

Lung Nodule Detection Using Convolutional Neural Networks

Jiaying Shi



Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2018-27

<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-27.html>

May 3, 2018

Copyright © 2018, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Lung Nodule Detection Using Convolutional Neural Networks

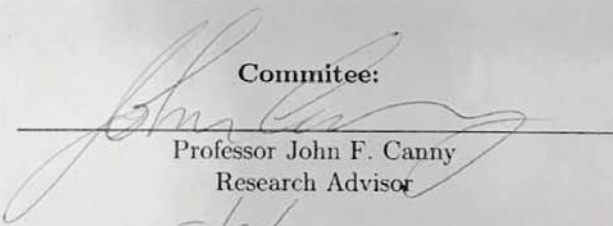
by Jiaying Shi

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences,
University of California at Berkeley, in partial satisfaction of the requirements for the
degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:


Committee:



Professor John F. Canny
Research Advisor

5/3/18

(Date)



Professor Shmuel S. Oren
Second Reader

5/3/2018

(Date)

May 2, 2018

Abstract

We present a novel procedure to apply deep learning techniques to medical image classification. With increasing popularity of computed tomography (CT) lung screening, fully manual diagnosis of lung cancer puts a burden on the radiologists who need to spend hours reading through CT scanned images to identify Region of Interests (ROIs) to schedule follow-ups. Accurate computer-aided diagnosis of lung cancer can effectively reduce their workload and help training new radiologists. However, lung cancer detection is challenging because of the varying size, location, shape, and density of nodules. Many studies have approached this problem using image-processing techniques with the intention of developing an optimal set of features. Convolutional neural network has demonstrated to learn discriminative visual features automatically and has beat many state-of-art algorithms in image-processing tasks, such as pattern recognition, object detection, segmentation, etc. In this report, we evaluate the feasibility of implementing deep learning algorithms for lung cancer diagnosis with the Lung Image Database Consortium (LIDC) database. We compare the performance of two 2.5D convolutional neural network models, two 3D convolutional neural network models, and one 3D convolutional neural networks with a 3D spatial transformer network module on the task of detecting lung nodules. The performance of our best model is comparable with the state-of-art results in the lung nodule detection task.

1 Introduction

Lung cancer is the number one cause of deaths in both men and women in the US. Statistics from the American Cancer Society estimated that in 2017, about 222,500 new lung cancer cases occurred in the U.S with estimated 155,800 deaths. It accounts for 1 in 4 cancer deaths [1]. The 5-year relative survival rate for lung cancer is 15% for men and 21% for women. However, the survival rate is 55% for early detected lung cancer that is at a localized stage which only takes only 16% of all cases. The prognosis of lung cancer is usually poor because patients tend not to feel sick until it is at an advanced stage. Though advances had been made in surgical, radiotherapeutic, and chemotherapeutic approaches have been made, the long-term survival rate remains low[2]. In the American Cancer Society report [1], it suggested that screening with low-dose spiral computed tomography(LDCT) has been shown to reduce lung cancer mortality by about 20% compared to standard chest x-ray among adults with at least 30 pack-year smoking history who were current smokers or had quit within 15 years.

Lung nodules are defined as rounded opacities, with moderately identifiable margins and diameters smaller than 3cm in [3]. Granulomatous disease and lung cancer are two common causes of lung nodules [4]. The contrast between the nodule and lung of CT is higher than that of chest radiography. CT also eliminates overlying structures such as the chest wall, mediastinum, and vessels [5]. However, limitations in nodule identification using CT have to be noticed. In [6], the overlooked nodules were small, faint in attenuation, adjacent to vessels, or adjacent to findings of prior tuberculosis. There are multiple ways to characterize lung nodules: by their morphology, densitometry, size and growth [7]. The development of CT that enables higher spatial and contrast resolution, more nodule characterization features are reflected on CT scans.

Lung cancer refers to those nodules in the lung that are malignant, which could grow

aggressively locally and enter the bloodstream to spread to other sites in the body. Lung cancer is more life-threatening than other types of cancers because it starts to spread very early after its formation. Efficiently and accurately identify lung cancer in the early stage is beneficial. Computer-aided diagnosis (CAD) can assist radiologists to efficiently identify lung cancer through CT scans. With the increasing popularity of regular CT screening for elderly people, the number of CT scans grow rapidly, and it takes a great amount of time for radiologists to manually read through all the CT scans to identify nodules. An initial automatic filtering system for nodule detection has the potential of reducing the work requested for radiologists, and such a system may be useful for training new radiologists in the future. The system can also significantly reduce the work for labeling large dataset that is useful in training more sophisticated models.

Early detection of lung cancer is challenging, due to the various shape, size, density, and location of nodules. Studies have adopted machine learning approaches such as segmentation, clustering, Artificial Neural Network and Supporter Vector Machine (SVM) to tackle this problem [8, 9, 10, 11, 12, 13]. However, little has been done in using Convolution Neural Network (CNN) approaches to classify nodules in CT images, primarily due to the limited size of labeled databases. CNN is inspired by the biological functions of neurons. A typical CNN framework consists of several convolutional layers, regularization layers, subsampling and fully connected layers. CNN has demonstrated to achieve better results in many image processing tasks, such as segmentation, object detection, etc. Compared to previous approaches that involve complicated image pre-processing and designing optimal vision features, CNN is able to learn discriminative visual features automatically and could potentially provide better classification accuracy in nodule detection.

In this project, we aim to experiment and implement various deep learning architectures in order to achieve human-level accuracy in CAD systems. In particular, we are interested in detecting nodules from lung CT scans in this project. We would like to classify nodule and non-nodule region of interests with high sensitivity and low false positive rate using various types of convolutional neural networks. This task can be considered as the first step in building CAD systems for lung cancer diagnosis. Moreover, by automating this task, we can significantly reduce the time for the radiologists to create large-scale labeled datasets of lung CT scans.

