Application of Modified Inception-ResNet and CondenseNet in Lung Nodule Classification

Jingzhi Fu

School of Computer Science and Technology University of Science and Technology of China
Suzhou, China

Abstract. Nowadays lung cancer is one of the most vital malignancies to human health. Restricted by current medical technology, the best solution to lung cancer is still early diagnosis and targeted treatment. Lung nodule is early clinical sign of lung cancer, and low-dose spiral computed tomography is widely considered to be the most effective approach for lung cancer early screening. Through increasing accuracy and stability of diagnosis, CAD can significantly improve quality and efficiency of medical image analysis, reduce the chance of wrong diagnosis caused by subjective factors and missed diagnosis. With rapid development of CNN in image processing, there emerge many kinds of CNN architectures which achieve outstanding performance in image classification. We select Inception-ResNet and CondenseNet as candidate networks due to their outstanding classification performance in ImageNet. Consider reliance among feature map channels and 3D nature of CT scans, 3D-SE-IRNet and 3D-SE-IRNet were designed to further improve accuracy of networks. Results of our experiment prove good performance of CNNs in lung nodule CT scans classification. What's more, introducing self-mechanism and 3D convolution can significantly improve network's accuracy.

Keywords: Inception-ResNet; CondenseNet; lung nodule classification; SENet; 3D convolution.

1. Introduction

Nowadays lung cancer is one of the most vital malignancies to human health. American Cancer Society's official periodical *CA* predict in [1] that there will arise 18.1 million new cases of cancer and 9.6 million cancer deaths worldwide in 2018, and among all kinds of malignancies, lung cancer is still the most commonly diagnosed one (11.6% of the total cases) and the leading cause of cancer death (18.4% of the total cancer deaths). Restricted by current medical technology, the best solution to lung cancer is still early diagnosis and targeted treatment, and clinical data also shows that success of treatment is highly correlated with stage of cancer in its diagnosis. The early detection and targeted treatment of lung cancer can effectively enhance survival rate of patients to 49%, so precautions and early diagnosis is one of the most important parts in current lung cancer controlling scheme. Lung nodule is early clinical sign of lung cancer, and among all kinds of diagnosis methods, low-dose spiral computed tomography is widely considered to be the most effective approach for lung cancer early screening.

Recent years, with rapid development of computer technology especially computer vision, Computer Aided Diagnosis (CAD) gets widely used in medical image analysis. Through increasing accuracy and stability of diagnosis and reducing time it needs, CAD can significantly improve quality and efficiency of medical image analysis, reduce the chance of wrong diagnosis caused by subjective factors and missed diagnosis brought by human eye oversight. It also lightens workload of radiologist considerably, so it is feasible diagnosis method for general examination and large-scale medical image analysis. In addition, imaging diagnosis is rightly the observation and analysis to massive medical images which excludes other diagnosis methods such as needle biopsy, which makes it suitable to use CAD as auxiliary screening approach. [2], [3] discuss feasibility and effectiveness of CAD and show that CAD can enhance performance for lung nodule screening.

Aim of this study is to design proper CNN architecture for lung nodule CT image classification based on existing CNN architectures. This network needs to capture features of lung nodules including the benign and the malignant, further realizing its suspiciousness classification. More specifically, our major contributions include:

1) This program proposes applying Inception-ResNet and CondenseNet into automatic lung nodule CT image suspiciousness classification to help radiologists screen malignant lung nodules from large-scale lung CT images.

2) We introduce self-attention mechanism in SENet into our CNN architecture to picture reliance among channels and adjust channels' relative feature strength adaptively through global loss function.

3) Considering the 3D nature of lung CT images, we apply 3D convolution kernels into our network to capture information hidden in reliance among sequent slices.

2. Related Work

2.1 Deep Convolutional Neural Network

Excellent performance of convolutional neural network (CNN) in image feature extraction makes it be mainstream algorithm in image classification tasks. Using CNN in image classification tasks can be traced back to LeNet [4] designed by LeCun for handwritten number recognition in 1998, and the framework proposed in this paper which contains convolutional, pooling and fully-connected layers is still the mainstream architecture of CNN. In 2011, Ciresan [5] propose the technical framework of deploying CNN in GPU, which provides hardware support for solving computing source problem of massive image analyzing. In 2012, Outstanding performance of AlexNet [6], champion of ILSVRC(ImageNet Large Scale Visual Recognition Competition), in this game(84% top-5 accuracy, which is way ahead of the second's 74%) validates CNN's powerful feature learning and expression ability and effectiveness of convolution operation in large dataset, so it brings turning point of computer vision research including targets classification algorithm. Research about application of CNN in computer vision starts blowout.

