


Article

3D Spatial Pyramid Dilated Network for Pulmonary Nodule Classification

Guokai Zhang ¹, Xiao Liu ¹, Dandan Zhu ¹, Pengcheng He ¹, Lipeng Liang ¹, Ye Luo ^{1,*} 
and Jianwei Lu ^{1,2,*}

¹ School of Software Engineering, Tongji University, Shanghai 201804, China; zhang.guokai@tongji.edu.cn (G.Z.); xiaoliu1990@hotmail.com (X.L.); dandanzhu@tongji.edu.cn (D.Z.); sherlockhpc@163.com (P.H.); gaosaroma@outlook.com (L.L.)

² Institute of Translational Medicine, Tongji University, Shanghai 201804, China

* Correspondence: yeluo@tongji.edu.cn (Y.L.); jwlu33@tongji.edu.cn (J.L.); Tel.: +86-21-6958-9585 (Y.L.)

Received: 6 August 2018; Accepted: 22 August 2018; Published: 1 September 2018



Abstract: Lung cancer mortality is currently the highest among all kinds of fatal cancers. With the help of computer-aided detection systems, a timely detection of malignant pulmonary nodule at early stage could improve the patient survival rate efficiently. However, the sizes of the pulmonary nodules are usually various, and it is more difficult to detect small diameter nodules. The traditional convolution neural network uses pooling layers to reduce the resolution progressively, but it hampers the network's ability to capture the tiny but vital features of the pulmonary nodules. To tackle this problem, we propose a novel 3D spatial pyramid dilated convolution network to classify the malignancy of the pulmonary nodules. Instead of using the pooling layers, we use 3D dilated convolution to learn the detailed characteristic information of the pulmonary nodules. Furthermore, we show that the fusion of multiple receptive fields from different dilated convolutions could further improve the classification performance of the model. Extensive experimental results demonstrate that our model achieves a better result with an accuracy of 88.6%, which outperforms other state-of-the-art methods.

Keywords: computer aided system; pulmonary nodule; dilated convolution; malignancy classification

1. Introduction

According to the statistics [1], lung cancer has been the leading cause of death among cancers. The reason for lung cancer's high mortality rate is because patients miss timely treatment opportunities at the early stage [2]. The early detection of lung cancer could efficiently improve the survival rate to 52% [3]. Currently, the most effective way to detect the cancer at the early stage is using low-dose computed tomography (CT) scans to detect the pulmonary nodules. These nodules are the small lung spherical tissue, which exhibits opacity. In clinical examinations, radiologists usually inspect all the images of CT scans to detect the pulmonary nodules. However, the CT images are three-dimensional, and the work of inspecting slices requires huge effort and time. A possible way to reduce the workload of radiologists is utilizing Computer Aided Detection (CAD) systems to detect the suspicious samples automatically. The CAD system (Figure 1) can not only help the radiologist detect the potential nodules candidates, but also reduce the missed diagnosis of nodules. In recent decades, many researchers have devoted great efforts towards achieving a more accurate and faster system to aid radiologists. In this paper, we focus on the work of designing an automatic classification model to classify the malignant and benign nodules. The classification task of malignant and benign nodules is an important part of the CAD system for diagnosing lung cancer. It could guide the doctor to assess the risk factor of each nodule and help design the following up treatment plan. Generally, the procedures of a CAD system

for lung cancer detection contain four stages: image processing, region of interest (ROI) extraction, feature selection, and classification. The feature selection and classification stages are usually regarded as the false positive reduction step. The main purpose of the false positive reduction step is to reduce the false positive as much as possible and improve the sensitivity and accuracy of the model. Therefore, it is the most important step in deciding the final effectiveness of a CAD system. Nowadays, most models are based on hand-crafted feature extraction and recognize the extracted features as the input of a classifier to classify, which is a main, basic way to carry out the false positive reduction step. Apart from the traditional hand-crafted feature extraction methods such as filter-based feature extraction, grey-level distribution, and intensity distribution, many other classical methods are also adopted in medical image analysis. For example, scale-invariant feature transform (SIFT) [4] extracts the scale-invariant features from the image, histogram of oriented gradients (HOG) [5] calculates the occurrences of gradient orientation in localized portions of the image, and local binary patterns (LBP) [6] learn the texture features from the input image. There are also some other attempts to combine different hand-crafted features to further improve the performance of the model. The other way to perform false positive reduction is to use convolution neural network (CNN). Compared with the hand-crafted feature extraction method, CNNs provide an end-to-end manner to automatically learn features that are more high-level and abstract from the original image data. The existing neural network architectures for image classification usually consist of convolution layers and pooling layers. The convolution layer is for the feature extraction with local connections and tied weights. The pooling layer can enlarge the receptive field of each filter and allow the network to extract more global and abstract features. For the large amounts of image data, CNN models usually show better performance than hand-crafted feature learning modes [7].

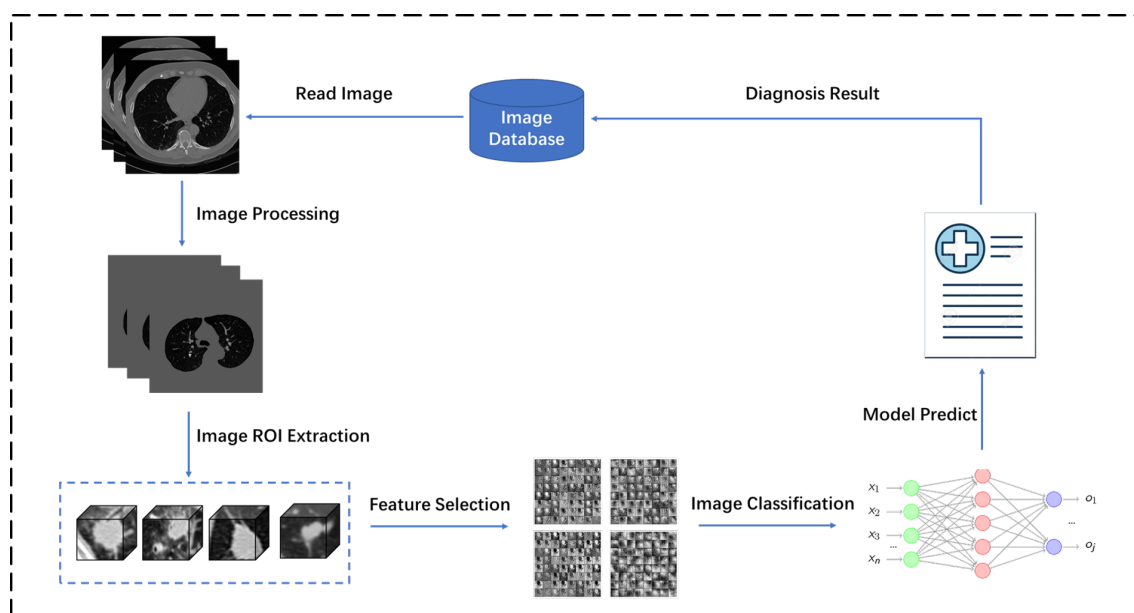


Figure 1. The flowchart of a Computer Aided Detection (CAD) system.