The remainder of the report is organized as follows. In Section 2, we review existing literature on automating the detection of lung nodules and lung cancer. In Section 3, we provide an overview of the dataset and present how the data is preprocessed. In Section 4, we present our convolutional neural network models for lung nodule detection and experiment results on those models. And Section 5 concludes the report.

2 Related Work

Many computer-aided systems have been developed to help radiologists better screen and classify lung cancer. We summarize these systems by the methods they adopt.

2.1 CNN-based Approaches

Ciampi et al. used a convolutional neural network out-of-the-box where they extracted features from a CNN pre-trained on the imageNet dataset and then applied a simple linear SVM classifier to classify peri-fissural nodules (a benign nodule in the lung with a high

false positive rate in lung cancer detection) [8]. They used the ensemble of classifiers for 3D slices of an object to obtain the final classifier and compared its performance with a bag of frequencies descriptor with similar accuracy of 86.8%. Hua explored a 3-layer CNN and a Deep Belief Network to approach the problem of nodule classification in computed tomography images [14]. Sun et al. considered three deep learning algorithms: CNN, Deep Belief Networks (DBNs) and Stacked Denoising Autoencoder (SDAE) to classify nodules. They showed that the performance on downsampled data of the three methods are similar and are around 80% [15]. Shen et al. proposed a hierarchical learning framework using Multi-scale CNN (MCNN). The method captures nodule heterogeneity by extracting features from stacked layers, especially with multi-scale nodule patches [16].

2.2 Neural Network Approaches

Many studies have focused on Artificial Neural Network approaches using an optimal combination of features [9, 11, 17, 18]. In particular, Kuruvilla and Gunavathi proposed an artificial neural network approach using statistical parameters like mean, standard deviation, skewness, kurtosis, fifth central moment and sixth central moment to achieve the classification accuracy of 91.1% [19]. Abdulla et al. used area, perimeter and shape as features to train an artificial neural network for classification of lung cancer with a classification accuracy of 90% [20]. Ada and Kaur carried out a computational procedure that effective sort images into groups by similarity. After using Histogram Equalization and morphology to pre-process the data, they extracted features, applied Principle Component Analysis (PCA) and trained a Neural Network Classifier with a single hidden layer [21].

2.3 Other Image Processing Approaches

Image processing techniques have been greatly explored in nodule classification for lung CT images. Numerous studies adopted segmentation, morphological operations and contour filter approaches for better nodule detection [10, 22, 12, 23, 13]. Nathaney and Kalyani developed a system to segment CT images and to classify focal areas in lung region in real time. They applied adaptive threshold method and ROI processing to detect lung cancer with least false negative rate [24]. Another image processing procedure LCDS is developed by Chaudhary and Singh, where they use smoothing, Gabor filter enhancement and Watershed segmentation in the pre-processing phase and then implemented morphology and colorimetric to extract features such as area, perimeter, and roundness. They compare the statistics of different features to identify the stage of lung cancer for each patient [25]. Sharma and Jindal first applied basic image processing techniques such as erosion, median filter, dilation, outlining, and lung border Extraction to detect lung region and then applied segmentation algorithms to detect cancer nodules and then applied a pre-specified classify to identify cancer nodules [26].

3 Data Set and Prepross

3.1 LIDC Dataset

The Lung Image Database Consortium (LIDC) dataset is one of the first nodule-specific performance benchmarks created as both a research and training resource [27, 28]. Numer-

ous studies and algorithms have been conducted and deployed with respect to the LIDC dataset and other publicly available data sets in order to objectively access the effectiveness. In this project, we train and test models on the LIDC dataset. Currently, there are 1018 cases available in the dataset. In the dataset, each case could have 0 to 6 nodules. A nodule represents a spectrum of abnormalities(irrespective of presumed histology), which is itself a subset of a boarder spectrum of abnormalities. The diameter of the nodules is between $3 - 30mm$. Nodules on the scans can represent primary lung cancer, metastatic disease, or non-cancerous processes.

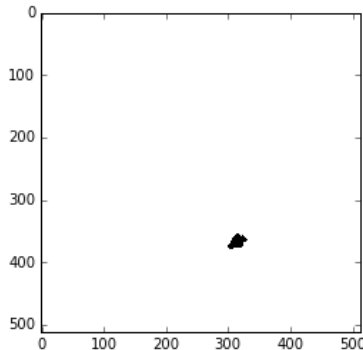


Figure 1: An Example of Nodule Annotation in LIDC

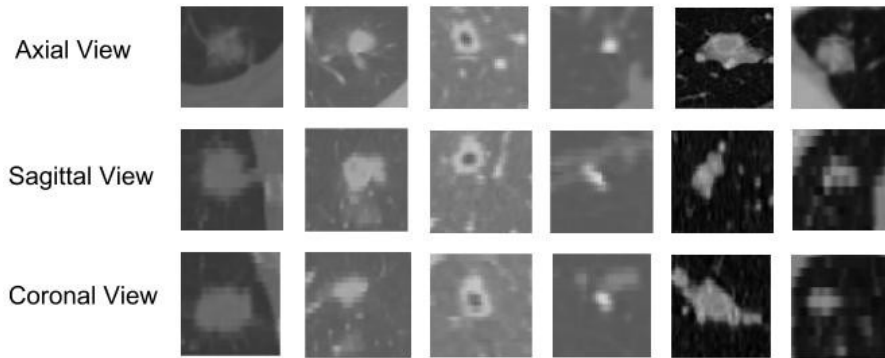


Figure 2: Examples of Nodules in Axial, Sagittal, and Coronal Views

The LIDC adopted a two-phase reading process. The first phase is blind-folded. For each case, there are four readers reading and annotating each case independently. The reading and annotation of each reader are done asynchronously and independently. For nodules with a diameter larger than $3mm$ but less than $30mm$, each reader drew a complete outline around the nodule in all sections in which it appeared, with the pixels that comprise the outline at the first pixel outside the nodule. For nodules whose diameter is smaller than $3mm$, the readers only indicate the approximate three-dimensional center-of-mass of any such nodule. Once the first phase results are collected from all four readers, the second

