UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# DEVELOPING MEDICAL IMAGE SEGMENTATION AND COMPUTER-AIDED DIAGNOSIS SYSTEMS USING DEEP NEURAL NETWORKS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

YUNZHI WANG
Norman, Oklahoma
2018

# DEVELOPING MEDICAL IMAGE SEGMENTATION AND COMPUTER-AIDED DIAGNOSIS SYSTEMS USING DEEP NEURAL NETWORKS

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

Dr. Bin Zheng, Chair

Dr. Le Gruenwald

Dr. Hong Liu

Dr. Joseph Havlicek

Dr. Liangzhong Xiang

# Acknowledgements

Firstly, I would like to thank my advisor, Dr. Bin Zheng. He provides me with a great research environment and spends lots of time discussing and revising my research works. My dissertation would have been unachievable without his support, guidance and suggestions throughout my Ph.D. studies.

I would like to thank my committee members, Dr. Le Gruenwald, Dr. Hong Liu, Dr. Joseph Havlicek and Dr. Liangzhong Xiang, for their great helps and suggestions. I would also like to thank my friends and colleagues in our research lab, Dr. Maxine Tan, Faranak, Gopi, Nafiseh, Morteza and Linsheng. Without them it would be more difficult to complete the dissertation.

I would like to thank my mentors and colleagues during my internships, Dr. Dashan Gao, Dr. Jiao Wang, Tianyi Zhao and Haichao Yu from 12 Sigma Technologies and Zhiqiang Hu, Jiahui Li and Shuang Yang from Sensetime, for their great guidance and collaborations.

I would like to thank my wife Mandy for her accompanies and encouragements throughout the past five years and in the future. I also thank my lovely cats Yuanxiao and Yuanbao. They provide me with a warm family here.

Finally, I would like to express my deepest gratitude to my parents for their unconditional love and support. They are the reason I work hard every day to become better.

# Table of Contents

# List of Tables

# List of Figures

## Abstract

Diagnostic medical imaging is an important non-invasive tool in medicine. It provides doctors (i.e., radiologists) with rich diagnostic information in clinical practice. Computer-aided diagnosis (CAD) schemes aim to provide a tool to assist the doctors for reading and interpreting medical images. Traditional CAD schemes are based on hand-crafted features and shallow supervised learning algorithms. They are greatly limited by the difficulties of accurate region segmentation and effective feature extraction. In this dissertation, our motivation is to apply deep learning techniques to address these challenges. We comprehensively investigated the feasibilities of applying deep learning technique to develop medical image segmentation and computer-aided diagnosis schemes for different imaging modalities and different tasks. First, we applied a two-step convolutional neural network architecture for selection of abdomen part and segmentation of subtypes of adipose tissue from abdominal CT images. We demonstrated high agreement between the segmentation generated by human and by our proposed deep learning models. Second, we explored to combine transfer learning technique with traditional hand-crafted features to improve the accuracy of breast mass classification from digital mammograms. Our results show that the ensemble of hand-crafted features and transferred features yields improvement of prediction performances. Third, we proposed a 3D fully convolutional network architecture with a novel coarse-to-fine residual module for prostate segmentation from MRI. State-of-art segmentation accuracy was obtained by using this model. We also investigated the feasibilities of applying fully convolutional network for prostate cancer detection based on multi-parametric MRI and obtained promising detection accuracy. Last, we proposed

a novel cascaded neural network architecture with post-processing steps for nuclear segmentation from histology images. Superiority of the model was demonstrated by experiments. In summary, these study results demonstrated that deep learning is a very promising technology to help significantly improve efficacy of developing computer-aided diagnosis schemes of medical images and achieve higher performance.

# Chapter 1. Introduction

## 1.1 Background

Diagnostic medical imaging is an important non-invasive tool in medicine. Imaging techniques including X-ray, ultrasound and magnetic resonance imaging (MRI) etc. provide rich information for radiologists to make diagnosis and/or treatment planning [1]. Traditionally, most image processing and interpretation processes were performed subjectively by radiologists. However, processing/analysis of medical images by human eyes and experience has a number of limitations including that (1) manual processes are always time-consuming and therefore cannot deal with current large-scale clinical datasets based data analysis tasks, and (2) the measurement or analysis may not be consistent because of the inter and intra reader variabilities in reading and interpreting medical images. As a result, development of computer-aided diagnosis (CAD) systems has been attracting great research interests in the last two to three decades. The motivation of developing CAD systems is to apply digital image processing and/or artificial intelligence techniques to assist the radiologists for more accurately and consistently reading, analyzing and interpreting the diagnostic images. CAD has been applied for a wide range of imaging modalities and diseases, as well as for conducting of variety of tasks (i.e., segmentation of regions of interest, detection of abnormalities, diagnosis of diseases or classification of malignant and benign lesions, and assessment of disease treatment results) [2].

For example, Lee et al. employed a CAD scheme embedded with a Support Vector Machine (SVM) classifier for brain tumor segmentation [3]; Pal et al. developed

a multi-stage Artificial Neural Network (ANN) based CAD system for micro-calcification detection from digitalized mammograms [4]; Gray et al. investigated the feasibility of using a CAD scheme with Random Forest based similarity measures for classification of Alzheimer's disease [5]; etc. In these CAD systems of medical images, machine learning, especially supervised learning method, is one of the key computerized technologies and has been extensively applied in assisting detection and/or diagnosis of different human diseases.

Traditional supervised learning based CAD systems usually consists of a few steps. The first step is region of interest (ROI) segmentation and common segmentation methods include region-growing, graph-cut and level-set etc. The second step is hand-crafted feature extraction, while the commonly used features include shape and texture features. Since the extracted features might be either redundant or unrelated to the classification task, the next step is to apply feature reduction method (e.g. principle component analysis (PCA) and recursive feature elimination (RFE)) to reduce the data dimensionality. The last step is to train a supervised learning model (e.g. SVM and ANN) to discriminate different classes. Although significant research efforts have been focused on the development of such CAD systems, the performance is greatly limited by the difficulties of (1) accurate ROI segmentation and (2) how to design effective features to discriminate different classes.

Recently, deep learning [6] technology has gained tremendous research efforts. The availability of large-scale data set and affordable high-speed computation resources (i.e. Graphics Processing Units (GPUs)) have greatly accelerated the development of deep learning methods in the last ten years. Compared to conventional machine learning

models, deep learning techniques provide a classification scheme that can automatically extract hierarchical feature representations from raw input data without knowledge of feature engineering [7]. Therefore, the difficulties of designing/selecting useful features can be avoided in developing deep learning based CAD systems. In computer vision area, deep learning has been proven to out-perform other conventional machine learning tools for ImageNet classification [8, 9] and object detection [10, 11] etc. Following the great success in computer vision area, deep learning is expected to be a promising tool for solving many difficult medical imaging problems, such as image segmentation, computer-aided diagnosis and image retrieval, and etc.

## 1.2 Introduction to deep learning

### 1.2.1 Logistic Regression and Artificial Neural Networks

Logistic Regression classifier is the simplest Neural Networks which contains only one single neuron. The concept of "a computational neuron" is a computation unit that takes the weighted summation of inputs and generates an output through a non-linear activation function. Logistic Regression was proposed as a supervised learning model, which aims to solve classification problems with discrete outputs (i.e. labels). By using a single layer of neurons with a sigmoid activation function, the logistic regression model can map an input feature vector to a continuous output vector with values in the range of [0, 1], indicating the probability of the input belonging to each class. The formula for a two-class Logistic Regression model can be written as:

$$f(x) = sigmoid(W^T x + b) = \frac{1}{1+\exp(W^T x+b)} \quad (1)$$

Where $W$ is the weight vector which has the same size with the feature vector $x$, and $b$ is the bias. The values of $W$ and $b$ are randomly initialized and then optimized by minimizing a cost function (e.g. negative log-likelihood) using gradient descent or stochastic gradient descent (SGD) methods. The model with optimized parameters can be applied to predict future inputs with unknown labels.

Logistic Regression is a "linear" classifier, which divided the input feature spaces into half-spaces and made predictions based on the linear combination of input features. Although linear classifiers have the advantages of low computational complexity and high robustness to over-fitting issues, they require sophisticatedly designed features that are linearly separable in the feature space, which is often quite difficult in many applications including medical imaging.

Alternatively, a non-linear classifier can be established by combining a number of simple neurons to form a Multiple Layer Perceptron (MLP) [12]. Due to the limitations of hardware and computerized techniques, traditional MLPs only have three layers. The first layer of the network is the input layer representing the input feature vectors. The second layer is called hidden layer since its values are not observable in the training set. Neurons in the input layer are fully connected to neurons in the hidden layer and the neurons in the hidden layer are connected to an output neuron with sigmoid activation function. The purpose of introducing a hidden layer is to non-linearly map from the input feature space to a hidden feature space, with the hope that the optimized hidden features are linearly separable. The formula of a two-layer MLP can be written as [13]:

$$f(x) = G(b^{(2)} + W^{(2)T}(s(b^{(1)} + W^{(1)T}x))) \quad (2)$$

Where *b(1)*, *W(1)*, *b(2)*, and *W(2)* are the bias and weight matrix of the first layer and second layer respectively; G is the activation function of the output layer, which is usually a sigmoid function and s is the activation function of the hidden layer. Typical choices for s include tanh, sigmoid or rectified linear unit (ReLU). The set of parameters are optimized by minimizing the cost function (e.g. log-likelihood) through back-propagation algorithm and SGD, where back-propagation is a special form of chain-rule derivation.

### 1.2.2 Convolutional Neural Network

In a regular MLP architecture, the units in the input layer are fully connected to units in the hidden layer. When using relatively large images (e.g. with size 96×96 pixels) as the inputs of network, there is a lot of parameters in the fully connected neural networks. The computation of a back-propagation algorithm for learning such large amount of parameters is computationally expensive and over-fitting may occur due to the relatively small training set in particular in the medical imaging field.

Inspired by cat's visual cortex that contains neurons with localized receptive fields [14], Convolutional Neural Network (CNN) was proposed as a variant version of standard MLP. Specifically, each hidden unit is only connected to a small sub-region of the previous layer (e.g. an 8×8 patch) instead of being fully connected to all the neurons in the previous layer. The "filters" corresponding to the local weight matrices are convolved with feature maps in the previous layer to obtain a new feature map representing feature activations at local positions in each image [12]. A standard CNN architecture consists of three types of layers including convolutional layer, pooling layer and fully connected layer. Convolutional layers are the core part of CNN. A number of

rectangular convolutional filters (i.e. weight matrices) are randomly initialized and learnable during the training process. The filters can be interpreted as local feature extractors that are automatically optimized from training data. Convolutional layer performs convolution operations between the input maps and the filters followed by a non-linear transformation to obtain output feature maps. The second type of layer is pooling layer, where max-pooling is most commonly used. Max-pooling operation takes the maximum values over sub-rectangular regions of features maps to form smaller feature maps. The spatial redundancy and number of parameters are greatly reduced by applying pooling layers. Several Convolutional-Pooling layer pairs are stacked to extract high level feature representations; these features are fed to standard MLP classifiers to generate prediction results. Different CNN architectures have been proposed such as LeNet 5 [15], Alex Net [8] and Google Net [16] etc. Standard back-propagation methods can be applied for training CNNs.

### 1.2.3 Transfer Learning

Training deep neural networks with many layers requires large scale training sets. However, in many real world applications including medical imaging, it is rare to have training set with sufficiently large size. The concept of transfer learning was proposed to overcome this limitation. Specifically, transfer learning consists of two steps including a "pre-training" step followed by a "fine-tuning" process. In the first step, the deep network was pre-trained using a large dataset with sufficient size but not necessarily similar to the target dataset (with limited size). The network for classifying target dataset was initialized with the parameters of the pre-trained network. Standard

back-propagation algorithm was performed on the network to fine-tune the parameters in high-level layers.

Transfer learning is motivated by the observation that low-level features at earlier layers of the network always contain generic features (e.g. edges or boundary in images) no matter how different the tasks are, while high-level features in the latter layers are more related to the specific task. By pre-training the network using large dataset, the parameters in earlier layers get sufficiently trained and can be transferred to new tasks without any changes. In the second step to fine-tune the network supervisingly, the parameters in later layers get to figure out how to integrate lower-level features effectively for the specific task. During fine-tuning process, the first several layers are always fixed. As a result, the size of parameters is smaller and risk of over-fitting issues can be reduced. Transfer learning has been extensively applied in many different area including natural language processing [17], visual recognition [18, 19] and medical image analysis [20], and etc.

### 1.3 Objective

As stated previously, traditional CAD systems are limited by the challenges of accurate ROI segmentation and effective feature designing and selection. In order to overcome these limitations, the goal of this dissertation is to comprehensively investigate the feasibility of applying deep learning technology to develop medical image segmentation and computer-aided diagnosis systems for different imaging modalities and diseases/cancers. Several state-of-art deep learning architectures were adopted for building the CAD systems and we further improved the CAD performance by proposing and testing novel network structures.

## 1.4 Organization of Dissertation

In this dissertation, we present five applications of applying deep learning technology for different medical image analysis tasks (e.g. semantic segmentation, instance segmentation and classification) and different imaging modalities (e.g. CT, MRI, mammography and pathology images). In Chapter 2, a two-step convolutional neural network architecture was developed for selection of abdomen part and segmentation of subtypes of adipose tissues from abdominal CT images. In Chapter 3, we combined deep transfer learning technique with traditional hand-crafted features based method for improving the accuracy of mammographic mass classification. In Chapter 4, we proposed a novel 3D fully convolutional network architecture with coarse-to-fine residual module for prostate segmentation from MRI. In Chapter 5, we investigated the feasibilities of applying fully convolutional network for prostate cancer detection based on multi-parametric MRI. In Chapter 6, we proposed a novel cascaded neural network architecture with post-processing steps for nuclear segmentation from digital histology images. Last, in Chapter 7, we summarized these development and application studies to present a conclusion of the whole work in this dissertation.

# Chapter 2. A two-step Convolutional Neural Network based Computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images

## 2.1 Introduction

Abdominal obesity is one of the most prevalent public health problems and over one third of adults were obese in the United States in recent years [21]. Obesity is strongly associated with many different diseases such as heart diseases, metabolic disorders, type 2 diabetes and certain types of cancers [21-23]. Inside a human body, there are subcutaneous fat areas (SFA) and visceral fat areas (VFA), which both contribute to the abdominal obesity. Studies have shown that in the clinical practice separate measurement or quantification of subtypes of adipose tissue in SFA and VFA is crucial for obesity assessment since visceral fat is more closely related to risk factors for hypertension, coronary artery disease, metabolic syndrome, and etc. [24, 25]. Other studies have also found that measurement of the total fat volume and/or the ratio between the VFA and SFA could generate useful clinical markers to assess response of cancer patients to the chemotherapies, in particular to many antiangiogenic therapies [26-28].

Among different imaging techniques for adiposity tissue detection and measurement, computed tomography (CT) has been most widely adopted because of its higher accuracy and reproducibility [29]. Accurate segmentation and quantification of SFA and VFA from CT slices is important for clinical diagnosis and prediction of disease (or cancer) treatment efficacy. Currently, manual or semi-automated

segmentation of SFA and VFA in a single subjectively chosen CT image slice has been adopted to determine fat areas and measure adiposity related features as demonstrated in the previous studies [27, 30]. However, this approach has a number of limitations including that 1) manual manipulations are time-consuming and cannot deal with large amount of data; 2) fat area measured from a single CT slice may not accurately correlate to the total fat volume of a human body; 3) measurement may also not be consistent due to the inter and intra reader variability in selecting CT slice and segmenting SFA and/or VFA areas. Therefore, developing a computer-aided detection (CAD) scheme for fully automated segmentation and quantification of SFA and VFA is necessary [10].

Recently, deep learning methods, especially deep convolutional neural networks (CNN), have gained extensive research interests and proven to be the state of art in a number of computer vision applications [7, 8, 15, 31, 32]. Following the tremendous applications of deep learning in computer vision area, there are a couple of previous works that successfully employed CNN methods to solve medical image analysis and CAD related problems [20, 33-39]. For example, Roth et al [38] developed a multi-level deep CNN model for automated pancreas segmentation from CT scans; Brebisson et al [33] used a combination of 2D and 3D patches as the input of CNN for brain segmentation; Yan et al [39] applied a multi-instance deep learning framework to discover discriminative local anatomies for body part recognition; etc.

In this study, we developed a two-step CNN based CAD scheme for automated segmentation and quantification of SFA and VFA from abdominal CT scans. The new CAD scheme consists of two different CNNs, while the first one is used to

automatically select and collect CT slices belonging to abdomen area from the whole CT scan series (i.e., the perfusion CT images acquired from ovarian cancer patients, which are scanned from lung to pelvis crossing the entire abdomen region), and the second CNN is used for automated segmentation of SFA and VFA in each single CT slice. While there has been a number of previously published studies that focused on automated quantification of visceral and subcutaneous adipose tissue by using combinations of traditional image processing techniques (e.g. thresholding, labelling and morphological operation) [25, 28, 40-45], previous works have a number of limitations including that: 1) the selection of CT slice range of interest (i.e. abdomen area in this study) is either manually processed or not mentioned; 2) the optimal values of some parameters (e.g. morphological operation kernel size) in some of these models may not be consistent for different patients and thus human intervention might be necessary for tuning these parameters. Our proposed CNN based CAD scheme aims to overcome these limitations and achieve fully automated segmentation since (1) the first CNN was developed for automated selection of CT slices belonging to abdomen area, and (2) there is no parameters needed to be tuned in our CAD scheme after the two CNNs are sufficiently trained. The details of this study including the development of CAD scheme and performance evaluation are presented in the following sections.

## 2.2 Materials and Methods

### 2.2.1 Overview of the CAD framework

Our proposed CAD framework consists of two steps with two CNNs, namely the Selection-CNN and Segmentation-CNN. The flowchart for demonstrating the whole process of this system is shown in Figure 1. The first Selection-CNN is applied to

automatically select and collect CT slices of interest (i.e. abdomen area in this study) for each patient. The selected slices will then be used as the input of the second Segmentation-CNN for automated segmentation of SFA and VFA. The SFA and VFA volumes segmented from all CT slices of interest will be combined to compute and measure the adiposity characteristics of the particular patient at the very end step. The proposed CAD system is fully automated and no human intervention is needed. In the next two sections we will discuss the details of the CNN architectures, training process and evaluation methods for the two CNNs, respectively.

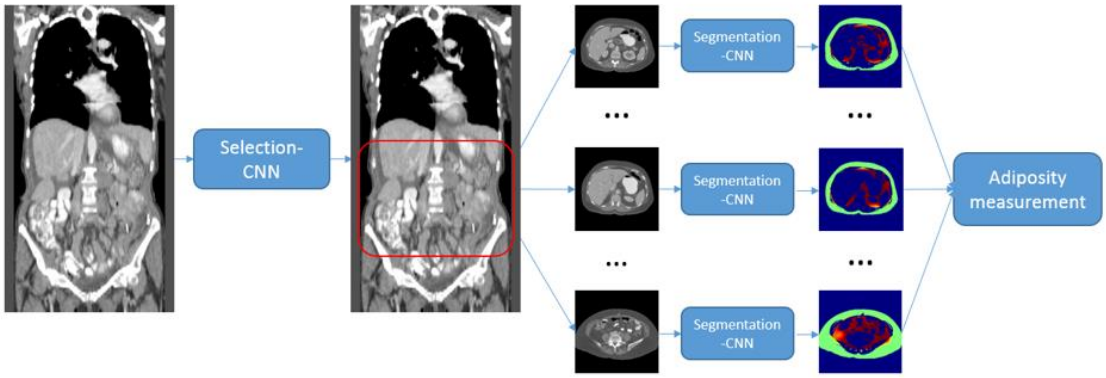


**Figure 1:** Overview of the two-step CNN based CAD scheme for adipose tissue quantification

### 2.2.2 A CT image dataset

In this study, we randomly assemble an image dataset, which consists of CT images acquired from 40 ovarian cancer patients who underwent cancer chemotherapy treatment in the Health Science Center of our University. The detailed image acquisition protocol has been reported in our previous publication [27, 28]. In brief, all CT scans were done using either a GE LightSpeed VCT 64-detector or a GE Discovery 600 16-detector CT machine. The X-ray power output was set at 120 kVp and a variable

range from 100 to 600mA depending on patient body size. CT image slice thickness or spacing is 5mm and the images were reconstructed using a GE "Standard" image reconstruction kernel. Next, these 40 patients were randomly and equally divided into two groups namely, a training patient group and a testing patient group. CT image data from training patient group were used to train two CNNs used in the CAD scheme and the data from the testing patient group were used to evaluate the performance of the trained CAD scheme.

### 2.2.3 Training and testing dataset for Selection-CNN

In order to train Selection-CNN model to "learn" how to discriminate CT slices as belonging to abdomen area or not, an observer manually identified CT slices that belong to the abdomen area for each of the twenty patients in the training patient group. Specifically, an upper bound was subjectively placed just below the lung area and a lower bound was placed at the umbilicus level. All CT slices between the two bounds were labeled as positive (i.e. belong to slices of interest) and other CT slices were labeled as negative (i.e. not belong to slices of interest). By doing this, we can collect a sample of 2,240 CT slices as the training set of Selection-CNN. Among them, 757 are "positive slices" located inside the abdominal region and 1,483 are "negative slices" located outside the abdominal region.

Although in order to optimally train a machine learning classifier, many previous studies chose to use "balanced" training datasets with the equal number of sample cases in two classes, the "balanced" approach has a limitation of potential sampling bias in selecting and removing part of sampling cases from the class with more samples. Thus, to maintain the diversity of all image slices, we used all 2,240

image slices in the training dataset although the number of "positive" and "negative" slices was different (or not balanced). Next, to verify the advantage of using all available training samples, we also conducted experiments to compare the classification accuracy between using "balanced" and "non-balanced" training datasets with 50 repeated training and validation tests in each training condition.

### 2.2.4 Training and testing dataset for Segmentation-CNN

Similar to Selection-CNN, Segmentation-CNN also needs some subjectively processed images as the training samples to train the classifier. For each of the twenty patients in the training patient group, 6 CT slices from the abdomen area were randomly selected for the training purpose (totally 120 CT images). A previously developed and tested CT image segmentation method was applied to remove the background (e.g. air and CT bed) and generate a body trunk mask. All the pixels inside the body trunk mask with CT number between the threshold -40 HU and -140 HU were defined as adipose tissue pixels [30]. An observer manually drew a boundary that contains the entire visceral area. The adipose tissue pixels inside the visceral area were labeled as VFA pixels, whereas the adipose tissue pixels outside the visceral area were labeled as SFA pixels.

A Matlab based graphic user interface (GUI) program is developed for implementation of image pre-processing step and manual segmentation process. Figure 2 shows an example of the working procedure of the GUI. After loading a CT image from the local hard drive, the original image is shown in the left figure of the program. The "Body Seg" button has been designed to implement the body trunk segmentation algorithm and display the body trunk image in the middle figure of the program. Then,

an observer can draw a boundary (the blue line in the middle image) that contains the entire visceral area. The segmented SFA/VFA result is then shown in the right figure of the GUI window.

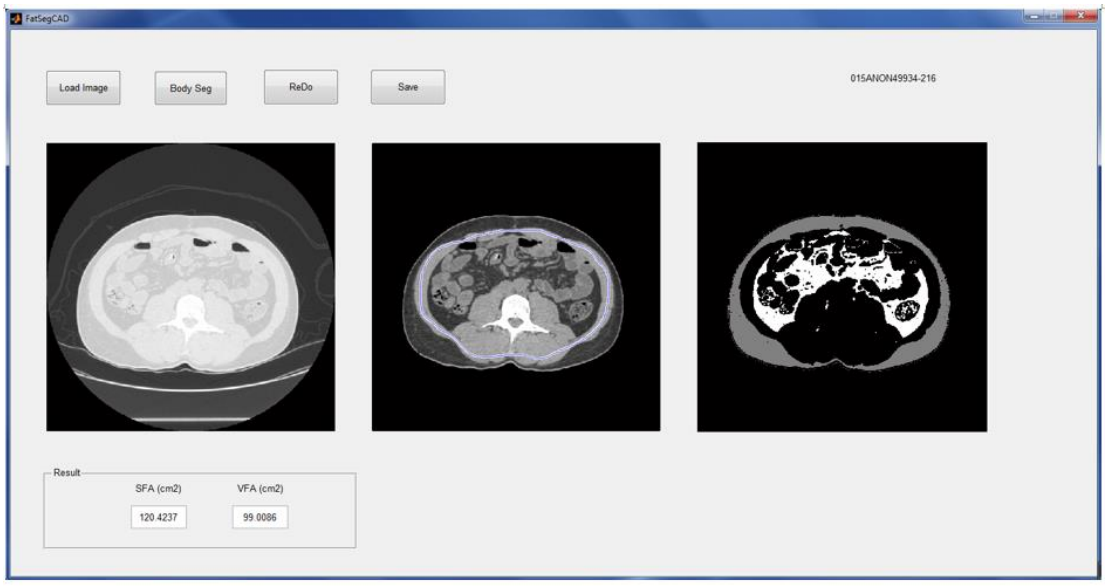


**Figure 2:** Demonstration of the Matlab GUI program for implementation of image pre-processing and manual segmentation process

Subsequently, 700 adipose tissue pixels (belonging to either VFA or SFA pixel class) were randomly selected from each of the 120 CT images for training Segmentation-CNN. By doing this, we can collect a sample of 84,000 adipose tissue pixels as the training set. Among them, 64,691 are labeled as SFA pixels and 19,309 are VFA pixels. The goal is to train the Segmentation-CNN to recognize or distinguish pixels with the CT number between -40 HU and -140 HU into SFA or VFA areas. By using the same segmentation criterion, 120 CT images from the 20 patients in the testing patient group were randomly selected and manually labelled for the purpose of evaluating performance of this segmentation-CNN.