While these CNN models have done well in many traditional image classification tasks, for pulmonary nodule classification, the loss of spatial acuity and feature maps resolution by using the pooling layers could hinder the model to achieve better performance [8,9]. Figure 2 illustrates different types of pulmonary nodules. In the figure, we can see that, in the process of diagnosing diverse characteristic nodules, it is crucial to precisely extract features from the fine scales or even at the pixel level. Nevertheless, if we adopt the pooling layer as one unit of the network, the loss of tiny and detailed information of nodules could be a handicap for classifying the nodules, especially for the small diameter nodules. Moreover, the receptive field of each filter plays a significant role in nodule

classification task [10]. Different size of receptive fields contains discriminative scale features which could also be important in dealing with object appearance variations on the classification problem. In this paper, we propose a novel 3D neural network model to classify the malignancy suspicious of pulmonary nodules. Instead of using the pooling layer, we adopt the 3D dilated convolution to keep the resolution unchanged and retain more detailed feature information. Meanwhile, to learn more multi-scale features from the input data, we fuse features output from dilated convolutions with different sizes of receptive fields to further enhance the classification performance of the model.

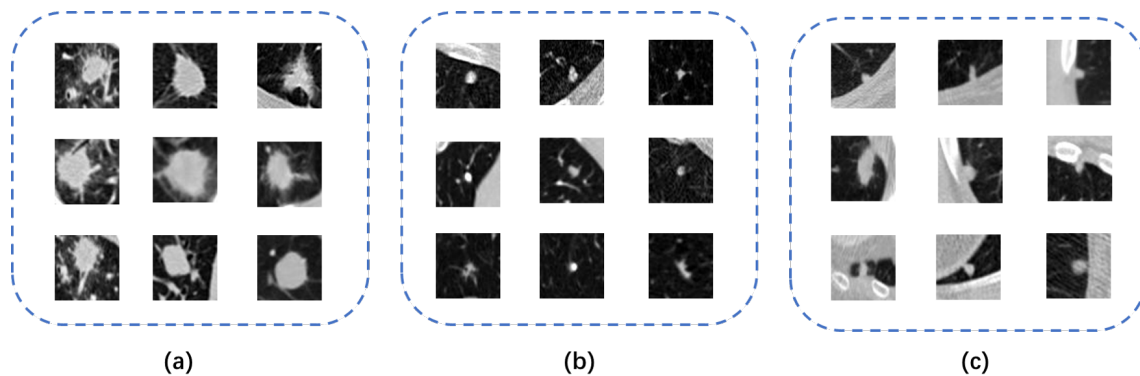


Figure 2. Samples of different pulmonary nodules for which diverse features are needed for the further diagnosis: (a) big diameter nodules; (b) small diameter nodules; and (c) nodules with adhesion.

The main contributions of this paper can be summarized as follows:

1. We propose a novel 3D spatial pyramid dilated network for the pulmonary nodule malignancy classification task. Compared with the 2D deep learning models, our model can capture more spatial and distinguishable information from the input data.
2. Unlike the previous work by using the pooling layer, we exploit the dilated convolution to alleviate the loss of tiny information and feature maps resolution.
3. A spatial pyramid dilated structure with multiple dilated rate is designed to learn the discriminative scale features from the nodule CT images. Extensive experiments show that our model has achieved a better result compared with other state-of-the-art methods.

2. Related Work

To detect the lung cancer at the early stage, many approaches have been tried to analyze the CT nodule images. One of the most frequently used feature extraction methods is the hand-crafted feature extraction method. Hand-crafted features such as texture, shape and intensity features are often used as the input into the classifier to classify the nodules. Krewer et al. [11] used 219 shape and texture features combined with the feature selection method to classify nodules on Lung Image Database Consortium (LIDC) dataset and the model achieved an accuracy of 90.91%. Uchiyama et al. [12] used grey-scale histogram feature to classify the pulmonary nodule; the model achieved a satisfying result at that time. Messay et al. [13] proposed a novel method to detect the nodules by combining intensity thresholding with morphological features; the best experimental result of the area under the curve (AUC) score was 0.89. Orozco et al. [14] proposed a supervised extraction way to select 11 features. Then, they combined in pairs as inputs to support vector machine (SVM). Finally, they obtained an accuracy of 82% with the sensitivity of 90.90%. Erdal et al. [15] used a novel method based on shape and texture features to classify nodules from CT images and the experimental result show that the AUC value increased from 96.39% to 96.79%. Han et al. [16] adopted three-dimensional texture features for pulmonary nodule classification; the model result demonstrated that the classification accuracy could achieve better results than the two-dimensional texture features classification methods. Although hand-crafted feature methods are