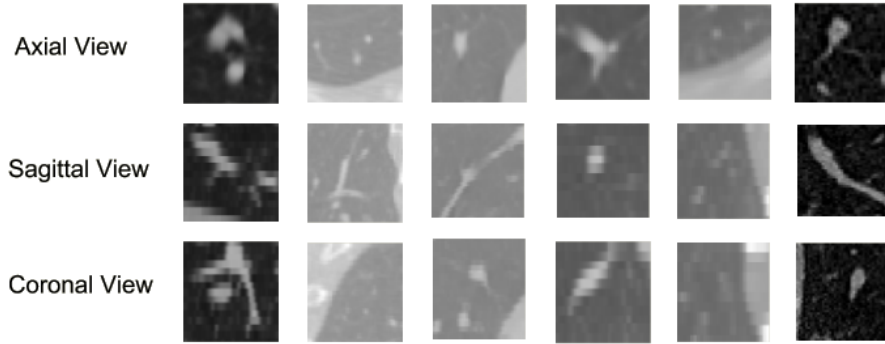


Figure 3: Examples of Non-Nodules in Axial, Sagittal, and Coronal Views

phase of reading is conducted. In the second phase, the first phase results are compiled and sent back out to the same reader so that they could see both their own markings as well as the markings from the other three readers. Each reader then, again independently, read each case, this time with the benefit of information as to what other readers saw/marked, and then made a final decision about the markings for that case. The annotation results of all readers are provided in the dataset. One example of the nodule annotation is shown in Figure 1. Examples of nodules in axial, sagittal, and coronal views are shown in 2. Examples of non-nodules are shown in 3. From the figures, we can see that there is no standard shape for nodules. Moreover, normal tissues on CT scans can look similar as nodules. In [29], statistical significance for lobe location was assessed by the Chi-square test. In the study, a significant asymmetry in overall lesion distribution was observed between the right and left lung. About 67.78% lesions are in the right lung. Lesions are also distributed unevenly between the lower and upper lobes. In the paper, the authors reported that malignant nodules were found to be significantly more common in the right upper lobe.

3.2 Preprocess

The preprocessing of the dataset includes four steps. We first parse the xml file to generate the masks for nodules with a diameter greater than $3mm$. Then, we extract the lung region from each DICOM image. In the third step, we rescale the 3D image along the axial, coronal and sagittal axis so that each voxel represents a $1mm \times 1mm \times 1mm$ volume. Last, we generate patches of images to generate training and test samples by cropping the rescaled 3D images.

In the CT slices, the intensity is proportional to the density of the tissue. The lung areas can be segmented out through a thresholding-based strategy. Thresholding-based methods segment the image by creating binary partitions that are based on image attenuation values. The procedure of lung mask extraction is illustrated in Figure 4. To extract the lung, we first extract the body part from the image using thresholding method. Then, the lung mask is generated given the body mask of the CT image. In [30], the authors point out that methods purely based on thresholding typically ignore the spatial characteristics of the lung. They are also sensitive to noise, and the thresholding interval is always set to

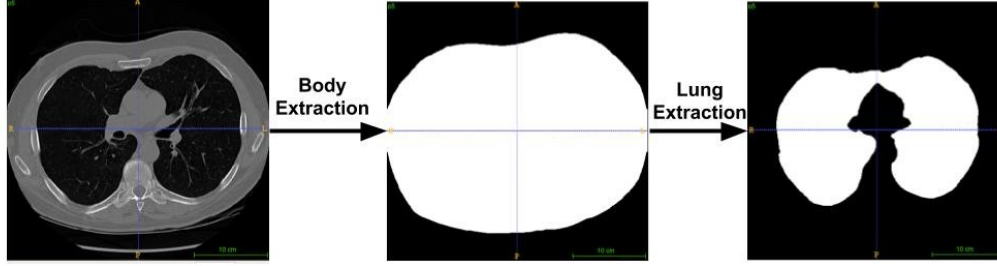


Figure 4: Lung Mask Extraction

exclude adjacent tissues from lung fields. However, the attenuation values of pathologic regions may locate in similar regions. We adopt additional morphological operations to correct the resulting segmentation. We applied connected component labeling to filter out noisy regions with low densities out of the lung. Dilation is adopted to remove noisy points within the lung part. Erosion is then utilized to adjust the boundary inaccuracy caused by dilation. An illustration of the morphological operations are shown in Figure 5.

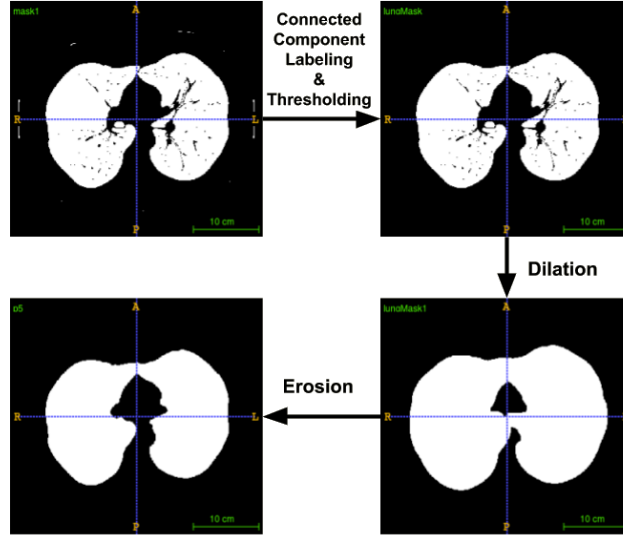


Figure 5: Morphological Operations in Lung Mask Extraction

In the data set, the CT scans are taken by machines with different settings. The image is calibrated differently. Scans have different voxel spacing, which can be problematic for convolutional neural networks. We rescale all CT scans to an isotropic resolution by interpolation so that the dimension of each voxel represents a $1mm \times 1mm \times 1mm$ volume. The training and test samples for all models are generated from the rescaled CT scans. The samples with dimension $41 \times 41 \times 41$ are generated by a sliding-window type approach.

