Convolutional Neural Network based Malignancy Detection of Pulmonary Nodule on Computer Tomography

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

in Partial Fulfillment of the Requirements

For the degree of Master of Science

in the Department of Electrical and Computer Engineering

University of Saskatchewan

Saskatoon, Saskatchewan

By

Yi Wang

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Electrical and Computer Engineering,
57 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A9
Canada

OR

Dean
College of Graduate and Postdoctoral Studies,
Room 116 Thorvaldson Building, 110 Science Place,
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Without performing biopsy that could lead physical damages to nerves and vessels, Computerized Tomography (CT) is widely used to diagnose the lung cancer due to the high sensitivity of pulmonary nodule detection. However, distinguishing pulmonary nodule in-between malignant and benign is still not an easy task. As the CT scans are mostly in relatively low resolution, it is not easy for radiologists to read the details of the scan image. In the past few years, the continuing rapid growth of CT scan analysis system has generated a pressing need for advanced computational tools to extract useful features to assist the radiologist in reading progress. Computer-aided detection (CAD) systems have been developed to reduce observational oversights by identifying the suspicious features that a radiologist looks for during case review.

Most previous CAD systems rely on low-level non-texture imaging features such as intensity, shape, size or volume of the pulmonary nodules. However, the pulmonary nodules have a wide variety in shapes and sizes, and also the high visual similarities between benign and malignant patterns, so relying on non-texture imaging features is difficult for diagnosis of the nodule types. To overcome the problem of non-texture imaging features, more recent CAD systems adopted the supervised or unsupervised learning scheme to translate the content of the nodules into discriminative features. Such features enable high-level imaging features highly correlated with shape and texture.

Convolutional neural networks (ConvNets), supervised methods related to deep learning, have been improved rapidly in recent years. Due to their great success in computer vision tasks, they are also expected to be helpful in medical imaging. In this thesis, a CAD based on a deep convolutional neural network (ConvNet) is designed and evaluated for malignant pulmonary nodules on computerized tomography. The proposed ConvNet, which is the core component of the proposed CAD system, is trained on the LUNGx challenge database to classify benign and malignant pulmonary nodules on CT. The architecture of the proposed ConvNet consists of 3 convolutional layers with maximum pooling operations and rectified linear units (ReLU) activations, followed by 2 denser layers with full-connectivities, and the architecture is carefully tailored for pulmonary nodule classification by considering the

problems of over-fitting, receptive field, and imbalanced data.

The proposed CAD system achieved the sensitivity of 0.896 and specificity of 8.78 at the optimal cut-off point of the receiver operating characteristic curve (ROC) with the area under the curve (AUC) of 0.920. The testing results showed that the proposed ConvNet achieves 10% higher AUC compared to the state-of-the-art work related to the unsupervised method. By integrating the proposed highly accurate ConvNet, the proposed CAD system also outperformed the other state-of-the-art ConvNets explicitly designed for diagnosis of pulmonary nodules detection or classification.

Acknowledgements

I would first like to express my deep sense of thanks and gratitude to my supervisor Dr. Seokbum Ko of the Department of Electrical and Computer Engineering at the University of Saskatchewan. I was extremely lucky to have a supervisor who always willing to help me in all the time of research and writing of this thesis. Without his patience and knowledge, I would not have been accomplished the work presented in the thesis.

I also would like to thank my lab members, Hao Zhang, Zhexin Jiang, Juan Yepez and Suganthi Venkatachalam who helped me throughout my experiments. In particular, I am grateful to Hao Zhang for providing precious comments on this work.

Last but not the least, I would like to thank my parents for their continuous supports and encouragements throughout the period of writing this thesis.

Contents

List of Tables

List of Figures

List of Abbreviations

AUC	Area Under the Curve
ANN	Artificial Neural Network
CAD	Computer-Aided Detection
CT	Computerized Tomography
ConvNet	Convolutional Neural Network
DNN	Deep Neural Network
DICOM	Digital Imaging and Communication in Medicine
HOG	Histogram of Gradient
kNN	K-nearest Neighbors
LBP	Local Binary Pattern
LD	Linear Discriminant
ILD	Interstitial Lung Disease
PCA	Principle Component Analysis
RBM	Restricted Boltzmann Machine
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristi
ROI	Region of Interest
SIFT	Scale-invariant Feature Transform
SVM	Support Vector Machine
VOI	Volume of Interest

Chapter 1

Introduction

Lung cancer is becoming the most common cancer disease in the world. According to the statistical data published by the American Cancer Society [1], lung cancer is the leading cause of cancer death. Early detection and diagnosis is a meaningful way to increase the survival rate of cancer. In the field of lung cancer diagnosis, a patient who is suspected of having lung cancer is recommended to perform a chest X-ray at first because of low cost and simplicity to take. Lung cancer comes from the pulmonary nodule described as masses of abnormal tissue inside the lung. When the pulmonary nodule projects into X-ray image, the mass have different areas of density compared to the healthy tissues. By noticing the changing of the area of the density, the radiologist can find the location of lung cancer. However, a chest X-ray is a 2-dimensional projection of the lung, and the size of the pulmonary nodule is usually less than 10 mm. Therefore, it is difficult for the radiologist to confirm the types of the nodule (i.e., benign or malignant) and the exact nodule location to perform a further medical treatment. If the radiologist notices any suspicious finding in the chest X-ray, a further computerized tomography examination is suggested to finalize the diagnosis.

1.1 Pulmonary Nodule on Computer Tomography

Computerized Tomography (CT) scan, due to its high sensitivity to detect pulmonary nodules, is widely used by radiologists. It shows a higher detection rate compared to other chest radiography methods and is thus helpful in reducing the lung cancer mortality [59]. To overcome the problem of the chest X-ray, CT scanner analyzes the patient's lung by performing many X-ray measurements with various angles and produces cross-sectional images of the lung. By combining the scanned CT images, the radiologist reviews the patient's lung in

3-dimensional space. In contrast of X-ray image, CT images have a sophisticated metric (as referring to Hounsfield Units) to help radiologist distinguish patterns in-between abnormal and normal tissues.

In order to perform a completed CT scan, the CT scanner slowly slides along the axial axis from the patient's head to the feet. As indicated in Figure 1.1(a), one CT scan generates a series of cross-sectional images where start from the top to the bottom of the lung. Each acquired cross-sectional image from CT scanner is called a slice. Figure 1.1(b) and Figure 1.1(c) show two different slices acquired at different axial positions from a 64-year-old female patient, who has lung cancer, respectively.

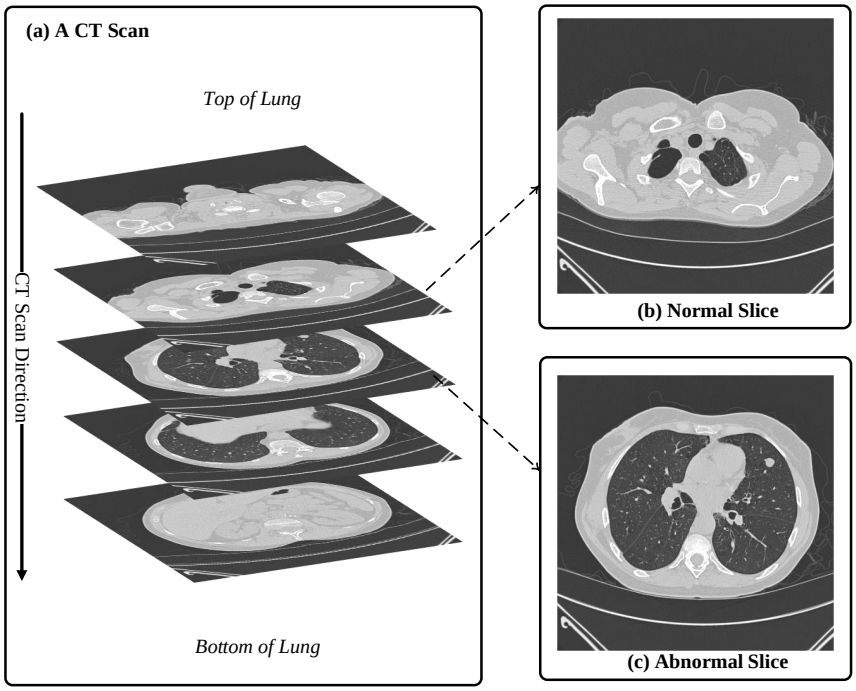


Figure 1.1: A lung CT scan from a 64-year-old female [26]. (a) direction of acquiring slices. (b) one normal slice. (c) one abnormal slice which contains a malignant nodule on the right top of the right lobe.

1.2 Computerized diagnosis of pulmonary nodule

In practice, although the CT scans have high sensitivity in the pulmonary nodule detection [60], it is still not easy for a radiologist to determine whether the nodule belongs to benign or

malignant. When the amount of patient cases is significant, it takes time for the radiologist to make the diagnosis. The seminal National Lung Screening Trial [60] reported mortality from lung cancer was reduced 20% by screening low-dose CT scans. However, the rate of positive low-dose CT scans was 24.2%, while a total of 96.4% of positive screenings showed a false positive. According to [60], a false positive diagnosis will lead to unnecessary follow up medical examinations which will delay the diagnosis while increased radiation exposure causes further damage to the patient.

Computer-Aided Detection (CAD) systems have been developed to reduce observational oversights by identifying suspicious features that a radiologist looks for during case review. Because the CAD system improves the accuracy of diagnosis by radiologists, the CAD system becomes a good choice for preliminary diagnosis [27, 44, 45, 11, 18, 4, 50, 31, 22, 69, 46, 2, 41, 25].

A typical CAD system for a lung CT scan consists of two stages: 1) nodule detection and 2) lung disease classification. The first stage refers to marked areas of concern which highlight suspicious nodule candidates. To generate marks, general approaches are applied through image processing methodologies such as double threshold metrics or morphology operations [27],[44]. However, marks generated by CAD systems generally have high sensitivities and false positive rates [11]. Therefore, the second stage aims to reduce false positive rates for marked areas of concern, while providing the capability of clinical diagnostic decision making. Typically, features from marked areas of concern are extracted, and machine learning schemes are used for nodule classification [27, 44, 45]. Figure 1.2 illustrates a completed CAD system to help clinical radiologist for diagnosis of pulmonary nodules.

1.3 Motivation

A computer-aided diagnosis tool related to malignancy grading of pulmonary nodule belongs to the second stage of the CAD system and focuses on reducing the false positive rate by distinguishing the nodule types in-between benign and malignant cases. Pulmonary nodules have a wide variety of shapes and sizes, as well as high visual similarities between benign and malignant patterns, so analyzing low-level non-textual characteristics such as shapes or

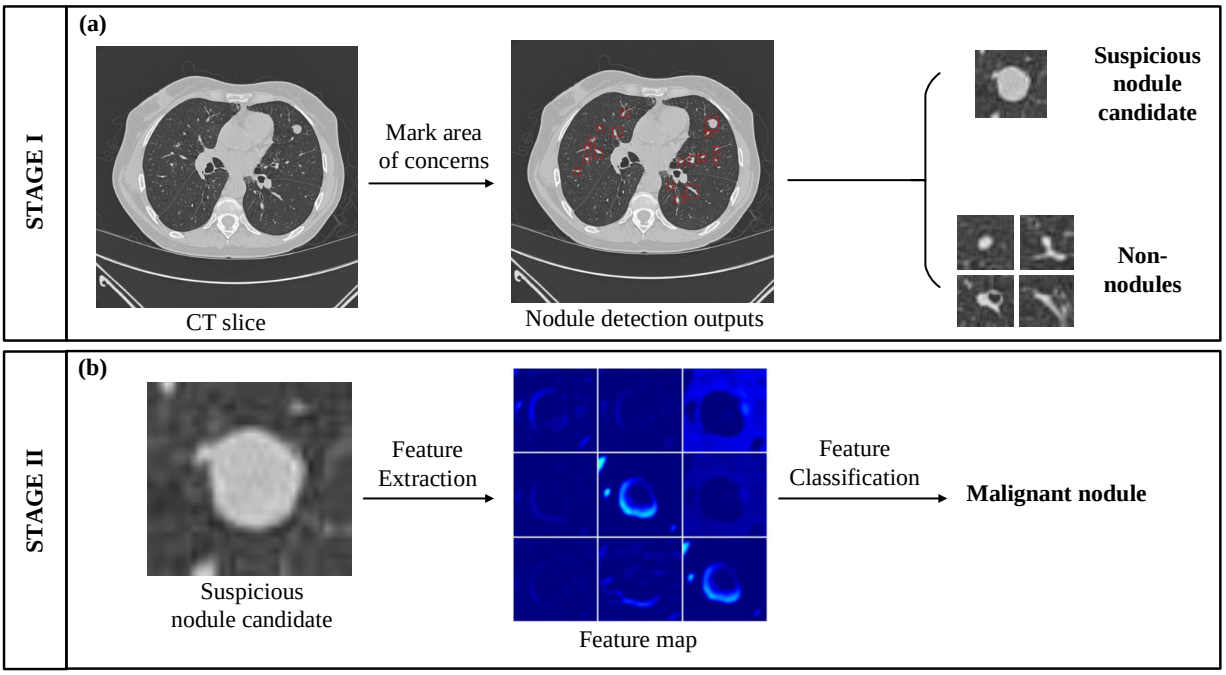


Figure 1.2: A typical CAD system to detect malignant pulmonary nodule from CT scan. (a) overview of stage 1 on nodule detection. (a) overview of stage 2 on lung disease classification.

sizes of the nodules always fail to provide a promising diagnostic performance. Early developed CAD system for pulmonary nodule detection applied conventional imaging processing methods to describe the low-level features of the pulmonary nodule in the higher dimensional spaces [18, 4, 50, 31]. However, such features had limited generalizations when new nodule patterns appeared due to low discriminative power [47].

In recent years, deep learning has achieved great success in image classification, objective detection, image segmentation, and natural language processing. In many of these fields, deep learning can achieve near human performance [53]. Convolutional neural network (ConvNet), the most popular deep learning architecture to perform image classification, has the ability to extract high-level discriminative features. In addition to clinical usage for pulmonary nodule detection, ConvNet uses the patch-based raw image without any additional information such as nodule segmentation or volume of the nodule, and the ConvNet is trained automatically end-to-end in a supervised manner without any additional feature extractor or classifier. Thus, ConvNet is expected to be helpful in improving the performance of CAD systems. Many research works have been done to utilize deep learning in medical field [36, 39, 13, 55,

37, 68] to achieve higher accuracy of diagnosis.

1.4 Aim of Thesis

This thesis aims to design a CAD system based on a ConvNet for diagnosis of the malignant pulmonary nodule on CT scan. The designed ConvNet is able to distinguish pulmonary nodules in-between benign or malignant through raw input image without the need of additional segmentation. To achieve the aim of this thesis, the designed ConvNet is carefully tailored for pulmonary nodule classification task by: 1) exploring optimal hyper-parameters including filter size, receptive field, and optimizer. 2) investigating the performance impact of the imbalanced training dataset. 3) comparing the classification performance with the state-of-the-art work without interference of ConvNet [46] and other related solutions [36, 37, 55, 68, 58].

1.5 Thesis Outline

The rest of the thesis is organized as follows: Chapter 2 introduces the LUNGx Challenge public database used to train and evaluate the proposed CAD system, followed by the theories of the ConvNet including layer operations and training optimization. Chapter 3 presents the current state-of-the-art works for malignancy detection of the pulmonary nodule on CT scan. Chapter 4 introduces the methodologies of the proposed CAD system for pulmonary nodule detection. Chapter 5 presents requirements to set up the experiments and the evaluation comparisons of unsupervised learning method and other ConvNet architectures from previous studies, followed by an analysis of imbalanced data problem that is commonly occurred in the medical imaging database. Finally, Chapter 6 concludes the thesis.

Chapter 2

Background

The purpose of this chapter is to introduce the lung CT database used to train and evaluate the proposed ConvNet, followed by relevant theories about the proposed convolutional neural network. The definition of the input database and ConvNet provide a better understanding regarding the implementation of the proposed CAD system.

2.1 LUNGx Challenge Database

LUNGx Challenge database [26] was used for LUNGx Challenge [3], which was a challenge to seek novel CAD systems for classification of pulmonary nodules on diagnostic computerized tomography scans as benign and malignant. During the LUNGx Challenge event, six experienced radiologists attended the event by manually performing nodule malignancy ratings. As reported in [3], the diagnostic area under the curve (AUC) from the six radiologists is ranged from 0.70 to 0.85 with a mean AUC of 0.79, and three of the six radiologists have statically better performance on malignancy detection compared to the submitted CAD systems during the LUNGx challenge event. Although LUNGx challenge was ended in 2016, malignancy detection of the pulmonary nodule on LUNGx database is still challenging.

The characteristics of the pulmonary nodules in LUNGx Challenge database is shown in Table. 2.1. The LUNGx Challenge database collect 41 malignant nodules, and the size of the malignant nodule is in the range between 5.7 mm and 45.0 mm with the average size of 18.6 mm. In addition to nodule solidity, 34 malignant nodules are categorized into the group of the solid nodules which its Hounsfield Units (HU) on CT are above -450. 2 malignant nodules demonstrate the characteristic of the non-solid nodule with lower HU intensity which has a range from -750 to -300. There have 5 malignant nodules which belong to part-solid nodules,

and the part-solid nodule consists of a solid and a non-solid part. On the other hand, in the LUNGx Challenge database, 42 nodules belongs to benign cases. The smallest benign nodule is 4.6 mm, and the largest one has the size of 34.6 mm. Among the entire benign cases, there are 35 solid, 2 non-solid and 5 part-solid nodules.

Table 2.1: Characteristics of the pulmonary nodules in LUNGx Challenge database

	Malignant Nodule	Benign Nodule
Number of nodules	41	42
Nodule size*		
Average nodule size	18.6 mm	15.8 mm
Minimum nodule size	5.7 mm	4.6 mm
Maximum nodule size	45.0 mm	34.6 mm
Nodule solidity*		
Number of non-solid	2	2
Number of part-solid	5	5
Number of solid	34	35

* All the data of nodule sizes and nodule solidities is provided by LUNGx organizers [3], and the nodule size is measured by following the Response Evaluation Criteria in Solid Tumors (RECIST) guidelines [61].

Each CT scan from the LUNGx Challenge database is obtained under 120kV or 140kV tube peak potential energy with tube current in the range from 240 to 500 mA and tube current-exposure time product of 200-325 mA. The CT scans are reconstructed to the digital imaging and communication in medicine (DICOM) format having the spatial size of 512 $\times$ 512 pixels. The DICOM files of each CT scan are reconstructed with no gap and consists of approximated 250 slices with the slice thickness of 1 mm.

