtained the nodules malignancy from the annotation files in the LIDC-IDRI dataset. Nodules with an average score higher than 3 were labeled as malignant and lower than 3 are labeled as benign. Some nodules were removed from the experiments in the case of the averaged malignancy score 3, ambiguous IDs, and rated by only one or two radiologists, which resulted in a total of 1004 nodules where there were 450 malignant nodules and 554 benign nodules.

3.2. Radiological categories annotation

For nodule radiological analysis, two experienced radiologists labeled nodules as SN, GGO, MGGO, CN according to the 3D radiological features of LDCT scans using ITK-SNAP [27]. The steps of data annotation are briefly listed as follows.

- 1) Two experienced doctors, respectively, marked the class of all nodules based on radiological characteristics.
- 2) Then, they carefully inspected and corrected the labeled results, respectively.
- 3) The final version was obtained by discussing and remark the different classes labeled in previous steps.

3.3. Preprocessing

The data preprocessing follows four steps.

- 1) **Normalization.** We clipped hounsfield units (HU) of the raw CT data into $[-1200, 600]$ and normalized them into $[0, 1]$.
- 2) **Extraction.** Foreground regions of normalized CT scans were extracted according to the ground truth masks provided by LUNA 16 challenge.
- 3) **Resample.** We resampled all CT volumes to have 1 mm spacing in the z -, y -, x -dimension.
- 4) **Crop.** The nodule regions used to train and test our method were cropped according to the experiment configurations (i.e. 2D/3D format and size)

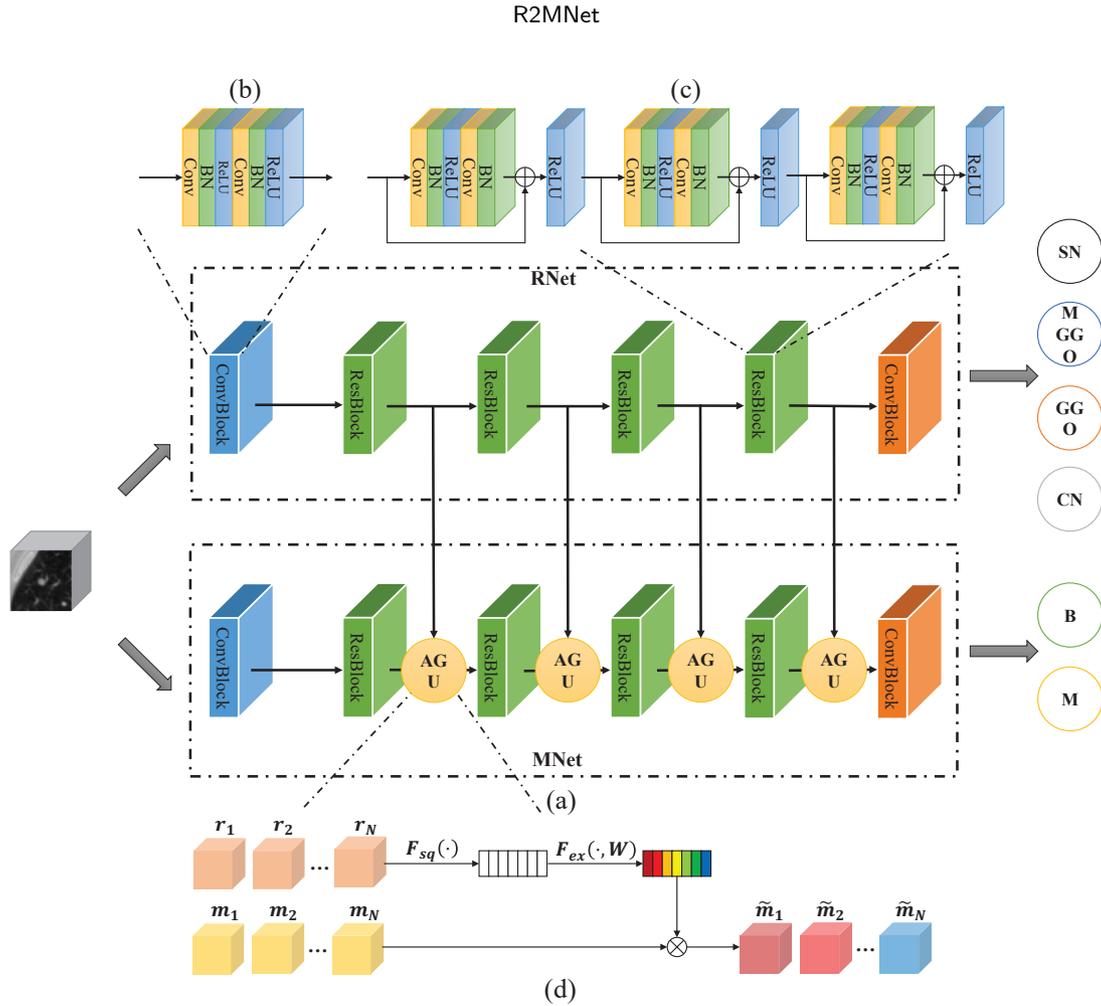


Figure 3: The diagram of the proposed method. (a) R2MNet. Note that we omit the four max pooling layers each of which is behind the Residual blocks for illustration convinience. (b) The convolutional block. (c) The residual blocks. (d) The AGU module.