For the 2.5D CNN models, the input images are obtained from the 3D samples so that we can compare the performance of the 2.5D CNN models and the 3D CNN models. We also follow these rules while generating the samples to avoid obvious non-nodule samples:

- The center voxel locates on the lung.
- The Hounsfield unit of the center voxel is above 100.
- The minimum Hounsfield unit of a box of size $5 \times 5 \times 5$ around the center voxel is above 0.

Since a nodule with diameter $3mm$ only takes about 0.1% of the total volume of the lung, the training samples are extremely imbalanced. We balance the dataset by subsample negative samples and augment the positive samples. The augmentation is a combination of 3D shifting and 3D rotation. We do not have zoom in/out in the augmentation since the size of the nodules contains important information. Similarly, we limit the angles of 3D rotations since the relative location of high attenuations to airways and vessels indicate whether it is a nodule.

4 Models and Results

4.1 Models

We study and compare the results of 6 different models: two 2.5D CNN models, two 3D CNN models, 1 3D CNN model with patch location information, and 3D CNN model with 3D spatial transformer network. The architectures of all models are similar. The fundamental building block of all models is the 2D/3D inception module as in Figure 6. The inception module was first proposed in [31] and then improved in [32]. The inception architecture improves the utilization of the computing resources inside the network. This was achieved by approximating the expected optimal sparse structure by smaller dense building blocks. This structure allows for increasing the depth and width of the network while keeping the computational budget constant.

2.5D CNN Models: We build and compare two 2.5D models. The input of the 2.5D models is 2D images. However, we combine the nine slices of axial, coronal and sagittal views in different ways. In the first 2.5D CNN model(2.5D-1) as shown in Figure 7, slices of all views are first concatenated, and there are 27 channels in the input image. In the second 2.5D CNN model(2.5D-2) as shown in Figure 8, slices for different views are first passed into CNNs with same structures and then concatenate the output of the CNNs. The number of filters in each convolution layer in the second 2.5D model. We have a fully connected layer and a pooling layer followed by the CNN modules of each view before concatenation to reduce the dimension of the embedding output. Although the CNNs in model 2 for different views has the same architecture, the parameters of the CNN for each view are different.

Comparing with purely feeding 2D slices of one view into CNNs, the 2.5D CNNs models provide are able to perceive more information of the 3D structures of lung tissues. The second model has CNN modules for different views. This might help identify nodules from other high attenuation parts of the lung. For example, nodules show as bright round shapes on slices of all three views, while vessels show as round/olive shapes in one or two slices but a belt shape on other views.

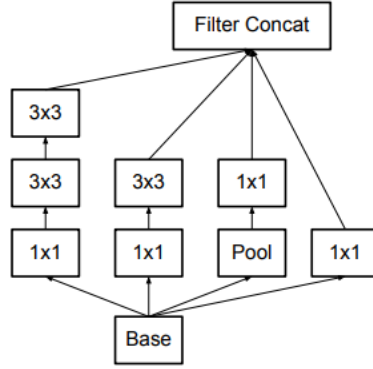


Figure 6: Inception Module from [32]

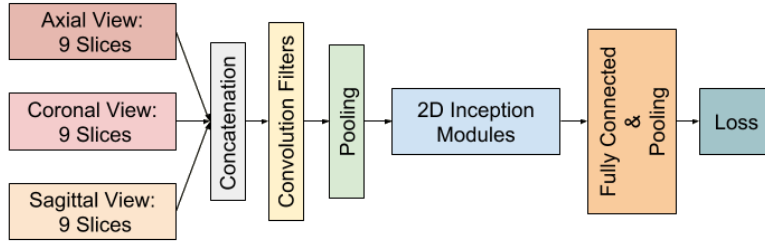


Figure 7: Architecture of 2.5D-1

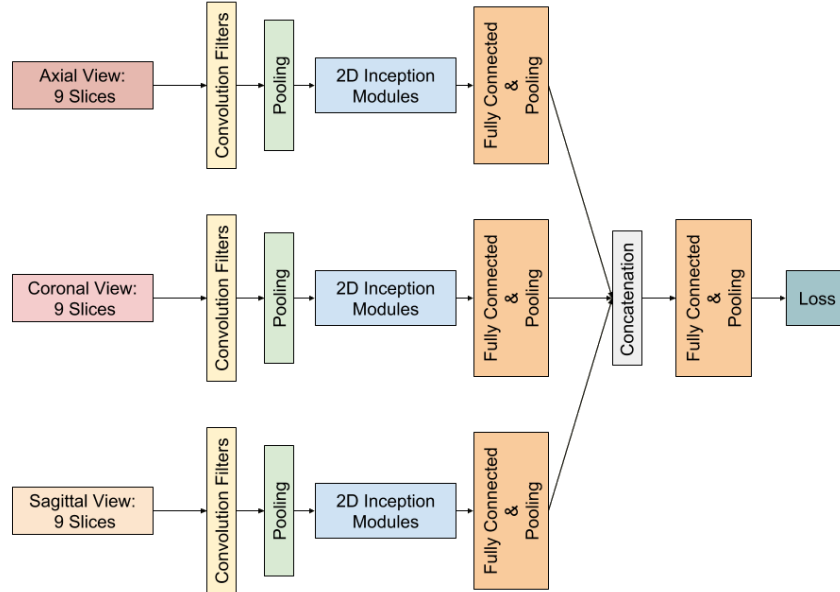


Figure 8: Architecture of 2.5D-2

3D CNN Models: We also build and compare two 3D CNN models. The architecture of the 3D CNN models is similar to that of the 2D-1. All convolution filters in the 3D CNN models are in 3D, and the pooling operations are also conducted in 3D. The input of the 3D models is 3D samples generated following the logic defined the previous section. We first studied two 3D CNN models (3D-1 and 3D-2) with the same architecture. The difference between the two models is that in each convolution layer 3D-2 has more filters. 3D-2 is more complicated than 3D-1 in the sense that it has more parameters. The 3D models classify the samples based on the 3D input directly without loss of the 3D information of the input data. We expect better performance from the 3D model than the 2D model.

