

Detection of Lung Nodules on CT Images based on the Convolutional Neural Network with Attention Mechanism

Khai Dinh Lai^{1,2,3}, Thuy Thanh Nguyen⁴ and Thai Hoang Le^{1,2,*}

¹Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam

laidinhkhai@sgu.edu.vn; lhthai@fit.hcmus.edu.vn

²Vietnam National University, Ho Chi Minh City, Vietnam

laidinhkhai@sgu.edu.vn; lhthai@fit.hcmus.edu.vn

³Faculty of Information Technology, Saigon University, Ho Chi Minh City, Vietnam

laidinhkhai@sgu.edu.vn

⁴VNU University of Technology, Ha Noi City, Vietnam

nguyenthanhthuy@vnu.edu.vn

*Correspondence: lhthai@fit.hcmus.edu.vn

Received: 29th January 2021; Accepted: 17th March 2021; Published: 1st April 2020

Abstract: The development of Computer-aided diagnosis (CAD) systems for automatic lung nodule detection through thoracic computed tomography (CT) scans has been an active area of research in recent years. Lung Nodule Analysis 2016 (LUNA16 challenge) encourages researchers to suggest a variety of successful nodule detection algorithms based on two key stages (1) candidates detection, (2) false-positive reduction. In the scope of this paper, a new convolutional neural network (CNN) architecture is proposed to efficiently solve the second challenge of LUNA16. Specifically, we find that typical CNN models pay little attention to the characteristics of input data, in order to address this constraint, we apply the attention-mechanism: propose a technique to attach Squeeze and Excitation-Block (SE-Block) after each convolution layer of CNN to emphasize important feature maps related to the characteristics of the input image - forming Attention sub-Convnet. The new CNN architecture is suggested by connecting the Attention sub-Convnets. In addition, we also analyze the selection of triplet loss or softmax loss functions to boost the rating performance of the proposed CNN. From the study, this is agreed to select softmax loss during the CNN training phase and triplet loss for the testing phase. Our suggested CNN is used to minimize the number of redundant candidates in order to improve the efficiency of false-positive reduction with the LUNA database. The results obtained in comparison to the previous models indicate the feasibility of the proposed model.

Keywords: *Attention convolutional network; triplet loss; nodules detection; false-positive reduction*

1. Introduction

Lung cancer is the leading cause of death from cancer for both men and women. According to the annual statistical report on the number of patients with lung cancer of the American Institute for Cancer Research, there were 2 million new cases in 2018 ¹. The development of disease is silent without symptoms or warning signs, making diagnosis very difficult. As patients reach the final

¹ Lung Cancer Statistic Homepage, <https://www.wcrf.org/dietandcancer/cancer-trends/lung-cancer-statistics>

stages of long-term lung disease, the life will come to an end. Regular cancer screenings for early detection and treatment are needed to reduce mortality.

There are many noninvasive imaging manners for detecting and diagnosing lung nodules such as Positron emission tomography (PET), computed tomography (CT), low-dose computed tomography (LDCT), and contrast-enhanced computed tomography (CE-CT). However, they are used in different contexts. PET scans are used to differentiate between benign and malignant nodules while CT and LDCT are used as the basis for early detection of nodules.

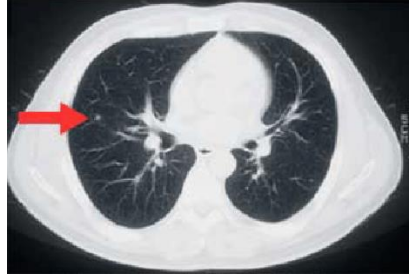


Figure 1. 2D CT scanning slice for an early stage nodule for lung cancer [1]

Computer-aided diagnosis (CAD) can help doctors interpret medical images, making for a more sensitive and accurate cancer diagnosis which is crucial for patients [2]. The aim is to develop a highly accurate CAD system, improve processing speed and reduce cancer screening costs for patients. This machine is capable of obtaining diagnostic images as input and rapidly recognizing patients with or without tumors, as well as being able to distinguish these tumors as benign or malignant. One of the first requirements of the system to satisfy the early stage cancer diagnosis criterion is to detect tiny nodules (< 10 mm in diameter). More complicated is a CT scan packed with ambient tissue, bone and air noises, which need first to be pre-processed in order to make the search for the CAD system successful. The typical CAD systems for lung cancer are made up of the following modules shown in Figure 1

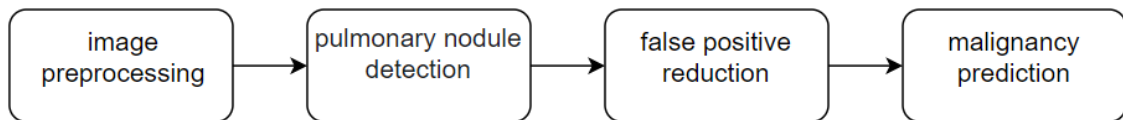


Figure 2. Four phases of a typical CAD system diagnose lung tumor

LUNA16² challenged researchers to focus on pulmonary nodule detection and false-positive reduction to establish a full nodule detection system based on data given by LIDC/IDRI. Pulmonary nodule detection schemes are a candidate detection phase designed to provide simpler representation of lung anatomy, meaning that aspects such as the lung wall and broad airways are omitted, leaving only data with a greater potential to be a nodule candidate. The aim of a false-positive reduction step is to exclude false-positive candidates from the pool of items obtained by the candidate detection phase. The removal mechanism should retain a relatively high sensitivity. Therefore, the false-positive reduction phase should be fitted with a strong discriminating capacity for the separation of nodule and non-nodule objects.