4. Methods

In this section, we introduce our R2MNet and detail its components. Then, the implemented details are presented. The proposed R2MNet are shown in Fig.3. The diagram of R2MNet is illustrated in Fig.3(a). R2MNet is composed of two CNN trained in multi-task learning manner (Section 4.1). The AGU transforms radiological features into a channel descriptor to facilitate malignancy evaluation (Fig.3(d)). CDAM are proposed for model explanation (Fig.4).

4.1. R2MNet

Here we present our R2MNet and provide an overview of the key components. The proposed R2MNet takes a 3D CT volume of as input and provides as outputs a radiology class and a nodule malignant score. Specifically, the R2MNet consists of two improved residual networks [28], i.e., RNet and MNet as illustrated in Fig.3 (a). MNet includes two convolutional blocks(Fig.3 (b)), four residual blocks each of which contains three residual units (Fig.3 (c)), four AGUs (Fig.3 (d)), and four max-pooling layers. The architecture of RNet is similar to MNet but without AGUs. The proposed

method can combine nodule radiological features for nodule malignancy evaluation. The RNet and MNet are trained simultaneously in a multi-task learning manner. This is different than current approaches that use directly a CNN for malignancy estimation [9, 10]. The goals of RNet are extracting radiological features of pulmonary nodules for nodule evaluation as well as providing the radiological characteristics as a reference for practice diagnosis. The outputs of RNet are four categories probabilities and radiological features. The radiological features are transformed into a channel descriptor by AGU (Section 4.2) to render the MNet focus on nodule area. MNet takes as inputs the CT volume data and the radiological features for pulmonary nodule malignancy evaluation. The loss function for training our networks is weighted cross-entropy (CE) loss.

$$L(Y_r, Y_m, \hat{Y}_r, \hat{Y}_m) = \lambda L_{CE}(Y_r, \hat{Y}_r) + (1-\lambda) L_{CE}(Y_m, \hat{Y}_m) \quad (1)$$

, where Y_r and Y_m are the ground truth, and \hat{Y}_r and \hat{Y}_m are predictions of the R2MNet.

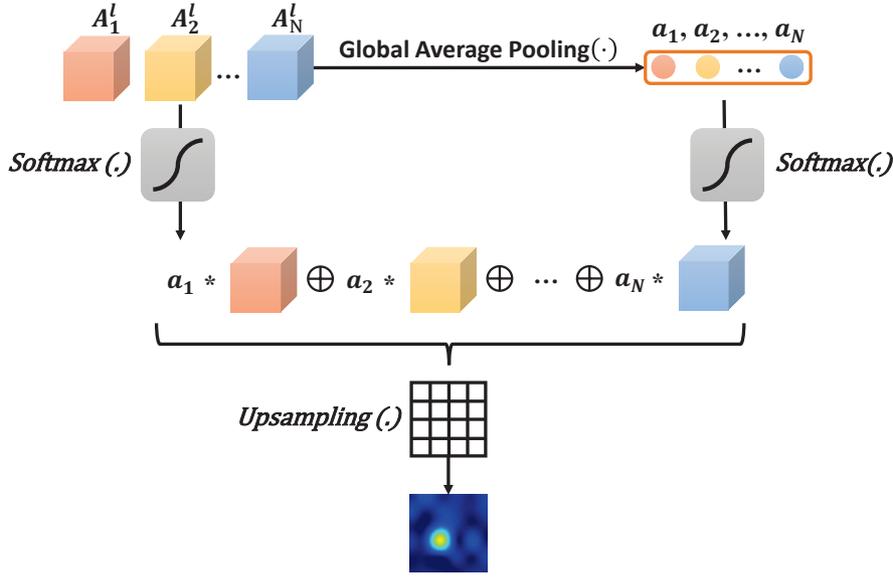


Figure 4: The diagram of the proposed CDAM. Activation maps are linearly weighted to generate visual explanation.

4.2. Assisted Gating Unit

The vanilla SE layer [21] adopted the input features to capture channel dependencies in 2D scenario. The AGU, instead, transforms the features extracted by RNet as a channel descriptor to capture channel dependencies of that by MNet in 3D scenario. The diagram of AGU is shown in Fig. 3 (d). Specifically, we transform radiological features into a channel descriptor to capture the channel dependencies of malignancy features. Similar to the SE block, we model channel interdependencies to recalibrate filter responses in two steps (i.e., squeeze and excitation) as discussed follows.