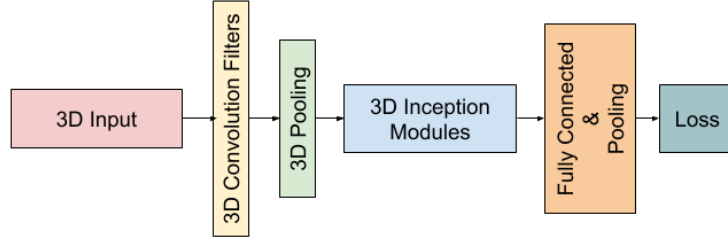


Figure 9: Architecture of the 3D CNN Models

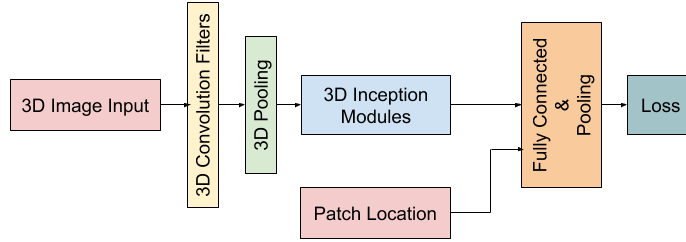


Figure 10: 3D CNN Model with Patch Location Information

3D CNN Model with Patch Location Information: We build a 3D CNN model with path location information as input (3D-LOC). The architecture of the model is shown in Figure 10. The location of the patch is computed using the coordinates of the center voxel divided by the dimension of the CT scan. The location information is hence a vector of 3 components in interval $[0, 1]$. The location information vector is concatenated with the output of the 3D inception modules and fed into the fully connected layer. The 3D-LOC model takes the relative position of a patch within a scan as input. If the nodule is location dependent, including such information will improve the performance of the model.

3D CNN Model with Spatial Transformer Network: We have another 3D model that has a spatial transformer network on top of the 3D CNN (3D-ST). Spatial transformer network is first introduced in [33]. It provides CNNs the ability to be spatial invariant to input data in a computationally efficient manner. Spatial transformer network has three components:

- A localization network that takes the input image and outputs the parameters of

spatial transformation.

- A grid generator that takes in the spatial transformation parameters and generates the sampling grid that the input image should be sampled to obtain the transformed output image.
- A sampler that produces the output map sampled from the input at the grid points.

The spatial transformer network can be considered as a component of the whole network and can be trained end-to-end with the remaining part of the network. In our model, we limit the 3D spatial transformation to the rotation. The 3D transformation matrix is then a 3×4 matrix with the first three columns being a 3D rotation matrix and the last column being all zeros. To preserve the size information of the nodules, we do not include scaling and shearing.

4.2 Experiments and Results

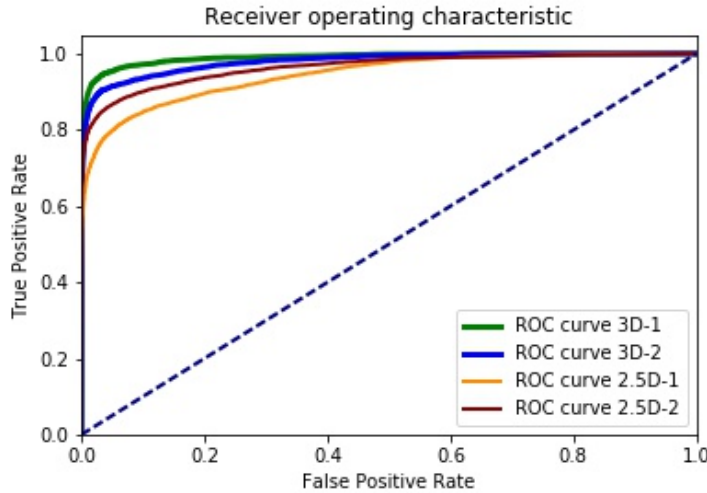


Figure 11: Receiver Operating Characteristic Curves

Training Methodology: In the experiment, we use samples generated from 815 scans in the LIDC dataset as the training set. Samples generated from the remaining scans are used as test set. The experiments are conducted on a desktop equipped with an Intel i5 processor, 32GB memory, and a GeForce GTX 960 GPU. We utilize RMSprop to train our models. The hyperparameters of the optimizer are similar to those of in [32]. For the 2.5D CNN models, we used a learning rate of 0.04. The decay of the RMSprop was 0.9 and $\epsilon = 1.0$. For the 3D CNN models, we conducted a random search for hyperparameter optimization. In the 3D CNN model with the best performance, we used 0.87 for the decay of the RMSprop, 1.0 for ϵ , and 0.002 for the learning rate. A random search on hyper-parameters is also conducted to achieve good performance of 3D-ST.

2.5D-1 v.s. 2.5D-2: We first compared the test results of the two 2.5D CNN models. As we expected, 2.5D-2 performs better than 2.5D-1 as shown in Figure 11. The AUC

of 2.5D-2 is 0.9443, while the AUC of 2.5D-1 is 0.9639. For 2.5D-2, the sensitivity of the classifier is above 0.9 when the false positive rate is above 0.2. For 2.5D-1, the sensitivity of the classifier is above 0.9 when the false positive rate is above 0.3. The results of the 2.5D CNN models demonstrate our conjecture that utilizing independent CNN modules to process slices of different views has better performance. This coincides with the fact that radiologists find nodules from CT scans by viewing the axial, coronal, and sagittal plains independently in order to infer the 3D shape of suspicious tissue.

