Computation-Efficient CNN System for High-Quality Lung Nodule Detection

Yijian Zhao

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science at

Concordia University

Montreal, Quebec, Canada

August 2023

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Yijian Zhao

Entitled: A Computation-Efficient CNN System for High-Quality lung nodule detection

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Electrical and Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. M. Omair Ahmad

_____ External Examiner
Dr. Hang Xu

_____ Internal Examiner
Dr. M. Omair Ahmad

_____ Supervisor
Dr. Chunyan Wang

Approved by: _____
Dr. Y.R.Shayan, Chair
Department of Electrical and Computer Engineering

_____
Dr. Mourad Debbabi, Interim Dean,
Gina Cody School of Engineering and Computer Science

Abstract

A Computation-Efficient CNN System for High-Quality Lung Nodule Detection

Yijian Zhao

Lung cancer diagnosis is a critical healthcare issue, and fully automated lung nodule detection is desirable for a timely diagnosis. However, due to the variability in shapes, sizes, textures, and locations in lung nodules, developing a computer vision system for this detection is a very challenging task.

In this thesis, a special CNN system is proposed for lung nodule detection. It consists of 2 stages, namely Stage A and Stage B. Stage A is designed to localize the nodule candidates, aiming at a high sensitivity in order to minimise the miss rate. Stage B is to identify the true nodules from the input samples. It can be used to identify falsely detected nodule samples from the output of Stage A, and also as a stand-alone lung nodule recognition system.

In Stage A, there are three blocks, i.e., a pre-processing block, custom-design CNN block and refinement block. The core of this stage is the custom-designed and U-net-based CNN block. The filtering modules in its convolution layers are specifically designed to suit the features produced in these layers. To reduce the data loss in the first four layers, Full-ReLU is used as the activation function. Furthermore, the refinement block is placed to reduce effectively the false positive rate. The computation complexity of Stage A is very low, as its total number of trainable parameters is only 0.16 M. Stage A delivers a high detection rate of 95.38% but the false positive rate is still as high as 6.9 FPs/scan. The output data will be applied to Stage B for further processing.

The design of Stage B is focused on distinguishing between the true nodules and their look-likes. Based on our analysis on the characters carried by nodules of different sizes, we propose to have 2 networks in Stage B for large and small nodule categories, respectively. The feature extraction in the 2 CNNs should be different, one targeting the variations in object regions of large nodules and the other looking more into nodule surroundings in case of small nodules. Two CNNs have been designed and each of them has a particular multi-branch feature extraction (FE)

block for the designated nodule category. Each CNN also involves fully-connected layers for classification. Stage B has been tested as a stand-alone lung nodule recognition system on LUNA 16 dataset. The results demonstrate that, with respect to similar systems found recently in literature, Stage B provides a good processing quality at a computation cost that is only a very small fraction of that needed by others.

The complete system for lung nodule detection, i.e., Stage A and Stage B combined, has also been tested with the same dataset. The results demonstrate the good functionality of the system. All these CNNs combined require 0.7M parameters, far less that other CNN systems performing the same task.

In summary, the proposed system has been custom-designed to optimize the computation efficiency, i.e., achieving a good detection quality at the lowest computation cost. To attain this goal, the design strategy is to decompose the complex task of lung nodule detection into subtasks so that the system can employs multiple simple CNNs, each performing a sub-task. In this way, each CNN can be structured to suit the characters of a particular kind of nodule data and optimised to meet specific performance requirements. The effectiveness of this strategy has been confirmed by the results of the performance evaluation. Because of its low computation cost, the proposed system can be very easily implemented in various environment.

Acknowledgements

I would like to express my best appreciation to my supervisor Dr. Chunyan Wang for her support during my study in Concordia University. She always gives me brilliant guidance. Moreover, her focus on details in research influences me deeply. It is my great honor and fortune to work under her supervision.

I would also like to thank my colleagues, Yunlong Ma, Yanming Sun, and Mingze Ni, for the help in the last two years.

Additionally, thank Compute Canada and Calcul Quebec for the powerful servers.

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

ADAM	Adaptive Moment Estimation
AFF	Adaptive Feature Fusion
ASCNN	Application-Specific Convolutional Neural Network
BCE	Binary Cross-Entropy
BN	Batch Normalization
CNN	Convolutional Neural Network
CT	X-Ray Computed Tomography
CPM	Competition Performance Metric
FCN	Fully Convolutional Network
FN	False Negative
FP	False Positive
FPN	Feature Pyramid Network
FROC	Free-Response Operation Characteristic
IN	Instance Normalization
LN	Layer Normalization
LUNA 2016	Lung Nodule Analysis
ReLU	Rectified Linear Unit
ROI	Region of Interest
RPN	Residual Network Applied to Feature Pyramid Network
SdcNet	Successive Depth-wise Network
SGD	Stochastic Gradient Descent
TP	True Positive
VGG	Visual Geometry Group

List of Symbols

l	Length of a 2D image
m	Width of a 2D image.
n	Number of consecutive axial slices
N	Number of batch size
N_s	Number of slices of each Stage B input
$P(y_i)$	Probability of y_i
$W(k_1, k_2)$	2D kernel
$x_{00}, x_{01}, \ldots, x_{33}$	Input of pooling operation
x	Input of activation function
$x(i, j)$	Input image of 2D convolution
x_{ij}	Input of the Full-ReLU
$x_{n\,ij}$	Output after applying Full-ReLU to the negative part of the input data
$x_{p\,ij}$	Output after applying Full-ReLU to the positive part of the input data
Xn	Normalized value of each element
$y_{00}, y_{01}, y_{10}, y_{11}$	Output of pooling operation
y_i	Output of the CNN
μ	Mean value of each element
σ	Standard deviation value of each element

1. Introduction

1.1 Background and Challenges

Lung cancer is one of the leading causes of cancer deaths[1] and lung screening can help to diagnose the disease in an early stage before any symptom appears. Lung nodules are clumps of cells inside a lung. As some of them can be cancerous, in particular in case of growing nodules, lung nodules are detected and monitored in lung screening. Computed tomography (CT) is widely used for lung screening. Manually detecting lung nodules from CT images is a difficult and time-consuming job, requiring knowledge of medical specialists. Limited resources for lung CT image examination may result in lengthy waiting for lung diagnosis results, affecting timely treatment of patients. Hence, it is important to develop automated lung nodule detection systems by computer vision to reduce the workload of medical doctors and improve the survival rate of lung cancer patients by early diagnosis and treatments.

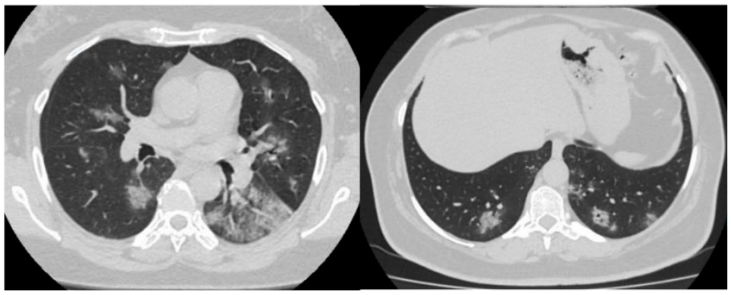


Figure 1.1 Examples of image slices sampled from a 3D CT scan.[2]

Human lung is a complex organ with varieties of vessels, bronchioles, muscles and other tissues appearing in its 3D structure. Two slices from a 3D lung CT image are shown in Figure 0.1, and one can see that there are a lot of variations in lung lobs and the image also includes other parts of the chest. A nodule can appear as a knobby mass of different sizes, shapes and solidities. A number of nodule samples are illustrated in Figure 0.2. While a lot of variations in patterns can be seen in large nodules, small nodules appear as simple spots in lung images. Nodule-like spots can be seen everywhere but most of them are not true ones. Hence, it is very difficult to achieve a high detection rate without including a large number of false positive objects.

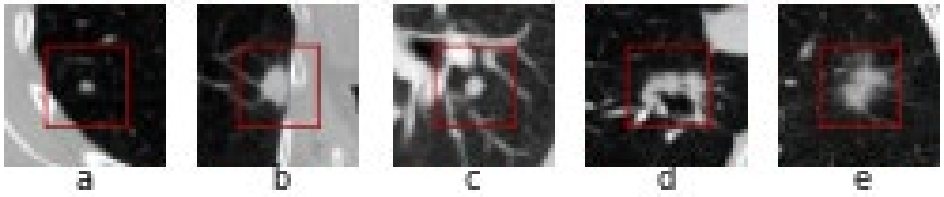


Figure 1.2 Samples of 5 different nodules.

Many lung detection systems so far reported have been developed with convolutional neural networks (CNN) [3]. The parameters of the convolutional kernels in a CNN are not determined manually, but by a training process with training samples. To handle the complex detection task, such as nodule detection, a CNN can have a large number of filtering layers and each layer has a very large number of filtering kernels. However, the larger number of the parameters, the more data samples are needed for a decent training of the system. If the quantity and quality of the training samples is limited, which is often the case, the quality of lung nodule detection of a CNN system does not necessarily get better if the system gets wider and deeper.

1.2 Motivation and Objective

As mentioned previously, lung nodule detection is a complex task. There are so many variations in appearances of nodules that it is very difficult to detect them by using knowledge-based filtering process. Many attempts to use CNNs in order to gain more computation power to do the job. It should be underlined that a lot of normal knobby tissues in lung often bear a high-degree resemblance to true lung nodule spots. Hence, in lung nodule detection, the problem of false positive is too serious to be ignored, which makes the design of nodule detection systems more challenging.

Computation power of a CNN is related to its scale of convolutions, often measured by the number of learnable parameters. The processing quality delivered by a CNN system is, however, related not only to its computation power, but also its computation efficiency regarding its processing task. Increasing the complexity of convolutions in a CNN does not necessarily lead to a better processing, but certainly makes the problem of insufficient training data more pronounced and the implementation of the system more restricted by the limitation of computation resources. Hence, achieving a high computation-efficiency, i.e., obtaining a high

processing quality while minimising the computation volume, is a critical issue in designing high-performance CNN systems.

The objective of the work presented in this thesis is to design and implement a CNN system detecting lung nodules from 3D CT lung images, aiming at a high processing quality at low computation cost. To achieve the objective, we propose a CNN system designed specifically for the lung nodule detection, and the computation efficiency is a key word throughout the design process.

1.3 Scope and Organization

The proposed system is an application specific CNN (ASCNN) system, fully custom-designed to detect the varieties of lung nodules. As there are 2 challenges in the detection, (i) very different appearances of the true nodules and (ii) similar appearances of true nodules and false ones, the system has 2 parts and each perform a specific task.

- Stage A is a CNN designed to detect nodules. Its main objective is to achieve a high detection rate, minimizing the low miss rate. In other words, this CNN is expected to catch nodule candidates, i.e., true nodules and their look-likes. It is thus a CNN for nodule candidate detection.
- Stage B is designed to recognize the true nodules from its input samples of nodule candidates. It can be used as an independent CNN system for lung nodule recognition. By doing so, each part operates with its own focus to handle one of the 2 challenges mentioned above. The computation structure in each part can be specifically designed to suit its task.

The thesis is organized as follows. In Chapter 2, basic elements in CNN are reviewed and existing lung nodule detection systems relevant to the proposed one are presented. The design of Stage A is described in Chapter 3. The work concerning Stage B is presented in Chapter 4 and its design is described in detail. Chapter 5 is dedicated to the performance evaluation of Stage A and Stage B. The work of the thesis is concluded in Chapter 6.

2. Background and Relevant Work

2.1 Introduction

The complexity of lung nodule detection requires a large scale of computation and CNN systems can be effective approach to a high performance. Various CNN designs have been reported in research publications, and some of them are relevant to the work presented in this thesis.

In this chapter, the basics of CNN is described in Sub-chapter 2.2 to provide the fundamentals in the design of the proposed system. In Sub-chapter 2.3, a number of CNN systems for lung nodule detection are reviewed. All of them are reported in recent years and relevant to the work of this thesis.

2.2 Fundamentals of CNN Design

A CNN consists primarily of convolution operations along with other critical elements such as normalization, non-linear activation functions, and pooling. Unlike traditional convolution operations, the kernels in a CNN are not predetermined but rather trained using a large number of samples.

2.2.1 2D Convolution and Convolution Modes

A 2D convolution is in fact a filtering operation, it is defined as follow:

$$y(i,j) = \sum_{k1=-n}^{n} \sum_{k2=-m}^{m} w(k_1,k_2)x(i-k_1, j-k_2) \tag{2.1}$$

where w is a 2D kernel; x and y are input and output and their dimensions are ($2n$+1, $2m$+1).

There are three types of 2D filter, high-pass filter, low-pass filter and band-pass filter. "Convolution in CNNs involves trainable parameters, but even after training, the convolution can still be categorized into one of three filters. To illustrate the effect of convolutional processing on images, a non-adaptive high-pass filter, specifically the Sobel filter, could be taken as an example. Figure 2.1 provides a visual representation of this.

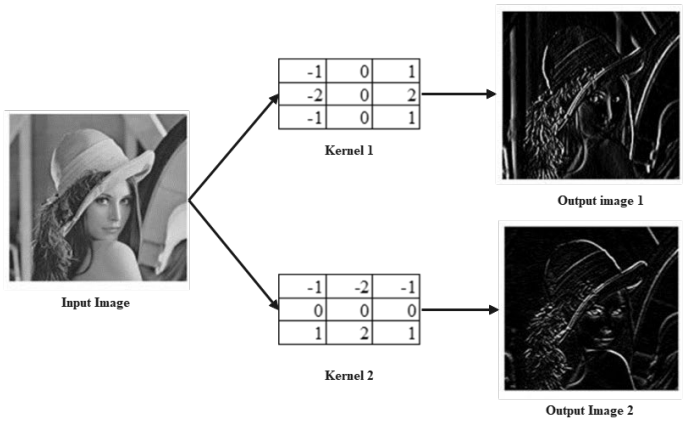


Figure 2.1 An example of 2D convolutions.