- 1) **Squeeze.** Squeeze operations are adopted for global information embedding. In R2MNet, radiological features $R = [r_1, r_2, \dots, r_N]$ are squeezed to a channel descriptor by using global average pooling (GAP). Noting that more sophisticated aggregation strategies could be employed here as well, we adopt GAP as used in [21]. The channel descriptor $T = [t_1, t_2, \dots, t_N] \in \mathbb{R}^C$ is computed as:

$$t_n = F_{sq}(r_n) = \frac{1}{D \times H \times W} \sum_{i=1}^D \sum_{j=1}^H \sum_{k=1}^W r_n(i, j, k) \quad (2)$$

where F_{sq} is the squeeze operation, and D , H and W denotes the depth, height and width of the feature maps.

- 2) **Excitation.** The following operation takes as input the information aggregated in the last step to capture channel dependencies, i.e. $S = [s_1, s_2, \dots, s_N] \in \mathbb{R}^C$. The excitation operation can be formulated as follows:

$$S = F_{ex}(T, W) = \sigma(g(T, W)) = \sigma(W_2 \delta(W_1 T)) \quad (3)$$

where σ is the sigmoid function and δ refers to ReLU activation, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$. Similar to [21], we form a bottleneck including a dimensionality-reduction layer with parameters W_1 with reduction ratio r , a ReLU activation, and then a dimensionality-increasing layer with parameters W_2 . Finally, the recalibrated malignant features $\tilde{M} = [\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_N] \in \mathbb{R}^C$ are obtained by rescaling the malignant features $M = [m_1, m_2, \dots, m_N] \in \mathbb{R}^C$ with the radiological channel descriptors T :

$$\tilde{m}_n = F_{scale}(t_n, s_n) = s_n \cdot t_n \quad (4)$$

where F_{scale} denotes channel-wise multiplication.

4.3. Channel-Dependent Activation Mapping

We propose CDAM for 3D features visualization motivating by CAM-based methods, as shown in Fig.4. CAM is a technique for identifying discriminative regions by a linearly weighted combination of activation maps of the last convolutional layer before the global pooling layer [22]. The motivation behind CAM is that each activation map of a CNN layer contains different spatial information about the input X and the importance of each channel is the weight of the linear combination of the fully connected layer following the global pooling. However, if there is no global pooling layer or there is no fully connected layers, CAM will not apply due to no definition of the weighted coefficients. Grad-CAM [23] and its variations [24] generalize CAM to models without global pooling layers by employing gradients as weights.

Instead of using weights of the fully connected layer or gradient information derived from external layers, CDAM

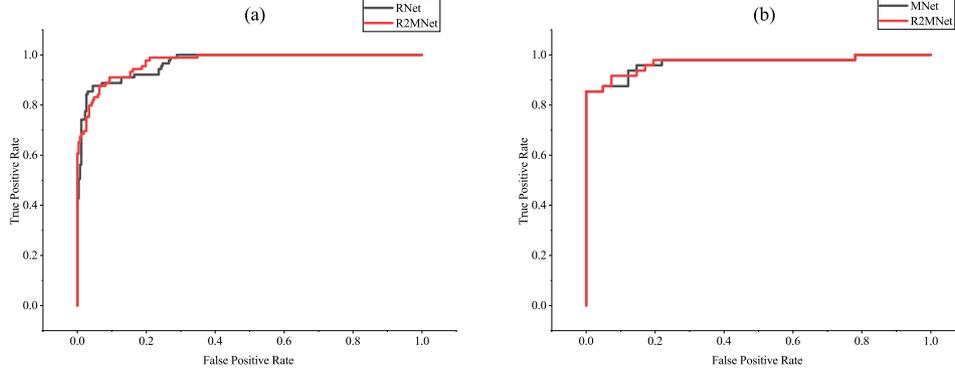


Figure 5: ROC curves of RNet and R2MNet on the radiology analysis (a), and malignancy evaluation (b).

employs activation maps themselves to obtain weights for a linear combination of activation maps. Formally, CDAM is defined as:

$$L_{CDAM} = \text{ReLU} \left(\sum_{i=1}^C \alpha_i A_i^l \right) \quad (5)$$

where A^l denotes the activations of the l th CNN layer, A_i^l refers to the activation map for the i th channel of A^l , and $a = [a_1, a_2, \dots, a_n] \in \mathbb{R}^C$ is defined as:

$$a_n = \frac{1}{D \times H \times W} \sum_{i=1}^D \sum_{j=1}^H \sum_{k=1}^W A_n^l(i, j, k) \quad (6)$$

, We apply a ReLU activation to the linear combination of maps because we are only interested in the features that have a positive influence. Both a_l and A_l are utilized after the Softmax activation because the relative output value after normalization is more reasonable to measure the relevance than the absolute output value. Furthermore, to capture voxel-wise importance, we up-sample L_{CDAM} to the input resolution using bicubic interpolation.