After AlexNet a series of creative CNN architectures are proposed: VGGNet [7] proposes replacing 5×5 and 7×7 convolutions with multiple 3×3 convolutions to reduce network parameters and strengthen network's ability of feature extracting and expressing through more nonlinear mapping. Compared with ZFNet [8] and VGGNet which enhance network's performance through increasing its depth and width, GoogLeNet [9] uses inception module based on the idea of NIN (Network in Network) [10]. Inception module applies integration of multiple convolution kernels which have different filter sizes rather than one single kernel to different visual patterns of different sizes. Inceptionv2 [11] replaces the 5×5 convolution of the inception module with two 3×3 convolutions and adopt Batch Normalization (BP) to solve the vanishing gradient problem in training. Inceptionv3 [12] substitutes one $n \times 1$ convolution and one $1 \times n$ convolution for the $n \times n$ convolutions in inception module to reduce number of parameters and lessen overfitting. What's more, this kind of asymmetric convolution splitting increases diversity of feature. Additionally, Inceptionv3 further improves structure of inception module.

Increasing network depth brings degradation problem where accuracy gets saturated and then degrades rapidly with network depth increasing, which makes it difficult to apply deep CNN. ResNet [13] adopts residual module which connects initial input with layer outputs directly which solves network degradation problem effectively. Simplicity and excellent performance make ResNet be the most popular CNN architecture in industry, and a lot of research into CNN architecture afterwards is based on ResNet. Inception-ResNet [14] introduces idea of ResNet into inception module. ResNeXt [15] uses the same topology structure in different branches of inception module to significantly reduce parameter number and adopts residual connections meanwhile. But this also kills inception module's feature of covering receptive fields in various scales. In DensNet [16], for each layer feature maps of all preceding layers are used as inputs, and its own feature maps are used as inputs into all subsequent layers. But training large-scale DenseNet needs too much graphics memory, to lessen connection redundancy problem of DenseNet, CondenseNet [17] proposes a novel module called learned group convolution where connections between layers for which this feature reuse is superfluous are moved. From the angle of polynomial, PolyNet [18] proposes more network architectures containing inception module and residual connections.

SENet [25] proposes squeeze and excitation operations to picture correlations between channels and it adjusts channels' relative feature strength adaptively through global loss function. Experiments show that this enhances accuracy of networks significantly while additional computational cost it brings is acceptable.

2.2 Lung Nodule CT Image Classification Algorithms

Currently research about applying CNN into lung nodule CT image classification is limited. A considerable part of study about lung nodule CT image classification [29] [30] [31] [32] is confined to traditional computer vision algorithms, which has a certain lag behind development of CNN in academia. Here are some where apply CNN into this study: to overcome difficulty brought by varying lung nodule size, [33] designs MCNN (Multi-scale Convolutional Neural Network) to process input lung nodule CT images from three scales and transmits them into three weight sharing CNNs. [34] introduces multi-crop pooling strategy to extract pattern of CT image in different scales. It is not one easy job to have an access to large lung nodule CT image dataset for training, so [35] proposes that we can train 3D CNN in nonprofessional medical image dataset and then transfer this CNN to professional dataset, which improves network's performance significantly. DeepLung [36] uses combination of dual path network [37] and gradient boosting machine to classify lung nodules. [38] adopts Multi-View-One-Network strategy in 3D CNN architecture for lung nodule classification, and realize 3D inception and 3D Inception-ResNet for this classification task.

3. Model and Framework

CNN possesses powerful ability in feature information extraction and classification of images, which benefits from outstanding performance of convolution operation in image pattern recognition. Besides, pixels in one image is more correlated with neighbors other than distant ones, and patterns of image aren't relied on their position. These two traits make room for designing deep CNN with sharing weights and local connectivity.
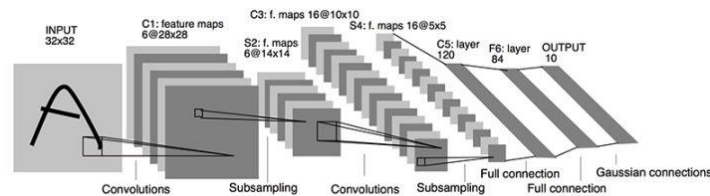