simple and fast, the parameters of each model need extra manual setting which could be empirical and subjective. Recently, inspired by the large number of image data and the good performance of parallelization computers, the convolution neural network has achieved remarkable success in computer vision field, as well as in medical image analysis [17–19]. Shi et al. [17] designed the Multimodal Stacked Deep Polynomial Networks to diagnose the Alzheimer’s Disease (AD). In experiments on brain image data, the model achieved promising results in classification accuracy. Kashif et al. [20] combined the hand-crafted features with the CNN learned features to detect the tumor nuclei. The combined features could contain more non-linear invariant texture features and achieve better detection results. Suk et al. [21] used the Deep Boltzmann Machine (DBM) for Alzheimer’s Disease and Mild Cognitive Impairment (MCI) classification on brain Magnetic Resonance Imaging (MRI) data. The model achieved a classification accuracy of 95.52% task, which outperformed other state-of-the-art methods. Zhu et al. [22] extracted the high-level feature representation by using the convolution neural network for prostate cancer detection; the experimental results showed that the proposed method achieved an averaged sensitivity of 91.51% and specificity of 88.47%. For the pulmonary nodule classification, the convolution neural network has also shown its effectiveness in malignant suspicious classification task. Kumar et al. [23] adopted an auto-encoder network to extract the features from nodule images, a binary decision tree was then used to classify the nodules as malignant and benign; it achieved a satisfactory performance at that time. Li et al. [24] applied a deep CNN for solid, semisolid and ground glass opacity nodules classification. Experiments were carried out on the 40,772 nodules and 21,720 non-nodules from the LIDC dataset, and the final result showed that the proposed method in terms of sensitivity and accuracy outperformed the competing methods. Chen et al. [25] used a multi-task learning architecture to find the relationship between different attributes; the experimental result showed that it outperformed other relevant studies on pulmonary nodule classification. Shen et al. [26] proposed a multi-crop 3D CNN architecture to automatically classify the malignant and benign pulmonary nodules from the raw CT images. Setio et al. [27] proposed a 2D multi-view convolution neural network to automatically learn the discriminative features from the training data. It achieved detection sensitivities of 85.4% and 90.1% at one and four false positive per scan, respectively. Yan et al. [28] designed a successive 3D convolution and max-pooling layers network architecture. The designed architecture could learn the spatial information and high-level representations from 3D nodule data; the experimental result showed that their method gained 87.4% accuracy on nodule malignancy classification. Jiang et al. [29] used multi-group 2D patches from the nodule volumes and fed them to a four-channel CNNs to classify the nodules, and achieved a low false positive rate. An overview of the representational related work is illustrated in Table 1.

Table 1. An overview of the representational related work.

Approach	Author	Year	Method
Hand-crafted feature	Uchiyama et al. [12]	2003	grey-scale histogram features
	Messay et al. [13]	2010	combined intensity thresholding with morphological processing to detect nodules
	Krewer et al. [11]	2013	texture features
	Orozco et al. [14]	2015	wavelet features and support vector machine
	Han et al. [16]	2015	three-dimensional texture features
Deep convolution feature	Li W et al. [24]	2016	2D CNN for solid, semisolid and ground glass opacity nodules classification
	Setio et al. [27]	2016	multi-view feature extraction for nodule classification
	Chen et al. [8]	2017	multiple semantic features for nodules classification by 2D CNN
	Shen et al. [26]	2017	multi-crop 3D CNN for nodule malignancy suspicious classification
	Yan et al. [28]	2017	compared the classification performance of 2D CNN with 3D CNN

3. Methodology

In this paper, we present a novel 3D spatial neural network for the malignancy assessment of pulmonary nodules. Different from the traditional convolution neural network architectures, which consist of successive convolution and pooling layer, we replace the pooling layer with 3D dilated convolution. The main architecture of the proposed model is illustrated in Figure 3. During the training process, the input image data are a $64 \times 64 \times 64$ nodule cube that first passes through a 3D convolution. After the first convolution operation, the center feature maps are cropped to extract more nodule-centric representations. As for capturing the discriminative scale and spatial acuity features, we designed a pyramid dilated structure with the various dilated rates. With the designed units, the tiny detailed information and more spatial representations could be preserved to leverage the model to achieve better performance. The extracted features from each dilated convolution are then fused to propagate to the subsequent layers.

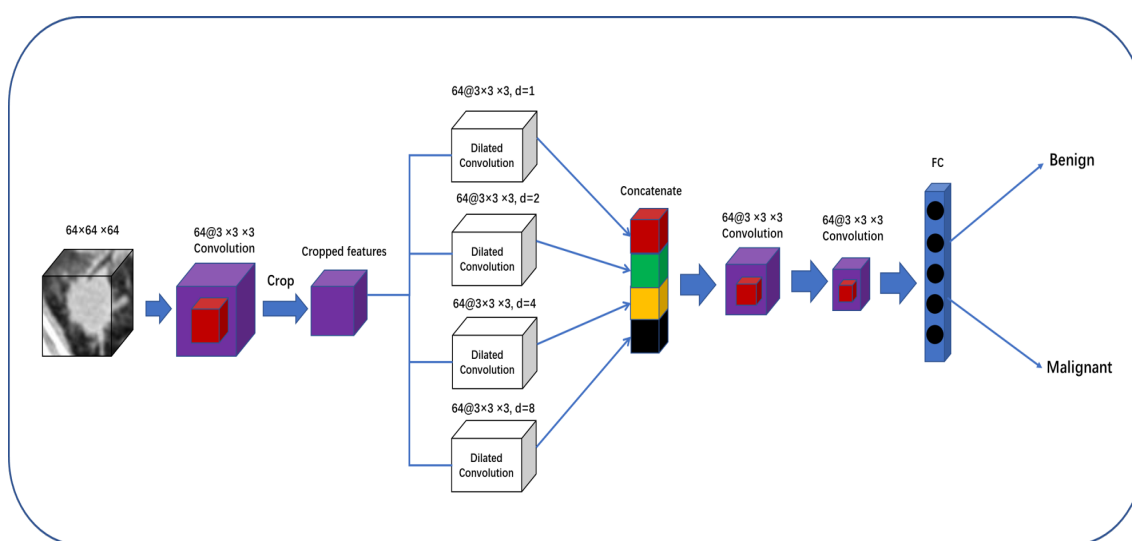


Figure 3. An overview of the proposed architecture for malignant and benign nodule classification. The input image size is $64 \times 64 \times 64$, the $(k@m \times m \times m, s)$ denotes the convolution has k kernels and the convolution size is $m \times m \times m$. In the dilated convolution, d is the dilated rate, and FC represents the fully connected layer.

3.1. 3D Spatial Convolution Layer

In 2D CNNs, the convolutions are only used on the 2D feature maps to gain features from the previous layers. However, pulmonary nodules usually have different shapes, sizes and spatial characteristic. For the purpose of computing the image features spatially and temporally, 3D convolution [30] is performed to extract the features from the 3D input cube image. The detailed comparison of 3D convolution operation and 2D operation is illustrated in Figure 4. By convolving the 3D kernel to the cube of stacking contiguous CT slices, the 3D convolution can learn more global and spatial information. During the convolution operation, the 3D kernel weights are duplicated across the cube.