From the LUNA16, we have a dataset of potential tumor candidates. The problem is how to reduce this number of candidates so that only the most potential candidates are achieved (most likely to contain tumors). In order to solve this challenge, it is necessary to create a classifier with the candidate input and to infer that the result is a candidate containing or not a tumor? Once this classification has been done, we will shorten to achieve the number of candidates most likely to be chosen. The classification of the candidate is seen in the figure 3.

To classify nodule candidates into nodules and non-nodules is performed using the following two broad categories of methods: (i) conventional feature-based classifiers and (ii) convolutional neural networks (CNNs). Conventional feature-based classification is performed using feature

² LuNa homepage, <https://luna16.grand-challenge.org/>

extraction and nodule candidate classification techniques. In recent years, Deep Learning has progressed rapidly, especially in image processing problems such as handling facial expressions, analyzing medical images, etc [3]. Many works applied deep learning in medical imaging processing, a large area of this are CT images. So, we would like to present a summary of some related research that used CNN for pulmonary nodule detection in section 2.

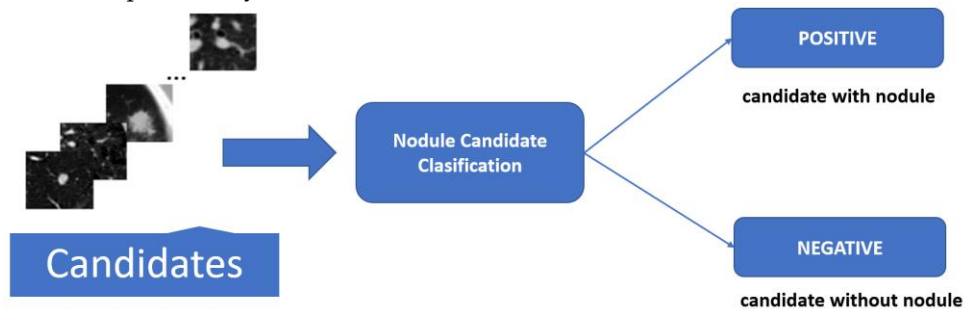


Figure 3. show the architecture of tumor classification system

In recent times, based on the candidate coordinate dataset given by LUNA16, the author³ captured them in 50x50-sized image frames and created a classification model to classify tumor candidates or not. This model has a precision of 89.3 %, recall of 71.2 % and specificity of 98.2 %. However, this model has not exploited the specific characteristics of medical images.

To address the drawbacks, in [4], we suggested the model was defined in Figure 4. This was called the Nodule Candidate Detection in general. The experimental findings were depended on data collected at a scale of 50x50 for each candidate by the author³.

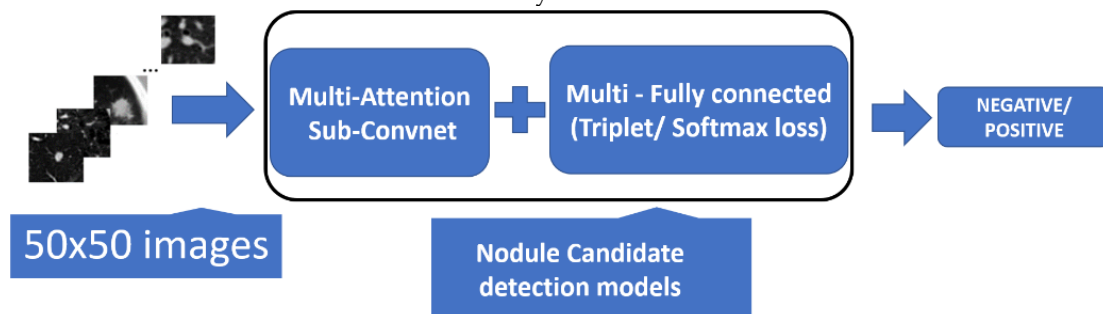


Figure 4. the position of nodule Candidate detection model in identification system

Based on the structure of the model in figure 4, three models were generated as in figures 4,5 and 6 in section 3. In more details, all of these models consisted of many sub-convnets combined using the attention mechanism which took into account the specific characteristics of data. We also considered selecting a softmax or triplet loss for the classification stage. On the same dataset, the models are compared with each other on performance, precision, recall and specificity, then best one will be selected. The final model chosen is compared with the one in the study³. The results show the feasibility of the proposed models. The order of our work is as follows: (1) we built a model with Attention convolutional layers with double triplet functions for both training and validation. So we call it ATT model; (2) possibly the ATT, except the triplet loss was replaced with softmax loss in the classification stage, meaning we've used the softmax loss both for training and validation. This is ASS model; (3) we trained a model with the name is AST that used ASS model as a pre-trained model and validated it with the triplet loss on the validating dataset; (4) we trained the model which was the best of all, on both training and validating subsets.

The following sections of the paper are organized as follows, section 2 describes literature review of paper. Section 3 describes in detail how the architectures of the proposed models are created. In section 4, we explain the step of preparing the data, then we present how to evaluate the models as well as the results obtained through the models. We also discuss and pick the best model for our problem of concern.

³ GitHub Swethasubramanian page, <https://git.io/lf9Og>

2. Literature review

2.1. Lung nodule database

Image data is taken from the Association for Research on Infectious Diseases and Lung Database [(LIDC / IDRI)]⁴. It is obtained from the biggest public lung nodule reference database, comprising 888 CT scans. But actually, initially the dataset consisted of 1018 CT scans. Then, scans with a thickness of more than 3 mm or scans with inconsistent slice spacing are discarded. These scans have been supported as MetaImage (.mhd) files. Each LIDC-IDRI scan is annotated and labelled as: nodule ≥ 3 mm; nodule < 3 mm; nodule (any other pulmonary abnormality) by experienced thoracic radiologists in a two-step reading process: blinded reading phase and unblinded reading phase. We use the reformatted version, LUNA16. This dataset includes CT scans with notes describing the coordinates of the nodule region and ground truth labels [5].