Figure 1: LeNet

Usually CNN is composed of convolutional, pooling and fully connected layer (Figure 1). Multiple convolution filters are used in convolutional layer for extract different features of the image, and through nonlinear transformation by activation function we obtain feature maps of this layer. Feature maps are transmitted to pooling layer for pooling operation where dimensions of feature map are properly reduced without too much feature information loss for lessen computational complexity. After several convolutional and pooling layers, feature maps enter fully connected layer for pattern integration from global perspective, further generating feature vector. At last class prediction are made based on the feature vector. Training of CNN adopts error backpropagation algorithm and uses loss function to measure network's accuracy. Following chain rule, we update network's parameters through gradient descent, further obtaining CNN for specific dataset.

3.1 Inception-ResNet

Inception (Figure 2, the left) adopts idea of NIN (Network in Network) and replaces the single convolutional filter with inception module containing multiple filters of different sizes to capture image features in different scales. It includes one pooling layer too. To decrease number of feature maps input after convolutional layer, 1×1 convolutions are placed before 3×3 and 5×5 convolutions for downsampling. Inceptionv2 (Figure 2, the middle) replaces 5×5 kernel in inception module with

two 3×3 kernels, which reduces parameters significantly and enhances nonlinear expression ability of network at the same time. V2 also introduces BP (Batch Normalization) into network to solve vanishing gradient problem in training and speed training of the whole network. Inceptionv3 replaces $n \times n$ filters in inception module with combination of one $n \times 1$ filter and one $1 \times n$ filter to save parameter number of network and lessen overfitting. What's more, this kind of asymmetric convolution splitting enhances diversity of pattern.
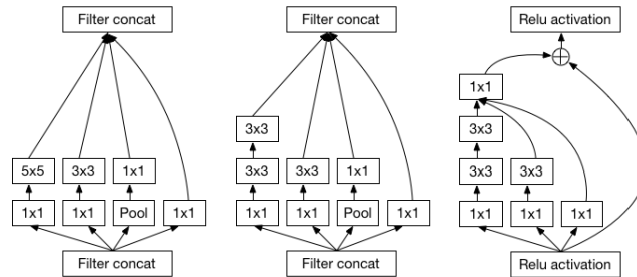

Figure 2: Inception series

Increasing of network depth brings blocked gradient backflow problem which restricts CNN's scale. Residual connection adopted by ResNet links initial input with outputs of all layers directly, guaranteeing backflow of gradient and further solving large-scale network's degradation problem effectively. Inception-ResNet (Figure 2, the right) introduces residual connection into inception module, which enhances network's pattern expression ability and accelerates its training dramatically at the same time.

3.2 CondenseNet

ResNet introduces shortcuts between initial inputs and convolutional layers into network to guarantee the input is transmitted into every layer without any information loss, which makes it possible to applying large-scale networks. Further for keeping information integrity of all layers' outputs (including initial input) when depth increases, in DenseNet
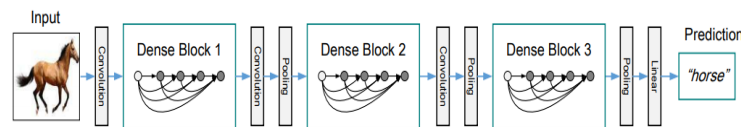

Figure 3: A DenseNet with three dense blocks

(Figure 3) for each layer feature maps of all preceding layers are used as inputs, and its own feature maps are used as inputs into all subsequent layers. This means adding L(L+1)/2 shortcuts in one L-layers network. Large-scale DenseNet needs too much graphics memory for training, so it adopts dense block piling for compromise.
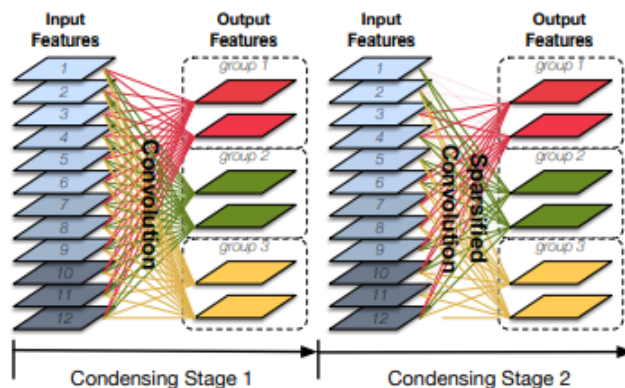