Deterministic kernels can extract specific features from input images, such as the vertical and horizontal edge detection kernels shown in Figure 0.3. However, designing a large number of deterministic kernels to extract variant features in Lung CT images is challenging. A CNN, on the other hand, uses multi-layer convolutions with multiple channels in each layer. Furthermore, unlike deterministic convolutions, CNN kernels are trainable and determined through a training process. This makes CNN a more practical choice for lung nodule detection.

As mention above, the convolution in CNN has multiple channels in each layer. There are three types of convolution modes for the channels in a layer. Assume the kernel size is 3×3, and the numbers of input and output channels are 8 and 16, respectively:

(1) Standard convolution. In this mode, each of the output channels is produced by all of the input channels, so the number of weights in this layer is 1152 ($3 \times 3 \times 8 \times 16$).

(2) Group convolution. Assume the input channels are divided into 2 groups, then a group convolution is similar to 2 small-scale standard convolutions. The number of weights in this layer is 576 ($3 \times 3 \times \frac{8}{2} \times \frac{16}{2} \times 2$).

(3) Depth-wise convolution. In this mode, each of the output channels is produced by only one of the input channels. Since the number of the output channels is twice that of the input channels, an input channel should produce two output channels. The number of weights in this layer is 144($3 \times 3 \times 16$).

2.2.2 Critical Components for CNN

To fully optimize a CNN structure, it's crucial to incorporate three essential components alongside the core convolutional kernel: normalization, a nonlinear activation function and Pooling.

(1) **Normalization**. Normalization is the process of standardizing the range and distribution of data within each layer of a network, which can significantly aid in the convergence of the network.

(i) **Batch normalization (BN)** [4]. A BN normalizes the feature maps by using the mean and standard deviation of a whole batch of data.

(ii) **Instance normalization (IN).** In contrast to BN, IN calculates the mean and variance for each channel of each sample across both spatial dimensions.

(iii) **Layer Normalization (LN)**. LN calculates the mean and variance for each individual sample across all channels and both spatial dimensions.

(2) **Non-linear activation function**. The relationship between input and output in image classification is non-linear. However, convolution, which is a fundamental operation in image classification, is a linear function. To enable a network to learn complex data and make accurate predictions, it's essential to introduce non-linearity through the use of activation functions.

(i) **Rectified Linear Unit (ReLU)** [5]. The ReLU activation function is defined as:

$$f(x) = \max(0, x) \tag{2.2}$$

It isa highly efficient non-linear function commonly used in neural networks. As depicted in Figure 2.2, the ReLU curve is defined for positive values only and facilitates backpropagation.

(ii) **Sigmoid[7]**. Sigmoid function is also called Logistic Activation Function. It is defined as:

$$f(x) = \frac{1}{(1 + e^{-x})} \tag{2.3}$$

Although the sigmoid function has a range of values from 0 to 1, input data with very high absolute values can result in a vanishing gradient problem. The curve of Sigmoid function is shown in Figure 2.3.

(3) **Pooling**. Pooling operations are used for down-sampling, increasing signal density, reducing data volume, and decreasing computation complexity. Max pooling and average pooling are common techniques used in practice.

(i) **Max pooling**. Setting the stride to 2, a max pooling operation reserves the highest value among four adjacent pixels while discarding the remaining three. This reduces the number of output pixels to one-fourth of the input pixels. Using the elements in Figure 2.4, Max-pooling can be expressed as

$$y_{00} = \max(x_{00}, x_{01}, x_{10}, x_{11}), \ldots, y_{11} = \max(x_{22}, x_{23}, x_{32}, x_{33}) \quad (2.4)$$

(ii) **Average Pooling**. During average pooling, the average value of four adjacent pixels is retained. Using the elements in Figure 2.4, Average-pooling can be expressed as:

$$y_{00} = (x_{00} + x_{01} + x_{10} + x_{11})/4, \ldots, y_{11} = (x_{22} + x_{23} + x_{32} + x_{33})/4 \quad (2.5)$$

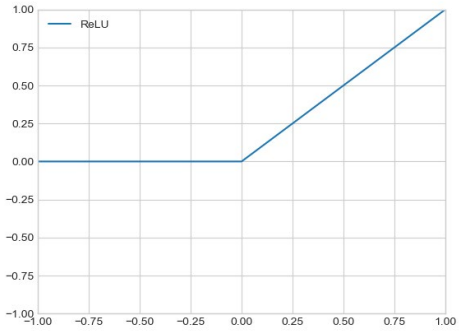


Figure 2.2 Curve of ReLU.

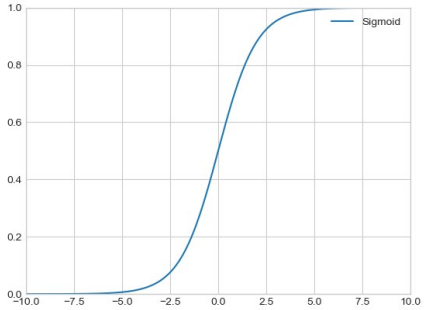


Figure 2.3 Curve of Sigmoid function.

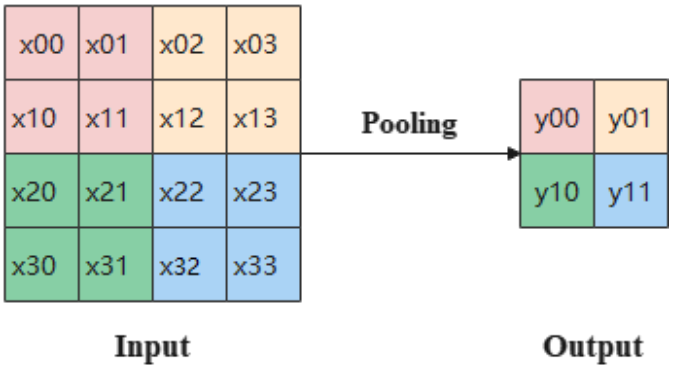
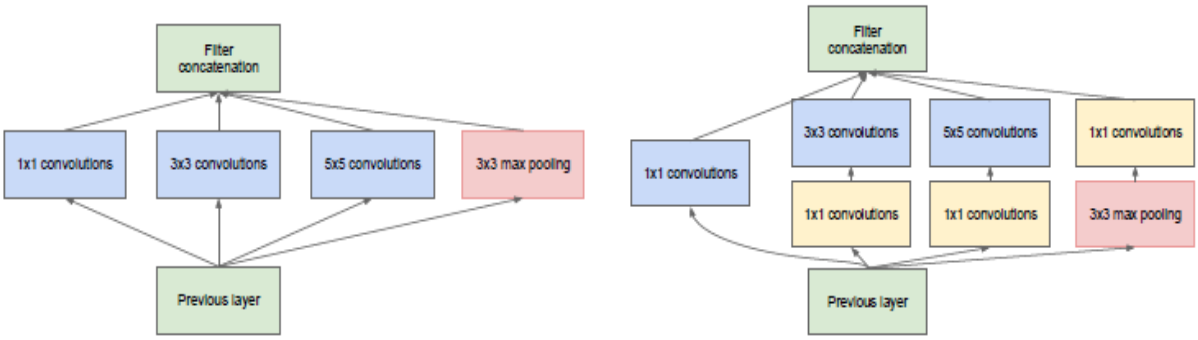


Figure 2.4 Input and Output of pooling, stride 2.

2.2.3 Commonly Used Basic Block in CNN Structures

In this subsection, some common basic blocks in CNN structures are presented. These blocks are extensively utilized to bolster the capabilities of CNN architectures.

The inception module[10], first utilized in GoodLeNet, incorporates three parallel convolution kernels of varying sizes. The larger kernel extracts global features, while the smaller kernel extracts local features, as illustrated in Figure 2.5(a). The use of different-sized convolution kernels results in different-sized perceptual fields, and combining these features through stitching achieves feature fusion across different scales. To reduce computational complexity, a 1x1 convolutional kernel is employed for dimensionality reduction, leading to the structure depicted in Figure 2.5(b).



(a) Inception module, naïve version (b) Inception module with dimension reductions

Figure 2.5 Architecture of inception module.

Deep residual block [11] was developed to address the challenges that accompany deep learning. Training very deep neural networks can be challenging, as they may lead to lower predictive accuracy. However, deep residual blocks utilize shortcut connections to enable the development of deeper networks. Figure 2.6 illustrates the architecture of this type of block.

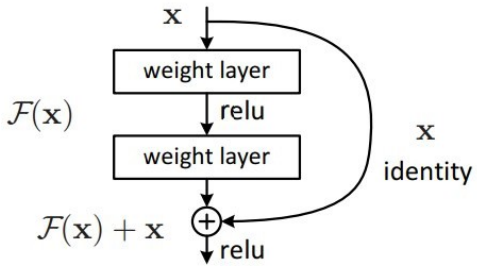


Figure 2.6 Architecture of deep residual block [11].

Densely connected block [12] has gained extensive popularity in recent years. The architecture of a densely connected block is presented in Figure 2.7. In this type of block, the feature-maps from all preceding layers are employed as inputs for all subsequent layers. This approach ensures that connections between layers are highly dense, thereby allowing for deeper network architectures.

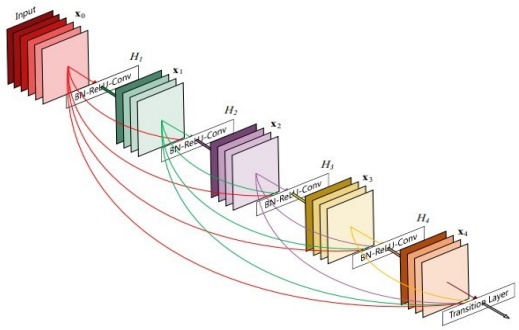


Figure 2.7 Architecture of the densely connected block [12].

Successive Depth-wise Convolution block [13] is a new configuration that conserves computational resources. The architecture of the Successive Depth-wise Convolution block is illustrated in Figure 2.8. This block represents the features of the input I in various ways through X, X' and X''. The output of the module is generated based on these representations. It is important to note that each channel in X can be formed from either a single channel or a group of multiple channels in I. When employed in the early processing phase of a network, this block

acquires low-level features. Conversely, when used in the latter half of the network, it produces vectors containing higher-order features, optimizing the overall feature extraction process.

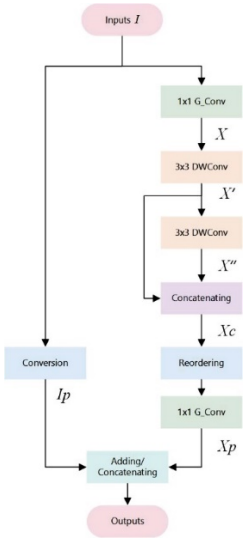


Figure 2.8 Architecture of SdcNet [13].

2.2.4 Training Process and Important Setups

After building a CNN architecture, the trainable parameters should be determined by training process. The training process includes forward propagation and backward propagation. (i) Forward propagation means that the CNN produces a predicted result based on the input data. (ii) Backward propagation means that a loss is calculated based on the predicted results and the ground truth, and the trainable parameters in the CNN are updated according to the loss.

There are many critical setups for training processing.

(1) Batch size. To capture variance features, input image batches should be sufficiently large. However, the maximum batch size is determined by the size of the dataset. Therefore, it is necessary to divide the dataset into appropriately sized batches to enable the network to update gradually.

(2) Data augmentation[14]. Effective CNN training relies on a sufficient number of features in the training set. However, many datasets have limited sample sizes, particularly in cases where positive samples are scarce. To overcome this challenge, data augmentation techniques such as flipping and rotating input images can be utilized to increase the number of training samples.

(3) Loss function. Binary Cross-Entropy loss is widely used in classification CNN[15]. It is defined as:

$$Loss = -\frac{1}{N}\sum_{i}^{N} y_i \times \log[P(y_i)] + (1 - y_i) \times \log[1 - P(y_i)] \quad (2.6)$$

Where y_i is the binary label 0 or 1, and $P(y_i)$ is the probability of y_i.

(4) Number of training epochs. To achieve optimal network performance, it is crucial to allow the training process to continue until the loss curve converges. Stopping the process too early may result in an insufficiently trained network, while stopping it too late would be a waste of time.

(5) Learning rate. Optimizing the learning rate is critical to achieving optimal network performance. In the initial stage, a larger learning rate can help quickly tune the network and decrease the loss function. However, in the final stage, a smaller learning rate is necessary to fine-tune the network and approach optimal performance.

(6) Optimizer. Stochastic Gradient Descent (SGD) [16] and Adaptive Moment Estimation (Adam) [17] are widely used in CNN. Adam is an efficient optimization algorithm that is well-suited for large networks and huge volumes of data.

2.3 CNNs for Lung Nodule Detection

A lot of CNN systems for lung nodule detection have been reported. Many of them are designed based on the famous VGG and U-net.

2.3.1 VGG and U-Net

VGG[8] is proposed by Visual Geometry Group at University of Oxford. The VGG architecture consists of two versions: VGG16 and VGG19. VGG16 comprises 13 convolution layers and 3 fully connected layers, as depicted in Figure 2.9. On the other hand, VGG19 is a deeper variant with a total of 19 layers. Each convolution layer in the VGG blocks utilizes a standard 3×3 kernel. Furthermore, the blocks incorporate the ReLU activation function and max pooling operation for feature extraction.