Lung Nodule Analysis 2016 (LUNA16 challenge), which is completely open, concentrate on a large-scale evaluation of autonomous nodule detection algorithms on the LIDC/IDRI dataset. LUNA16 offers data and annotations to every participant to ensure that all participants are equitable while training models and evaluating the performance of algorithms in the same database. LUNA16 challenge encourages the science group to take part in one or two of the following challenge tracks: complete Nodule detection (NDET) and False positive reduction (FPRED) corresponding to phases 2 and 3 of the CAD system.

In the complete nodule detection track, participants are asked to construct a complete nodule detection system. It is known that the input of images is taken from the database given and the output of the system identifies the candidates as nodules.

In the false positive reduction track, the research teams are expected to identify some of the positions in the scans as nodules or non-nodules. In particular, this is a prerequisite to conduct a two-class classification (nodule/non-nodule) problem, and this task encourages research groups to also have experience in classification problems without background in the analysis of medical images.

2.2. False positive reduction systems

The false positive reduction competition attracted a variety of study groups to participate, and the findings were reported on the LUNA16 website. However, LUNA16 decided to stop accepting new submissions in 2018. In this section, we briefly cover some of the impressive outcomes of the False-positive-reduction challenge track over the last few years.

2.2.1. CUMedVis

Dou *et al.* [6] were using multi-level contextual 3D ConvNets to build a framework called CUMedVis in 2016. A framework consisting of three separate 3D ConvNets architectures (Archi-1, Archi-2, Archi-3) was proposed to solve the challenges resulting from differences in nodule size, form, and geometry characteristics.

Table 1. Architectures of three models from [6] with C: convolution, M: max-pooling, FC: fully-connected

Archi-1			Archi-2			Archi-3		
layer	kernel	channel	layer	kernel	channel	layer	kernel	channel
Input	-	1	Input	-	1	Input	-	1
C1	5x5x3	64	C1	5x5x3	64	C1	5x5x3	64
M1	1x1x1	64	M1	2x2x1	64	M1	2x2x2	64
C2	5x5x3	64	C2	5x5x3	64	C2	5x5x3	64
C3	5x5x1	64	C3	5x5x3	64	C3	5x5x3	64
FC1	-	150	FC1	-	150	FC1	-	150
FC2	-	2	FC2	-	2	FC2	-	2
Softmax	-	2	Softmax	-	2	Softmax	-	2

⁴ LIDC-IDRI homepage, <https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI>; Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, *et al.*, "Data from LIDC-IDRI. The Cancer Imaging Archive", 2015, DOI:10.7937/K9/TCIA.2015.LO9QL9SX.

Each network utilizes an input image of a different receptive field to incorporate different levels of contextual information surrounding the pulmonary nodules. In the last stage of this method, three subsystem architectures are fused with a weighted linear combination to achieve the final classification outcome for the candidate.

2.2.2. DIAG_CONVNET

Setio *et al.* [7] proposes a multiview convolutional network-based lung nodule detection system obtained by combining three candidate detectors specifically designed for solid, subsolid, and large nodules. This system is called as DIAG_CONVNET on the luna16 website. The authors extract nine 65×65 patches of 50×50 mm from various views corresponding to different planes of symmetry in cubes for each candidate. These views are processed using a stream of 2D ConvNets composed of 3 convolutional layer and max-pooling layers. The convolutional layers are of turn size: 24 kernels of 5×5 ; 32 kernels of 3×3 ; and 48 kernels of 3×3 . The final detection step is performed using multiple streams of 2-D convolutional networks and a dedicated fusion method. Their method reaches high detection sensitivities of 85.4% and 90.1% at 1 and 4 false positives per scan, respectively. Authors in [7] need high cost to have a dataset which classified subsolid nodules, solid nodules, large subsolid nodules. Besides, the training for each detection is quite time-consuming. On the other hand, the cost of mix-fusion detection and operating the system is high.

2.2.3. 3D DCNN

Ding *et al.* [8] proposed a lung nodule detection system involved the application of a region-based CNN for nodule detection on image slices and employed a 3-D CNN to reduce false-positive per scan. First, a deconvolutionary framework is applied to the Faster Region-based Convolutional Neural Network (Faster R-CNN) for candidate detection on axial slices. A three-dimensional DCNN is then presented for the resulting false-positive reduction. It was evaluated using the Lung Nodule Analysis Challenge (LUNA16) dataset and achieved a high sensitivity (94.4%) with only four FPs/scans. Authors took advantage of VGG 16 and used the regional proposal network to attain high efficiency, but the expense of developing the area proposal network is not low.