4.4. Implemented details

The network were performed on PyTorch [29]. The models were trained via Adam optimizer [30] with standard back-propagation. Data augmentation operations i.e., scaling, flip, and rotation were also employed in the experiments. The learning rate was set as a fixed value of $1e-4$ and the number of epochs was 100. The networks were trained on a single NVIDIA GeForce GTX 1080Ti.

5. Experiments and results

In this section, we evaluate the proposed R2MNet on the LIDC-IDRI database and show the results. First, we performed nodule characteristics identification and malignancy evaluation individually. Then, we combined two tasks in

Table 1

Performance comparison of RNet and R2MNet on radiology analysis.

Model	SN	MGGO	GGO	CN	AUC
RNet	95.50	89.88	91.01	96.63	95.21
R2MNet	96.63	92.13	91.01	97.75	97.08

multi-task learning where radiology analysis assisted malignancy evaluation. For model explanations, we visualized the feature maps and analyzed their characteristics. Experimental results show that the proposed method achieved higher performance compared to the baseline.

5.1. Nodule Radiology analysis

Nodule radiology analysis aims to classify nodules as SN, MGGO, GGO, and CN nodules. Identifying these characteristics renders the model to learn radiological features for facilitating malignant evaluation. In addition, these characteristics can assist radiologists in determining nodule attributes as well. Experimental results of nodule characteristics classification are listed in Table 1. Both RNet and R2MNet achieved accuracy higher than 90% among the four categories. After combined with MNet, the performance of R2MNet either remained the accuracy level of RNet (GGO, CN) or was higher than that of RNet (SN, MGGO). Also, the area under curve (AUC) of R2MNet is larger than that of RNet. According to the Fig.5 (b), the ROC curve of R2MNet nearly surrounds that of the RNet.

5.2. Nodule malignancy evaluation

Radiological features of pulmonary nodules can assist CNN for malignant classification because the inference procedure conforms to the diagnosis process. To testify the effectiveness of the proposed method, we conducted experiments of nodule malignant classification. As shown in Table 2, R2MNet outperforms MNet with an accuracy gain of 1.72% and an AUC gain of 1.38%, respectively. Moreover, the accuracy and AUC of R2MNet are more stable

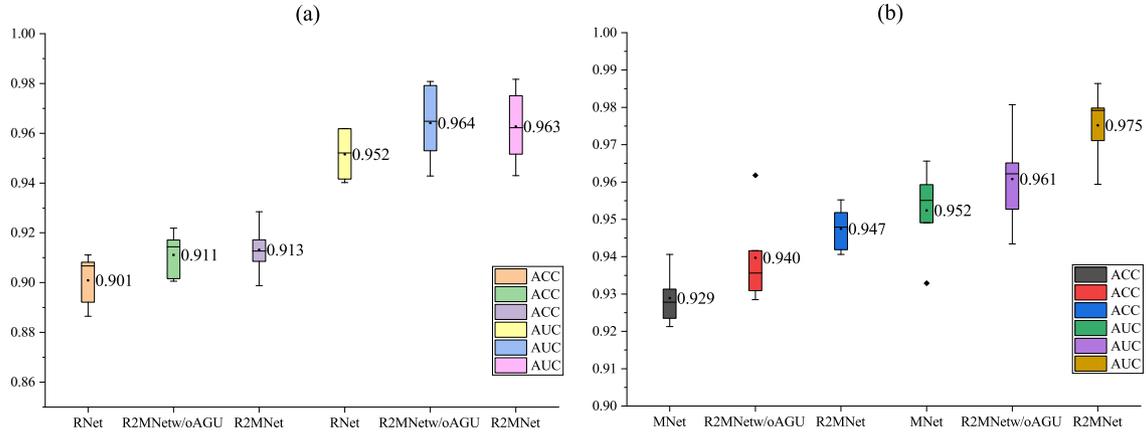


Figure 6: Comparison among RNet, MNet, R2MNetw/oAGU, and R2MNet with accuracy and AUC on radiology analysis (a) and malignancy evaluation (b), respectively. The first three columns are the accuracy boxes and the remaining are AUC ones. Each scalar in the left of the corresponding boxes is the average value.

Table 2

Performance comparison measured by accuracy and AUC ($mean \pm s.d.\%$) for MNet, R2MNetw/oAGU, and R2MNet on radiology analysis and malignancy evaluation.