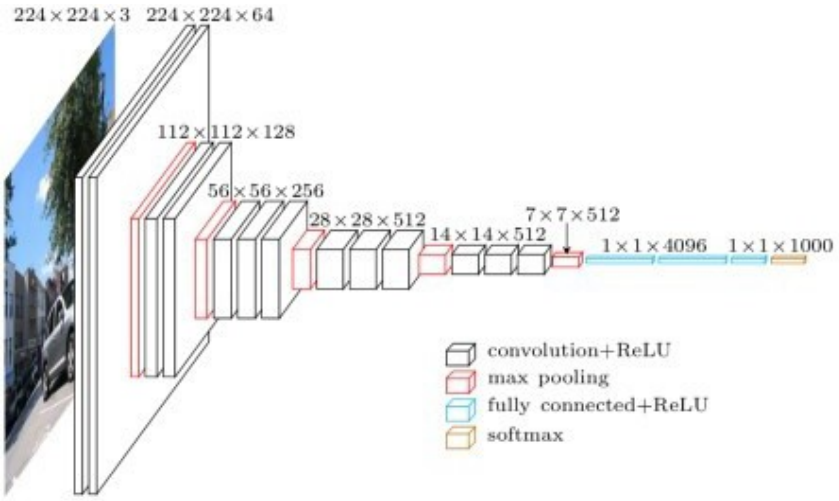


Figure 2.9 Architecture of VGG [8].

U-Net, based on FCN[18], has become a popular choice in medical image processing, particularly for tasks such as lung nodule detection. The architecture of U-Net is illustrated in Figure 2.10, featuring a VGG-like contracting path and an expansive path that utilizes copies of the feature data generated in the contracting path through concatenation in the convolutions.

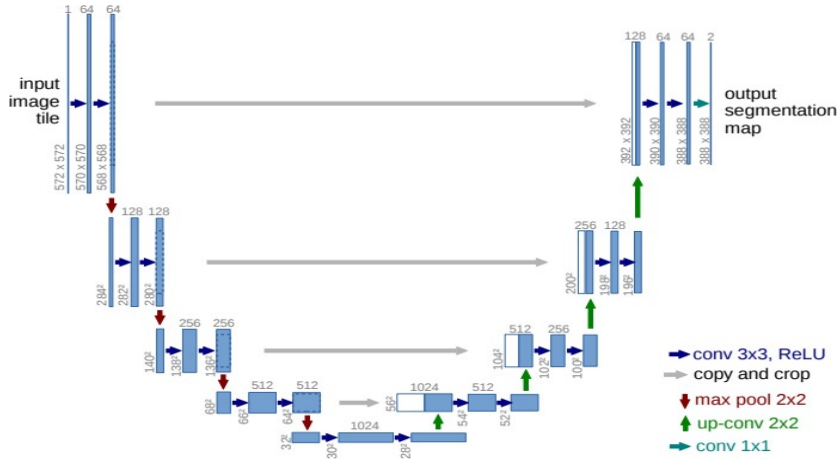


Figure 2.10 Architecture of U-net [19].

2.3.2 CNNs for Detecting Lung Nodule Candidates

The network shown in Figure 2.11 utilizes VGG-16 as a crucial component for detecting candidate nodules. In a similar manner to the U-Net structure, the VGG structure serves as the contract path, also known as the encode path. The initial 5-group convolutions of VGG-16 are specifically employed to extract an ample number of features required for the subsequent processing of regions of interest (ROIs).

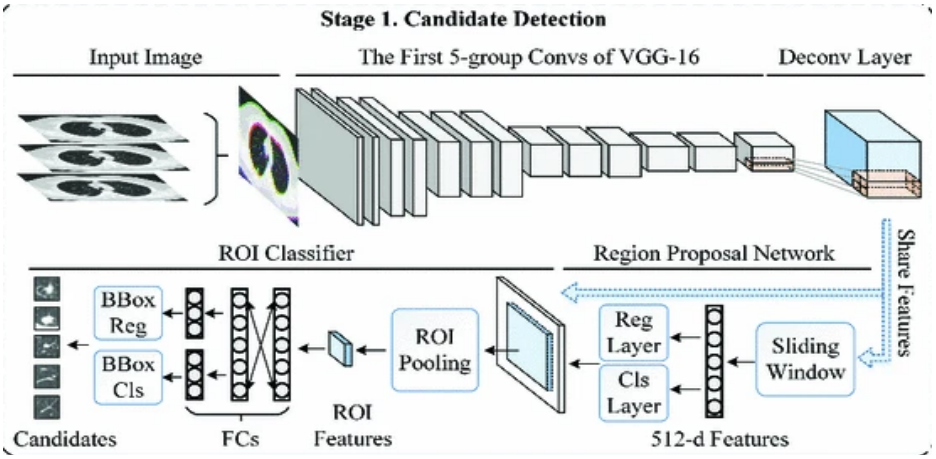


Figure 2.11 Architecture of the network presented in [22].

U-net-based convolutional neural network (CNN) systems have emerged, focusing on enhancing the quality of lung nodule detection. One particular approach to enhance the CNN, as illustrated in Figure 2.12, involves substituting the standard convolution blocks within the original U-net with Residual blocks. This modification aims to optimize the detection performance and overall effectiveness of the system.

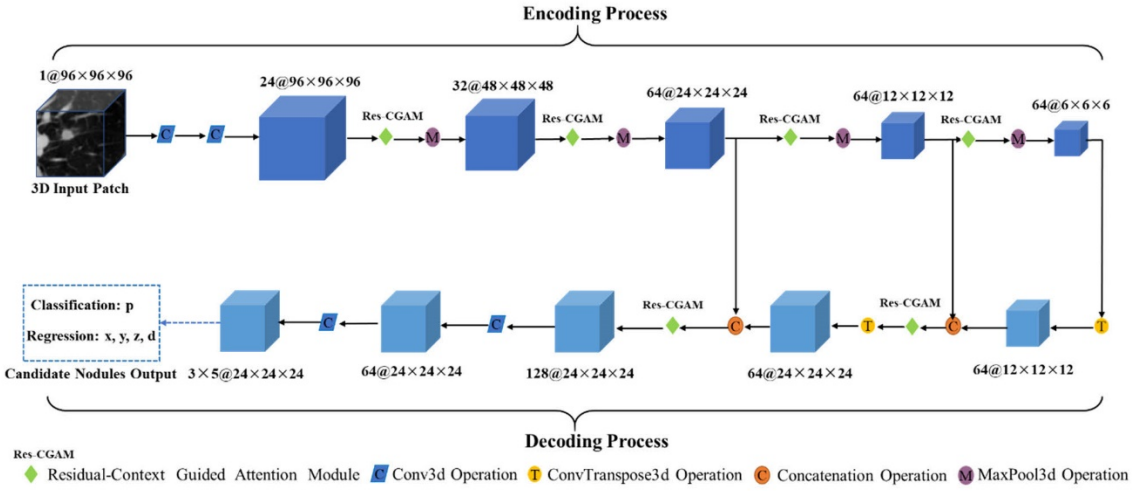


Figure 2.12 The improved 3-D Residual U-Net network structure[20].

Some other identification mechanisms, such as the adaptive feature fusion network [21] illustrated in Figure 2.13, are introduced into U-Net based structure. These structures maintain the overall symmetrical structure and basic convolution block without altering them. However, they incorporate an additional operation called RPN to complement and enhance the network's capabilities.

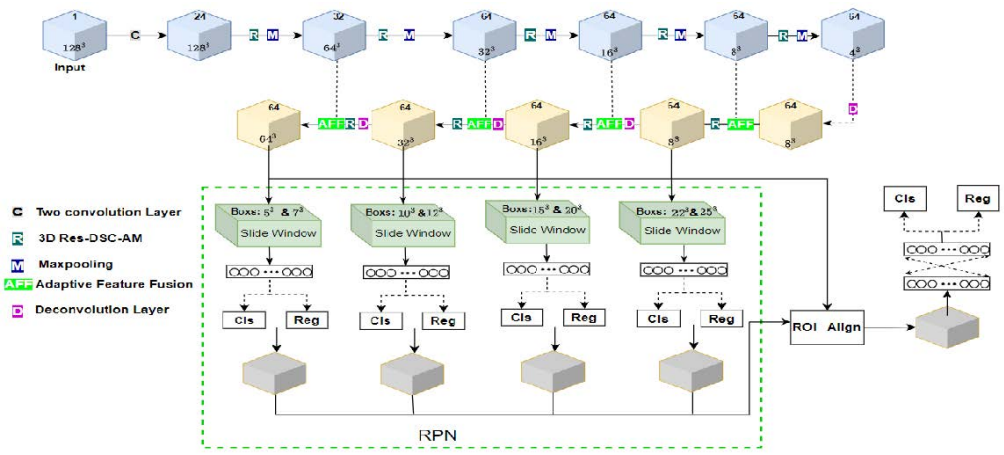


Figure 2.13 Overall framework of U-Net based network with AFF structure[21].

In certain networks, such as the multiscale aggregation network [23] shown in Figure 2.14, the base convolution blocks are substituted with inception modules and residual blocks. Additionally, the link between the encode and decode path of the U-Net is modified. These modifications contribute to an enhanced network capacity achieved through more intricate connections, thereby maximizing the utilization of feature maps from the encode path.

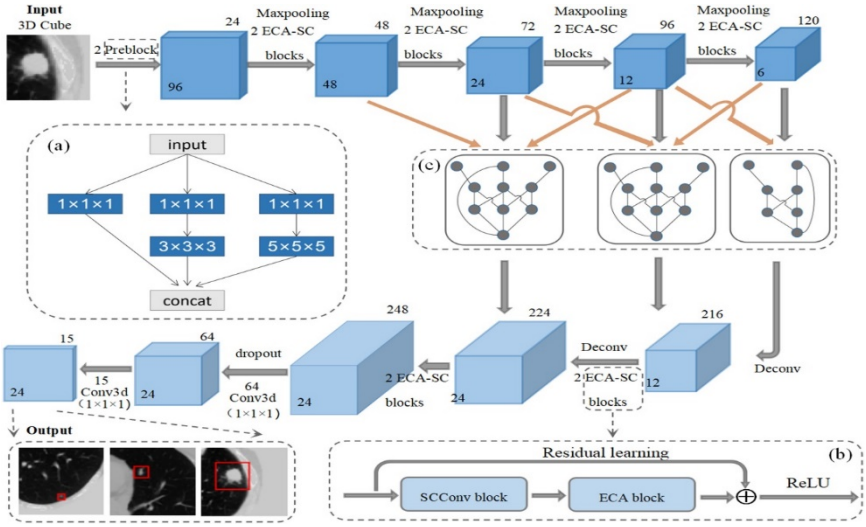


Figure 2.14 Overall framework of MSANet [23].

2.3.3 CNNs for False Positive Reduction

To facilitate the false positive reduction stage, recognition networks are employed to process the candidates, with many of them being built upon the VGG structures. To provide an illustration, the network utilized in the second stage is depicted in Figure 2.15.

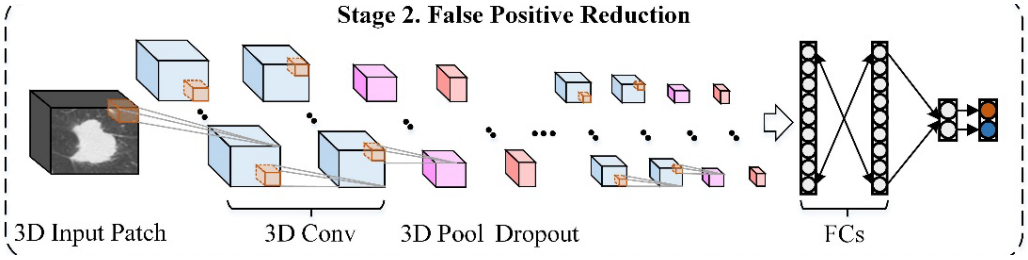


Figure 2.15 Architecture of the network presented in [22].

Similar to the networks in Stage A, one notable approach to enhancing the VGG architecture, as exemplified in Figure 2.16, entails replacing the conventional convolution blocks in the original VGG with Residual blocks.

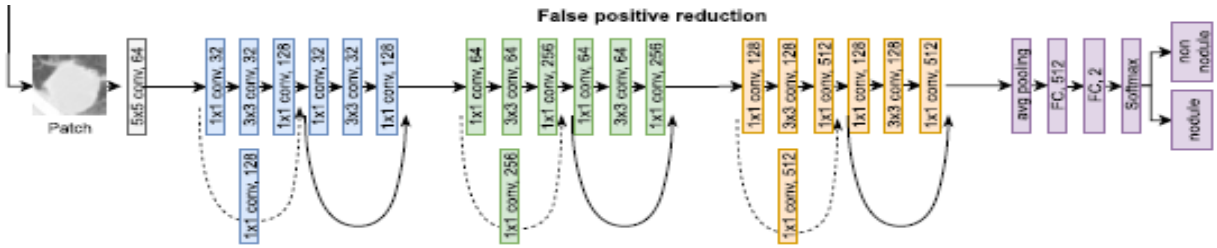


Figure 2.16 Architecture of the network presented in [24].

In addition to employing more intricate convolution blocks, improvements can also be made at the network structure level. One such enhancement involves feeding multi-scale data into the network and subsequently merging the extracted feature maps for further processing, as the Figure 2.17 shown.

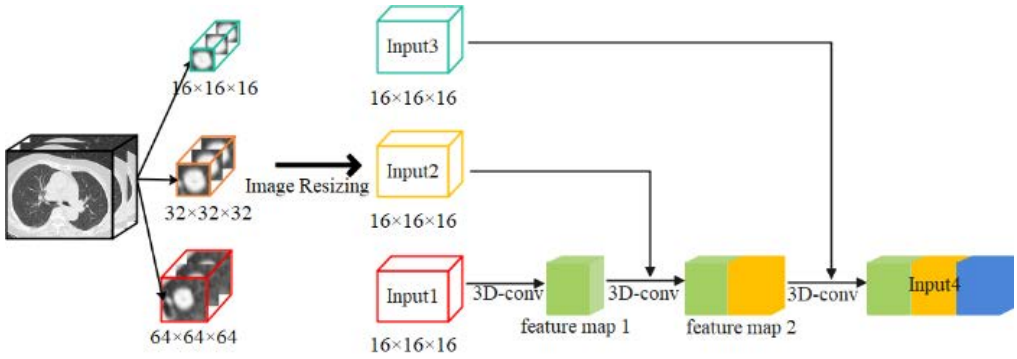


Figure 2.17 Architecture of the network presented in [25].