### 2.2.5 Selection-CNN

In the first step, Selection-CNN is developed and used to select the CT slices belonging to abdomen area, which will be collected for the computation of adipose tissue volume and features. We formulate this task as a binary classification problem. Specifically, we use each single CT slice image as the input of a classifier (Selection-CNN) and determine whether the CT slice belongs to the abdomen area or not. In this way, each CT slice is processed independently; the location and spatial consistency are not considered. In order to overcome this limitation, a simple post-processing step is performed to ensure the spatial consistency and dependence after obtaining the raw Selection-CNN output.

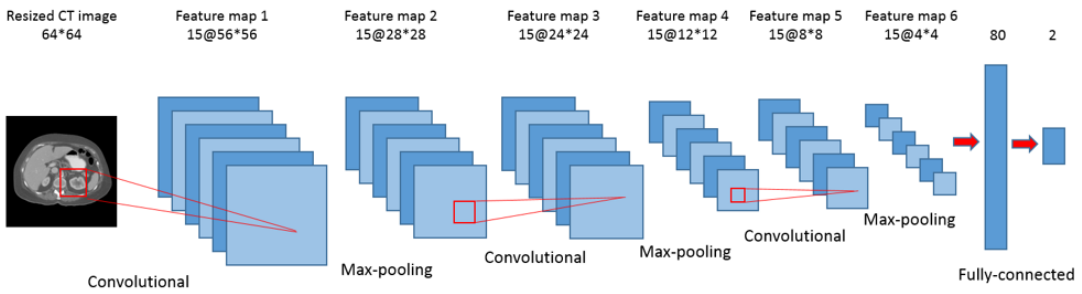


**Figure 3:** Architecture of Selection-CNN

The architecture of Selection-CNN is shown in Figure 2.3 and it was developed based on LeNet, which was designed by LeCun et al. for image recognition [15]. First, the original 512×512 CT images are resized to 64×64 using an 8×8 averaging kernel. This step can be interpreted as either a down-sampling based image pre-processing step or a "pooling" layer in CNN architecture. By doing this, the input size and number of parameters of Selection-CNN is greatly reduced, which can potentially improve the

training efficiency and reduce the risk of over-fitting. The down-sampled CT images are then used as input into a standard CNN architecture consisting of three convolutional-pooling layers, one fully connected layer, and one soft-max layer. Convolutional layers are the core part of CNN architecture. They consist of a number of rectangular convolutional filters; the parameters of these filters are randomly initialized and learnable during the training process. Each convolutional layer performs two-dimensional convolutional operations between the input image maps and the convolutional filters followed by a non-linear transformation. The convolutional layers can be interpreted as automatic feature extractors that are optimized from training data, and thus the outputs of convolutional layers are referred as "feature maps".

Following convolutional layers, max-pooling layers are commonly performed in a number of CNN architectures [7, 8, 15]. The operation of max-pooling layer is to take the maximum values over sub-windows of feature maps, which can greatly reduce the spatial redundancy and the number of parameters. Several convolutional-pooling layer pairs can be stacked to get high-level feature representations. These features are then used as input into a standard Multi-Layer Perceptron (MLP) classifier, which consists of a fully connected hidden layer and a soft-max layer. The Selection-CNN developed in this study contains three convolutional-pooling layers. The numbers of feature maps are 15 for all the three layers and the filter sizes are 9×9, 5×5 and 5×5, respectively. A tanh function is applied for non-linear transformation. The size of max-pooling is 2×2 for all layers. The fully connected layer contains 80 hidden neurons and the soft-max layer contains 2 output neurons (i.e. positive or negative).

As a result, the Selection-CNN architecture maps each 512×512 CT image slice to a vector of two continuous numbers between 0 and 1, indicating the probabilities of the input image belonging to positive or negative classes. Considering that the size of training set of Selection-CNN is relatively small compared to many of other computer vision datasets using deep learning (e.g. MNIST and ImageNet), we take following measures to avoid the potential over-fitting problem, which include that (1) the numbers of filters for each convolutional layer and hidden layer are set to be smaller than the commonly used CNNs, and (2) an L2 regularization term is adopted as a part of loss function.

### 2.2.6 Segmentation-CNN

After CT slices belonging to abdomen area are selected, the second step is to develop and apply a Segmentation-CNN based scheme to segment SFA and VFA depicted on each single CT slice. This task is also formulated as a binary classification problem. Specifically, an image pre-processing step including a previously developed body trunk segmentation scheme [46] and a well-defined thresholding process for adipose tissue identification [30] is applied to identify all the pixels belonging to the fat area (namely adipose tissue pixels) in each CT image; a classifier (Segmentation-CNN in this study) is then trained and applied to classify each adipose tissue pixel as belonging to SFA or VFA by using their neighborhood pixels and location information as input.

There have been a couple of previous works that applied CNN methods for medical image segmentation tasks [33, 34, 36-38]. The basic method is a patch-wise classification process. Specifically, in order to classify each single pixel into its correct

class, a 2-D rectangle patch centered at the specific pixel is extracted, resized and used as the input of a CNN classifier. The outputs of CNN indicates the probability of the center pixels belonging to each class. The advantage of patch-wise classification is that by extracting separate and possible overlapping image patches, we can collect a training set that is large enough for the requirement of deep CNN architectures. However, spatial consistency and pixel location information are lost to some degree in this basic method. Many approaches have been applied to overcome this limitation, including post-processing, increasing the size of image patches and using location information as part of the input of CNN, etc. [33, 34, 37, 38].

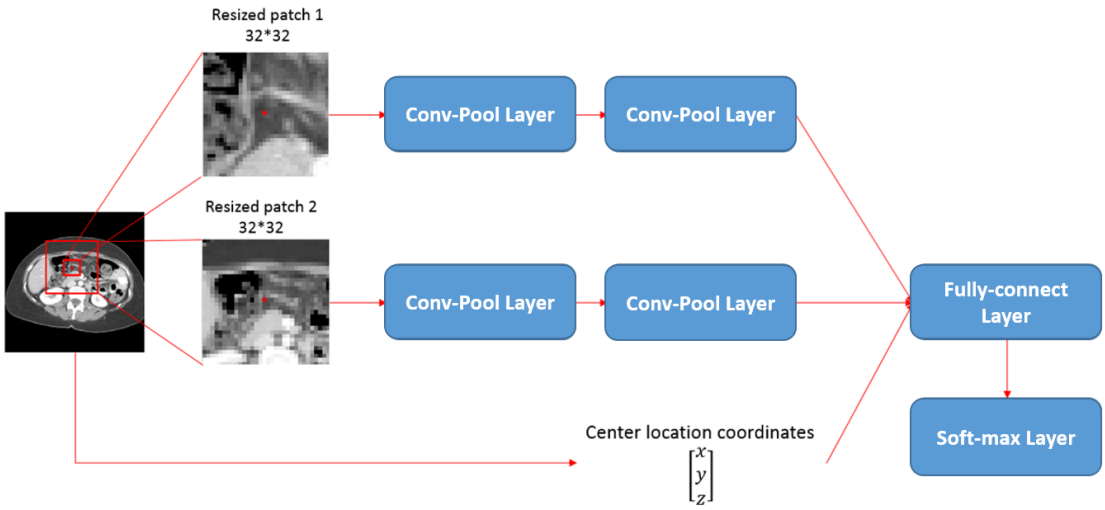


**Figure 4:** Architecture of Segmentation-CNN

In this study, we developed the Segmentation-CNN architecture as shown in Figure 4. The objective of Segmentation-CNN is to classify an adipose tissue pixel as belonging to either SFA or VFA. It is a 2-scale CNN architecture where the patches with smaller size are used to represent fine details around the target pixel and patches with larger size are used to maintain global consistency. The patch sizes of 64×64 and

128×128 are used because experimental results indicated that using these two patch sizes yield better results than others. Multi-scale CNN architecture is not considered because it will generate larger network and take longer time to train and execute. In addition, it may also suffer from the over-fitting problems.

The input of Segmentation-CNN consists of three parts. The first part is a 32×32 rectangular patch obtained by down-sampling a 64×64 patch centered at the adipose tissue pixel depicted on the CT slice, which is the representation of fine details around the target adipose tissue pixel. Two convolutional-pooling layer pairs are stacked to get high-level image patch features. The numbers of feature maps are 50 for both two layers. The filter sizes are 5×5 and 3×3, respectively, while the sizes of max-pooling were 2×2. The second part of Segmentation-CNN input is a 32×32 rectangular patch obtained by down-sampling a 192×192 patch centered at the adipose tissue pixel, which is designed to preserve more spatial consistency. Same convolutional-pooling layer pairs are employed to get high-level features. The third part is a 3-D vector that contained the normalized and adjusted spatial location coordinates of the adipose tissue pixel. A fully connected layer with 2,000 hidden neurons is used to fuse the information and high-level features from the three parts of input. The dimension of the input of fully-connected layer is 50 (feature maps) × 6×6 (size of each feature map after two conv-pool layers) × 2 (two convolutional channel) + 3 (location coordinates). A soft-max layer is finally applied to generate the likelihood based prediction scores.

In addition, we also evaluate the performance of a baseline CNN for comparison. The baseline CNN only consists of one channel which is a resized to a 32×32 rectangular patch obtained from a 64×64 patch in the image. Two convolutional-

pooling layer pairs and one fully-connected layer were stacked to build the network. The numbers of feature maps were 50 for both the convolutional layers. The filter sizes were 5×5 and 3×3, respectively, and the fully-connected layer contained 1,000 hidden neurons.

### 2.3 Experiments and evaluation

#### 2.3.1 Evaluation of Selection-CNN

The loss function of Selection-CNN is set to be the summation of a negative log-likelihood term and a L2 regularization term. Mini-batch stochastic gradient descent (SGD) methods are applied to minimize the loss function. SGD is one of the most popular and widely used training methods in machine learning (especially in deep learning) applications [16]. It is more efficient for large-scale learning problems compared to some other training methods such as the second-order approaches. Specifically, the training set is split into a number of batches; the gradient of loss function was estimated over each batch instead of the whole training set. By using mini-batch SGD, the parameters get more frequent updates and training efficiency can be greatly improved. In this study, we set mini-batch size equal to 50 and iteratively trained the Selection-CNN for 50 epochs aiming to obtain the optimal parameters.

After obtaining the raw Selection-CNN output, a post-processing step is performed to check and ensure the spatial consistency. Specifically, a one-dimensional median filter is performed to smooth the outputs (i.e. the probabilities of each CT slice belonging to positive or negative class). The longest consecutive CT slices with probabilities of being positive greater than 0.5 are predicted as positive and the remaining slices were predicted as negative.

For each CT case in the testing group, we compare the manually processed labels and the results generated by Selection-CNN with the post-processing step. We compute the prediction accuracy, sensitivity and specificity for each testing case and averaged them over all 20 cases in the testing group for the selection-CNN performance evaluation.

### 2.3.2 Evaluation of Segmentation-CNN

Similar to Selection-CNN, a mini-batch SGD is employed to train the Segmentation-CNN network to get optimal parameters. The loss function of Segmentation-CNN is solely a negative log-likelihood function since we have collected enough training samples. The size of mini-batch was set to 500 and the iteration time was set to 400 epochs.

For each input (i.e. a 512×512 CT image) in the testing dataset, CAD scheme applies the body trunk segmentation algorithm [46] to remove the background. The pixels inside the body trunk are scanned one by one. If the pixel has a CT number between the threshold -140 HU and -40 HU, a neighborhood patch and location information are extracted and used as input of Segmentation-CNN. A likelihood score generated by Segmentation-CNN is used to label the pixel as belonging to either SFA or VFA. In order to evaluate the performance of this Segmentation-CNN based scheme, 120 CT images from the 20 patients in the testing patient group are randomly selected and manually labelled. For each CT image, the segmentation results generated subjectively and by Segmentation-CNN based scheme are compared. Pixel-wise prediction accuracy and dice coefficients (DC) of SFA/VFA are calculated for each image and averaged over the 120 CT images for performance evaluation. Dice coefficient is a similarity

measurement index commonly used to evaluate the performance of image segmentation tasks. It calculates the ratio of overlapping volume between two segmented areas. The formula of DC with respect to two segmentation areas *A* and *B* is shown as below.

$$DC = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

**2.4 Results**

The training and evaluation process of CNNs were performed on a Dell T3610 workstation equipped with a quadcore 3.00GHz processor, 8 Gb RAM and a NVidia Quadro 600 GPU card. The models were implemented in Python using Theano library [47]. Figure 5 shows an example of the comparison between the abdomen area labelled subjectively by an observer and generated by the optimized Selection-CNN. Table 1 summarized and compared the quantitative performance evaluation results (i.e. accuracy and DC) of Selection-CNN with and without post-processing. Specifically, by using individual CT scans independently as the input of Selection-CNN, we can obtain averaged prediction accuracy and Dice Coefficients over 90%. Adding the post-processing steps enabled to further improve performance by taking the spatial consistence into account. Finally, the study results yielded a mean prediction accuracy equal to 0.9582 with standard deviation 0.0268, mean sensitivity equal to 0.9481 with standard deviation 0.0595 and mean specificity equal to 0.9625 with standard deviation 0.0521, respectively. The improvement is statistically significant by using paired t-tests ($p < 0.005$ for all three evaluation indices).

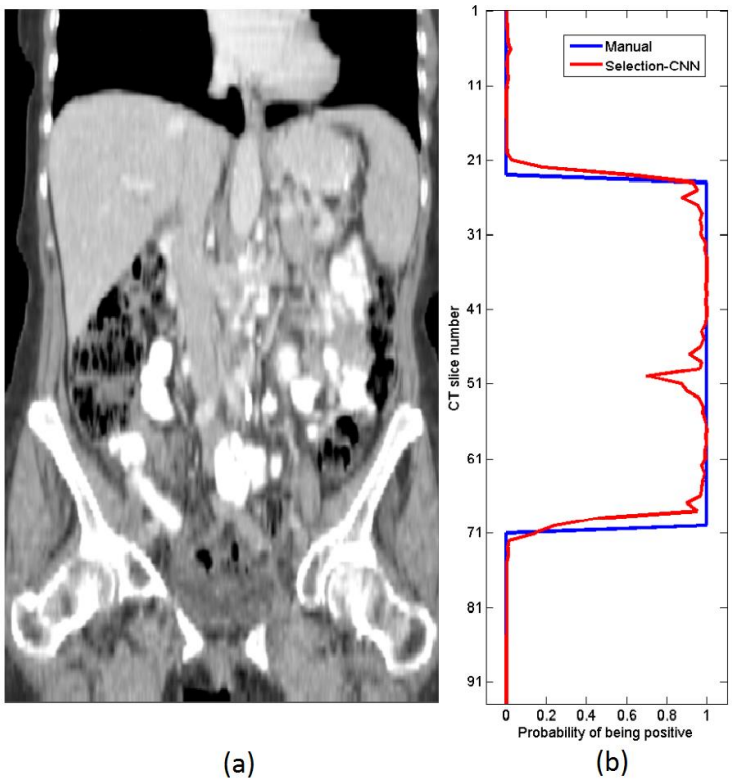


**Figure 5:** Examples of abdomen area selected by Selection-CNN. (a) A patient CT scan from vertical view. (b) Comparison of abdomen area selected by an observer (blue line) and by optimized Selection-CNN (red line)

**Table 1:** Summarization of quantitative performance evaluation of Selection-CNN without and with post-processing

| | Prediction accuracy | Sensitivity | Specificity |
|---|---|---|---|
| Selection-CNN | 0.9352 ± 0.0344 | 0.9287 ± 0.0690 | 0.9362 ± 0.0722 |
| Selection-CNN with post-processing | 0.9582 ± 0.0268 | 0.9481 ± 0.0595 | 0.9625 ± 0.0521 |

Table 2 summarized and compared the performance evaluation results of the proposed Segmentation-CNN and a baseline CNN architecture which only used the 64×64 neighborhood patches as input. It shows that the performance of Segmentation-CNN is statistically significantly better than the baseline CNN ($p < 0.005$ for all the

three evaluation methods). Figure 6 shows two examples of segmentation results generated by Segmentation-CNN for segmenting SFA and VFA depicting on CT image slices. Figure 7 shows two scatter plots of the volumes of manually labelled SFA/VFA and CAD measured SFA/VFA among all the 120 testing CT images. The correlation coefficients are 0.9980 with 95% confidence interval (CI) (0.9972, 0.9986) for SFA and 0.9799 with 95% CI (0.9712, 0.9859) for VFA respectively.

**Table 2:** Summarization of quantitative performance evaluation of baseline CNN and Segmentation-CNN

| | Pixel-wise prediction accuracy | SFA Dice coefficient | VFA Dice coefficient |
|---|---|---|---|
| A baseline CNN | 0.9535 ± 0.0253 | 0.9696 ± 0.0175 | 0.8890 ± 0.0563 |
| Segmentation-CNN | 0.9682 ± 0.0218 | 0.9797 ± 0.0145 | 0.9150 ± 0.0624 |

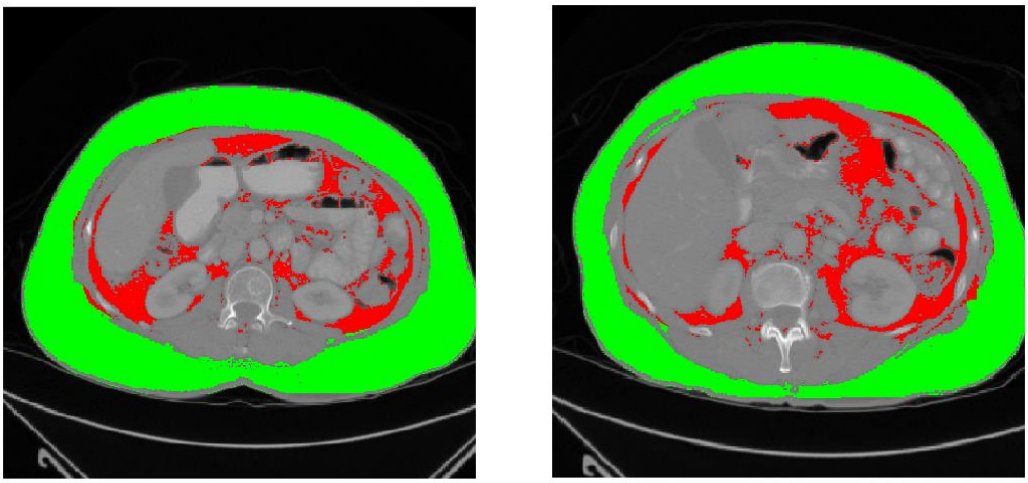


**Figure 6:** Examples showing the segmentation of VFA and SFA generated by Segmentation-CNN in two CT image slices. In these two images, SFA is shown in green color and VFA is represented by red color.

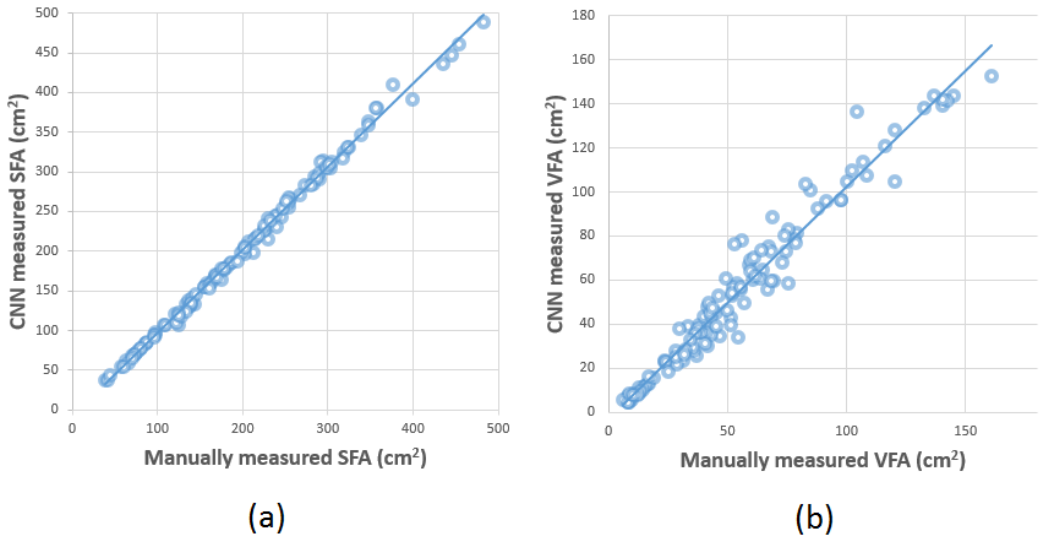


**Figure 7:** The scatter plots of the manually and automatically (i.e. by Segmentation-CNN) measured (a) SFA and (b) VFA volume.

To evaluate the stability of our CAD scheme, following are the results of our two experiments. First, by repeated training the Selection-CNN with different random initialized weights for 50 times, the Selection-CNNs without post-processing steps yielded a mean prediction accuracy of 0.9324, which is quite similar to the result reported in Table 1 (0.9352). The highest, median and lowest accuracies among the 50 experiments are 0.9456, 0.9328 and 0.9167, respectively. Second, the 50 pairs of repeated experiment that compared the performance of the trained Selection-CNN using balanced versus unbalanced datasets yielded an equal mean prediction accuracy of 0.9324. However, the standard deviations were 0.0119 and 0.0066, when using 757 pairs of "balanced" training samples and all available 2,240 "unbalanced" training samples, respectively, which indicates that using all available training samples increase diversity of training dataset and yielded more stable testing results.

## 2.5 Discussion

In this study, for the first time we developed a deep learning based fully-automated CAD scheme and demonstrated its feasibility to automatically segment volumetric SFA and VFA data without any human intervention. This study and the new CAD scheme have a number of unique characteristics. First, a Selection-CNN with post-processing step was developed for automated selection of CT slices belonging to abdomen areas. This CNN based process can not only overcome the limitation of manual selection in most previous studies, which were quite difficult and time-consuming to deal with large-scale datasets, and also generated high segmentation accuracy as compared to manually processed results or "ground-truth" (i.e. yielding a prediction accuracy and DC greater than 0.95). Therefore, the Selection-CNN for automated selection of abdomen CT slices is reliable and can be used to replace manual selection, which provide the capability of managing large-scale dataset based medical data analysis studies with high efficiency.

Second, in most of the previously developed schemes, the segmentation of SFA and VFA was obtained by detecting visceral masks or abdomen wall masks using sequences of traditional image processing techniques such as morphological operations, pixel labelling and thresholding [25, 28, 40-45]. Therefore, the segmentation results might be sensitive to the selection of parameters (e.g. morphological operation kernels and distance thresholds); these parameters were mostly subjectively determined and the optimal values might be different for different patients. In this study, we developed a Segmentation-CNN based scheme to segment SFA and VFA, which was based on machine learning classifiers and thus provided a more "intelligent" way by considering

the location coordinates and neighborhood information. After sufficiently trained and optimized, the CNN model is free of parameters and no human intervention is required for getting optimal segmentation results. Therefore, the Segmentation-CNN provided a reliable CAD scheme for fully-automated segmentation of SFA and VFA from single CT slices.

Third, the two tasks in this study (i.e. selection and segmentation) were both formulated as binary classification problems and Convolutional neural networks were employed as the classifier to solve the problems. In traditional machine learning classifiers (e.g. Support Vector Machines and random forests) based systems, how to design and select effective and discriminative features is a crucial but difficult task. The advantage of convolutional neural network is that it can automatically learn hierarchical feature representations from its raw input images and therefore, no manual feature extraction and selection process is needed [7]. Following the success of CNNs in many other computer vision and medical image analysis areas, this study demonstrated that CNN models are effective for recognizing CT slices that belong to abdomen areas and segmenting SFA and VFA from single CT slices.

Fourth, we further investigated and employed new approaches of adding post-processing and a 2-scale network to maintain global consistency and improve prediction accuracy. The study results showed that after using these new approaches CAD scheme enabled to yield significantly higher segmentation accuracy than the raw CNN outputs. Thus, this study provided an example of how to optimally apply CNN with consideration of spatial or location information in developing a deep learning based scheme. The stability of the deep learning based CAD scheme has also been tested and

approved by the experiments using repeated random initializations and different training datasets with balanced and unbalanced training samples in two classes.

