MEDICAL IMAGE SEGMENTATION USING RANDOM FOREST AND CONVOLUTIONAL NEURAL NETWORKS

DUAN YOU
SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

2018

MEDICAL IMAGE SEGMENTATION USING RANDOM FOREST AND CONVOLUTIONAL NEURAL NETWORKS

DUAN YOU

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

2018

Table of Contents

Abstract

Medical image analysis has received attention in recent years. It is an important part of disease analysis, diagnosis and treatment. Due to the limited number of doctors, some patients cannot receive timely treatment. In recent years, supervised learning algorithms have developed rapidly and have made major breakthroughs in the field of medical image analysis. Among them, Convolutional Neural Network (CNN) and Random Forest (RF) are proved to be powerful tools for computer vision tasks, including target recognition, segmentation and localization.

In this dissertation, I used the CNN and RF methods to process the Nuclei Segmentation image dataset, which include big number of targets and boundaries among these targets are not obvious, and then analyze and compare with the previous experimental phenomena, then use CNN and RF methods to process the digital retinal image (DRIVE) database, whose target is not obvious and light is unevenly distributed, and compare with state-of-the-art experiment.

This dissertation found that Convolutional Neural Network (CNN) works well in the field of image segmentation. Random Forest (RF) can also obtain good results when processing image segmentation, but it will be affected by the image noise.

Acknowledgments

The thesis would not have been completed without the instructions and support from my supervisor, Assoc.Prof Ponnuthurai Nagaratnam Suganthan. I thank him sincerely for giving me this valuable opportunity and memorable research experience.

Also, I heartily thank Rakesh Katuwal, for providing me a number of constructive comments and suggestions throughout the dissertation. He also taught me about research experience, which benefitted me a lot. Without his constant guidance this dissertation would not have been possible.

My gratitude also goes to the School of Electrical and Electronics Engineering at Nanyang Technological University, Singapore, for providing me with this opportunity to pursue my dissertation and for the facilities provided along with constant guidance and support, answering all the querries that I had during my dissertation.

Thank you to my parents and friends for their continuous support.

List of Figures

List of Tables

Chapter 1 Introduction

1.1. Background

In recent decades, the deep learning has made the breakthrough results of the image classification, object detection, segmentation, high-resolution image generation and many other areas. Machine learning has great potential in the field of medical image processing. With the design of algorithms that Are able to to generalize from observed evidences, and to make predictions about unseen data, machine learning has been applied in many fields such as computer aided diagnosis, detection and segmentation. Among these algorithm, the potential of Random Forests (RF) and Convolutional Neural Network (CNN) in medicine image processing has been recognized by researchers.

Random forests now become a popular supervised learning algorithm. This method has achieved very good performance in computer vision applications. Random forests consist of a large number of single decision trees. Random forests are very good models and provide a good probability framework for solving different learning tasks. Following the divide-and-conquer strategy, they effectively create partitions of the high-dimensional feature space and model probability distributions in each of these partitions. Therefore, they allow the classification, regression, or clustering of tasks to approximate any function or density.[1]

Another well-known algorithm is the Convolutional Neural Network (CNN). Its neurons can respond to a part of the surrounding cells in the covered area and are very powerful when dealing with large-scale images. A convolutional neural network consists of many convolutional layers (conv layers) and top-level fully connected layers (meaning classical neural networks). It also contains related weights and pool layers. This special structure allows the convolutional neural network to use the two-dimensional structure of the input data. Convolutional neural networks provide good performance in image and object recognition compared to other machine learning structures[2]

1.2. Motivation

Medical imaging is an important area of artificial intelligence in the coming years. In many parts of the world, medical facilities are incomplete, and the number of doctors is not enough. When it comes to some cases such as tumor recognition which need doctors to show high medical skills and rich experience, many people have to go to hospitals in big cities, which is time consuming and laborious, and puts a high demand on doctors' work.

However, if we have a method that can enable patients to enjoy the treatment of top doctors at local hospitals, the diagnosis time is shortened, and the diagnostic accuracy is improved. What will people do? Of course, they will choose this method. AI medical imaging is such a method.

AI medical imaging is a method of predicting the condition based on the analysis of a large numbers of expert diagnosis results and a large number of patient data using machine learning algorithms.

In a large numbers of machine learning algorithms, because of the limited number of datasets, the features of medical image data are complex, which puts high requirements on the choice of machine learning algorithms. After numerous experiments by researchers, it was proved that Convolutional Neural Network (CNN) and Random Forest (RF) performed well when dealing with such issues.

1.3 Objective

The main goal of this paper is to use convolutional neural network (CNN) and random forest (RF) to perform image segmentation on two medical image datasets. Since the number of datasets is small and the image definition is not good, high requirements are imposed on the algorithm's implementation and generalization capabilities.

In this article, we try to optimize the existing state-of-the-art results and try solutions that have not been used before. For the Nuclear segmentation dataset, first we implement the previous experiment by Marcelo Cicconet et al.(2016) [3] , and adjust the parameters to get the best performance. Then the

FCN network was used to process the nuclear segmentation dataset and good results were obtained. For the Digital Retinal Images for Vessel Extraction (DRIVE) database, we implement the previous experiment by Olaf Ronneberger et al.(2017)[4], then attempted to process this dataset with RF to obtain good result.

1.4 Organization of the Dissertation

The next chapter, chapter 2, is a review of the literature. We discuss what computer vision is, what image segmentation is, and what are the main problems encountered, and the principles of convolutional neural networks and random forest classification methods and their advantages and disadvantages, as well as the data set to be used in the experiment.

Chapter 3 discusses the equipment used and the experimental process.

In Chapter 4, we will introduce and discuss experiments and results. The discussion will include how to conduct the experiment, how to use two algorithms for image segmentation on two medical image datasets, and compare the states of the segmentation results on these datasets.

Finally, in Chapter 5, we conclude and provide suggestions for the scope of future work.

Chapter 2 Literature Review

This chapter describes the computer vision and image segmentation techniques, as well as convolutional neural networks, principle, structure, advantages and disadvantages of random forests. Finally, Markov random field technology for optimizing the spatial structure of three-dimensional images is introduced.

2.1 Computer Vision and Image segmentation

2.1.1 What is Computer vision

Computer vision is a science that studies how to make a machine "see". Further, it refers to the use of computers or other facilities rather than human eyes to identify, track and measure machine visions, and further to do graphics processing to make computers. As a scientific discipline, computer vision studies related theories and techniques and attempts to establish an artificial intelligence system capable of acquiring 'information' from images or multidimensional data. The information referred to here is defined by Shannon and can be used to help make a decision message. Because perception is treated as a process of extracting information, computer vision can also can be seen as the science of extracting information from images or multidimensional data through algorithms[5].

Computer vision is a simulation of biological vision using computers and related equipment. Its main task is to process the collected pictures or videos to obtain the three-dimensional information of the corresponding scenes, just as humans and many other types of creatures do every day.

Computer vision is the use of various computer vision systems as input data. The computer will replace the brain for learning and interpretation in the future. The ultimate goal of the research is to enable computers to understand things in a manner similar to the human visual system and to adapt to different environments. The goal that can be achieved through long-term efforts. Therefore, before the final goal is achieved, the medium-term goal of people's efforts is to establish a visual system that can accomplish certain tasks based on visual intelligence and feedback to some degree of intelligence. For example, an important application area of computer vision is the visual navigation of autonomous vehicles, and there is no condition to realize systems that can recognize and understand any environment like human beings and complete autonomous navigation. Therefore, the goal of people's hard work is to achieve a visual assisted driving system that has the ability to track roads on highways and avoid collisions with vehicles ahead. One point to be pointed out here is that the computer plays the role of replacing the human brain in the computer vision system, but it does not mean that the computer must complete the visual information processing according to the human visual method. Computer vision can and should process visual information based on the characteristics of the computer system. However, the human visual system is by far the most powerful and complete visual system known to people. As will be seen in the following sections, the study of human visual processing mechanisms will provide inspiration and guidance for computer vision research. Therefore, the use of computer information processing methods to study the human visual mechanism, the establishment of human visual computing theory, is also a very important and interesting areas of research. This research is called Computational Vision. Computational vision can be considered as a research area in computer vision[6].

2.1.2 Image Segmentation

Image segmentation means that the image is divided into several non-overlapping regions according to features such as grayscale, color, texture, and shape, and the features are similar in the same region, and appear clearly in different regions. Segmentation method, then we will carry out detailed understanding and research on individual methods.

2.1.2.1 Thresholding Method

The threshold method is one of the most basic method in image segmentation methods. It separates image pixels from their intensity levels. For example, if the pixel threshold is set to 128, the pixels in the picture that have more than 128 pixels are filtered out. This method is suitable for images with different background and object colors. The choice of these methods may be manual or automatic, that is, information that may be based on previous knowledge or image features[7] .

2.1.2.2 Edge Based Segmentation Method

Edge detection technology is a relatively common technique in image processing. The edge-based segmentation method is based on the great change of the intensity values of the image, one reason is that a single intensity value cannot provide enough features about the edges and it is easy to lose the features. Edge detection techniques have intensity methods to deal with edges and the first derivative of intensity is greater than a certain threshold, because it can detect small changes of edge. In the edge-based segmentation method, all edges are first detected and then connected together by some means to form an object boundary to divide the desired area. To detect edges, basic edge detection techniques such as the sobel operator method can use, and this method will also be used frequently [9].

2.1.2.3 Region Based Segmentation Method

The region-based segmentation method is a method of segmenting an image into regions having similar features. We introduce two technologies [10].

Region growing method: A segmentation method based on region growth is a method of dividing an image on different regions which are based on the growth of seeds. The selection of these seeds can be manual or automatic. The seed is then controlled by the connection between the pixels. Because there is some prior knowledge, special moments can be stopped [11] [12].

Region segmentation methods and merge methods: Segmentation methods based on region segmentation and merging use two basic techniques, namely segmentation and merging to segment an image into regions. Splitting means that the image is iteratively divided into regions with similar features. Merging helps to combine adjacent similar regions[13].

2.1.2.4 Clustering Based Segmentation Method

Clustering-based methods are working for segmenting an image into clusters with similarly characterized pixels. Data clustering is a important way of dividing data features into clusters, making features in the similar features. It has two types of clustering methods: hierarchical methods and partition-based methods. The layering method is based on the meaning of a tree. Here, the root of the tree represents the whole data and the internal nodes means clusters. Another thing is that the partition-based method uses an optimization method to iteratively minimize the objective function, and this method will also be used frequently [14] [15].