The proposed CAD system is trained and evaluated by using LUNGx Challenge database. The database provides 60 test sets of full thoracic coverage CT scans with 10 calibration sets. The purpose of the calibration sets is to train the CAD systems designated for detecting

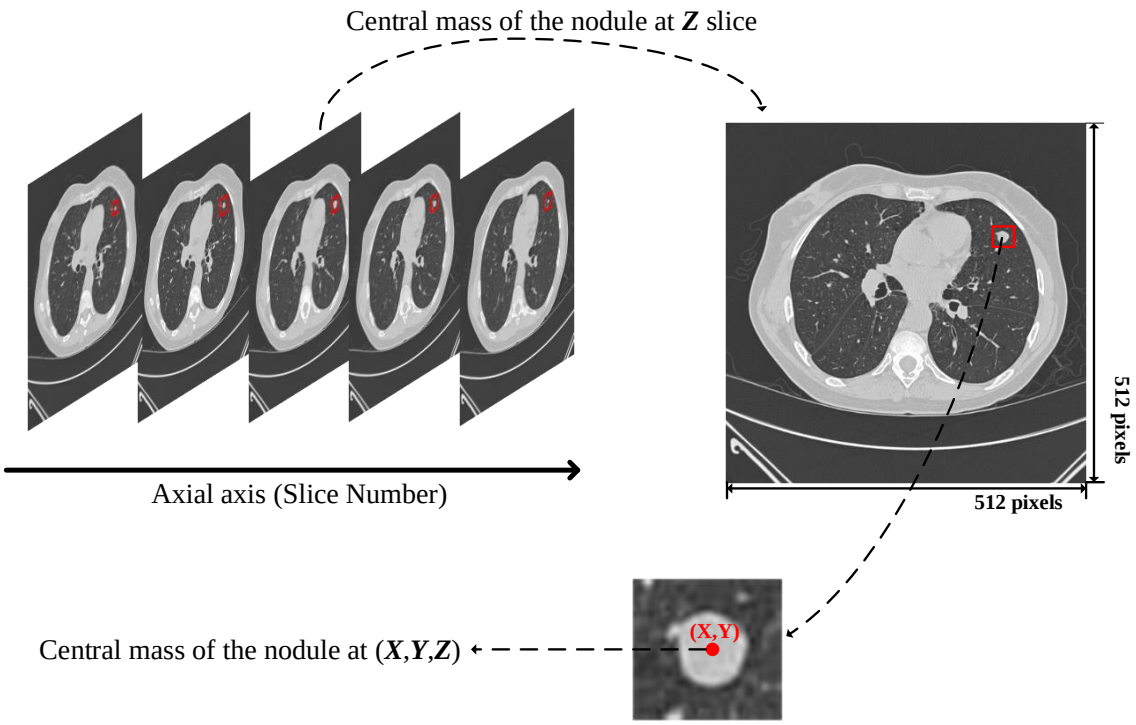


Figure 2.1: Description of the spatial coordinate of a malignant nodule labeled by the six radiologists from LUNGx group. The central coordinate of the malignant nodule in the volume of CT scan is labeled as (x,y,z).

malignancy of pulmonary nodule, and the performances of the trained CAD systems are evaluated by using the 60 testing dataset as the inputs. Six experienced radiologists confirmed the 83 pulmonary nodules which consist of 42 benign nodules and 41 malignant nodules from the entire LUNGx Challenge database. The reference standard for each confirmed nodule was set via spatial coordinates of the approximate center of each nodule and diagnosis decision between benign and malignant. Figure 2.1 demonstrates the process of how the radiologists label each patient case on CT scan. The malignant nodule as shown in Figure 2.1 is labeled by finding the slice where appeared nodule is located at the center mass of the nodule volume. Because each slice is a cross-sectional image of the CT scan in the axial view, the nodule appears in a series of slice images, so the center mass of the nodule volume has an axial coordinate as referring to the slice number. After locating the slice number, the nodule exhibits as a 2-D image in the slice. By locating the center of the nodule in the slice, the completed central mass of the nodule is labeled as (x,y,z). As refer to Figure 2.1, the central

mass of the malignant nodule is located at the z^{th} slice image where the nodule appears at (x,y) in the spatial domain of the z^{th} slice. By taking advantage of the labeled data which has already provided by LUNGx organizers, the nodule samples can be extracted. The detail of extracting the nodule samples is described in section 4.1.1.

2.2 Neural Networks

The artificial neural system brings the strong interest of emulating human brain with a size and complexity comparable model. Artificial neural network (ANN) has originally been interested in modeling neocortex on the biological nervous system but has recently applied to machine learning tasks such as voice recognition, computer vision, robotic control, and data mining. Deep neural network (DNN) is an improved ANNs model and capable of predicting data as accurate as human performance. DNN model consists of more layers than typical ANN; thus, DNN is able to extract high-level abstraction in data [35]. ConvNet is one type of DNN by employing convolution operations to extract image features. Since ConvNet is introduced in early 1990's [34], it has demonstrated impressive image recognition on ImageNet [15] benchmark. In 2015, ResNet [24] has classified 1.2 million high-resolution images from ImageNet database that contains 1000 different classes with error rates of 3.57%.

2.2.1 Artificial Neuron

When we see an object via the eye, the light receptors in our eye send biological current through the optic nerve. Neurons in our brains process the biological current and let us know what the eye sees. Figure 2.2 illustrates a simplest artificial neuron which consists of three input nodes and produces a binary output.

Because the computer has different representations of the visual system than human, computer perceives image by constructing the image patterns with pixels. In Figure 2.2, three image pixels (x_0, x_1 and x_2) through unique synapses paths to generate synapses current. Each synapse path contains a learnable weight (w_0, w_1 and w_2) which interacts multiplicatively with the input pixel. When our eye focuses on one object, our eye automatically ignores other objects where are surround us. Likely, learnable weights control synapses

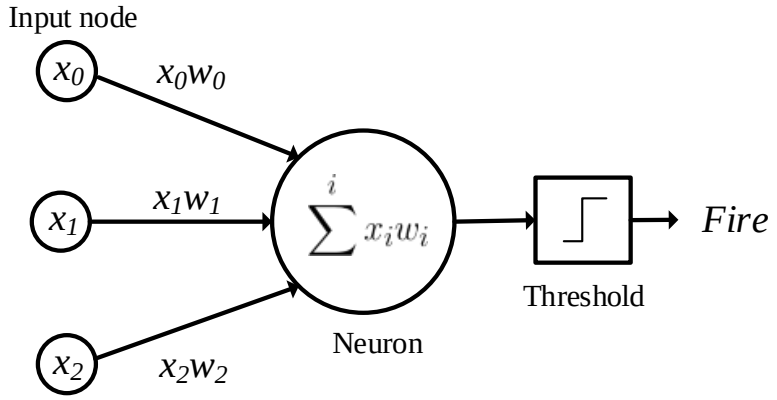


Figure 2.2: A mathematical model of artificial neuron with three input nodes, x_0, x_1 and x_2.

current to related image patterns. For example, by setting negative weights, inhibitory (negative) synapses currents are generated to ignore unrelated image patterns. When input nodes consist of relating image patterns, excitatory (positive) synapses currents are generated. In the final stage of the model, the neuron receives and sums all synapses currents. If summed value exceeds a certain threshold, the neuron is fired. The output of the artificial neuron can be mathematically expressed as:

$$y = f(\sum^{i} x_i w_i + b) \tag{2.1}$$

x_i is the i$^{\text{th}}$ input, w_i is the weight corresponding to i$^{\text{th}}$ input, b is bias and f is activation function. Bias is a constant parameter summed before activation function. The idea of bias is that bias allows the cluster of the classifier to fit the data accurately [19]. For example, an output of the neuron cannot be fired regardless of changing learnable weights if the inputs are all zeros. However, most medical images are stored as gray-scale. In the representation of a gray-scale image, a pixel which has zero intensity corresponds to the black color. Therefore, by adding bias, the neuron can be fired even all input pixels are zeros.

Activation function uses to control the firing rate of the neuron by changing the threshold. Because activation function is a non-linear function, it statically compresses the output of the neuron in a finite range regardless of how large or small the output of the neuron is. Most common activation function are: sigmoid function (Eq. 2.2), tanh function (Eq. 2.3) or

rectified linear unit (Eq. 2.4).

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

$$f(x) = tanh(x) \tag{2.3}$$

$$f(x) = max(x, 0) \tag{2.4}$$

2.2.2 Artificial Neural Networks

A collection of artificial neurons forms the layers of the ANN. The neurons from the previous layer build full-connections to each neuron in the next layer via synapse paths. A typical ANN consists of an input layer in which the size of neurons matches the number of input features, and it has one output layer used to generate scores of different classes. One or two hidden layers are employed in-between the input layer and the output layer to perform feature extraction [65]. The architecture of the DNN is inspired by integrating with deeper and denser hidden layers. Because DNN builds more complex synapse paths compared to ANN, increased capacity of learnable weights enables better learning abilities so that DNN is able to extract more features in higher level abstraction.

Figure 2.3 shows a simple ANN which takes two input features (x_0 and x_1) and consists of three neurons (h_0, h_1 and h_2) in the hidden layer. Such a 3-layer ANN structure is capable of generating a single class score (y). For example, the risk of getting lung canner could depend on family history of lung cancer and smoking history [43]. To evaluate the risk via ANN, the 3-layer ANN extracts the patient histories as two input features and outputs the score for the risk of getting lung cancer.

The ability to extract input features is that neurons are organized into layers, and the hidden layer uses to extract input features. In Figure 2.3, the hidden layer extracts three features regarding both input features. The extracted features are uncorrelated due to none of the synapse path with each other, but synapse paths build full-connections from two input features to each neuron in the hidden layer, so the extract features are correlated to both

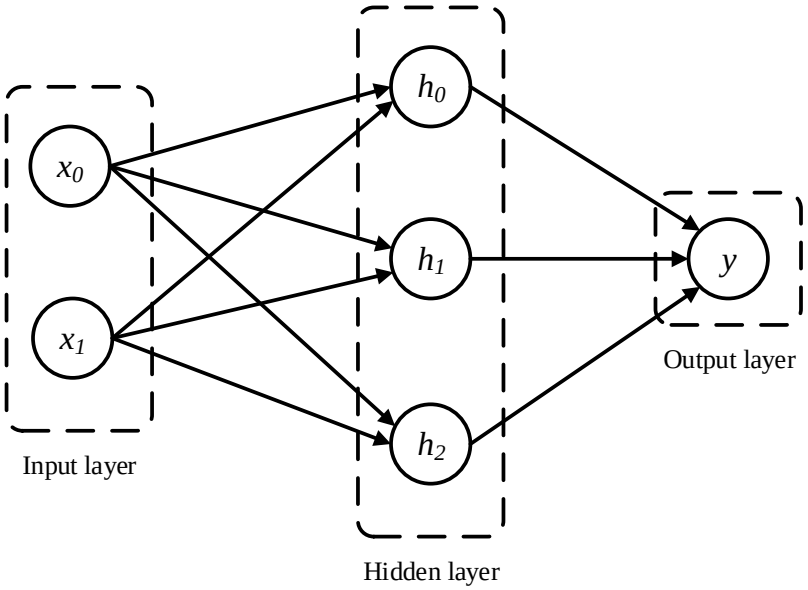


Figure 2.3: A 3-layer artificial neural network which consists of one input layer with two neurons, one hidden layer with three neurons and one output layer with a single neuron.

input features. Then, neurons from the hidden layer pass the extracted features via synapse paths to the output neuron.

The process of propagating input features from the hidden layer (layers) to output layer is defined as feed-forward propagation [19]. The feed-forward propagation for 3-layer ANN can be defined as:

$$\begin{cases} h_i = f(\sum^j x_j w_{ij}^1 + b_i) \\ y = f(\sum^i h_i w_i^2 + b_y) \end{cases} \tag{2.5}$$

where h_i is the i^{th} neuron in the hidden layer, x_j is the j^{th} input neuron, w_{ij}^1 is the weight in the synapse path where is from j^{th} input neuron to i^{th} neuron in the hidden layer, and y is the output neuron. The synapse path, from i^{th} neuron in the hidden layer to the output neuron, has weight of w_i^2. b_i and b_y are the biases for i^{th} neuron in the hidden layer and the output neuron respectively. f refers to activation function.

Deep Neural Networks

By propagating more stages (hidden layers) before reaching the output layer, such an ANN structure can be expressed as a DNN. Empirically, DNN, employs deeper hidden layers com-

pared to ANN, has more representational power for feature extraction and better generalization for feature learning [19]. Figure 2.4 shows increasing the depth and width of each hidden layer yields increased test accuracy, which the experiment data is collected from [19].

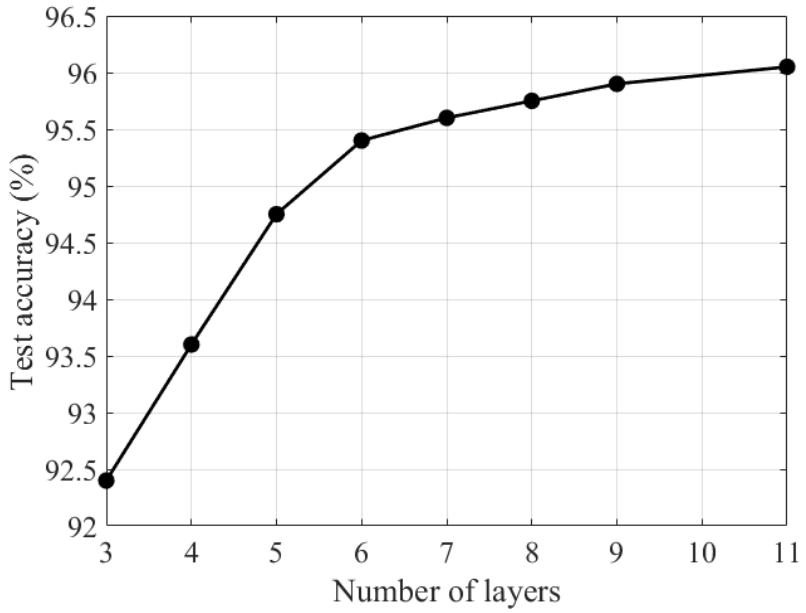


Figure 2.4: Classification performance on the different number of layers [19].

Because deeper hidden layers statically increase the number of synapse paths, the deeper hidden layers extract more features via increased learnable parameters. For example, the 3-layer ANN, showed in Figure 2.3, adopts 4 neurons to extract input features, and it builds 9 synapse paths which consist of 9 learnable weight parameters with additional 4 bias parameters for a total of 13 learnable parameters. By inserting additional two hidden layers with double size of neurons, the new architecture contains 16 neurons which build up 36 more synapse paths and 12 more bias parameters compared to the original 3-layer ANN. The total number of learnable parameters is increased by approximately 7 times. Therefore, DNN outperforms ANN for the statistical reason because adding and expanding hidden layers increase the number of learnable parameters dramatically.

Problems of Deeper Neural Network

When a shallow DNN with few hidden layers is trained, inserting more hidden layers does indeed to have better classification accuracy on the test dataset. However, the classification

performance eventually is saturated when the number of hidden layers exceeds a certain threshold (as referring to Figure 2.4).

As DNN consists of deeper hidden layers, such a model builds a more complicated function (connection) in-between input features and outputs. On the other hand, over-fitting [33] occurs when the deeper layers easily accumulate an extensive collection of learnable parameters. The learnable parameters with high capacity force the model to fit the training data perfectly by tunning the parameters to fit the noises in the training data. As a consequence, the model losses generalization on additional data.

Moreover, DNN builds full connectivities between two adjacent layers. Such a connection scheme forces the network to extract features based on entire input features. Therefore, trained model is lack of generalizing spatial invariance. For instance, trained images which apply rotation transformation result in inferior classification performance because the input features are shuffled via rotation.

2.2.3 Convolutional Neural Network

ConvNet is a particular type of DNN that employs convolution operation to extract features within the local spatial region of the image. Instead of building fully-connectivities in-between hidden layers, ConvNet forces hidden neurons to "see" local information and combine them to form high-level feature. Different features usually appear at various local regions because neighboring pixels are more correlated than faraway pixels. Therefore, ConvNet adopts a collection of local feature detectors to achieve tremendous success in image classification.

ConvNet stacks a sequence of layers to transform input images from pixel values to class scores. General ConvNet has four types of different layers: convolutional layer, activation layer, pooling layer and fully-connected layer.

Convolutional Layer

The convolutional layer adopts learnable filter banks. Each filter convolves full depth of the input volume within fixed-scale of spatial area, depending on filter size along with its width and height. Instead of building full connectivity from whole neurons on input volume to output neuron such as ANNs, the connection of the convolutional layer inspires feature

extraction within local regions of the input volume. On the other hand, a set of filter banks produce overlapping features extracted from a particular local region due to output depth of each filter. Such overlapping features offer various representations of the particular local region. Therefore, the output of the convolutional layer generates a group of feature maps, and each feature map corresponds to different local regions from input volume.

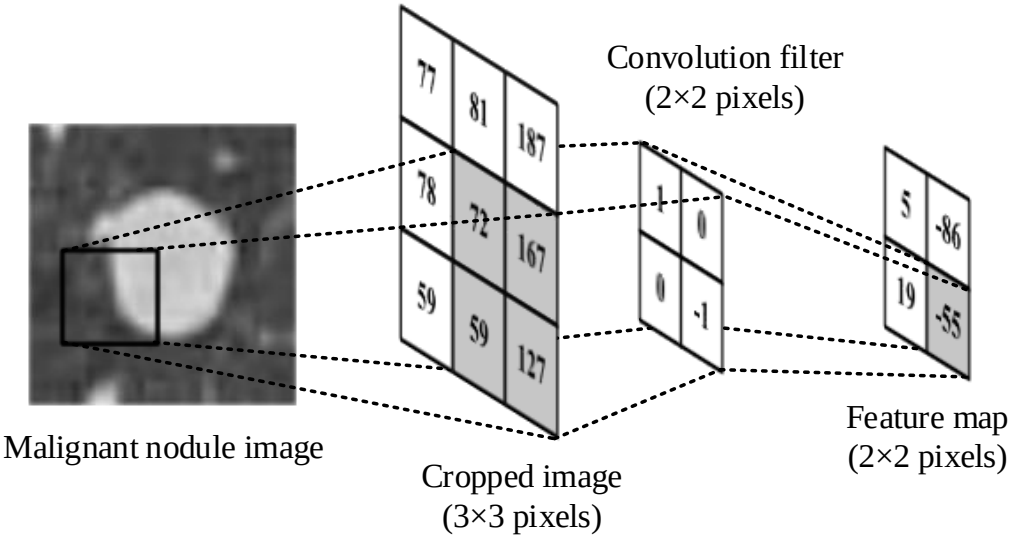


Figure 2.5: An example of convolutional layer to extract features within local region of the cropped image. The convolutional layer consists of one filter of size 2×2 and one output feature map.

Figure 2.5 shows a convolution operation for the cropped image by sliding the convolution filter along the spatial matrix of size 3×3. For a convolutional layer, hidden neurons are organized into each feature map. Unlike the DNN connection scheme (as referring to Figure 2.3), each hidden neuron in the convolutional layer builds the connection to a particular local region of the input features. For example, the feature map, in Figure 2.5, consists of 4 hidden neurons to form a 2×2 matrix of the feature map, and each matrix element in the feature map represents the output of each hidden neuron. By convolving the 2×2 filter without flipping, the highlight area of the input features is extracted to a unique feature which is equal to -55 in the feature map. Hence, a local region of input features with the size of 2×2 corresponds to one feature in the feature map. In general, the output of each hidden neuron can be expressed as Equation 2.6 [19].

$$y_{i,j} = (x \circledast w)_{(i,j)} + b_{i,j} \quad \text{or equivalently} \quad y_{i,j} = \sum_m \sum_n x(i+m, j+n) w(m,n) + b_{i,j} \tag{2.6}$$

In order to extract the feature $y_{i,j}$ at i^{th} row and j^{th} column of the feature map, weights of the filter w with size of $m \times n$ perform dot production on the input features x from i^{th} row and j^{th} column to $(i+m)^{\text{th}}$ row and $(j+n)^{\text{th}}$ column while the result of the dot product adds the bias term, $b_{i,j}$.

Although the feature extraction scheme of the convolutional layer is inspired by the feed-forward propagation in ANN (as referring to Equation 2.5), the learnable parameters are dramatically reduced because a single feature map is generated by convolving the same filter, so each hidden neuron shares the learnable parameters from the same filter. Hence, ConvNet has fewer chances of having over-fitting due to sharing the learnable parameters. Moreover, ConvNet takes advantage of sharing parameters. Instead of relying on a denser hidden layer to extract as many features as possible, the convolutional layer avoids the trained model fit the input noises while learning the invariant relationships within the local spatial region by stacks a series of filters (filter banks).

Activation Layer

The convolutional layer is often followed by an activation layer. Instead of extracting learned features, the activation layer is used to introduce nonlinearity to the ConvNets in order to accelerate convergence during the training phase. Although a large amount of data and high-performance parallel computing solutions spurs ConvNets to integrate more convolutional layers in order to achieve better classification result, long training time is still an important concern. Most modern ConvNets architectures employed activation layer to reduce training time while maintaining similar performance. Typical activation layer produces non-linearity via sigmoid function, tanh function or ReLU.