2.3. Background

2.3.1. Attention mechanism

In 2017, Jie Hu *et al.* [9] proposed an attention mechanism in order to improve the accuracy of image classification deep networks. The motivation is that inside a feature map with C channels, not all C are crucial to the final decision of a deep network. In which, there are some important ones that should be focused on. Therefore, it needs an attention (self-attention) mechanism so as to emphasize to those essential ones. Authors proposed a block (as ResNet block [10]), called Squeeze-and-Excitation block. This block can be built upon a transformation F_{tr} to map the input feature $X \in R^{H' \times W' \times C'}$ to output feature $U \in R^{H \times W \times C}$. Then, the output feature is squeezed by applying a Global Average Pooling operation to form a vector $z_c = F_{sq}(U) \in R^{1 \times 1 \times C}$. Subsequently, the vector z_c is nonlinearly transformed by a Fully-connected layer, which maps z_c to $s = \sigma(F_{ex}(z_c, W)) \in R^{1 \times 1 \times C}$. In which, $W \in R^{C \times C}$ is the weight of the Fully-connected layer, and σ is the sigmoid function. The final sigmoid function constraints the value range of the vector s within $[0,1]$. Here, s can be seen as the attention gate. More specifically, each element s_i corresponds to a channel U_i . If the value of s_i is close to 1, the channel U_i is more important. After obtaining the attention gate s , the attentioned output feature can be computed as: $\hat{X} = s \times U$. By multiplying the output feature U with the vector s , less crucial channels will be suppressed, while other important ones will be remained.

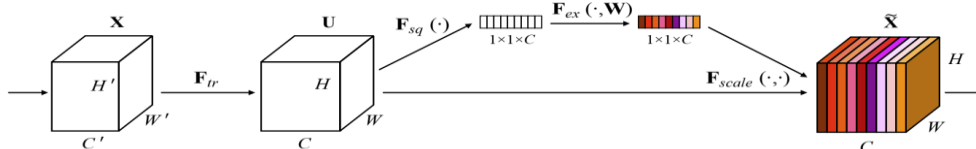


Figure 5. Squeeze-and-Excitation block [9].

2.3.2. Loss function: Triplet loss and Softmax loss

2.3.2.1. Triplet loss

The main purpose of triplet loss [11] is to distinguish identities by minimizing the distance between an anchor and a positive, both of which have the same identity, and maximizing the distance between the anchor and a negative of a different identity. Thus, desirable criteria is

$$L = \sum_i^N \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \right]_+, \quad (1)$$

Where x_i^a is an anchor image, x_i^p is a positive image and x_i^n is a negative image. α is a margin that is made compulsory between a pair of positive and negative. T is the set of all possible triplets in the training set. The loss is computed with the formular (1), where N is number of candidates in training set.

2.3.2.2. Softmax loss

Softmax loss takes a vector of K real numbers elements as its input, which normalizes it into a probability distribution made up of K probabilities equal to the exponential origin values. That is, prior to applying softmax, some vector components could be random numbers; and might not sum to 1; but after applying softmax, each component will be in the interval (0,1) and sum to 1. Softmax loss with a cross-entropy loss that has the form:

$$L_i = -\log \left(\frac{e^{f_i}}{\sum_j e^{f_j}} \right) \quad (2)$$

Where f_j mean the j-th element of the vector of class scores f .

3. Proposal methods

3.1. Attention sub-Convnet

As described section 2, the "attention" approach was used to improve the precision of the image classification in deep-networks. The channels were chosen in the feature maps which were useful. We developed a small convolution network called Attention sub-Convnet. Its architecture is mentioned in table 2. Specifically, a sub-convnet is made up of components. The first component is the convolutional layer that accepts HxW-sized images with C' initial channels. Then use convolutional multiplication by a kernel of KxK size to get the output of feature maps of size HxW with C channels. Second, the pooling layer conducts average pooling for HxW-sized feature maps, and produces C 1x1-size feature maps for C channels. The scalars channels together form a vector z of length C. Third, The fully connected layer on vector z generates C neurons of output. In the next step, the sigmoid function is used to normalize the components of the vector z to the corresponding values of the domain value [0; 1]. Finally, the sub-convnet performs the normalized vector z multiplication with feature maps of size HxWxC where C is the number of channels.

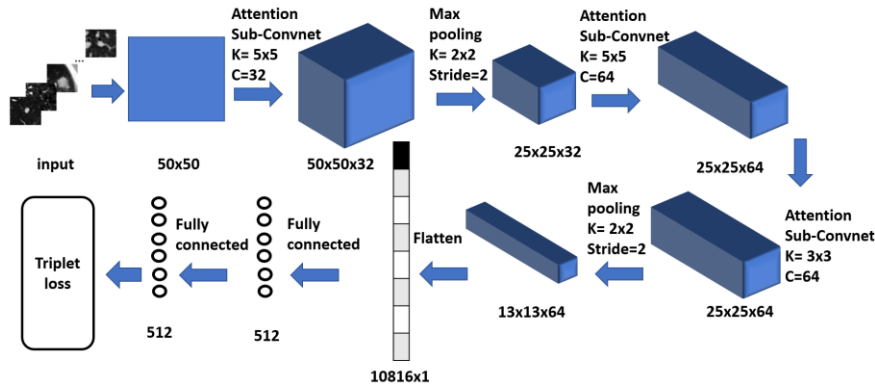
Table 2. An overview of the Attention sub-Convnet architecture included: layers, input size, output size and parameters

Layer	Input size	Output size	Parameter
Input	HxWxC'		
Convolution	HxWxC'	HxWxC	Kernel KxK
Average Pooling	HxWxC	1x1xC	Kernel HxW
Fully connected	1x1xC	1x1xC	
Sigmoid	1x1xC	1x1xC	
Multiply	HxWxC, 1x1xC	HxWxC	
Output		HxWxC	