3D CNN Models v.s. 2.5D CNN Models: From Figure 11, we can see that both 3D models perform better than 2.5D models as we expected. Comparing with 2.5D CNN models, the input is lossless regarding spatial information. The AUC of 3D-1 is 0.9894, and the AUC of 3D-2 is 0.9792. Though 3D-2 has more filters and is more complicated, the performance is worse than that of 3D-1. We observe overfitting for 3D-2. The training accuracy of 3D-2 is higher than that of 3D-1 while the test accuracy of 3D-2 is the lower one.

3D CNN Models v.s. 3D-CNN-LOC: We trained the 3D-LOC model using 3 sets of

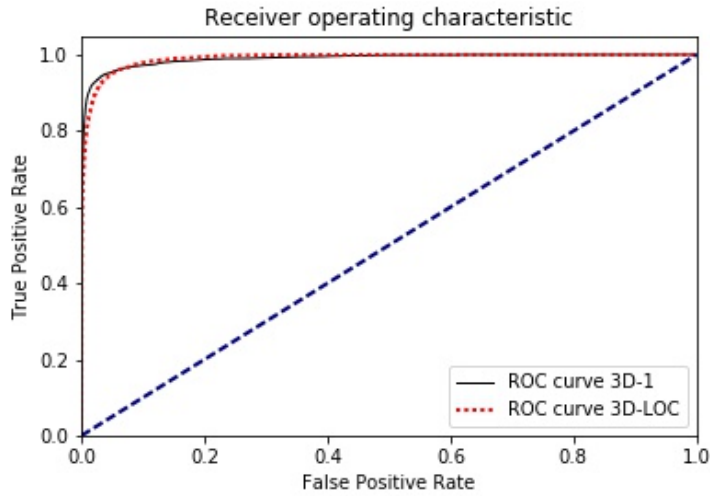


Figure 12: Receiver Operating Characteristic Curves of 3D-1 and 3D-LOC

hyperparameters and compared the 3D-LOC model that had the best performance with 3D-1. The ROC curves of 3D-LOC and 3D-1 are shown in Figure 12. The AUC of 3D-LOC is 0.9907 which is higher than that of 3D-1. From the ROC curves, we can see that when the false positive rate is smaller than 0.1, 3D-1 performs better. However, when the false positive rate is higher than 0.1, 3D-LOC performs better.

We also compared our nodule candidate detection results with the some of best LUNA nodule candidate detection results summarized in [34]. The results are shown in Table 1. In the competition, both sensitivity and the number of candidates per scan which indicates the selectivity are reported and compared. In addition to achieving higher sensitivity, another aspect of the goal is to propose as few candidates as possible, which indicates the selectivity of the model. We can see that the sensitivities of 3D-1 and 3D-LOC are slightly better than the proposed model in [34] while generating similar number of candidates per

scan.

Table 1: Comparison with LUNA Competition Results

System	Sensitivity	Candidates/Scan
LUNA-ISICAD	0.856	335.9
LUNA-ETROCAD	0.929	333
Results from [34]	0.946	15.0
3D-1	0.9475	14.9
3D-LOC	0.9439	15.0

3D-1 v.s. 3D-ST: The 3D-ST model performs worse than 3D-1. To understand why the 3D-ST model performs worse than our best 3D CNN model, we tried two approaches: the first is to check whether the spatial transformer is trained; the second is to check whether the interpolation in the grid sampling module of spatial transformer introduces addition loss.

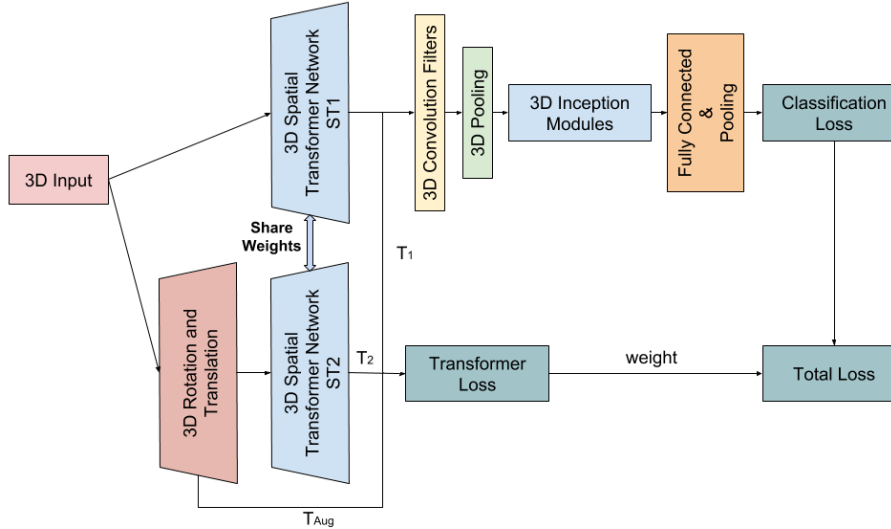


Figure 13: Architecture of 3D-ST in training

We first add an additional path in the 3D-ST model to force the 3D spatial transformer to converge first. To avoid the artifacts on the boundaries caused by 3D spatial transformation, we set the input images of the spatial transformer network module to be $73 \times 73 \times 73$. We then crop the output of the 3D spatial transformer module to $41 \times 41 \times 41$ and take the cropped image without artifacts as the input of the following 3D CNN module. The dimension of the input of the second path is $127 \times 127 \times 127$ to avoid artifacts. It is then cropped into $73 \times 73 \times 73$ after the 3D random rotation. The architecture of 3D-ST in the training process is shown in Figure 13. The parameters of the fully connected layers that generate the spatial transformation parameters are shared between the two paths. The loss function has two parts: the classification loss and the weighted spatial transformer loss. The transformer loss measures the difference between T_1 and the composition of