In addition, based on the experimental results, we also made several observations. For example, (1) although direct comparison of segmentation accuracy between using this new deep learning scheme (3D data) and previous semi-automated schemes (2D data) is difficult, the new automated scheme can achieve high segmentation accuracy as comparing to the general visual segmentation. (2) Unlike the conventional machine learning methods, which should be optimally trained using a balanced dataset with equal number of training samples in two classes, a deep learning based CAD scheme can be optimally trained without such a restriction. Thus, the deep learning scheme may have an advantage to build a more stable classification model using all available training samples with increased diversity.

Despite the encouraging results, this is a preliminary technology development study with a number of limitations. First, we applied and evaluated a commonly used CNN architecture, activation functions, loss functions and the training methods in this study. The performance of this CAD system can be potentially improved by employing other advanced deep learning models and methodologies. Second, it took a relatively long time (i.e. a few minutes) to segment SFA and VFA in each single CT slice. This is because that each adipose pixel is used as an independent input of the deep network and the CAD scheme needs to scan all the pixels in the image. More research efforts should be devoted to investigate how to improve the computational efficiency of this system. For example, applying an optimal sampling and/or super-pixel concept might be helpful for reducing the amount of pixel-wise classification process in the system. Thus, at

current stage, this deep learning based CAD scheme can only be used offline to process the images and segment SFA and VFA from the images. Last, the clinical potential of this new technology needs to be evaluated in the clinical studies to validate whether using volumetric adiposity-related image features can significantly improve accuracy in predicting disease prognosis (e.g. response of ovarian cancer patients to chemotherapies) as compared to previous manual or semi-automated methods that measured adiposity from one selected single CT slice.

In summary, in order to overcome the limitation of estimating SFA and VFA from one subjectively selected single CT image slice, we developed and tested a new CAD scheme for adipose tissue segmentation and quantification based on a sequential two-step process including (1) selecting CT slices belonging to abdomen areas and (2) segmenting SFA and VFA from each of the selected CT slices. We demonstrated that applying this new deep learning CNN based CAD scheme enabled to recognize specific body parts (abdomen in this study) from volumetric CT image data and segment SFA and VFA from each selected CT slice with high accuracy (i.e. >90%) or agreement with manual segmentation results. As a result, this study provided researchers a new and reliable CAD tool to assist processing volumetric CT data and quantitatively computing a new adiposity related imaging marker in the future clinical practice.

# Chapter 3. Combining Transfer Learning and Hand-crafted Features for Breast Mass Classification from Mammograms

## 3.1 Introduction

Breast cancer is the most common cancer occurred in women population worldwide with high mortality rates. According to the 2016 cancer statistics provided by American Cancer Society [48], breast cancer accounted for 29% of women cancer diagnosis and the death rates were 36% in the United States. Scientific evidence has indicated that mortality and recurrence rates of breast cancer can be greatly reduced by early cancer detection and treatment [49]. Currently, mammography is an only clinically accepted population-based breast cancer screening tool aiming to early detect breast cancer. However, reading and interpreting mammograms is a difficult task for radiologists, in particular to accurately classify between benign and malignant lesions, because of the dense fibro-glandular tissues overlapping as well as the large heterogeneity of breast lesions, in particular the mass-like lesions [50]. As a result, currently the false positive recall rate of screening mammography and the associated negative biopsy rate are high in breast cancer screening practice, which generate both physical and psychosocial harms to many cancer-free women participating in the routine mammography screening [51]. Hence, how to reduce the false positive rates and improve the efficacy of screening mammography is an important clinical challenge for early breast cancer detection.

In the past twenty years, developing machine learning based computer-aided detection and/or diagnosis (CAD) systems of mammograms has gained extensive

research efforts [52]. The objective of developing CAD systems is to provide a "second reader" to assist radiologists in their decision making process when reading and interpreting mammograms. CAD schemes for mammographic lesion detection have been commercially available since later 1990s. Such CAD systems aim to detect suspicious areas containing either micro-calcification clusters or mass-like lesions depicting on the mammograms. These CAD schemes also generate a relatively higher false positive cueing rates. Therefore, developing CAD based classification schemes to distinguish between benign and malignant mammographic masses to reduce false-positive recall rates has also been continuously attracting much research interest during last two decades [53-57]. Most of the previous CAD based classification schemes rely on hand-crafted feature extraction and traditional machine learning classifiers.

Methodologically, the framework of such systems usually consists of four steps namely, lesion segmentation, feature extraction, feature selection and training a supervised machine learning model. First, automated image segmentations algorithms (e.g. region-growing and level-set) are applied to segment the suspicious lesion area from the background. Then hand-crafted features (e.g. morphology and/or texture features), which are subjectively designed with the hope that they can jointly distinguish different classes, are computed and extracted from the segmented area. Feature selection methods are subsequently applied to select the subset of extracted features which are most discriminatory. The selected feature subset is finally used to train a machine learning classifier (e.g. a support vector machine or artificial neural network) to classify suspicious lesions into benign and malignant classes. Despite significant research efforts have been focused on developing traditional machine learning based

CAD systems, how to design and select effective features to classify benign/malignant lesions still remains great challenge. No computer-aided lesion classification schemes are currently accepted and used in clinical practice.

Compared to traditional machine learning classifiers that have "shallow" architectures and cannot process the data in their original form [6], deep learning techniques can automatically learn hierarchical feature representations from raw input data. Therefore, deep learning techniques provide a classification scheme that does not require precise lesion segmentation, hand-crafted feature extraction and selection [7]. One outstanding issue of applying deep learning (e.g. CNN) for medical imaging tasks is that training a deep network usually requires a large dataset to avoid the potential over-fitting issue. For medical image segmentation tasks, the training set can be collected by cropping separate and partially overlapping image patches to satisfy the requirement of deep network architectures [58-61]. However, for image classification tasks such as mammographic mass-like lesion classification, which is the focus of this article, it is difficult to collect a training set that is large enough to train a complete deep learning network.

Previous works investigated the feasibility of optimizing a relatively shallow CNN architecture (e.g. contain 2 or 3 Convolutional layers) to classify between benign and malignant masses [62-64]. Although employing shallow network with few or small number of parameters can reduce the risk of over-fitting, it may not be capable of extracting high level representations from the raw images. Instead of training a CNN from scratch, other works applied the concept of "transfer learning" to solve similar tasks. Transfer learning includes a supervised pre-training step that optimizes the

parameter of a deep CNN with an independent dataset of sufficient size (e.g. ImageNet) [65-67]. The pre-trained network can either be fine-tuned with the target dataset or fixed as a feature extractor, depending on the size of the specific data in hand. Transfer learning is motivated by the fact that earlier layers in the network usually learn some generic image features (e.g. edge detector), while later layers tend to learn the features that are more related to a specific task.

Thus, the transfer learning based CAD schemes for mass classification have also been investigated in several previous studies. For example, Levy et al. compared the performance of training a baseline CNN with three convolutional layers and an Alex-Net [8] architecture pre-trained on ImageNet dataset [68]. Their results show that the pre-trained Alex-Net can yield significantly higher performance than the baseline CNN with a relatively shallow architecture; Jiao et al. pre-trained a deep learning network with five convolutional layers on the ILSVRC datasets and then fine-tuned the network on mammogram images [69]. In another example, Huynh et al. employed a pre-trained Alex-Net as a fixed feature extractor and trained a support vector machine (SVM) classifier with the transferred CNN features [70].

Although preliminarily satisfactory results were reported in these studies, we identified and/or highlighted several unsolved issues of how to optimally applying transfer learning for mammographic mass classification. First, the natural images (e.g. ImageNet) are color images with three channels (i.e. RGB), while mammograms are gray scale images with only one channel. Although gray scale images can be simply converted to color images by duplicating (as most previous studies did), it may provide redundant information to the pre-trained deep network. Second, large natural image

datasets (e.g. ImageNet) usually contain hundreds or thousands of image categories. The features learned in the mid-level layers may represent heterogeneous shape and color information, while most of them may not be actually related to mammogram images. Third, the interpretability of transfer learning is still poor for mammographic mass classification application.

In this study, we developed a new transfer learning based CAD system for mammographic mass classification. Specifically, we employed a pre-trained Alex-Net architecture as a fixed feature extractor; we generated pseudo-color images by combining the original mammogram images with their morphological and texture variations to incorporate more prior knowledges; we applied additional "filters" to eliminate un-related transferred CNN features in the network. Meanwhile, a traditional machine learning based CAD scheme was also performed in this study for (1) comparing the performance of traditional CAD and transfer learning based CAD systems; (2) investigating whether combination of these two types of classifiers can further improve the performance. The detailed methods and evaluation results of this study are presented as follows in the chapter.

### 3.2 A patient dataset

In this study, a reference dataset was retrospectively collected from an existing full-field digital mammography (FFDM) database in our laboratory. The detailed information regarding our FFDM image database has been reported in our previous studies (e.g., [56, 57]). Mammogram images acquired from 301 women participated in mammography screening were included in the dataset. All of these women were previously recalled by radiologists because suspicious soft tissue masses were detected

in their mammograms. According to the biopsy and pathological diagnosis, 149 of them were negative (i.e. benign masses) and 152 cases were positive (i.e. malignant masses). The cranio-caudal (CC) view mammograms of these cases were used in this study and the center of each mass has been identified by the radiologists.

### 3.3 A transfer learning based CAD system

In this study, we independently developed a transfer learning based mass classification system and a hand-crafted feature engineering based system for comparison purpose. The transfer learning based CAD system is introduced in this section. Building and running this system consists of four steps. First, Regions of Interest (ROIs) are extracted from the CC view mammogram images and pseudo-color ROIs with three channels are generated from the original gray-scale ROIs. Second, an Alex-Net which was pre-trained on the ImageNet dataset is employed and fixed as a feature extractor. Transferred CNN features are obtained by feeding the pseudo-color ROIs into the pre-trained Alex-Net. Third, dimensionality reduction methods are performed to eliminate the un-related transferred features and improve the system robustness. Last, a linear SVM is trained with the surviving features to discriminate between benign and malignant cases. Each of the four steps is further discussed accordingly in the following sections.

#### 3.3.1 Generation of three-channel pseudo-color ROIs

For each digital mammogram (image), a 64×64 patch centered at the point that was previously marked by the radiologists on the image is extracted as an ROI. However, Alex-Net takes color images with RGB channels as the input while the ROIs extracted from mammograms only have one channel. Two approaches are commonly

employed in order to overcome this issue in previous transfer learning based mass classification schemes. For the first approach, the gray-scale ROIs were duplicated and fed into the three channels simultaneously to obtain a three-channel image [70]. While for the second one, the natural images were transferred to gray-scale images and the Alex-Net was pre-trained with the transferred natural images [69]. Although these approaches can successfully fit the gray-scale ROIs to the Alex-Net architecture, the prediction power of Alex-Net is significantly reduced due to the information redundancy or parameter reducing. Alternatively, we propose an approach to generate pseudo-color ROIs by combining the original gray-scale ROIs with their variations. Similar approaches were previously applied for lung cancer diagnosis [71]. The advantages of using pseudo-color ROIs include (1) pseudo-color ROIs can fit the Alex-Net architecture without providing redundant information or reducing the number of parameters in the model; (2) prior knowledges or additional information can be leaked to the deep network by designing different ROI variations.

Specifically, for each ROI, we first generate a binary image to indicate the segmented mass area by using a previously developed region-growing based mass segmentation scheme [57]. The segmentation scheme applies a basic region-growing algorithm on a Gaussian-constraint image and calculated circularity and sharpness of the segmented area for parameter selection (Please refer to [57] for further details). Then, a texture image is generated using a local standard deviation filter, where each pixel in the texture image contains the standard deviation of a 3×3 local patch centered at the corresponding pixel in the original ROI. The original ROI, segmentation ROI and texture ROI are fed into RGB channels, respectively, to generate a pseudo-color image.

Figure 8 shows an example of the three input ROIs and the finally combined pseudo-color ROI. It should be noted that the segmentation ROI contains the information of the size and shape of the masses, while the texture ROI highlights the homogeneity and contrast characteristics of the masses. Therefore, by feeding the pseudo-color ROIs into the deep CNN network, we can provide additional information about the suspicious masses, which may potentially improve the performance of the transfer learning based system.

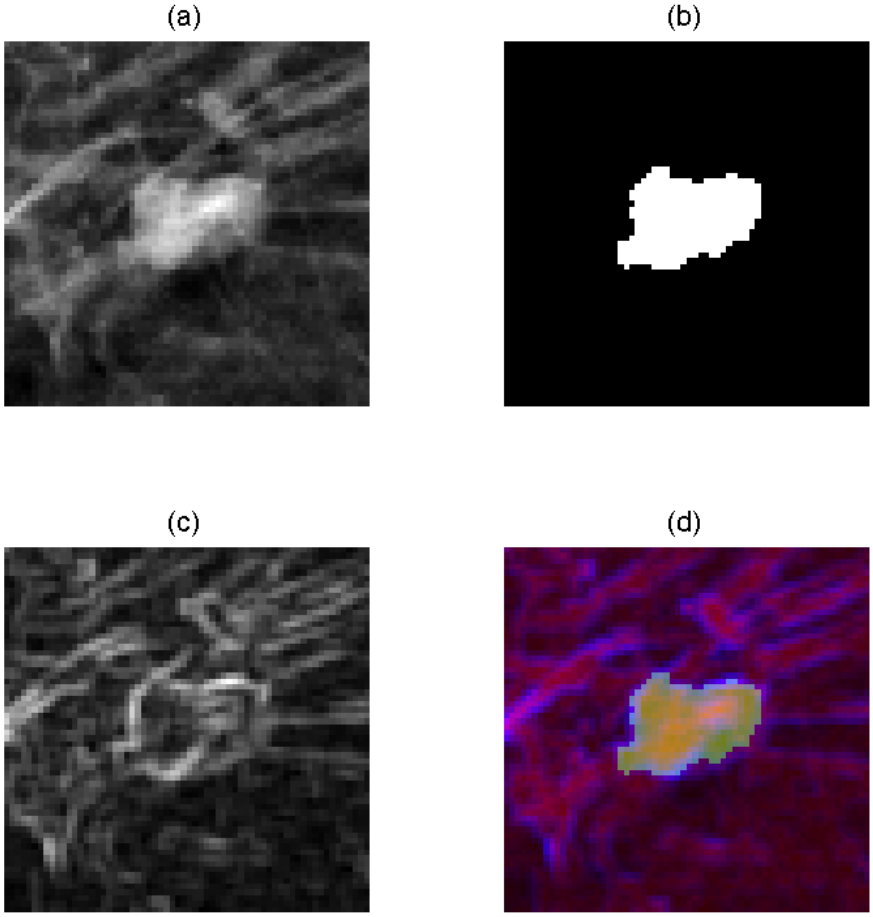


**Figure 8:** An example of: (a) Original ROI, (b) Segmentation ROI, (c) Texture ROI and (d) Combined pseudo-color ROI

### 3.3.2 Deep CNN architecture

The next step of this system is to extract transferred CNN features using a pre-trained Alex-Net. In traditional CAD systems, how to design effective and discriminatory hand-crafted features is a significant issue and great research efforts have been focused on it. Deep learning, especially deep CNN model provides an alternative approach for image classification where the optimal features can be automatically learned from the raw images without the process of feature engineering. Alex-Net is one of the most popular deep CNN architectures. It was originally developed for ImageNet classification and obtained state-of-art performance [8]. Figure 9 shows the architecture of Alex-Net. It includes following steps. First, the Alex-Net takes inputs of color images with RGB channels and size of 227×227. Second, five Convolutional Layers (namely Conv 1-5) with different filter size are stacked to extract high level local features. Three max pooling layers are applied after Conv1, Conv2 and Conv5 respectively for redundancy reduction. Third, two fully connected layers (namely FC6 and FC7) with rectified linear units (ReLU) activation are applied consecutively after the last Convolutional layer (i.e. Conv5) to generate high level global features. Finally, the last layer (i.e. FC8) of Alex-Net contains a soft-max layer with 1000 units, which stand for the 1000 categories in ImageNet dataset.

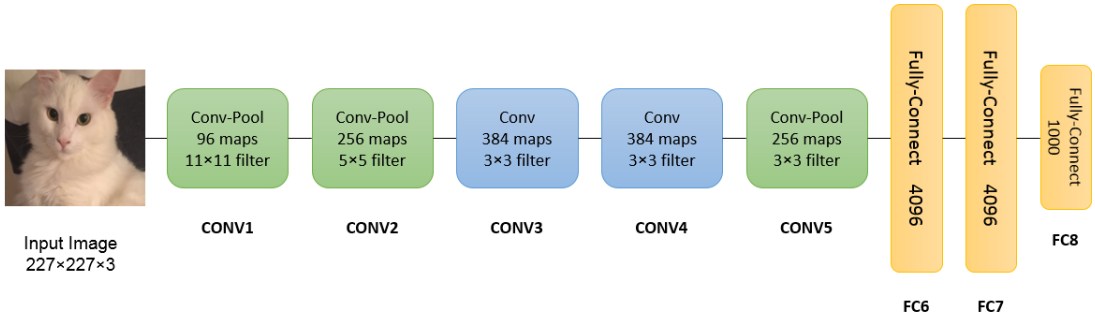


**Figure 9:** Architecture of Alex-Net

Although Alex-Net has the capability of modelling highly complex non-linear function, it requires a large dataset to train the network to avoid over-fitting issues. In medical imaging background, it is often difficult to collect a dataset that is large enough for training an Alex-Net from scratch. As introduced in earlier part, transfer learning is a commonly used approach for addressing this problem. The early layers in a deep CNN usually represent generic image features and therefore can be transferred between different tasks. In this study, we used the Caffe software tool [72] and a publicly available Alex-Net model that is pre-trained on the ImageNet dataset (https://github.com/BVLC/caffe/tree/master/models/bvlc_alexnet) as a fixed feature extractor. Supervised fine-tuning is not performed here because our mammogram dataset only contains 301 cases and fine-tuning the parameters may lead to over-fitting concerns. Specifically, the pseudo-color ROIs obtained in previous steps are resized to 224×224 and fed into the pre-trained Alex-Net. The activations of the hidden neurons in the FC6 and FC7 layers are calculated and extracted as transferred CNN features. We evaluate the performance of using the feature vector extracted from FC6 (i.e. 4096 features), FC7 (i.e. 4096 features) and the combination of them (i.e. 8192 features), respectively. The activations in the Convolutional and pooling layers in Alex-Net are

not considered for feature extraction because they contain relatively low-level local information with relatively high dimensionalities.

### 3.3.3 Dimensionality reduction

ImageNet dataset contains 1000 image categories with significantly different shape, texture and color. The neurons in the hidden layer of the pre-trained network may encode a large heterogeneity of features related to these characteristics. However, mammogram ROIs are relatively homogeneous to each other and therefore, many hidden neurons are actually not "activated" by the inputs of mammogram ROIs. Since ReLU activation function is applied in Alex-Net, we define that a hidden neuron or feature is "activated" if it has a strictly positive value and not activated if it has a zero value. Thus, we propose a simple activation-based feature selection approach for dimensionality reduction of transfer learning applied in CAD scheme. The hypothesis is that the transferred features that are activated by a majority of ROIs in the training set are more effective to describe the main characteristics of the input ROIs. Accordingly, the strategy of the activation-based approach is to eliminate the transferred features that are activated by less than one third of the ROIs in the training set, since these features are considerably un-related to the input ROIs.

As a comparison, traditional univariate feature selection methods were also applied to select most discriminative transferred features based on univariate statistical tests (i.e. Student t-test in this study). It should be emphasized that although the transferred features do not substantially follow normal distributions, the t-tests are still quite robust since the sample sizes of the two classes are nearly equal and fairly large (i.e. >30) [73]. Multi-variate feature selection methods such as Joint Mutual Information

[74] and Quadratic Programming Feature Selection [75] are not applied in this study because it is difficult and time-consuming to model the mutual relationship of feature vectors with over 4000 dimensions.

### 3.3.4 Training and evaluation

The last step of the CAD system is to train a machine learning classifier using the selected transferred CNN features. Considering that the feature vectors have a relatively high dimensionality and the size of training set is relatively small, a linear SVM classifier is trained to classify benign and malignant masses. A leave-one-case-out (LOCO) training and testing process is applied to evaluate the performance of the transfer learning based system. In each LOCO cycle, mammograms from 300 patients are collected for training the classifier, while the remaining one is used for performance evaluation. The processes of pseudo-color ROI generation, transferred CNN feature extraction and activation-based feature selection are performed consecutively over the training set that contains 300 ROIs. A linear SVM is then trained using the selected pseudo-color ROI based transferred CNN features and applied to the testing ROI to obtain a likelihood score. We repeat this cycle by 301 times and calculate the overall predication accuracy and area under receiver operating characteristic (ROC) curve (AUC) as performance evaluation indices of the system.

## 3.4 A Traditional machine learning CAD scheme

### 3.4.1 Hand-crafted feature extraction

A traditional CAD system based on hand-crafted features is also implemented in this study. The system consists of four steps: (1) mass segmentation, (2) feature

extraction, (3) feature selection and (4) classifier optimization. For the first step, a region-growing based segmentation algorithm described in section 3.2.1 and [57] is also applied here to segment the suspicious area from the background. Subsequently, six categories of commonly used hand-crafted features for mass classification are calculated from the segmented area, including shape, contrast, size, spiculation, homogeneity and gray-level co-occurrence (GLCM) texture features. The following are brief descriptions of the computed features [57].

1. Mass size: We compute 3 image features in this category. The first one is the mass area, which is computed by automatically counting the total number of pixels inside the segmented mass region, and then multiplying the pixel size (or spatial resolution of mammogram). In addition, since in clinical practice, radiologists use the radial length of the lesion to measure mass size (based on RECIST guidelines [76]), we compute the second feature that is the normalized mean radiant length computed by the mean radial length divide by the total number of pixels inside the mass region [77], and the third feature that is the maximum radial length.

2. Mass shape: Radiologists typically rate mass shape into round, oval, or irregular. To quantify these ratings, we computed two most commonly used shape-related features [77]. The first one is a shape factor ratio defined as $P^2/A$, where $P$ and $A$ are the perimeter and area of a lesion region, respectively. The second one is a radiant length coefficient of variation. Using the radial length ($r_i$) that is the distance between the mass center and pixel ($i$) located at the mass boundary, this feature is defined as the coefficient of variation of $r_i$, which can be computed by standard deviation of $r_i$ divided by mean value of $r_i$.

3. Mass contrast: In order to compute the contrast related features, different types of mass outside surrounding area can be selected, which will have different impact on the computational results [78]. In this study, we use the method reported by te Brake *et al* [79] to define the mass surrounding area. First, a morphological dilation operation with a spherical kernel of size $0.6R$ is performed on the segmented mass region, where $R$ is the mean radial length ($\bar{r}$) of the mass region. Then the pixels inside the dilated region but outside the mass region are labelled as "outside surrounding area." Figure 10 shows an example of the outside surrounding area that was used to compute the contrast features. Next, three contrast related features are defined and computed [79]. The first one is computed by the difference between the average pixel values inside the segmented mass region and its surrounding (outside) area. The second one is computed based on a distance measure between the two pixel intensity histograms, which can be computed as:

$$distance\ based\ contrast = \frac{(E(I)-E(O))^2}{Var(I)+Var(O)} \quad (4)$$

Where $I$ is denoted as the set of pixels in the mass region, $O$ is denoted as the set of pixels in the outside surrounding area. The last contrast related feature is computed based on the average gradient vector magnitude of the boundary pixels, which is related to the sharpness of the boundary.

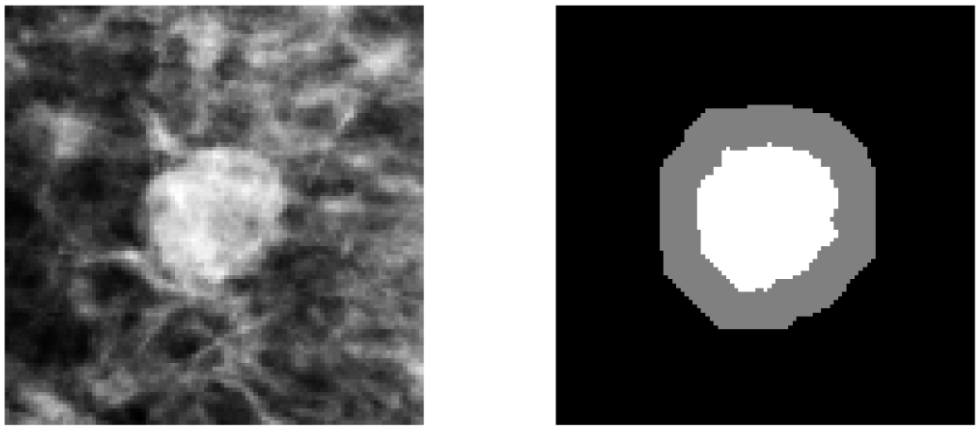


**Figure 10:** An example of a suspicious mass (a) and computed mass outside surrounding area (b), where the gray area is the outside surrounding area and the white area is the segmented mass area.