Pooling Layer

It is common to insert a pooling layer between successive convolutional layers to reduce the spatial dimensionality of each feature map. Full connectivity structure used in ANN requires

a large number of weight parameters to match high definition input scale which often leads to over-fitting. Also, the convolutional layer still has the chance to produce specific redundant information because the local region still builds full connectivity. Therefore, pooling layer retains the essential information and reduces chances of over-fitting. General pooling operations are max-pooling or average-pooling by replacing input values to maximum or average value.

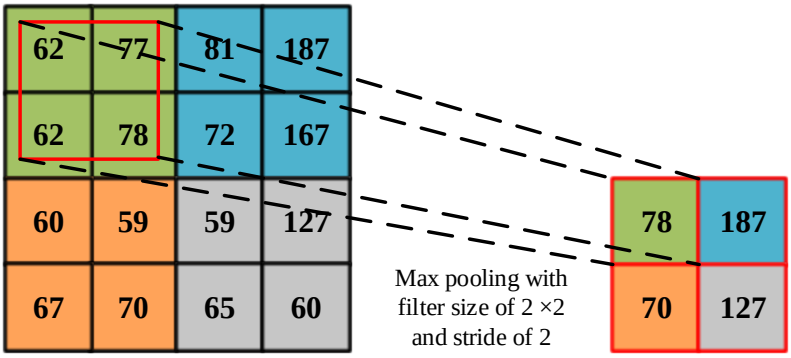


Figure 2.6: An example of max pooling operation. The 2×2 filter with a stride of 2 takes 4 inputs and perform down-sampling by retaining the maximum value.

Figure 2.6 illustrates a max pooling layer down-samples the input feature map. The feature map of size 4×4 is pooled by retaining the maximum value within the pooling filter which has the size of 2×2 and stride of 2. By keeping the maximum value of neighboring features, the feed forward path of the ConvNet has less redundancy, and the path having max activation is also retained.

Fully-connected Layer

Fully-connected layer builds full connections to previous pooling layer. The fully-connected layer has a similar structure as ANN which performs classification. After last fully-connected layer, it is common to normalize the output vectors via SoftMax [7] function. The SoftMax can be defined as:

$$\widehat{y}_i = \frac{e^{y_i}}{\sum^j e^{y_j}} \tag{2.7}$$

where the SoftMax function suppresses the output vector at i^{th} neuron to a value in-between zero and one via an exponential function. Then, by dividing the suppressed output vector at i^{th} neuron with the sum of all suppressed output vectors in the layer, the normalized output vector at i^{th} neuron is generated and referred as $\widehat{y}_i$. The normalized output vector represents the probability distribution among all possible categories [7].

2.2.4 Training Convolutional Neural Network

Training the ConvNet corresponds to optimize the learnable parameters in convolutional layers and fully-connected layers so that the trained model fits the training data. Because the output of ConvNet produces the probabilistic distribution of the classes via SoftMax function, the input features that belong to a particular class most likely has the highest probability by feed-forward propagating the well-trained model. A standard optimization method of the ConvNet is to minimize the losses of the loss function.

Loss Function

In order to evaluate the classification performance in the training phase, a loss function is used to measure the errors between prediction scores and ground truths. For example, a malignant nodule image patch is fed into a ConvNet model, but the output of the ConvNet result in a higher score toward the class of benign nodule. To measure the error of the misclassification, the loss function intuitively produces a large loss regarding the misclassification. By utilizing the learnable parameters, the loss is gradually reduced while the ConvNet gets more confidence to classify the image patch as malignant. Because the malignancy detection of pulmonary nodule relates to binary classification, the loss function for the malignancy detection can be defined as the binary cross-entropy loss function:

$$L_i = -f_{x_i} \cdot \log(\widehat{f}_{x_i}) - (1 - f_{x_i}) \cdot \log(1 - \widehat{f}_{x_i}) \tag{2.8}$$

where the cross-entropy loss for i^{th} input volume which consists of feature x_i is L_i. f_{x_i} refers to the ground truth label, and $\widehat{f}_{x_i}$ means the class score corresponding to the i^{th} input volume.

Optimizers

A loss function quantifies the quality of how well the model fits the training data. In order to achieve low losses, it is crucial to apply an appropriate optimizer to fine-tune learnable parameters so that the loss function is minima. A naive way of fitting the training data is to generate a large collection of random values, and the best model is obtained by trial-and-error. Instead of exhaustive searching the best learnable parameters, the gradient of the loss function provides a direction toward to minimal loss. The gradient of the loss function can be expressed as:

$$\theta = \theta - \mu \cdot \nabla_\theta L(\theta) \tag{2.9}$$

where the loss function is minimized with regards of a parameter, θ for an entire training dataset. By moving θ in a descent direction of the gradient with constant steps, the losses for the entire training dataset is gradually minimized. $\nabla_\theta L(\theta)$ refers to the gradient of the loss function, and μ is the step size as referring to the learning rate.

Because the cross-entropy function is the part of Kullback-Leibler divergence [52] which is a convex function, applying a relatively large learning rate leads to high chances of missing the optimal point of the model. Conversely, a relatively small learning rate leads to a continuous reduction in the loss, but the model takes time to reach the optimal point with regards to convergence.

On the other hand, such a gradient descent algorithm (as referring to Equation 2.9) updates the learnable parameters by calculating the gradients of an entire dataset. When the size of the training data is large, a single update still requires a long time [49].

Stochastic Gradient Descent

Stochastic gradient descent (SGD) updates the gradient of the loss function based on each training sample or batch [8]. In contrast of the conventional gradient descent which performs redundant computations among the entire training dataset, SGD is much faster by performing frequent computations within a limited batch size. Instead of finding the global minimum point of the loss for the entire training dataset, SGD optimizes the model by converging to the local minimum points for each training sample or batch. Because the activation layer

introduces non-linearity, the cross-entropy loss has a shape of a non-convex surface, and the non-convex surface usually consists of several basins as referring to local minimal points. Thus, frequent updates of the local gradients in SGD leads the model to jump to a new local minima with a better loss, and the model eventually converges to the global minma. SGD with regard to a parameter update θ can be defined as:

$$\theta = \theta - \mu \cdot \frac{1}{n} \cdot \nabla_\theta \sum_n L(\theta, x_n, y_n) \tag{2.10}$$

where the loss function L computes the average of the cross-entropy losses from n training samples $\{x_1, ..., x_n\}$ with ground truth labels $\{y_1, ..., y_n\}$.

SGD refers to a batch based gradients, and the gradient of a batch loss mathematically represents as a non-convex surface so that the SGD contentiously keep overshooting (SGD fluctuation [49]). It is common to apply a step-down learning rate to control the direction of the gradient. A step-down learning rate applies a relatively large learning rate in the early stage of the training phase, so the speed of the convergence is accelerated. After several training iterations, the learning rate is decreased by order of magnitude. A reduction of the learning rate leads to consistent finding the best local minima which has the lowest loss. In order to have a smooth reduction of the learning rate, an exponential decay algorithm gradually decrease the learning after each epoch, and each epoch refers to the total number of iterations for the model to train the entire training dataset. The exponential decay of the learning rate can be expressed as:

$$\mu = \mu \cdot e^{-\gamma t} \tag{2.11}$$

where γ is a hyper-parameter that refers to the decay rate. t is the iteration number.

Adaptive Optimizers

The hyper-parameters of the learning rate in SGD has to predefine before the training phase. Hence optimizing the learning rate in SGD is an expensive time cost progress. Adaptive optimizers (e.g., Adam [32], AdaGrad [16] or RMSProp [62]) adaptively adjust the learning rate based on the magnitude of the gradients [49]. Adaptive optimizers dynamically change the learning rate to perform more efficient gradient updates compared to non-adaptive optimizer

such as SGD. However, the adaptive optimizers lead the learning rate to shrink in one of the local minimal points so that the model will never reach the best optimal loss.

Back-propagation

Optimizers such as SGD point out the appropriate direction to achieve minimal loss of the loss function. In order to alter the weights and biases in each layer, a back-propagation provides an efficient way to find the gradients of each layer and recursively computes the gradients from the loss function to the first layer of the ConvNet. The gradient for each layer is estimated via chain rule of the derivation. Algorithm 1 presents the algorithm of back-propagation. The back-propagation is employed after computing the gradient of the loss function through optimizer such as SGD.

Algorithm 1 Back-propagation

```
while current iteration < maximum iterations AND gradient of the loss function < desired criterion do
    for Input sample from training dataset do
        Compute the gradient of the loss function through optimizer
        for each hidden neurons in the output layer do
            Compute gradient as error signal
        end for
        for each layer flow backward from output layer do
            Compute gradient of the layer
            Update bias parameter
            for each hidden neurons in the layer do
                Compute gradient of the hidden neuron
                Update weight parameter
            end for
        end for
    end for
end while
```

Chapter 3

Related works

Pulmonary nodule patterns are generally presented as unique texture inside lung lobe. Before deploying classification schemes, most CAD systems apply feature extraction operations within the marked area of concern. Depending on the input scale of the classifier, the marked area of concern typically is generated as two different modalities: local regions of interest (ROIs) or volumes of interest (VOIs). By sliding fixed-scale of classifier over ROIs or VOIs, the diagnostic decision for pulmonary nodules is made.

3.1 Traditional Feature Extraction Approaches

Early studies demonstrated the usefulness of discriminative features to detect pulmonary nodule over ROIs or VOIs, such as first order gray-level thresholding operations [18, 4], histogram of gray-level [50] and histogram of CT density [31]. Since more modern texture descriptions were implemented, recently proposed CAD systems provided a new perspective of feature extractions in higher dimensional spaces. Such systems employed local binary pattern (LBP) [22], histogram of gradient (HOG) and scale-invariant feature transform (SIFT) [69].

3.2 Unsupervised Learning

The previously presented approaches for feature extractions relied on hand-crafted features which had lack of adaptabilities when new nodule patterns appeared. To achieve promising results against new data, more recent studies agreed with unsupervised feature learning. Such features allowed the systems to discover customized features regardless of input features. Such systems are k-means [13], principal component analysis (PCA) followed by convolution

and pooling operations [46], and restricted Boltzmann machine (RBM) [25] which is an unsupervised learning scheme adopted architecture of the artificial neural network (ANN).

Regardless of extracted features, it is also crucial to choose an appropriate classifier that optimally translates the features into correct classes. General feature-based classifiers applied for pulmonary nodule detection are linear discriminant (LD) [50, 2], k-nearest neighbors (kNN) [45] and support vector machine (SVM) [46, 69]. Also, ANN is not only able to extract learned features, but also able to perform classification due to output vector transformed from the input dimension to one-dimensional space [4, 41].

A successful attempt for malignancy detection of the pulmonary nodule is adopted by unsupervised learning schemes with linear SVM [46]. Under testing within LUNGx database, the previous work [46] has demonstrated superior performance among other hand-crafted features, such as histogram of CT density, LBP on the three orthogonal planes (axial, coronal and sagittal CT images) and LBP with random sampling. The feature extraction is done over VOIs, a 3-D cubic bounding box that contains entire pulmonary nodule and surrounding tissues. During feature extraction, multiple stages were applied as follows: 1) PCA over VOIs. 2) multiple kernels of 3-D convolution. 3) pooling operations over convolved features. In the first stage, the extracted VOIs were converted into 3-D kernels with high correlated unsupervised features by employing PCA. The following stage used 3-D convolution operations over the kernels generated from the previous stage. Higher-level features were extracted on the second stage. The last stage employed the combination of max-pooling and min-pooling which reduced volumes of convolved features into one-dimensional feature vectors while extracted features were maintained via pooling operations. The one-dimensional feature vectors were used to train linear SVM and evaluate malignancy detection of pulmonary nodules over the classifier.

3.3 Supervised Learning

Regarding significant success in large-scale image classification [15], ConvNets outperformed the state-of-the-art in the field of computer vision by taking advantage of supervised learning schemes [33, 54]. To overcome the issue of the unsupervised learning schemes which require

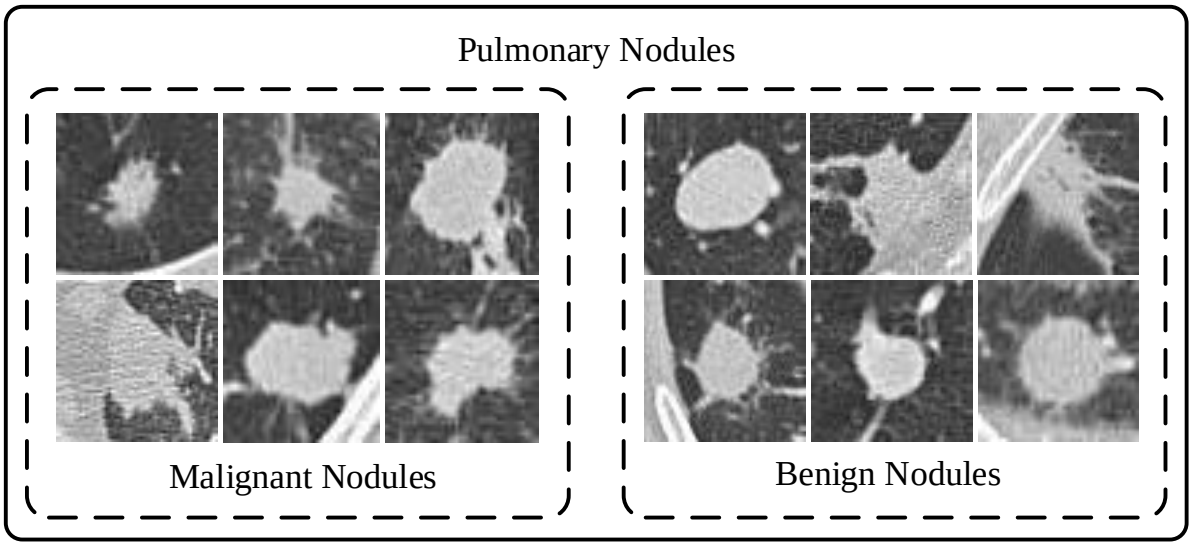


Figure 3.1: Visual representations of the pulmonary nodule patches in axial view. The pulmonary nodules consists of two main categories: benign and malignant nodule. The nodule images are captured under LUNGx database [26].

to transform representations of input features, ConvNets can learn and extract numerous amount of high-level discriminative features from the raw image at multiple levels of abstraction while whole networks are trained in a supervised manner to perform classification due to similar output structure of ANNs. As a consequence, the characteristics of ConvNet are well suited to pulmonary nodule detection when low-level features always resulted in inaccurate prediction and entire classification process should be done in an automatic fashion.

3.3.1 Shallow ConvNet

Early attempts have made to overcome classification of the lung disease patterns via ConvNets. A shallow ConvNet was proposed to classify patch-based images with interstitial lung disease (ILD) [36]. Like pulmonary nodule patterns as illustrated in Figure 3.1, ILD patterns also have high variation texture within the same class while different classes often have similar visual representation. The previous ConvNet [36] showed the remarkable capability of extraction on high-level discriminative features in comparisons of LBP, SIFT, and RBM. However, the previous ConvNet, using single convolutional layer, captures high-level features insufficiently because such shallow structure cannot demonstrate the full potential of feature

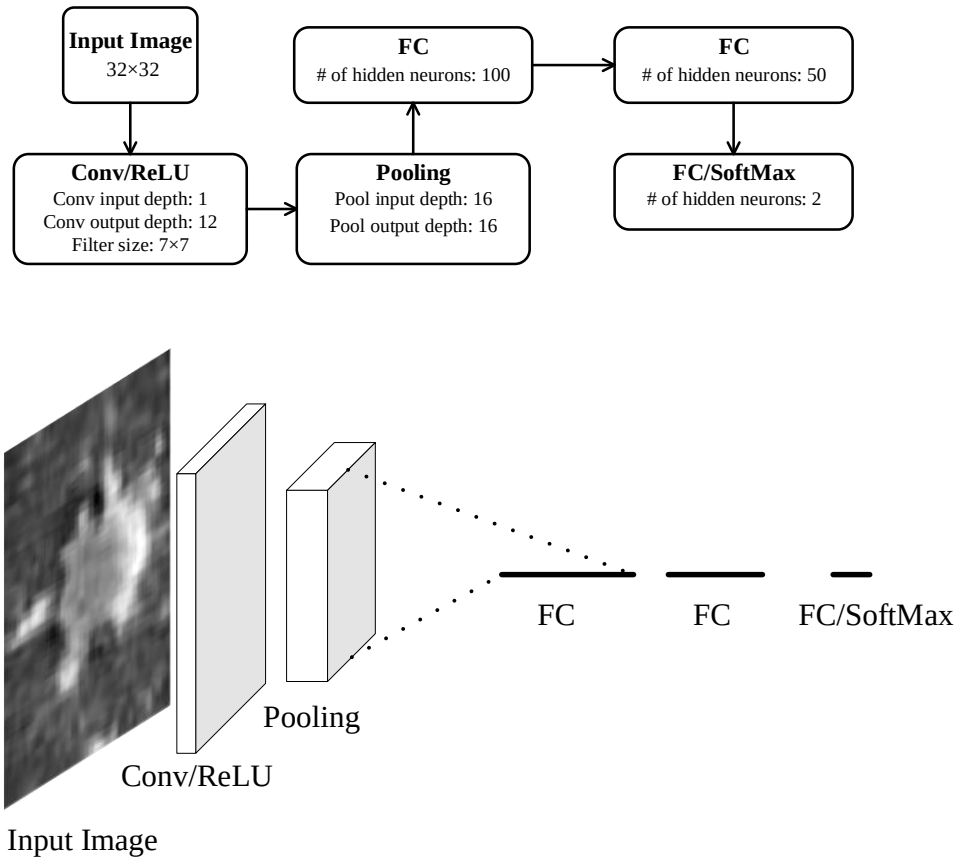


Figure 3.2: Schematic of the Shallow ConvNet [36]. Conv/ReLU, convolutional layer followed by rectified linear unit; pooling, maximum pooling layer; FC, fully-connected layer; FC/SoftMax, fully-connected layer followed by SoftMax.

extraction based on the hierarchy of abstraction which usually requires a deeper structure. The schematic of the Shallow ConvNet is shown in Figure 3.2.

3.3.2 ConvNet with Two Convolutional Layers

In contrast with the shallow ConvNet [36], Song et al. [55] adopted a deep ConvNet structure with two convolutional layers (as shown in Figure 3.3). The ConvNet [55] achieved statically higher precision on malignancy detection of the pulmonary nodule compared to ANN and Stacked Autoencoder [64] which is a neural network based unsupervised learning algorithm. The architecture of ANN and SAE mainly consists of fully-connected layers, so such an architecture with fully-connectives in-between hidden layers always fail to analyze invariant features within the local region, and the pulmonary nodule patterns are usually

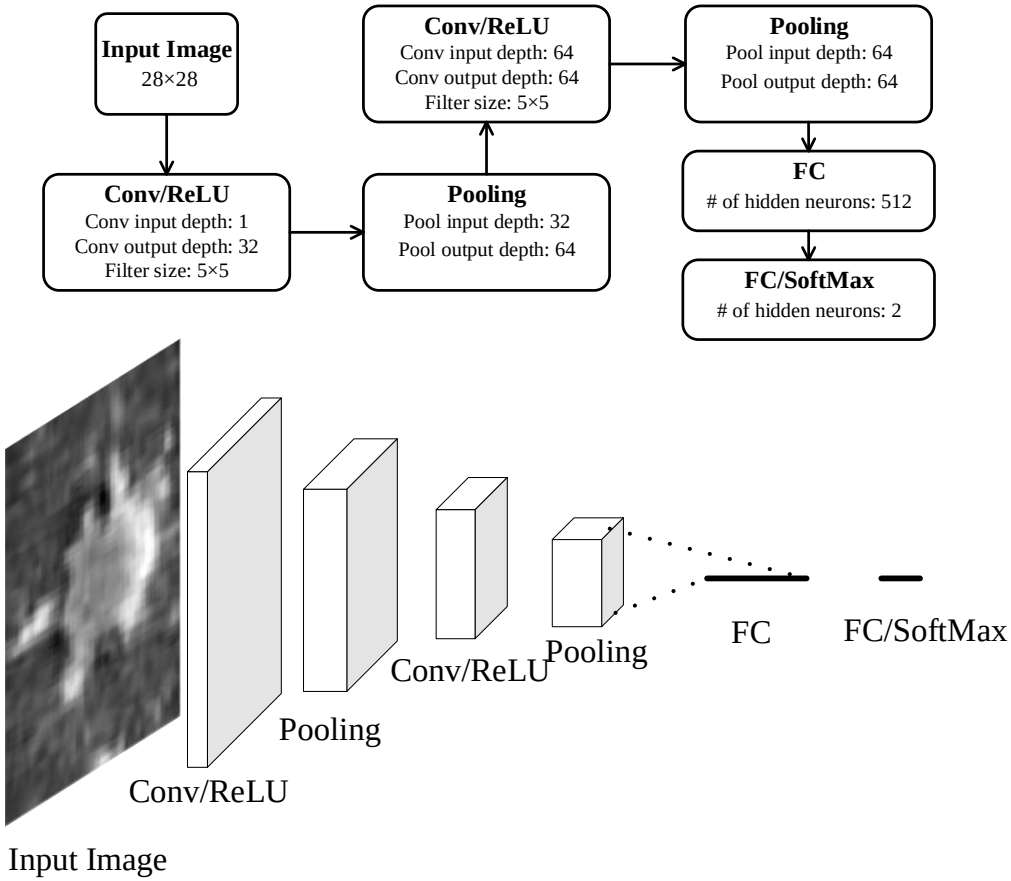


Figure 3.3: Schematic of the ConvNet with two convolutional layers [55]. Conv/ReLU, convolutional layer followed by rectified linear unit; pooling, maximum pooling layer; FC, fully-connected layer; FC/SoftMax, fully-connected layer followed by SoftMax.

highly correlated in-between benign and malignant nodule.