Furthermore, there are networks based on the U-Net architecture. For example, as depicted in Figure 2.18, the network utilizes U-Net to generate supplementary losses, which play a crucial role in optimizing the identification process.

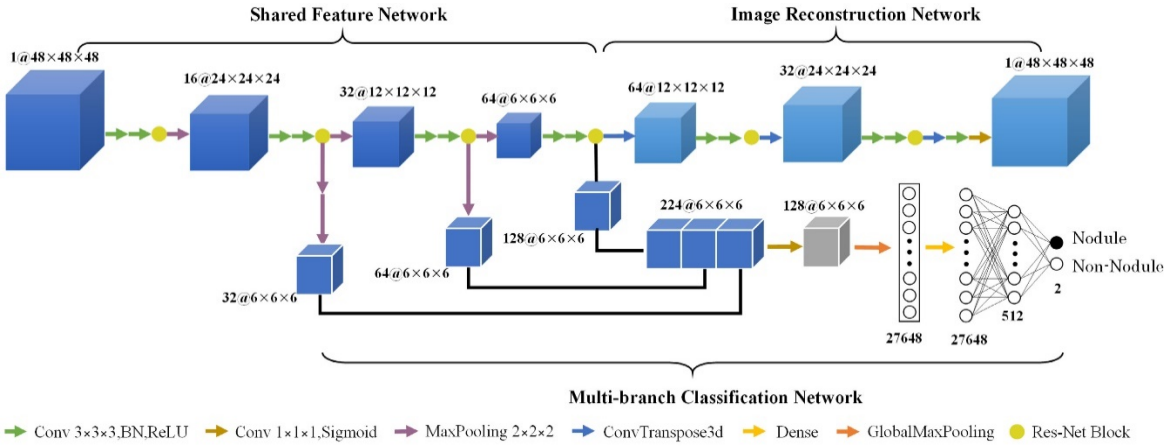


Figure 2.18 Structure of multi-branch 3-D classification network based on multi-task learning.

2.4 Summary

Convolutional neural networks can handle complex signal detection or recognition problems. Thus, the preparation and literature search for the work presented in this thesis have been focused on CNN systems for automatic detection of lung nodules. This Chapter is dedicated to the review of the CNN basic, including elements in a CNN architecture and training, and CNN design examples in the topic area of lung nodule detection. Most of these design examples have been published very recently, and can represent the latest development.

With a view to developing CNN systems of better overall performance, with respect to the existing ones, we need a novel design approach to making CNNs of high computation efficiency. As mentioned earlier, the approach of custom-designing CNN structures to suit particular tasks can be promising. In Chapter 3 and Chapter 4, new CNN structures of Stage A and Stage B are proposed and presented in great detail to illustrate the effectiveness of this design approach.

3. Stage A for Lung Nodule Candidate Detection

3.1 Introduction

Stage A is designed to localize the candidates of lung nodules in a 3D CT lung image. The focus is to achieve a high detection rate. The output data will then be processed in Stage B for a false positive reduction.

This stage consists of 3 blocks, for preprocessing, coarse nodule detection by a U-Net-based CNN, and postprocessing, respectively.

3.2 Pre-Processing Block

The operations in this block are performed to uniformize the data samples, in terms of signal range and resolution.

- Uniformizing the resolution of the data in the 3rd dimension, by bilinear interpolation. The resolution in the 3rd dimension, known as "spacing" between 2 consecutive axial slices, of the input data may vary, depending on the sources of the data samples. It is important to uniformize the resolution before applying the data samples from different sources to the system.
- Uniformizing the gray level distribution of the data samples by sample-wise normalization of the input data. It is performed with the data of each individual 3D sample. The range of the data is firstly scaled down to [0,1] and then X_n, the normalized value of each element is calculated by a standard normalization function $X_n = (X-\mu)/\sigma$. In this calculation, μ and σ are the average value and the standard deviation of the scaled data of the 3D sample.

3.3 Custom-Designed CNN for Lung Nodule Candidate Detection

The function of this CNN is to detect the lung nodule candidates in an input of 3D lung CT image. It is designed to achieve a high sensitivity in detection, and its output will then be applied to Stage B specifically designed to remove false candidates. The structure of this CNN, as shown

in Figure 3.1 and, is based on U-Net [19], featuring a contracting path for feature extraction and an expanding path processing feature data of different levels. The details of structure are shown in Table 3.1. The convolution modules are custom-designed to perform the specific task with the 3D images of lung CT scans. The following issues are taken in the design of this particular U-Net-based CNN.

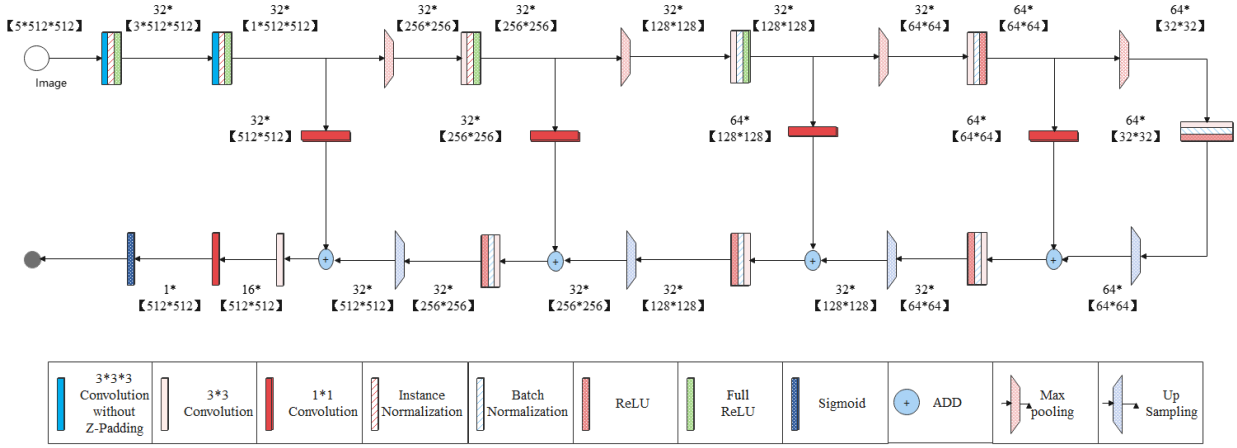


Figure 3.1 Detailed block diagram of the CNN detecting lung nodule candidates. The details of Full-ReLU module are illustrated in Figure 0.22.

Table 3.1 Details of the CNN Configuration

Layer #	Kernel Size	Number of input channels	Number of output channels	Size of input channel	Size of output channel	Number of parameters
1	3×3×3	1	32	512×512×5	512×512×3	448
2	3×3×3	32	32	512×512×3	256×256	27680
3	3×3	32	32	256×256	128×128	9248
4	3×3	32	32	128×128	64×64	9312
5	3×3	32	64	64×64	32×32	18624
6	3×3	64	64	32×32	64×64	37056
7	3×3	64	32	64×64	128×128	18528
8	3×3	32	32	128×128	256×256	9312
9	3×3	32	32	256×256	512×512	9312
10	3×3	32	16	512×512	512×512	4641
Seven 1×1 Convolutions						18169
Total						157,729

- **Input samples and the two 3D convolution layers.** The size of each input sample of the CNN shown in Figure 0.21 is (512x515x5), i.e., 5 consecutive slices of 512 x 512 pixels. The convolutions in the first 2 layers are performed with 3 x 3 x 3 kernels to obtain the original feature information from the 3D input. Thirty-two 3 x 3 x 3 kernels are used in first layer and 64 in the second layer. It should be noted that the option of no-padding is used in the 3rd dimension in each of these convolution layers, and the output channels of the 2nd layer are 2D feature maps of 512 x 512 pixels. Even though the data dimension is reduced to 2D, the original 3D information is included in these feature maps and they can then be processed in the succeeding 2D convolution layers.
- **Activation function of Full-ReLU to preserve signal data**. The operations in the first 4 layers are mainly for feature extraction by means of high-pass filtering. It is known that each high-pass kernel can be used to detect gray-level variations in the 2 opposite directions, indicated by the positive or negative signs of the output data elements. The activation function of Full-ReLU [27] are applied in these 4 layers to eliminate the risk of losing the information carried by the negative data elements. The function of Full-ReLU is to separate the positive and negative elements of a data map $[x]$, and place them into 2 maps $[x_p]$ and $[x_n]$, defined as follow.

$$\begin{cases} x_{n\,ij} = \left|\min(0, x_{ij})\right| \\ x_{p\,ij} = \max(0, x_{ij}) \end{cases} \tag{3.1}$$

$[x_p]$ and $[x_n]$ can then be used separately or combinedly. In this CNN, they are grouped with learnable weights, by means of 1x1 convolution, as shown in Figure 3.2, to generate new feature maps applied to the next layers.

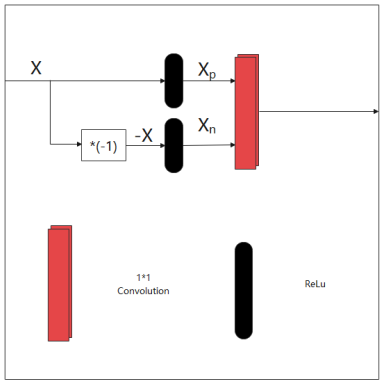


Figure 3.2 Implementation of Full-ReLU.

- **Skip lines**. here are 4 skip lines, used to transmit the feature data, generated respectively by the 4 convolution layers, i.e., from 2nd to 5th, in the contracting path, to the corresponding layers in the expanding path. A 1 x 1 convolution is applied to scale the feature data, with learnable weights, and neither normalization nor activation function are used after it to preserve the feature information.
- **Expanding path**. This path features 1) bilinear up-sampling, instead of deconvolution, for data map expanding, 2) the scaled feature data modulating the data maps by means of addition operations, and 3) simple1 x 1 and 3 x 3 convolutions being applied to generate the output of the network, as shown in Figure 0.21.

3.4 Refinement Block

As mentioned previously, the CNN in Stage A is designed to be sensitive to likely nodule candidates, minimizing the risk of missing a true nodule, while resulting in a very large number of false nodule candidates. The post-processing block in Stage A is used to detect and to remove some of the false candidates to reduce the load of Stage B.

Some of false nodules can be easily detected by checking 2 indications.

- Nodule locations. A nodule can appear in many locations inside a lung lobe, like the example in Figure 3.3, and it is unlikely found outside lobes. Thus, the 2 ends of a 3D lung scan, i.e., it's very top or bottom section, is out of the region of interest. Of the 888 3D lung CT scans available in LUNA16, no true lung nodule is found in the sections of the lowest 10% and the highest 4% of the axial lung slices. Thus, all the candidates found in these sections are considered as false nodules and are removed.
- Nodule thickness. A lung nodule is lump-like and the thickness is ranged 3~32 mm. For the data samples used in this work, a nodule should appear in at least 2 consecutive slices. A nodule candidate appearing only in one slice is unlikely a true one.

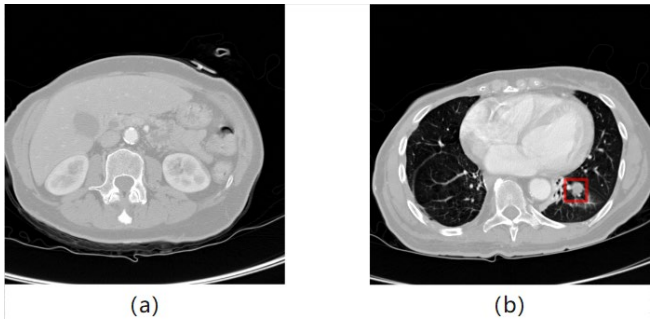


Figure 3.3 (a) Slice sampled from a section outside lung lobes.
(b) Slice sampled from a section of lung lobes.

By this simple detection, checking the location and the thickness of the candidates, performed in this post-processing block, one third of the false candidates can be identified and removed. The risk of removing true nodule candidates in this block is very small.

3.5 Summary

The objective of the data processing in Stage A is to localize the candidates of lung nodules in a 3D CT lung image, aiming at achieving a high detection rate. In this chapter, the design of Stage A is presented in detail. The core of Stage A is the custom-designed and U-net-based CNN. The filtering modules in its convolution layers are specifically designed to suit the features produced in these layers. There is a pre-processing block to uniformize the input data distribution and also a refinement block to remove some of the false nodules from the data produced by the CNN.

As mentioned previously, in a 3D CT lung image, there are many nodule-look-likes, a U-net based CNN is hardly able to distinguish true nodules from false ones accurately. Hence, to obtain a high-quality detection, we proposed a CNN system, referred to as Stage B, for nodule recognition. The design of Stage B is presented in Chapter 4.

4. Stage B for Lung Nodule Recognition

4.1 Introduction

The outputs of Stage A are the lung nodule candidates. Each is defined by a predicted nodule size in diameter and location represented in coordinates in the 3D lung image. As only a minority of the candidates are true nodules, the false positive rate of the first stage is very high. Thus, Stage B is designed to identify the true nodules from all the candidates so that a good final detection result, i.e., a high true positive rate and low false positive rate, can be obtained.

As mentioned previously, Stage B can, in fact, be seen as an independent CNN system for lung nodule recognition. Its input data are 3D nodule-candidate samples cropped from the 3D lung images. Each sample is a cuboid consisting of n consecutive axial slices of (l x m) pixels. The values of l, m, and n are determined according to the maximum dimensions of the nodule candidates so that the entire nodule candidate is centered and completely covered in the cuboid. To each input sample, the processing in this stage will result in a binary output, nodule or non-nodule.

4.2 Problem Statement and the Design Strategy

It is known that lung nodules vary extensively in gray levels, sizes, shapes, patterns and surroundings. Moreover, there are a lot of spots of different patterns looking like nodules in lung images. To precisely identify true nodules, one needs to investigate different features of nodules in order to find an optimal way to distinguish the true ones from false ones.