Figure 4: Learned group convolution

To solve connection redundancy problem of DenseNet, CondenseNet proposes trimming unnecessary weights when training, and applied learned group convolution when executing 1×1 convolutions (Figure 4). Besides, CondenseNet points out that feature reuse of layers near output layer is more important, so it adopts exponentially increasing growth rate. What's more, it adds fully connected layer between dense blocks (Figure 5) to realize better feature reuse.
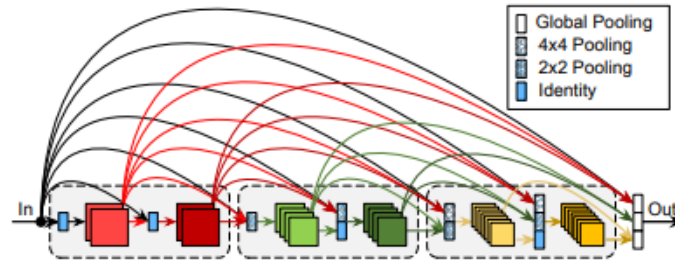

Figure 5: DenseNet variant

3.3 Self-attention Mechanism and 3D Nature of CT Images

For further improving classification performance of convolutional neural network, we introduce self-attention mechanism (Figure 6) in SENet: 1), Squeeze, execute global average pooling for every channel of outputted feature map and we obtain one feature vector about channels. 2), Excitation, learn feature weight of every channel through this feature vector and convert initial feature map into weighted feature map. This mechanism considers reliance among different channels and adjusts channels' relative feature strength adaptively, which further enhances performance of the network. Experiments show that compared to improvement of network performance, computational cost it brings is acceptable.
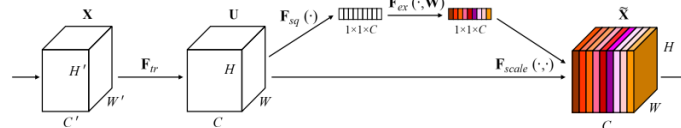

Figure 6: Self-attention mechanism

Besides CT images are natural 3D image sequence, and there are already relevant studies that show better classification performance can be obtained if we input lung nodule CT images as 3D cube. This is because 3D CT image contains reliance among consecutive slices, and we can obtain image features of lung nodule in a deeper level, further realizing better classification. This study will also use 3D CNN to increase accuracy of lung nodule classification.

4. Experiments

4.1 LIDC-IDRI Dataset

We evaluate our CNN architectures in LIDC-IDRI dataset. It is composed of 1018 cases, and for each case there is a series of corresponding thoracic CT scans. It also contains one XML file which records medical annotations provided by 4 experienced chest radiologists, including information about coordinates of lung nodules, edge and so on. We removed cases whose slice thickness > 3mm and where distance between slices varies. Then we obtain 888 target cases.

According the nodule collection report of LIDC-IDRI dataset, we obtain malignancy suspiciousness of each nodule rated from 1 to 5 by 4 experts, where degree of malignancy suspiciousness increases along with the suspiciousness score. We calculate average score of each lung nodule and remove nodules whose average rating are 3. We labelled nodule as benign one if its average score is lower than 3 and considered it as malignant nodule if its average rating is higher than

3. At last, we extracted 1186 lung nodules from this dataset for our experiment. Among them we see 650 benign nodules and 536 malignant ones.

4.2 Experiment Details

We cropped 64×64 image and 48×48×9 3D volume data of lung nodules from their CT scans based on the annotated center and divided them into two parts, where 75% of them are used for training validation and the rest is treated as testing set. To overcome problem of limited nodule samples, we augmented nodules by translation, rotation and flip operations. Such augmentation is also beneficial to extract features of lung nodule CT scans which don't change along with these three operations.

Besides traditional Inception-ResNet and CondenseNet, we introduce self-attention mechanism and 3D convolution into them and obtain our new networks: 3D-SE-IRNet and 3D-SE-CDNet. We adopt Mini-batch Gradient Descent in our experiment to minimize the loss function and learn weights of CNNs. During the training process, we use Gaussian distribution to initialize weights randomly and utilize standard Back Propagation to update networks' parameters. Learning rate is initially set as 0.1, and it decays 5% every 2000 epochs. Batch size and momentum are initialized as 64 and 0.9. We keep similar data distribution between training set and testing set to avoid excessive or insufficient feature expression caused by unbalanced distribution.