2.1.2.5 Watershed Based Methods

The watershed-based approach uses the meaning of topology interpretation. Here, the intensity means the basin in the smallest hole out of the water. When the water line reaches the basin boundary, adjacent basins merge. In order to hold the isolation between the basins, dams are necessary and are the boundaries of the subdivisions. These methods were built with extensions. The watershed method treats the gradient of the image as the terrain surface. This method will also be used frequently[16].

2.1.2.6 Partial Differential Equation Based Segmentation Method

The method based on partial differential equations is a practical method of segmentation. These apply to time-critical things. There are two basic PDE methods: nonlinear isotropic diffusion filters (for enhanced edges) and convex non-quadratic changes (for noise removal). The result of this method is blurry edges and borders, and this method is also be used frequently [17].

2.1.2.7 Artificial Neural Network Based Segmentation Method

Segmentation method based on neural network simulating human brain learning strategy for decision making. This way is now mainly used for the medical images segmentation. It is used to separate the desired image from the background and foreground. A neural network consists of lots of connected nodes, each with a specific weight. In this issue is converted into a problem, which is solved using neural networks. The method has two steps: extract features through a neural network and segment through a neural network[18].

Table below shows the compare among seven segmentation ways, including the description, advantages and disadvantages.

Segmentation technique	Description	Advantages	Disadvantages
Thresholding Method	based on the histogram peaks of the image to find threshold values	no need of previous information, simplest method	highly dependent on peaks, spatial details are not considered
Edge Based Method	based on discontinuity detection	good for images having better contrast between objects	not suitable for wrong detected or too many edges
Region Based Method	based on partitioning image into homogeneous regions	more immune to noise, useful when it is easy to define similarity criteria	expensive method in terms of time and memory
Clustering Method	based on division into homogeneous clusters	fuzzy uses partial membership therefore more useful for real problems	determining membership function is not easy
Watershed Method	based on topological interpretation	results are more stable, detected boundaries are continuous	complex calculation of gradients
PDE Based Method	based on the working of differential equations	fastest method, best for time critical applications	more computational complexity
ANN Based Method	based on the simulation of learning process for decision making	no need to write complex programs	more wastage of time in training

Table 2.1.2(a): Comparison of Various Segmentation Techniques

2.1.3 Medical Image Processing

The history of medical imaging technology dates back to 1895 when German physicist Roentgen discovered X-rays and used it for medical diagnosis and invented the X-ray imaging technology, which for the first time provided humans with anatomical pictures of human internal organ tissues in a non-destructive manner, thereby providing important clinical information for doctors in clinical diagnosis. This triggered a revolution in medical diagnostic technology. It is an interdisciplinary discipline that uses advanced technologies such as physics, electronics, and computer science to diagnose and treat diseases.

With the fast development of microelectronic technology, computer network technology, computer graphics and image processing technology, artificial intelligence and automatic control technology, modern medical imaging technology has become one of the fastest growing technical fields in the 21st century. With the advent of new imaging diagnostics and therapeutic methods such as ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), intervenient radiology, and positron emission tomography (PET), medicine imaging from scratch, from small to large, has experienced a rapid development process. In particularly, the emergence of introductory radiology has enabled the development of a simple diagnostic radiology room to become a large-scale clinical medical imaging department integrating diagnosis and treatment. Undoubtedly, in the new century, medical imaging technology will develop faster and be used in the medical field. An increasingly important role.

Here are some of the major types of medical imaging:

First, based on the radar and sonar technologies, a variety of ultrasound imaging technologies were developed using the principle of echolocation, and A-type, B-type, and M-type ultrasound diagnostic instruments were developed. Currently (transmission type) Ultrasound Computed Tomography (UCT) technology has matured. Ultrasound imaging has the advantages of no harm on body and high sensitivity. The observation of soft tissue does not require pre-imaging such as injection of contrast media, and the imaging is rapid and the equipment is cheap. It can reflect both the anatomical image of the organ and the functional status. Therefore, ultrasound imaging is currently the most widely used and fastest-developing technology in all imaging technologies.

Second, CT computed tomography (CT) is the use of X-rays to scan a certain range of the human body layer-by-layer, obtain information, after computer processing to obtain a reconstructed image (transverse anatomical map), obtained by computer processing three-dimensional reconstruction image. The cross-sections, faults, and digital images generated by CT solve the main defects of three-dimensional structure overlap, poor soft-tissue resolution, and low information efficiency in traditional images, achieving epoch-making innovations. However, before the successful development of multi-slice CT, CT was once at a relatively stagnant stage. Multi-layer CT technology has entered a new phase of the transition from peak to peak. Its major breakthroughs are: acquisition speed (scanning speed), imaging quality (spatial resolution and density resolution), and data acquisition range (scanning range). The relationship between constraints, so that the coordination of the best value through the improvement of technical methods has become an important research topic in the development of CT technology [19].

Third, MRI Magnetic resonance imaging (MRI), also known as nuclear magnetic resonance (NMR), is a new medical imaging technology developed rapidly in recent years. It is considered to be the most advanced and promising in the 20th century. In 1946 American scholars Bloch and Purcell first discovered the phenomenon of nuclear magnetic resonance, resulting in the discipline of nuclear magnetic resonance spectroscopy. Magnetic resonance imaging is compatible with the characteristics of ray technology and nuclear medicine. It not only displays morphological anatomy, but also shows the chemical structure of various tissues, as well as physiological, biochemical and dynamic information. Such as water content, fat content, F, Na, P and other elements. MRI is an electronic method to adjust the gradient field to achieve scanning. Therefore, it is not only possible to directly display a section of an arbitrary resolution, but also to obtain an infinite number of sections and use these sections for three-dimensional visualization[20].

In clinical application, compared with CT, MRI has no radiation damage, no bone artifacts, can be multi-faceted, multi-parametric imaging, a high degree of soft tissue resolving power, no need to use contrast media to display the unique vascular structure advantage. Almost all different diseases of the system for the body, such as tumors, inflammation, trauma, the implementation of pathological changes and a variety of congenital diseases. The reality of brain, spine and myelopathy is better than CT. It does not need angiographic contrast agent, which shows the structure of the blood vessels, so it has its own uniqueness in the identification of blood vessels, tumors, lymph nodes and vascular structures. It also has a soft tissue resolving power that is several times higher than the CT and sensitively detects changes in the water content in the tissue components. Therefore, it is often more effective and early than the CT to find that the lesion MRI can clearly and comprehensively show the heart chamber, the heart muscle, the pericardium and other small structures in the heart are reliable methods for diagnosing various heart diseases and heart function tests.

2.2 Convolutional Neural Networks

Convolutional Neural Network (CNN) is a feed-forward neural network and its neurons can respond to the surrounding cells in the coverage area and have excellent performance for large image processing. is a feed-forward neural network whose artificial neurons can respond to a part of the surrounding cells in the coverage area and have excellent performance for large image processing. It includes a convolutional layer and a pooling layer.

2.2.1 CNN Structure

2.2.1.1 Convolutional neural network

First, we introduce the neural network briefly, and introduce each unit of the next neural network as follows:

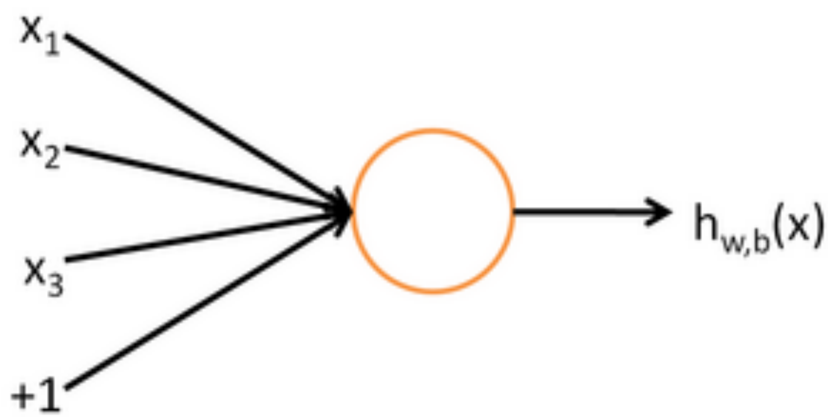


Figure2.2.1(a) Neural Network Structure

The formula shows as follows:

$$h_{W,b}(x) = f(W^T x) = f(\sum_{i=1}^{3} W_i x_i + b) \tag{1}$$

Among them, this unit can also be called a Logistic regression model. When multiple units are combined with each other, a hidden layer is formed to form a typical neural network model. The following figure shows a neural network with a hidden layer.

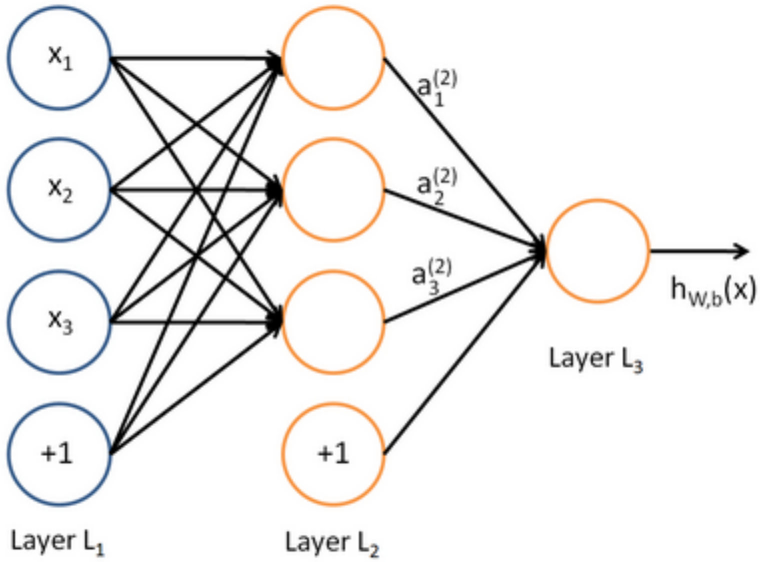


Figure2.2.1(b) Neural Network with Hidden Layer[21]

The corresponding formula is as follows:

$$
\begin{aligned}
a_1^{(2)} &= f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)}) \\
a_2^{(2)} &= f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)}) \\
a_3^{(2)} &= f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)}) \\
h_{W,b}(x) = a_1^{(3)} &= f(W_{11}^{(2)}a_1^{(2)} + W_{12}^{(2)}a_2^{(2)} + W_{13}^{(2)}a_3^{(2)} + b_1^{(2)})
\end{aligned}
\tag{2}
$$

2.2.1.2 Local perception

There are two types of artifacts inside convolutional neural network to reduce the parameters numbers. The first type of artifact is called a local perception field. It is thought that people's perception of the things is from the local to the global, and the image relation is also the local pixel is more closely linked, while the farther pixel correlation is weaker. Therefore, each neuron actually does not need to perceive the global image, but only needs to get the local images, and then the local information is integrated at a better one to obtain the global information. The idea of connecting parts of the network is also inspired by the structure of the visual system in biology. Neurons in the visual cortex are locally informed. As shown in the following figure: The left picture shows the full connection, and the right picture shows the partial connection[21].