Task	Radiological analysis		Malignant evaluation	
	Accuracy	AUC	Accuracy	AUC
MNet	90.82% \pm 1.09	95.15% \pm 1.05	92.89% \pm 0.76	95.24% \pm 1.25
R2MNet_w/oAGU	91.11% \pm 0.95	96.41% \pm 1.64	93.97% \pm 1.33	96.08% \pm 1.40
R2MNet	91.13% \pm 1.10	96.27% \pm 1.60	94.74% \pm 0.62	97.52% \pm 1.04

compared to MNet according to the standard deviation. The ROC curves of MNet and R2MNet are depicted in Fig.5 (c). To compare the overall performance of MNet and R2MNet through five-fold cross-validation, we also illustrated the box plots with accuracy and AUC in Fig.6. As shown, compared to MNet, R2MNet achieved more stable and higher results.

5.3. Ablation study

We conducted an ablation study to investigate the individual contributions of R2MNet and AGU module. We implemented the experiments both on radiology analysis and malignant evaluation. The experiments were performed from two ends; on the one hand, we just included the radiology analysis in nodule malignant evaluation, which resulted in a fundamental version of R2MNet (i.e., R2MNetw/oAGU). On the other hand, the AGU modules were introduced into the preliminary R2MNet to construct the final version of the proposed method (i.e., R2MNet).

In nodule radiology analysis, a comparison was made among RNet, R2MNetw/oAGU, and R2MNet. As indicated in Table 2 the accuracy and AUC scores of the R2MNetw/oAGU are similar to that of R2MNet. Both of them slightly outperforms RNet. Results are shown in Fig.6(a).

In nodule malignancy evaluation, a comparison was implemented among MNet, R2MNetw/oAGU, and R2MNet.

The results of the five-fold cross-validation are listed in Table 2. We can observe from the table that combining radiological analysis with malignant evaluation improves performance over doing the latter only. Further, when AGU is introduced into R2MNet, the synergy between these two components generates the best performance. The illustration of these results is shown in Fig.6 (b).

5.4. Model interpretation

Direct approaches that classify pulmonary nodule as benign or malignant from input CT data to the malignant probabilities lack of interpretation. To build explainable models, we provided visual explanations using the proposed CDAM. The experiments were performed both on malignant evaluation and radiology analysis to investigate voxel-wise importance regions which the models focus on in different tasks. Specifically, we employed the feature maps with a size of $256 \times 6 \times 6 \times 6$ after the last residual block in our model as activation maps. Since the activation maps are volume data, we adopted the center slice for visualization convenience. Fig.7 shows the CDAM features and its corresponding probabilities of MNet, R2MNetw/oAGU, and R2MNet concerning nodule malignant evaluation, respectively. The value below each sub-figure is the probability predicted by the corresponding model. Besides, we illustrated the CDAM features

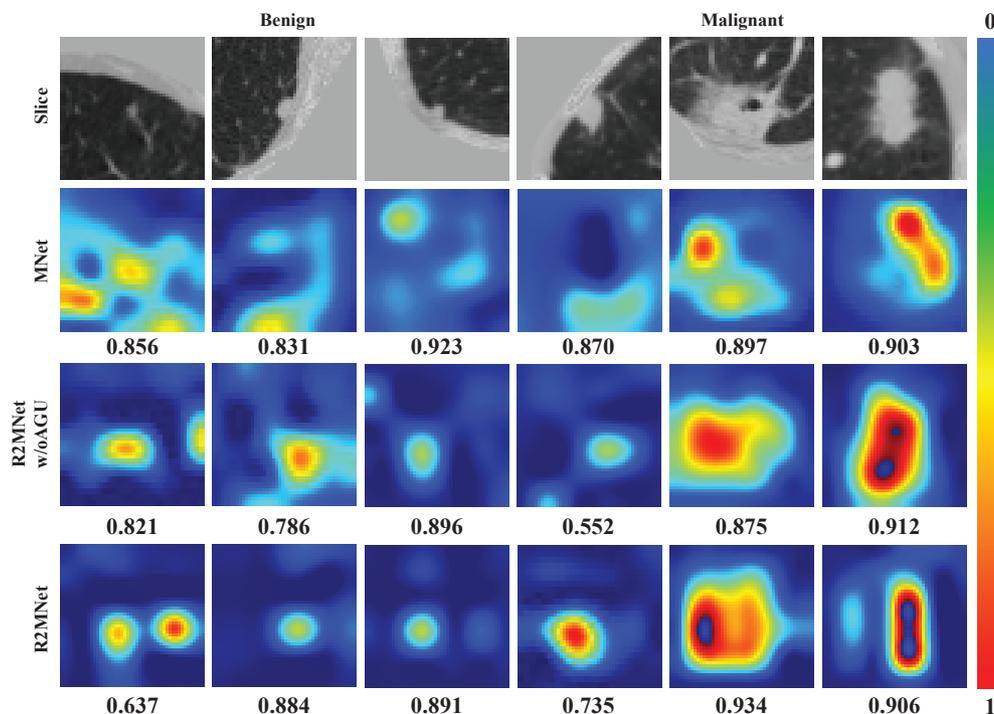


Figure 7: Visualization of CDAM features derived from MNet, R2MNetw/oAGU, and R2MNet regarding malignant evaluation, respectively. The value under each sub-figure is probability predicted by the corresponding model. Note that we show the central slice only for visualizing convenience. Figure best viewed in color.

with respect to nodule radiology analysis in Fig.8.