4. Mass homogeneity: The degree of mass density heterogeneity phenotypes contains biologically important tumor development patterns including the degree of tumor stiffness variation and necrosis. To quantify mass density homogeneity, we compute four features. These are 1) standard deviation of pixel intensities inside the mass region. 2) Kurtosis of pixel intensities inside the mass region. 3) Average local pixel intensity fluctuation in the mass region as defined in our previous study [77], where the local pixel intensity fluctuation of a pixel is defined as the maximum absolute difference between the pixel intensity and the intensity of pixels inside a 5×5 square kernel centered at that pixel. 4) Standard deviation of the local pixel intensity fluctuation inside the segmented mass region.
5. Mass spiculation: The degree of mass boundary spiculation is another primary characteristic indicating mass malignancy. In this study, we use two radial edge-gradient analysis based features [80] to measure the spiculation of a segmented

mass. First, a 3×3 mean filtering is performed on the mammogram as a preprocessing step. A morphological dilation and erosion operation is then applied to the segmented mass region, respectively. The difference between the dilated and eroded image is extracted as the "lesion boundary area." For each pixel inside the lesion boundary area, the maximum gradient at that pixel and the radial direction from the mass center to the boundary pixel are computed, respectively. Next, the "radial angle" is obtained by the angle between the two vectors (i.e. maximum gradient and radial direction). The radial angles of all pixels in the lesion boundary are collected to form a radial angle histogram. Based on the a priori knowledge that if a mass boundary is not spiculated, the radial angle histogram will tend to be compact and accumulate near 0°, we extracte the kurtosis of the distribution as the first spiculation-related feature. Then the number of pixels whose radial angles are between 60° and 120° or -60° and -120° are counted as "spiculated" pixel number. The spiculated pixel number divided by total pixel number inside the mass boundary area is calculated as the second feature. Figure 11 shows two examples of radial angle histograms, where the first one is from a less-spiculated mass and the second one is from a spiculated mass.

6. Texture: GLCM texture features are commonly used for mammographic mass classification. GLCM calculates the frequencies of combinations of pixel intensities co-occurred in the images along different angles. 16 GLCM texture features are extracted, including the contrast, correlation, energy and homogeneity statistics of GLCMs along four different angles (i.e. 0º, 45º, 90º and 135º).

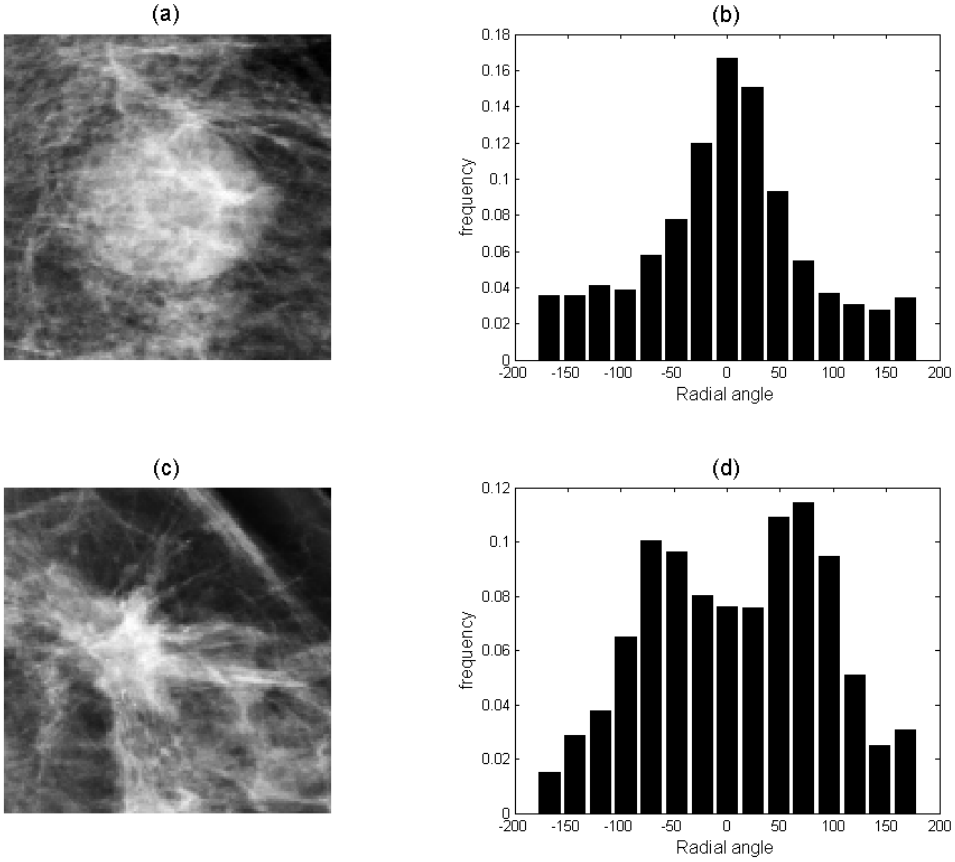


**Figure 11:** Example of radial angle histogram of two mammographic masses where (a) shows a less-spiculated mass with its radial angle histogram (b); while (c) shows a spiculated mass with its radial angle histogram (d).

After feature extraction, a joint mutual information (JMI) [74] based multi-variate feature selection algorithm is applied to eliminate redundant and useless features. The algorithm starts from an empty feature set and iteratively includes new features based on the measurement of conditional mutual information (MI). Similarly to the transfer learning based system, a LOCO process is performed to evaluate the performance of this traditional CAD scheme. In each training/testing iteration, JMI algorithm is performed for feature selection over the 300 training cases. A linear SVM is then optimized using the selected feature subsets and applied to predict the testing

case. Prediction accuracy and AUC value are finally obtained for assessment and comparison.

## 3.5 Results

### 3.5.1 Gray-scale ROIs vs. pseudo-color ROIs

We evaluated and compared the performance of applying the original gray-scale ROIs (i.e. feed the gray-scale ROIs into RGB channels respectively) and pseudo-color ROIs as the inputs of the pre-trained Alex-Net for extracting transferred CNN features. Notably, the dimensionality reduction approaches were not performed to this end. The features extracted from the FC6 layer, FC7 layer of the Alex-Net and their combination were evaluated respectively using linear SVM classifiers and LOCO process. Figure 12 shows the performance assessment of using different ROIs (i.e. gray-scale vs. pseudo-color) and different features (i.e. FC6, FC7 and FC6+FC7). It can be observed that the pseudo-color ROIs based classifiers outperform the gray-scale ROIs based classifiers for both the two assessment indices and the three feature sets. It also shows that the transferred CNN features extracted from the FC6 layer have better prediction performance than the features from FC7 layer. Combining the features from FC6 layer and FC7 layer slightly decreased the performance level as compared to using FC6 layer only. Therefore, we only applied the pseudo-color ROI based CNN features extracted from the FC6 layer for the following analysis and results.

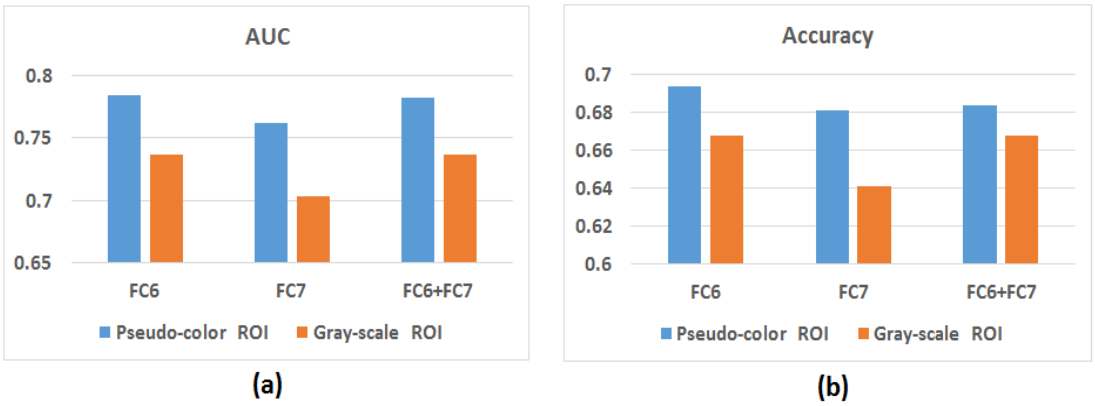


**Figure 12:** (a) AUC values and (b) Predication accuracies of using gray-scale ROIs and pseudo-color ROIs with features extracted from different layers of Alex-Net.

### 3.5.2 Effects of dimensionality reduction

In this section we evaluate whether dimensionality reduction approaches can improve the performance of the classifiers by removing un-related transferred features. Two methods were implemented and compared including a traditional t-test based univariate feature selection method and a novel activation based method. Table 3 summarizes the performance of the classification system without feature selection, using t-test feature selection with different p-value thresholds and using activation-based feature selection respectively. Notably, the features used in this section were extracted from the FC6 layer of the pre-trained Alex-Net that took pseudo-color ROIs as inputs. The results show that the activation-based feature selection method obtained best performance in terms of both AUC and prediction accuracy, while t-test method did not substantially improve the performance of the classifier. Therefore, traditional univariate statistics based methods may not be effective for dimensionality reduction of transfer learning systems.

**Table 3:** Performance of the transfer learning based classification system with different dimensionality reduction methods

| **Dimensionality reduction method** | **AUC** | **Accuracy** |
|---|---|---|
| Original feature set (no feature selection) | 0.784 | 0.694 |
| T-test (p-value threshold = 0.05) | 0.765 | 0.688 |
| T-test (p-value threshold = 0.1) | 0.782 | 0.704 |
| T-test (p-value threshold = 0.2) | 0.761 | 0.681 |
| Activation-based method | 0.792 | 0.711 |

For better visualization, we also plotted the distribution of the transferred features according to the number of non-zero activations over the 301 cases as shown in Figure 13. We observed that near three quarters of the transferred features are activated by less than 100 pseudo-color ROIs and about half of the features are activated by less than 40 cases. It demonstrated that only a small fraction of the transferred features is strongly related to the input ROIs and contributes to the classification between benign and malignant masses.

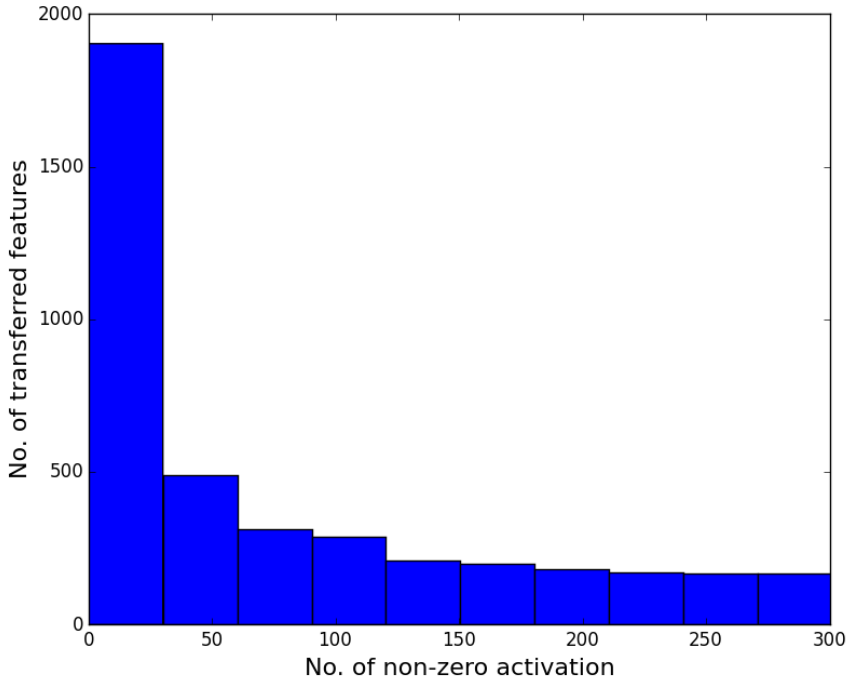


**Figure 13:** Distribution of the transferred features according to the number of non-zero activations over the 301 cases

### 3.5.3 Comparison with hand-crafted features based CAD system

We independently optimized a transfer learning based CAD system and a traditional hand-crafted featurebased CAD system for comparison. An ensemble classifier was also developed by simply averaging the two classification scores generated by the two classifiers. Figure 14 shows the ROC curves of the two classifiers as well as the ensemble classifier. The AUC values and prediction accuracies are shown in Table 4. The results indicate that the transfer learning based classification system can yield better results (i.e. higher AUC) than traditional hand-crafted features based classifier. The performance can be further improved by combining the classification scores generated by the two classifiers. The AUC value of the ensemble classifier is

statistically significantly higher than the AUC value obtained by the traditional classifier ($p$<0.02), while the difference between the ensemble classifier and the transfer learning based scheme is not statistically significant ($p$>0.05).

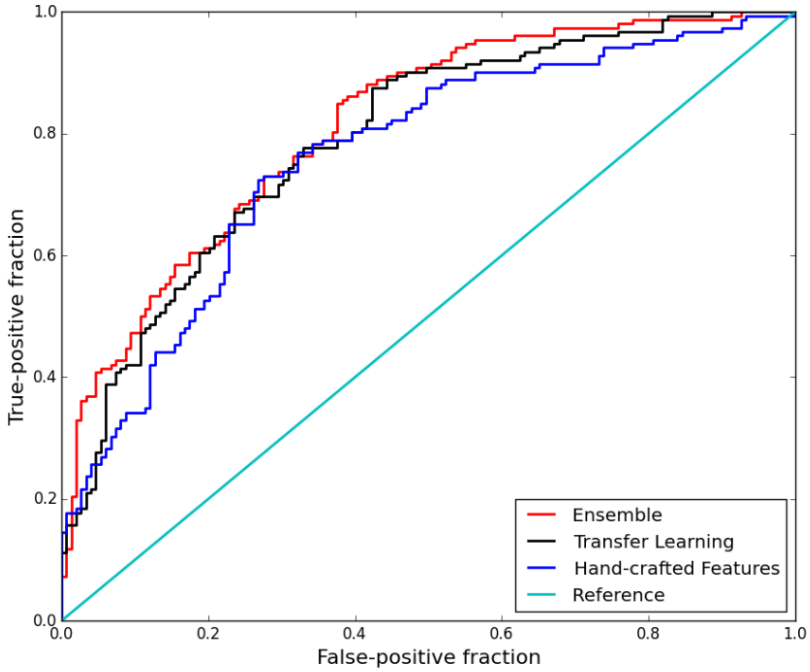


**Figure 14:** ROC curves of the transfer learning based classifier, hand-crafted features based classifier and the ensemble classifier.

**Table 4:** Performance assessment of the transfer learning based classifier, hand-crafted features based classifier and the ensemble classifier.

| | **AUC** | **Accuracy** |
|---|---|---|
| Transfer learning | 0.792 | 0.711 |
| Hand-crafted features | 0.762 | 0.718 |
| Ensemble | 0.813 | 0.718 |

### 3.6 Discussions

Developing CAD systems for mammographic mass classification may play a crucial role assisting radiologists in breast cancer diagnosis by reducing high false-positive recall rates and the unnecessary biopsies. Traditional machine learning classifier based CAD systems are limited by the difficulties of designing and selecting the effective hand-crafted image features to capture the intra-class variations of the mass-like lesions. Although deep learning technology can overcome this challenge by automatically learning feature representations, it cannot be directly applied for mass classification due to the limited size of mammogram dataset. In this study, we investigated the feasibility of developing a deep transfer learning based CAD system for benign and malignant breast mass classification. The new CAD system employed a deep CNN model (i.e. Alex-Net) that is well-trained with a large natural image dataset (i.e. ImageNet). Therefore, we can take advantages of automatic high-level feature extraction of deep networks without considering the limitation of image dataset size. Comparing to previous deep transfer learning based classification schemes, the new CAD scheme reported in this study has several unique characteristics.

First, instead of directly feeding the gray-scale ROIs into the pre-trained Alex-Net, we investigated a new approach to obtain pseudo-color ROIs to fill in RGB channels from the original gray-scale ROIs. Our results demonstrated that the pseudo-color ROIs based classifier can yield significantly higher performance than the gray-scale ROI based classifier. There are two possible reasons: (1) redundant information was provided to the Alex-Net by using gray-scale images and therefore we cannot take full advantages of the pre-trained parameters; (2) we incorporated additional shape and

texture characteristics into the pseudo-color ROIs and these characteristics are helpful for extracting effective image features. We also demonstrated that the features extracted from FC6 layer are more effective than the features from FC7 layer which are high-level but more specifically related to the natural images in ImageNet dataset.

Second, a simple activation-based feature selection approach was applied to eliminate un-related transferred features and the results show that the new method outperforms traditional univariate statistical test based feature selection methods. The activation-based method is motivated by the observation in which a large fraction of hidden neurons in the pre-trained Alex-Net may not necessarily represent the characters of the pseudo-color mammographic ROIs (as shown in Figure 3). We demonstrated that the classifier trained with about 1200 strongly related transferred features outperformed the classifier trained using the whole transferred feature set.

Third, how to combine deep learning technologies with prior knowledge and hand-crafted features is also attracting significant research interests recently. In medical imaging area, two previous studies have shown that the prediction performance was improved by combining the traditional hand-crafted features with the automatic features learned by deep network [70, 81]. In this study, we investigated to incorporate prior knowledge into the transfer learning system in two ways: (1) the subjectively designed morphological or texture variations of the original ROI were used to form the pseudo-color ROIs; (2) an ensemble classifier was developed by averaging the classification score generated by the hand-crafted features based classifier and the transfer learning based classifier. We demonstrate that incorporating additional information into deep learning systems potentially enables to obtain the improved performance.

In summary, we proposed and tested a novel deep transfer learning based CAD system for mammographic mass classification in this study. We investigated to incorporate prior knowledge and hand-crafted features into the CAD system, which enables to obtain significantly higher classification performance (e.g., AUC value) than using the traditional CAD scheme to classify between benign and malignant mammographic masses. Hence, this study provides a reliable framework for developing transfer learning systems based on gray-scale images and/or small image datasets.

Although the experiment result is encouraging, this study still has a number of limitations and can be further improved in the future. First, the dataset used in this study only contains 301 mammogram images. As a result, the pre-trained Alex-Net was fixed as a feature extractor in the CAD system. The performance of the system can be potentially improved by collecting a larger dataset and performing a fine-tune process on the pre-trained network. Second, a relatively simple method (i.e. averaging) was applied to combine the transferred CNN features and hand-crafted features in this study. More sophisticated fusion approaches need to be developed and further evaluated in future studies.

# Chapter 4: Automated prostate segmentation in MR images using 3D Fully Convolutional Network with a coarse-to-fine residual module

## 4.1 Introduction

Prostate cancer is the most common cancer occurred and second common cause of cancer mortality in men population in United States [48]. It was estimated that 180,890 cases were diagnosed with prostate cancer in 2016, which accounts for 21% of new cancer diagnoses [48]. Accurate segmentation of prostate area from Magnetic Resonance (MR) images is an important pre-processing step for diagnosis and treatment planning of prostate cancer as well as other prostate diseases [82]. Manual segmentation of prostate area from MR slices is time-consuming and suffers from large intra- and inter-reader variability. As a result, developing automatic or semi-automatic prostate segmentation schemes has potentially clinical utility and thus has gained significant research interests in the recent decade. However, it is difficult to accurately segment the prostate area from MR images due to the large variation of prostate size and shape between patients, as well as the different scanning protocols and unclear prostate boundaries [83]. Traditional methods for automatic MR prostate segmentation include multi-atlas based [84, 85] and deformable model based [86] approaches. These methods are limited by the challenges of how to design effective features to identify correspondence between images or discriminate prostate areas and background tissues [87], and therefore cannot achieve the performance level of human segmentation.

Recently, deep convolutional neural network (CNN) models have gained great research efforts in machine learning and computer vision society. CNN models contain

many hidden layers that can automatically extract high level representations from the raw input images [6]. As a result, the challenges of designing effective hand-crafted features can be avoided by employing deep learning methods (e.g. CNN). State-of-art performances have been obtained and reported by using deep CNN models in many applications such as natural image classification [9], object detection [11] and segmentation [88]. Following that, deep learning technique has been demonstrated to outperform traditional machine learning algorithms for solving various medical imaging problems [1, 89, 90]. As a result, deep learning is also expected as a promising and powerful tool for developing an automatic prostate segmentation system.

In this study, our motivation is to investigate new approaches based on Fully convolutional network (FCN) architecture [91] to improve the performance of existing MR prostate segmentation schemes. Specifically, we modify the basic 3D U-Net structure [92] by adding a coarse-to-fine residual module together with a deep supervision training strategy to improve the segmentation performance as well as the training efficiency. We then apply a densely connected convolutional module [93] to refine the initial segmentation results in an auto-context manner. The performance of the proposed method is evaluated using a publically available prostate MR image dataset, which aims to demonstrate accurate segmentation and fast convergence.

## 4.2 Related works

### 4.2.1 CNN for medical image segmentation

Deep learning has been demonstrated its feasibility to overcome difficulties of the traditional segmentation systems in a number of applications such as brain segmentation [58, 94] and pancreas segmentation [95, 96] etc. In these frameworks, the

segmentation problem is modeled as a pixel/voxel-wise classification task. Early works applied a sliding window to extract small patches near each pixel/voxel and applied CNN to predict the label of the targeted pixel/voxel. Such approach has the drawback of low computational efficiency due to the repeating computation of the overlaps between image patches [89]. Alternatively, Fully Convolutional Network (FCN) provides an end-to-end neural network architecture that can generate dense predictions with high efficiency by significantly reducing the repeating computations [91]. Based on FCN, U-net [92] was proposed to improve the segmentation performance by adding skip-connections between up and down paths, which has been extensively applied in medical image segmentation studies.

### 4.2.2 Prostate segmentation

Multi-atlas based segmentation methods have been commonly adopted for prostate segmentation from the volumetric MR images [85, 97]. These methods are based on an image set with pre-labelled masks of prostate area. Non-rigid registration methods are employed to register the template images with respect to targeted images, and the corresponding registered prostate masks are fused together to generate the target segmentations. Deformable model based approach is another type of popular prostate segmentation method. For example, Toth et al. proposed an Active Appearance Model (AAM) based segmentation method that utilized the level-set implementation [86]. There are a number of recent studies that focused on applying deep learning for prostate segmentation from MR images. For example, Milletari et al. proposed a V-Net structure with a novel dice coefficient objective function [98]; Yu et al. added long and short mixed residual connections to FCN structure [99]; Cheng et al. developed a holistically

nested network based prostate segmentation system [100]. 3D convolutional networks were applied in these studies since they can process volumetric MR images and take full use of 3D spatial information.

## 4.3 Materials and Methods

### 4.3.1 Dataset and pre-processing

The MR image data acquired and downloaded from the existing NCI-ISBI 2013 Challenge – Automated Segmentation of Prostate Structure (https://wiki.cancerimagingarchive.net/display/Public/NCI-ISBI+2013+Challenge+-+Automated+Segmentation+of+Prostate+Structures) are used in this study to train and evaluate the proposed deep neural network [101]. The training set contains prostate MR images acquired from 60 patients, which were obtained with different equipment and scanning protocols (e.g. 1.5T vs. 3T and endo-rectal receiver coil vs. surface coil). Another 10 cases in the "Leaderboard" set were used to evaluate the performance of the model. For each case, T2-weighted (T2W) MRI axial pulse sequences were obtained with different size and spacing. The corresponding prostate masks marked by radiologists are also provided. It should be noted that in this study, we aimed to segment the whole prostate area from the background, which was different from the original motivation of NCI-ISBI 2013 challenge.

It is necessary to pre-process the image data due to the variations of the gray value range, image size and spacing. Linear interpolation is applied to resize the images and masks to obtain a fixed spatial resolution of 1×1×3 millimeters. A 3D patch with size of 128×128×24 is then cropped for each patient. A z-score normalization is finally performed to standardize the voxel gray values with zero mean and unit standard

deviation. The normalized 3D patches are then used as the input of the proposed neural network.

### 4.3.2 The proposed network

We address the prostate segmentation problem using a two-stage FCN. The first stage is a 3D U-Net architecture which takes the pre-processed 3D MR images as input and generates a binary mask with the same size for segmentation. Deep supervision and residual module strategies are investigated to accelerate the training process and improve the segmentation accuracy. The second stage is a densely connected convolutional module which takes the MR images together with the segmentation results generated by the first stage as the input. The motivation of adding the second stage is to refine the initial segmentation results by incorporating auto-context features. Details of the U-Net structure, residual module with deep supervision, and refinement network are described in the following sections.