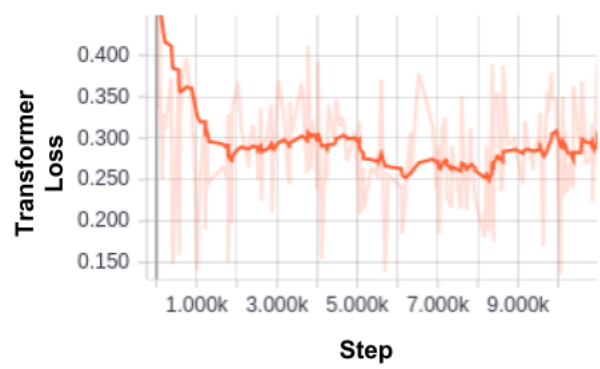


Figure 14: Transformer Loss in Training with Hyperbolic Tangent Activation Function

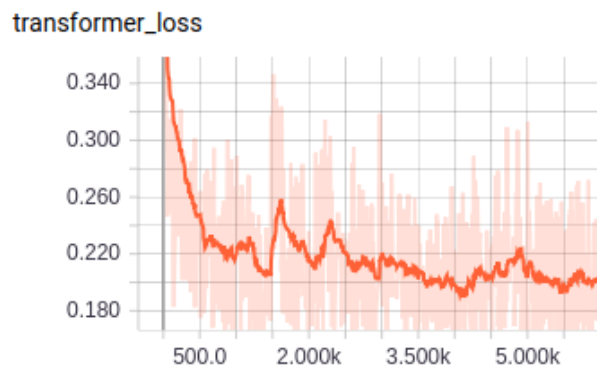


Figure 15: Transformer Loss in Training with Leaky ReLU Activation Function

T_{Aug} and T_2 . The transformer loss when we adopt hyperbolic tangent activation function (3D-ST-tanh) for spatial transformer network is shown in Figure 14. The weight of the spatial transformer loss is a hyperparameter which is tuned in the training process. We can see that the transformer loss decreased from 0.45 to 0.26. One possible reason that the transformer loss cannot be further decreased is that the neurons of the localization network are saturated. The gradient of updating the parameters of the localization network is very small however the update for the 3D-CNN network is more significant. To address this issue, we tried to utilize leaky ReLU as the activation function when generating 3D spatial transformation parameters. The transformer loss when we have leaky ReLU activation function is shown in Figure 15. The transformer loss can be decreased from around 0.45 to 0.19. However, the performance of the best 3D-ST is still worse than that of 3D-1.

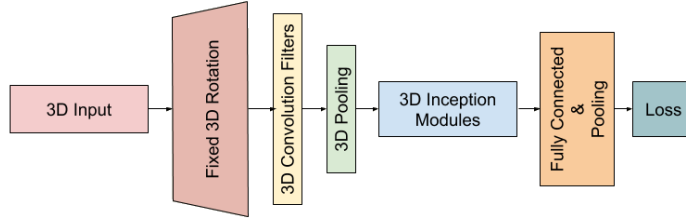


Figure 16: 3D CNN with Rotated Input (3D-RT)

We also tried to train the 3D CNN model with fixed 3D rotation to test whether the bilinear interpolation module in grid sampling of the spatial transformer networks introduced additional loss. The architecture of the 3D CNN with fixed 3D rotation is shown in Figure 16. The input of 3D-RT is of size $73 * 73 * 73$ to avoid the artifacts on boundary voxels introduced by rotation. After rotating the input image, the 3D images are cropped into images of size $41 * 41 * 41$ as the input of the 3D CNN module. We trained a 3D-RT model with same hyperparameters as in 3D-1. The ROC curves are compared in Figure 17. The

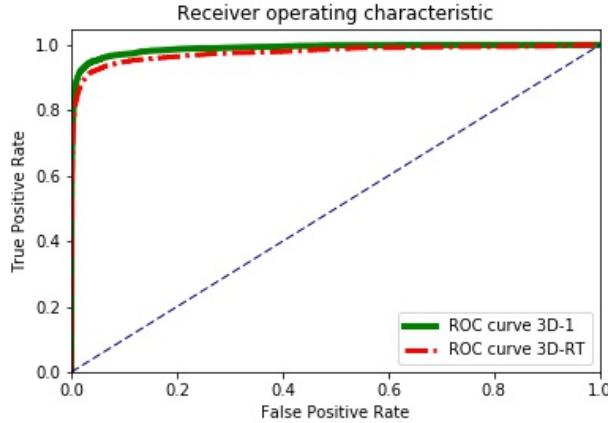


Figure 17: Comparison of ROC curves of 3D-1 and 3D-ROT

results shows that 3D-RT performs worse, which indicates that the bilinear interpolation

introduces loss.

5 Conclusions

We present a novel CNN-based procedure to detect nodules in Lung CT images. We study and compare two 2.5D CNN models, two 3D CNN models, a 3D CNN model with patch location information, a 3D CNN model with fixed spatial transformation, and a 3D CNN model with 3D spatial transformer network module. Results show that 2.5D CNN model that has independent CNN modules for slices of each view performs better than the model that uses one CNN module for all views. The 3D CNN model with fewer parameters performs best among all models. In our experiment, the 3D CNN model with 3D spatial transformer network performs worse than the best 3D CNN model. By comparing the 3D CNN model and the corresponding model with fixed rotation, with the fixed rotation, the performance of the model is worse. We suspect that the bilinear interpolation in grid sampling of spatial transformer network will introduce noise in the model and influence the performance of the model. To conclude whether spatial transformer will be helpful and what kind of spatial transformation should be included requires further research. The 3D CNN model we developed can be used to detect lung nodules from CT scans. Accurate automatic nodule detection can be applied as the initial step to identify ROIs in CT images to reduce the manual work required for radiologists. Also, this kind of system can help train new radiologists with a wide selection of images and corresponding labels. Such a system can also speed up the process of creating a new labeled medical dataset for future research.