3.3.3 ConvNet with Two Grouped Convolutional Layers

A grouped convolution based ConvNet was proposed to detect pulmonary nodule [37]. The grouped convolution was initially introduced in AlexNet model [33] to reduce computation costs when the output depth of convolutional layer is deep. Grouped convolution also improves classification accuracy when a deep convolutional layer divides into a group of two light convolutional layer [67]. As shown in Figure 3.4, the grouped ConvNet has two groups of two convolutional layers in a series.

3.3.4 ConvNet with Three Convolutional Layers

Another recent successful CAD system for malignancy detection of the pulmonary nodule employs three successive convolutional layers [68]. After trained with 1,018 real clinical pulmonary nodules, the trained model achieved diagnosis on malignancy detection of the pulmonary nodule as well as a real radiologist. Figure 3.5 shows the architecture of the ConvNet with three convolutional layers.

3.3.5 Transfer Learning on AlexNet

Instead of training ConvNet model from scratch, using the pre-trained model to train new data has demonstrated nonperformance in the classification of medical image tasks [6, 63]. Training pre-trained weights to adopt new data is defined as progress of transfer learning. Transfer learning reduces risks of over-fitting and issue of convergence. Especially, medical images always have difficulty to optimize clusters of the classifier due to similarities of its features. More recent studies applied pre-trained deeper ConvNet (AlexNet [33]) with transfer learning to classify pulmonary embolism patterns [58]. The AlexNet consists of 5 convolutional layers, and the architecture of the AlexNet is shown in Figure 3.6. By applying transfer learning, the AlexNet inherited capability of feature extraction from the pre-trained model which had shown state-of-the-art performance on the natural image. However, doubt has raised between transfer learning and customized architecture regarding the significant differences between natural images and medical images.

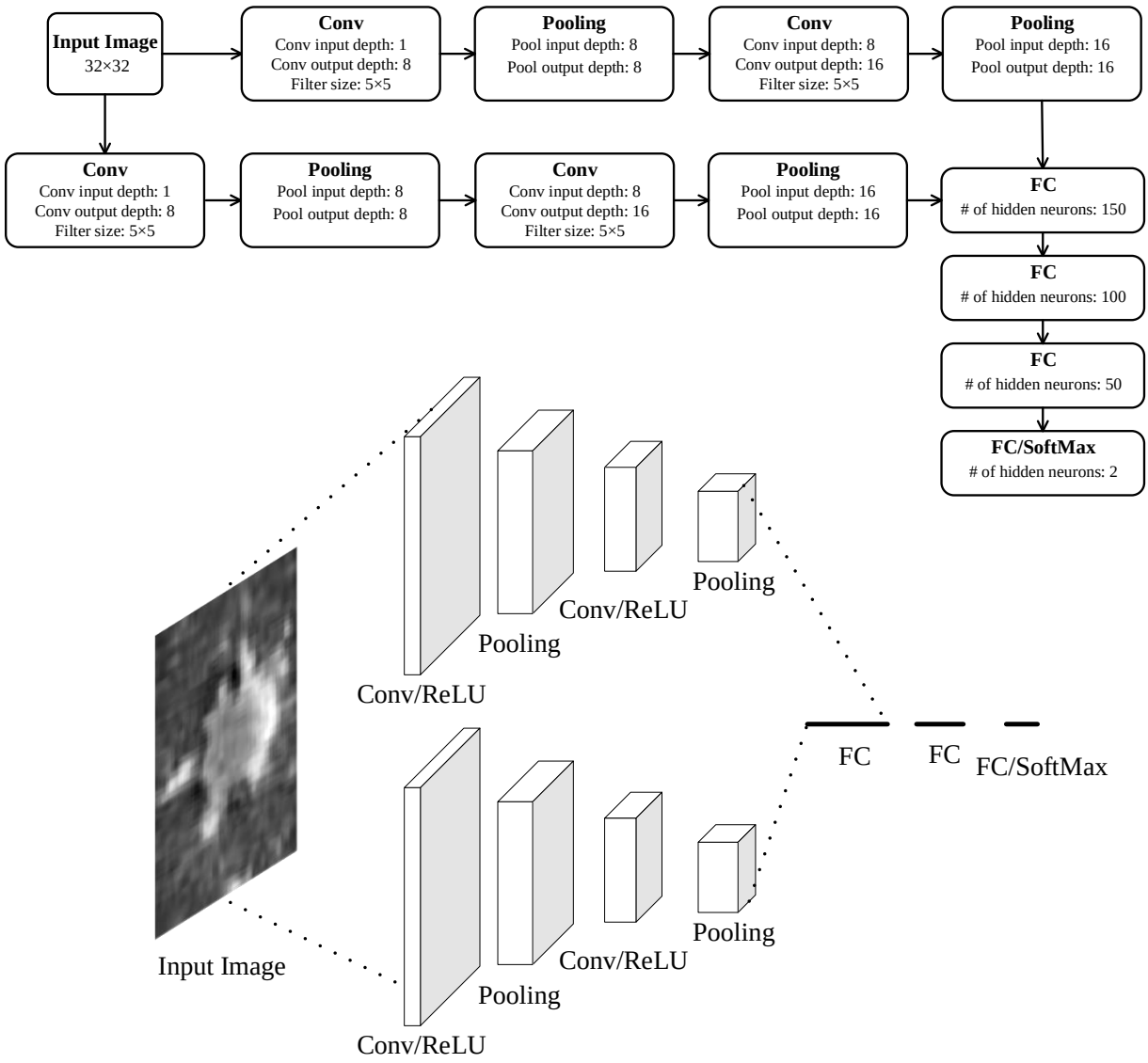


Figure 3.4: Schematic of the ConvNet with two grouped convolutional layers [37]. Conv/ReLU, convolutional layer followed by rectified linear unit; pooling, maximum pooling layer; FC, fully-connected layer; FC/SoftMax, fully-connected layer followed by SoftMax.

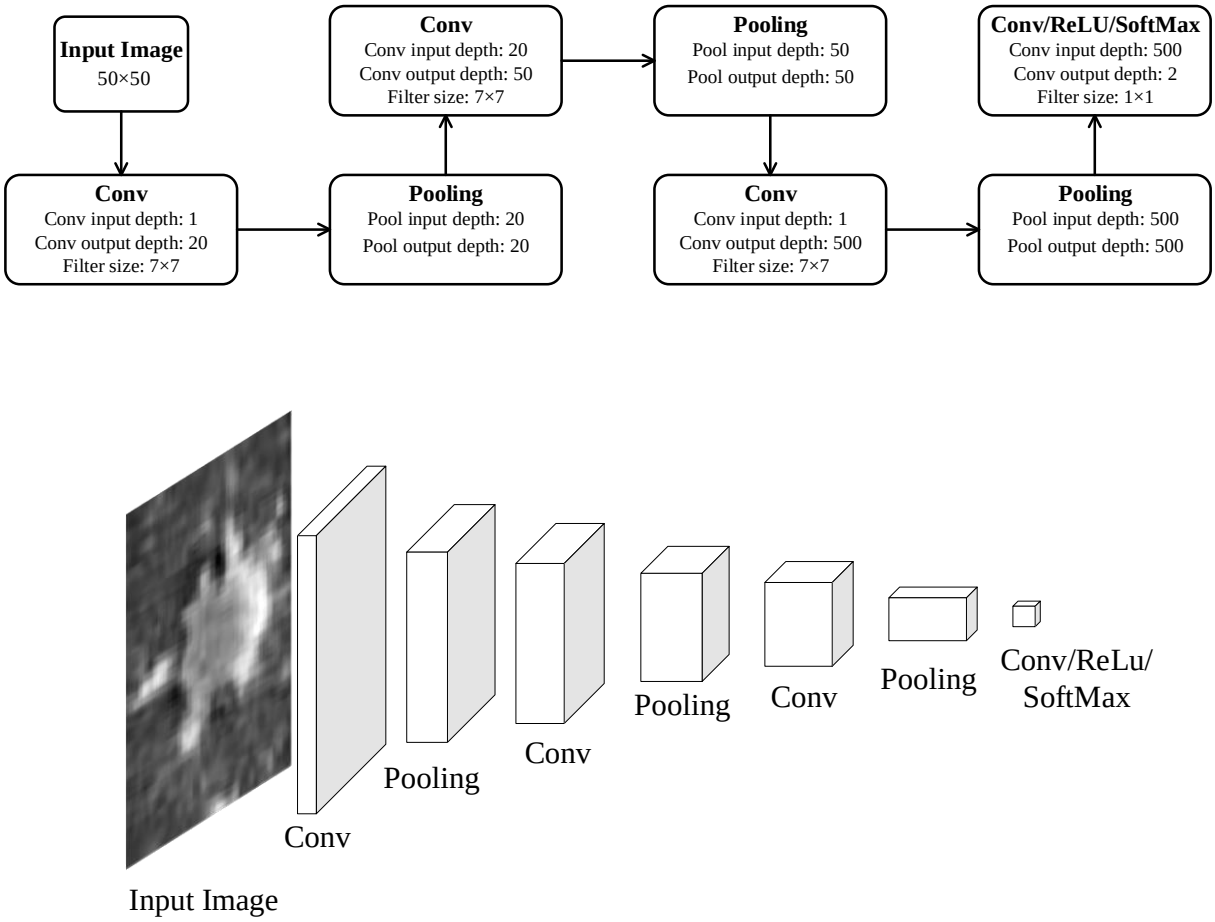


Figure 3.5: Schematic of the ConvNet with three convolutional layers [68]. Conv, convolutional layer; pooling, maximum pooling layer; FC, fully-connected layer; Conv/ReLU/SoftMax, convolutional layer followed by rectified linear unit and SoftMax.

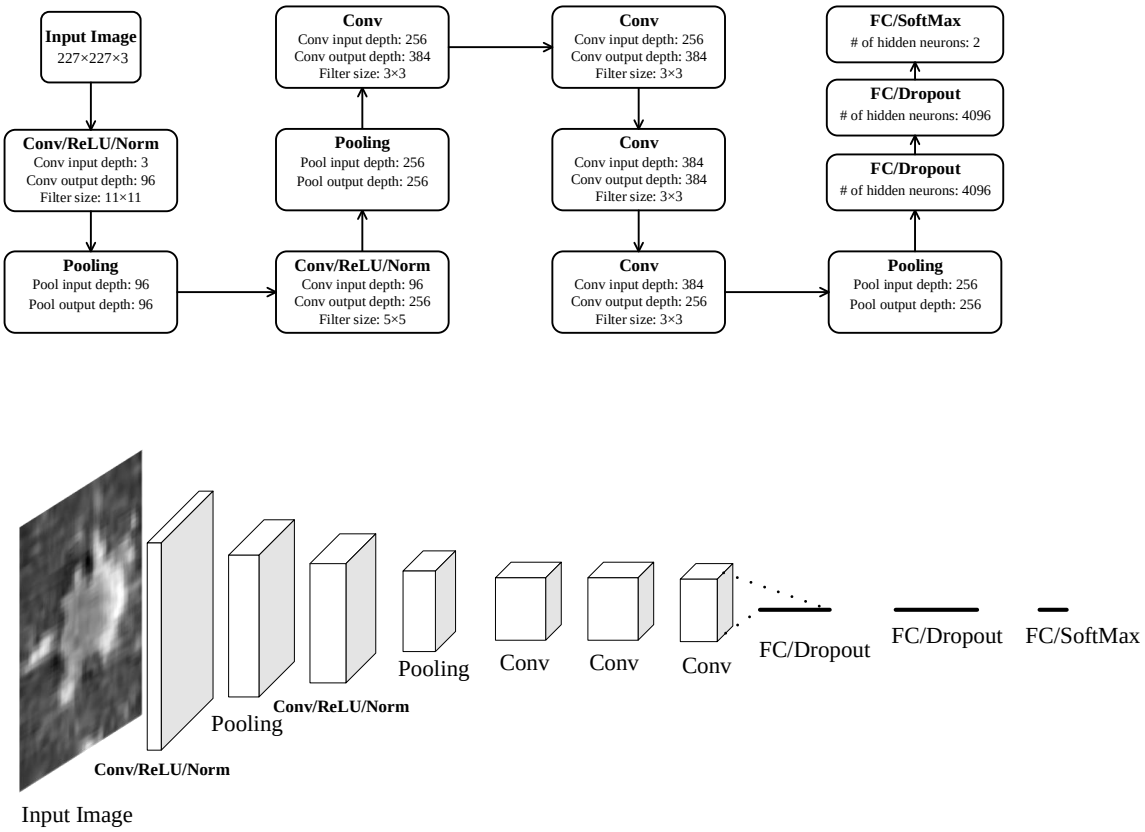


Figure 3.6: Schematic of the AlexNet [33]. Conv/ReLU/Norm, convolutional layer followed by rectified linear unit and batch normalization; pooling, maximum pooling layer; FC/Dropout, fully-connected layer followed by dropout; FC/SoftMax, fully-connected layer followed by SoftMax.

Chapter 4

Proposed Method

This chapter aims to provide the detailed definition of the proposed CAD system. Flowchart of the proposed CAD system is shown in Figure 4.1(a). The proposed CAD system is a novel patch-based malignancy detection tool to learn and capture high-level discriminative features of the pulmonary nodule in a supervised manner and learned features are used for nodule diagnosis without any additional classifier. The system consists of two main stages as follows: 1) data preparation aims to provide ROIs in conditions of spatial transformations. 2) ConvNet performs feature extraction and classification by taking advantage of more relevant information provided from the first stage.

4.1 Data Preparation

Common CAD systems for pulmonary nodule detection aim to analyze the marked area of concern. Data preparation is an important stage to enhance texture description on the nodule that improves feature extraction in the following stage. The data preparation stage is shown in Figure 4.1(b). The data preparation stage consists of three steps: 1) extract ROIs from CT images. 2) over-sampling ROIs. 3) contrast normalization.

The extracted ROI presents the pulmonary nodule as a 2-D image patch, and the image patch highlights the pulmonary nodule within a relatively large region compared to the normal CT slice image. Because it is common that medical imaging database usually has limited data size in contrast of natural imaging database, extracted ROIs are over-sampled via spatial transformation, so the increased number of training samples reduce the risk of over-fitting. On the other hand, different CT scanners lead various changes in illumination of the CT slice image. The contrast normalization is used to reduce the illumination effect by

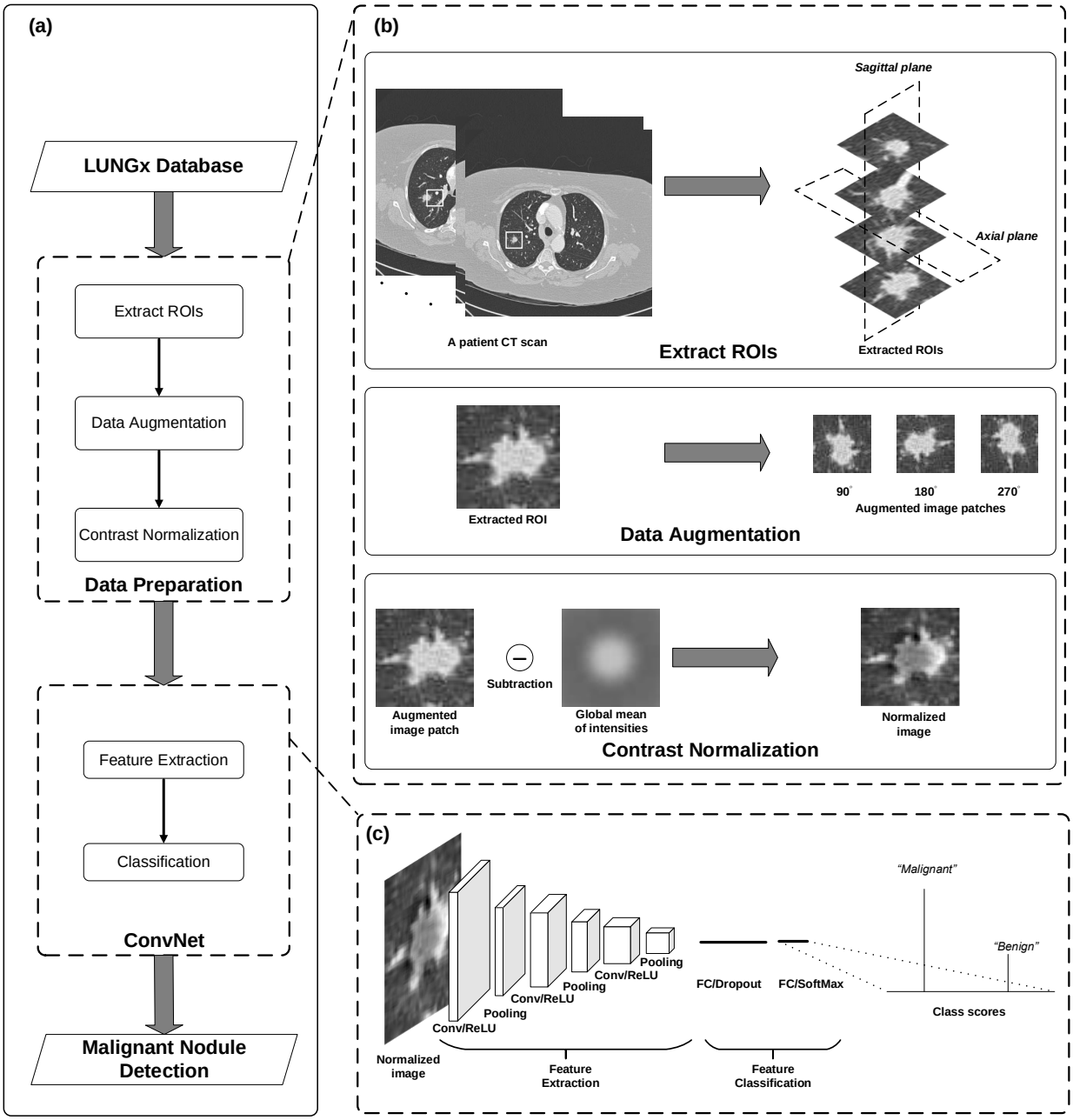


Figure 4.1: Overview of the proposed CAD system. (a) The flowchart of the proposed CAD system. (b) Illustration of the data preparation stage. (c) Schematic of the ConvNet. Conv/ReLU, convolutional layer followed by rectified linear unit; pooling, maximum pooling layer; FC/Dropout, fully-connected layer followed by dropout; FC/SoftMax, fully-connected layer followed by SoftMax; ConvNet, Convolutional neural network.

normalizing the intensities of the ROIs. In the following sections, the detailed method of ROI extraction is described in section 4.1.1, followed by the data argumentation (section 4.1.2) and contrast normalization (section 4.1.3).