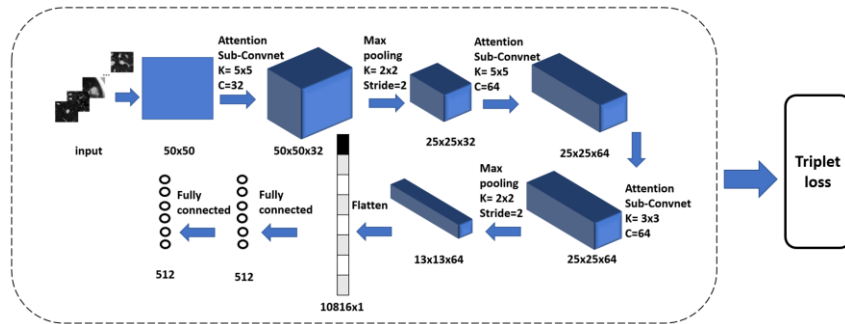
3.2. ATT model

In this scenario, we used three Attention SubConvnets. Concurrently, for both training and validating, we take triplet loss in the classification stage. Attention sub-Convnet is based on the

standard frame of the model referred to in section 2.1 with specific values: H and W are equal to 50 pixels in height and width of the image. The initial channel number 'C' is 32. We use a 5x5 kernel with a 2-stroke to perform a convolution multiplication on each image. Since using Attention sub-Convnets for the first time, features maps size 50x50x32 can pass through the max pooling layer with a 2x2 kernel to transform to feature maps size 25x25x32. The second Attention sub-Convnet is applies with C equal to 64 and the kernel having a size of 5x5 to receive feature maps with size 25x25x64. The feature maps continue to move through the third Attention sub-Convnet with the same number of channels, except this time the kernel is 3x3 in size. Then feature maps are transformed by the max-pooling layer to reduce the size to 13x13x64 before being flattened out to a vector of 10816 dimensions. To eventually obtain the 512 neurons, we use double fully connected layers. (see in figure 6)



(a) Training phase of ATT



(b) Validating phase of ATT

Figure 6. The architectures of ATT

3.3. ASS model

We use Attention sub-Convnets for convolution layers, as well as ATT. However, softmax loss is used as an alternative to triplet losses during the classification phase in all three training, validating and testing processes. (see in figure 7). In this architecture, the output consists of only 2 neurals representing the probability distribution of the two classes: positive (nodule) and non-nodule layers (negative).

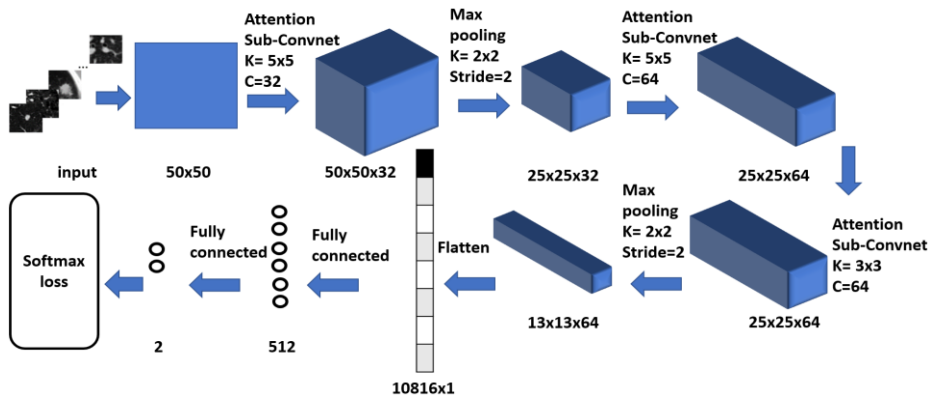


Figure 7. The architectures of ASS

3.4. AST model

ASS model was used as pre-trained model. we used the triplet loss to adjust the parameters of the model. Finally, the triplet loss was used once again for testing. (see in figure 8)

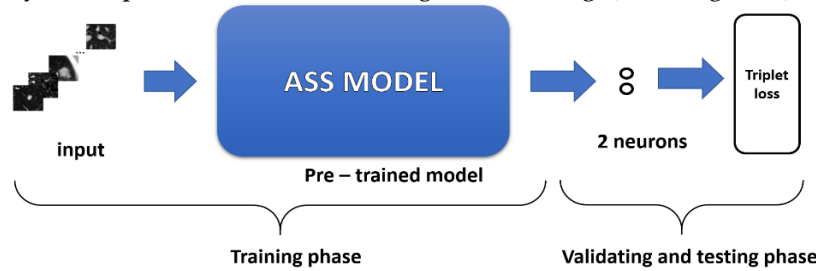


Figure 8. The architectures of AST

4. Experimental results

4.1. Dataset

The main data is from the Association for Research on Infectious Diseases and Lung Database [(LIDC / IDRI)]. But we use the reformatted version, LUNA16. This dataset includes 888 CT scans with notes describing the coordinates of the nodule region and whether or not the image is labeled as a nodule. Each CT scan has a size of $512 \times 512 \times n$, where n is the number of scans, there are about 200 images in each CT scan.