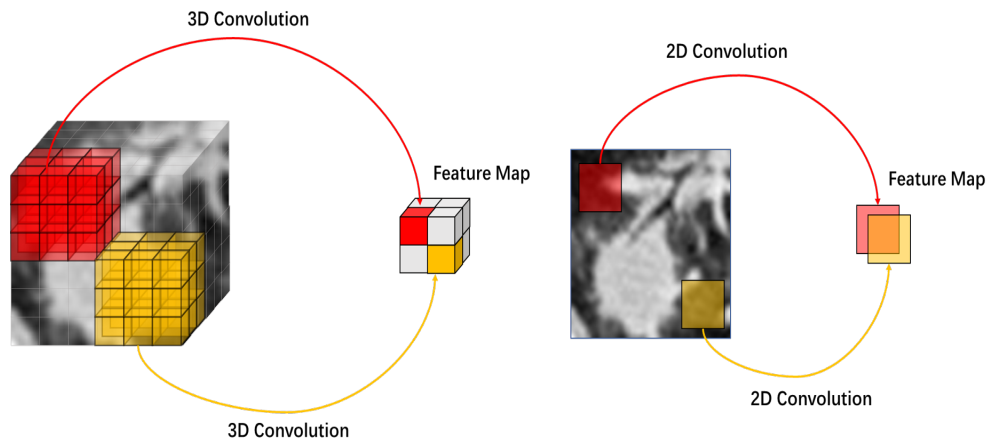


Figure 4. The comparison of 3D convolution operation and 2D convolution operation.

Let us define (P_i, Q_i, R_i) as the size of the 3D kernel. w_{ijm}^{pqr} is the (p, q, r) th value of the kernel which connect to the m th feature map in the previous layer and b_{ij} is a bias. v_{ij}^{xyz} is the value at position (x, y, z) in the i th layer of the j th feature map. Then, the output of 3D convolution at position (x, y, z) on the j th feature map in the i th layer can be defined as follows:

$$v_{ij}^{xyz} = f(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}) \quad (1)$$

3.2. Pyramid Dilated Structure

Before inputting the features into the pyramid dilated convolution structure, first the center features are cropped from the previous layer. The motivation for cropping the center features is based on the fact that the size of nodules usually vary hugely. Therefore, the cropped features can effectively learn the overall characteristic information of small diameter nodules and the salient centric features of big diameter nodules.

The loss of spatial information by using the consecutive pooling layer could be a handicap for improving the classification performance. Especially, if the tiny and spatial information is decayed or even lost in the previous layers, there will be little hope for the network to recover them in the subsequent layers. A simple way to preserve those features is to remove the pooling layers of the network. However, with the disappearance of pooling layers, the receptive field of each filter will also be limited. To handle this challenge, 3D dilated convolution [31] with different dilated rates is used to increase the receptive field of filters exponentially. It also has the benefit of preserving the spatial and tiny information of the nodule. f is a discrete function defined as: $Z^3 \rightarrow R$. Let g be a 3D discrete filter, the convolution operator $*$ is defined as

$$(f * g)(n) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(n - \tau). \quad (2)$$

In the dilated convolution, adding a dilated rate l to the convolution operator $*$ and the dilated convolution operator $*_l$ can be defined as

$$(f *_l g)(n) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(n - l\tau). \quad (3)$$

Consider $f_0, f_1, \dots, f_{n-1} : Z^3 \rightarrow R$ as discrete functions, and g_0, g_1, \dots, g_{n-2} as discrete $\{3 \times 3 \times 3\}$ filters, the dilated factor $l = 2^i$ and the f_{i+1} can be expressed as

$$f_{i+1} = f_i *_{2^i} g_i, \quad i = 0, 1, \dots, n-2. \quad (4)$$

Thus, the size of each receptive field can be $(2^{i+2} - 1) \times (2^{i+2} - 1) \times (2^{i+2} - 1)$. By using the dilated convolution, the receptive field expands exponentially; it also allows the network to efficiently extract the tiny and spatial information with high feature map resolution. If we consider the two-dimensional convolution operation, the receptive field of different dilated convolution rates is illustrated in Figure 5.

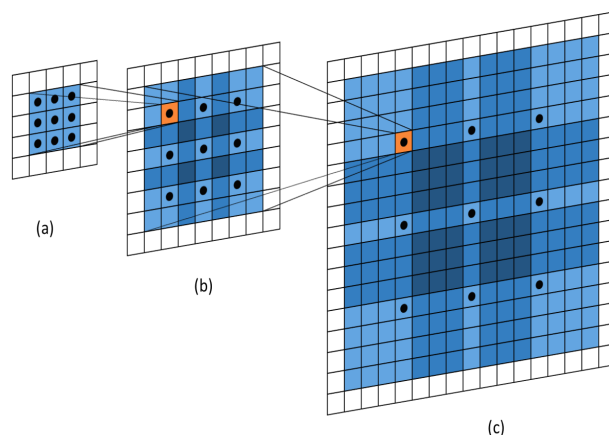


Figure 5. The exponential expansion of the receptive field by the dilated convolution: (a) the first dilated convolution with dilated rate 1 and the receptive field of the filter is 3×3 ; (b) the second dilated convolution with dilated rate 2 and the receptive field of the filter is 7×7 ; and (c) the third dilated convolution with dilated rate 4 and the receptive field of the filter is 15×15 .

3.3. Multiple Receptive Field Feature Fusion

The receptive field of the network plays an important role in image classification task, different sizes of receptive fields usually have multiple scale features. If the size of the receptive field is small, there will be more local characteristic information for the network to train. On the contrary, if the size of the receptive field is big, more global and abstract feature information will be captured during the training. For the purpose of learning more discriminative scale features from various sizes of nodules, the pyramid dilated convolution structure is applied with different dilated rates to achieve different receptive fields of the filters. In this structure, the dilated rate is set as $\{1, 2, 4, 8\}$, and the corresponding size of each receptive field is $\{3 \times 3 \times 3, 7 \times 7 \times 7, 15 \times 15 \times 15, 31 \times 31 \times 31\}$. Figure 6 shows the feature maps of different dilated convolutions. The four different sizes of the receptive fields can provide more local and global representations, which can be crucial in classifying the various sizes of nodules. On this condition, the four feature maps are concatenated together to enrich the discriminative scale features of the model. After fusing the multi-scale feature maps, we feed them into two convolutions to reduce the dimensions and finally classify the nodule malignancy by a softmax function.