4.1.1 ROI Extraction

The proposed CAD system extracts ROIs in the axial plane, and the extracted ROIs contain the full size of the pulmonary nodules. Typical ROI extraction is done by a manual segmentation [2, 41] or automatic detection via the CAD systems [27, 44, 45]. Since LUNGx Challenge database has already provided approximated central mass of each confirmed nodule, ROIs are manually extracted by cropping the axial CT images with bounding boxes. Such bounding boxes have fixed cropping size for each nodule, depending on the visual size of the nodule while keeping the same aspect ratio to prevent information loss when rescaling to match input size of the proposed ConvNet.

Samples of bounding boxes with representations of nodules are shown in Figure 4.2(a) and Figure 4.2(c), and the extracted ROIs demonstrate that the sizes of the pulmonary nodules maintain the major portion of the image patch. If the presence of the pulmonary nodule is in a small region of the ROI, the feature information obtained by the ConvNet is not enough to determine whether there is a pulmonary nodule in the input image patch due to the ConvNet extracts image features within numerous local regions. This can also lead to the decrease in classification accuracy. In addition to the architecture of the proposed ConvNet (as described in section 4.2), three maximum pooling layers are employed to reduce redundant local features by sub-sampling the input feature maps. Therefore, if the size of the pulmonary is too small to be recognized in the ROI, the presence of the pulmonary nodule in the feature maps located in-between the last pooling layer and the first FC layer is too small to provide enough information.

In terms of the ROI size, extracted ROIs are resized to 64×64 pixels to match the input size of the proposed ConvNet. From the practice of the ConvNet, the ConvNet which has relatively large input size requires a lot more parameters to process the input images. With too many learnable parameters, the classification performance of the ConvNet is degraded because of over-fitting. Also, each ROI is extracted in the axial plane which represents

a cross-sectional slice of the CT scan, so each nodule generates various numbers of ROIs, depending on the size of the nodule volume. The samples of the ROIs extracted from single benign and malignant nodules are illustrated in Figure 4.2(b) and Figure 4.2(d).

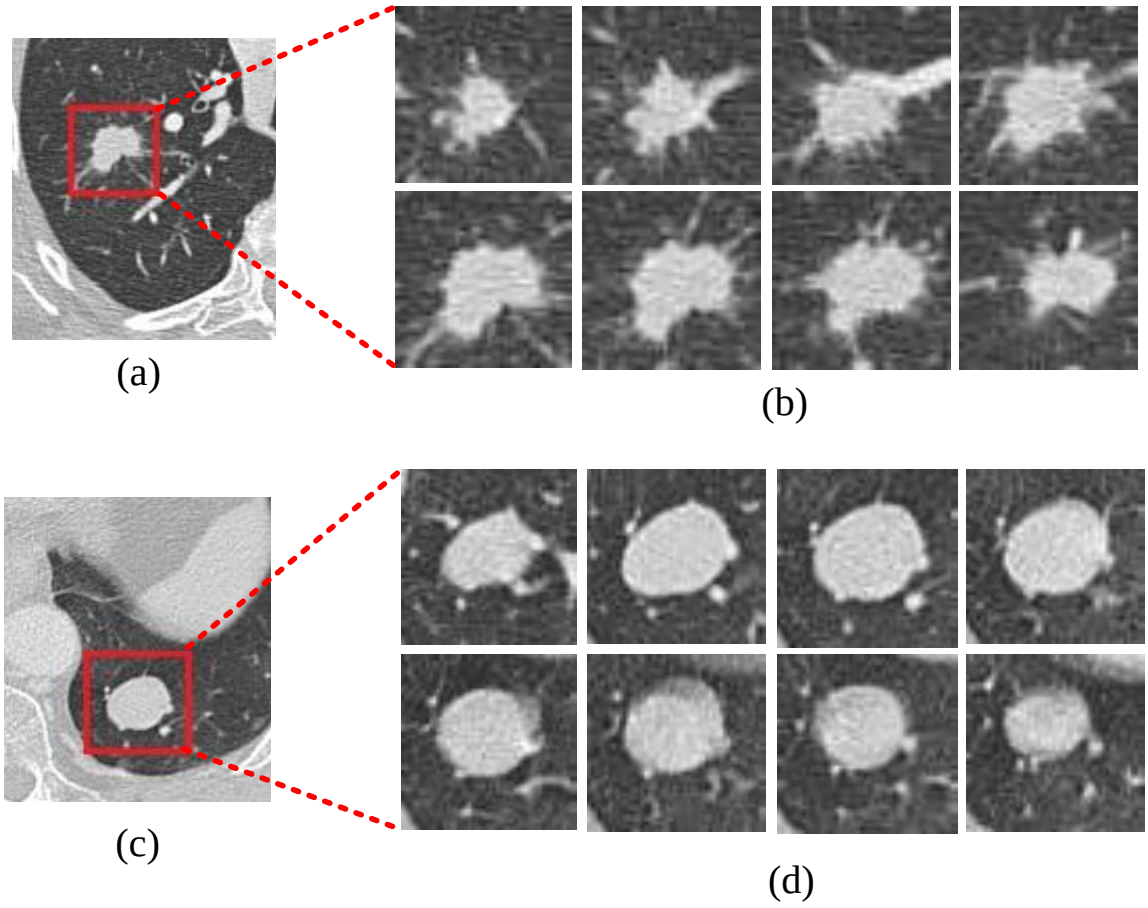


Figure 4.2: Extracted ROIs from LUNGx Challenge database [26]. Images (a) and (c) show the bounding boxes of the pulmonary nodules. Image (a) shows a malignant nodule from 68-year-old female. Image (b) shows extracted ROIs from the malignant nodule. Image (c) shows a 79-year-old female with benign nodule. Image (d) shows samples of the extracted ROIs from the benign nodule.

4.1.2 Data Augmentation

To achieve promising classification results for the diagnosis of the malignant pulmonary nodule, ConvNets rely on massive numbers of learned parameters. However, deep structures always increase the chance of over-fitting when dataset used to train is limited. Because

limited training dataset misleads learned parameters to achieve local optima, the correct prediction only occurs when input data is very close or exact as training dataset. To reduce the risk of over-fitting, data augmentation is used to generate spatial invariance on extracted ROIs. Natural images usually contain geometrical information. For example, an image of the residential house consists of the high-level geometrical structure such as the sky. By applying spatial transformation such as rotation or flipping, the transformed natural image fails to demonstrate the geometrical information. In contrast of natural images, geometrical information is trivial to describe nodule. It is still valid to present nodule in any orientation. In this thesis, each extracted ROI is over-sampled to 3 additional artificial image patches by rotating 90, 180 and 270 degrees.

4.1.3 Contrast Normalization

After rotation transformation, contrast normalization is applied to augmented image patches. By subtracting the global average of intensities which is computed from the entire dataset, the augmented image patches achieves zero-centering, and the zero-centering refers to the zero mean of the entire dataset. Thus, the normalized image has fewer sensitives for illumination changes which usually happened when a CT examination is taken on different CT scanner. For example, a dataset consists of n image patches, and each image presents as an image tensor $x_{i,j}$ in the spatial domain, so the global mean of the entire dataset ($\overline{x}$) can be expressed as:

$$\overline{x} = \frac{1}{n} \sum_{i,j} x_{i,j} \tag{4.1}$$

Contrast normalization also improves performance and reduce training time [29]. Because the learnable parameters are optimized by gradient descent algorithm, normalized images have less spatial variance due to zero-centering, so the contrast normalization avoids steeper gradient update which can significantly affect the training performance because a steeper gradient leads a larger number of update in learning parameters. Hence, the training without contrast normalization has the higher risk to miss the global minima of the loss function, and the trained model is easy to over-fitting.

Moreover, the proposed ConvNet adopts activation layers to introduce non-linearity in-between the convolutional layers. During the training phase, the weights in convolutional filters are initialized as a collection of random numbers. By applying contrast normalization, the output feature maps convolved by randomly initialized filters have zero-centered distributions. As a consequence, a small step of update in the weight parameter leads to significant changes in the activation layer because of the activation function is a non-linear function which has the steepest slope near the origin. This also leads the model to have a faster convergence speed.

4.2 Proposed ConvNet

After obtaining nodule image patches via the data preparation stage, the next stage is to perform classification task by analyzing the features of input regions in various levels using ConvNet. The schematic of the proposed ConvNet is shown in Figure 4.1(c).

It is crucial to choose an appropriate filter size for each convolutional layer due to local connectivity within fixed-scale of spatial area. Natural images are usually constructed with high-level structures among whole image region such as arbitrary colors, multiple objects or geometrical information. In contrast of natural images, nodule image patches are characterized by local textural features. Such features are stochastic patterns which repeat within a small region. In order to capture the local textural features, small filters are applied so that the convolutional layer takes advantage of the filters adopted with small size by analyzing fine-grained local regions.

In order to optimize the architecture of the proposed ConvNet, the receptive field should also be considered. The receptive field is defined as the particular region in input volume that ConvNet performs extraction within this region. As illustrated in Figure 4.3, an input feature map with the size of 4×4, convolved with the filter size of 2×2 and stride of 1, produces 3×3 output feature map. By applying the same filter size of convolution, the final output feature map size is reduced to 2×2. Because each feature on the last feature map requires 2×2 features from the first feature map where is acquired from 3×3 input features, the output feature map is able to describe the local structures on 3×3 local regions of the

input feature map. Hence, an appropriate total receptive field should be considered to pass enough local sparse structures from input to output of the ConvNet while preventing final output features exceed input receptive field.

In this thesis, a group of 5×5 fixed scale filters is used for convolution operation along with input width and height, which its size is relatively small by comparing with the input size of 64×64. Also, the proposed ConvNet consists of 3 convolutional layers and 3 maximum pooling layers, and total receptive field is limited to 36×36 so that approximately half of the input receptive field has been seen by the ConvNet while maintaining the specific local textural features. Table 4.1 shows the size of the receptive field respected to each learnable layer in the proposed ConvNet. The receptive field of each learnable layer gradually increases while the input data propagates via the 3 convolutional layers and 3 pooling layers which are employed in the proposed ConvNet.

Table 4.1: Receptive field of each learnable layer in proposed ConvNet

	Input	Conv	Pool	Conv	Pool	Conv	Pool
Filter size		5×5	2×2	5×5	2×2	5×5	2×2
Receptive field	64×64	5×5	6×6	14×14	16×16	32×32	36×36

Conv, convolutional layer; Pool, maximum pooling layer

4.2.1 Architecture

The architecture of the proposed ConvNet is shown in Figure 4.4. The processed image patches from the previous stage are rescaled to 64×64 pixels to match the input data size. The ConvNet to be applied consists of consecutive 3 convolutional layers followed by 2 fully-connected layers. All the three convolutional layers use 5×5 filters. The same filter will be used to slide around one input feature map to generate one output feature map. The first convolutional layer contains 1×12 filters of size 5×5, which receives the 1 input feature map and generates 12 output feature maps. Similarly, the second and third convolutional layer contain 12×24 and 24×48 filters respectively.

The ReLUs are used as the activation function of each convolutional layer. The activation

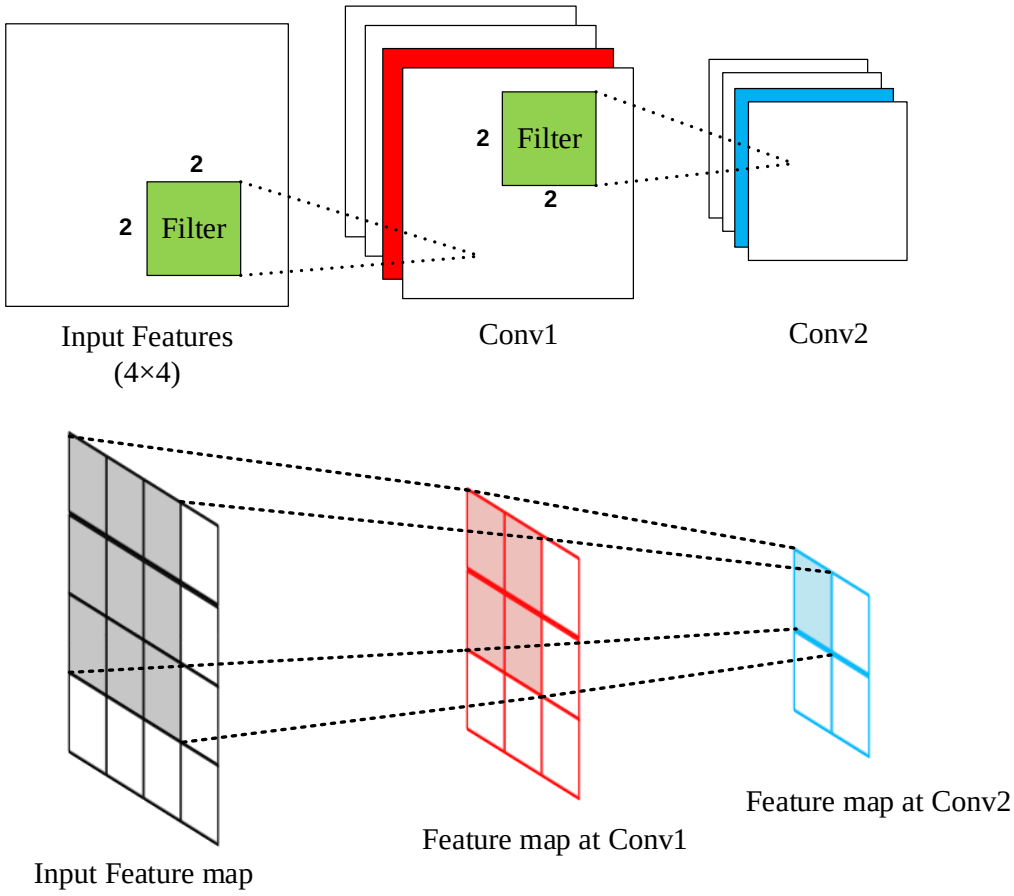


Figure 4.3: Example of the receptive fields in a network with two convolutional layers. Conv, convolutional layer

function can introduce non-linearity to the network. Especially, ReLU, which has shown better gradient changes than sigmoid and tanh functions [33], suppresses negative values to 0 and keeps positive values. Besides, the ReLU function, $f(x) = max(x, 0)$, is easy to implement which can help improve the speed performance of both the training and inference process of the network.

Max-pooling layer is inserted after each convolutional layer. Max-pooling operation maintains maximum value within the non-overlapping filter of size 2×2 and stride of 2 on each feature map. The pooling layers are used to reduce the dimensionality of the following layers of the network in order to maintain more relevant local features. Moreover, pooling operation increases the receptive field size so that the network can learn more complex local sparse structures from the input. In the proposed ConvNet, the output dimensionality will be half

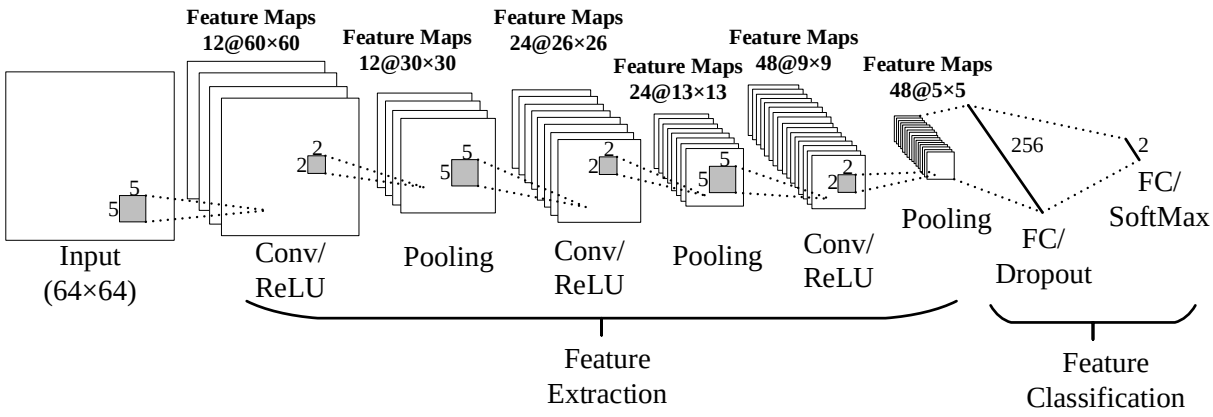
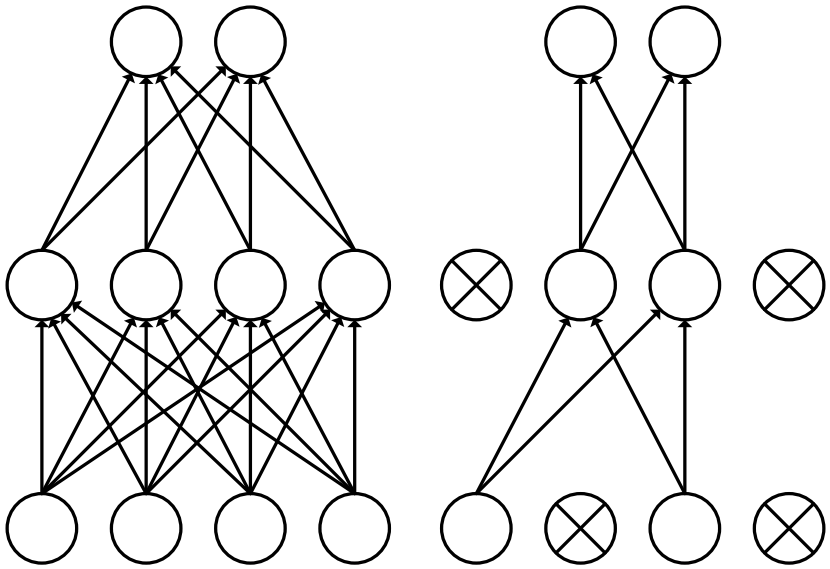


Figure 4.4: Architecture of proposed ConvNet. Conv/ReLU, convolutional layer followed by rectified linear unit; pooling, maximum pooling layer; FC/Dropout, fully-connected layer followed by dropout; FC/SoftMax, fully-connected layer followed by SoftMax.

of that of the input after each max-pooling operation.

Fully-connected (FC) layers construct the last part of the proposed ConvNet after the max-pooling layer. First FC layer has 256 output neurons to build full connectivity and shares all local sparse structures from 48 feature maps generated by the last max-pooling layer. The last FC layer consists of 2 output neurons in order to match the number of output classes which are benign and malignant.

The denser FC layer often leads over-fitting due to full connectivity structure which requires a large number of parameters. Here in the proposed ConvNet architecture, in order to avoid over-fitting, the dropout [56] scheme is applied to the first FC layer. The output of the last FC layer is set to zero with a probability of 0.5. Figure 4.5 shows a network with three FC layers employs a dropout with a ratio of 0.5. After applying dropout, the two hidden neurons in both first and second FC layers act inhibitory. As a result, the total number of the synapse paths are reduced from 24 to 8. The dropped out neurons do not contribute to the network calculation on both of forwarding pass and backward propagation. Therefore, each time an input is present during training, the network performs the forward inference with a different architecture. This can effectively reduce the co-adaptation of the neurons and thus avoid over-fitting.



(a) Standard FC layers (b) Dropout with a ratio of 0.5

Figure 4.5: Network with three fully-connected layers applies dropout with a ratio of 0.5. Image (a) illustrates the connections without applying dropout. Image (b) shows the connections with dropout of 0.5.

4.3 Training Strategy

To express the full potential of classification performance in the proposed ConvNet, training strategy is applied to perform a promising classification result on malignancy detection of the pulmonary nodule. The training strategy can be viewed as three parts: 1) reorganized database 2) choice of the optimizer. 3) selection of hyper-parameters.