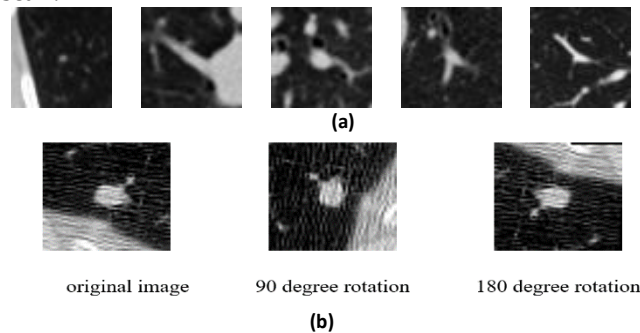


Figure 9. (a) the 50x50 pixel images (b) original image and its rotations

There are 551065 annotations totaling. 1351 are labeled as a positive in all these notes, often known as a nodule, while the rest are labeled as negative, not a nodule. The author³ cropped the images around the coordinates provided in the captions for the new images were created with a size of 50×50 pixels in gray scale to train, validate and test the CNN model. However, the data is out of balance because the number of images containing the tumor is too small for the entire data. The solution to this problem was that the author³ made a rotation of 90 degrees and 180 degrees to create more images with the tumor (see in figure 7). The last dataset that we used in this study includes 8106 images with 50×50 pixel size divided into 3 directories, in particular:

The training folder includes 5187 images, with 845 positive labels and 4342 negative labels. The validating folder includes 1297 images with 224 positive labels and 1073 negative labels. The testing folder contains 1622 images, with 282 positive labels and 1340 negative labels. Therefore, the ratio for the two classes is 20:80.

4.2. Environment for experiments

Models were implemented in the same environment, in particular the same training, evaluation and testing datasets, were calculated on the same computer system. Initializing a set of weights based on a Gaussian distribution with mean of zero, std (standard deviation) as 0.02, bias as zero. We chose adam [12], an algorithm for gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. This method requires little memory, is simple to implement, has computational efficiency, is invariant to diagonal rescaling of the gradients. We used Adam with learning rate is $1e-4$ and weight decay is $1e-6$. For each case, we performed about 50,000

steps corresponding to 50 epochs with a batch size of 8 images. We implemented early stopping technique for our convolutional neural networks. This helps us not only to avoid spending time training after the performance has converged but also help avoid overfitting.

4.3. Evaluation

We used the Holdout method to evaluate models. To get objective results, we splitted the data into three independent directories with ratio between training set, validating set and testing set, respectively 64%, 16% and 20%. We performed as follows on each model sequentially, with four different steps. Firstly, to use data from training to identify parameters and create models. Second, to use cross-validation method to determine the accuracy of the models. If the validation accuracy is poor, then the back-propagation method should change it. After the final model has been obtained, an accuracy test will be conducted with the testing data.

Our task is a binary classification problem that mean we only get results in two cases: positive and negative. The result is positive if the image contains tumor and vice versa, the result is negative if the image has no tumor, we determined the results of the model testing process were split into four groups in confusion matrix. They were true positive (TP) , false positive (FP), true negative (TN), and false negative (FN), respectively. True positive: the image containing the tumor is classified as positive; false positive: the image without the tumor is classified positive; true negative: the image without the tumor is classified negative; false negative: the image containing the tumor is classified negative. The measurements we used for our comparative analyses were precision, recall, specificity, AUC, loss of validation [13].

The specificity is that the measurement indicates the proportion of negatives that are correctly defined. Properly defined as the percentage of segmented slices without the nodule cancer.

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \quad (3)$$

4.4. Experimental results

After implementing these models in the section 3, the results were obtained visually in the diagrams and matching tables (see in figure 10 and table 3,4). Observing the results in the tables, it was clear that for all the comparison measures ATT has yielded worse results than ASS and AST. We have some clarity why at the section 4.5. Experiments showed that AST, the proposed softmax network trained with a combination of multi-Attention sub-convnets with double fully connected and validated by triplet loss gave us the best result. Therefore, we decided to re-train the AST model, but this time on both the training and validating datasets. We calculated the area under the ROC- in the final model, scoring as high as 0.9923 (see in figure 11). The FP value dramatically decreased whereas the TP and TN ratio deals better. We matched our results with the results in the study³, as they are performed in the same datasets and comparison results are shown in Table 5.

Table 3. The Confusion matrix determined specific values for TP, FP, TN and FN.

Model	TP	FN	FP	TN
ATT	200	82	59	1281
ASS	259	23	49	1291
AST	261	21	56	1284

Table 4. Displays results for Precision, Recall, Specificity of models

Model	Precision	Recall	Specificity
ATT	0.710	0.773	0.933
ASS	0.917	0.840	0.984
AST	0.924	0.823	0.986

Table 5. The comparison between the method in the study³ with our method

Model	Precision	Recall	Specificity
method ³	0.893	0.712	0.982
Our method	0,950	0,864	0,989