6. Discussion

Automatic pulmonary nodule malignancy evaluation is an essential component of a CAD system for lung cancer diagnosis. Deep learning-based methods have demonstrated promising results on this task. Table 3 summarizes the related works from the literature. Shen et al. introduced a Multiscale CNN for nodule malignancy diagnosis and achieved an accuracy of 86.84% on a selected LIDC-IDRI dataset [31]. Nibali et al. adopted ResNet with multiview inputs for benign/malignant classification [9]. They evaluated their method on the dataset derived from the LIDC-IDRI and achieved an accuracy of 89.90%. Al-Shabi et al. employed non-local blocks to model nodule global features and residual blocks to capture local features of nodule [10]. They estimated the model on the selected LIDC-IDRI database with accuracy of 88.46%. However, classifying lung nodules as benign or malignant directly from the CT volume (or slice) lack clinical basis and explanations of the features extracted by the CNN. Therefore, the results are short of confidence level. Hussein et al. empirically established the significance of different high-level nodule attributes for malignancy determination [32]. Furthermore, they adopted CNNs to learn a series of features for nodule attributes then fused these features to predict the malignancy of pulmonary nodule in a multi-

Table 3

Overview of previous methods for pulmonary nodule evaluation. Abbreviations: Information Processing in Medical Imaging (IPMI), International Symposium on Biomedical Imaging (ISBI), International Journal of Computer Assisted Radiology and Surgery (IJCARS).

Methods	Accuracy
MCNN [31], IPMI	86.84%
TumorNet [32], ISBI	82.47%
TumorNet (Attributes) [32], ISBI	92.31%
Nodule-ResNet [9], IJCARS	89.90%
MIT-3DCNN [16], IPMI	91.26%
PN-SAMP [17], ISBI	97.58%
Local-Global Networks [10], IJCARS	88.46%
R2MNet, ours	94.74%

task learning manner [16]. Similarly, Wu et al. proposed a multi-task learning CNN that integrated pulmonary nodule segmentation attributes and malignancy prediction [17]. Their approach simultaneously predicted the malignancy of lung nodules, segmented the nodule areas and learned nodule attributes, and aimed to tackle the problem of model interpretability. Note that it can be difficult to pursue an objective cross-study comparison due to the differences in datasets, initialization methods, and experimental settings.

Our method leveraged radiological features as a channel descriptor to assist lung nodule evaluation in a multi-task

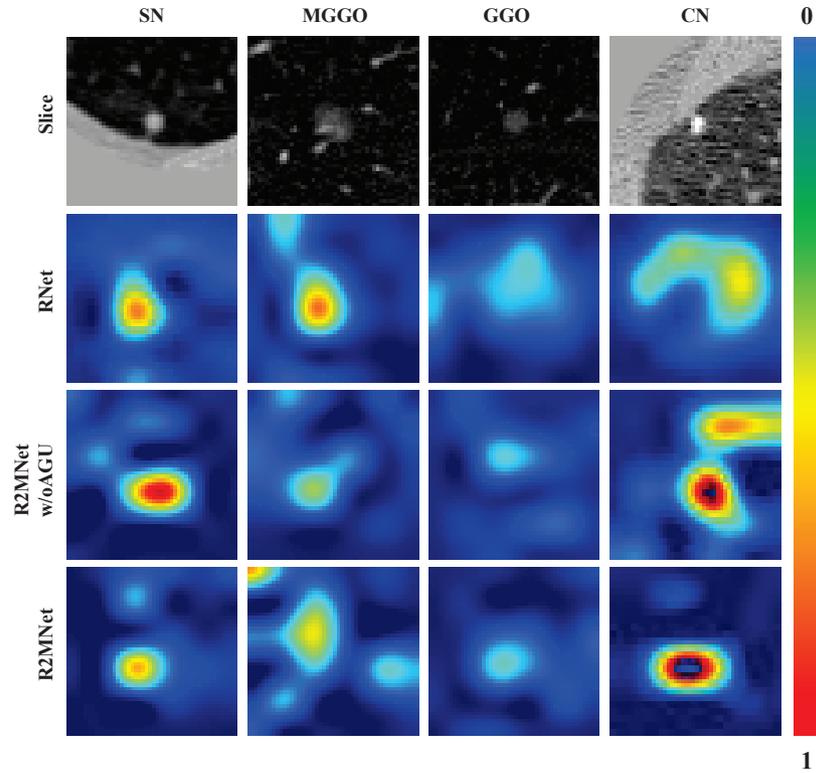


Figure 8: Visualization of CDAM features derived from RNet, R2MNet_w/oAGU, and R2MNet concerning radiology analysis, respectively. Note that we show the central slice only for visualizing convenience. Figure best viewed in color.