4.3.1 Dataset Distribution

Diagnostic performance of malignant pulmonary nodule under the proposed CAD system is evaluated based on train-validation-test scheme. The LUNGx organizers stated that the calibrations sets are not relevant for the evaluation of difficulty levels in the test sets, while ConvNet generally needs a large amount of data to learn enough discriminative features in order to have full potential in pulmonary nodule classification to distinguish malignant and benign classes. The training scheme of the proposed CAD system is different from LUNGx

Challenge training scheme.

To ensure the performance of proposed CAD system, instead of using LUNGx calibration sets to train the proposed ConvNet, 54 cases in the LUNGx testing sets are reserved for training and validation purposes, and leaving additional 6 cases in the LUNGx testing sets and entire LUNGx calibration cases to test the performance of the proposed ConvNet. Here in the proposed CAD system, the training sets consist of 60 nodules with 30 benign and 30 malignant cases while the validation sets include 3 benign nodules and 2 malignant nodules. During the testing stage, the proposed CAD system is evaluated by 9 benign nodules and 9 malignant nodules.

Figure 4.6 illustrates the dataset distribution for the training-validation-test scheme of the proposed ConvNet by reorganizing the LUNGx Challenge dataset, so the proposed ConvNet is trained by 6 times more nodules compared to LUNGx Challenge training data size, and the rest of nodules are divided into validation dataset and testing dataset.

After the data preparation stage, the proposed ConvNet is trained by augmented image patches where 2,225 image patches belong to malignant nodules and 1,496 image patches belong to benign nodules. The proposed ConvNet uses 112 image patches as the malignant class and 76 image patches as the benign class for validation. After the training stage, the proposed ConvNet is tested by 432 image patches as malignant and 344 image patches as benign. Because the training, validation and testing sets of the proposed CAD system are organized based on nodule cases, this setup guarantees augmented image patches for validation and testing that have never been "seen" by the trained model during the training stage. The data distribution of the proposed CAD system is shown in Table 4.2.

4.3.2 Optimizer

The proposed ConvNet is trained with the back-propagation algorithm by minimizing weight losses to achieve an optimal model. The weight loss is measured by cross-entropy error and normalized by SoftMax to represent the probabilistic distribution of classes. The weights are optimized by using stochastic gradient descent (SGD) which is a non-adaptive gradient descent optimizer. Although adaptive optimizers such as Adam, AdaGrad or RMSProp can significantly reduce training time due to faster convergence by comparing with SGD, adaptive

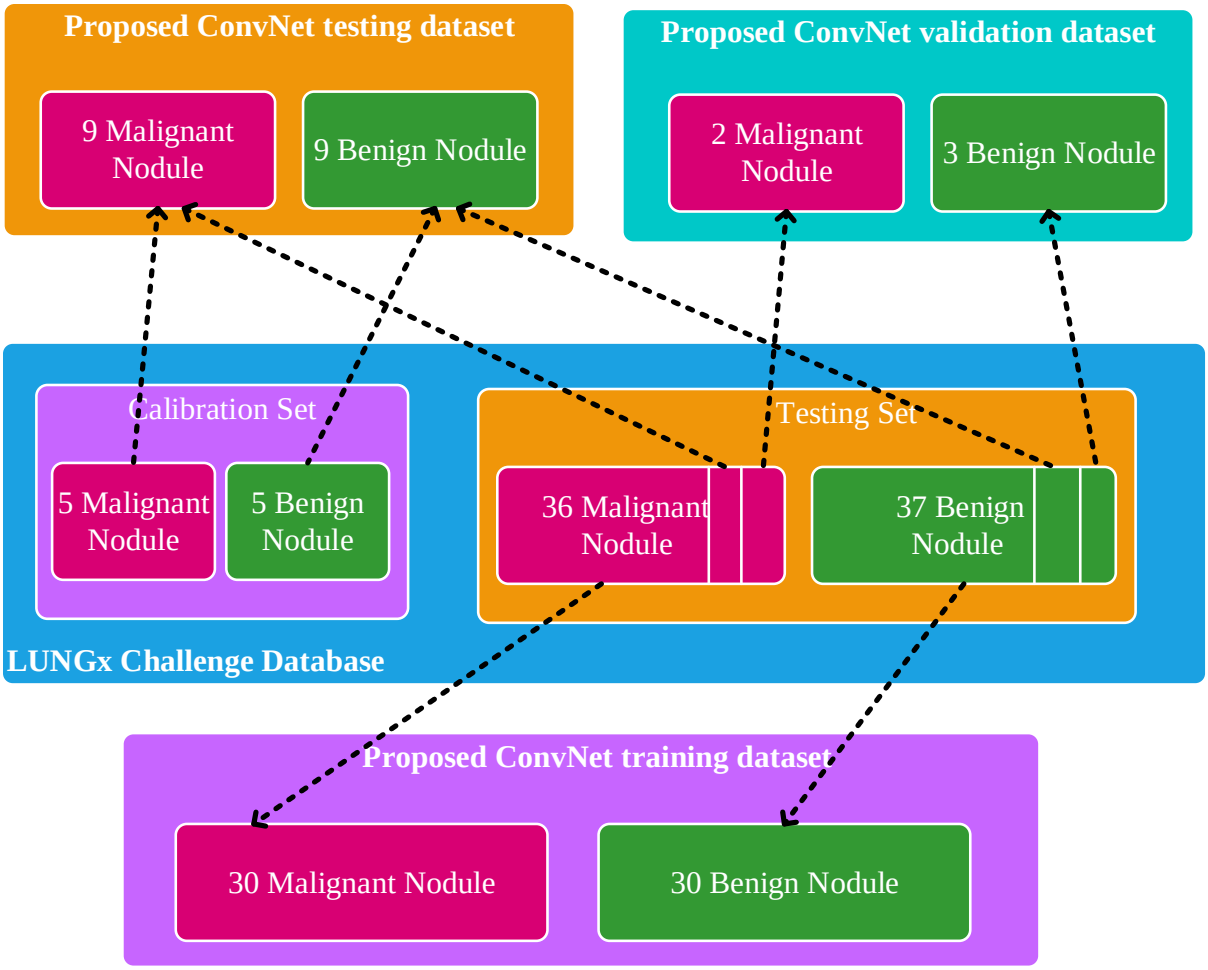


Figure 4.6: Dataset distribution for the training-validation-test scheme of the proposed ConvNet compared to the LUNGx Challenge training-test scheme.

optimizer tends to have insufficient generalization to perform classification on new data. Wilson et al.[66] pointed out that models optimized by SGD outperformed classification on new data by comparing with adaptive optimization methods. Adaptive optimizers produced inaccurate classification results with probability arbitrarily close to half. Therefore, SGD is applied to optimize the proposed ConvNet.

4.3.3 Hyper-parameters

Hyper-parameters define the complexity of the ConvNet such as filter size or input size. Hyper-parameters are also the parameters defined before training such as learning rate, batch

Table 4.2: Dataset distribution

	Data set	Training set	Validation set	Testing set
Number of malignant nodules	41	30	2	9
Number of benign nodules	42	30	3	9
Proportion of nodules	100%	72%	6%	22%
Number of malignant patches	2,769	2,252	112	432
Number of benign patches	1,916	1,496	76	344
Proportion of patches	100%	79%	4%	17%

size or initialized weight distribution. An exponential decay learning scheme is used to update the learning rate with the initial rate of 0.0001. 16 input samples per batch are employed to update weights while weights are initialized as the Gaussian distribution with a standard deviation of 0.005 and constant bias of 0.1. The predefined number of training epochs is 1,000, but the optimal trained model is obtained when the validation result at the epoch achieves the best overall accuracy by comparing with validation results at next 10 epochs. The overall accuracy is the average of corrected predictions from both malignant and benign classes.

Chapter 5

Results and Analysis

By integrating the proposed highly accurate ConvNet, the proposed CAD system achieves a promising diagnostic accuracy for detection of malignant pulmonary nodules on CT. In section 5.3, a set of experiments justifies the proposed ConvNet architecture from choices of hyper-parameters, layer components and training strategy. Then, the proposed CAD system performance for diagnosis of the malignant pulmonary nodule on CT is compared with state-of-the-art work in section 5.4, followed by performance comparisons with other ConvNet architectures in section 5.5.

5.1 Experiment Environment

The proposed ConvNet is implemented under Caffe [30] deep learning framework. Other experiments related with convolutional neural networks are also performed under Caffe environment. Training and testing networks are processed under Ubuntu 16.04 Linux system with Intel Xeon E5-2630 @ 2.60 GHz, 32 GB RAM and Nvidia K40 GPU.

5.2 Experiment Metrics

5.2.1 Accuracy

Accuracy measures the average of correct predictions on benign and malignant nodules during the validation and testing phase. Although accuracy is trivial to evaluate the diagnostic performance for malignancy detection of the pulmonary nodule, it is a good estimator to evaluate the performance of the classifier. Because the proposed CAD system consists of

a ConvNet to classifier the pulmonary nodules in-between benign and malignant nodules, the performance of the classifier in the proposed ConvNet can be evaluated via accuracy. In section 5.3, the classification performances of the ConvNet with different configurations of the hyper-parameters are analyzed based on the accuracy metric. The accuracy of the classifier in the proposed ConvNet can be expressed as:

$$Accuracy = \frac{Number\ of\ correct\ predicition}{Total\ number\ of\ inputs} \tag{5.1}$$

5.2.2 Sensitivity and Specificity

To evaluate the diagnostic performance for malignancy detection of the pulmonary nodule, the criterion of the estimators are sensitivity and specificity. In the field of medical research, the sensitivity refers to the probability of the diagnostic decision for a particular disease correctly identified, and the specificity is the ability of correct diagnostic decision for the patient without the particular disease. In this thesis, the sensitivity represents a percentage of malignant nodules classified as malignant; whereas, the specificity is a percentage of benign nodules predicted as benign. The sensitivity and specificity can be expressed as:

$$Sensitivity = \frac{Number\ of\ maligant\ nodule\ classified\ as\ malignant}{Total\ number\ of\ malignant\ nodules} \tag{5.2}$$

$$Specificity = \frac{Number\ of\ benign\ nodule\ classified\ as\ benign}{Total\ number\ of\ benign\ nodules} \tag{5.3}$$

5.2.3 Statical Analysis: Receiver Operating Characteristic

Since confidence level of classification for classes of the malignant or benign nodule is based on normalized scores between those two classes, the receiver operating characteristic (ROC) [71], a statistical analysis, is used to evaluate overall diagnostic performance for detection of the malignant pulmonary nodule. The ROC is the plot of sensitivity (true positive rate) and specificity (false positive rate) at different decision thresholds of the classifier. An area under the ROC curve (AUC), a classification analysis, measures the area the entire ROC curve, and AUC provides the classification performance based on the true positive rate and false positive

rate on all possible classification thresholds. In this thesis, AUC is calculated under DeLong test [14] in order to compare performances between different classifiers. By using 1,000 bootstraps, sensitivity and specificity at the optimal cut-off of the ROC are obtained and used as supplemental estimators. In the section 5.4, a comparison with state-of-the-art work by using unsupervised method is evaluated based on sensitivity and specificity at the optimal cut-off point of ROC curve; also, AUCs are compared. In section 5.5, the classification performance of the proposed ConvNet is also compared with other ConvNet architectures by performing ROC analysis. Instead of using accuracy to measure the classification performance in section 5.6, the classification result is analyzed by AUC under ROC because the experiments are done under different testing datasets.

5.3 Tuning of Hyper-parameters

Table 5.1: Performance of the ConvNets with Different Configurations

Input Scale	Dropout	Filter Size	Optimizer	⋆Accuracy
64×64	0	5×5	SGD	0.8505
64×64	0.5	7×7	SGD	0.8402
64×64	0.5	3×3	SGD	0.8479
32×32	0.5	3×3	SGD	0.8763
64×64	0.5	5×5	Adam	0.8595
64×64	**0.5**	**5×5**	**SGD**	**0.8866**

⋆ **Accuracy**: average of correct predictions on benign and malignant nodules from the testing sets.

In order to choose optimal ConvNet architecture, relevant experiments are performed based on different configurations of the hyper-parameters. The malignant pulmonary nodule detection performances of different ConvNet architectures are evaluated by using the testing sets, and the results are shown in Table 5.1. According to the Table 5.1, the proposed ConvNet uses a group of 5×5 fixed scale filters and SGD optimizer while dropout with a ratio of 0.5 is applied on the first FC layer.

A group of 5×5 fixed scale filters is proposed in order to limit the total of receptive field (36×36) into approximately half of the input receptive field. By increasing the size of the filter to 7×7, the overall accuracy results in a drop of roughly 5%. The total receptive field is enlarged to 50×50 which is an insufficient size to pass enough local sparse structures from the first to the latest layer by comparing with the proposed filter size of 5×5. A smaller filter, size of 3×3, is also compared with the proposed filter size. Using a filter size of 3×3 reduces overall accuracy by approximately 4%. In this thesis, the input receptive is limited to 64×64 pixels. Using a filter size of 3×3 is too small to capture enough invariant information due to the highly correlated neighboring pixels from the input data. The input image patches are then rescaled to 32×32. The overall accuracy improves by 3% while using a filter size of 3×3. However, the performance of using 3×3 filter is still inferior to the proposed filter size.

Also, a dropout with a probability of 0.5 increases approximately 4% overall accuracy compared with the same ConvNet architecture without dropout operation. Performance differences for the ConvNets trained by SGD or Adam are also compared. The SGD gradient optimizer surpassed the Adam by increasing the overall accuracy of about 4/

5.3.1 Analysis of the Proposed CAD System's Performance

This section provides additional analysis for the performance of the proposed ConvNet. During the training phase, the training loss and validation accuracy for each epoch are shown in Figure 5.1. As illustrated in Figure 5.1, the training loss of the proposed ConvNet started to converge until roughly 60 epochs. The best validation accuracy achieved at 208 epochs. Figure 5.2 shows classified nodule samples and some difficult cases that nodule samples were misclassified by the proposed ConvNet.

5.4 Comparison with Unsupervised Method

The proposed CAD system performance for diagnosis of the malignant pulmonary nodule on CT is compared with previous state-of-the-art work [46]. The previous work adopted unsupervised learning scheme by applying Principle Component Analysis (PCA). The previous work was trained and evaluated under testing dataset of the LUNGx Challenge database.

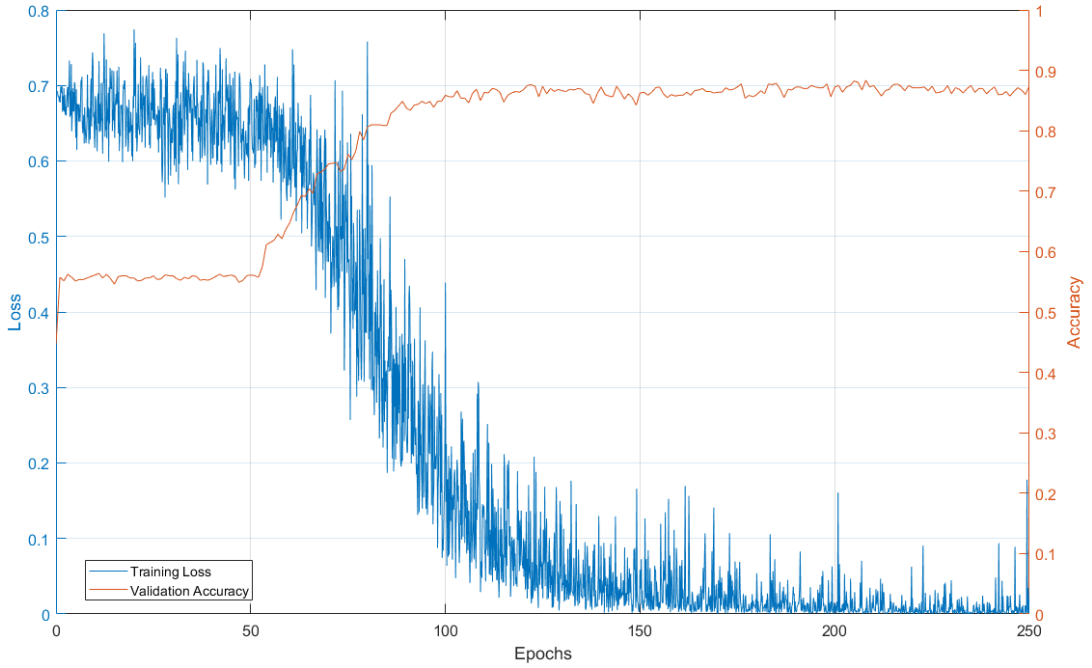


Figure 5.1: Training losses and validation accuracies during the training of the proposed system.

However, the authors did not provide the data distribution for training and testing datasets. To perform a fair comparison in-between the proposed ConvNet and the state-of-the-art work, the classification performance in the proposed ConvNet is further justified via additional nodule samples acquired from calibration set of the LUNGx Challenge database (as referring to Figure 4.6).

The comparison with the previous work is shown in Table 5.2. Nishio et al. [46] applied a combination of PCA, convolution and pooling operations as a multi-stage feature extractor, followed by a linear SVM to perform classification in-between malignant and benign pulmonary nodules. The system [46] was evaluated under LUNGx testing sets and achieved a sensitivity of 0.867 and specificity of 0.744 at the optimal cut-off point of the ROC curve with AUC of 0.837. In contrast with Nishio's approach, the proposed CAD system outperforms pulmonary nodule detection in-between benign and malignant cases. The best-trained model of the proposed ConvNet achieved AUC of 0.920 with a sensitivity of 0.896 and specificity of 0.878.

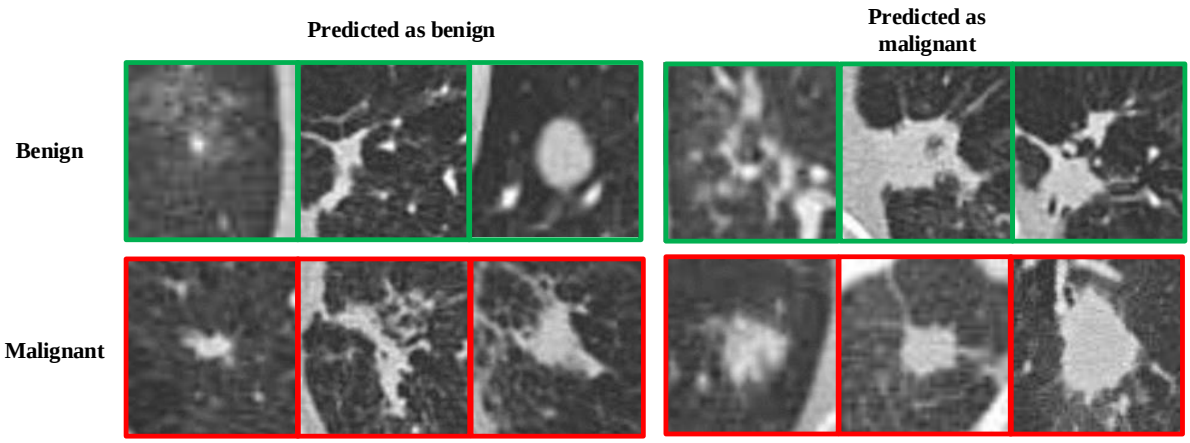


Figure 5.2: Example of ROIs that were classified or misclassified by the proposed ConvNet.

The proposed CAD system surpassed unsupervised feature extraction that involves multi-stages as described in the previous work [46] for three reasons: 1) principal components are statistically based on the covariance matrix. However, the covariance matrix cannot be evaluated in an accurate manner. PCA transforms input features into high dimensional spaces. Such input transformation often leads to the difficulty to calculate a large size covariance matrix within a limited training database. In contrast with PCA, the proposed CAD system, applied ConvNet structure, extracts features in high dimensional spaces without the need for the covariance matrix. 2) PCA has high sensitivity on illumination and local structure changes such as shifting or rotation. Instead of providing invariant samples for training the classifier explicitly, the proposed CAD system can recognize invariant changes by inserting pooling layer after each convolutional layer which maintains relevant local features. 3) the PCA followed by single convolution operation has a shallow hierarchical layered structure by comparing with the proposed CAD system. It is expected that the proposed CAD system surpasses the previous work [46] due to more complex extracted features. The proposed CAD system is a hierarchical layered structure to extract high-level discriminative features based on the hierarchy of abstractions. By comparing with the previous work [46], more complex (higher-level) internal representations of the input image are hierarchically extracted from the first to the last convolutional layer.