Hardware environment of our experiment is based on NVIDIA Tesla P100. Software environment is keras 2.1.0 and tensorflow-gpu 1.12.0 built on the basis of Ubuntu 16.04 operation system. Language used is Python 3.6.7.

4.3 Evaluation Metrics and Results

According to given labels and classification results, we use accuracy (ACC), sensitivity (True Positive Rate, TPR) and specificity (SPE) as primary indicators to access performance of proposed CNNs. Specifically, we can use TP (True Positive), FN (False Negative), FP (False Positive) and TN (True Negative) to calculate ACC, TPR and PPV: ACC=(TP+TN)/(TP+TN+FP+FN), TPR=TP / (TP+FN), SPE=TN/(FP+TN).

Results of our experiment are listed in Table 1:

Table 1. performance of different cnns

Model	ACC (%)	TPR (%)	SPE (%)
Inception-ResNet(64×64)	87.23	82.41	89.72
CondeneNet(64×64)	87.56	81.68	88.82
3D-SE-IRNet (48×48×9)	89.51	84.26	91.68
3D-SE-CDNet(48×48×9)	89.93	83.34	91.05

From this table we can see that accuracies of 2D Inception-ResNet and CondeneNet achieve 87.23% and 87.56% correspondingly, which proves feasibility of using CNN assist doctors in early diagnosis of lung cancer. After we introduce self-attention mechanism and 3D convolution into them, there is an obvious improvement in the whole accuracy for both of them, where 3D-SE-IRNet's ACC achieves 89.51% and 3D-SE-CDNet's achieves 89.93%. What's more, not only ACC, TPR and SPE also get enhanced significantly. This is because self-attention mechanism considers reliance among channels of feature map and 3D convolution utilizes reliance among consecutive slices, which makes networks' feature extraction and expression ability better.

5. Conclusion

To improve performance of neural networks in challenge of classification of lung nodule malignancy based on CT scans, this paper briefly reviewed history of CNN and introduces self-attention mechanism and 3D convolution into Inception-ResNet and CondenseNet, which obtain 3D-SE-IRNet and 3D-SE-IRNet correspondingly. We first claimed feasibility of applying CNN into lung nodule malignancy classification and then reviewed development of CNN. Among various kinds of

CNN architectures, we select Inception-ResNet and CondenseNet as candidates due to their outstanding classification performance in ImageNet. Considering reliance among feature map channels and 3D nature of CT scans, 3D-SE-IRNet and 3D-SE-IRNet were designed to further improve accuracy of networks.

Experiment results show that 2D Inception-ResNet and 2D CondenseNet achieve good performance in lung nodule classification. And after introducing self-attention mechanism and 3D convolution into them, ACC of them get further improved where 3D-SE-IRNet's achieves 89.51% and 3D-SE-CDNet's achieves 89.93%.

Acknowledgments

References

[1]. Bray F., Ferlay J., Soerjomataram I., et al. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68, 394-424, 2018.

[2]. T. Kobayashi, X.-W. Xu, H, MacMahon, C. E. Metz, and K. Doi, Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs. Radiology, 199, 843-848, 1996.

[3]. H. MacMahon, R. Engelmann, F. M. Behlen, et al. Computer-aided diagnosis of pulmonary nodules: Results of a large-scale observer test. Radiology, 213, 723-726, 1999.

[4]. LeCun Y., Bottou L., Bengio Y., et al. Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86(11), 2278-2324, 1998.

[5]. D. C. Ciresan, U. Meier, L. M. Gambardella, et al. Deep, big, simple neural nets for handwritten digit recognition. Neural Computation, 22(12), 32073220, 2010.

[6]. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, 1106–1114, 2012.

[7]. Simonyan K., Zisserman A., Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[8]. Zeiler M. D., Fergus R., Visualizing and understanding convolutional networks. In European Conference on Computer Vision, 818-833, 2014.

[9]. C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions. In Conference on Computer Vision and Pattern Recognition, 1-9, 2015.

[10]. M. Lin, Q. Chen, S. Yan. Network in network. arXiv preprint arXiv:1312.4400, 2013.

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

[12]. C. Szegedy, V. Vanhoucke, S. Ioffe, et al. Rethinking the inception architecture for computer vision. In Conference on Computer Vision and Pattern Recognition, 2818–2826, 2016.

[13]. K. He, X. Zhang, S. Ren and J. Sun. Deep residual learning for image recognition. In Conference on Computer Vision and Pattern Recognition, 770-778, 2016.