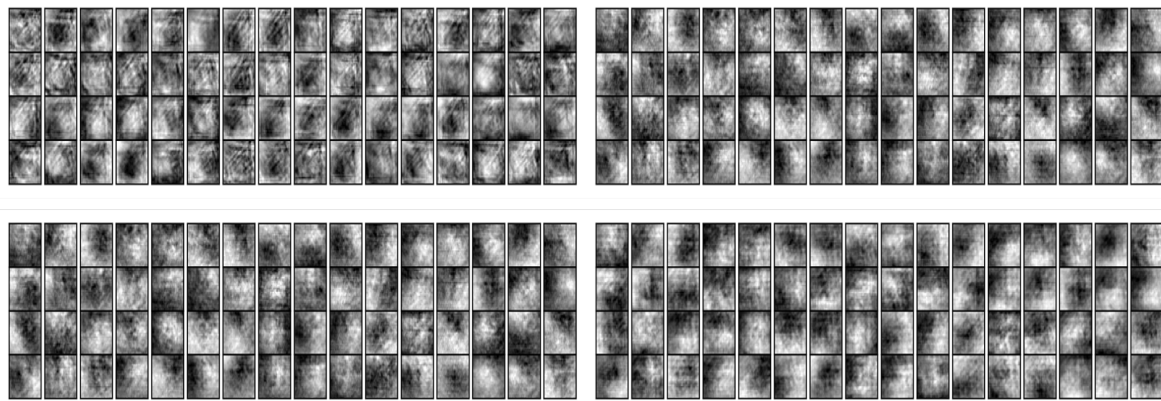


Figure 6. The feature maps of different dilated convolutions. Top left: The feature maps with dilated rate of 1; Top right: The feature maps with the dilated rate of 2; Bottom left: The feature maps with the dilated rate of 4; Bottom right: The feature maps with the dilated rate of 8.

3.4. Fully Connected Layer

During the traditional convolution operations, the neural network adopts a partial activation rule to prevent the training being over-fitting. Unlike the convolution operation, the neurons of the fully connected layer are all activated. By adopting the fully connected layer, the network has the ability to learn more high-level and low-level features from multi-layers. The output features are flattened from the previous layer to a vector, and then made a matrix vector multiplication as:

$$h^f = \sigma(b^f + W^f h^{f-1}), \quad (5)$$

where $h^{(f-1)}$ denotes the feature vector of f-1 layer, W^f is the weight of the matrix, and b^f is the bias of the f th layer. Meanwhile, σ is the Relu activation function.

3.5. Softmax Layer

Softmax activation function is a generalization of the logistic function, which is used in the multiple classification tasks. It maps a K-dimensional output vector to a K-dimensional mapping vector where each value of the mapping vector is in the range (0,1), and all values of the mapping vector add up to 1. Each value of the mapping vector can be regarded as the probability of each class. The highest probability is just the class which the input belongs to. Supposing that the input matrix is a vector z with k dimensions, and the softmax function is actually a normalized index function, which can be formulated as follows

$$y_c = \frac{e^{z_c}}{\sum_{d=1}^k e^{z_d}} \quad \text{for } c \in [1, k], \quad (6)$$

the output of the softmax function can be a vector y with k dimensions, and each value of the vector is between 0 and 1, and y_c is the output probability of the c th class.

3.6. Training Details

The model was validated based on five-fold cross validation. We randomly split the whole data into five same size groups. Four of them were used to fine tune the network, while the last group was taken as the validation set to validate the performance of the model. We adopted cross-entropy as the loss function to optimize the parameters; the initial learning rate was 1.0×10^{-3} . We used the data augmentation method to avoid the training being over-fitting. The training process was stopped when the val-loss was not improved over five epochs. The network structure was designed by Tensorflow in the programming language Python.

3.7. Evaluation Metrics

To efficiently evaluate the performance of our model, we relied on four evaluation metrics to quantify the validation performance. True positives (TP) can be defined as the samples that are correctly classified as positive. Analogously, false positives (FP) can be defined as samples that are incorrectly classified as positive. True negatives (TN) and false negatives (FN) are denoted similarly.

1. The classification accuracy rate is the ratio of the number of samples correctly sorted by the classifier to the total number of samples when the probabilistic output of the classifier is threshold at t .

$$Accuracy(t) = \frac{TP(t) + TN(t)}{TP(t) + FP(t) + TN(t) + FN(t)} \quad (7)$$

2. Sensitivity or the true positive rate (TPR) is an empirical value used to measure the percentage of actual positives which are correctly identified. Sensitivity or TPR is defined as the function of threshold t , with the expressions below

$$TPR(t) = \frac{TP(t)}{TP(t) + FN(t)} \quad (8)$$

3. The false positive rate (FPR) which is defined as

$$FPR(t) = \frac{FP(t)}{FP(t) + TN(t)} \quad (9)$$

The FPR is an empirical value which is used to measure the percentage of actual negatives which are correctly classified.

4. Finally, we draw the receiver operating characteristic (ROC) curve which shows the fit degree of ground-truth label and classifier-predicted label. The AUC score is the area under the ROC curve. It is an empirical value representing the probabilistic output of the classifier which is greater for a positive example than for a negative example.

4. Experiment

4.1. Data Description

We evaluated our model on the opening LIDC dataset [32]. The total number of the CT images is 1010. The specific and detailed attributes (coordinate, diameter, and malignancy) of each pulmonary nodule was annotated by four professional radiologists. The malignancy score has 5 levels. A nodule malignancy score below 3 is regarded as benign, and the nodule malignancy score above 3 is regarded as malignant. We did not consider the malignancy score of 3. Due to the diversities of different CT images, we first resampled the spacing of each CT image to $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ by using the spline interpolation. We center-cropped the nodule voxels from the CT images and each malignancy class of the nodule is based on the voting results from the four radiologists. There are 353 total training and testing data. Due to the limited data, we adopted data augmentation methods to augment our data. The augment methods were mainly random flipping, translation and rotation. The ranges of the rotation were from 40° to 120° . We also adopted the zoom in and zoom out methods to get different scales of input image. By augmenting the existing data samples, the model could better accommodate the geometric variations and transformations [33]. The sample data of the augmentation operation are illustrated in Figure 7. Finally, there were 781 total the nodule images, with 397 malignant nodules and 384 benign nodules.