In contrast of the hand-crafted feature extractors such as histogram of density, LBP-TOP

Table 5.2: Comparison of the proposed CAD system with the previous work

Methods	Feature Extractor	Classifier	Sensitivity	Specificity	AUC
Nishio [46]	Hisogram	SVM	0.867	0.488	0.640
	LBP-TOP	SVM	0.900	0.558	0.688
	RLBP	SVM	0.800	0.674	0.725
	Multi-stages	SVM	0.867	0.744	0.837
Proposed	ConvNet		0.896	0.878	0.920

Histogram, histogram of CT density; LBP-TOP, local binary pattern on the three orthogonal planes; RLBP, LBP with 3-D random sampling; Multi-stages, combination of PCA, convolution and pooling operations; SVM, support vector machine.

and RLBP, the proposed CAD system achieved superior classification result for detection of malignant pulmonary nodule due to a hierarchical layered structure. Such hand-crafted features are histogram transformations limited in spatial domain with lack of complexity in terms of data representations which often fail to adopt new data.

5.5 Comparison with Other ConvNet Architectures

Table 5.3: Comparison with other ConvNet architectures

Methods	Sensitivity	Specificity	AUC	95% Confidence Interval
Li et al. [36]	0.785	0.860	0.868	0.842 to 0.891
Li et al. [37]	0.824	0.853	0.884	0.865 to 0.902
Song et al. [55]	0.835	0.851	0.843	0.821 to 0.864
Yang et al. [68]	0.850	0.838	0.876	0.856 to 0.895
Tajakhsh et al. [58]	0.792	0.913	0.897	0.874 to 0.918
Proposed Method	0.896	0.878	0.920	0.898 to 0.938

Some ConvNets are already applied for various medical imaging applications. To show the superiority of the proposed ConvNet when performing pulmonary nodule classification, the ConvNet architectures, proposed in [36, 37, 55, 68, 58], are also trained and tested for pulmonary nodule classification under the same training and testing datasets. Their performances are compared with that of the proposed ConvNet, and experiment results are summarized in Table 5.3.

Shallow ConvNet

The shallow ConvNet proposed in [36], had inferior experiment result with AUC of 0.868 because Li et al. [36] applied a single convolutional layer followed by a single max-pooling layer and three FC layers. Because quantities of learned features are based on hierarchical layered structure, such shallow structure cannot capture enough high level features. Also, the shallow ConvNet had demonstrated the over-fitting issue due to redundant hidden neurons in the FC layer. However, the proposed ConvNet showed similar generalizations for both positive and negative samples (e.g., sensitivities and specificities). In contrast with the shallow ConvNet, the proposed ConvNet adopted approximately 4% less hidden neurons and dropout method to reduce complexities of the connectivities in FC layers.

ConvNet with Two Grouped Convolutional Layers

Li et al. [37] proposed a ConvNet with two grouped convolutional layers and achieved an AUC of 0.884. Although the number of the convolutional layers is doubled compared with the shallow ConvNet [36], the filter size for each convolutional layer is relatively large with respect to its input receptive field. In [37], the authors used 5×5 filters while the input receptive field is 32×32. To overcome this problem, the proposed ConvNet uses the same filter size with larger input receptive field size to pass enough invariant information and prevent losses of local sparse structures.

ConvNet with Two Convolutional Layers

Song et al. [55] proposed a ConvNet which has a similar architecture as [37]. The ConvNet adopts two convolutional layers with the filter size of 5×5. The ConvNet proposed in [55] had

inferior experiment result with AUC of 0.829 because the input receptive is further reduced to 28×28. As a result, such large filter sizes fail to capture detailed local information. Moreover, the proposed ConvNet employs dropout scheme to reduce the complexity of the fully-connections in FC layer. However, [55] has to tune two times more parameters in the FC layer, so it is easier to be over-fitted.

ConvNet with Three Convolutional Layers

The ConvNet [68], which employed with successive three convolutional layers, had AUC of 0.876, but the result is still inferior compared with the proposed ConvNet. Adding an FC layer translates local features extracted by convolutional layers into unique representations of different classes, and translated information is used to perform classification. Although the last layer in [68] transforms the extracted features into a 1-D vector by performing convolution, the last convolutional layer only builds two hidden neurons that have full connections from the previous layer. As a result, all extracted features from previous layers rely on the two neurons; unlikely, the proposed ConvNet adopts 256 neurons to translate local extracted features from the previous layer and learn the non-linear combination of those features effectively.

Transfer Learning on AlexNet

In addition, during ImageNet contest [15], AlexNet showed approximately 17% error rate among 1.2 million natural images over 1,000 different classes. AlexNet consists of 5 convolutional layers and 3 FC layers with over 61 million learnable parameters. In our experiment, the AlexNet model is evaluated by applying transfer learning in order to compare pulmonary nodule classification performance with the proposed ConvNet composed with fewer layers. In order to match the input size of AlexNet, all image patches are resized to 224×224, and RGB channels are encoded by duplicating with same gray level intensities. A layer-wise fine-tuning strategy [58] is applied to gradually re-adjust the pre-trained weights in order to achieve the optimal result. After an adequate fine-tuning, the best model achieved AUC of 0.897.

However, the proposed ConvNet still has the superior performance by comparing with the fine-tuned AlexNet model. By considering the full connectivity in-between the last pooling

and the first FC layer on AlexNet, 37 million learnable parameters had to be tuned during training so that this full connectivity is too redundant in limited training datasets. Also, it is still possible that the huge number of parameters lead the trained model to achieve local optima.

ROC Analysis

ROC curves for different ConvNet architectures are shown in Fig. 5.3. To test the statistical significance of the AUC differences among all analyzed ConvNets, the significance level under DeLong test is performed [17]. Significance level refers to the probability of null hypothesis occurred. If the significance level is less than .05 ($\rho \leqslant .05$), the null hypothesis is rejected. In the presence of this thesis, the null hypothesis is that the proposed ConvNet has a statically similar diagnostic performance for malignancy detection of the pulmonary nodule. According to the resulted ROCs, the proposed ConvNet achieve statically significant ($\rho \leqslant .05$), so the proposed ConvNet is statistically confirmed to be the best model against all models.

5.6 Investigation of Imbalanced Data Problem

To train a ConvNet under natural imaging database, the training dataset is generated by randomly selecting a fixed portion of data from the entire augmented database. However, augmented medical imaging database usually have high spatial correlations within each lesion sample, so a random selection in nodule image patches leads unreliable high performance on the diagnosis of pulmonary nodule because the testing or validation dataset might contain the coherent augmented image patches that have already "seen" by the model. Hence, the datasets of the proposed ConvNet are quantified depending on the number of nodules. The proposed ConvNet is trained by 30 benign and 30 malignant. However, each nodule generates a various number of ROIs based on different sizes of nodule volume, so the training data size is imbalanced for each class (1,496 benign images and 2,252 malignant images). In general, a model trained by an imbalanced dataset tends to have poor classification performance compared to the model trained by a balanced dataset. According to previous studies on imbalanced medical imaging database, the imbalanced problem has occurred when the

majority class has 1,000 times more samples than the minority class [20, 21]. Although the database used in the proposed CAD system is significantly smaller, the imbalanced problem for the proposed database is still ambiguous.

As the way to deal with imbalanced class size in the training phase is very important, additional experiments are done to explore the effects of the imbalanced training data. The imbalanced data problem is investigated by evaluating the classification performance of the proposed ConvNet trained by a larger imbalanced or manually balanced dataset via over-sampling or under-sampling. In section 5.6.2, the results are compared to the proposed model trained by the original training dataset. The weighted cross-entropy is a popular method to overcome imbalanced problem [10]. The weighted cross-entropy function is applied to the proposed ConvNet, and the classification performances between weighted and conventional cross-entropy are compared in section 5.6.3.

5.6.1 Preparation of a Larger Imbalanced Dataset

To justify whether the proposed ConvNet has data imbalance problem, the number of difference between malignant and benign class is increased on purpose. This is done by performing additional data augmentation so that the training dataset is more imbalanced. During data augmentation, each extracted ROI is further over-sampled by flipping along horizontal and vertical axises; flipped image patches are also applied rotation transformations by rotating 90, 180 and 270 degrees. The data distributions for the original database and further augmented database are presented in Table. 5.4. In contrast with the original training dataset, the difference between malignant and benign nodule samples is increased approximately 3 times. In addition to the testing dataset, the number of data samples is increased due to the additional data augmentation among all ROIs.

An important question that needs to be considered at first is whether the proposed ConvNet trained with the original training dataset keeps the same performance when the testing data is further augmented. Based on the testing result under further augmented testing dataset, the proposed ConvNet trained with original training dataset achieved a sensitivity of 0.880 and specificity of 0.876 at the optimal cut-off point of ROC curve with AUC of 0.920 (as shown in Table 5.5). In contrast to the original testing dataset, further augmented

Table 5.4: Data distribution for original database and further augmented database

	Number of malignant samples	Number of benign samples
Training dataset		
Original	2,252	1,496
Further augmented	6,756	4,488
Testing dataset		
Original	432	344
Further augmented	1,296	1,032

testing dataset resulted in the same AUC when the model was trained by the original training dataset. However, it is noticed that the sensitivity and specificity decreased. This is in a situation of moved classification threshold respected to the distribution changes of prior probabilities in different classes when additional spatial transformations are applied to the testing samples. To demonstrate the discriminative power of the classifier, AUC is more intuitive to be used as the criterion of the estimating the classification performance because the AUC measures an aggregate performance of classifier regarding all possible classification thresholds. Therefore, the proposed ConvNet trained with original training dataset and tested with further augmented testing dataset shows equally discriminative power when the trained model is tested by the original testing dataset.

5.6.2 Under-sampling and Over-sampling

According to the previous studies of medical image analysis [42, 40], AUC analysis is applied to analyze the performances of the classifiers trained by imbalanced datasets. The impact of data imbalance problem in the proposed method is explored by comparing the AUCs between the model trained with further augmented training dataset and the model trained with original training dataset. By comparing with the model trained by the original training dataset, the trained model under further augmented training dataset demonstrates data imbalanced problem due to deterioration of AUC from 0.920 to 0.901. To overcome the imbalanced data problem, under-sampling and over-sampling are applied.

Under-sampling is capable of dealing with the imbalance problem by randomly remove samples in the majority class until the numbers matched the minority class [28]. To overcome the imbalance problem when the training dataset is further augmented, the malignant classes are reduced to the same number of the benign classes by randomly selecting the augmented malignant nodules. Once the training dataset is balanced, the trained model achieves an AUC of 0.918, but still inferior to the model trained with the original dataset. A more robust way to deal with the imbalance problem is over-sampling [38]. over-sampling is performed by maintaining all augmented malignant nodule patches while randomly duplicating samples from benign class to make the training dataset be balanced. The model trained with this over-sampled database has an AUC of 0.919.

The classification performances of the proposed ConvNet trained with original training dataset, a larger imbalanced dataset and balanced dataset via over-sampling or under-sampling are compared. All the experiment results are summarized in Table 5.5. The under-sampling and over-sampling are indeed effective to overcome data imbalance problem occurred when the proposed ConvNet is trained by further augmented training dataset. In contrast with the model trained by original training dataset, sampling based approaches resulted in a drop AUC. Because the way of under-sampling is to randomly discard the samples within majority class, particular nodule patches could be all discarded in the worst case, or a large portion of the particular nodule patches are discarded. As a consequence, the under-sampling results ineffective learning process which is against the propose of data augmentation. In addition to over-sampling, over-sampling leads to downgrade the classification performance due to the redundancy of duplication. Hence, the original data distribution resulted in no impact of data imbalance problem, and the proposed ConvNet trained by original training dataset demonstrated statically better generalization than the other models trained by larger imbalanced or manually balanced dataset. Moreover, [9] stated that the classification performance could be reduced by balancing the dataset with over/under-sampling if the original dataset is small.

Table 5.5: Comparison with imbalanced and balanced datasets

	Sensitivity	Specificity	AUC
Imbalanced	0.863	0.878	0.901
Balanced			
Under-sampling	0.887	0.849	0.918
Over-sampling	0.847	0.874	0.919
Proposed method	0.880	0.876	0.920

Imbalance, an imbalanced training dataset generated by additional spatial transformations. Under-sampling, a balanced training dataset via under-sampling. Over-sampling, a balanced training dataset via over-sampling. Proposed method, proposed model trained by original training dataset and tested by the further augmented testing dataset with additional augmented samples such as flipping.

5.6.3 Weighted Cross-entropy

In addition to balancing the training dataset via under-sampling or over-sampling, the weighted cross-entropy is applied in the proposed ConvNet to overcome the imbalance problem by introducing the weight into the conventional cross-entropy function. Because the conventional cross-entropy is not capable of rescaling the loss, the conventional cross-entropy could fail to measure the loss for minority class when the amount of the majority class significantly exceeds the minority class. As a consequence, the model stops updating the learnable parameters corresponding to the minority class, and over-fitting occurs.

Unlike conventional cross-entropy function, weighted cross-entropy is able to rescale the loss by up-scaling or down-scaling the positive error with regard to the left term of the cross-entropy function (Equation 2.8). Under further augmented training dataset, the model trained with weighted cross entropy has better AUC performance compared with the model trained with conventional cross-entropy loss function. The best testing result under AUC is 0.916 when weight is adjusted to 1.5. However, the testing result under AUC is still inferior compared with the model trained with original training dataset and conventional

cross-entropy loss function. According to [10], the weighted cross entropy is used to stabilize training loss when the model is trained by an extreme imbalanced dataset. In conclusion, weighted cross entropy is able to increase performance when training data is exceeded a certain portion of imbalance ratio. However, the original training dataset does not have this problem. Moreover, evaluated sensitivity and specificity for the proposed method are fairly close, so the trained model has well-distributed weights among two classes. Therefore weighted cross entropy might not be the best option to train the ConvNet under the original training dataset.

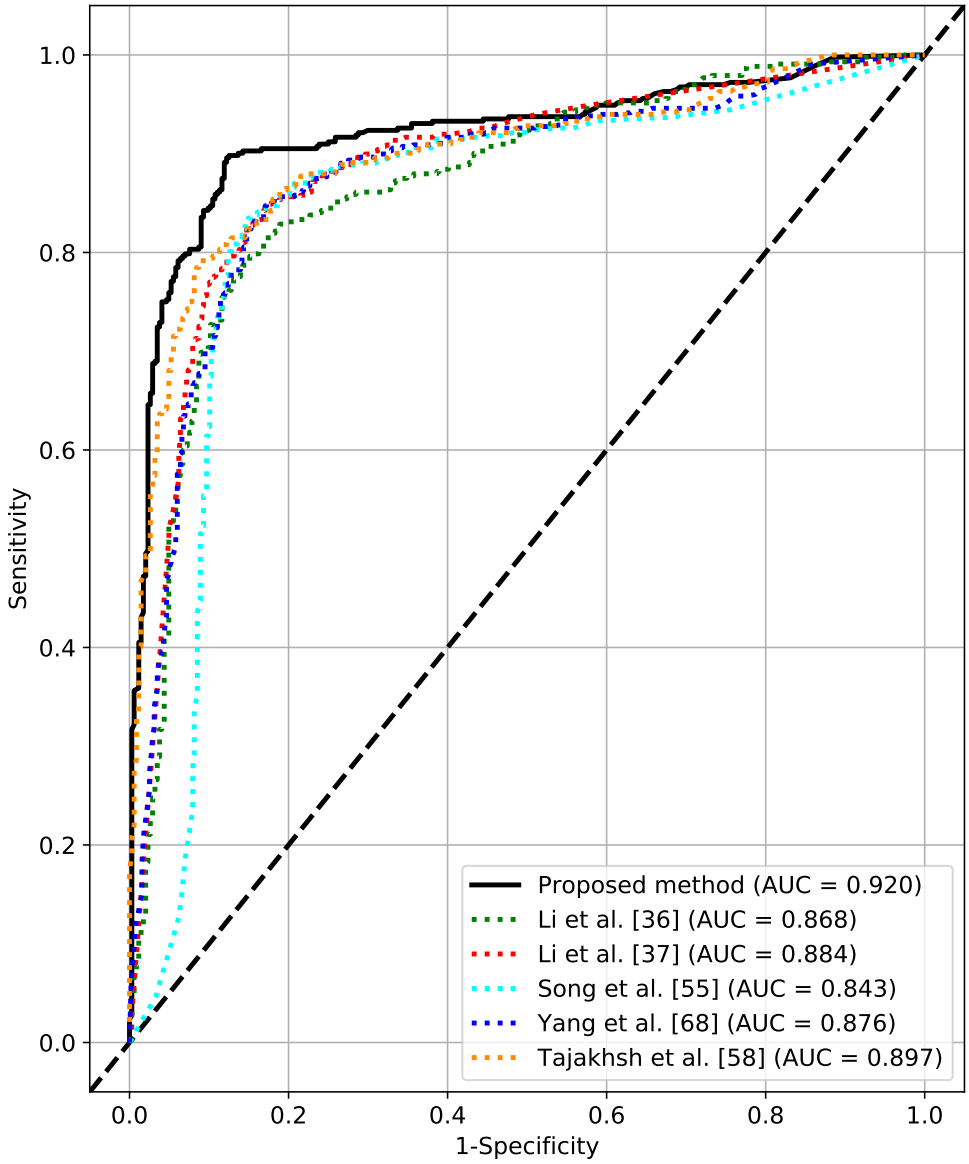


Figure 5.3: ROC curves for different ConvNet architectures. The AUCs of ROC curves are: Li et al. [36], 0.868; Li et al. [37], 0.884; Song et al. [55], 0.843; Yang et al. [68], 0.876; Tajakhs et al. [58], 0.897; proposed method, 0.920. ROC, receiver operating characteristic; AUC, area under the curve.