The size of a nodule, often presented by the diameter appearing in the center slice of its 3D sample, varies a lot. It can be measured in number of pixels. The distribution of nodule diameters shown in Figure 4.1 is given by the 1557 true nodule samples in LUNA dataset [25]. The diameter is ranged from 3 pixels to more than 30 pixels, whilst many are smaller than 8 and the most frequently occurring value is around 6. One can observe significant differences, between small and large nodules.

- Most small nodules appear as simple spots, as shown in Figure 4.2. They look very similar to non-nodule spots, e.g., those of bronchioles. Hence, the clues to distinguish true nodule spots from their look-likes may be found not only in the spots themselves, but also in the surrounding areas.
- Large nodules have richer variations, with respect to small nodules, as some examples illustrated in Figure 4.3. The difficulties in the detection are due to the complexity in nodule shapes and patterns.

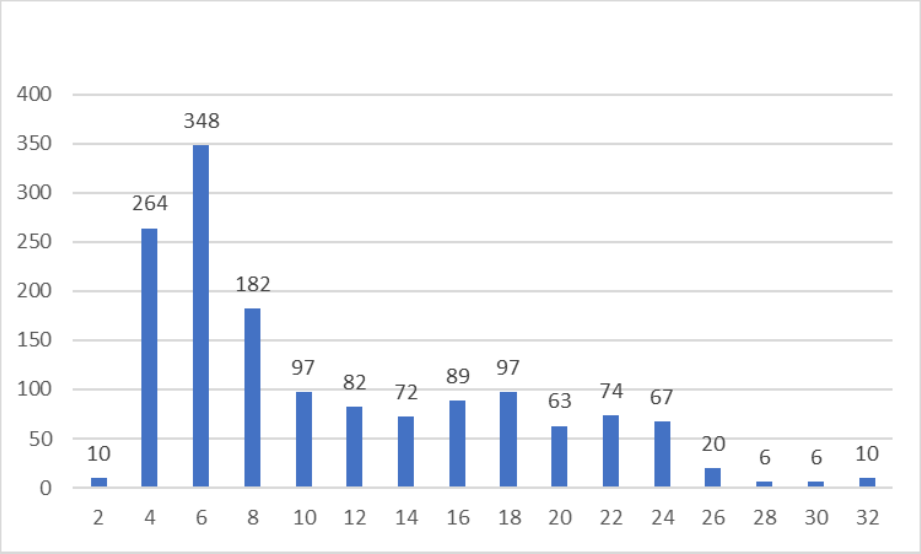


Figure 4.1 Distribution of nodules diameters, measured in number of pixels, of the 1557 true nodule samples of LUNA16.

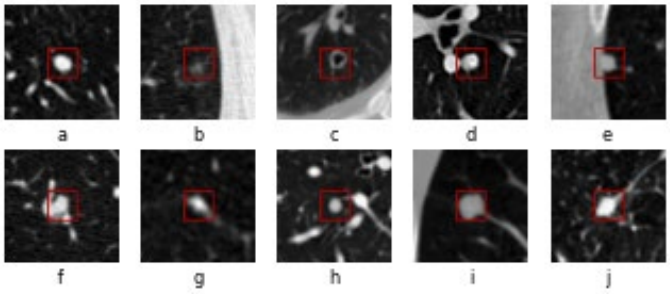


Figure 4.2 Samples of small nodules. The nodule spot in each sample is red-framed.

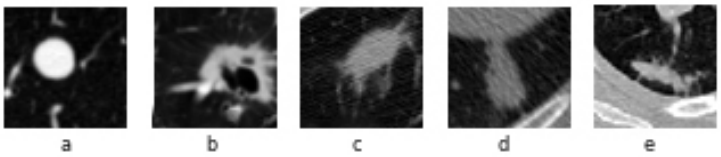


Figure 4.3 Samples of large nodules of different patterns.

It seems that one needs to handle 2 categories of lung nodules, i.e., large ones and small ones. Large nodules can identify themselves by their own features, whereas small ones also by their neighbors' features. The image features in the first category can be incoherent with those in the second category. Each of the categories has enormous variations, and the appearances of the two combined are even much more diverse and complex.

A CNN can handle complex detection task with its large number of filters, but there is a limit. Moreover, an excessive number of learnable filtering parameters may lead to excessive randomness in training, which may produce adverse effects. Hence, we consider to have 2 simple CNNs aiming at the 2 nodule categories, respectively. To this end, on one hand, each of the 2 should be designed specifically to enhance the capability of identifying nodules of its designated category. On the other hand, both CNNs should be compatible with each other in terms of recognizing nodules of median sizes, as their characters can be found in both nodule categories.

As mentioned earlier, each input of Stage B is a series of n consecutive axial slices of (l x m) pixels, cropped from 3D lung images. As nodules are mostly irregularly lump-shaped, the brightness and size of a nodule spot reduce gradually from the center slice to each of the 2 ends, as the example shown in Figure 4.4. The shape of nodule spot also varies gradually from slice to slice. The n consecutive slices should include all the image features needed for the nodule recognition. However, for the 2 nodule categories, the contents in these slices are not exactly the same.

- In case of a small nodule, the information of its near neighborhoods is needed. Thus, the series of slices should cover the nodule region and its surroundings. It can have 3 sections, each having, for example, 5 slices. The center section covers most pixels, if not all, in the nodule region. The other 2 sections cover the 2 adjacent regions in front and back of the nodule, respectively, as shown in Figure 4.5. Hence, each section carries the information of its own object, whereas all the objects from the 3 sections are needed to find an indication of

a true nodule. Thus, it may be beneficial to detect the features in different sections separately and then process the features together for the recognition.

- In case of a large nodule, the identification of true nodule can be based mainly on the features extracted from the nodule region. If the number of slices in the series is similar to that of the small nodule category and the series is also divided into 3 sections, the head, center and tail of the nodule will appear in the 3 sections, respectively. One can also extract the features from the 3 sections and then process them collectively for the recognition.

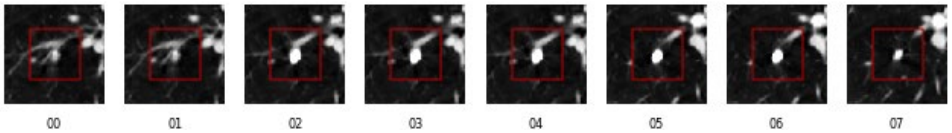


Figure 4.4 Series of 8 slices covering a small nodule.

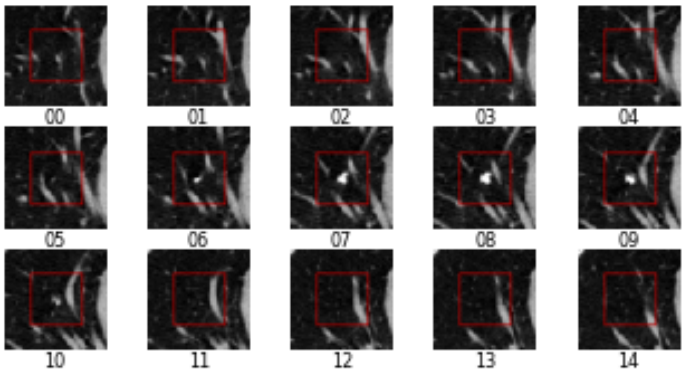


Figure 4.5 Series of 15 consecutive slices of a true nodule sample.

Based on the observations and analysis described above, we propose a design strategy to achieve a high quality of lung nodule recognition, and it is summarized as follows.

(1) *Multiple simple CNNs for nodule recognition*. Each simple CNN should be designed to focus on features of one category of nodules.

(2) *Particular 3D data processing to identify true nodule cases.* This particularity is first to divide the series of input slices into multi sections to facilitate the extraction of their 3D sematic features, separately and collectively, and then to use the data generated by the multi sets to make the final identification decision. This idea can be interpreted as a block diagram illustrated in Figure 4.6.

To apply this strategy, more analysis of the input data in the 2 categories is need in order to find the most suitable way to prepare the input data and to find the structures of convolution layers for each of the 2 CNNs. The design details are presented in the following chapter.

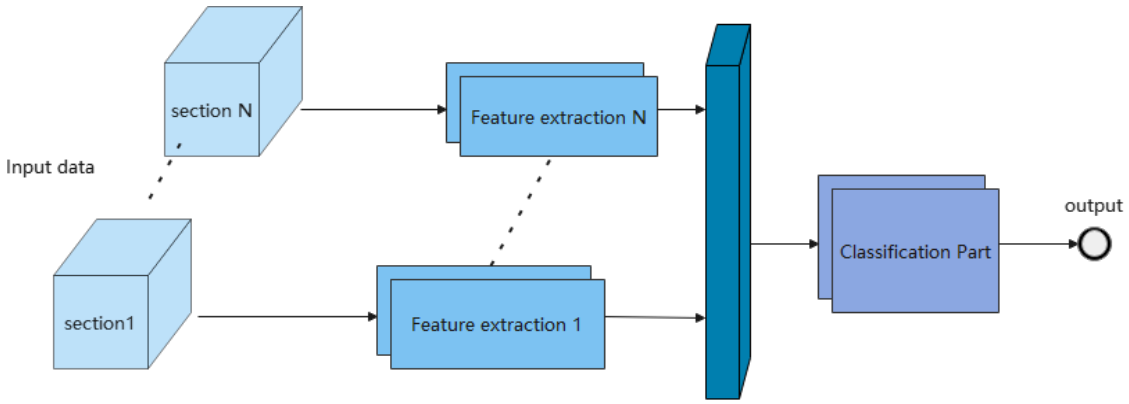


Figure 4.6 Proposed CNN framework for nodule detection. A 3D input image is divided into multiple sections and their features are extracted, respectively, in the multiple designated branches. The results are then merged for the classification.

4.3 Design of the Two CNNs for Nodule Recognition

As mentioned previously, we propose a system, consisting of 2 simple CNNs, for lung nodule recognition, as nodules of different sizes have very different characters. One of the 2 CNNs is designed to optimize the feature extraction of nodules in small size category and the other for the features often seen in large nodules. However, the 2 CNNs should be compatible with each other so that nodules having characters found in both categories can be recognized by one or both CNNs. Each input data sample is applied to both CNNs, resulting in 2 outputs, each representing the probability of the sample involving a true nodule. The final binary result is obtained by merging the 2 probability values.

The structures of the 2 CNNs are based on the scheme shown in Figure 4.8. However, the data preparation is done differently and the convolution units are also configurated differently. The detailed design of the CNN for small nodule category is found in Sub-subsection 4.3.1 and that of the other CNN in Sub-subsection 4.3.2.

4.3.1 CNN Aiming at Small Lung Nodule Recognition

A majority of lung nodules are small, i.e., found in the lower part of a typical nodule distribution shown in Figure 0.24. Though it is difficult to define a clear boundary point of this part, it is assumed that, in the category of small nodules, the diameters of nodules, appearing in its center slices, are not more than 10 pixels. This CNN is designed to identify lung nodules of this category. More specifically, the preparation of the input data and the configuration of convolution units are made particularly according to the characters of such nodules.

As mentioned previously, each input of Stage B is a 3D samples consisting of N consecutive axial slices of (l x m) pixels, cropped from the 3D lung image. As the diameters of the targeted objects are no more than 10 pixels in their center slices, the slice size is chosen to be 32x32 pixels so that, in each slice, the entire nodule area and its surroundings in the 4 sides are included.

Now, let us determine the size of the 3rd dimension, i.e., the number of the consecutive axial slices in each input sample. It is observed that, in this small nodule category, the main body of a true nodule appear in no more than 5 slices. Thus, these 5 slices constitute the central section, referred to as C-section, of the N slices. Additional 8 slices, of which 4 located before the central section and the other 4 after it, are also included in the input sample so that it covers the surroundings in all the 6 sides of the 3D nodule region. Hence, each sample is defined as 13 slices of 32 x 32 pixels, partitioned into 3 sections, namely H (head), C (core) and T (tail) sections.

It should be noted that, as the nodule region takes only a tiny part of the central section, the feature data in this section are insufficient to yield anything meaningful to identify the nodule. Hence, the sections, of which the image features are correlated, should be combined to extract nodule-relevant features. Of the 3 sections, the 2 adjacent sections, H and C sections, or C and T sections, are more correlated than nonadjacent H and T sections. Hence, the 3 sections are combined into 2 sets, i.e., the 9 slices from H and C sections making up the first set and the 9 slices from C and T sections the other. The data of C section are, in fact, included in both sets.

To implement the idea described above, we propose a CNN structure illustrated in Figure 0.30 and the configuration details are found in Table 4.1. This CNN consists of feature extraction (FE) block and classification block. The slices in the input series are indexed [0, 12]. The FE block

comprises 2 branches. The first 9 slices, indexed 0 to 8, from H and C sections are applied to the upper branch, and the other set of 9 slices, indexed 4 to 12, to the lower branch. There are 2 particular points in this block.

- The data in each input sample are normalized with the average value μ and the standard deviation σ obtained from its core of 16 x 16 x 9 pixels so that the data normalization can be more dominated by the pixels in the nodule area and its nearest surroundings.
- The input data of the first convolution layer in each branch are nine 2D channels, corresponding to the 9 slices. In this way, a 2D convolution layer can generate its output data involving the 3D information of its 9 input channels.

As shown in Figure 0.30, each branch in the FE block comprises seven 2D convolution layers. Full-ReLU activation, instead of regular ReLU, is used after the first 3x3 convolution to minimize feature data loss. The convolved data channels are down-sampled using max-pooling after the 2nd and 4th layers, and thus the channel size is reduced from 32 x 32 to 8 x 8. The last three layers of the FE block perform convolutions without padding to exclude irrelevant border pixel values from calculations, resulting in a reduction of feature maps from 8 x 8 to 2 x 2. Each branch generates 64 channels and the data of the 64 x 2 channels are then applied to the classification block.

As the shown Figure 4.7, the classification block involves 3 fully-connected (FC) layers as the main processing units. One special aspect of this block is the use of 2 x 2 depth convolution instead of the standard global average pooling method. The final result is produced by applying the sigmoid function directly to the output data generated by the third fully connected (FC) layer, without going through batch normalization or ReLU.