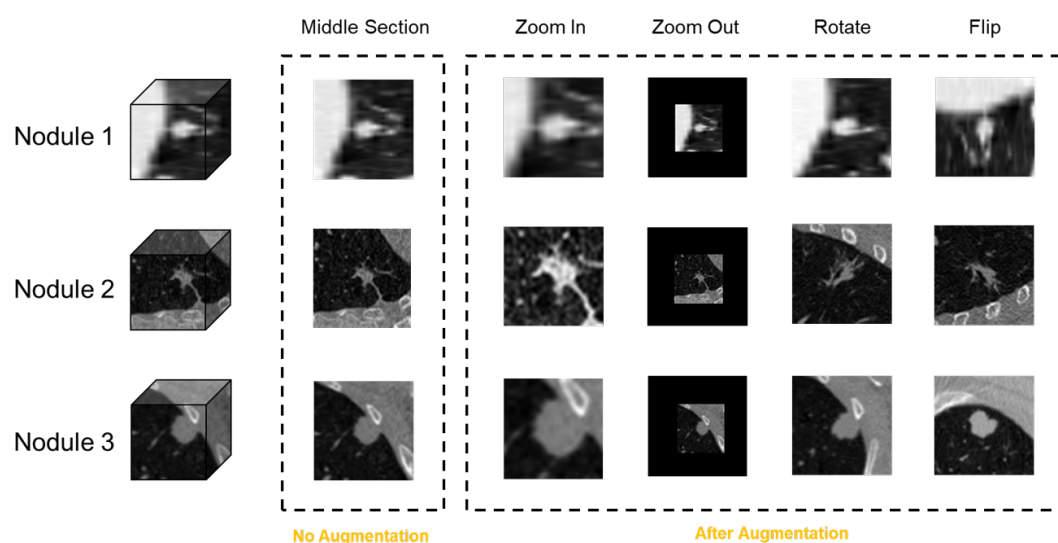


Figure 7. The sample data of augmentation operation.

4.2. The Influence of Different Sample Sizes

The sample size of the data is crucial in fine-tuning the model. Too small of a dataset may cause the network being over-fitted, while adding more data adds more time to training the network. Thus, we first explored the influence of the data sample size on the model performance. We divided the data into three groups to conduct the comparison experiment. Their sizes were 177, 353 and 706, i.e., half, equal, and double the size of the original dataset. The comparison result is shown in Table 2. The result demonstrates that adding more data could improve the performance of the model smoothly. In the following experiments, we adopted the image data size of 706 as the experimental data.

Table 2. Comparison with various data size.

Datasize	177	353	706
Accuracy (%)	81.2	85.5	88.6
Sensitivity (%)	82.2	84.2	86.3
Specificity (%)	75.1	80.2	90.3
AUC	0.721	0.832	0.883

4.3. Compared with State-of-the-Art Method

We evaluated the performance of our model by comparing three basic state-of-the-art methods (2D CNN, 2D Resnet and 3D CNN). The 2D CNN has the same structure as the previous work [34], and it has achieved high classification result in their work. The Resnet [35] structure also has shown its effectiveness in image feature learning. Nevertheless, the limitation of the 2D images as the input to the network is that it could lose the spatial and global characteristic information of the nodules. Those characteristic information could play an important role in identifying the malignant and benign nodules. The third basic method we compared is based on the hierarchical network [36] which consists of a series of 3D convolution and pooling layers. In contrast with the 2D CNN methods, the 3D CNN method is capable of capturing the spatial and internal characteristic information from the nodules. The ROC comparison result is illustrated in Figure 8. Our model achieves 0.883 AUC score, which represents better performance than the other three methods.

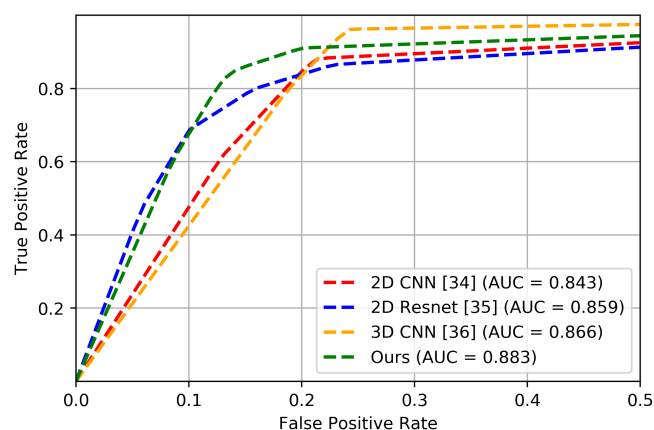


Figure 8. Receiver operating characteristic (ROC) comparison with two basic methods.

Other comparison metrics are evaluated, and the detailed result is shown in Figure 9. From the experimental result, we can see that our model has achieved an accuracy of 88.6% which is the best among the methods. Furthermore, the sensitivity score of our model is 86.3%, which is the best performance compared with the other models, and the sensitivity result is with great value in the clinical diagnosis. Meanwhile, we found that the overall performance of 3D network is better than the 2D network. This could be explained as 3D convolution can extract more spatial and global features from the nodule cube than 2D convolution (Figure 4).

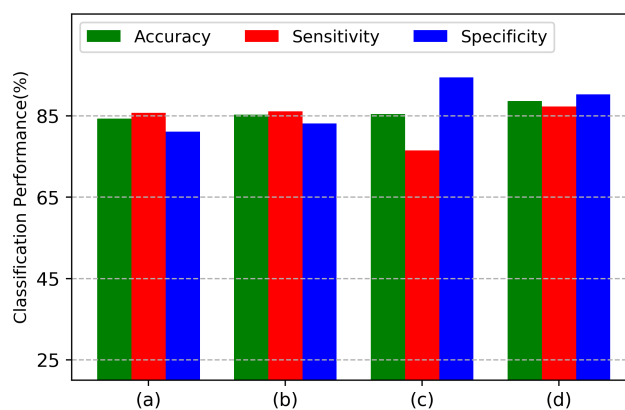


Figure 9. Compared accuracy, sensitivity, and specificity with different methods: (a) 2D CNN; (b) 2D Resnet; (c) 3D CNN; and (d) our proposed model.

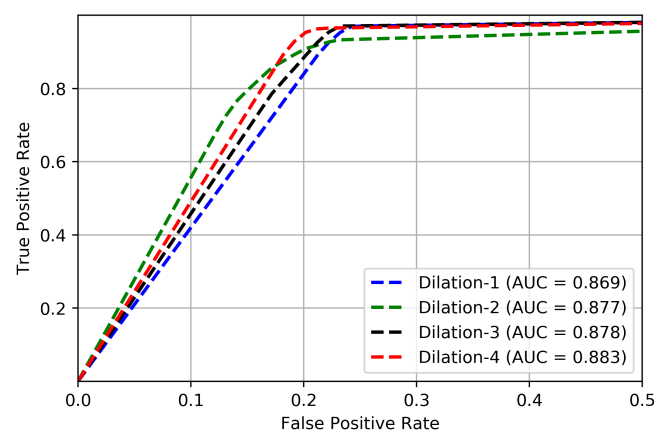
We further evaluate our model with other methods on the same opening LIDC dataset. The comparison result is illustrated in Table 3 and the experimental result shows that our model has achieved a competitive performance on that dataset.