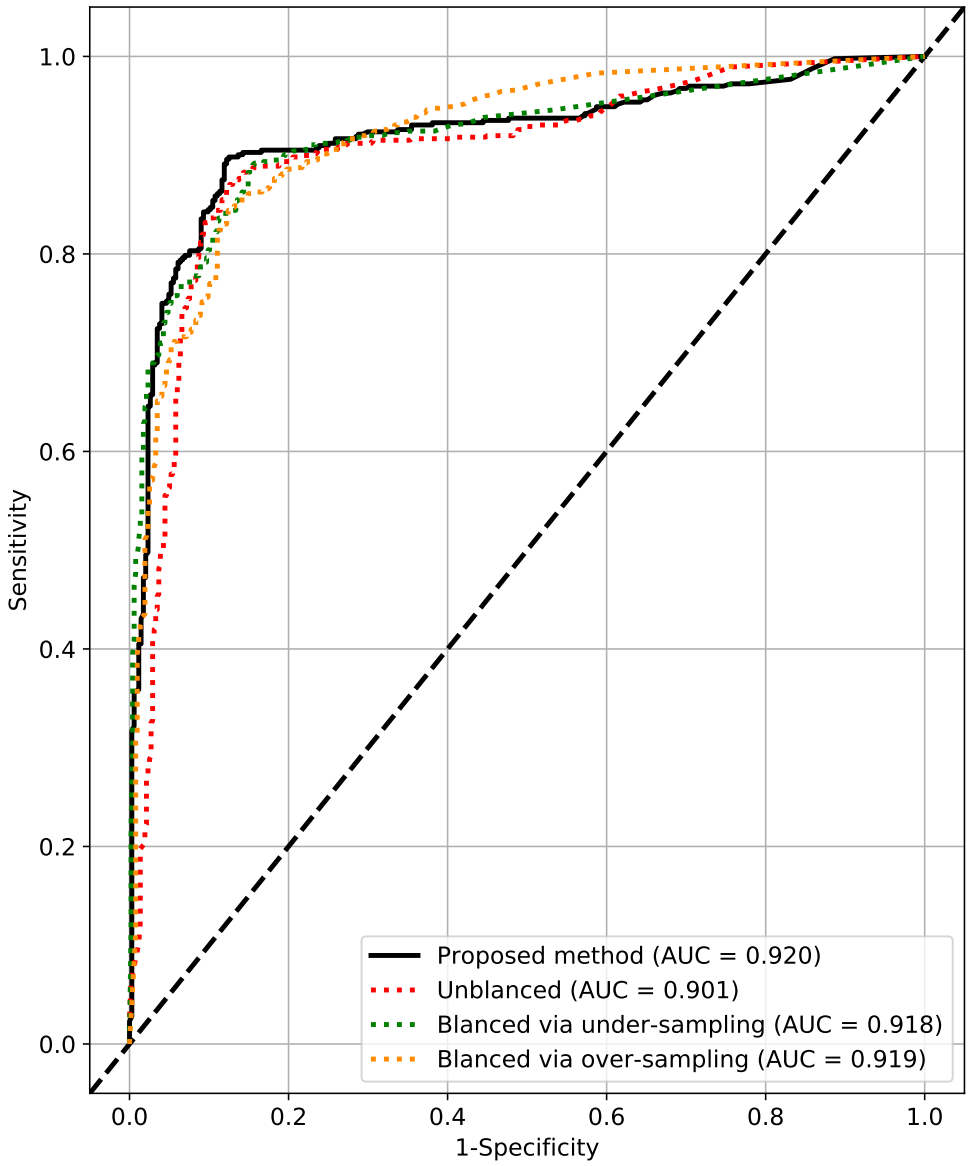


Figure 5.4: ROC curves for different distributions of the training dataset. Imbalance, an imbalanced training dataset generated by additional spatial transformations. Under-sampling, a balanced training dataset via under-sampling. Over-sampling, a balanced training dataset via over-sampling. Proposed method, proposed model trained by original training dataset and tested by the further augmented testing dataset with additional augmented samples such as flipping.

Chapter 6

Conclusion

6.1 Conclusion

In this thesis, a novel computer-aided detection (CAD) system is proposed for the diagnosis of the malignant pulmonary nodule on computerized tomography (CT) by using patch-based imaging features. Instead of designing a particular feature extraction scheme to describe the high-level features of the pulmonary nodule, the proposed CAD system directly learns them from the process of supervised learning. The architecture of the proposed Convolutional Neural Network (ConvNet) is carefully tailored to capture high-level discriminative features of the pulmonary nodule. The proposed ConvNet is composed of a series of 3 convolutional layers with the filter size of 5×5, 3 max-pooling layers and 3 ReLU activation layers, followed by 2 fully-connected layers. The model is trained automatically end-to-end in a supervised manner by minimizing cross-entropy via stochastic gradient descent (SGD), a non-adaptive optimized. The performance for diagnosis of the malignant pulmonary nodule on CT is evaluated under LUNGx Challenge database. The proposed CAD system illustrated better diagnostic performance than the unsupervised approach [46] and other previous works adopted ConvNets [36, 37, 55, 68, 58].

During the LUNGx Challenge event, the submitted CAD systems resulted in worse diagnostic performance than the experienced radiologists for malignancy detection of the pulmonary nodule on CT. The LUNG Challenge database had been challenged until Nishio [46] proposed a CAD system, an unsupervised method, and the system achieved an area under ROC (AUC) of 0.837. Instead of using conventional feature extraction approaches such as histogram of CT density, local binary pattern (LBP), the unsupervised solution [46] showed high-level discriminative feature extraction ability and better diagnosis of the ma-

lignant pulmonary nodule on CT by using a combination of principal component analysis (PCA), convolution and pooling operations. However, the proposed CAD system in this thesis demonstrated more promising diagnostic results in a supervised way. The proposed CAD system achieved a sensitivity of 0.896 and specificity of 0.878 at the optimal cut-off point of the ROC curve with AUC of 0.920. By comparing with the unsupervised method [46] under the same database, the evaluated results for the diagnosis of malignant pulmonary nodule achieve 10% AUC improvement.

In contrast with other ConvNet solutions designed for the classification of lung patterns [36, 37, 55, 68, 58], the proposed CAD system in this thesis outperformed the malignancy detection of the pulmonary nodule on CT under same LUNGx Challenge database. The malignancy detection performance for the shallow ConvNet [36], adopted a single convolutional layer, achieved 4% AUC improvement compared to the unsupervised method proposed by Nishio [46], but the classification performance of the shallow ConvNet was inferior to the proposed ConvNet. The shallow ConvNet cannot express the full potential of high-level discriminative feature extractions through the hierarchical layered structure by employing a single convolutional layer. Also, the redundant hidden neurons in Fully-connected (FC) layer resulted in over-fitting. The ConvNet [37] with two grouped convolutional layers improved AUC by approximately 2% compared to the Shallow ConvNet, but the proposed ConvNet still demonstrated better AUC because the proposed ConvNet employs a relatively small filter respected to the input size in order to pass enough invariant information and prevent losses of the local sparse structure. The ConvNet [55] with two convolutional layers showed the worse AUC among all ConvNet solutions because the total receptive field in the ConvNet [55] exceeds the input receptive field, so the filters failed to prevent the network to capture non-local information. In contrast with the ConvNet adopted the same amount of convolutional layers [68], the proposed ConvNet improved AUC by 5% due to none of the FC layer. In addition to the transfer learning by fine-tuning AlexNet model [58], the proposed ConvNet surpassed the AlexNet model by improving 2% AUC. The AlexNet model consists of 37 million learnable parameters, the enormous numbers of parameters lead the trained model to achieve local optima.

In addition to clinical usage, a CAD system does not only need to provide a promising

diagnostic performance, but it also has to be easy to be used. In contrast with the proposed CAD system, the previous unsupervised approach for diagnosis of the malignant pulmonary nodule on CT requires multiple system blocks such as feature transformation, feature extraction, and feature classification. Since the proposed ConvNet is trained end-to-end in a supervised manner, the proposed system is easy to perform diagnosis of the malignant pulmonary nodule on CT in a single system. On the other hand, the proposed CAD system takes raw images as inputs. Unlikely, the previous works [2, 41] require segmented nodule as inputs, or additional low-level imaging features of the pulmonary nodules (e.g., density, size or shape) are employed as inputs [12]. Regarding the input format, the proposed CAD system performs diagnosis of the malignant pulmonary nodule on CT without any manual intervention.

6.2 Future Work

Although the proposed CAD system achieved more accurate classification result than the current state-of-the-art works, augmented samples used to train ConvNet are still significantly less than natural image database (e.g., ImageNet). The limited number of training samples can be expanded by adopting other lung databases. By increasing the training samples, there will be uncertainty that the proposed ConvNet would have better performance or worse. According to the experiments conducted by Zhu et al. [70], increasing the complexity of ConvNet with a large amount of training data would not be the optimal solution. The future work would be worth to include evaluating the diagnostic performance for the malignancy detection of the pulmonary nodule on CT under other lung databases. For example, Lung Image Database Consortium (LIDC) public database [5] consists of 1,018 patient cases with 2,636 pulmonary nodules. Besides, the LIDC database has various of slice thicknesses from 0.6 mm to 5.0 mm, so it also would be worth to justify the classification performance of the proposed ConvNet when the representations of the inputs are not standardized.

In addition to the future work for the architecture of the proposed ConvNet, a deeper architecture would be worth to evaluate by adopting ideas of residual blocks [24] or inception blocks [57]. In the sense of training methods, hard negative mining, related to the imbalanced

problem, would also be worth to be further justified by using larger lung database. Moreover, it would be also worth to compare the classification performances with different shapes of the input data since some recent studies demonstrate state-of-the-art performances for malignancy detection by analyzing the volumetric patterns of the pulmonary nodule instead of patch-based features or using 2.5D based input nodule data by concatenating the nodule image patches in different views such as sagittal, coronal and axial views [51, 70].

The cropping sizes of the extracted ROIs in this thesis are determined by manual selections. The way of extracting ROI could be done in an automatically fashion in the future. Region-based convolutional neural networks (R-CNNs) could be the one to replace the proposed ROI extraction approach. R-CNN is a fully convolutional network that predicts boundaries of objects by applying region proposal networks (RPNs). RPNs share convolutional layers with region proposal methods for state-of-the-art object detection. For example, SPPnet [23] and Fast R-CNN [48] have demonstrated inexpensive computational cost and obtained promising detection accuracy.

References

[1] American Cancer Society. Cancer facts & figures 2017. Available online: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html, 2017.

[2] Masahito Aoyama, Qiang Li, Shigehiko Katsuragawa, Feng Li, Shusuke Sone, and Kunio Doi. Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose ct images. *Medical Physics*, 30(3):387–394, 2003.

[3] Samuel G. Armato, Karen Drukker, Feng Li, Lubomir Hadjiiski, Georgia D. Tourassi, Justin S. Kirby, Laurence P. Clarke, Roger M. Engelmann, Maryellen L. Giger, George Redmond, and Keyvan Farahani. Lungx challenge for computerized lung nodule classification. *Journal of Medical Imaging*, 3:3–9, 2016.

[4] Samuel G. Armato, Maryellen Lissak Giger, Catherine J. Moran, Heber MacMahon, and Kunio Doi. Automated detection of pulmonary nodules in helical computed tomography images of the thorax. *Proc.SPIE*, 3338:4, 1998.

[5] Samuel G. Armato, Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng Zhao, Denise R. Aberle, Claudia I. Henschke, Eric A. Hoffman, Ella A. Kazerooni, Heber MacMahon, Edwin J. R. van Beek, David Yankelevitz, Alberto M. Biancardi, Peyton H. Bland, Matthew S. Brown, Roger M. Engelmann, Gary E. Laderach, Daniel Max, Richard C. Pais, David P-Y Qing, Rachael Y. Roberts, Amanda R. Smith, Adam Starkey, Poonam Batra, Philip Caligiuri, Ali Farooqi, Gregory W. Gladish, C. Matilda Jude, Reginald F. Munden, Iva Petkovska, Leslie E. Quint, Lawrence H. Schwartz, Baskaran Sundaram, Lori E. Dodd, Charles Fenimore, David Gur, Nicholas Petrick, John Freymann, Justin Kirby, Brian Hughes, Alessi Vande Casteele, Sangeeta Gupte, Maha Sallam, Michael D. Heath, Michael H. Kuhn, Ekta Dharaiya, Richard Burns, David S. Fryd, Marcos Salganicoff, Vikram Anand, Uri Shreter, Stephen Vastagh, Barbara Y. Croft, and Laurence P. Clarke. The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Med Phys*, 38(2):915–931, Feb 2011.

[6] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan. Chest pathology detection using deep learning with non-medical training. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 294–297, Apr 2015.

[7] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[8] Léon Bottou. Online learning in neural networks. chapter On-line Learning and Stochastic Approximations, pages 9–42. Cambridge University Press, New York, NY, USA, 1998.

[9] Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *CoRR*, abs/1710.05381, 2017.

[10] Zhaowei Cai, Quanfu Fan, Rogério Schmidt Feris, and Nuno Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. *CoRR*, abs/1607.07155, 2016.

[11] Ronald A. Castellino. Computer aided detection (cad): an overview. *Cancer Imaging*, 5(1):1719, 2005.

[12] Hui Chen, Yan Xu, Yujing Ma, and Binrong Ma. Neural network ensemble-based computer-aided diagnosis for differentiation of lung nodules on ct images: Clinical evaluation. *Academic Radiology*, 17(5):595–602, 2010.

[13] Francesco Ciompi, Bartjan de Hoop, Sarah J. van Riel, Kaman Chung, Ernst Th. Scholten, Matthijs Oudkerk, Pim A. de Jong, Mathias Prokop, and Bram van Ginneken. 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. *Medical Image Analysis*, 26(1):195–202, 2015.

[14] Elizabeth R. DeLong, David M. DeLong, and Daniel L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. 44(3):837–845, 1988.

[15] J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009.

[16] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, July 2011.

[17] Bradley Efron. *An introduction to the bootstrap*. Monographs on statistics and applied probability (Series) ; 57. Chapman & Hall, New York, 1993.

[18] Maryellen L. Giger, Kyongtae T. Bae, and Heber Macmahon. Computerized detection of pulmonary nodules in computed tomography images. *Investigative Radiology*, 29(4):459465, 1994.

[19] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[20] Jerzy W. Grzymala-Busse, Linda K. Goodwin, Witold J. Grzymala-Busse, and Xinqun Zheng. An approach to imbalanced data sets based on changing rule strength. In *Rough-Neural Computing: Techniques for Computing with Words*, 2004.

[21] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73:220–239, 2017.

[22] Yong Fan Hancan Zhu, Hewei Cheng. Random local binary pattern based label learning for multi-atlas segmentation. *Proc.SPIE*, 9413:8, 2015.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *CoRR*, abs/1406.4729, 2014.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[25] Kai-Lung Hua, Che-Hao Hsu, Shintami Chusnul Hidayati, Wen-Huang Cheng, and Yu-Jen Chen. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Onco Targets Ther*, 8:2015–2022, Aug 2015.

[26] Armato III, Lubomir Samuel G.and Hadjiiski, Georgia D.and Karen Drukke Tourassi, Maryellen L. Giger, Feng Li, George Redmond, Keyvan Farahan, Justin S. Kirby, and Laurence P. Clarke. Spie-aapm-nci lung nodule classification challenge dataset, 2015.

[27] Colin Jacobs, Eva M. van Rikxoort, Thorsten Twellmann, Ernst Th. Scholten, Pim A. de Jong, Jan-Martin Kuhnigk, Matthijs Oudkerk, Harry J. de Koning, Mathias Prokop, Cornelia Schaefer-Prokop, and Bram van Ginneken. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Medical Image Analysis*, 18(2):374–384, 2014.

[28] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intell. Data Anal.*, 6(5):429–449, Oct 2002.

[29] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In *2009 IEEE 12th International Conference on Computer Vision*, pages 2146–2153, Sept 2009.

[30] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[31] Ayano Kamiya, Sadayuki Murayama, Hisashi Kamiya, Tsuneo Yamashiro, Yasuji Oshiro, and Nobuyuki Tanaka. Kurtosis and skewness assessments of solid lung nodule density histograms: differentiating malignant from benign nodules on ct. *Japanese Journal of Radiology*, 32(1):14–21, Jan 2014.

[32] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[33] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1097–1105, USA, 2012. Curran Associates Inc.

[34] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Comput.*, 1(4):541–551, Dec 1989.

[35] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.

[36] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen. Medical image classification with convolutional neural network. In *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, pages 844–848, Dec 2014.

[37] Wei Li, Peng Cao, Dazhe Zhao, and Junbo Wang. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Comput Math Methods Med*, 2016:6215085, Dec 2016.

[38] Charles X. Ling and Chenghui Li. Data mining for direct marketing: Problems and solutions. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, KDD'98, pages 73–79. AAAI Press, 1998.

[39] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Snchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[40] Marcus A. Maloof. Learning when data sets are imbalanced and when costs are unequal and unknown. 2003.

[41] Yuichi Matsuki, Katsumi Nakamura, Hideyuki Watanabe, Takatoshi Aoki, Hajime Nakata, Shigehiko Katsuragawa, and Kunio Doi. Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution ct. *American Journal of Roentgenology*, 178(3):657–663, Mar 2002.

[42] Maciej A. Mazurowski, Piotr A. Habas, Jacek M. Zurada, Joseph Y. Lo, Jay A. Baker, and Georgia D. Tourassi. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*, 21(2):427–436, 2008. Advances in Neural Networks Research: IJCNN 07.

[43] Annette McWilliams, Martin C. Tammemagi, John R. Mayo, Heidi Roberts, Geoffrey Liu, Kam Soghrati, Kazuhiro Yasufuku, Simon Martel, Francis Laberge, Michel Gingras, Sukhinder Atkar-Khattra, Christine D. Berg, Ken Evans, Richard Finley, John Yee, John English, Paola Nasute, John Goffin, Serge Puksa, Lori Stewart, Scott Tsai, Michael R. Johnston, Daria Manos, Garth Nicholas, Glenwood D. Goss, Jean M. Seely, Kayvan Amjadi, Alain Tremblay, Paul Burrowes, Paul MacEachern, Rick Bhatia, Ming-Sound Tsao, and Stephen Lam. Probability of cancer in pulmonary nodules detected on first screening ct. *New England Journal of Medicine*, 369(10):910–919, 2013.

[44] Temesguen Messay, Russell C. Hardie, and Steven K. Rogers. A new computationally efficient cad system for pulmonary nodule detection in ct imagery. *Medical Image Analysis*, 14(3):390–406, 2010.

[45] K. Murphy, B. van Ginneken, A.M.R. Schilham, B.J. de Hoop, H.A. Gietema, and M. Prokop. A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification. *Medical Image Analysis*, 13(5):757–770, 2009.

[46] Mizuho Nishio and Chihiro Nagashima. Computer-aided diagnosis for lung cancer: Usefulness of nodule heterogeneity. *Academic Radiology*, 24(3):328–336, 2017.

[47] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors (Basel)*, 16(1):115, Jan 2016.

[48] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

[49] Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016.

[50] II Samuel G. Armato, Maryellen L. Giger, Catherine J. Moran, James T. Blackburn, Kunio Doi, and Heber MacMahon. Computerized detection of pulmonary nodules on ct scans. *RadioGraphics*, 19(5):1303–1311, 1999.

[51] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Snchez, and B. van Ginneken. Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, 35(5):1160–1169, May 2016.

[52] Jonathon Shlens. Notes on kullback-leibler divergence and likelihood. *CoRR*, abs/1404.2000, 2014.

[53] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484 –489, Jan 2016. Article.

[54] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. volume abs/1409.1556, 2014.

[55] Qingzeng Song, Lei Zhao, Xingke Luo, and Xuechen Dou. Using deep learning for classification of lung nodules on computed tomography images. *Journal of Healthcare Engineering*, 2017:7, Aug 2017.

[56] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, January 2014.

[57] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.

[58] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, May 2016.

[59] The International Early Lung Cancer Action Program Investigators. Survival of patients with stage i lung cancer detected on ct screening. *New England Journal of Medicine*, 355(17):1763–1771, 2006.

[60] The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011.

[61] Patrick Therasse, Susan G. Arbuck, Elizabeth A. Eisenhauer, Jantien Wanders, Richard S. Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T. van Oosterom, Michaele C. Christian, and Steve G. Gwyther. New guidelines to evaluate the response to treatment in solid tumors. *JNCI: Journal of the National Cancer Institute*, 92(3):205–216, 2000.

[62] T. Tieleman and G. Hinton. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.

[63] B. van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 286–289, Apr 2015.

[64] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, Dec 2010.

[65] Sun-Chong Wang. *Artificial Neural Network*, pages 81–100. Springer US, Boston, MA, 2003.

[66] Ashia C. Wilson, Rebecca Roelofs, Mitchell Stern, Nathan Srebro, and Benjamin Recht. The marginal value of adaptive gradient methods in machine learning. *CoRR*, abs/1705.08292, 2017.

[67] Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *CoRR*, abs/1611.05431, 2016.

[68] H. Yang, H. Yu, and G. Wang. Deep Learning for the Classification of Lung Nodules. *ArXiv e-prints*, Nov 2016.

[69] F. Zhang, Y. Song, W. Cai, M. Z. Lee, Y. Zhou, H. Huang, S. Shan, M. J. Fulham, and D. D. Feng. Lung nodule classification with multilevel patch-based context analysis. *IEEE Transactions on Biomedical Engineering*, 61(4):1155–1166, April 2014.

[70] Xiangxin Zhu, Carl Vondrick, Charless C. Fowlkes, and Deva Ramanan. Do we need more training data? *CoRR*, abs/1503.01508, 2015.

[71] M H Zweig and G Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, 1993.