### 4.3.3 3D U-Net

U-Net was proposed by Ronneberger et al. for various biomedical image segmentation tasks [92]. The original U-Net is a 2D convolution operation based FCN. Recent studies have demonstrated the effectiveness of 3D FCN for accurate segmentation of volumetric medical images by fully utilizing the 3D spatial information [94, 98, 99, 102]. Therefore, a U-Net with 3D convolution units (i.e. 3D U-Net) is employed in this study as the basic network structure. Figure 15 shows the architecture of the 3D U-Net. The network takes the MR image patches with size 128×128×24 as input. The left side of the network is a contracting path that gradually reduces the resolution of the feature maps. At each resolution level, two 3D convolutional layers

with rectified-linear unit (ReLU) non-linearity are stacked, followed by a max-pooling layer with 2×2 sliding window to down-sample the feature maps. Small kernel size (i.e. 3×3×3 voxels) is adopted in the convolutional layers and the number of filters is doubled as the resolution halving. Batch Normalization mechanism is introduced between convolutional operations and ReLU, in order to accelerate the training process.

The motivation of using contracting path is to increase the receipt field of each voxel in the feature maps and thus incorporate more spatial information. An expansive path in the right side of the network is applied to generate high-resolution feature maps for voxel-level prediction. De-convolution operations with 2×2×2 trainable kernels are employed to increase the size of input feature maps by a factor of 2. The output of de-convolution layers is concatenated with the corresponding feature maps generated in the contracting path with same resolution (shown as the horizontal connection in Figure 15).

The connection from the contracting path aims to provide complementary high-resolution information, since the de-convolution layers only take coarse features from the low-resolution layers as input. The concatenation is then processed by stacking a number of convolutional layers, ReLU and Batch Normalization to generate high-resolution image feature maps. A convolution operation with kernel size of 1×1×1 is performed on the feature maps with highest resolution to generate a single feature map, which is fed into a sigmoid activation function to generate voxel-wise binary classification probabilities. Due to the limitation of the size of the training image dataset, we apply a relatively small U-Net architecture. The network consists of 15 convolutional layers, 3 max-pooling layers and 3 de-convolutional layers. The first

convolutional layer has the same resolution with the input images containing 16 convolutional kernels.

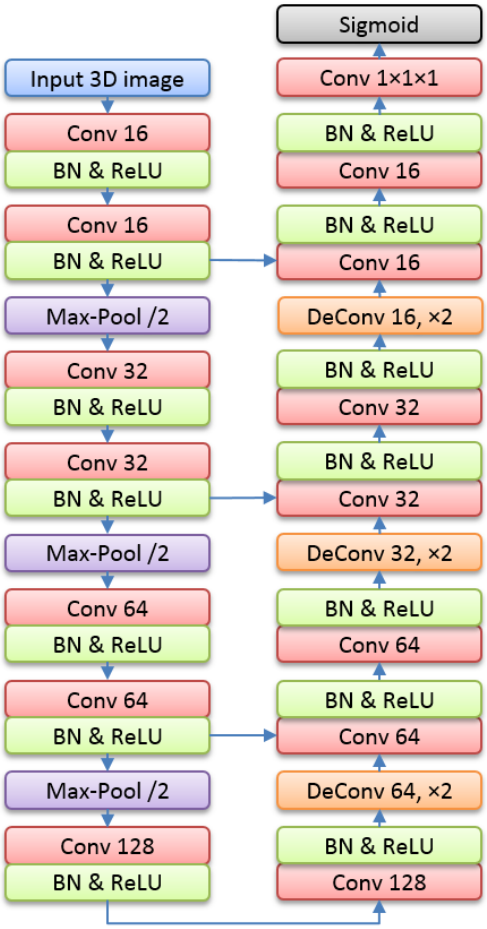


**Figure 15**: U-Net architecture

### 4.3.4 Deep supervision and residual module

Deep neural networks are always difficult to train with the gradient descent based algorithms because of the vanishing gradient problem [103, 104], which makes the convergence rate slow in early neural layers. This challenge can be addressed by adding auxiliary classifiers in the intermediate layers to increase the back-propagated gradient signals in early layers [16]. Based on this strategy, Dou et al. integrated a deep

supervision mechanism in a FCN for medical image segmentation [102]. Specifically, auxiliary convolutional layers are injected to the hidden layers of the FCN at different resolution level to generate auxiliary predictions; the auxiliary predictions are then interpolated to the size of input images for computing losses. It has demonstrated that the deep supervision mechanism can accelerate the training process and improve the segmentation accuracy simultaneously in various previous studies [94, 102]. However, the auxiliary coarse predictions at low resolution level are not effectively propagated to the final segmentation in these approaches, since the auxiliary convolutional layers and corresponding predictions are simply abandoned in the testing phase and the convolutional layers in the main path need to re-formulate the information encoded in the auxiliary layers.

Inspired by the studies of deep residual learning [9] and image super-resolution [105], we propose a new 3D U-Net based FCN architecture with coarse-to-fine residual module and deep supervision, as shown in the left part of Figure 16 (i.e. stage-1). The left side of the stage-1 network is a basic 3D U-Net. Auxiliary convolutional layers with kernel size of 1×1×1 are connected to the hidden feature maps with different resolution in the expansive path of the U-Net, in order to generate single feature maps which are then up-sampled and fed into a sigmoid function to obtain auxiliary predictions. The single feature maps are obtained by the summations of the up-sampled single feature maps from lower resolution levels and residual feature maps generated by the current resolution levels. In this way, the coarse auxiliary predictions can be effectively propagated to higher resolution levels and the convolutional layers in the main path are applied to extract the residual features which contain rich fine details of the images.

Therefore, our proposed coarse-to-fine architecture takes the advantages of both residual learning and deep supervision, in attempt to improve the segmentation performance and training efficiency.

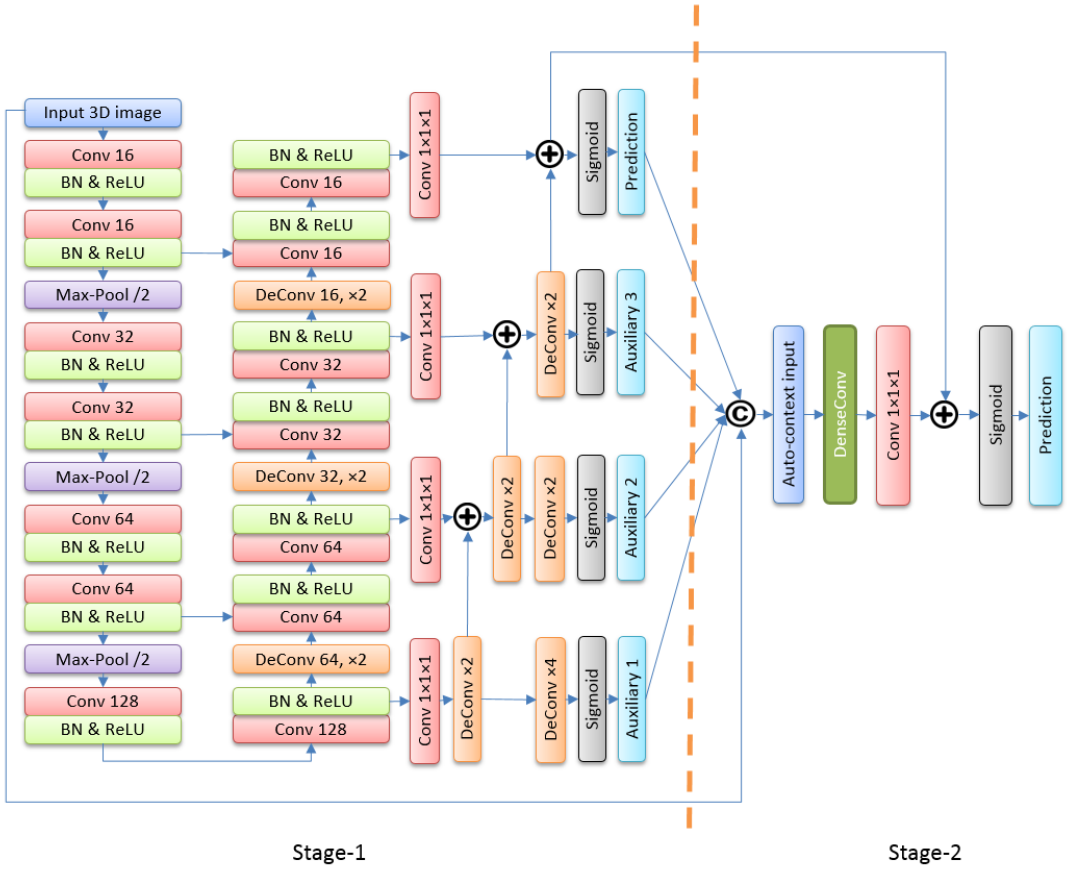


**Figure 16:** Proposed FCN architecture. Stage-1 is a 3D U-Net with deep supervision and coarse-to-fine residual module; stage-2 is a densely connected convolutional module for refinement.

### 4.3.5 Auto-context refinement

The stage-2 network shown in Figure 16 aims to refine the segmentation generated by the stage-1 network using an auto-context [106] strategy, which integrates low-level voxel values and high-level context information to generate new feature maps [94]. A residual learning strategy similar to the previous work of Xu et al. [107] is

adopted in this study. Specifically, the original input MR images are concatenated with the segmentation results of the stage-1 network at different resolution levels, forming 4D images with 5 channels. The 4D images are then fed into a densely connected convolutional (DenseConv) module to extract auto-context features. The densely connected architecture has the advantages of feature re-using and propagation strengthening [93]. Figure 17 shows the architecture of DenseConv module, which consists of 6 convolutional layers. The single feature map generated in main prediction path of the stage-1 network is summated with the feature map obtained by the DenseConv module, and then fed into a sigmoid function to generate final segmentation. In this way, the DenseConv module learns a residual function that highlights the refinement of the initial segmentation results.

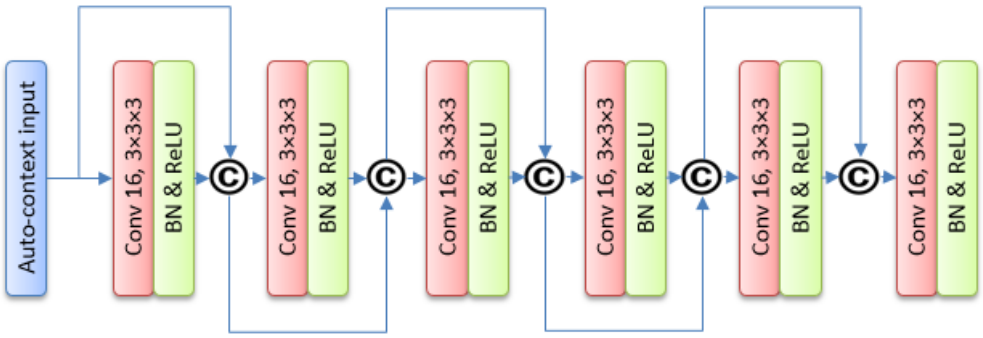


**Figure 17:** Architecture of densely connected convolutional (DenseConv) module for refinement

### 4.3.6 Implementation

Due to the size limitation of the prostate MR image dataset, an on-the-fly data augmentation strategy is applied to enlarge the training set and avoid the potential over-fitting problems. The original images are randomly shifted (± 4 mm), rotated (± 15 º) and scaled (a factor of 0.9 – 1.1) to generate augmented images as the input of the network for every two training epochs. All weights used in the network are randomly initialized with a normal distribution. We adopt dice coefficient (DC) loss [98] as the objective function to train the network. The DC loss between the ground truth segmentation and predicted segmentation is defined as:

$$DC\ loss = \frac{2\sum_i^N g_i p_i}{\sum_i^N g_i + \sum_i^N p_i} \quad (5)$$

Where N is the total number of voxels, $g_i$ is the ground truth label of a particular voxel and $p_i$ is the predicted probability of the corresponding voxel. The advantage of DC loss over cross-entropy or weighted cross-entropy loss is that it can address the problem of imbalance between the number of background and foreground voxels without setting any hyper-parameters [98]. Adam algorithm [108] is implemented to optimize the parameters and minimize the objective function. The neural network models are implemented in Python with TensorFlow library [109].

We adopt a 3-step process to train the two-stage FCN, which is similar to the previous work by Xu et al. [107]. First, the stage-1 network is trained with the objective function set as the weighted summation of DC losses of the main prediction and the auxiliary predictions. The weights are 0.3, 0.3, 0.6 and 1 for the predictions from coarse to fine. Second, we fix the parameters of the stage-1 network and optimize the

parameters of the stage-2 network with respect to the DC loss until convergence. Last, we fine-tune the entire network using a smaller learning rate. In the testing phase, 10 cases of unseen prostate MR images were fed into the two-stage network in a feed-forward manner. The outputs of the stage-2 network are binarized with a threshold of 0.5 to obtain the predicted masks of prostate area.

### 4.4 Results

We first compared the convergence rate and segmentation performance of the baseline 3D U-Net and our proposed stage-1 network, which is a U-Net with deep supervision and coarse-to-fine residual module. Figure 18 shows the curve of testing losses of these two networks at different training epochs. It demonstrates that our proposed network structure achieves faster convergence and lower DC loss in the testing dataset than the original 3D U-Net. Comparing to the baseline 3D U-Net that may require approximately 300 epochs to yield a relatively stable and converged result, the convergence was achieved after approximately 100 epochs using the new module.

Table 5 summarizes and compares the dice coefficients of different network components including the baseline U-Net, stage-1 network and the entire two-stage network with DenseConv module for refinement. Specifically, by integrating the deep supervision and coarse-to-fine residual mechanism, the average DC of the 3D U-Net was improved from 0.892 to 0.902. The stage-2 network which employs a DenseConv module further improved the performance and yielded an average DC of 0.909 with a standard deviation of 0.011. Figure 19 shows three qualitative examples of the prostate MR segmentation results generated by the proposed new methods.

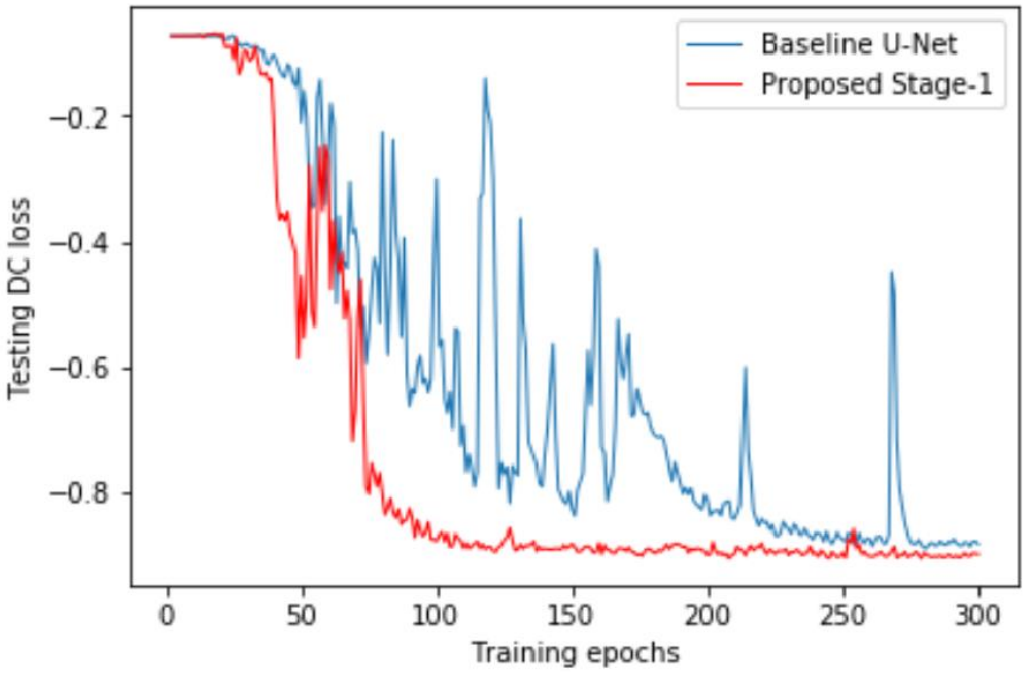


**Figure 18:** Testing loss of 3D U-Net and our proposed Stage-1 network with deep supervision and residual module

**Table 5:** Dice coefficients of different network components

| Method | DC mean |
|---|---|
| 3D U-Net | 0.892 ± 0.018 |
| Stage-1 network | 0.902 ± 0.017 |
| Entire network | **0.909** ± **0.011** |

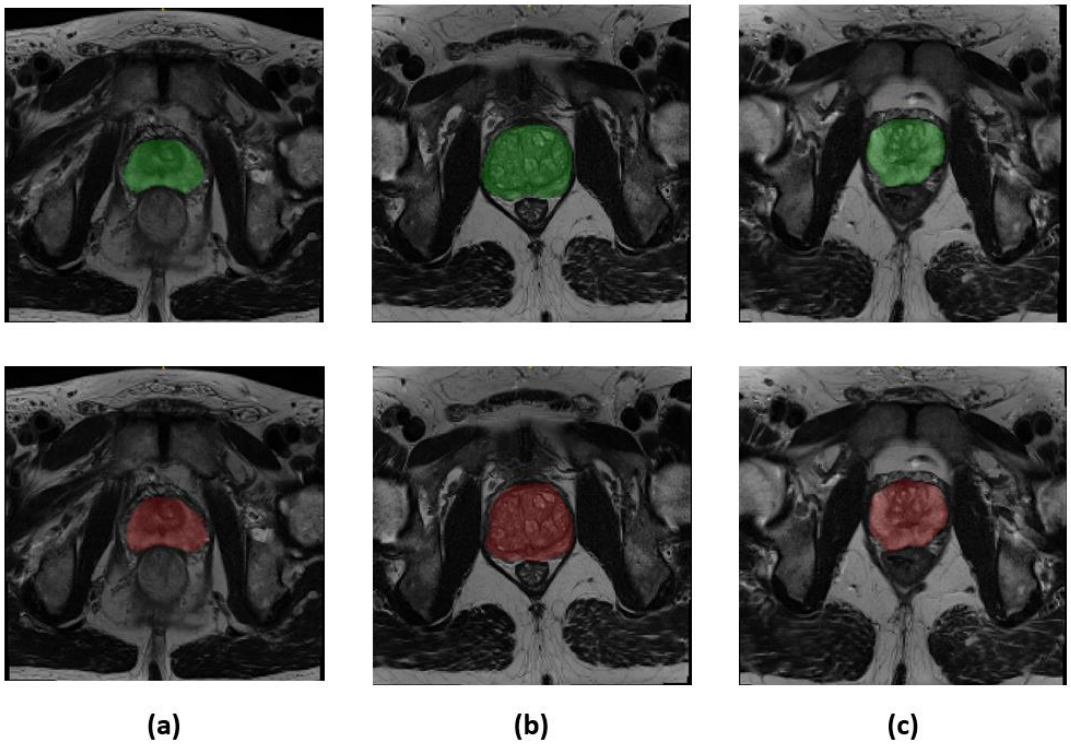


**Figure 19:** Three examples showing the segmentation of prostate area obtained by radiologists (top, green) and by the proposed two-stage network (bottom, red)

We also performed a comprehensive quantitative comparison between our proposed approach and existing state-of-art prostate segmentation systems using deep neural networks. Four metrics were used to evaluate and compare the performance of the systems, which include the Dice coefficient (DC), the Intersection over Union (IoU), the percentage of the absolute difference between volumes (aRVD) and the 95% Hausdorff distance (95HD). Three state-of-art deep neural network architectures were implemented and evaluated with the same prostate MR image dataset used in this study. The networks include V-Net [98], Holistically-Nested Networks (HNN) [100] and Volumetric ConvNet [99]. Table 6 shows the segmentation performance of different networks and it demonstrates that the new method tested in this study outperforms the three existing methods for all the four evaluation metrics by yielding not only higher

accuracy values, and also smaller standard deviations, which indicates that the results generated by the new method are more robust. Furthermore, we performed paired t-tests to statistically analyze the differences of dice coefficients between our proposed methods with the baseline U-Net, V-Net, HNN and Volumetric ConvNet. The results also demonstrate that our method yields significant improvement compared to baseline U-Net and previous state-of-art methods with $p$-value < 0.03.

**Table 6:** Segmentation performance of different network structures

| Method | DC | IoU | aRVD | 95HD (mm) |
|---|---|---|---|---|
| Our method | **0.909 ± 0.011** | **0.833 ± 0.018** | **0.061 ± 0.033** | **5.065 ± 0.998** |
| V-Net | 0.866 ± 0.019 | 0.764 ± 0.030 | 0.114 ± 0.075 | 6.477 ± 1.154 |
| HNN | 0.865 ± 0.044 | 0.765 ± 0.065 | 0.236 ± 0.136 | 9.425 ± 4.755 |
| Volumetric ConvNet | 0.885 ± 0.017 | 0.794 ± 0.027 | 0.086 ± 0.048 | 7.781 ± 5.305 |

Note: For DC and IoU, higher values are better; for 95HD and aRVD, lower values are better.

### 4.5 Discussions

Developing automated schemes for prostate segmentation from MR images plays an important role in computer-aided diagnosis of prostate cancer and other prostate diseases. Traditional segmentation methods are greatly limited by the challenges of how to design effective hand-crafted features. Recently, with the availability of big data and fast parallel computation resources, deep learning techniques have been applied for addressing various challenges in computer vision and medical image analysis field. In this study, we proposed a novel neural network architecture for the automated prostate MR image segmentation. The proposed model consists of two stages of FCNs, where the first stage applied a 3D U-Net based network structure to obtain an initial segmentation and the second stage employed a densely connected

convolutional module to refine the details of the segmentation results. The proposed method demonstrates a number of unique characteristics. First, we propose a novel coarse-to-fine residual module and integrate it with the basic 3D U-Net structure. The residual module enables effective propagation from coarse segmentation to fine segmentation and also avoids feature re-formulating in the expansive path of U-Net. Auxiliary convolutional layers and predictions are applied to address the vanishing gradient problems and provide additional regularization as well. As a result, we demonstrate by experiments that our proposed structure can greatly accelerate the training efficiency of U-Net and yield better segmentation accuracy.

Second, we adopt an auto-context strategy by developing a stage-2 network to combine features from input images and initial segmentations at different resolution level. The stage-2 network aims to integrate low level features with high level features and refine the segmentation of stage-1 network by learning a residual/refinement function. We apply a state-of-art densely connected convolutional module, which yields further improvement of the segmentation performance.

Last, we compare the proposed method with previous state-of-art 2D (i.e. HNN) and 3D (i.e. V-Net and Volumetric ConvNet) deep neural network architectures for prostate segmentation. We highlight the main limitations of the previous methods compared to our new method. (1) V-Net does not adopt batch normalization strategy, which limits the prediction power of the network; (2) HNN is a 2D convolutional operation based network and cannot integrate 3D information; (3) Volumetric ConvNet adopts standard cross-entropy as loss function and might be affected by data-unbalance of prostate area and background area. Our method enables to overcome these limitations

and integrate novel modules to improve the performance. As a result, our experimental results demonstrate that our method outperforms previous models by a large margin for all the four evaluation metrics with higher segmentation accuracies and smaller standard deviations.

While applying deep learning for computer vision problems usually requires large amount of training data to avoid over-fitting, the limitation of MR image set (i.e. 60 cases for training) is not a big issue in our study since (1) the objective function is obtained over each voxel of a 3D image for medical image segmentation problems and each volumetric MR image scan contain large number of voxels and rich appearances of background; (2) we perform extensive on-the-fly data augmentation strategy to increase the diversity of training data and avoid over-fitting.

In summary, we presented and tested a novel deep neural network architecture for accurate prostate segmentation from MR images in this study. We demonstrate state-of-art segmentation performance using the proposed network. Our method can be potentially applied as a pre-processing step for development of new computer-aided detection/diagnosis systems of MR images of prostate cancer and other prostate diseases. The proposed new deep learning framework is not only limited to segment prostate regions from MR image, it can also be easily adopted and applied to fulfill other medical image or natural image segmentation tasks.

# Chapter 5. Applying a fully convolutional neural network for prostate segmentation and cancer detection using multi-parametric magnetic resonance images: an initial investigation[1]

## 5.1 Introduction

Since prostate cancer is one of the leading causes of cancer mortality in men in the United States [48], early detection and diagnosis of prostate cancer is important for reducing mortality rate and improving treatment efficacy. In current clinical practice, prostate specific antigen (PSA) and transrectal ultrasound (TRUS) biopsy are commonly used for prostate cancer detection and diagnosis. Such methods are invasive and yield relatively high false-positive rates. Alternatively, magnetic resonance imaging (MRI) provides a noninvasive imaging tool that enables more accurate detection and diagnosis of prostate cancer [110]. Especially, multi-parametric MRI (mpMRI) with various MRI modalities including T2-weighted (T2W), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced (DCE) etc. has demonstrated to be more effective for prostate cancer detection than using single MRI modalities [111-113]. However, reading and interpreting MR images for prostate cancer detection by radiologists is often difficult because it requires substantial expertise of the radiologists and the reading process is also tedious and time consuming due to the large amount of 3D MR slices. Therefore, developing computer-aided detection (CAD) schemes for automated detection of prostate cancer from the MR images has attracted great research interests in the recent twenty years [114]. Such CAD schemes usually consist of two

[1] Works of this chapter were done during an internship at 12 Sigma Technologies.

steps, which first segment the prostate area from the background tissues and then detect lesions within the prostates.