Table 3. Performance comparison with different nodule classification methods on the LIDC dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Han et al. [16]	-	89.4	86.0	0.941
Kumar et al. [23]	75.0	83.3	-	-
Chen et al. [25]	86.8	60.3	95.4	-
Shen et al. [26]	87.1	77.0	93.0	0.930
Messay et al. [37]	75.0	83.3	-	-
Proposed Model	88.6	86.3	90.3	0.883

4.4. The Effectiveness of Dilated Convolution Setting

The receptive field plays an important role in object classification task, and different sizes of receptive fields could capture various local and global features from the images. In our model, the dilated convolution setting allows the network to increase the receptive field exponentially without losing the feature maps resolution, which could greatly help the network to better learn the small object features. In this case, we explore the influence of different numbers of dilated convolutions on the performance of the model. The experimental number of dilated convolutions is 1, 2, 3 and 4, respectively. Given different numbers of dilated convolutions, the model could capture various receptive fields of features. The detailed comparison result is illustrated in Figure 10. From the result, we see that the best AUC score is achieved by the Dilation-4 structure which has four different dilated convolutions with various dilated rates. It achieved an AUC score of 0.883, which proves that the receptive field plays an important role in classifying the nodules.

**Figure 10.** Receiver operating characteristic (ROC) comparison with different numbers of dilated convolutions.

The evaluation criteria of accuracy, sensitivity, and specificity were also compared. The detailed result is shown in Figure 11. The four dilated convolutions setting has achieved competitive performance, especially in the sensitivity score, which could be explained as the diverse receptive fields of the network can efficiently improve the ability to capture scale relevant features of the nodules.

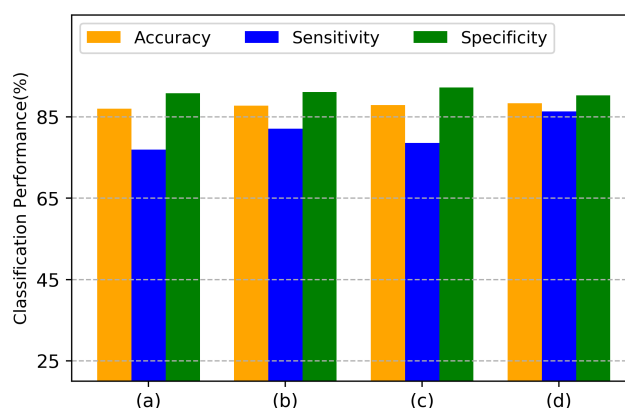


Figure 11. Compared accuracy, sensitivity, specificity with different numbers of dilated convolutions: (a) Dilation-1; (b) Dilation-2; (c) Dilation-3; and (d) Dilation-4.

4.5. Comparison with Different Feature Cropping and Fusion Modes

In this section, we try to explore the effectiveness of feature cropping and fusing operation. The detailed comparison result of feature cropping is illustrated in Table 4. The result demonstrates that the classification accuracy of cropping features mode is better than the no cropping features mode. It could be because cropping the central features from the first convolution layer helps the network to further capture the centric feature information of the pulmonary nodules, and also enriches the feature diversities of the model. After cropping features process, we used four different dilated convolutions with different receptive fields for learning multi-scale features of nodules. The features from the four dilated convolutions are then fused into the following layers. A comparison of different fusion methods is also given in Table 3. Three fusion methods (maximum, average, and concatenate) were attempted to explore the influence of that configuration. We found that the concatenate mode works best, which achieved the classification accuracy score of 88.6%.

Table 4. Evaluation on different fusion modes.

Method	Accuracy (%)
Crop Features + Dilated Fusion (Average)	87.0
Crop Features + Dilated Fusion (Maximum)	88.2
Crop Features + Dilated Fusion (Concatenate)	88.6
No Crop Features + Dilated Fusion (Average)	83.2
No Crop Features + Dilated Fusion (Maximum)	84.1
No Crop Features + Dilated Fusion (Concatenate)	84.5

5. Conclusions

In this paper, we propose a novel 3D spatial pyramid dilated network to classify the malignant and benign pulmonary nodules. Since the traditional pooling layer can reduce the resolution of feature maps, thus, the tiny feature information will be missing or decayed. Instead of using the pooling layer, we utilize the 3D dilated convolution to learn the nodule feature without losing the resolution. Moreover, with the diverse setting of dilated rates, the network could capture multi-scale features from different receptive fields. Extensive experiments on the opening dataset LIDC have demonstrated that our model outperforms the state-of-the-art methods. Future work will be focused on exploring the relationship between other nodule attributes and the malignancy rating.

Author Contributions: G.Z. conceived and designed the experiments; G.Z. performed the experiments; G.Z., P.H., L.L. and X.L. analyzed the data; Y.L., D.Z. and J.L. contributed reagents/materials/analysis tools; and G.Z. wrote the paper.

Funding: This work was supported by the General Program of National Natural Science Foundation of China (NSFC) under Grant No. 61572362. This research was also partially supported by the General Program of National Natural Science Foundation of China (NSFC) under Grant No. 81571347. Fundamental Research Funds for the Central Universities under Grant No. 22120180012.