learning manner. Specifically, Table 1 indicates the results of the radiological analysis. Although radiology analysis is an auxiliary component in the R2MNet, R2MNet increased the accuracy among four nodule categories and the AUC score compared with RNet. Moreover, the ROC curves in Fig.5(b) where the curve of R2MNet nearly surrounds that of RNet illustrate that the classification performance of R2MNet is better than that of RNet. In nodule malignancy evaluation, Fig.5(c) depicts the ROC curves of MNet and R2MNet in which the curves of the latter are higher than that of the former. As indicated in Table 2, in general, joint learning of radiology analysis and malignancy evaluation improved the performance compared to each individual. Combined learning facilitates communication between different tasks. We can conclude that these two tasks reinforce each other. Furthermore, comparing R2MNetw/oAGU with MNet, we can view the accuracy and AUC gain both in the two tasks. The performance of R2MNet on radiological analysis is nearly equal to that of R2MNetw/oAGU. It is reasonable because the AGU module adopted radiological features to facilitate nodule malignancy evaluation. Indeed, the performance gain was obtained by R2MNet in malignancy estimation. On the other hand, Fig.6 depicts the box plots with average values and data distribution. The accuracy scores and AUC scores increase gradually among MNet, RNet, R2MNetw/oAGU, and R2MNet, which further proves the effectiveness of the proposed methods. Viewing the boxes of MNet/RNet and R2MNetw/oAGU,

one can conclude that although multi-task learning can bring performance gain, the results tend to fluctuate due to the hard convergence of the networks. However, the results of R2MNet are stable compared to others because introducing AGU into R2MNetw/oAGU enables the R2MNet to employ radiological features and then improve the adaptability of the model to different data.

Although performance improvement is one of a great purpose in developing deep learning-based methods, interpretability is essential as well. According to the experiences of radiologists, the shape and density of nodule regions are two critical factors that influence a nodule to be inferred as malignant. Fig.7 shows the CDAM features of MNet, R2MNetw/oAGU, and R2MNet concerning nodule malignancy evaluation, respectively. MNet tended to be disturbed by the background noise and confused with benign and malignant features. In contrast, both R2MNet and R2MNetw/oAGU can focus on nodule regions except that they yielded a wrong identification in the first benign nodule. Furthermore, these two architectures paid higher attention to malignant nodules and lower attention to benign ones, which conforms to the risk of the nodules. According to the last two columns of benign and malignant nodules in Fig.7, even though the MNet generated high probabilities, similar to other models, the concerning regions of MNet slightly deviate from the ground truth. On the contrary, R2MNet predicted low scores when it falsely located the nodule region, whereas MNet still gen-

erated high probability (Fig.7, the first column). We can conclude that incorporating malignancy evaluation with radiology analysis can render the network emphasize nodule regions and characterize the shape and density features of nodules. Besides, the density of nodules plays a key role for nodule radiology analysis. As shown in Fig.8, even though the four classes of nodules have different densities, the boundaries among them are confused, which led both the RNet and R2MNet/oAGU to locate the nodule regions inaccurately. On the contrary, the R2MNet accurately located the nodules and lay different emphasis on these regions according to their densities, conforming with the clinical basis. Therefore, we can conclude that even though the results of R2MNet and R2MNet/oAGU are similar, the inference process of R2MNet is more reasonable.

A major limitation of this work is that the input data depend on pulmonary nodule detection. The input data are derived from either manually choosing by radiologists or automatic detection by nodule detectors. Previous researches integrated multi-models into a synthetic system whose components were trained separately to performed different tasks. For example, Bonavita et al. developed a lung cancer classification pipeline that integrated a 3D CNN with an existing nodule detection framework [33]. Liao et al. adopted a 3D Faster R-CNN for patch-based nodule detection and integrated the leaky noisy-OR model into neural networks to solve lung cancer prediction [7]. Similarly, Zhu et al. build a DeepLung system to identify suspicious nodules and predict nodule malignancy [6]. Ozdemir et al. introduced a CAD system that included two sub-systems for nodule candidates segmentation and malignancy prediction [8]. An end-to-end explainable CAD system for lung cancer diagnosis that integrates nodule detection, segmentation and malignancy prediction is of extensive clinical application value. This will be considered as our future work.

7. Conclusion

In this paper, we proposed the R2MNet that evaluated pulmonary nodule malignancy resorting to radiology analysis instead of directly infer malignant probability, which conformed to the clinical diagnosis procedure and increased the confidence of prediction results. Specifically, the radiological features were transformed into a channel descriptor that emphasized the informative malignant features and suppressed the less useful ones, so that the network could estimate the malignant risk based on radiological characteristics as did an experienced doctor to a patient. Besides, model explanations with CDAM shed light on the voxel-wise nodule regions which CNNs focussed on when they estimated nodule malignancy risk. The experimental results on the LIDC-IDRI database demonstrated the effectiveness of the proposed R2MNet.