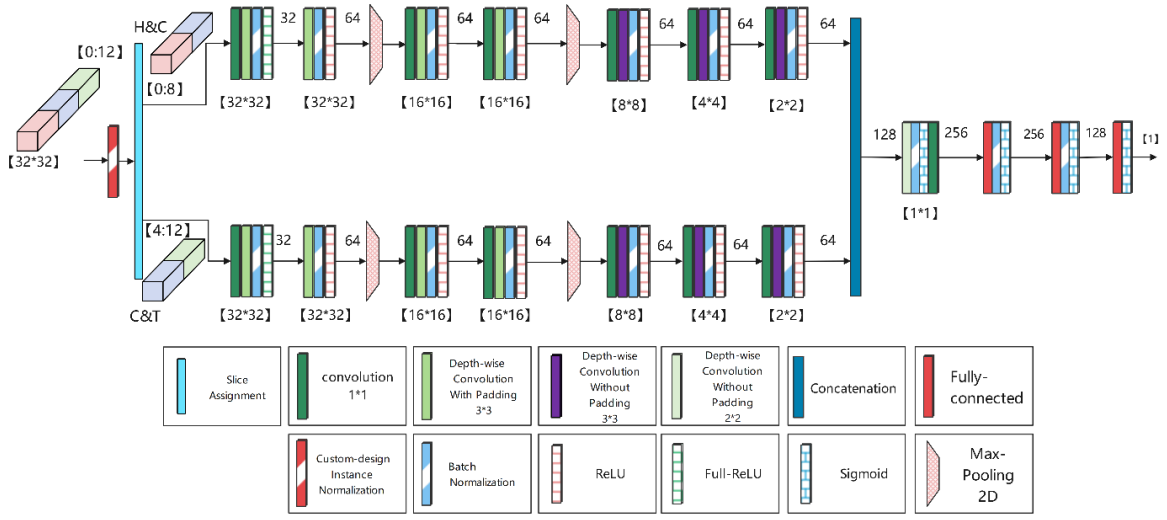


Figure 4.7 Block diagram of the CNN designed to focus on small nodules.

Table 4.1 Detail in convolution layers of CNNs aiming at small nodule

Features Extraction Parts (Input channels of first layer: H&C:9; C&T:9;)					
Layer		Activation function	Size of features map	Number of Output channels	Number of parameters
1	H&C	Full ReLU	32x32	32	1408
	C&T	Full ReLU	32x32	32	1408
2	H&C	ReLU	32x32	64	2496
	C&T	ReLU	32x32	64	2496
3	H&C	ReLU	16x16	128	4928
	C&T	ReLU	16 x 16	128	4928
4	H&C	ReLU	16 x 16	128	4928
	C&T	ReLU	16 x 16	128	4928
5	H&C	ReLU	6 x 6	256	4928
	C&T	ReLU	6 x 6	256	4928
6	H&C	ReLU	4 x 4	256	4928
	C&T	ReLU	4 x 4	256	4928
7	H&C	ReLU	2 x 2	256	4928
	C&T	ReLU	2 x 2	256	4928
Classification Part					
2*2		Sigmoid	1 x 1	512	33920
FC1		Sigmoid	1	128	132608
FC2		Sigmoid	1	64	65920
FC3		N/A	1	1	129
Total					289,665

4.3.2 CNN Aiming at Large Lung Nodule Recognition

As mentioned earlier, a majority of the lung nodules are small, i.e., the nodule diameters are concentrated around 6 pixels, and the CNN described in the preceding subsection is designed for such cases. On the contrary, the diameters of large nodules can span much more, from 10 to more than 30 pixels, and are less concentrated. Hence, the CNN aiming at large nodule recognition should be made to target such a wide range of nodules.

It should be mentioned that a large nodule can be identified mainly by the features in its own nodule regions. If the nodule size is reduced, the density of the nodule information in a data sample is also reduced and one has to use the features in its surroundings to make the decision. Considering the upper side of nodule diameters can be 30 pixels or more, the slice size is chosen to be 64 x 64 pixels and each data sample should be a cuboid be 64 x 64 x 15 pixels cropped from lung image.

The number of slices per sample is chosen to be 15. It is sufficient to cover a nodule of fairly large size, and in case of a medium-sized nodule, also to include its surroundings. A very large nodule can show up in more than 15 slices, however the image features found in the central 15 slices are usually sufficient for the recognition. Increasing the number of slices reduces the density of nodule information in the sample, making the detection of smaller nodules more difficult.

Each slice in a data sample is then down-sized, by average pooling, from 64 x 64 to 32 x 32 pixels. By doing so, the slice format of this CNN is identical to that in the CNN for small nodules, and it will hardly affect the processing quality, as the information density is not reduced. If a data sample containing a very small nodule, the down-sizing may make the nodule region more difficult to detect in this CNN, but it will more likely be identified in the CNN for small nodules.

The series of 15 slices is divided into head (H), core (C) and tail (T) sections. The central 7 slices are assigned to C section. Similar to the cases of small nodules, the first 4 slices of each sample are assigned to H section and the last 4 to T section. In case of a fairly large nodule, the features in H and T sections give the information of nodule head and tail. For nodules of smaller sizes, these 2 sections will contain features of neighborhoods.

The block diagram of the proposed CNN for large lung nodule recognition is illustrated in Figure 4.8 and the configuration details are found in Table 4.2. In the feature extraction block of this CNN, there are 3 branches placed to process the input data from the 3 sections, respectively. There is one-slice-overlap between 2 adjacent sections. Thus, the first 5 slices of the input series are applied to the upper branch is 5, the last 5 to the lower one, and the 7 slices in the center to the branch in the middle. As these 7 slices cover the core section of the input data and there are a lot of variations in this section, the number of kernels in the convolution layers in the middle branch is significantly larger than that in the other 2 branches to enhance the feature extraction.

The features data produced by the 3 branches are applied to the classification block in this CNN. The structure of this block is the same as that in the CNN for small nodules, except a small difference in the number of kernels in one of the layers.

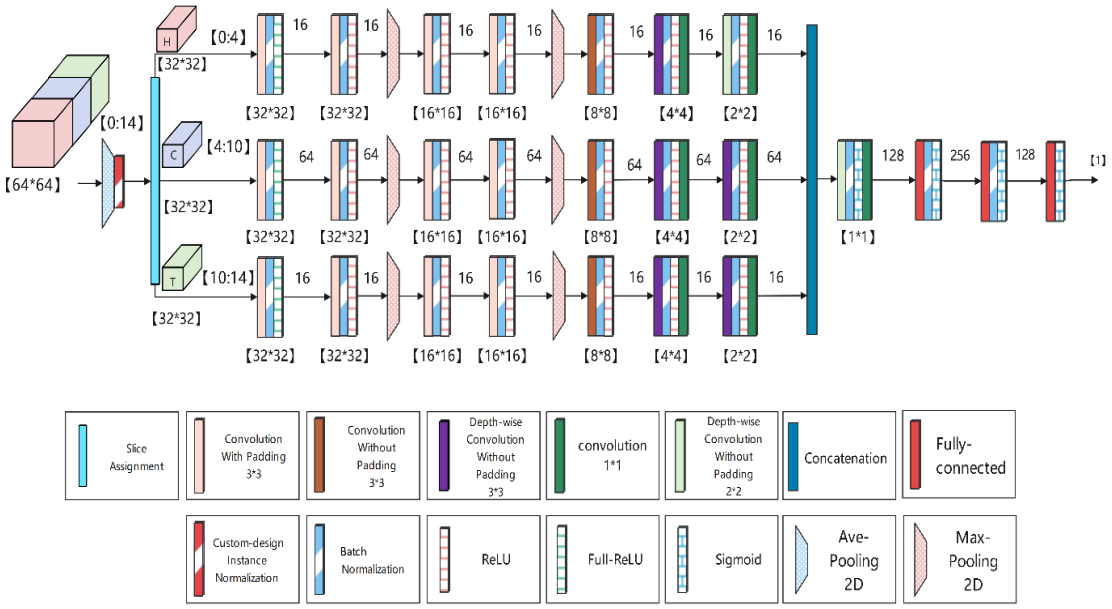


Figure 4.8 Block diagram of the CNN designed to focus on three sections.

Table 4.2 Detail in convolution layers of CNNs aiming at large nodules

Features Extraction Parts (Input channels of first layer: H:5; C:7; T:5;)					
Layer		Activation function	Size of feature map	Number of Output channels	parameters
1	H	Full ReLU	32 x 32	16	656
	C	Full ReLU	32 x 32	64	6272
	T	Full ReLU	32 x 32	16	656
2	H	ReLU	32 x 32	16	2352
	C	ReLU	32 x 32	64	37056
	T	ReLU	32 x 32	16	2352
3	H	ReLU	16 x 16	16	2352
	C	ReLU	16 x 16	64	37056
	T	ReLU	16 x 16	16	2352
4	H	ReLU	16 x 16	16	2352
	C	ReLU	16 x 16	64	37056
	T	ReLU	16 x 16	16	2352
5	H	ReLU	6 x 6	16	2352
	C	ReLU	6 x 6	64	37056
	T	ReLU	6 x 6	16	2352
6	H	ReLU	4x4	16	768
	C	ReLU	4x4	64	49278
	T	ReLU	4x4	16	768
7	H	ReLU	2x2	16	768
	C	ReLU	2x2	64	49278
	T	ReLU	2x2	16	768
Classification					
2*2		Sigmoid	1x1	128	13088
FC1		Sigmoid	1	128	33536
FC2		Sigmoid	1	64	33152
FC3		N/A	1	1	129
Total					266,241

4.4 Summary

Stage B is designated to perform lung nodules recognition. Nodules are very difficult to be identified from a lung CT image, due to (i) their enormous variations nodule regions and (ii) high similarity to other objects in the image. In this section, based on our analysis on the different characters of large and small nodules, a design strategy has been proposed. It is to have 2 networks in Stage B, one targeting variations in object regions of large nodule category and the other looking more into nodule surroundings in case of small nodules. Two CNNs have been

designed, as an application of this strategy. Each of them has a particular multi-branch feature extraction (FE) block and a classification block involving fully-connected layers. The multiple branches in the FE block of each CNN are specifically designed for the features of its designated nodule category.

The input data sample is applied to each of the 2 CNNs to obtain a value indicating how certain the sample involves a true nodule. The 2 outputs are then merged to make the final recognition decision. The test results of Stage B, together with other tests, are presented in Chapter 5.

5. Performance Evaluation

5.1 Introduction

In this section, we present the performance evaluation of the proposed lung nodule detection system with LUNA16 dataset. The information about the dataset and the tools used in the implementation of the system is found in Sub-chapter 5.1. The details about the training process are described in Sub-chapter 5.2. The test results, in comparison with those reported in recent journal publications are found in Sub-chapter 5.3.

5.2 Dataset and Training

5.2.1 Dataset

The LUNA16 dataset employed in the performance evaluation is derived from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) database. This dataset comprises 888 3D chest CT scans, all with a diameter larger than 3mm and a slice thickness smaller than 2.5mm. In the 888 scans, there are 1,186 nodules are identified by medical specialists. Each of them is indicated by its coordinates and sizes in diameters in the 3D scan.

To argument the number of training samples, rotations of 90°, 180°, and 270° are applied to each data sample.

5.2.2 Training Details

Network training is a critical process in the implementation of the proposed CNNs. Each detail in this training may make difference in determining the filtering parameters and, consequently, the network performance. Here are the most important hyperparameters used in the training process.

- Batch size. The batch size is chosen to be 24 to train the CNN for nodule candidate detection, while each training sample is a 3D lung scan of 512x512x5 pixels. In case of the CNN for small nodules, samples have 32 x 32 x 13 pixels for small nodule CNN and 64 x 64 x 15, and the batch size is both chosen to be 48.
- Loss function. BCE Loss function is used in the training.

- Learning rate. Cosine annealing is employed for learning rate scheduling, with a maximum rate of 0.001 and a minimum rate of 0.0001.
- Optimizer. The chosen optimizer for network training is Adaptive Moment Estimation (ADAM).
- Number of training epochs. The proposed CNNs requires no more than 100 epochs to minimize the loss and complete the training process.
- Kernel initialization. The initial weights are chosen to be truncated normal distribution with 0.1 standard deviation, and the initial biases are chosen to be 0.1.

In the following subsections, the results of the performance evaluation, including those of an ablation study, are presented.

5.2.3 Performance Evaluation Metric

The performance metrics of nodule detection systems involve 2 kinds of measures, the detection quality and the network complexity.

The detection quality can be measured in Sensitivity, also referred to as true positive rate, i.e., the ratio of the number of true nodules detected over the total number of true nodules in all the samples. It can be expressed as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.1}$$

where *TP* is the number of true positive nodules detected and *FN* is that of the false negative ones (i.e., missed nodules). FROC curve, i.e., plot of Sensitivity versus the average number of false-positives per scan is also used to present the detection quality.

The network complexity is related to the computation cost, and can be measured by the number of trainable parameters of the network. Fewer parameters, less computation is needed in training process and also facilitates its implementation in computation resource restricted environment. Excessively employing trainable parameters in a network may lead to more randomness in the network training and cause problem of reproducibility.

5.3 Test Results and Performance Comparison

The proposed system has been trained and tested with the LUNA16 datasets. As there is no specific validation set in the datasets, the 10-fold cross-validation method has been applied to generate the test results. The results have been compared with those reported in reputed journals in the topic area.

The test procedures and results of Stage A is found in Sub-chapter 5.3.1 and those of Stage B in Sub-chapter 5.3.2. The test results of the entire lung detection system, i.e., Stage A and Stage B combined, are found in Sub-chapter 5.3.3.

5.3.1 Test Results of Stage A for Nodule Candidate Detection

Stage A, involving a U-Net-based CNN and a simple refinement block, has been designed to localize the nodule candidates in 3D CT scans. It is focused on minimising the miss rate, i.e., false negative rate, as its output will be applied to Stage B for final detection result.