Acknowledgments: The authors are grateful for the comments and reviews from the reviewers and editors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics. *CA Cancer J. Clin.* **2016**, *66*, 7–30. [[CrossRef](#)] [[PubMed](#)]
2. Stewart, B.; Wild, C.P. *World Cancer Report 2014*; International Agency for Research on Cancer: Lyon, France, 2014.
3. Henschke, C.I.; McCauley, D.I.; Yankelevitz, D.F.; Naidich, D.P.; McGuinness, G.; Miettinen, O.S.; Libby, D.M.; Pasmantier, M.W.; Koizumi, J.; Altorki, N.K.; et al. Early Lung Cancer Action Project: Overall design and findings from baseline screening. *Lancet* **1999**, *354*, 99–105. [[CrossRef](#)]
4. Farag, A.; Ali, A.; Graham, J.; Farag, A.; Elshazly, S.; Falket, R. Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest. In Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, 30 March–2 April 2011; pp. 169–172.
5. Song, Y.; Cai, W.; Zhou, Y.; Feng, D.D. Feature-based image patch approximation for lung tissue classification. *IEEE Trans. Med. Imaging* **2013**, *32*, 797–808. [[CrossRef](#)] [[PubMed](#)]
6. Sorensen, L.; Shaker, S.B.; De Bruijne, M. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Trans. Med. Imaging* **2010**, *29*, 559–569. [[CrossRef](#)] [[PubMed](#)]
7. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [[CrossRef](#)] [[PubMed](#)]
8. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
9. Yu, F.; Koltun, V.; Funkhouser, T. Dilated residual networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; doi:10.1109/CVPR.2017.75. [[CrossRef](#)]
10. Dou, Q.; Chen, H.; Yu, L.; Qin, J.; Heng, P.A. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1558–1567. [[CrossRef](#)] [[PubMed](#)]
11. Krewer, H.; Geiger, B.; Hall, L.O.; Goldgof, D.B.; Gu, Y.; Tockman, M.; Gillies, R.J. Effect of texture features in computer aided diagnosis of pulmonary nodules in low-dose computed tomography. In Proceedings of the 2013 IEEE International Conference on Systems, Man and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 3887–3891.
12. Uchiyama, Y.; Katsuragawa, S.; Abe, H.; Shiraishi, J.; Li, F.; Li, Q.; Zhang, C.T.; Suzuki, K.; Doi, K. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. *Med. Phys.* **2003**, *30*, 2440–2454. [[CrossRef](#)] [[PubMed](#)]
13. Messay, T.; Hardie, R.C.; Rogers, S.K. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med. Image Anal.* **2010**, *14*, 390–406. [[CrossRef](#)] [[PubMed](#)]
14. Orozco, H.M.; Villegas, O.O.V.; Sánchez, V.G.C.; Dominguez, H.D.J.O.; Alfaro, M.D.J.N. Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine. *Biomed. Eng. Online* **2015**, *14*, 9. [[CrossRef](#)] [[PubMed](#)]
15. Erdal, U.A. Shape and texture based novel features for automated juxta-pleural nodule detection in lung cts. *J. Med. Syst.* **2015**, *39*, 46.
16. Han, F.; Wang, H.; Zhang, G.; Han, H.; Song, B.; Li, L.; Moore, W.; Lu, H.; Zhao, H.; Liang, Z. Texture feature analysis for computer-aided diagnosis on pulmonary nodules. *J. Digi. Imaging* **2015**, *28*, 99–115. [[CrossRef](#)] [[PubMed](#)]
17. Shi, J.; Zheng, X.; Li, Y.; Zhang, Q.; Ying, S. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer’s disease. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 173–183. [[CrossRef](#)] [[PubMed](#)]

18. Maninis, K.K.; Pont-Tuset, J.; Arbeláez, P.; Gool, L.V. Deep retinal image understanding. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016; pp. 140–148.
19. Song, Y.; Zhang, L.; Chen, S.; Ni, D.; Lei, B.; Wang, T. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 2421–2433. [[CrossRef](#)] [[PubMed](#)]
20. Kashif, M.N.; Raza, S.E.A.; Sirinukunwattana, K.; Arif, M.; Rajpoot, N. Handcrafted features with convolutional neural networks for detection of tumor cells in histology images. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 1029–1032.
21. Suk, H.I.; Lee, S.W.; Shen, D.; Alzheimers Disease Neuroimaging Initiative. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* **2014**, *101*, 569–582. [[CrossRef](#)] [[PubMed](#)]
22. Zhu, Y.; Wang, L.; Liu, M.; Qian, C.; Yousuf, A.; Oto, A.; Shen, D. MRI-based prostate cancer detection with high-level representation and hierarchical classification. *Med. Phys.* **2017**, *44*, 1028–1039. [[CrossRef](#)] [[PubMed](#)]
23. Kumar, D.; Wong, A.; Clausi, D.A. Lung nodule classification using deep features in CT images. In Proceedings of the 12th Conference on Computer and Robot Vision, Halifax, NS, Canada, 3–5 June 2015; pp. 133–138.
24. Li, W.; Cao, P.; Zhao, D.; Wang, J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Comput. Math. Methods Med.* **2016**. [[CrossRef](#)] [[PubMed](#)]
25. Chen, S.; Qin, J.; Ji, X.; Lei, B.; Wang, T.F.; Ni, D.; Cheng, J.Z. Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images. *IEEE Trans. Med. Imaging* **2017**, *36*, 802–814. [[CrossRef](#)] [[PubMed](#)]
26. Shen, W.; Zhou, M.; Yang, F.; Yu, D.; Ding, D.; Yang, C.; Zhang, Y.; Tian, J. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognit.* **2017**, *61*, 663–673. [[CrossRef](#)]
27. Setio, A.A.A.; Ciompi, F.; Litjens, G.; Gerke, P.; Jacobs, C.; Riel, S.J.V. Wille, M.M.W.; Naqibullah, M.; Sánchez, C.I.; Ginneken, B.V. Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* **2016**, *35*, 1160–1169. [[CrossRef](#)] [[PubMed](#)]
28. Yan, X.; Pang, J.; Qi, H.; Zhu, Y. Bai, C.; Geng, X.; Liu, M.; Terzopoulos, D.; Ding, X. Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2D and 3D strategies. In *Computer Vision—ACCV 2016 Workshops*; Springer: Switzerland, Sweden, 2017.
29. Jiang, H.; Ma, H.; Qian, W.; Gao, M.; Li, Y. An automatic detection system of lung nodule based on multi-group patch-based deep learning network. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 1227–1237. [[CrossRef](#)] [[PubMed](#)]
30. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [[CrossRef](#)] [[PubMed](#)]
31. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. In Proceedings of the International Conference on Learning Representations 2016, San Juan, PR, USA, 2–4 May 2016.
32. Armato, S.G., III; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B.; Henschke, C.I.; Hoffman, E.A.; Kazerooni, E.A.; et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med. Phys.* **2011**, *38*, 915–931. [[CrossRef](#)] [[PubMed](#)]
33. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 764–773.
34. Tajbakhsh, N.; Suzuki, K. Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs. *Pattern Recognit.* **2017**, *63*, 476–486. [[CrossRef](#)]
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

36. Shen, W.; Zhou, M.; Yang, F.; Yang, C.; Tian, J. Multi-scale convolutional neural networks for lung nodule classification. In Proceedings of the 24th International Conference on Information Processing in Medical Imaging (IPMI 2015), Isle of Skye, UK, 28 June–3 July 2015; pp. 588–599.
37. Messay, T.; Hardie, R.C.; Tuinstra, T.R. Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset. *Med. Image Anal.* **2015**, *22*, 48–62. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).