2.2.1.3 Parameter sharing

We can think of these parameters (that is, convolution operations) as a way to extract features that are position-independent. The basic principle is that the statistical features including the image level are the same as other statistical properties. This provides a basis for us to use features in this section as well as share them to other parts, so we can share the same learning features at all different locations on the image.

More intuitively, when a small block is randomly selected from a large-size image, say 8x8 as a sample, and some features are learned from this small block sample, we can use the features learned from this 8x8 sample as detector is applied to any place in this image. In particular, we can use the features learned from the 8x8 samples to convolve with the original large-size image to obtain a different feature activation value for any position on this large-size image.

As shown in the figure below, the process of convolving a 3×3 convolution kernel on a 5×5 image is shown. Each convolution is a feature extraction method, just like a sieve, which filters out the parts of the image that meet the conditions (the larger the activation value is, the more qualified the conditions are).

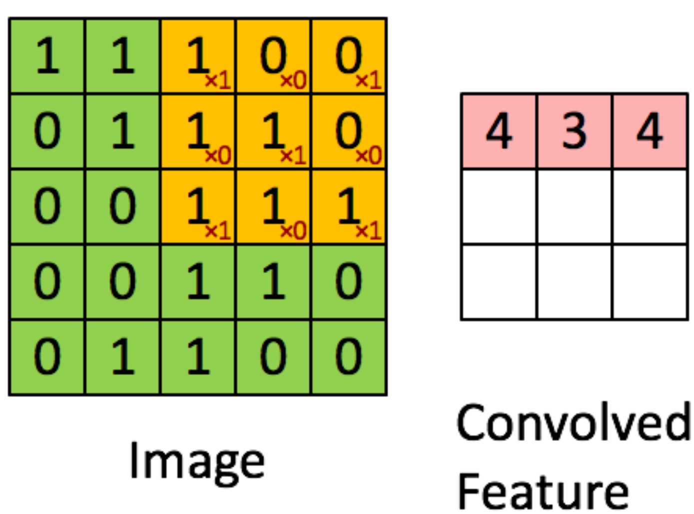


Figure2.2.1(c) How Convolution Works[21]

2.2.1.4 ReLU

ReLU is an element-wise operation (applied to each pixel) and replaces all the pixel which is negative values in the feature map with zeros. Now people use more ReLU rather than sigma euation. The purpose of ReLU is to introduce non-linearity in our convolutional layers because we want most of the actual data our ConvNet learns to be non-linear (convolution is a linear operation - element-wise matrix multiplication and addition, so we introduce Nonlinear functions like ReLU to solve non-linear problems).

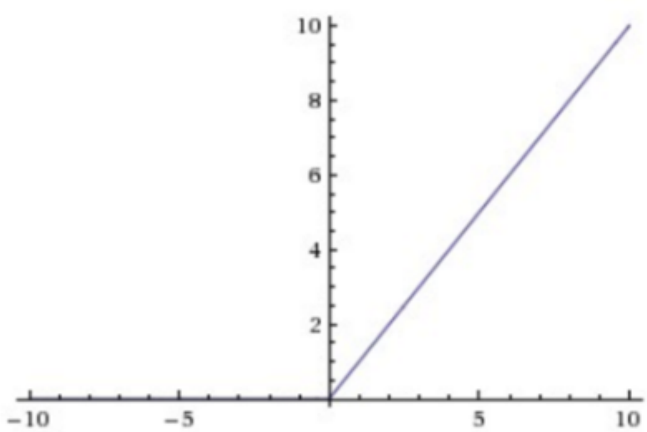


Figure 2.2.1(d) the ReLU operation[21]

2.2.1.5 Down-pooling

Down-pooling reduces the dimensionality of each feature map but still holds the most significant features. down pooling can be of different types: Max pooling, Average pooling.

For Max Pooling, we define a spatial neighborhood for testing (a 2x2 window, of course, other dimensions can be selected), and take the largest element from the correction feature map in the window and extract it. We can take the average (average summary) or the sum of all the elements in this window instead of using the largest element. In fact, Max Pooling found out after testing by many researchers that the results were very good.

The following figure shows an example of a maximum pool operation on a rectification feature map (save after convolution + ReLU operation) using a 2×2 window.

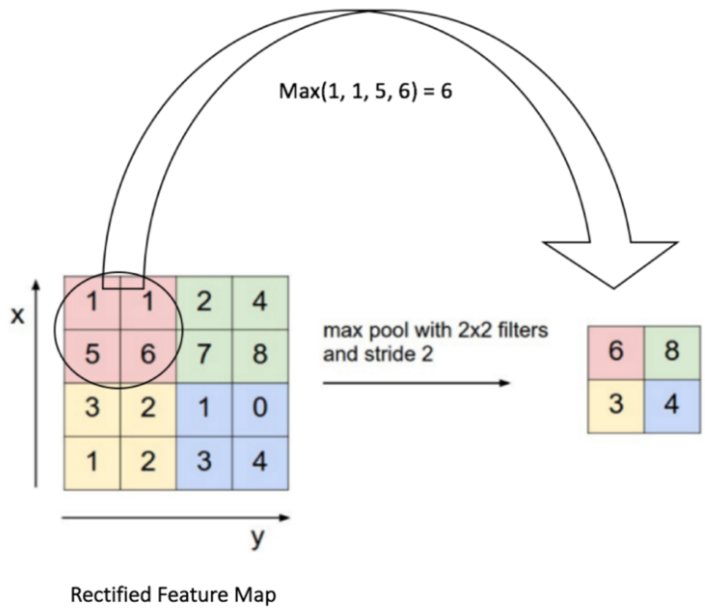


Figure 2.2.1(e) Max Pooling[21]

2.2.1.6 AlexNet structure

The following figure shows Alex's CNN structure. It should be noted that this model uses a 2-GPU parallel structure, that is, the 1st, 2nd, 4th, and 5th convolutional layers all train the model parameters into 2 parts. Here, further, the parallel structure is divided into data parallelism and model parallelism. Data parallelism means that the model structure is the same on different GPUs, but the training data is divided and trained separately to obtain different models, and then the models are merged. The model parallelism is to segment the model parameters of several layers. Different GPUs use the same data for training, and the results obtained are directly connected as the input of the next layer.

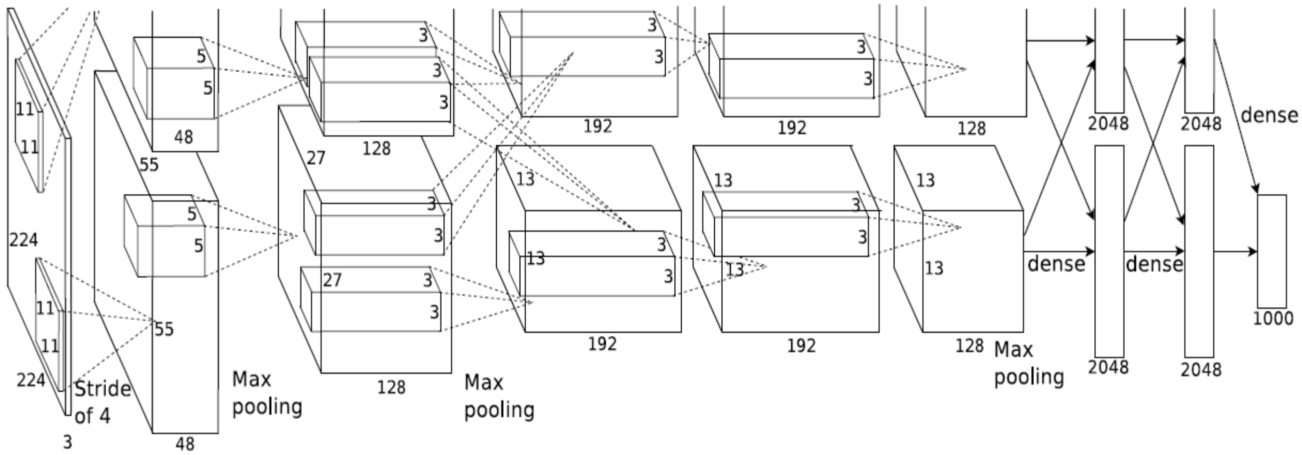


Figure2.2.1(g) AlexNet Structure[22]

The basic parameters of the above model are:

Input: 224×224 size picture, 3 channels

The first layer convolution: 96 convolution kernels of 11×11 size, 48 on each GPU.

The first layer max-pooling: 2x2 cores.

The second layer convolution: 256 5×5 convolution kernels, 128 on each GPU.

The second layer max-pooling: 2x2 core.

The third layer convolution: It is fully connected with the upper layer, with 3*3 convolution kernels 384. Divided into two GPUs on the 192.

The fourth level of convolution: 384 3×3 convolution kernels and 192 GPUs. This layer is connected to the last layer without going through the pooling layer.

The fifth layer convolution: 256 convolution kernels of 3×3, 128 on two GPUs.

The fifth layer max-pooling: 2x2 core.

The first layer is fully connected: 4096 dimensions, and the output of the fifth layer max-pooling is connected as a 1D vector, which is used as the input of the layer.

The second full connection: 4096 dimensions

Softmax layer: The output is 1000. Each dimension of the output is the probability that the picture belongs to which class.

2.2.2 FCN Structure

The Fully Convolutional Networks (FCN) transforms the fully connected layer in the traditional CNN into a convolutional layer. As shown in the figure below, in the traditional CNN layers, the front 5 layers are convolutional layers, the 6th and 7th layers are a one dimention vector, whose length is 4096. The 8th layer is a one dimention vector whose length is 1000. FCN represents these 3 layers as a convolutional layer, and the size (channel number, width, height) of the convolution kernel is (4096, 1, 1), (4096, 1, 1), $(1000, 1, 1)$ respectively. These added convolutional layers are called full convolutional networks.