Traditional machine learning based CAD systems for prostate cancer detection usually adopt a two-stage framework, which consists of a candidate generation stage followed by a false-positive reduction stage. For example, Litjens et al. extracted voxel based features for candidate generation and then calculated region based statistical features for candidate classification [115]. Linear discriminant classifier and ensemble learning models were adopted for both two stages. A number of recent studies also investigated to apply deep learning techniques to address the detection task. Tsehay et al. applied a holistic nested network with deep supervision strategy for pixel-level prediction [116]. Zhu et al. employed stacked auto-encoders to extract high level representations from image patches and trained a random forest for patch classification [117]. Kohl et al. proposed a network architecture with FCN and adversarial network to improve detection performance [118]. The multi-parametric MRI (mpMRI) with different modalities was used in these studies.

In this study, we investigate the feasibility of developing a new prostate cancer detection scheme implemented with a deep neural network model. The proposed scheme consists of two consecutive networks for prostate segmentation and tumor detection, respectively. Fully convolutional network (FCN) architecture is adopted for both steps since it can be trained end-to-end and provide efficient pixel/voxel level prediction. The scheme extracts information from mpMRI for both prostate region segmentation and lesion detection, since a number of previous studies have demonstrated the effectiveness of mpMRI based CAD schemes [115-117].

The motivation of conducting this study is because (1) to our best knowledge, this is the first study that applies mpMRI for prostate area segmentation, which is different from the previous studies only focusing on T2W based segmentation; (2) we are able to introduce a state-of-art deep learning model to address the challenges of prostate cancer detection. Thus, we conduct following experiments to evaluate performance of prostate region segmentation and tumor detection using our proposed method with a prostate MR dataset.

## 5.2 MpMRI based Prostate segmentation

### 5.2.1 Dataset and pre-processing

A dataset containing volumetric prostate MR images acquired from 195 patients was used for development of prostate segmentation step. The dataset was randomly separated into a training set with 160 cases and a testing set with 35 cases. For each case, multi-parametric modalities including T2W, T1 and DWI with highest b-value were obtained. The masks of prostate areas were marked in T2W images by an experienced radiologist as ground truth. Standard affine registration was performed to register T1 and DWI images to T2W images for correction of patient movement and magnetic distortion. The volumes of interest together with the ground truth masks were resized to a fixed spacing of 1×1×3 millimeters. A 3D patch of 128×128×24 was cropped for each MR imaging modality and the voxel values were normalized to zero mean and unit variation. The pre-processed patches from different modalities were concatenated together to form a 4D matrix with 3 channel, which is the input of our proposed FCN.

### 5.2.2 Network architecture

We adopt a 3D FCN architecture which is similar to V-Net developed by Milletari et al. [98]. Figure 20 shows the architecture of the network. The input of the network is a multi-channel MR volume data with four dimensions representing the height, width, depth and channel. The network consists of a contracting path in the left and an expansive path in the right. In the contracting path, convolutional layers with 3×3×3 convolution kernels and restricted linear unit (ReLU) activation function are stacked to extract high-level features. Convolution layers with strides greater than 1 are applied to gradually reduce the resolution of feature maps and increase the receptive field to incorporate more spatial information. At each resolution level, the input signals are directly added to the output feature maps, which enables the stacked convolutional layers to learn a residual function [9]. The contracting path generates feature maps with low resolution and the expansive path is applied to up-sample the feature maps to generate dense prediction. De-convolution operations are employed to increase the resolution of feature maps. The feature maps generated in the contracting path are concatenated with the output of de-convolutional layers to incorporate high-resolution information. Residual connections are also adopted in the expansive path. The final feature maps which have the same size with the input MR volumes are processed by a convolutional layer with 1×1×1 kernel and sigmoid activation function to generate binary predictions. Dice loss [98] is adopted as the objective function at the training phase. Post-processing steps are then applied to refine the initial segmentation generated by FCN. Specifically, we use a 3D Gaussian filter to smooth the predicted probability maps and a connected component analysis to remove small isolated components.

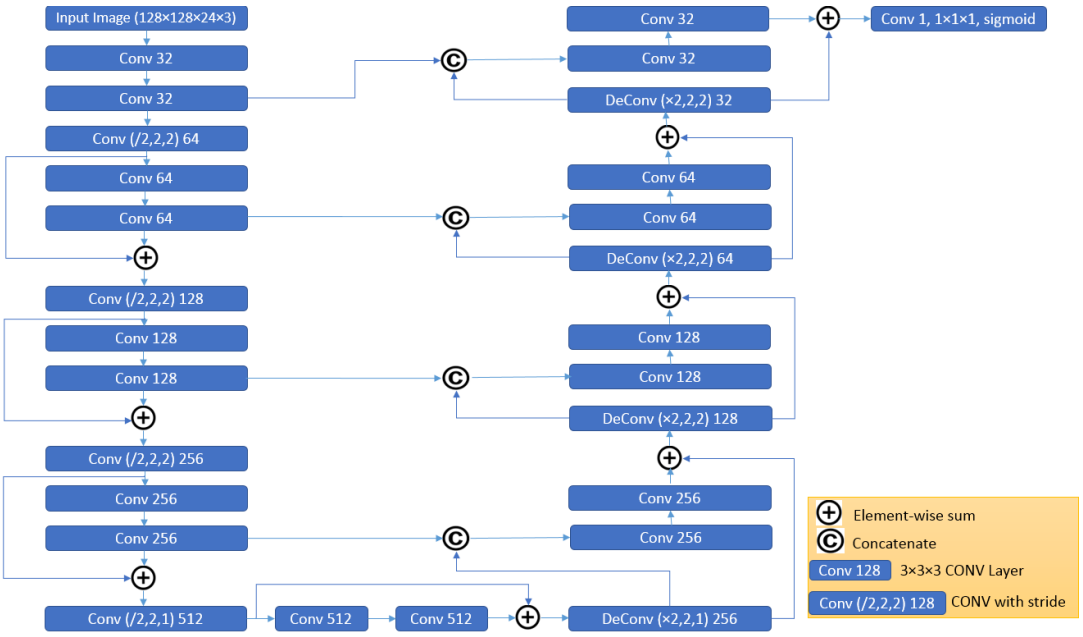


**Figure 20:** The FCN architecture applied for mpMRI based prostate segmentation

### 5.2.3 Implementation and evaluation

We implemented the network with Tensorflow library [109]. Since the dataset is relatively limited, we adopted an on-the-fly data augmentation process to randomly shift, rotate and scale the original volumes during the training phase. The weights were randomly initialized and Adam optimizer [108] was used to train the network. When testing, the optimized network took the mpMRI volume data in the testing set as input and performed feed-forward calculation. The output probability maps were then refined by the post-processing steps and binarized with a threshold of 0.5. We calculated the dice coefficients between the ground truth and predicted segmentation for performance evaluation.

## 5.3 Tumor detection

### 5.3.1 Dataset and pre-processing

The dataset used for development of tumor detection step contains volumetric prostate MR images acquired with multi-modalities including T2W, DWI, apparent diffusion coefficient (ADC) and K-trans from 79 patients. ADC maps were calculated from DWI and K-trans images were obtained using dynamic contrast enhanced (DCE) MR perfusion. Similar to the previous step we applied affine registration to register different modalities to T2W volume data. For each patient, at least one tumor was identified by radiologists and the boundary of the tumor area was also marked.

The dataset was randomly divided into a training set with 60 cases and a testing set with 19 cases. Different from the segmentation step, we adopted 2D network for tumor detection since the sizes of marked tumors are relatively small with respect to the resolution in vertical direction. 400 slices from the 60 cases in the training set were selected for training the network. Among them, 260 slices contain tumors and the other 140 slices are negative. For each slice, a patch containing prostate area was cropped and interpolated to a fixed size of 176×176. Z-score normalization was applied to normalize the pixel values. We enlarged the training set by 30 times using random shift, rotation and scaling.

### 5.3.2 Network architecture

Figure 21 shows the architecture of FCN used for tumor detection. It is similar to the previous network for prostate segmentation, since both networks aim to generate voxel-level predictions. Due to the limitation of annotated data, we apply a relatively

small network with 3 down-sampling layers in the contracting path. The network takes 2D images with 4 channels (i.e. T2W, DWI, ADC and K-trans) as input and outputs the predicted tumor masks. Weighted cross-entropy loss is employed as the objective function to train the network. Specifically, voxel-level cross entropy loss was first calculated for each single voxel. The losses of the positive voxels were multiplied by a hyper-parameter *w* and then summed with the losses of negative voxels to form the objective function. By tuning *w*, we can adjust the weights between losses of positive and negative predictions. We also apply L2 regularization to avoid overfitting.

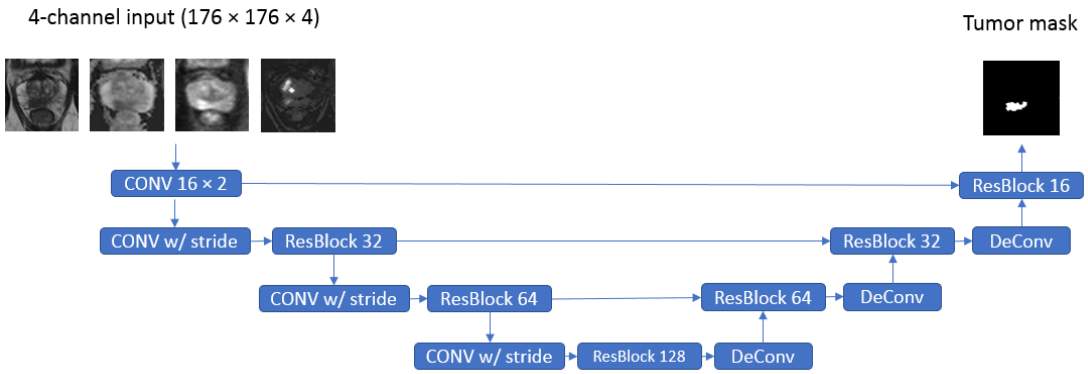


**Figure 21:** The FCN architecture applied for mpMRI based tumor detection

### 5.3.3 Implementation and evaluation

We applied a cascaded two-stage process to train the network. First, we used a large *w* value to train the network for a number of epochs. The motivation is to obtain initial detection with relatively high sensitivity. For the second stage, we calculated the objective function over the candidates (i.e. positive voxels generated by the initial detection) and used a smaller *w* value to train the network, aiming to further reduce the number of false-positives. In the testing phase, we fed 2D mpMRI slices into the

network and stacked 2D predictions to obtain 3D detection. We plot a performance curve (i.e. detection rate versus false positive rate) operated at different thresholds to evaluate the proposed detection method. Specifically, we calculated the detection rate according to the overlap between the ground truth masks and the predicted masks. If over 10% of voxels in a tumor are labelled as positive, it is considered as a success detection. The false positive rate is simply calculated by dividing the number of false positive voxels by the number of all negative voxels inside the prostate.

## 5.4 Results

### 5.4.1 Segmentation

We trained the mpMRI based FCN for 500 epochs for prostate segmentation. We also built another FCN with same architecture which only takes T2W image data as input. Table 7 summarizes the performance of mpMRI and T2W based FCNs. It demonstrates that the segmentation performance can be improved by combining information from multiple MR modalities as compared to using single T2W image data only. Meanwhile, mpMRI based FCN also achieves faster convergence than T2W based FCN.

**Table 7:** Segmentation performance of T2W and mpMRI based FCN

| FCN input | Training epochs | Average Dice coefficient |
|---|---|---|
| T2W | 2000 | 0.8818 |
| T2W, DWI, T1 | 500 | 0.8935 |

### 5.4.2 Tumor detection

Figure 22 shows two performance curves, where the red one is obtained by training the network with the first stage and the blue one is obtained by the two-stage training strategy. It shows that the detection performance can be improved by the cascaded training process with candidate generation stage and false-positive reduction stage, which yield 100% detection rate with false-positive rate smaller than 0.2. Figure 23 shows two examples of detection results that are generated at the level of detection rate of 0.85.

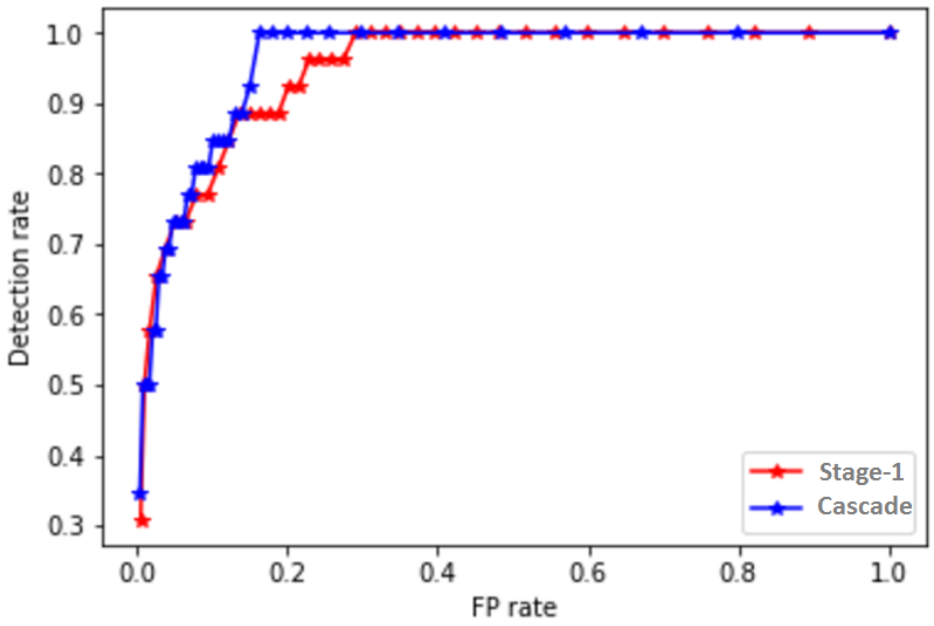


**Figure 22:** Performance curves of the proposed FCN. Red line represents the single-stage training and blue line represents the cascaded training strategy

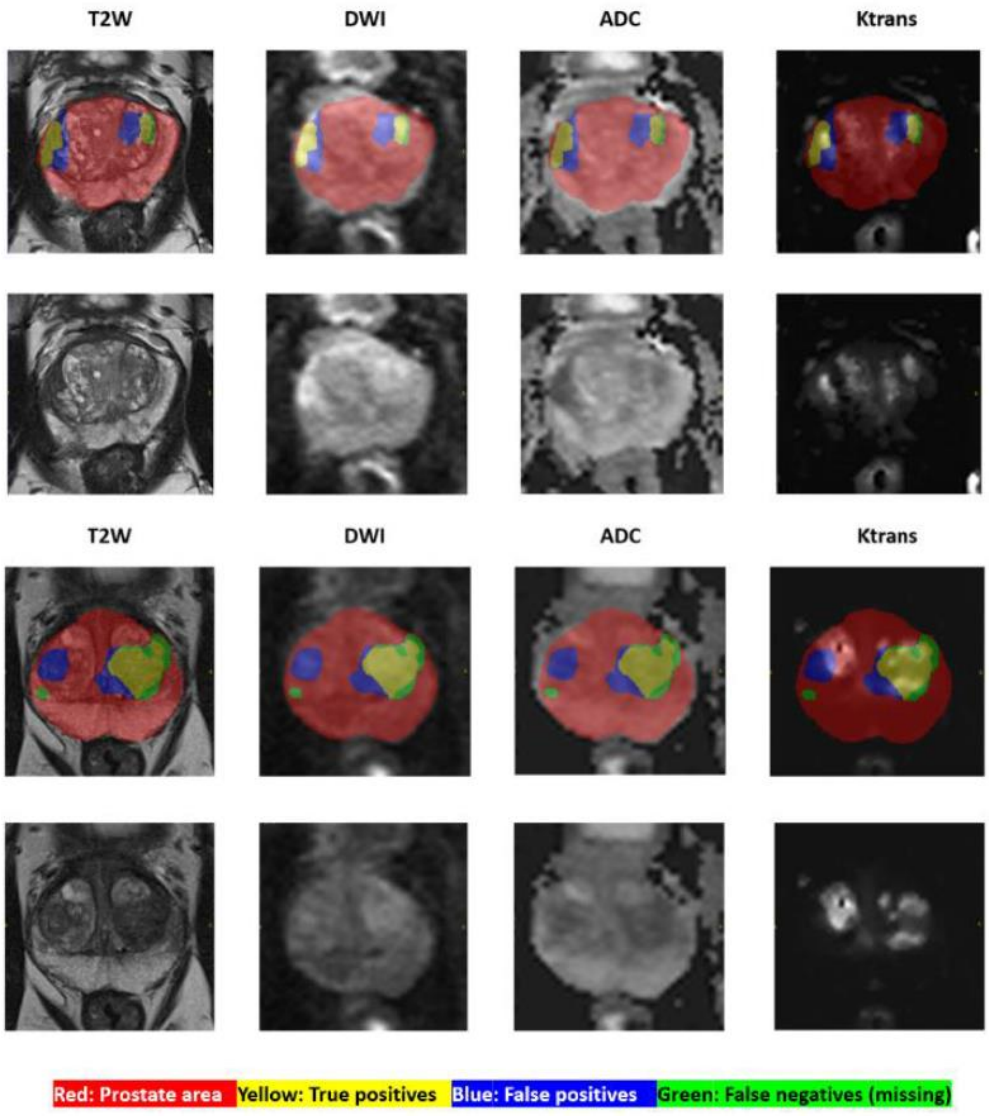


**Figure 23:** Two examples of detection results generated by the FCN with cascaded training. The upper example demonstrates two successful detections, while the lower one demonstrated one successful detection, one false-positive and one undetected tumor

### 5.5 Conclusions

In this study, we developed and tested a mpMRI based CAD scheme for prostate cancer detection. The CAD scheme consists of a prostate segmentation stage and a tumor detection stage. We employed state-of-art deep learning models to address the challenges of hand-crafted feature extraction. The proposed scheme demonstrates a

number of unique characteristics. First, instead of using single T2W modality for segmentation, we developed a FCN model to combine imaging information from multiple MR modalities, which enables faster convergence and more accurate segmentation performance. It demonstrates that T1 and DWI volumes may provide complementary information for defining prostate boundary. Second, we investigated the feasibility of applying a FCN based network architecture to generate voxel-level prediction for prostate tumor detection. We proposed a cascaded training strategy and demonstrated its effectiveness by experiments. Despite the encouraging results, this is a preliminary study with a relatively small dataset. The performance of the proposed CAD scheme needs to be further evaluated with a larger dataset in the future.

# Chapter 6: Cascaded fully convolutional network for nuclear segmentation from histology images[2]

## 6.1 Introduction

### 6.1.1 Background

In current clinical practice, histopathology data analysis plays an important role for cancer diagnosis and prognosis. Tissue histology images can be saved as digital images using whole slide imaging scanner. These images provide rich diagnostic information for different aspects of the diseases and cancers [119]. Subjective reading and interpreting histology specimens using microscopes or digital images generated by whole slide scanners by pathologists has relatively limited reproducibility due to the inconsistency to identify the regions of interest in the large image searching area of the whole slide, as well as inter and/or intra-observer variability [119, 120]. Therefore, in the recent two decades, developing computer-aided detection and diagnosis (CAD) schemes of digital pathology has attracted extensive research interests to overcome the limitation of manual assessment [121]. Among the many CAD tasks in digital pathology, accurate segmentation of nuclear is an important prerequisite step in a number of computational pathology pipelines. Some characteristics of nuclear such as density, size and nucleus-to-cytoplasm ratio etc. are informative for cancer grading and assessment of treatment effectiveness [122, 123]. Cell nuclei counting is also helpful for diagnosis of a few cancerous diseases [124]. Therefore, development of automated nuclei segmentation algorithm from histology images is important in computational

[2] Works of this chapter were done during an internship at Sensetime.

pathology. However, accurate detection and segmentation of nuclei is difficult because of (1) the great variation of appearances among different diseases and organs, and (2) the challenges of accurately separating crowded and adjacent nuclei into individual ones.

### 6.1.2 Related work

Traditional image processing based segmentation methods have been extensively explored in previous nuclei segmentation studies. Commonly used techniques include watershed segmentation algorithm, active contour model, level-set algorithm and graph based models etc. [125-129]. The performance of such methods is relatively poor because of the large variation of appearances between different nuclei and different images. The challenges can be addressed by applying machine learning based techniques, which can significantly improve the segmentation accuracy by recognizing different appearances of nuclei through training. Shallow classifiers with hand-crafted features were commonly used in earlier works. For example, Kong et al. extracted local neighborhood based color-texture features and applied an expectation maximization linear discriminant analysis (EMLDA) classifier for cell segmentation [130]. However, how to design effective hand-crafted features to discriminate nuclei and background still remains a great challenge.

Recently, deep learning models, especially Convolutional Neural Networks (CNNs) have been extensively investigated for object detection and segmentation for both natural images and medical images. Standard CNN for binary classification cannot be directly applied for nuclear segmentation because it cannot separate touching nuclei. There are a number of recent studies that proposed CNN based models with extra post-

processing steps for nuclei segmentation [131-133]. Kumar et al. developed a CNN architecture with three convolutional layers to classify each pixel as belonging to nuclei internal, nuclei contour or background, in order to segment nuclei area and separate adjacent nuclei [131]. The efficiency of this method is low because of repeated computations. Among CNN architectures, fully convolutional network (FCN) is extensively used for semantic segmentation because it is efficient for pixel-level prediction [91]. Therefore, in another work, Chen et al. proposed a multi-tasked FCN model namely deep contour-aware network (DCAN) for the nuclear segmentation task [132]. The DCAN model enables to predict the nuclei areas and contours simultaneously. Instead of directly predicting the contour, Naylor et al. proposed a FCN architecture for regression of distance maps and applied post-processing (i.e. watershed transformation) to separate touching nuclei [133].

### 6.1.3 Objective

In deep learning based nuclear segmentation frameworks, how to accurately separate adjacent nuclei still remains a great challenge although various approaches have been proposed to address it. In this study, our motivation is to investigate novel approaches to improve the segmentation accuracy of nuclear from histology images. Specifically, we proposed a deep neural network architecture with cascaded two-stage FCNs. For the first FCN, we employ the state-of-art image classification and segmentation model namely deep layer aggregation (DLA) [134] to predict the nuclei mask and an intermediate direction mask for additional supervision. The second FCN is a standard U-Net [92], which is used to generate the adjacent contour information. The

performance of the proposed method is evaluated using a public histology dataset and we demonstrated promising experimental results.

### 6.2 Dataset and pre-processing

In this study, we used the data acquired from the image dataset published by Kumar et al. [131] for evaluation of our proposed nuclear segmentation method. The dataset consists of hematoxylin and eosin (H&E) stained tissue images captured at 40x magnification. 30 whole slide images (WSIs) were downloaded from The Cancer Genomic Atlas (TCGA) website. The histology images were obtained from different organs and different hospitals to maximize the variation of the dataset. Tissue samples from seven organs are included in the dataset, including breast, liver, kidney, prostate, bladder, colon and stomach. A sub-image with size of 1000 x 1000 was cropped for each WSI and the nuclear boundaries were annotated as ground truth. Totally there are about 21,000 annotated nuclei in the 30 histology images [131].

The images are normalized using the mean and standard deviation values obtained from ImageNet dataset for each color channel. Due to the limitation of the size of the dataset, extensive data augmentation is employed to avoid the potential over-fitting issues. The data augmentation includes random cropping of 512 x 512 patches from the original images, contrast-limited adaptive histogram equalization (CLAHE), scaling, up/down and right/left flip, rotation, color jitter, Gaussian noise and elastic transformation. Figure 24 shows an example of an original histology image patch and its augmented images. The augmentation is performed on-the-fly with the training process, in order to maximize the variation of training set.