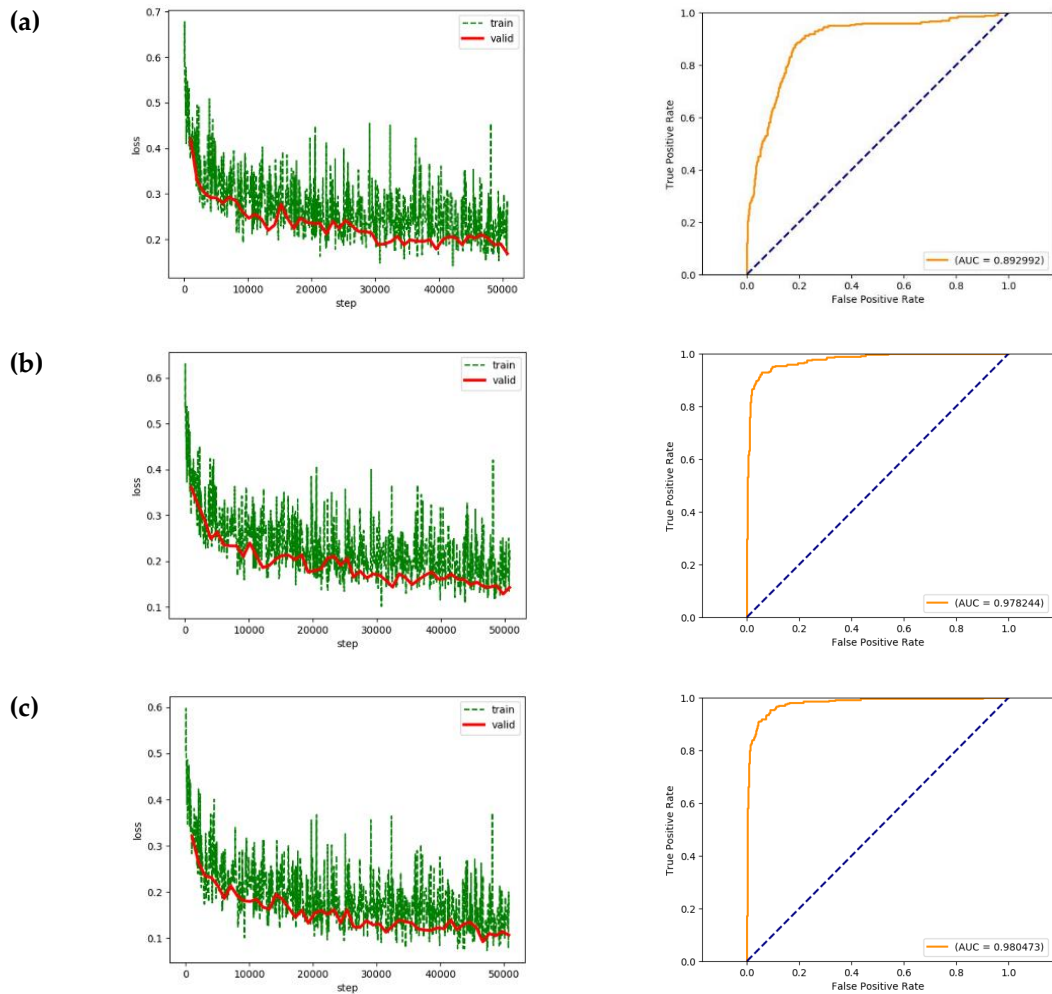


Figure 10. These models were stopped at step 50000 (a) ATT, loss=0.17; AUC \approx 0.892, (b) ASS, loss=0.15; AUC \approx 0.978, (c) AST, loss = 0.12; AUC \approx 0.980.

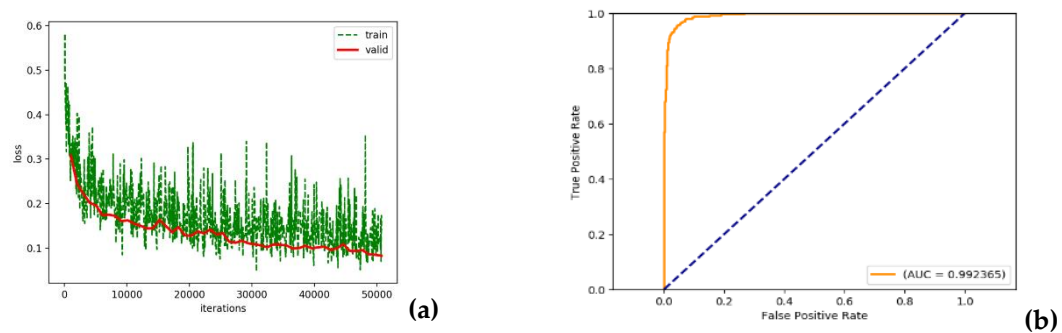


Figure 11. The diagrams displays the results of the final model with (a) loss = 0.08; (b) AUC \approx 0.992

4.5. Discussion

The aim of the training phase is to change the parameters to achieve the optimational function according to the training data and to be highly efficient when tested. For classification point, ATT used the triplet loss by minimizing the distance between it and the co-label samples and maximizing the distance between it and other samples. If the samples with the same label are identical in form, the evaluation of the model by a triplet loss is simple and highly efficient. Unfortunately, the peculiarities of the nodule dataset are very complex, the samples are not the same. This variation appears in images of both nodule and non-nodule. That is, while two samples containing the same

label are positive, their form is quite different. In other words, the distance between co-label samples is not even short. Therefore, when conducting training by using a triplet loss may fall into the case of overfitting, the explanation is that the triplet attempts to change the parameters to match the training dataset the best. In fact, the nodule images are so complicated as mentioned above, so that the system can easily make mistakes during the test process. Additionally, the data is so complex that target value is hard to reach, which is why ATT has the highest loss in the models. Nevertheless, the use of the triplet was successful in the testing phase, as the two classes were now explicitly partitioned. Softmax is flexible in weight adjustment as long as the input is provided into classes without depending on the data characteristics. The ability to prevent overfitting by softmax loss is evaluated in the papers [14]. In specific applications, a suitable loss will be chosen in order to obtain the highest performance.

5. Conclusions

In this paper, we suggest a new approach to solve one of the two problems of lung detection, in particular it is the false-positive reduction challenge. we are interested in classifying whether a candidate is a nodule or not as a nodule. We have proposed the general model include many Attention Sub-Convnets. From the general model, we generated 3 models ATT, ASS, AST using Attention sub-convnet. The choice of softmax loss or triplet loss has demonstrated the effectiveness of each model through both experimental and theoretical assessments. Therefore, it is proposed to select the most effective AST model for the nodule detection problem on LUNA16. Our final result has a precision of 95%, recall of 86.4% and specificity of 98.9 %. After classification, the system dramatically decreased non-tumor candidates.