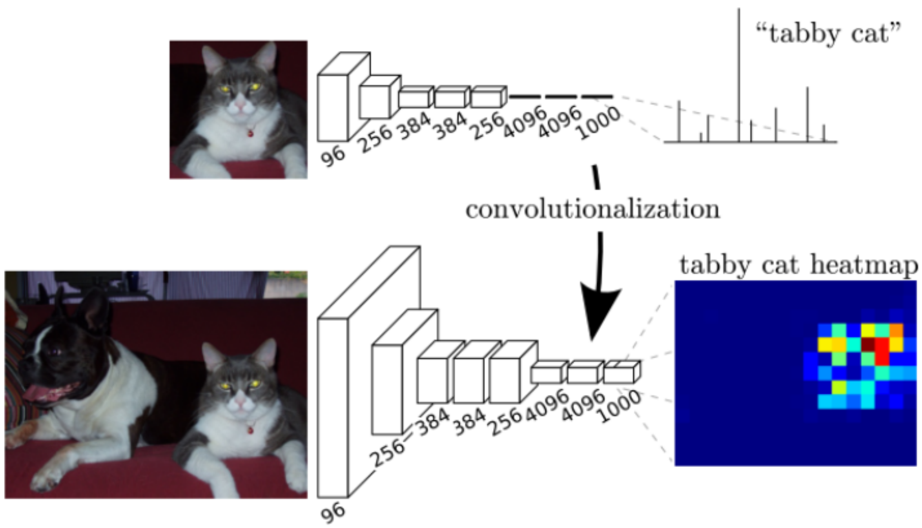


Figure 2.2.2(a) How CNN and FCN Works[22]

We know for image segmentation, the FCN training input is an image, and the output is also an image, semantically segmented pixels to pixels. This method exceeds the current highest accuracy without further processing. To the best of our knowledge, this is the first job to train end-to-end FCNs for pixel prediction, as well as work from pre-training. Another advantage is that the input set is scientifically derived from predictive output of any size input. Both learning and inference perform full image processing through feed-forward calculation and reverse propagation. The intra-upsampling layer allows pixel prediction and learning to share the network with subsampling pools[22].

It can be found that after multiple convolutions (and pooling), the output image is becoming smaller and smaller, in addition, the resolution is becoming lower (rough images). So how does FCN get the category of each pixel in the image? The restore the original image resolution from this coarse resolution image, FCN uses upsampling. Such as, after 5 times of convolution and pooling process, the resolution of the image has been reduced by 2, 4, 8, 16, 32 times. For the last layer of the output image, 32 times upsampling is needed to get the exact size as the initial image.

This upsampling is achieved through deconvolution. The output of the 5th layer is deconvolved to the initial image matrix size. The result obtained is still not good enough and some small parts can not be recovered. So some people deconvolved the output of the fourth layer and the output of the third layer one by one, which required 16 times and 8 times the upsampling, respectively, and the result was finer. The following figure shows the process of this convolution and deconvolution upsampling:

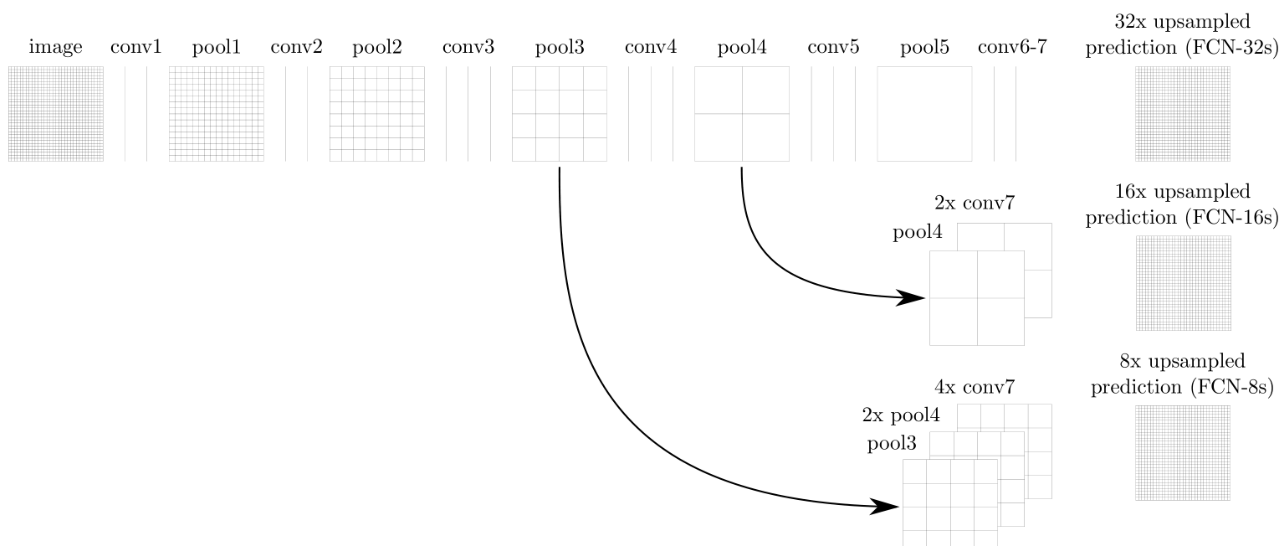


Figure 2.2.2(b) Comparison of 32, 16 and 8 times upsampled results[22]

2.2.3 Advantage and Disadvantage of FCN

Compared with the traditional method of image segmentation using CNN, FCN has advantages in image segmentation. It can accept any size input image, it recovers the category to which each pixel belongs from the abstract features, thereby achieving a classification that extends from image-level classification to pixel-level classification. It does not perform any preprocessing on the input of an entire image. The whole process is to send an entire picture to the end-to-end training in the network.

At the same time, the disadvantages of FCN are also obvious: First, the upsampling results are blurry and smooth, and are not sensitive to the details of each image. The second is the classification of each pixel. The relationship between pixels and pixels is not fully considered. The spatial regularization steps used in the usual pixel-based segmentation methods are ignored and lack spatial consistency[23].

2.3 Random Forest

Decision trees are a type of model used for both classification and regression. Trees answer sequential questions which send us down a certain route of the tree given the answer. The model behaves with “if this than that” conditions ultimately yielding a specific result. This is easy to see with the image below which maps out whether or not to play golf.

Random Forest is an easy-to-use machine learning algorithm that produces good results in most cases even without hyperparameter adjustment. It is also one of the most commonly used algorithms in image segmentation because it is simple and can be used for classification and regression tasks. In this article, I will introduce the application of random forest algorithm in processing medical image processing.

Random Forest is a type of supervised learning algorithm which builds a forest by lots of decision trees and makes it somehow random. The forest is an ensemble of decision trees, most of the time trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result.

To say it in simple words: Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

One advantage of random forest is, that it can be used for both classification and regression problems. I will talk about random forest in classification. Below you can see how a random forest would look like with two trees:

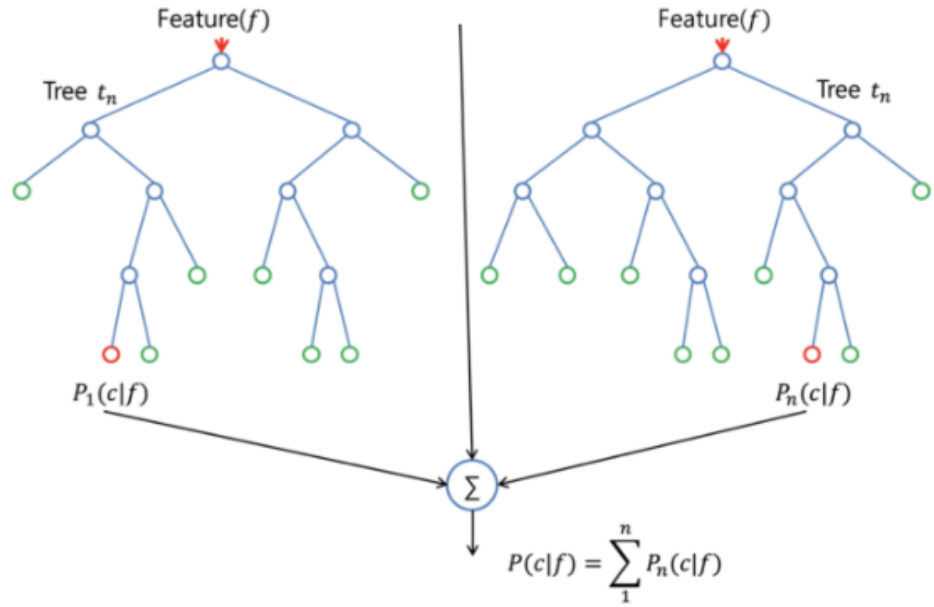


Figure 2.3(a) Random forest with two trees

With a few exceptions, the random forest classifier has all the hyperparameters of the decision tree classifier and all hyperparameters of the bagging classifier to control the collection itself. Instead of building a bagged classifier and passing it to a decision tree classifier, you can use a random forest classifier class, which is more convenient and optimized for decision trees. Note that there is also a random forest regressor for regression tasks[24].

The random forest algorithm gives the model additional randomness as the tree grows. Instead of searching for the best features when splitting the nodes, the best features are searched in a random subset of features. This process produces a wide variety of diversity, which usually leads to better models.

Therefore, when you grow a tree in a random forest, only random subsets for splitting nodes are considered. You can even make the tree more random by using a random threshold on each feature instead of searching for the best possible threshold (as in a normal decision tree) [25].

2.3.1 Random Forest Structure

Here we introduce the working principle of the random forest algorithm. The random forest algorithm has two stages. One is the creation of a random forest, and the other is a random forest classifier created from the first stage to perform the prediction. The entire process is shown below, and it is easy to understand using this diagram.

1. Use the characteristics of the test, plus the rules of each randomly created decision tree to predict the result and store the result of the prediction.
2. Calculate the number of votes for each forecasted goal
3. Consider high voting prediction targets as the final prediction of the random forest algorithm.

The process is easy to understand, but it’s somehow efficient.