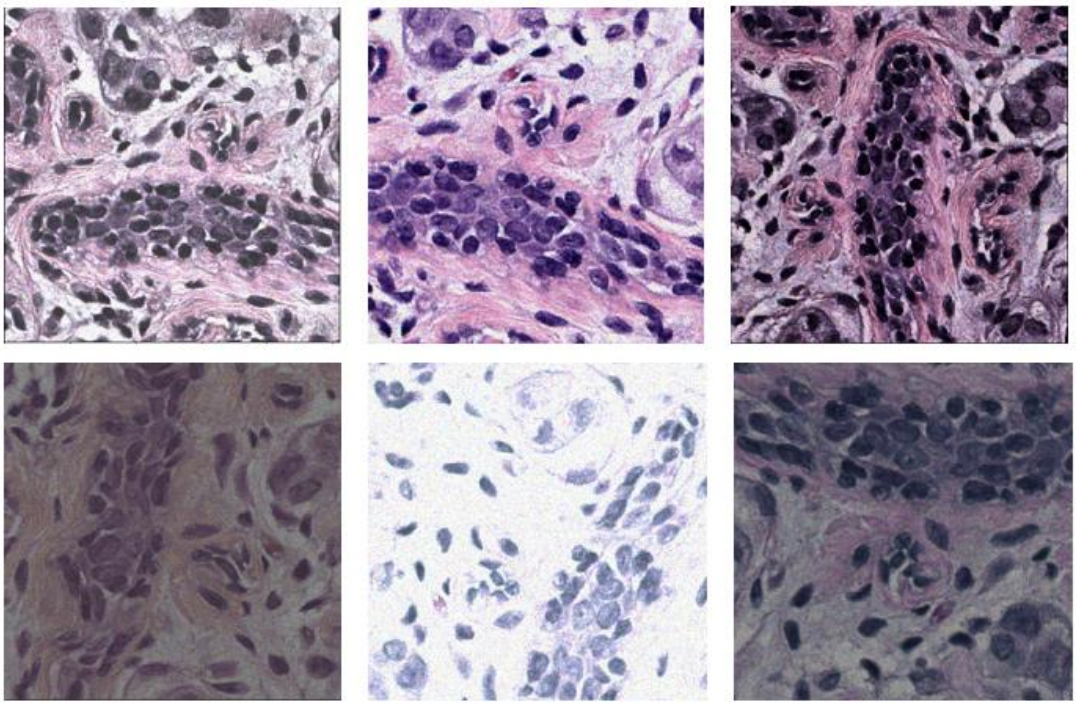

**Figure 24:** An example of an original image patch (up-left image) and its augmented images.

### 6.3 The proposed two-stage FCN

In this section we introduce our novel approach for nuclear segmentation. Previous studies have demonstrated the effectiveness of building multi-task deep neural networks to predict nuclei area and nuclei contours simultaneously [131, 132]. The nuclei contours can be further divided into two classes, including nuclei/background boundary and the boundary between adjacent nuclei. While most of the contour pixels belong to the first class, the pixels belonging to the second class are what we need to separate touching nuclei. Therefore, one potential limitation of the previous studies is that the pixels of nuclei/background boundary may dominate the training of the deep networks, leading to some failure for recognizing the boundary between adjacent nuclei. Instead of predicting the whole nuclear contour, we investigated to only predict the

boundary between adjacent nuclei in this study. In order to generate the ground truth label of such boundary, we apply a morphological dilation operation to enlarge the nuclei area, followed by a watershed algorithm to separate different nuclei. The dilated watershed lines are adopted as the label of boundary between adjacent nuclei. Figure 25 shows an example of an original image patch, the entire contour and the adjacent-boundary obtained by this process.

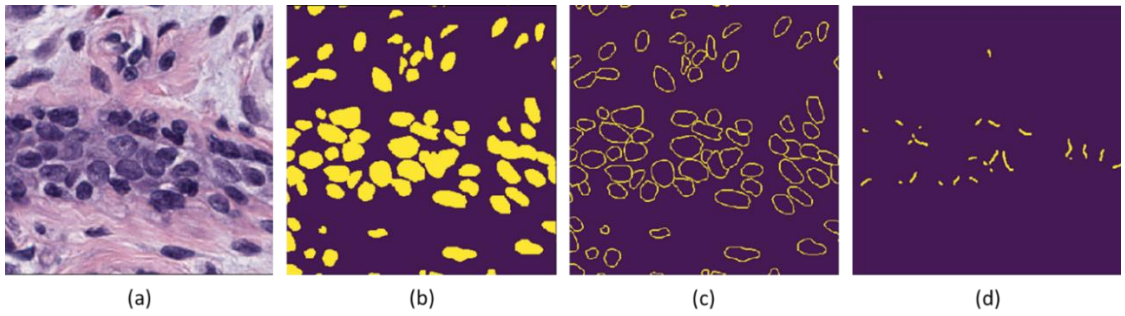


**Figure 25:** An example of (a) original image patch, (b) nuclei area, (c) nuclei contour and (d) boundary between adjacent nuclei

While predicting the nuclei area (i.e. semantic segmentation) is relatively straightforward, directly learning the boundary between adjacent nuclei is more complex and challenging. Inspired by the work of deep watershed transform for instance segmentation [135], we propose a cascaded two-stage FCN architecture with additional supervision of intermediate signals. Specifically, we adopt the state-of-art semantic segmentation architecture namely fully convolutional deep layer aggregation (DLA) [91] as the first stage FCN to predict the nuclei areas and the direction vectors from positive pixels to the centers of nuclei they belong to. The direction vectors are used as auxiliary supervision to assist the network for prediction of adjacent boundaries.

The feature maps in the last layer of stage-1 FCN as well as the outputs of stage-1 FCN are then fed into a stage-2 FCN, which is a standard U-Net [92] to predict the boundaries between adjacent nuclei. A few post-processing steps are applied on the predicted nuclei areas and adjacent boundaries to obtain the final instance segmentation results. Figure 26 shows an overview of the entire cascaded FCN architecture.

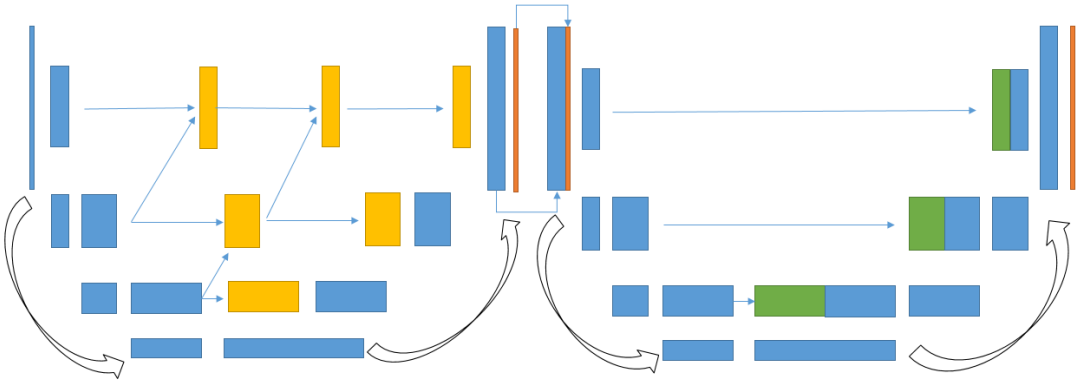

**Figure 26:** Our proposed cascaded FCN architecture. Blue blocks are general convolution blocks; yellow blocks are aggregation nodes; orange blocks are prediction layers and green block are copying nodes.

### 6.3.1 Stage-1 FCN

Deep layer aggregation model was proposed by Yu et al. with state-of-art performance for classification and segmentation [134]. Therefore, we adopt DLA model as our stage-1 network. In this section we first briefly introduce the architecture of DLA. The term “aggregation” is defined as the combination of feature maps from different layers. The skip connections in ResNet [9], DenseNet [136] and U-Net are considered as shallow aggregation because they are simple and linear. Instead of shallow aggregation, hierarchical deep aggregation (HDA) and iterative deep

aggregation (IDA) were developed in DLA, where the aggregation is non-linear and organized as a tree structure.

The architecture of a DLA for classification applications or backbone of detection/segmentation applications is shown in Figure 27 (a). The network consists of a number of stages and different stages correspond to different resolutions of feature maps. Inside single stages, HDA aggregates features from shallower and deeper hidden layers into a tree structure to generate richer information hierarchy. Between stages, IDA is applied to progressively integrate information from earlier stages and later stages. The residual blocks with cardinality which was proposed in ResNeXt model [137] are applied as the basic convolution blocks in DLA. The aggregation node is a concatenation operation followed by a standard convolution – batch norm – ReLU operation. Figure 27 (b) shows the comparison between a standard U-Net with shallow aggregations and the fully convolutional DLA architecture with additional IDAs for semantic segmentation.

In U-Net, the aggregations from encoder to decoder are skip connections which are simple and linear, resulting that the high resolution features aggregated to the decoder are relatively shallow. In fully convolutional DLA, the convolution blocks of standard DLA for classification are used as backbone and additional IDAs are applied to increase depth and resolution for pixel-level prediction. In this way, the high-resolution features from the encoder are aggregated more for obtaining stronger semantic information. We refer the readers to reference [134] for more details of DLA and fully convolutional DLA.

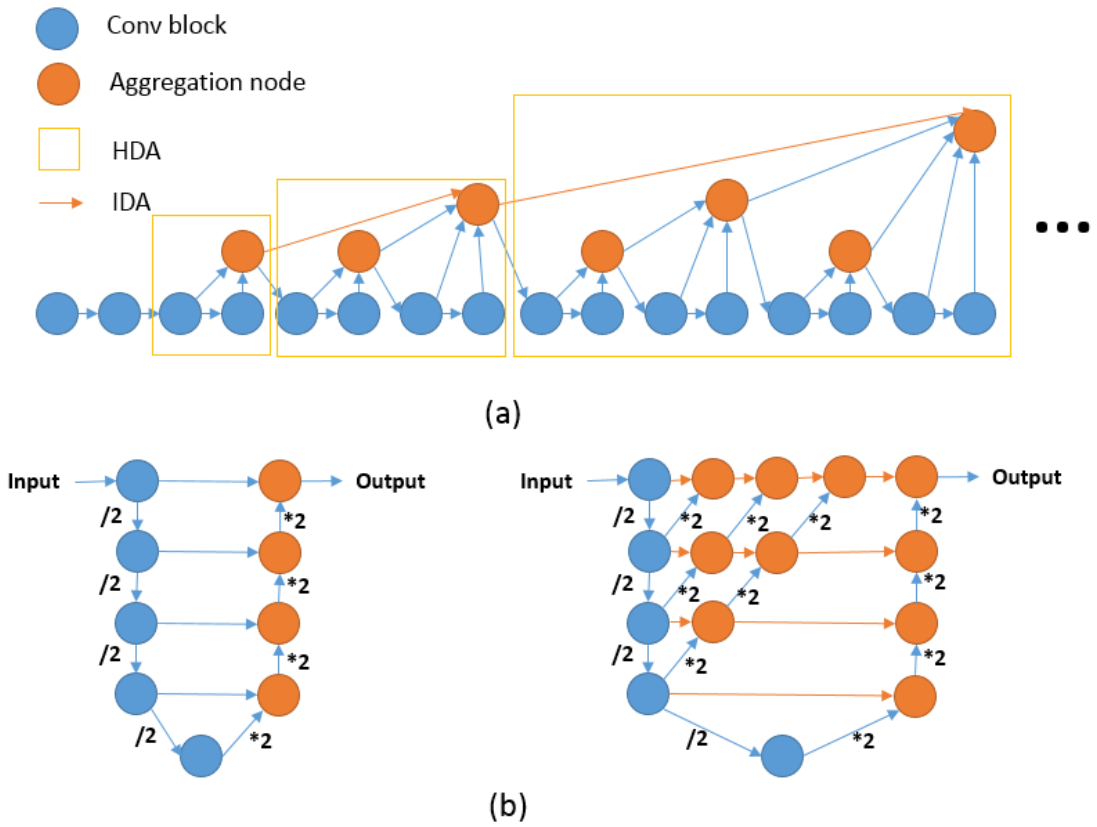


**Figure 27:** (a) Architecture of DLA. (b) Comparison of U-Net (left) and fully convolutional DLA (right)

While the deep watershed transform model takes the semantic mask as the input, our proposed model aims to generate the semantic segmentation and instance segmentation results simultaneously. The stage-1 FCN which is a fully convolutional DLA takes the augmented histology image patches as input and predicts a three-channel output. The first channel is the semantic mask of the nuclei area and the remaining two channels correspond respectively to the first and second dimension of the unit direction vector pointing from each positive pixel to the center of the nuclei it belongs to. The direction signals are not directly used for obtaining the instance segmentation results, but they provide auxiliary information to supervise the fully convolutional DLA to learn

useful features. The pixels which are close to each other but belongs to adjacent nuclei will have opposite direction vectors. Assigning such pixels to wrong nuclei center will cause a large regression error and get a heavy penalty [135]. Hence, the fully convolutional DLA is forced to put more attention on the pixels belonging to the boundary of adjacent nuclei. The learned features are then fed into the stage-2 FCN to predict the boundary of adjacent nuclei. Due to the limitation of training set, we initialized the encoder of fully convolutional DLA using the parameters of a classification DLA pre-trained on ImageNet dataset. This transfer learning approach greatly improves the efficiency of training the network.

### 6.3.2 Stage-2 FCN

The feature maps from the last layer of the fully convolutional DLA are concatenated with the three-channel output of the stage-1 FCN as the input of stage-2 FCN. The motivation of stage-2 FCN is to learn the boundary of adjacent nuclei from the hidden features and intermediate outputs of stage-1 FCN. Since the fully convolutional DLA in stage-1 has already extracted very deep features with strong semantic information and large receptive field, we only apply a relatively simple network for the second stage, which is a U-Net architecture. Compared to the original U-Net proposed by Ronneberger et al. [92], we made a few changes to improve the performance, including batch normalization and spatial dropout [138]. The difference between spatial dropout and standard dropout is that the spatial dropout randomly drops the entire feature maps instead of single neurons. The U-Net is randomly initialized without any pre-trained parameters, since the input has a relatively large dimensionality.

### 6.3.3 Implementation

We train the fully convolutional DLA in stage-1 and U-Net in stage-2 jointly. The loss function is the summation of three parts: the segmentation loss of semantic mask from the output of stage-1, the pixel-level regression loss of direction vectors from the output of stage-1, and the segmentation loss of adjacent boundary from the output of stage-2. Due to the imbalance of the positive area and background area, we use the summation of a cross entropy loss term and an intersection over union (IOU) loss term as the segmentation loss. The regression loss is a standard mean square error (MSE) term. The neural network is implemented in Python with PyTorch library. Adam optimizer with a learning rate of 1e-4 is applied to optimize the two-stage network to minimize the objective function.

### 6.3.4 Post-processing

By thresholding the probability maps generated by the proposed network, we can obtain a predicted nuclei mask and a predicted adjacent-boundary mask for each input image. The next step is to apply post-processing steps to obtain instance segmentation results. First, we aim to generate a marker mask. In an ideal marker mask, one nuclear corresponds to exactly one connected component namely a marker, and the markers of different nuclei are not connected to each other. The marker mask can be simply obtained by subtracting the adjacent-boundary mask from the nuclei mask. However, in some cases, the prediction of boundary is not accurate enough to separate adjacent nuclei. Figure 28 (a) shows such an example, where red color is the prediction of nuclei area and yellow is the prediction of boundary between adjacent nuclei. Therefore, we apply a few image processing steps to address this problem. Specifically,

we calculate the shortest distance from each positive pixel to the background and normalize it with the radius of the connected component it belongs to. Next, we threshold the normalized distance map and only keep the positive pixels which are relatively far away from the background as the marker mask. Figure 28 shows an example of the process for generating marker mask. A connected component analysis is then applied on the marker mask to label different markers with different IDs. Last, the instance segmentation results are obtained by applying a marker-based watershed algorithm that takes the nuclei mask and labelled marker mask as input.

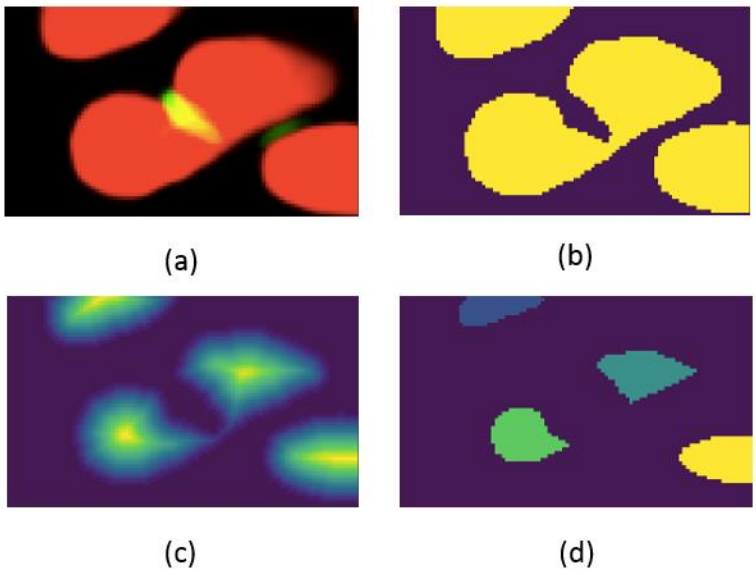


**Figure 28:** Illustration of post-processing. (a) An example of a probability map. (b) Marker mask obtained by subtracting adjacent boundary from the nuclei mask. (c) Normalized distance map. (d) Labelled marker mask by thresholding on normalized distance map.

### 6.4 Experiments and Results

In order to compare our model with previously reported methods using the same dataset, we use Aggregated Jaccard Index (AJI) as the evaluation metric [131]. In order to compute AJI, each ground truth nuclear is associated with a predicted nuclear which has the maximum IOU. The intersection of the two nuclear areas is added to the aggregated intersection while the union is added to the aggregated union. The areas of all unpaired predicted nuclei are also added to the aggregated union to penalize the false positives. The readers are referred to reference [131] for more details of AJI.

A 3-fold cross validation process is performed to evaluate the performance of our method. Table 8 shows the comparison of our proposed method with two previously published studies, including the 3-class CNN [131] and deep distance regression model [133]. We also implemented a few other neural network models for comparison purpose or ablation experiments. The models include: (1) Mask R-CNN [10], which is the state-of-art end-to-end model for instance segmentation in natural images; (2) Multi-task U-Net, which applies a standard U-Net to predict nuclei and adjacent-boundary simultaneously; (3) Multi-task fully convolutional DLA, which applies a pre-trained fully convolutional DLA to predict nuclei and adjacent-boundary simultaneously. Notable that the multi-task U-Net/fully convolutional DLA is a single-stage model without auxiliary supervision of direction signals, which is an ablated version of our proposed model. The results demonstrate that our proposed two-stage FCN outperforms all the other models. The fully convolutional DLA yields better results than standard U-Net, because of the superiority of network architecture. Using intermediate direction vector for supervision can further improve the segmentation accuracy of fully

convolutional DLA. We also demonstrate that Mask R-CNN performs relatively poor for the problem of nuclear segmentation, due to the differences between the natural images and histology images. The differences include the large amount of objects with small size and the unclear or fuzzy boundary between the touching instances in the digital histology images. Therefore, how to adapt Mask R-CNN architecture to the problem of nuclear segmentation still remains a great challenge and needs further investigations. Figure 29 shows several visualization examples of applying our method for nuclear segmentation.

**Table 8:** Comparison of different methods for nuclear segmentation

| Method | Train/test set | AJI |
|---|---|---|
| 3-class CNN [131] | 14 images for testing | 0.508 |
| Deep distance regression [133] | 14 images for testing | 0.560 |
| Multi-task UNet | 3-fold cross validation | 0.588 |
| Multi-task FC-DLA | 3-fold cross validation | 0.610 |
| Two-stage FCN | 3-fold cross validation | 0.621 |
| Mask R-CNN | 3-fold cross validation | 0.576 |

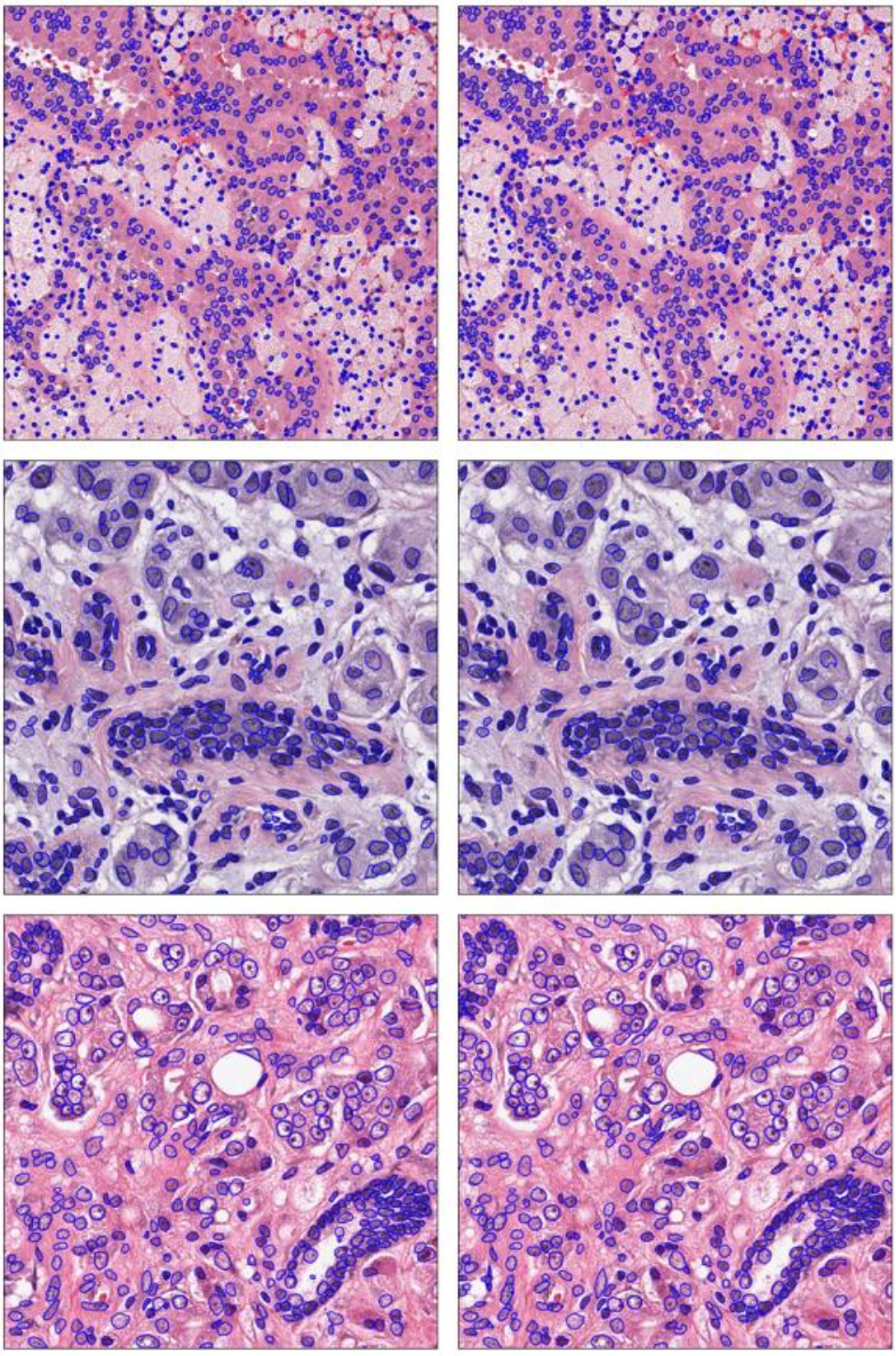

**Figure 29:** Three examples of the segmentation results. The left figures show the ground truth annotations while the right figures show the nuclei masks predicted by the two-stage FCN.

## 6.5 Conclusion

Developing automated CAD-type scheme for nuclear segmentation from histology images plays an import role in computational pathology. Due to the differences between natural images and histology images, the state-of-art instance segmentation models (e.g. Mask R-CNN) cannot be directly employed for nuclear segmentation. Alternatively, the feasibility of applying multi-task semantic segmentation models with post-processing steps has been demonstrated in previous studies. In this study, we continue the investigation of applying semantic segmentation models for addressing the task of nuclear segmentation. We proposed a novel network architecture which consists of two cascaded FCNs. Our method demonstrates a number of advantages over previous semantic segmentation based approaches. First, most previous works train the network to learn the entire contour of the nuclei, resulting that the majority of pixels belonging to the nuclei-background boundary dominate the training. Instead of using the entire contour, we investigate to focus on predicting the boundaries of adjacent nuclei. Second, we adopt the fully convolutional DLA, which is a state-of-art architecture for semantic segmentation, as our stage-1 FCN. We demonstrate by experiments that the fully convolutional DLA out-performs standard U-Net in a large margin. Last, we introduce intermediate supervision signals which are the direction vectors pointing from nuclei area to nuclei centers. This auxiliary supervision can also greatly improve the segmentation accuracy according to our results. Therefore, our study provides a reliable neural network architecture for instance segmentation in medical images.