[14]. C. Szegedy, S. Ioffe, V. Vanhoucke, Inception-v4, inception-resnet and the impact of residual connections on learning. In AAAI Conference on Artificial Intelligence, 4-12, 2017.

[15]. Xie S., Girshick R., Dollár P., et al. Aggregated residual transformations for deep neural networks. In Conference on Computer Vision and Pattern Recognition, 5987-5995, 2017.

[16]. Huang G., Liu Z., Weinberger K. Q., et al. Densely connected convolutional networks. In Conference on Computer Vision and Pattern Recognition, 2261-2269, 2017.

[17]. Huang G., Liu S., Maaten L., et al. CondenseNet: An Efficient DenseNet using Learned Group Convolutions. In Conference on Computer Vision and Pattern Recognition, 2752-2761, 2018.

[18]. Zhang X., Li Z., Loy C. C., et al. Polynet: A pursuit of structural diversity in very deep networks. In Conference on Computer Vision and Pattern Recognition, 3900-3908, 2017.

[19]. M. Wang, B. Liu, H. Foroosh. Factorized Convolutional Neural Networks. In International Conference on Computer Vision Workshops, 545-553, 2017.

[20]. Howard A. G., Zhu M., Chen B., et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

[21]. Sandler M, Howard A, Zhu M., et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Conference on Computer Vision and Pattern Recognition, 4510-4520, 2018.

[22]. Chollet F.. Xception: Deep learning with depthwise separable convolutions. In Conference on Computer Vision and Pattern Recognition, 1800-1807, 2017.

[23]. Zhang X., Zhou X., Lin M., et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Conference on Computer Vision and Pattern Recognition, 6848-6856, 2018.

[24]. Ma N., Zhang X., Zheng H. T., et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design. arXiv preprint arXiv:1807.11164, 2018.

[25]. Hu J., Shen L., Sun G.. Squeeze-and-excitation networks. In Conference on Computer Vision and Pattern Recognition, 7132-7141, 2018.

[26]. Zoph B., Le Q. V.. Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578, 2016.

[27]. Zoph B., Vasudevan V., Shlens J., et al. Learning transferable architectures for scalable image recognition. In Conference on Computer Vision and Pattern Recognition, 8697-8710, 2018.

[28]. Liu C., Zoph B., Shlens J., et al. Progressive neural architecture search. arXiv preprint arXiv:1712.00559, 2017.

[29]. G. Wei, H. Ma, W. Qian, et al. Lung nodule classification using local kernel regression models with out-of-sample extension. Biomedical Signal Processing and Control, 40, 1–9, 2018.

[30]. G. Wei, H. Cao, H. Ma, et al. Content-based image retrieval for lung nodule classification using texture features and learned distance metric. Journal of Medical Systems, 42(1), 13, 2017.

[31]. M. Sergeeva, I. Ryabchikov, M. Glaznev, et al. Classification of pulmonary nodules on computed tomography scans:Evaluation of the effectiveness of application of textural features extracted using wavelet transform of image. In Conference of Open Innovations Association and Seminar on Information Security and Protection of Information Technology, 2016.

[32]. J. Ma, Q. Wang, Y. Ren, et al. Automatic lung nodule classification with radiomics approach. In SPIE Medical Imaging, 9789, 978906, 2016.

[33].   W. Shen, M. Zhou, F. Yang, et al. Multi-scale convolutional neural networks for lung nodule classification. In International Conference on Information Processing in Medical Imaging, 588–599, 2015.

[34].   W. Shen, M. Zhou, F. Yang, et al. Multi-crop convolutional neural networks for lung nodulemalignancy suspiciousness classification. Pattern Recognition, 61, 663–673, 2017.

[35].   S. Hussein, K. Cao, Q. Song, et al. Risk stratification of lung nodules using 3d cnn-based multi-task learning. In International Conference on Information Processing in Medical Imaging, 249–260, 2017.

[36].   W. Zhu, C. Liu, W. Fan, et al. Deeplung: 3d deep convolutional nets for automated pulmonary nodule detection and classification. arXiv preprint arXiv:1709.05538, 2017.

[37].   Y. Chen, J. Li, H. Xiao, et al. Dual path networks. arXiv preprint arXiv:1707.01629, 2017.

[38].   G. Kang, K. Liu, B. Hou, et al. 3D multi-view convolutional neural networks for lung nodule classification. PloS one, 12(11), e0188290, 2017.