There are 888 3D CT scans, i.e., 888 patient cases, available for training and testing. In these scans, there are 1186 nodules identified by medical specialists and used as the ground truth data. The method of 10-fold cross-validation has been applied to test Stage A. Figure 5.1 illustrates the 10 values of Sensitivity and obtained in the 10 folds. The values of Sensitivity and False-Positive given by the 10-fold cross-validation are presented in Table 5.1. One can see that Stage A delivers a good detection of nodule candidates and the simple refinement helps to reduce effectively the false positive rate, without causing problems in Sensitivity.

The processing quality of Stage A is compared with those given by five other CNN systems reported in recent years. The results are presented in

Table 5.2. Compared with others, the proposed Stage A is able to detect nodules with a good sensitivity and a significantly lower false positive rate. Moreover, it has been achieved with a much lower computation complexity, as Stage A requires only 0.16M trainable parameters whereas the others may use tens of millions of parameters.

The time required to complete a test for one patient case is approximately 26 seconds. It is under the condition that RTX 2060 is used.

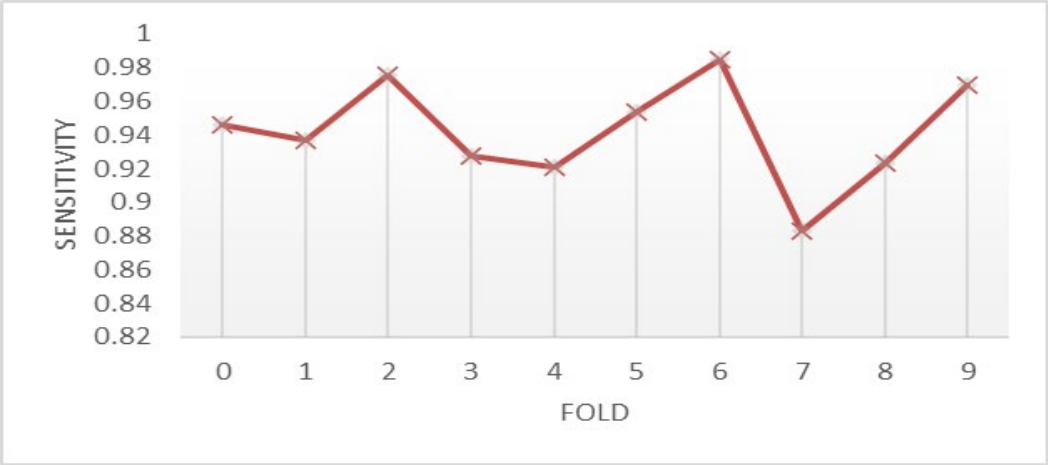


Figure 5.1 Sensitivity values given by the 10 folds in the cross-validation when the false positive per scan is 6.97.

Table 5.1 Test results of Stage A by the 10-fold cross-validation

	Average sensitivity (%)	FPs /scans	Num of Parameters
After CDCNN	95.43%	11.68	157K
After Refinement-block	95.38%	6.97	157K

Table 5.2 Comparison of Stage A results in recent years

Methods	Year	Types	Sensitivity	FPs/Scan	Parameters
Zheng et al.[28]	2020	2D	95.4%	20.4	⩾1.3M
Pereira et al. [29]	2021	3D	98.1%	49.67	N/A
Yuan et al. [20]	2022	3D	94.0%	50.25	>3.88M
Zhou et al. [30]	2022	3D	95.9%	12.21	47.39M
Zhao et al. [31]	2023	3D	94.7%	13.9	N/A
Proposed Stage A	**2023**	**2.5D**	**95.4%**	**6.97**	**0.157M**

5.3.2 Test Results of Stage B for Lung Nodule Recognition

Stage B is designed for lung nodule recognition. It can be used as a stand-alone recognition system or combined with Stage A to perform nodule detection. In this sub-sub-chapter, the test results of Stage B under the stand-alone condition are presented.

Stage B consists of two CNNs designed for the 2 size-based nodule categories, respectively. As there is no clear boundary point to divide the data samples, each of the 2 CNNs is trained

with the same training samples of all sizes. Then, each of the testing samples are applied to both CNNs to produce 2 results and the final decision is based on their average value.

From LUNA16 dataset, 754975 samples, including 1557 true nodule ones, have been used for training and testing Stage B. The process of the 10-fold cross-validation has also been conducted. In order to reduce the randomness in training/testing, the process of the same 10-fold cross-validation has been repeated 3 times. Figure 5.2 illustrated the FROC curves obtained in the 10 folds. The red FROC curve illustrated in Figure 5.3 is obtained by averaging those presented in Figure 5.2. The other 2 curves are given by the 2 CNNs in Stage B, and demonstrate that the 2 CNNs combined deliver better results than each one acts alone. The values used to plot the red FROC curve is presented in Table 5.3.

Stage B has been compared with 5 other lung nodule recognition systems reported recently. The comparison results are presented in

Table 5.4 is demonstrated that the processing quality of Stage B is at least comparable to that of the other system. It should be mentioned that all the values by Stage B are the means obtained in the 3 repeated experiments with the same training and testing conditions, as there are small deviations in the results, as shown in Figure 5.2. In general, the more trainable parameters in a network, the more randomness in training, and the more deviations in the results if the network is retrained.

In terms of computation complexity, Stage B requires only 0.55 M trainable parameters. The network is much simpler and its computation efficiency is much higher than the others. It can, thus, be implemented much easily in computation resources restricted environments.

The time required to complete a test for one patient case is approximately 0.2 seconds. It is under the condition that RTX 2060 is used.

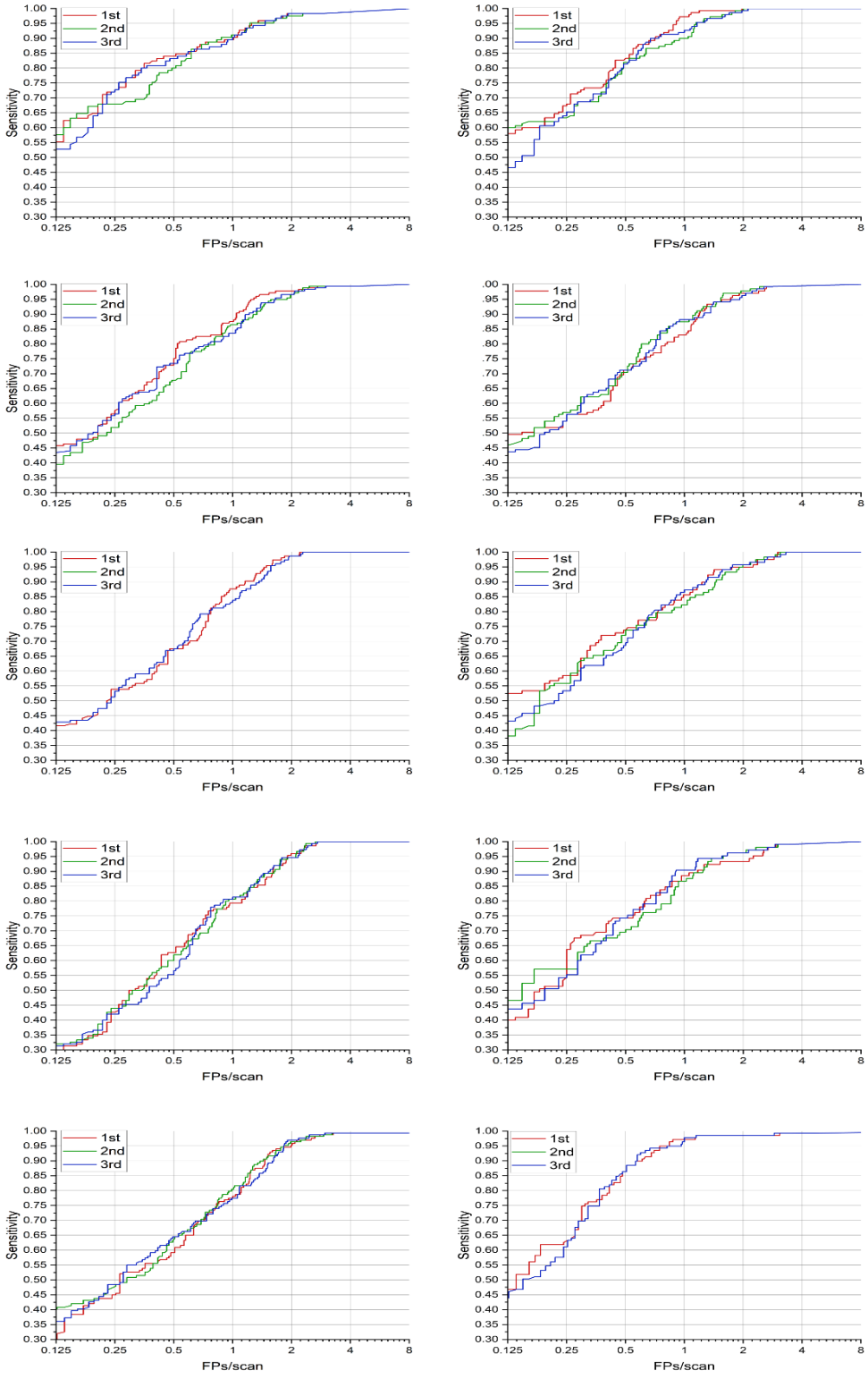


Figure 5.2 FROC curves of Stage B, obtained in the 10 folds in the 10-fold cross-validation. Each fold has been repeated 3 times, resulting in 3 curves.

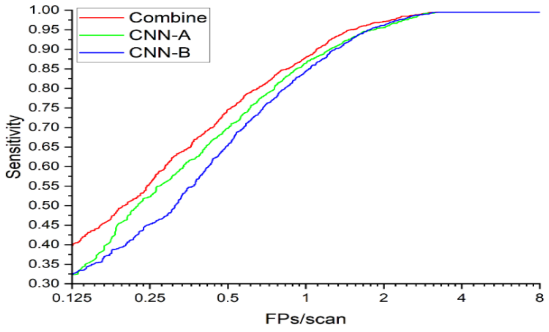


Figure 5.3 FROC curves resulting from the process of the 10-fold cross-validation repeated 3 times. The blue curve is given by the CNN for small nodules, the green one by that for large nodules, and the red one by the 2 CNNs combined.

Table 5.3 Average values of Sensitivity of Stage B, at 7 different False Positive per Scan (FPS), resulting from the 10-Fold cross-validation repeated 3 times

FPs/scan / fold	0.125	0.25	0.5	1	2	4	8
0	0.54	0.70	0.82	0.90	0.97	0.98	0.99
1	0.54	0.65	0.82	0.93	0.98	0.98	0.99
2	0.42	0.55	0.72	0.86	0.97	0.98	0.99
3	0.45	0.56	0.71	0.86	0.97	0.98	0.99
4	0.40	0.53	0.67	0.84	0.98	0.99	0.99
5	0.45	0.56	0.72	0.85	0.95	0.98	0.99
6	0.30	0.43	0.60	0.80	0.95	0.98	0.99
7	0.43	0.57	0.73	0.88	0.95	0.98	0.99
8	0.36	0.47	0.63	0.79	0.96	0.98	0.99
9	0.45	0.62	0.87	0.95	0.97	0.98	0.99
Average	0.40	0.55	0.72	0.88	0.97	0.98	0.99

Table 5.4 Comparison of Stage B

FPs/scan / system	0.125	0.25	0.5	1	2	4	8	Params
Xie et al.[32] ***	0.73	0.74	0.76	0.80	0.82	0.83	0.83	>1.08M
Cai et al[33]***	0.47	0.70	0.85	0.88	0.89	0.89	0.89	N/A
SANet[34]*	0.71	0.80	0.86	0.90	0.94	0.95	0.96	N/A
Zhou et al.[30]*	0.74	0.84	0.90	0.92	0.94	0.95	0.95	2.27M
Zhao et al. [31]***	0.66	0.75	0.83	0.92	0.95	0.97	0.98	N/A
Proposed**	0.57	0.72	0.85	0.91	0.97	0.97	0.99	0.55M
Proposed****	0.46	0.58	0.72	0.87	0.97	0.98	0.99	0.55M

* Test results on one-tenth of the data samples in LUNA16 dataset.
** Test results (average values) on one-tenth of the data samples in LUNA16 dataset, repeated 3 times.
*** Test results of 10-fold cross-validation.
**** Test results (mean values) of the 10-fold cross-validation repeated 3 times.

5.3.3 Test Result of the Proposed Lung Nodule Detection System

The complete system for lung nodule detection, proposed in this thesis, consists of Stage A and Stage B. As mention previously, Stage A is to detect the locations of nodule candidates and Stage B is to identify true nodules from the candidates. The test of the entire system has been conducted and described in this sub-sub-section.

The testing procedure involves 2 steps. In the first step, Stage A is tested in the same way as described in Sub-sub-section 5.3.1. Then, from the locations identified as the nodule points in the first step, 7320 data samples of 64 x 64 x 15 pixels are taken and applied to Stage B as its inputs. It should be noted that, a vast majority in the 7320 data samples are false nodules.

In this test, Stage B is trained with the same dataset, i.e., LUNA 16. The 10-fold cross-validation has been applied. Each of the 10 folds results in a trained Stage B, and the 7320 input samples are then applied to this trained Stage B to produce its output. Figure 5.4 illustrates 10 sets of FROC curves obtained from the 10 folds, respectively. As the 10-fold cross-validation has been repeated 3 times to reduce the effect of randomness in the test, there are 3 curves in each set. The overall test results of the 10-fold cross-validation on the complete system are also presented in Table 5.5. Each value of Sensitivity in the table is the mean of the 3 values obtained from the 3 repeated 10-fold cross-validations.

It should be noted that the total number of trainable parameters in Stage A and Stage B combined is 707K. It is a very small percentage of that used by others. Thus, the proposed system delivers a good detection result at an extremely low computation cost.