References

- [1] American Cancer Society. *Cancer facts and figures 2017*. Tech. rep. American Cancer Society, 2017. URL: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf>.
- [2] National Lung Screening Trial Research Team et al. “Reduced lung-cancer mortality with low-dose computed tomographic screening”. In: *N Engl J Med* 365 (2011), pp. 395–409.
- [3] JH Austin et al. “Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society.” In: *Radiology* 200.2 (1996), pp. 327–331.
- [4] Blair A Keagy et al. “Major pulmonary resection for suspected but unconfirmed malignancy”. In: *The Annals of thoracic surgery* 38.4 (1984), pp. 314–316.
- [5] Claudia I Henschke et al. “Early Lung Cancer Action Project: overall design and findings from baseline screening”. In: *The Lancet* 354.9173 (1999), pp. 99–105.
- [6] Ryutaro Kakinuma et al. “Detection failures in spiral CT screening for lung cancer: analysis of CT findings”. In: *Radiology* 212.1 (1999), pp. 61–66.
- [7] Jane P Ko and David P Naidich. “Lung nodule detection and characterization with multislice CT”. In: *Radiologic Clinics* 41.3 (2003), pp. 575–597.

- [8] Francesco Ciompi et al. “Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box”. In: *Medical image analysis* 26.1 (2015), pp. 195–202.
- [9] Sheeraz Akram et al. “Artificial neural network based classification of lungs nodule using hybrid features from computerized tomographic images”. In: *Applied Mathematics & Information Sciences* 9.1 (2015), p. 183.
- [10] Anindya Gupta et al. “A tool for lung nodules analysis based on segmentation and morphological operation”. In: *Intelligent Signal Processing (WISP), 2015 IEEE 9th International Symposium on*. IEEE. 2015, pp. 1–5.
- [11] Lin Lu et al. “Hybrid detection of lung nodules on CT scan images”. In: *Medical physics* 42.9 (2015), pp. 5042–5054.
- [12] Temesguen Messay, Russell C Hardie, and Timothy R Tuinstra. “Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset”. In: *Medical image analysis* 22.1 (2015), pp. 48–62.
- [13] Atsushi Teramoto et al. “Automated detection of lung tumors in PET/CT images using active contour filter”. In: *Medical Imaging 2015: Computer-Aided Diagnosis*. Vol. 9414. International Society for Optics and Photonics. 2015, p. 94142V.
- [14] Kai-Lung Hua et al. “Computer-aided classification of lung nodules on computed tomography images via deep learning technique”. In: *OncoTargets and therapy* 8 (2015).
- [15] Wenqing Sun, Bin Zheng, and Wei Qian. “Computer aided lung cancer diagnosis with deep learning algorithms”. In: *Medical Imaging 2016: Computer-Aided Diagnosis*. Vol. 9785. International Society for Optics and Photonics. 2016, 97850Z.
- [16] Wei Shen et al. “Multi-scale convolutional neural networks for lung nodule classification”. In: *International Conference on Information Processing in Medical Imaging*. Springer. 2015, pp. 588–599.
- [17] Weisheng Wang et al. “Data analysis of the lung imaging database consortium and image database resource initiative”. In: *Academic radiology* 22.4 (2015), pp. 488–495.
- [18] Koichiro Yasaka et al. “High-resolution CT with new model-based iterative reconstruction with resolution preference algorithm in evaluations of lung nodules: Comparison with conventional model-based iterative reconstruction and adaptive statistical iterative reconstruction”. In: *European journal of radiology* 85.3 (2016), pp. 599–606.
- [19] Jinsa Kuruvilla and K Gunavathi. “Lung cancer classification using neural networks for CT images”. In: *Computer methods and programs in biomedicine* 113.1 (2014), pp. 202–209.
- [20] Azian Azamimi Abdullah and Syamimi Mardiah Shaharum. “Lung cancer cell classification method using artificial neural network”. In: *information engineering letters* 2.1 (2012).
- [21] Rajneet Kaur Ada¹. *Early detection and prediction of lung cancer survival using neural network classifier*. 2013.

- [22] Anindya Gupta et al. “Methods for increased sensitivity and scope in automatic segmentation and detection of lung nodules in CT images”. In: *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on*. IEEE. 2015, pp. 375–380.
- [23] Sudipta Mukhopadhyay. “A segmentation framework of pulmonary nodules in lung CT images”. In: *Journal of digital imaging* 29.1 (2016), pp. 86–103.
- [24] Ms Gangotri Nathaney and Kanak Kalyani. “Lung Cancer Detection System on Thoracic CT Images Based on ROI Processing”. In: *Lung Cancer* 4.4 (2015), pp. 173–176.
- [25] Anita Chaudhary and Sonit Sukhraj Singh. “Lung cancer detection on CT images by using image processing”. In: *Computing Sciences (ICCS), 2012 International Conference on*. IEEE. 2012, pp. 142–146.
- [26] Disha Sharma and Gagandeep Jindal. “Identifying lung cancer using image processing techniques”. In: *International Conference on Computational Techniques and Artificial Intelligence (ICCTAI)*. 2011, pp. 115–120.
- [27] Samuel G Armato et al. “The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans”. In: *Medical physics* 38.2 (2011), pp. 915–931.
- [28] S Armato III et al. *Data from LIDC-IDRI. The cancer imaging archive*. 2015.
- [29] Simone Perandini et al. “Distribution of solid solitary pulmonary nodules within the lungs on computed tomography: a review of 208 consecutive lesions of biopsy-proven nature”. In: *Polish journal of radiology* 81 (2016), p. 146.
- [30] Awais Mansoor et al. “Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends”. In: *RadioGraphics* 35.4 (2015), pp. 1056–1076.
- [31] Christian Szegedy et al. “Going deeper with convolutions”. In: *Cvpr*. 2015.
- [32] Christian Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2818–2826.
- [33] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. “Spatial transformer networks”. In: *Advances in neural information processing systems*. 2015, pp. 2017–2025.
- [34] Jia Ding et al. “Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2017, pp. 559–567.