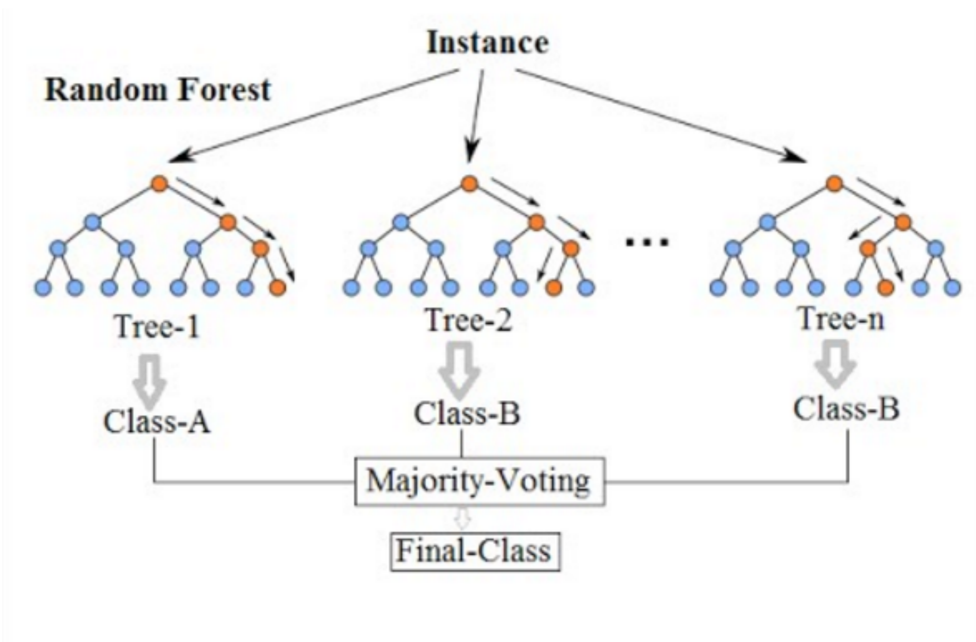


Figure 2.3.1(a) How Random Forest Works[25]

2.3.2 Advantage and Disadvantage of RF

First, we talk about the merits of random forests. It is one of the most accurate learning algorithms in classification algorithms in supervised learning. For many data sets, it generates highly accurate classifiers and can run very efficiently on large databases. At the same time, it can handle thousands of input variables without the need for variable deletion, and can estimate which variables are important in the classification.

With the progress of forest construction, it will produce an inherent unbiased estimate. It has an effective method to estimate the lost data and maintain accuracy when most of the data is lost, and it has a method to balance the error of unbalanced data in the class of the school population.

Calculate prototypes to provide information on the relationship between variables and classifications. It calculates approximations that can be used for clustering, locating outliers, or (by scaling) pairs of situations that give an interesting view of the data, and the above capabilities can be extended to unlabeled data, leading to unsupervised clustering, data views, and outliers Testing. Another thing is that it provides an experimental method for detecting the interaction of variables [26] .

The disadvantage is that random forests have been observed to suit certain datasets with noise classification/regression tasks [27] .

Chapter 3 Methodologies

This chapter deals with the flow of the experimentation and the states the resources used for this dissertation.

3.1 Equipment Used

In this dissertation, all experiments were carried out on the workstation available in the IoT Lab in Nanyang Technological University. The workstation has an Intel ® Xeon® CPU E5-2697 v2 @ 2.7 GHz with 512 GB RAM and is a 64-bit Windows Operating System. Some scripts were written in Matlab R2017a, using a few built-in functions and experiments were run on Matlab on the workstation, and some were in Python 2.7, including some library, including Numpy, PIL, opencv, h5py, ConfigParser , scikit-learn, either directly or using a Windows Remote Desktop connection.

3.2 Flow of Experimentation

First, we download the Nuclei Dataset, from the link below,
https://nucleisegmentationbenchmark.weebly.com/
Nuclear segmentation dataset is a common dataset, mainly a data set of observation data under a digital microscope, which can extract high-quality features for nuclear morphometry and other analyses in computational pathology.

Another dataset is DRIVE(Digital Retinal Images for Vessel Extraction) dataset, we can download it from the link below,
http://www.isi.uu.nl/Research/Databases/DRIVE/
The DRIVE database is also a very important database, the data is mainly retinal images, the purpose of the database is to analyze the blood vessels in the retina. The data set includes 40 photographs with 20 training pictures and 20 test pictures, respectively. Ground truth is the result of manual segmentation by an experienced doctor.

The second step of dissertation was to configure the language environment.

For Matlab, some necessary function should be download.

For Python, the neural network is developed with the Keras library, we refer to the Keras repository for the installation, the following dependencies are needed:

- numpy >= 1.11.1
- PIL >=1.1.7
- opencv >=2.4.10
- h5py >=2.6.0
- ConfigParser >=3.5.0b2
- scikit-learn >= 0.17.1

3.3 Inspection Standards

Sørensen-Dice similarity coefficient and Jaccard similarity coefficient are widely used to represent the similarity coefficient for image segmentation[28]. For these two standard, if the score is large, it means that the similarity is high.

3.3.1 Dice Similarity Coefficient

Sørensen's original formula was intended to be applied to binary data. Given two sets, X and Y, it is defined as Dice Similarity Coefficient(DSC):

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{3}$$

where |X| and |Y| are the cardinalities of the two sets. The Sørensen index equals twice the number of elements common to both sets divided by the sum of the number of elements in each set and the $\cap$ sign stands for set intersection, and |X| stands for the cardinality of the set X, basically the number of elements in the set. This coefficient measures the similarity between sets X and Y. If the two sets are identical (i.e. they contain the same elements), the coefficient is equal to 1.0, while if X and Y have no elements in common, it is equal to 0. Otherwise it is somewhere in between.

3.3.2 Jaccard similarity coefficient

The Jaccard index is also referred to as the intersection point and Jaccard's similarity coefficient (the coefficient originally created by Paul Jaccard). The larger the value, the higher the similarity between the two comparison objects. This is a statistic used to compare the similarity and diversity of sample sets. The Jaccard coefficient measures the degree of similarity between a finite set of samples and is defined as the size of the union of the intersecting parts divided by the sample set[29]:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{4}$$

Chapter 4 Experimentation and Result Discussion

In this chapter, we introduced the experimental results in the order described in the previous chapter. When the results are presented, the dice score and Jaccard similarity coefficients are used to compare the similarity between the predicted image and groundtruth.

The dataset used in the experiment is the Nuclear Segmentation Dataset and DRIVE (Digital Retinal Images for Vessel Extraction) database, which are cell image segmentation and digital retinal image segmentation respectively. The boundary of the cell images is not obvious, and the digital retinal images have more noise, so the two datasets are very convincing in detecting the performance of RF and CNN.

This chapter presents two experiments. The first experiment is medical image segmentation experiment using RF (Random Forest) and CNN (Convolutional Neural Network) in Nuclei Segmentation Dataset. For the Nuclei segmentation dataset, we obtain 60 training images and 10 validation images. And 10 test images.

The second experiment is medical image segmentation experiment using RF (Random Forest) and CNN (Convolutional Neural Network) in DRIVE(Digital Retinal Images for Vessel Extraction) dataset. For the DRIVE data set, we obtained 20 training images, 10 verification images and 20 test images.

4.1 RF and CNN application on Nuclei Segmentation Dataset

In this section, we will discuss how RF and CNN work on a Nuclei segmentation dataset. First we implement the RF segmentation experiment by Marcelo Cicconet et al.(2016)[3], and adjust the parameters to get the best performance. Then the FCN network was used to process the nuclear segmentation dataset.

As we know, dataset images consist primarily of multiple cells and tissue fluids. There is a lighter cell structure around the cells and the tissue fluid has the lightest color. The goal of this section is to use algorithms for image segmentation of cell and tissue fluids. Figure 4.1(a) shows the Nuclei Segmentation Dataset example.

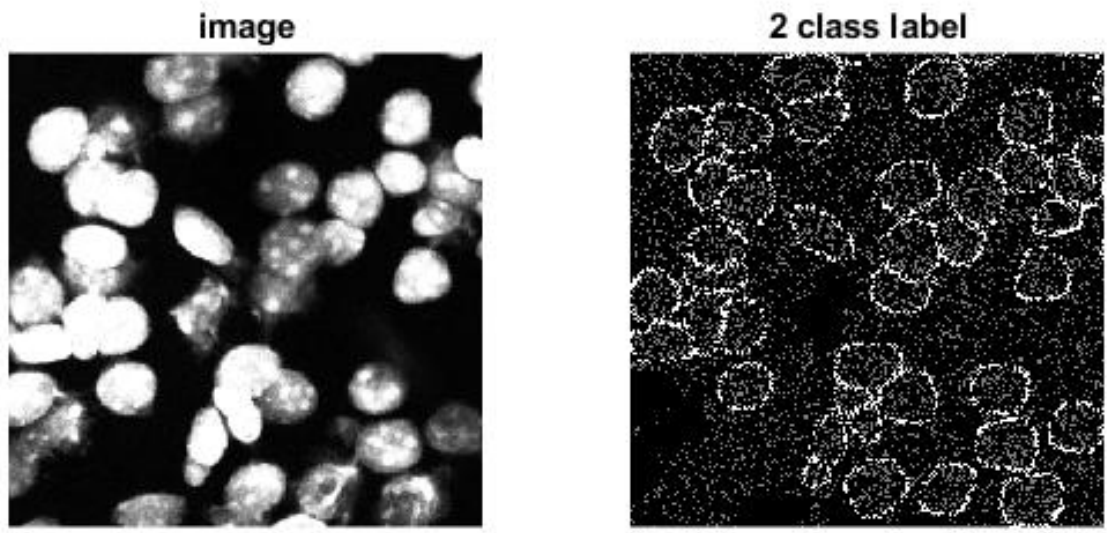


Figure 4.1(a) Nuclei Segmentation Dataset Example

4.1.1 RF method

The code is in matlab, first, set the parameters and use the training image to train the random forest model, then save the model, and use this model to predict the test output. First let me explain some important parameters in the code.

```
nImages = 60;
sigmas = 2;
offsets = 5;
edgeLikFeatOn = true;
probMapELIndex = 2;
circFeaturesOn = true;
probMapCFIndex = 2;
radiiRange = [12 24];
pmSigma = 2;
nTrees = 20;
minLeafSize = 60;
rfModelFolderPath = 'Model';
```

I will explain several important parameters,

labels=[1,2] means that class labels present background and foreground;

sigmas=2 means image features are simply derivatives (up to second order) in different scales;

offsets=5 means offset features from probability maps;

radiiRange=[12 24] means range of radii on which to compute circularity features.

Then we train the treeBag, and treeBag is an ensemble of decision trees for either classification or regression. Bagging stands for bootstrap aggregation. Every tree in the ensemble is grown on an independently drawn bootstrap replica of input data. Observations not included in this replica are "out of bag" for this tree.

```
[treeBag,featImp,oobPredError] = train(ft,lb,nTrees,minLeafSize);
treeBag=TreeBagger(ntrees,rfFeat,rfLbl,'MinLeafSize',minleafsize,
'oobvarimp','on');
```

After we train the treeBag, we save tree bag, and ready to pack model,

```
rfModel.nLayers = nLayers;
rfModel.labels = labels;
rfModel.nLabels = nLabels;
rfModel.sigmas = sigmas;
rfModel.offsets = offsets;
rfModel.edgeLikFeatOn = edgeLikFeatOn;
rfModel.probMapELIndex = probMapELIndex;
```

```
rfModel.circFeaturesOn = circFeaturesOn;
rfModel.probMapCFIndex = probMapCFIndex;
rfModel.radiiRange = radiiRange;
rfModel.pmSigma = pmSigma;
rfModel.nImageFeatures = nImageFeatures;
treeBags = cell(1,nLayers);
```

Now we can save the rfModel. Then, we can use this model to predict the test images. We put test image inside the model layers, and we can get the output.