Acknowledgement

This work was supported by the Natural Science Foundation (Grant No. 2020J01472) and Provincial Science and

Technology Leading Project (Grant No.2018Y0032) of Fujian Province, China. This work was also supported by Fujian Key Laboratory of Cardio-Thoracic Surgery (Fujian Medical University)

References

- [1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012, *International journal of cancer* 136 (2015) E359–E386.
- [2] I. Sluimer, A. Schilham, M. Prokop, B. Van Ginneken, Computer analysis of computed tomography scans of the lung: a survey, *IEEE transactions on medical imaging* 25 (2006) 385–405.
- [3] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [4] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: learning dense volumetric segmentation from sparse annotation, in: *International conference on medical image computing and computer-assisted intervention*, Springer, 2016, pp. 424–432.
- [5] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: *2016 fourth international conference on 3D vision (3DV)*, IEEE, 2016, pp. 565–571.
- [6] W. Zhu, C. Liu, W. Fan, X. Xie, Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification, in: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2018, pp. 673–681.
- [7] F. Liao, M. Liang, Z. Li, X. Hu, S. Song, Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network, *IEEE transactions on neural networks and learning systems* 30 (2019) 3484–3495.
- [8] O. Ozdemir, R. L. Russell, A. A. Berlin, A 3d probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans, *IEEE Transactions on Medical Imaging* 39 (2019) 1419–1429.
- [9] A. Nibali, Z. He, D. Wollersheim, Pulmonary nodule classification with deep residual networks, *International journal of computer assisted radiology and surgery* 12 (2017) 1799–1808.
- [10] M. Al-Shabi, B. L. Lan, W. Y. Chan, K.-H. Ng, M. Tan, Lung nodule classification using deep local-global networks, *International journal of computer assisted radiology and surgery* 14 (2019) 1815–1819.
- [11] 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 (lide) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans, *Medical physics* 38 (2011) 915–931.
- [12] 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, B. van Ginneken, Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks, *IEEE transactions on medical imaging* 35 (2016) 1160–1169.
- [13] Q. Dou, H. Chen, L. Yu, J. Qin, P.-A. Heng, Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection, *IEEE Transactions on Biomedical Engineering* 64 (2016) 1558–1567.
- [14] Z. Wu, R. Ge, G. Shi, L. Zhang, Y. Chen, L. Luo, Y. Cao, H. Yu, Md-ndnet: a multi-dimensional convolutional neural network for false-positive reduction in pulmonary nodule detection, *Physics in Medicine & Biology* 65 (2020) 235053.
- [15] M. Winkels, T. S. Cohen, Pulmonary nodule detection in ct scans with equivariant cnns, *Medical image analysis* 55 (2019) 15–26.
- [16] S. Hussein, K. Cao, Q. Song, U. Bagci, Risk stratification of lung nodules using 3d cnn-based multi-task learning, in: *International conference on information processing in medical imaging*, Springer, 2017, pp. 249–260.
- [17] B. Wu, Z. Zhou, J. Wang, Y. Wang, Joint learning for pulmonary

- nodule segmentation, attributes and malignancy prediction, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 1109–1113.
- [18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.
- [19] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7794–7803.
- [20] Y. Cao, J. Xu, S. Lin, F. Wei, H. Hu, Gcnet: Non-local networks meet squeeze-excitation networks and beyond, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019, pp. 0–0.
- [21] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
- [22] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2921–2929.
- [23] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.
- [24] A. Chattopadhyay, A. Sarkar, P. Howlader, V. N. Balasubramanian, Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018, pp. 839–847.
- [25] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, X. Hu, Score-cam: Score-weighted visual explanations for convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 24–25.
- [26] 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, *Medical image analysis* 42 (2017) 1–13.
- [27] P. A. Yushkevich, Y. Gao, G. Gerig, Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016, pp. 3342–3345.
- [28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, in: Advances in neural information processing systems, 2019, pp. 8026–8037.
- [30] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
- [31] W. Shen, M. Zhou, F. Yang, C. Yang, J. Tian, Multi-scale convolutional neural networks for lung nodule classification, in: International Conference on Information Processing in Medical Imaging, Springer, 2015, pp. 588–599.
- [32] S. Hussein, R. Gillies, K. Cao, Q. Song, U. Bagci, Tumornet: Lung nodule characterization using multi-view convolutional neural network with gaussian process, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE, 2017, pp. 1007–1010.
- [33] I. Bonavita, X. Rafael-Palou, M. Ceresa, G. Piella, V. Ribas, M. A. G. Ballester, Integration of convolutional neural networks for pulmonary nodule malignancy assessment in a lung cancer classification pipeline, *Computer methods and programs in biomedicine* 185 (2020) 105172.