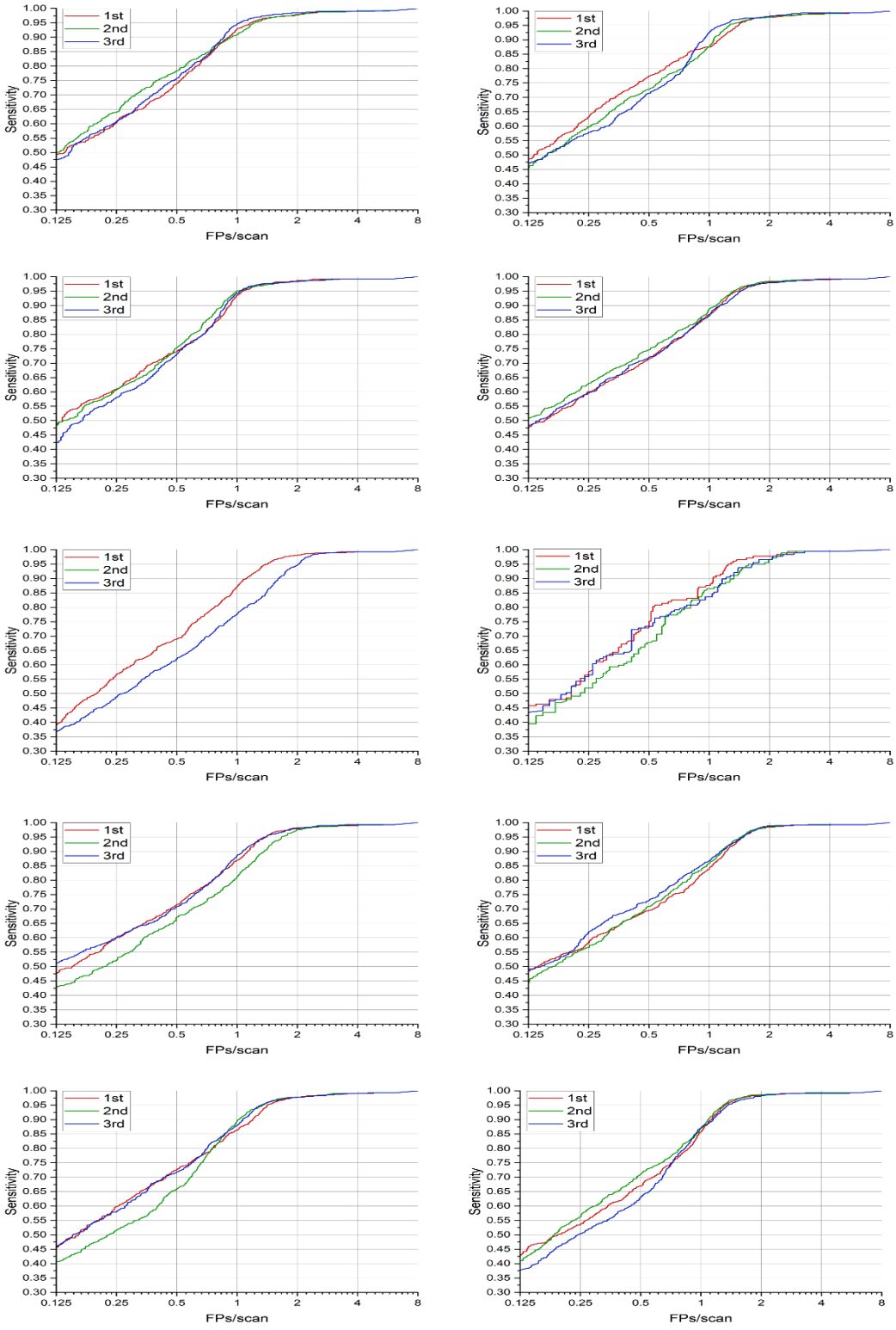


Figure 5.4 FROC curves of entire system, obtained in the 10 folds in the 10-fold cross-validation. Each fold has been repeated 3 times, resulting in 3 curves.

Table 5.5 Ten Test for Evaluate the Performance of Entire System

fold \ FPs/scan	0.125	0.25	0.5	1	2	4
0	0.50	0.64	0.77	0.95	0.97	0.98
1	0.48	0.63	0.76	0.93	0.97	0.98
2	0.48	0.57	0.75	0.87	0.97	0.98
3	0.51	0.62	0.75	0.87	0.96	0.98
4	0.40	0.56	0.68	0.87	0.97	0.99
5	0.45	0.58	0.73	0.85	0.97	0.99
6	0.51	0.60	0.71	0.88	0.97	0.99
7	0.48	0.63	0.73	0.86	0.97	0.99
8	0.41	0.47	0.64	0.80	0.96	0.98
9	0.46	0.60	0.72	0.90	0.98	0.99
Average	0.46	0.58	0.72	0.87	0.97	0.98

5.4 Summary

In this chapter, the performance evaluation of Stage A, Stage B and the complete system for lung nodule detection has been presented. The performance measurement has been done by 10-fold cross-validation. The results demonstrate that all of the proposed networks deliver good processing quality while the computation complexity is very low. Thus, they can be implemented easily in various systems.

6. Conclusion

Lung nodules are hard to detect due to (i) their enormous variations in nodule regions and (ii) high similarity to other objects in the image. Developing computer vision systems to handle this kind of detection, one needs to deal with the localization of nodule regions and reduction of false detection.

The objective of the work presented in this thesis is to develop CNN systems for high-quality lung nodules detection at the lowest computation cost, with a view to facilitating their implementation and practical applications. To achieve this, instead of using more complex network structures, we propose a system of multi-CNNs specifically for this specific detection.

The proposed system consists of 2 stages, namely Stage A and Stage B. Stage A is designed to localize the nodule candidates and the design aims at a high sensitivity in order to minimize the miss rate. Stage B is to identify the true nodules from the input samples. It can be used to identify falsely detected nodule samples from the output of Stage A, and also as a stand-alone lung nodule recognition system.

In Stage A, there are three blocks, i.e., a pre-processing block, custom-design CNN block and refinement block. The core of this stage is the custom-designed and U-net-based CNN block. The filtering modules in its convolution layers are specifically designed to suit the features produced in these layers. To reduce the data loss in the first four layers, Full-ReLU is used as the activation function. Furthermore, the refinement block is placed to reduce effectively the false positive rate. The computation complexity of Stage A is very low, as its total number of trainable parameters is only 0.16 M. Stage A delivers a high detection rate of 95.38% but the false positive rate is still as high as 6.9 FPs/scan. The output data will be applied to Stage B for further processing.

The design of Stage B is focused on distinguishing between the true nodules and their look-likes. Based on our analysis on the characters carried by nodules of different sizes, we propose to have 2 networks in Stage B for large and small nodule categories, respectively. The feature extraction in the 2 CNNs should be different, one targeting the variations in object regions of large nodules and the other looking more into nodule surroundings in case of small nodules. Two

CNNs have been designed and each of them has a particular multi-branch feature extraction (FE) block for the designated nodule category. Each CNN also involves fully-connected layers for classification. Stage B has been tested as a stand-alone lung nodule recognition system on LUNA 16 dataset. The results demonstrate that, with respect to similar systems found recently in literature, Stage B provides a good processing quality at a computation cost that is only a very small fraction of that is needed by others.

The complete system for lung nodule detection, i.e., Stage A and Stage B combined, has also been tested with the same dataset. The results demonstrate the good functionality of the system. All these CNNs combined require 0.7M parameters, far less that other CNN systems performing the same task.

In summary, the proposed system has been custom-designed to optimize the computation efficiency, i.e., achieving a good detection quality at the lowest computation cost. To attain this goal, the design strategy is to decompose the complex task of lung nodule detection into subtasks so that the system can employs multiple simple CNNs, each performing a sub-task. In this way, each CNN can be structured to suit the characters of a particular kind of nodule data and optimised to meet specific performance requirements. The effectiveness of this strategy has been confirmed by the results of the performance evaluation. Because of its low computation cost, the proposed system can be very easily implemented in various environment.

References

[1] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D. and Bray, F. (2015), Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer, 136: E359-E386

[2] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C. , Qing, D. P. Y. , Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Casteele, A. V., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., Clarke, L. P. (2015). **Data From LIDC-IDRI** [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX

[3] Torres, E Lopez et al. “Large scale validation of the M5L lung CAD on heterogeneous CT datasets.” *Medical physics* vol. 42,4 (2015): 1477-89. doi:10.1118/1.4907970 H. Yuan, Y. Wu, J. Cheng, Z. Fan and Z. Zeng, "Pulmonary Nodule Detection Using 3-D Residual U-Net Oriented Context-Guided Attention and Multi-Branch Classification Network," in IEEE Access, vol. 10, pp. 82-98, 2022, doi: 10.1109/ACCESS.2021.3137317.

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

[5] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,”

[6] *Proc. of International Conference on Machine Learning*, pp. 807–814, 2010.

[7] M. N. Gibbs and D. J. MacKay, “Variational gaussian process classifiers,” *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1458–1464, 2000.

[8] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale imagerecognition,” *arXiv:1409.1556*, 2014.

[9] C. Szegedy et al., "Going deeper with convolutions," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2015, pp. 1-9.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2016, pp. 2818-2826.

[11] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2016, pp. 770-778.

[12] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, July 2017, pp. 2261-2269.

[13] Y. Ma and C. Wang, "SdcNet: A Computation-Efficient CNN for Object Recognition," 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), Shanghai, China, 2018, pp. 1-5, doi: 10.1109/ICDSP.2018.8631567.

[14] D. Ciregan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642-3649

[15] Y. LeCun, Y. Bengio, and G. Hinton, "*Deep learning*," *Nature*, vol. 521, no. 7553, p. 436-444, 2015.

[16] Robbins, Herbert; Monro, Sutton. "A Stochastic Approximation Method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400-407, 1951.

[17] Diederik P. Kingma, Jimmy Ba, "Adam: A method for stochastic optimization", *arXiv:1412.6980*, 2014.

[18] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2015, pp. 3431-3440.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Proc. MICCAI, Nov. 2015, pp. 234–241.

[20] H. Yuan, Y. Wu, J. Cheng, Z. Fan and Z. Zeng, "Pulmonary Nodule Detection Using 3-D Residual U-Net Oriented Context-Guided Attention and Multi-Branch Classification Network," in *IEEE Access*, vol. 10, pp. 82-98, 2022, doi: 10.1109/ACCESS.2021.3137317.

[21] G. Zhang, H. Zhang, Y. Yao and Q. Shen, "Attention-Guided Feature Extraction and Multiscale Feature Fusion 3D ResNet for Automated Pulmonary Nodule Detection," in IEEE Access, vol. 10, pp. 61530-61543, 2022, doi: 10.1109/ACCESS.2022.3182104.

[22] Ding, Jia, et al. "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks." Medical Image Computing and Computer Assisted Intervention− MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20. Springer International Publishing, 2017.

[23] Z. Guo, L. Zhao, J. Yuan and H. Yu, "MSANet: Multiscale Aggregation Network Integrating Spatial and Channel Information for Lung Nodule Detection," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 6, pp. 2547-2558, June 2022, doi: 10.1109/JBHI.2021.3131671.

[24] C. C. Nguyen, G. S. Tran, V. T. Nguyen, J. -C. Burie and T. P. Nghiem, "Pulmonary Nodule Detection Based on Faster R-CNN With Adaptive Anchor Box," in IEEE Access, vol. 9, pp. 154740-154751, 2021, doi: 10.1109/ACCESS.2021.3128942.

[25] B. Liu, H. Song, Q. Li, Y. Lin and J. Yang, "3D ARCNN: An Asymmetric Residual CNN for Decreasing False Positive Rate of Lung Nodules Detection," 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, 2022, pp. 1644-1647, doi: 10.1109/BIBM55620.2022.9994973.

[26] Https://Luna16.Grand-Challenge.Org/Hom

[27] Sun, Yanming, and Chunyan Wang. "A computation-efficient CNN system for high-quality brain tumor segmentation." Biomedical Signal Processing and Control 74 (2022): 103475.

[28] S. Zheng, J. Guo, X. Cui, R. N. J. Veldhuis, M. Oudkerk and P. M. A. van Ooijen, "Automatic Pulmonary Nodule Detection in CT Scans Using Convolutional Neural Networks Based on Maximum Intensity Projection," in IEEE Transactions on Medical

Imaging, vol. 39, no. 3, pp. 797-805, March 2020, doi: 10.1109/TMI.2019.2935553.

[29] F. R. Pereira, J. M. C. De Andrade, D. L. Escuissato and L. F. De Oliveira, "Classifier Ensemble Based on Computed Tomography Attenuation Patterns for Computer-Aided Detection System," in IEEE Access, vol. 9, pp. 123134-123145, 2021, doi: 10.1109/ACCESS.2021.3109860.

[30] Z. Zhou, F. Gou, Y. Tan and J. Wu, "A Cascaded Multi-Stage Framework for Automatic Detection and Segmentation of Pulmonary Nodules in Developing Countries," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 11, pp. 5619-5630, Nov. 2022, doi: 10.1109/JBHI.2022.3198509.

[31] Zhao, Dandan, et al. "An attentive and adaptive 3D CNN for automatic pulmonary nodule detection in CT image." *Expert Systems with Applications* 211 (2023): 118672.

[32] Xie, Hongtao, et al. "Automated pulmonary nodule detection in CT images using deep convolutional neural networks." Pattern Recognition 85 (2019): 109-119.

[33] L. Cai, T. Long, Y. Dai and Y. Huang, "Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis," in IEEE Access, vol. 8, pp. 44400-44409, 2020, doi: 10.1109/ACCESS.2020.2976432.

[34] Mei, Jie et al. “SANet: A Slice-Aware Network for Pulmonary Nodule Detection.” IEEE transactions on pattern analysis and machine intelligence vol. 44,8 (2022): 4374-4387. doi:10.1109/TPAMI.2021.3065086.