In figure 4.1(a), OOB error means when number of trees equal to 20, the out-of-bag classification error can be minimum. OOB is the mean prediction error on each training sample, using only the trees that did not have training sample in their bootstrap sample.

In figure 4.1(b) Left image is groundtruth, right image is RF prediction results.

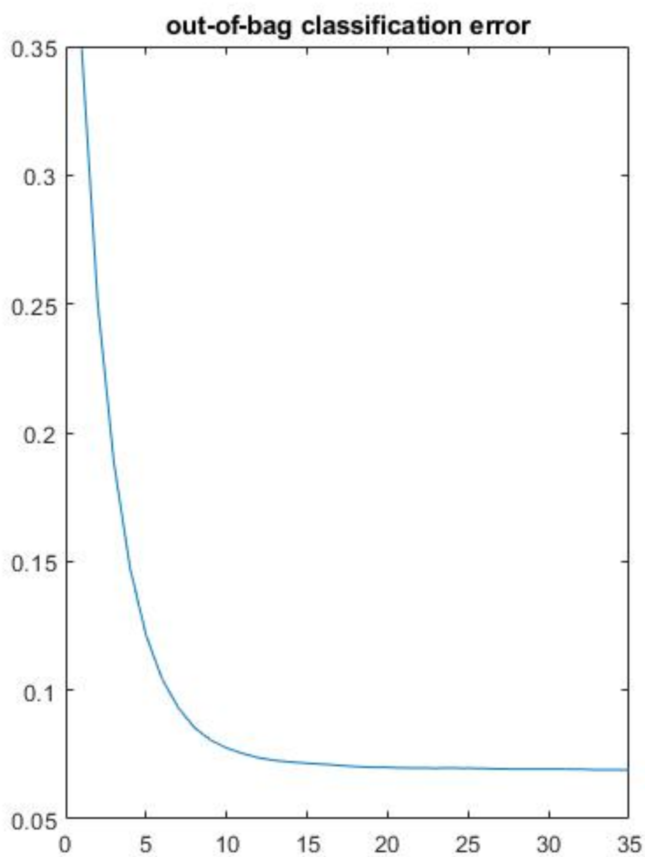


Figure 4.1.1(a) OOB Error

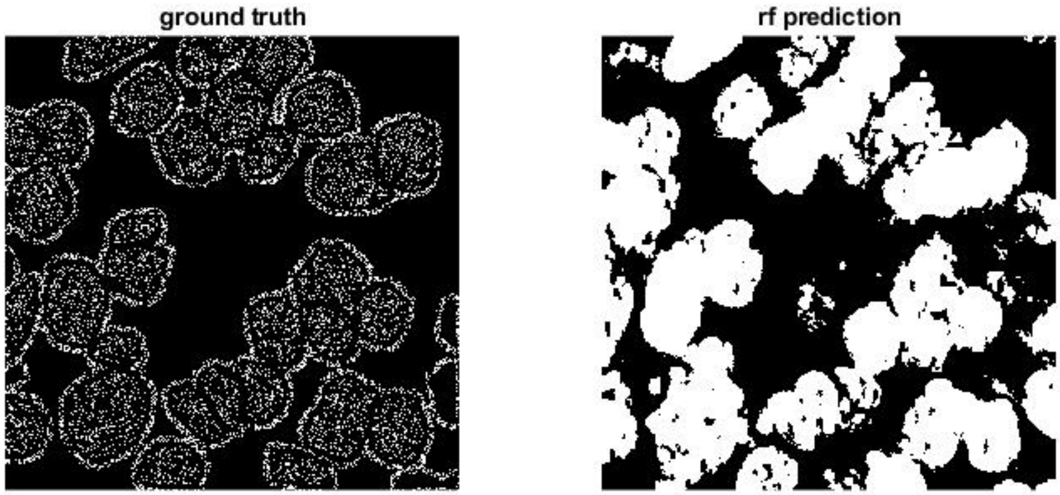


Figure 4.1.1(b) Groundtruth and RF prediction results

Now we change the number of trees, and get the result below. We know the best trees depth number is 40. When the number of trees increase, the score does not increase, but the running time increase, so we set 40 is the best number of trees depths.

Depth of Trees	Train Dice Score	Train Jaccard Score	Test Dice Score	Test Jaccard Score
10	0.6563	0.6851	0.5961	0.5632
20	0.7739	0.7149	0.6327	0.5864
30	0.7812	0.7322	0.6613	0.6177
40	0.7811	0.7317	0.6612	0.6177

Table 4.1.1(a) Change depth of trees result

Now we change the minimum leaf size, and get the result below. We know the best min leafsize is 8.

minLeafSize	Train Dice Score	Train Jaccard Score	Test Dice Score	Test Jaccard Score
5	0.7569	0.7226	0.6314	0.5923
8	0.7812	0.7322	0.6613	0.6177
10	0.7628	0.7198	0.6552	0.6111
15	0.7588	0.7213	0.6615	0.6096

Table 4.1.1(b) Change minLeafSize result

After change trees depth and minLeafSize, we get the best score when trees depth is 40, and minLeafSize is 8.

Average_dice = 0.6613.

Average_Jaccard = 0.6177.

4.1.2 CNN method

The neural network structure is derived from the U-Net architecture. The number of convolution layers is 20, including 4 times downsampling, and 4 times upsampling. The input image is larger than the output image because the input image was mirrored in this paper, which is the same method as FCN(Fully Convolutional Neural Network).

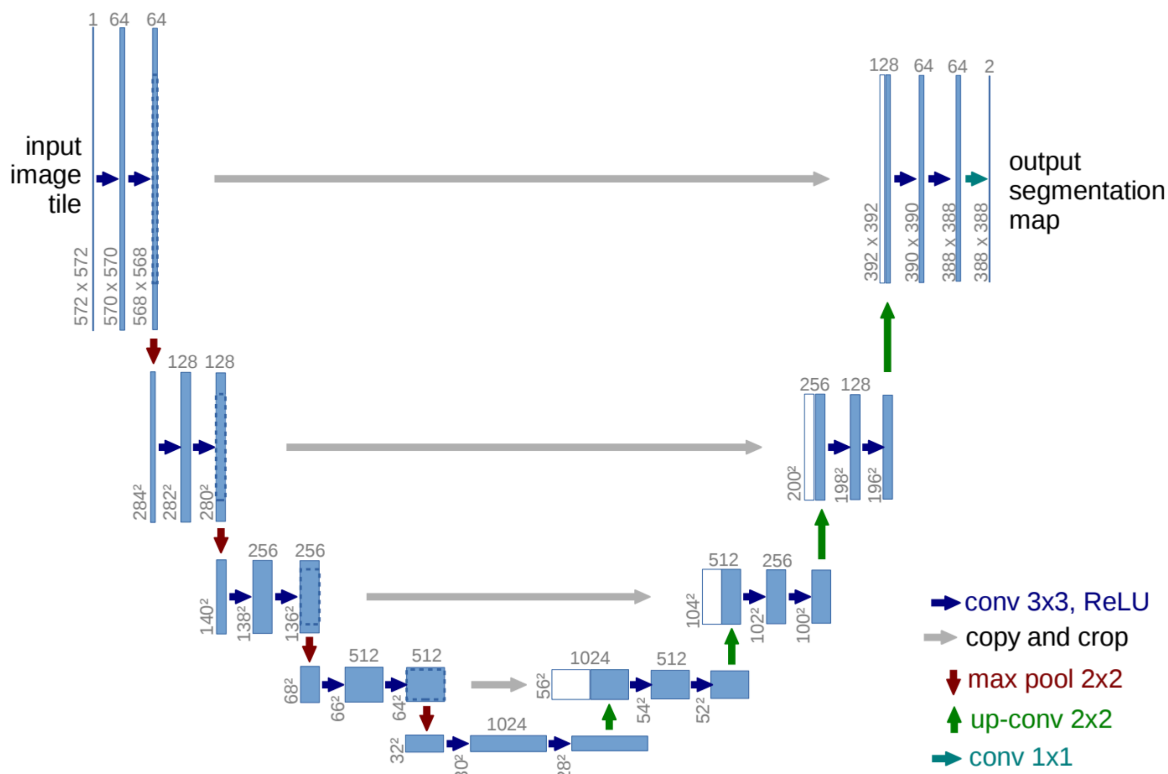


Figure 4.1.2(a) U-net Architecture (example for 32x32 pixels in the lowest resolution).

The network structure is shown in the figure. Blue represents the convolution and activation function. Gray represents replication, red represents downsampling, green represents upsampling and convolution processing, and conv 1*1 represents a 1*1 convolution operation. This is also an end-to-end image, the input is an image and the output is also an image.

The entire network structure looks like a "U" type and is therefore called U-Net. The idea of the entire network is similar for FCN. One difference is that it does not use the VGG model as a pre-training model, because U-Net does binary image segmentation of medical images, and it is not necessary to use ImageNet's pre-training model, and we can freely deepen the U-Net network structure according to their own data sets, such as when dealing with goals with greater receptive fields; another difference is that U-Net used in the shallow feature fusion, it is used a superposition approach, not a summation operation in the FCN. That is, the white module in the figure above is directly superimposed from the blue module on the left (if implemented in Caffe, the u-net is the Concat layer.

Because of the dataset size, before training, we do pre-processing with images. The data format is tif file. We prepare the hdf5 datasets of the DRIVE database using,

```
python prepare_datasets_DRIVE.py
```

We also do pre-processing by defining these functions, which are used to histogram equalization, Contrast Limited Adaptive Histogram Equalization, normalize over the dataset, adjust gamma value respectively.

```
def histo_equalized(imgs):
def clahe_equalized(imgs)
def dataset_normalized(imgs)
def adjust_gamma(imgs, gamma=1.0)
```

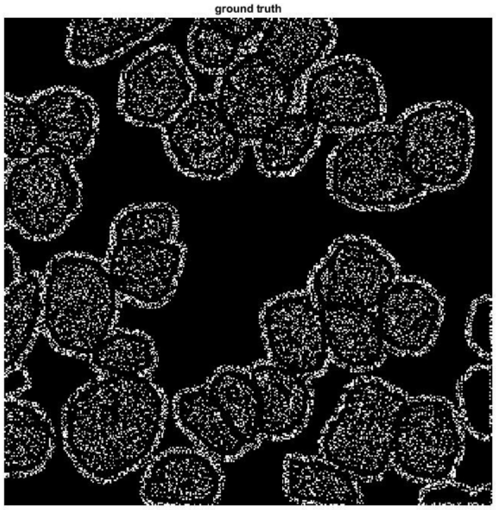
After all these process, we get the train data we want, then we train the input data inside the network.

```
python run_training.py
```

After set the test data paths and and experiment name correctly, we began torun testing by:

```
python run_testing.py
```

The next figure shows the test groundtruth, amd prediction results.