# Chapter 7. Conclusions and future works

## 7.1 Conclusions

Deep learning techniques have revolutionized many image or vision based areas including computer vision and medical image analysis. From the studies reported in this dissertation, we demonstrated the feasibility of applying deep learning for a wide range of medical image analysis tasks. One important step of developing computer-aided diagnosis (CAD) scheme is ROI segmentation, including organ/mass from radiology images and gland/nuclei from pathology images. Development of accurate automated image segmentation scheme can greatly alleviate the tedious and repeating works of the doctors. Patch-based classification methods were commonly adopted in earlier deep learning based segmentation methods. We investigated the feasibility of this method for adipose tissue segmentation in Chapter 2 and further demonstrated that the multi-scale context and position information can improve the segmentation accuracy. One outstanding issue of patch-based segmentation scheme is that the execution is time-consuming due to large amount of repeating computations. Therefore, fully convolutional network architecture is usually considered as a better model for semantic segmentation. Different from computer vision applications where the natural images are 2D images, many radiology images are 3D images. Optimally extracting, computing and including the information from all dimensions in CAD-type image processing and feature analysis schemes is clinically important. In Chapter 4 we adopted a 3D fully convolutional network architecture for prostate segmentation from MRI, which is an important pre-processing step for detection of prostate cancer. In order to improve the performance, we proposed a coarse-to-fine residual module to utilize the information

from low-solution outputs and a densely connected convolutional module to combine auto-context information. In Chapter 5, we formulated the problem of prostate tumor detection as a semantic segmentation problem. While the tumor segmentation is visually more difficult, we applied a two-step training process with hard negative mining to address this challenge. Different from organ segmentation problems which are semantic segmentation tasks, nuclear segmentation is an instance segmentation problem, where we not only need to predict the label of each pixel, but also need to separate adjacent instances. In Chapter 6, we illustrated that the state-of-art instance segmentation (i.e. Mask RCNN) is not suited for nuclear segmentation due to the differences between natural images and histology images. We demonstrated the superiority of a cascaded fully convolutional network architecture with intermediate signal for auxiliary supervision. This approach adopts state-of-art semantic segmentation models with post-processing steps for the problem of instance segmentation.

Other than ROI segmentation, another challenge of developing CAD scheme is how to design and compute effective image features for cancer/disease classification. Although using deep learning can avoid design and/or identification of hand-crafted image features, it is often difficult to directly apply deep learning models to extract effective features for medical image classification because of the limitation of medical image dataset sizes. In Chapter 3, we explored to combine transfer learning technique with traditional hand-crafted features to improve the classification accuracy of breast masses from mammography. We demonstrated that the hand-crafted image features and the image features transferred from ImageNet dataset may provide complementary information to each other. In Chapter 6, we also adopted the concept of transfer learning

by initializing the parameters with a pre-trained model, in order to improve the training efficiency.

In summary, this dissertation presents a number of new approaches to make deep learning technologies being optimally applied to develop CAD schemes of medical images and demonstrates the superiority of applying deep learning based CAD for addressing a wide range of medical image analysis problems. These studies also provide a number of novel and reliable deep neural network frameworks for organ segmentation, dense instance segmentation and image classification with small datasets.

### 7.2 Future works

Training deep learning models usually requires a large training set. However, in the field of medical image analysis, it is difficult and expensive to obtain high quality annotations. As a result, the medical image datasets are relatively small compared to natural image datasets. How to effectively apply deep neural networks on limited medical image datasets still remains a great challenge and needs future research efforts. Transfer learning is considered as a common approach for tackling the problems of small training set. For example, Liu et al. [139] proposed a 3D Anisotropic Hybrid Network to transfer 2D features to 3D features recently. Due to the differences between natural images and medical images (e.g. color v.s. gray-scale, 2D v.s. 3D etc.), how to effectively transfer knowledges from natural images to medical images needs to be further explored in the future.

In another aspect, semi-supervised learning and weakly-supervised learning techniques also provide a potentially promising machine learning mechanism that can extract useful information from both annotated image data and un-annotated/weakly-

annotated image data. In medical imaging field, Zhu et al. [140] proposed a DeepEM algorithm to use weakly-labelled data to improve the performance of pulmonary nodule detection system. More research efforts should also be devoted for development of novel and powerful semi-supervised and weakly-supervised learning algorithms, especially for medical imaging applications.

# References


[1] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging,* vol. 35, pp. 1153-1159, 2016.

[2] M. N. Wernick, Y. Yang, J. G. Brankov, G. Yourganov, and S. C. Strother, "Machine learning in medical imaging," *IEEE signal processing magazine,* vol. 27, pp. 25-38, 2010.

[3] C.-H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, "Segmenting brain tumors with conditional random fields and support vector machines," in *International Workshop on Computer Vision for Biomedical Image Applications*, 2005, pp. 469-478.

[4] N. R. Pal, B. Bhowmick, S. K. Patel, S. Pal, and J. Das, "A multi-stage neural network aided system for detection of microcalcifications in digitized mammograms," *Neurocomputing,* vol. 71, pp. 2625-2634, 2008.

[5] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, and A. s. D. N. Initiative, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage,* vol. 65, pp. 167-175, 2013.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, pp. 436-444, 2015.

[7] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning,* vol. 2, pp. 1-127, 2009.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[10] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017, pp. 2980-2988.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91-99.

[12] A. Ng, J. Ngiam, C. Y. Foo, Y. Mai, and C. Suen, "UFLDL tutorial," ed, 2012.

[13] U. o. M. LISA lab. (2015). *Deep Learning Tutorial Release 0.1*. Available: http://deeplearning.net/tutorial/

[14] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of physiology*, vol. 195, pp. 215-243, 1968.

[15] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278-2324, 1998.

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[17] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, pp. 2493-2537, 2011.

[18] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717-1724.

[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.

[20] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, pp. 1285-1298, 2016.

[21] C. L. Ogden, M. D. Carroll, B. K. Kit, and K. M. Flegal, "Prevalence of childhood and adult obesity in the United States, 2011-2012," *Jama*, vol. 311, pp. 806-814, 2014.

[22] A. H. Kissebah, N. VYDELINGUM, R. MURRAY, D. J. EVANS, R. K. KALKHOFF, and P. W. ADAMS, "Relation of Body Fat Distribution to Metabolic Complications of Obesity*," *The Journal of Clinical Endocrinology & Metabolism*, vol. 54, pp. 254-260, 1982.

[23] E. C. Mun, G. L. Blackburn, and J. B. Matthews, "Current status of medical and surgical therapy for obesity," *Gastroenterology,* vol. 120, pp. 669-681, 2001.

[24] J.-P. Després, I. Lemieux, J. Bergeron, P. Pibarot, P. Mathieu, E. Larose, *et al.*, "Abdominal obesity and the metabolic syndrome: contribution to global cardiometabolic risk," *Arteriosclerosis, thrombosis, and vascular biology*, vol. 28, pp. 1039-1049, 2008.

[25] S. Makrogiannis, G. Caturegli, C. Davatzikos, and L. Ferrucci, "Computer-aided assessment of regional abdominal fat with food residue removal in CT," *Academic radiology,* vol. 20, pp. 1413-1421, 2013.

[26] B. Guiu, J. M. Petit, F. Bonnetain, S. Ladoire, S. Guiu, J.-P. Cercueil, *et al.*, "Visceral fat area is an independent predictive biomarker of outcome after first-line bevacizumab-based treatment in metastatic colorectal cancer," *Gut,* vol. 59, pp. 341-347, 2010.

[27] K. N. Slaughter, T. Thai, S. Penaroza, D. M. Benbrook, E. Thavathiru, K. Ding, *et al.*, "Measurements of adiposity as clinical biomarkers for first-line bevacizumab-based chemotherapy in epithelial ovarian cancer," *Gynecologic oncology,* vol. 133, pp. 11-15, 2014.

[28] Y. Wang, T. Thai, K. Moore, K. Ding, S. Mcmeekin, H. Liu, *et al.*, "Quantitative measurement of adiposity using CT images to predict the benefit of bevacizumab‐based chemotherapy in epithelial ovarian cancer patients," *Oncology Letters,* vol. 12, pp. 680-686, 2016.

[29] S. Rössner, W. Bo, E. Hiltbrandt, W. Hinson, N. Karstaedt, P. Santago, *et al.*, "Adipose tissue determinations in cadavers--a comparison between cross-

sectional planimetry and computed tomography," *International journal of obesity,* vol. 14, pp. 893-902, 1990.

[30] T. Yoshizumi, T. Nakamura, M. Yamane, A. H. M. Waliul Islam, M. Menju, K. Yamasaki, *et al.*, "Abdominal Fat: Standardized Technique for Measurement at CT 1," *Radiology*, vol. 211, pp. 283-286, 1999.

[31] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *Neural Networks, IEEE Transactions on*, vol. 8, pp. 98-113, 1997.

[32] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *null*, 2003, p. 958.

[33] A. Brebisson and G. Montana, "Deep Neural Networks for Anatomical Brain Segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 20-28.

[34] C. Cernazanu-Glavan and S. Holban, "Segmentation of bone structure in X-ray images using convolutional neural network," *Adv. Electr. Comput. Eng,* vol. 13, pp. 87-94, 2013.

[35] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, ed: Springer, 2013, pp. 411-418.

[36] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, et al., "Brain tumor segmentation with deep neural networks," Medical image analysis, vol. 35, pp. 18-31, 2017.

[37] A. Dubrovina, P. Kisilev, B. Ginsburg, S. Hashoul, and R. Kimmel, "Computational mammography using deep neural networks," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 6, pp. 243-247, 2018.

[38] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, *et al.*, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2015*, ed: Springer, 2015, pp. 556-564.

[39] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, S. Zhang, *et al.*, "Multi-Instance Deep Learning: Discover Discriminative Local Anatomies for Bodypart Recognition," *IEEE transactions on medical imaging*, vol. 35, pp. 1332-1343, 2016.

[40] H. Chung, D. Cobzas, L. Birdsell, J. Lieffers, and V. Baracos, "Automated segmentation of muscle and adipose tissue on CT images for human body composition analysis," in *SPIE Medical Imaging*, 2009, pp. 72610K-72610K-8.

[41] S. Hussein, A. Green, A. Watane, G. Papadakis, M. Osman, and U. Bagci, "Context Driven Label Fusion for segmentation of Subcutaneous and Visceral Fat in CT Volumes," *arXiv preprint arXiv:1512.04958*, 2015.

[42] Y. J. Kim, S. H. Lee, T. Y. Kim, J. Y. Park, S. H. Choi, and K. G. Kim, "Body fat assessment method using CT images with separation mask algorithm," *Journal of digital imaging*, vol. 26, pp. 155-162, 2013.

[43] S. D. Mensink, J. W. Spliethoff, R. Belder, J. M. Klaase, R. Bezooijen, and C. H. Slump, "Development of automated quantification of visceral and

subcutaneous adipose tissue volumes from abdominal CT scans," in *SPIE Medical Imaging*, 2011, pp. 79632Q-79632Q-12.

[44] D. Romero, J. C. Ramirez, and A. Mármol, "Quanification of subcutaneous and visceral adipose tissue using CT," in *Medical Measurement and Applications, 2006. MeMea 2006. IEEE International Workshop on*, 2006, pp. 128-133.

[45] B. Zhao, J. Colville, J. Kalaigian, S. Curran, L. Jiang, P. Kijewski, *et al.*, "Automated quantification of body fat distribution on volumetric computed tomography," *Journal of computer assisted tomography*, vol. 30, pp. 777-783, 2006.

[46] J. K. Leader, B. Zheng, R. M. Rogers, F. C. Sciurba, A. Perez, B. E. Chapman, *et al.*, "Automated lung segmentation in X-ray computed tomography: development and evaluation of a heuristic threshold-based scheme1," *Academic radiology*, vol. 10, pp. 1224-1236, 2003.

[47] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, *et al.*, "Theano: a CPU and GPU math expression compiler," in *Proceedings of the Python for scientific computing conference (SciPy)*, 2010, p. 3.

[48] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: a cancer journal for clinicians*, vol. 66, pp. 7-30, 2016.

[49] L. Tabár, B. Vitak, H. H. T. Chen, M. F. Yen, S. W. Duffy, and R. A. Smith, "Beyond randomized controlled trials," *Cancer*, vol. 91, pp. 1724-1731, 2001.

[50] J. J. Fenton, J. Egger, P. A. Carney, G. Cutter, C. D'Orsi, E. A. Sickles, *et al.*, "Reality check: perceived versus actual performance of community mammographers," *American Journal of Roentgenology*, 2012.

[51] J. Brodersen and V. D. Siersma, "Long-term psychosocial consequences of false-positive screening mammography," *The Annals of Family Medicine,* vol. 11, pp. 106-115, 2013.

[52] R. M. Nishikawa, "Current status and future directions of computer-aided diagnosis in mammography," *Computerized Medical Imaging and Graphics,* vol. 31, pp. 224-235, 2007.

[53] Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, and C. E. Metz, "Computerized classification of benign and malignant masses on digitized mammograms: a study of robustness," *Academic Radiology,* vol. 7, pp. 1077-1084, 2000.

[54] W. K. Lim and M. J. Er, "Classification of mammographic masses using generalized dynamic fuzzy neural networks," *Medical physics,* vol. 31, pp. 1288-1295, 2004.

[55] N. R. Mudigonda, R. Rangayyan, and J. L. Desautels, "Gradient and texture analysis for the classification of mammographic masses," *IEEE transactions on medical imaging,* vol. 19, pp. 1032-1043, 2000.

[56] M. Tan, J. Pu, and B. Zheng, "Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model," *International journal of computer assisted radiology and surgery,* vol. 9, pp. 1005-1020, 2014.

[57] Y. Wang, F. Aghaei, A. Zarafshani, Y. Qiu, W. Qian, and B. Zheng, "Computer-aided classification of mammographic masses using visually sensitive image features," *Journal of X-ray science and technology,* vol. 25, pp. 171-186, 2017.

[58] A. de Brebisson and G. Montana, "Deep neural networks for anatomical brain segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 20-28.

[59] A. Dubrovina, P. Kisilev, B. Ginsburg, S. Hashoul, and R. Kimmel, "Computational mammography using deep neural networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1-5, 2016.

[60] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, *et al.*, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE transactions on medical imaging*, vol. 35, pp. 1322-1331, 2016.

[61] Y. Wang, Y. Qiu, T. Thai, K. Moore, H. Liu, and B. Zheng, "A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images," *Computer Methods and Programs in Biomedicine*, 2017.

[62] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Convolutional neural networks for mammography mass lesion classification," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 797-800.

[63] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Computer methods and programs in biomedicine*, vol. 127, pp. 248-257, 2016.

[64] Y. Qiu, S. Yan, M. Tan, S. Cheng, H. Liu, and B. Zheng, "Computer-aided classification of mammographic masses using the deep learning technology: a preliminary study," in *SPIE Medical Imaging*, 2016, pp. 978520-978520-6.

[65] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, *et al.*, "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition," in *Icml*, 2014, pp. 647-655.

[66] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in neural information processing systems*, 2014, pp. 3320-3328.

[67] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 806-813.

[68] D. Lévy and A. Jain, "Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks," *arXiv preprint arXiv:1612.00542*, 2016.

[69] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221-231, 2016.

[70] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging*, vol. 3, pp. 034501-034501, 2016.

[71] W. Sun, B. Zheng, and W. Qian, "Automatic Feature Learning Using Multichannel ROI Based on Deep Structured Algorithms for Computerized Lung Cancer Diagnosis," *Computers in Biology and Medicine*, 2017.

[72] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675-678.

[73] S. S. Sawilowsky and R. C. Blair, "A more realistic look at the robustness and Type II error properties of the t test to departures from population normality," *Psychological bulletin,* vol. 111, p. 352, 1992.

[74] H. H. Yang and J. E. Moody, "Data Visualization and Feature Selection: New Algorithms for Nongaussian Data," in *NIPS*, 1999.

[75] I. Rodriguez-Lujan, R. Huerta, C. Elkan, and C. S. Cruz, "Quadratic programming feature selection," *Journal of Machine Learning Research,* vol. 11, pp. 1491-1516, 2010.

[76] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, *et al.*, "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)," *European journal of cancer,* vol. 45, pp. 228-247, 2009.

[77] B. Zheng, A. Lu, L. A. Hardesty, J. H. Sumkin, C. M. Hakim, M. A. Ganott, *et al.*, "A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment," *Medical physics,* vol. 33, pp. 111-117, 2006.

[78] B. Zheng, Y. H. Chang, and D. Gur, "On the reporting of mass contrast in CAD research," *Medical physics,* vol. 23, pp. 2007-2009, 1996.

[79] G. M. te Brake, N. Karssemeijer, and J. H. Hendriks, "An automatic method to discriminate malignant masses from normal tissue in digital mammograms1," *Physics in Medicine and Biology*, vol. 45, p. 2843, 2000.

[80] Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, P. Lu, D. E. Wolverton, *et al.*, "Analysis of spiculation in the computerized classification of mammographic masses," *Medical Physics,* vol. 22, pp. 1569-1579, 1995.

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

[82] G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, *et al.*, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," *Medical image analysis,* vol. 18, pp. 359-373, 2014.

[83] D. Mahapatra and J. M. Buhmann, "Prostate MRI segmentation using learned semantic knowledge and graph cuts," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 756-764, 2014.

[84] Y. Ou, J. Doshi, G. Erus, and C. Davatzikos, "Multi-atlas segmentation of the prostate: A zooming process with robust registration and atlas selection," *MICCAI Grand Challenge: Prostate MR Image Segmentation,* vol. 2012, 2012.

[85] S. Klein, U. A. Van Der Heide, I. M. Lips, M. Van Vulpen, M. Staring, and J. P. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Medical physics,* vol. 35, pp. 1407-1417, 2008.

[86] R. Toth and A. Madabhushi, "Multifeature landmark-free active appearance models: application to prostate MRI segmentation," *IEEE Transactions on Medical Imaging,* vol. 31, pp. 1638-1650, 2012.

[87] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE transactions on medical imaging,* vol. 35, pp. 1077-1089, 2016.

[88] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561,* 2015.

[89] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, *et al.*, "A survey on deep learning in medical image analysis," *arXiv preprint arXiv:1702.05747,* 2017.

[90] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering,* 2017.

[91] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

[92] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241.

[93] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," *arXiv preprint arXiv:1608.06993,* 2016.

[94] H. Chen, Q. Dou, L. Yu, and P.-A. Heng, "Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation," *arXiv preprint arXiv:1608.05895*, 2016.

[95] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, *et al.*, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 556-564.

[96] H. R. Roth, L. Lu, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested networks for automated pancreas segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 451-459.

[97] J. A. Dowling, J. Fripp, S. Chandra, J. P. W. Pluim, J. Lambert, J. Parker, *et al.*, "Fast automatic multi-atlas segmentation of the prostate from 3D MR images," in *International Workshop on Prostate Cancer Imaging*, 2011, pp. 10-21.

[98] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*, 2016, pp. 565-571.

[99] L. Yu, X. Yang, H. Chen, J. Qin, and P.-A. Heng, "Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images," in *AAAI*, 2017, pp. 66-72.

[100] R. Cheng, H. R. Roth, N. S. Lay, L. Lu, B. Turkbey, W. Gandler, *et al.*, "Automatic magnetic resonance prostate segmentation by deep learning with

holistically nested networks," *Journal of Medical Imaging,* vol. 4, p. 041302, 2017.

[101] J. Kirby, "NCI-ISBI 2013 challenge-automated segmentation of prostate structures," 2013.

[102] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, *et al.*, "3D deeply supervised network for automated segmentation of volumetric medical images," *Medical Image Analysis,* 2017.

[103] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249-256.

[104] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Artificial Intelligence and Statistics*, 2015, pp. 562-570.

[105] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646-1654.

[106] Z. Tu, "Auto-context and its application to high-level vision tasks," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.

[107] N. Xu, B. Price, S. Cohen, and T. Huang, "Deep Image Matting," *arXiv preprint arXiv:1703.03872,* 2017.

[108] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[109] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[110] K. Kitajima, Y. Kaji, Y. Fukabori, K. i. Yoshida, N. Suganuma, and K. Sugimura, "Prostate cancer detection with 3 T MRI: Comparison of diffusion-weighted imaging and dynamic contrast-enhanced MRI in combination with T2-weighted imaging," *Journal of Magnetic Resonance Imaging*, vol. 31, pp. 625-631, 2010.

[111] P. Kozlowski, S. D. Chang, E. C. Jones, K. W. Berean, H. Chen, and S. L. Goldenberg, "Combined diffusion-weighted and dynamic contrast-enhanced MRI for prostate cancer diagnosis—Correlation with biopsy and histopathology," *Journal of Magnetic Resonance Imaging*, vol. 24, pp. 108-113, 2006.

[112] M. A. Haider, T. H. Van Der Kwast, J. Tanguay, A. J. Evans, A.-T. Hashmi, G. Lockwood, *et al.*, "Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer," *American Journal of Roentgenology*, vol. 189, pp. 323-328, 2007.

[113] N. Hara, M. Okuizumi, H. Koike, M. Kawaguchi, and V. Bilim, "Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modality for the precise detection and staging of early prostate cancer," *The Prostate*, vol. 62, pp. 140-147, 2005.

[114] S. Wang, K. Burtt, B. Turkbey, P. Choyke, and R. M. Summers, "Computer aided-diagnosis of prostate cancer on multiparametric MRI: a technical review of current research," *BioMed research international*, vol. 2014, 2014.

[115] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "Computer-aided detection of prostate cancer in MRI," *IEEE transactions on medical imaging*, vol. 33, pp. 1083-1092, 2014.

[116] Y. K. Tsehay, N. S. Lay, H. R. Roth, X. Wang, J. T. Kwaka, B. I. Turkbey, *et al.*, "Convolutional Neural Network Based Deep-learning Architecture for Prostate Cancer Detection on Multiparametric Magnetic Resonance Images," in *SPIE Medical Imaging*, 2017, pp. 1013405-1013405-11.

[117] Y. Zhu, L. Wang, M. Liu, C. Qian, A. Yousuf, A. Oto, *et al.*, "MRI-based prostate cancer detection with high-level representation and hierarchical classification," *Medical physics*, vol. 44, pp. 1028-1039, 2017.

[118] S. Kohl, D. Bonekamp, H.-P. Schlemmer, K. Yaqubi, M. Hohenfellner, B. Hadaschik, *et al.*, "Adversarial Networks for the Detection of Aggressive Prostate Cancer," *arXiv preprint arXiv:1702.08014*, 2017.

[119] M. N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE reviews in biomedical engineering*, vol. 2, p. 147, 2009.

[120] H. Llewellyn, "Observer variation, dysplasia grading, and HPV typing: a review," *Pathology Patterns Reviews*, vol. 114, pp. S21-S35, 2000.

[121] D. N. Louis, M. Feldman, A. B. Carter, A. S. Dighe, J. D. Pfeifer, L. Bry, *et al.*, "Computational pathology: a path ahead," *Archives of pathology & laboratory medicine*, vol. 140, pp. 41-50, 2015.

[122] A. H. Beck, A. R. Sangoi, S. Leung, R. J. Marinelli, T. O. Nielsen, M. J. Van De Vijver, *et al.*, "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Science translational medicine*, vol. 3, pp. 108ra113-108ra113, 2011.

[123] P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, "Computer-Aided Breast Cancer Diagnosis Based on the Analysis of Cytological Images of Fine Needle Biopsies," *IEEE Trans. Med. Imaging*, vol. 32, pp. 2169-2178, 2013.

[124] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, 2008, pp. 284-287.

[125] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, pp. 2405-2414, 2006.

[126] M. Veta, P. J. Van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. Pluim, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PloS one*, vol. 8, p. e70221, 2013.

[127] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Processing,* vol. 122, pp. 1-13, 2016.

[128] Y. Gao, V. Ratner, L. Zhu, T. Diprima, T. Kurc, A. Tannenbaum, *et al.*, "Hierarchical nucleus segmentation in digital pathology images," in *Medical Imaging 2016: Digital Pathology*, 2016, p. 979117.

[129] P. Guo, A. Evans, and P. Bhattacharya, "Segmentation of nuclei in digital pathology images," in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2016 IEEE 15th International Conference on*, 2016, pp. 547-550.

[130] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting," *IEEE transactions on medical imaging,* vol. 30, pp. 1661-1677, 2011.

[131] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE transactions on medical imaging,* vol. 36, pp. 1550-1560, 2017.

[132] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Medical image analysis,* vol. 36, pp. 135-146, 2017.

[133] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of Nuclei in Histopathology Images by deep regression of the distance map," *IEEE Transactions on Medical Imaging,* 2018.

[134] F. Yu, D. Wang, E. Shelhamer, and T. Darrell, "Deep layer aggregation," *arXiv preprint arXiv:1707.06484*, 2017.

[135] M. Bai and R. Urtasun, "Deep watershed transform for instance segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2858-2866.

[136] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *CVPR*, 2017, p. 3.

[137] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 2017, pp. 5987-5995.

[138] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 648-656

[139] S. Liu, D. Xu, S. K. Zhou, O. Pauly, S. Grbic, T. Mertelmeier, et al., "3D Anisotropic Hybrid Network: Transferring Convolutional Features from 2D Images to 3D Anisotropic Volumes," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, pp. 851-858.

[140] W. Zhu, Y. S. Vang, Y. Huang, and X. Xie, "DeepEM: Deep 3D ConvNets With EM For Weakly Supervised Pulmonary Nodule Detection," arXiv preprint arXiv:1805.05373, 2018.