Note

This manuscript is an extended version of the paper entitled "The Combination of Attention Sub-Convnet And Triplet Loss For Pulmonary Nodule Detection In CT Images", 3rd International Conference on Emerging Technologies in Computing 2020 (iCETiC '20), London Metropolitan University, London, UK [4].

References

- [1] Wafaa Alakwaa, Mohammad Nassef and Amr Badr, "Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN)", *International Journal of Advanced Computer Science and Applications*, Print ISSN: 2158-107X, Online ISSN: 2156-5570, Vol. 8, Issue 8, 2017, DOI: 10.14569/IJACSA.2017.080853.
- [2] S. Ignatious and R. Joseph, "Computer aided lung cancer detection system", *2015 Global Conference on Communication Technologies (GCCT)*, Thuckalay, India, Electronic ISBN: 978-1-4799-8553-1, CD: 978-1-4799-8552-4, pp. 555-558, 2015, DOI: 10.1109/GCCT.2015.7342723.
- [3] D. M. Vo and T. H. Le, "Deep generic features and SVM for facial expression recognition", *3rd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS)*, Danang, Print ISBN: 978-1-5090-2098-0, Electronic ISBN: 978-1-5090-2100-0, pp. 80-84, 31st October 2016, DOI: 10.1109/NICS.2016.7725672, Available: <https://ieeexplore.ieee.org/document/7725672/similar#similar>.
- [4] Lai K.D., Cao T.M., Thai N.H. and Le T.H., "The Combination of Attention Sub-convnet and Triplet Loss for Pulmonary Nodule Detection in CT Images", In *Miraz M.H., Excell P.S., Ware A., Soomro S., Ali M. (eds) Emerging Technologies in Computing. iCETiC 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Print ISSN: 1867-8211, Online ISSN: 1867-822X, pp 227-238, Vol. 332, Published by Springer, Cham, DOI: 10.1007/978-3-030-60036-5_16, Available: https://link.springer.com/chapter/10.1007%2F978-3-030-60036-5_16.
- [5] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas de Bel, Moira S.N. Berens, Cas van den Bogaard *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge", *Medical Image Analysis*, ISSN 1361-8415, pp. 1-13, Vol. 42, 2017, 13rd July 2017, Published by Elsevier B.V, DOI: 10.1016/j.media.2017.06.015, Available: <https://www.sciencedirect.com/science/article/abs/pii/S1361841517301020?via%3Dihub>.
- [6] Dou Q, Chen H, Yu L, Qin J and Heng PA, "Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection", *IEEE transactions on bio-medical engineering*, Print ISSN: 0018-9294,

- Electronic ISSN: 1558-2531, pp. 1558-1567, vol. 64, no. 7, July 2017, DOI: 10.1109/TBME.2016.2613502, Available: <https://ieeexplore.ieee.org/document/7576695>.
- [7] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs *et al.*, "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks", *IEEE Transactions on Medical Imaging*, Print ISSN: 0278-0062, Electronic ISSN: 1558-254X, pp 1160–1169, Vol. 35, Issue: 5, May 2016, Published by the Institute of Electrical and Electronics Engineers (IEEE), DOI: 10.1109/TMI.2016.2536809, Available: <https://ieeexplore.ieee.org/document/7422783>.
- [8] Jia Ding, Aoxue Li, Zhiqiang Hu and Liwei Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks", *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Print ISBN 978-3-319-66178-0, Online ISBN 978-3-319-66179-7, pp. 559–567 (2017), Vol. 10435, 04th September 2017, Published by Springer, Cham, DOI: 10.1007/978-3-319-66179-7_64, Available: https://link.springer.com/chapter/10.1007/978-3-319-66179-7_64.
- [9] J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu, "Squeeze-and-Excitation Networks", *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Print ISBN:978-1-5386-6421-6, Electronic ISBN:978-1-5386-6420-9, pp. 7132-7141, 17th December 2018, Published by the Institute of Electrical and Electronics Engineers (IEEE), DOI: 10.1109/CVPR.2018.00745, Available: <https://ieeexplore.ieee.org/document/8578843>.
- [10] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, Print ISBN:978-1-4673-8852-8, Electronic ISBN:978-1-4673-8851-1, pp. 770-778, 12 December 2016, DOI: 10.1109/CVPR.2016.90, Available: <https://ieeexplore.ieee.org/document/7780459>.
- [11] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, Print ISSN: 1063-6919, Electronic ISSN: 1063-6919, pp. 815-823, 12th June 2015, DOI: 10.1109/CVPR.2015.7298682, Available: <https://ieeexplore.ieee.org/document/7298682>.
- [12] Diederik P. Kingma and Jimmy Ba, "Adam: a method for stochastic optimization", *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. 2015, 13 pages.
- [13] Paula Branco, Luis Torgo and Rita Ribeiro, "A Survey of Predictive Modeling on Imbalanced Domains", *ACM Computing Surveys*, ISSN: 0360-0300, 50 pages, Vol. 49, No. 2, Article 31, Publication date: August 2016, DOI: 10.1145/2907070, Available: <https://dl.acm.org/doi/10.1145/2907070>.
- [14] Xiaobin Li and Weiqiang Wang, "Learning discriminative features via weights-biased softmax loss", *Pattern Recognition*, Vol. 107, 2020, 107405, ISSN 0031-3203, 31st May 2020, DOI: 10.1016/j.patcog.2020.107405, Available: <https://www.sciencedirect.com/science/article/abs/pii/S0031320320302089>.



© 2021 by the author(s). Published by Annals of Emerging Technologies in Computing (AETiC), under the terms and conditions of the Creative Commons Attribution (CC BY) license which can be accessed at <http://creativecommons.org/licenses/by/4.0>.