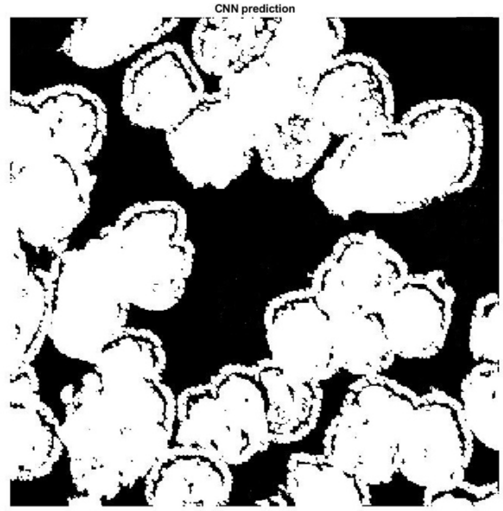


Figure 4.1.2(b) Groundtruth and CNN prediction results

Now we change the learning rate, and get the result below. We know the best learning rate is 0.003.

Learning Rate	Train Dice Score	Train Jaccard Score	Test Dice Score	Test Jaccard Score
0.01	0.8297	0.7654	0.6921	0.6627
0.005	0.8821	0.8126	0.7347	0.7125
0.003	0.8971	0.8257	0.7524	0.7341
0.001	0.8875	0.8178	0.7497	0.7215

Table 4.1.2(a) Change learning rate result

Now we change the batch size, and get the result below. We know the best batch size is 150.

Batch size	Train Dice Score	Train Jaccard Score	Test Dice Score	Test Jaccard Score
50	0.8424	0.7765	0.7144	0.7021
100	0.8751	0.8052	0.7321	0.7154
150	0.8971	0.8257	0.7524	0.7341
200	0.8832	0.8128	0.7463	0.7225

Table 4.1.2(b) Change batch size result

After change learning rate and batch size, we get the best score when learning rate is 0.003, and batch size is 150.

Average_dice = 0.7524.

Average_Jaccard = 0.7341.

4.1.3 RF+CNN method

In this chapter, we use CNN+RF method to do image segmentation.

From paper [31] , we learn that the score map with pixel-wise predictions is used as a feature map which is learned from multimodal MRI training dataset using the FCN. The learned features are then applied to random forests to classify each MRI image voxel into normal brain tissues and different parts of tumor. We get to know that using RF+CNN can generate better performance.

Our method is comprised of four major steps(pre-processing, FCN, feature extraction, random forest classification) that are depicted in Figure4.1.3(a).

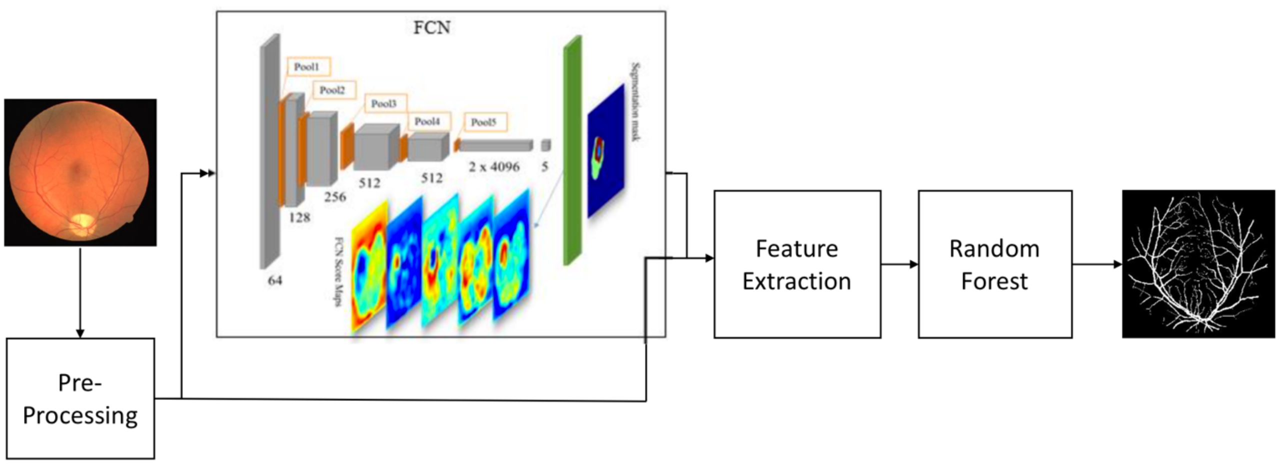


Figure 4.1.3(a) CNN+RF Structure

We get the feature map in penultimate FCN layer, and collect them.

```
from keras import backend as K

# with a Sequential model
get_last_layer_output = K.function([model.layers[0].input],
                                  [model.layers[10].output])
layer_output = get_last_layer_output([x])[0]
```

We use pre-processing images and penultimate feature map to do feature extraction. Feature extraction is a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector. This approach is useful when image sizes are large and a reduced feature representation is required to quickly complete task.

Then, after feature extraction on pre-processing images and penultimate feature map, we use random forest to do classification.

The figure below shows the test groundtruth, amd CNN+RF prediction results.

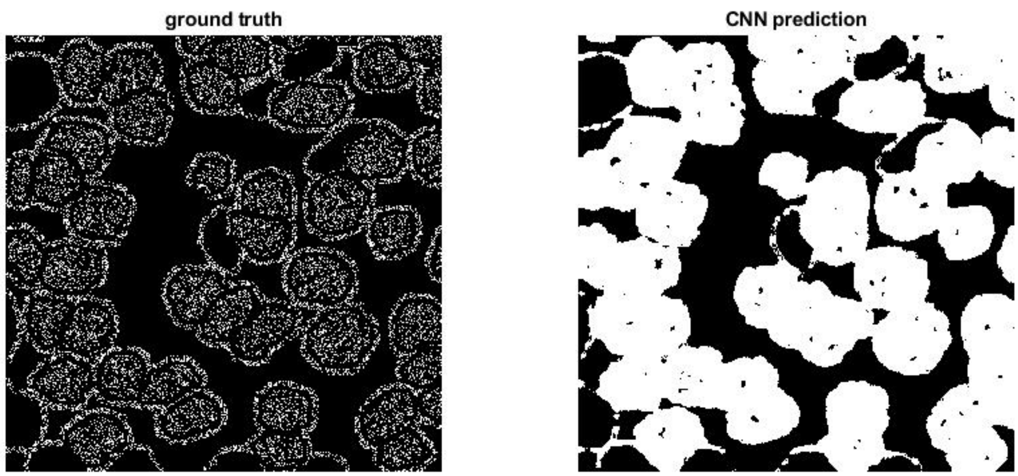


Figure 4.1.3(b) CNN+RF, Groundtruth and prediction results

4.1.4 Compare three methods result

Now we compare the four results. First is the ground truth, second is RF prediction result, third is CNN prediction result, fourth is RF+CNN prediction results. We find CNN is better than RF in dealing with Nuclei segmentation dataset when using seperately, and the combination of CNN and RF can generate even better performance. Figure 4.1.4(a) below shows the four images, and Table 4.1.3(a) shows comparision among previous experiment, RF,CNN,RF+CNN results, we can find that RF+CNN performances better.

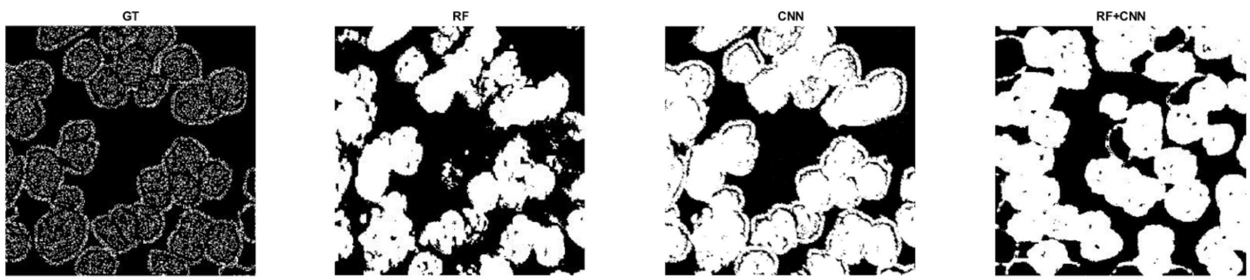


Figure 4.1.4(a) Ground truth, RF result, CNN result, RF+CNN result

	Previous Experiment		My RF Improved		My CNN Method		RF+CNN Method	
	training	testing	training	testing	training	testing	training	testing
Dice	0.7802	0.6557	0.7812	0.6613	0.8971	0.7524	**0.9102**	**0.7866**
Jaccard	0.7298	0.6096	0.7322	0.6173	0.8257	0.7341	**0.8456**	**0.7518**

Table 4.1.3(a) Compare 3 results

4.2 RF and CNN application on DRIVE Dataset

In this section, we will discuss how RF and CNN work on a DRIVE Dataset. First, we implement the previous experiment on CNN by Olaf Ronneberger et al.(2017)[4] , then we attempted to do experiment by using RF.

Before training, the 20 images of the DRIVE training datasets are pre-processed with gray-scale conversion and standardization. The figure of digital retinal image and ground truth show below.

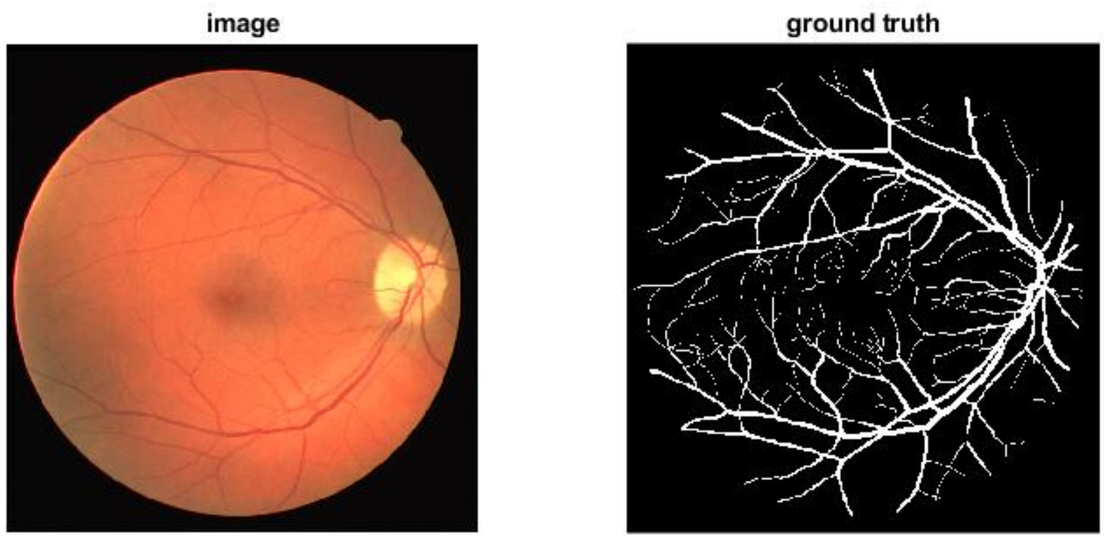


Figure 4.2(a) STARE Dataset images and Ground Truth

4.2.1 RF method

The code is in matlab, first, set the parameters and use the training image to train the random forest model, then save the model, and use this model to predict the test output. First let me explain some important parameters in the code.

```
nImages = 20;
sigmas = 2;
offsets = 5;
edgeLikFeatOn = true;
probMapELIndex = 2;
```

```
circFeaturesOn = true;
probMapCFIndex = 2;
radiiRange = [12 24];
pmSigma = 2;
nTrees = 100;
minLeafSize = 10;
rfModelFolderPath = 'Model';
```

Some parameters are the same as last random forest project, so I will explain several important parameters,

nTrees=100, I train the model many times, and find when the number of trees equal to 100, the out of bag error is minimum, and the training time is also acceptable.

Then we train the treeBag,

```
[treeBag,featImp,oobPredError] = train(ft,lb,nTrees,minLeafSize);
treeBag=TreeBagger(ntrees,rfFeat,rfLbl,'MinLeafSize',minleafsize,
'oobvarimp','on');
```

Now we can save the rfModel, its structure shows next.

```
rfModel =

  struct with fields:

           nLayers: 1
            labels: [1 2]
           nLabels: 2
            sigmas: 2
           offsets: 5
     edgeLikFeatOn: 1
    probMapELIndex: 2
    circFeaturesOn: 1
    probMapCFIndex: 2
        radiiRange: [12 24]
           pmSigma: 2
    nImageFeatures: 8
          treeBags: {[1×1 TreeBagger]}
```

Then, we get rfModel, we can use this model to predict the test images. We put test image inside the model layers, and we can get the output.

In figure 4.2.1(a), OOB error means when number of trees equal to 100, the out-of-bag classification error can be minimum. OOB is the mean prediction error on each training sample, using only the trees that did not have training sample in their bootstrap sample.

In figure 4.2.1(b) Left image is original image, middle image is groundtruth, right image is RF prediction results.

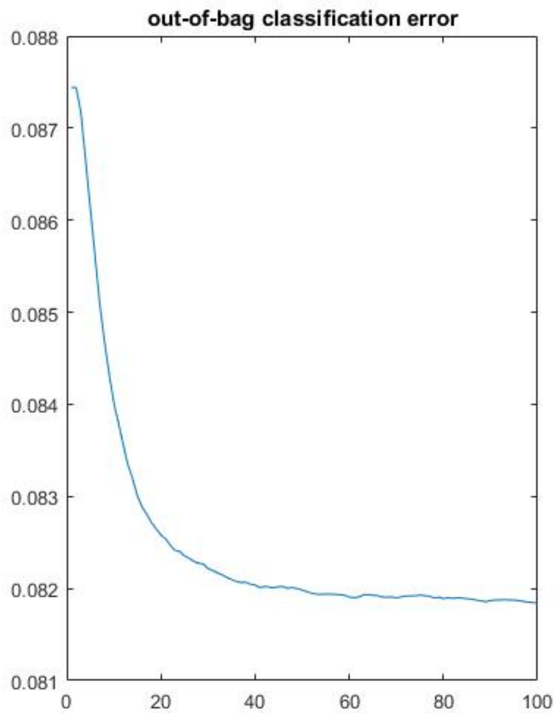


Figure 4.2.1(a) OOB Error

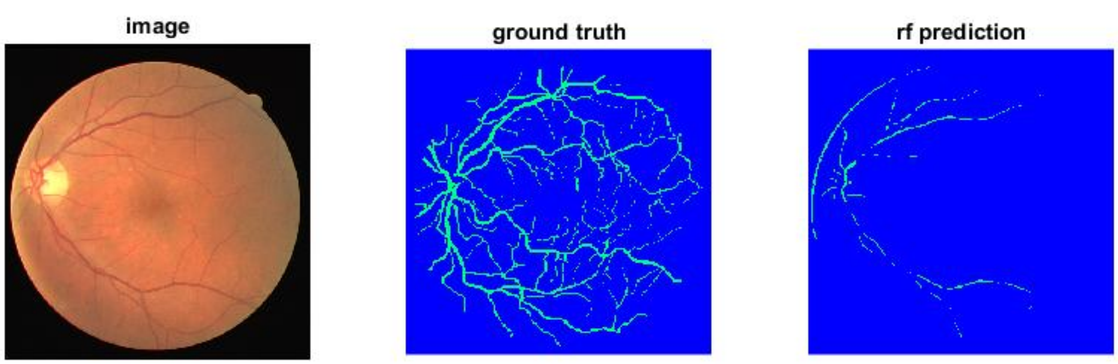


Figure 4.2.1(b) Initial image, Groundtruth and RF prediction results

We can see that the effect of segmentation is not very good. Random forest can only identify thicker retinal blood vessels, but for the retinal blood vessels, the random forest will be integrated with the liquid in the eyeball and cannot be identified. The dice score shows in the table below.

Now we change the number of trees, and get the result below. We know the best trees number is 120. When the number of trees increase, the score does not increase, but the running time increase, so we set 120 is the best number of trees depths.

Depth of Trees	Train Dice Score	Train Jaccard Score	Test Dice Score	Test Jaccard Score
80	0.8641	0.8369	0.8514	0.8098
100	0.8978	0.8674	0.8791	0.8317
120	0.9217	0.8745	0.8912	0.8461
150	0.9183	0.8701	0.8874	0.8322

Table 4.2.1(a) Change depth of trees result

Now we change the minimum leaf size, and get the result below. We know the best min leafsize is 13.

minLeafSize	Train Dice Score	Train Jaccard Score	Test Dice Score	Test Jaccard Score
5	0.8714	0.8573	0.8594	0.8114
10	0.8876	0.8701	0.8823	0.8355
13	0.9217	0.8745	0.8912	0.8461
15	0.9156	0.8678	0.8827	0.8402

Table 4.2.1(b) Change minLeafSize result

After change trees depth and minLeafSize, we get the best score when trees depth is 120, and minLeafSize is 13.

Average_dice = 0.8912.

Average_Jaccard = 0.8461.

4.2.2 CNN method

The library contains an implementation of a convolutional neural network for segmenting blood vessels in the retinal fundus image. This is a binary classification task: Judge the neural network to predict whether each pixel in the fundus image is a blood vessel. Before training, the 20 images of the DRIVE training dataset were preprocessed with the following conversions:

1. Gray-scale conversion
2. Standardization
3. Contrast-limited adaptive histogram equalization (CLAHE)
4. Gamma adjustment

The neural network training is performed on the sub-image (patch) of the preprocessed complete image. Every patch, whose size is 48*48, is obtained by randomly selecting its center in the entire image. In addition, patches that are partially or completely outside the field of view (FOV) are selected. In this way, the neural network learns how to distinguish between the visual field boundaries and the blood vessels.

A set of 190,000 patches was obtained by randomly extracting 9500 patches in each of the 20 DRIVE training images. Although patches overlap, for example, different patches may contain the same portion of the original image, no further data enhancements are performed. The first ninety percents of the data set was used for training (171,000 patches), and the last 10% was used for verification (19,000 patches).

The code is written in Python, it is possible to replicate the experiment on the DRIVE database by following the guidelines below. The neural network is developed with the Keras library, we refer to the Keras repository for the installation.

It is convenient to create HDF5 datasets of the ground truth, masks and images for both training and testing. In the root folder, just run:

```
python prepare_datasets.py
```

After all parameteres have been configured, you can train the neural network with:

```
python run_training.py
```

After set test data paths and and experiment name correctly, we began torun testing by:

```
python run_testing.py
```

The next figure shows the test images, test groundtruth, and prediction results.

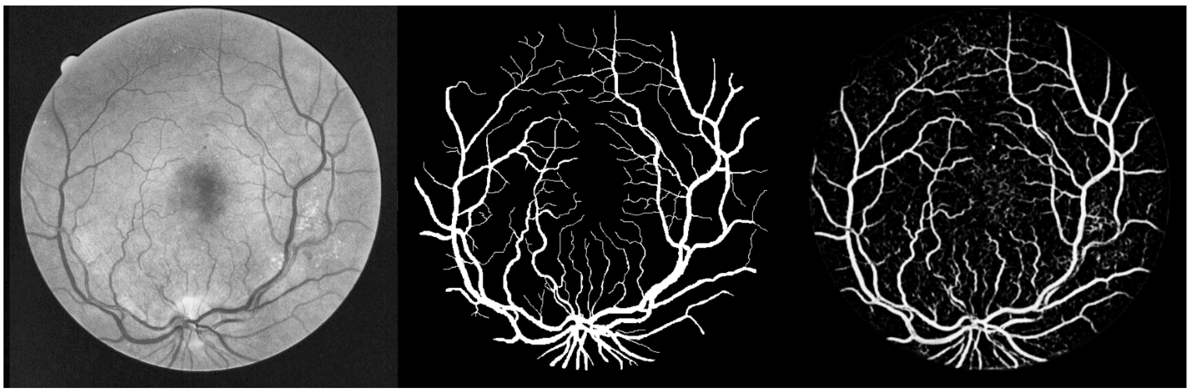


Figure 4.2.2(a) CNN, Initial image, Groundtruth and CNN prediction results

Now we change the learning rate, and get the result below. We know the best learning rate is 0.1.

Learning Rate	Train Dice Score	Train Jaccard Score	Test Dice Score	Test Jaccard Score
0.05	0.9452	0.9007	0.9218	0.8251
0.08	0.9721	0.9206	0.9547	0.8762
0.1	0.9896	0.9275	0.9638	0.9045
0.12	0.9774	0.9108	0.9559	0.8987

Table 4.2.2(a) Change learning rate result

Now we change the batch size, and get the result below. We know the best batch size is 35.

Batch size	Train Dice Score	Train Jaccard Score	Test Dice Score	Test Jaccard Score
25	0.9427	0.9152	0.9233	0.8364
30	0.9766	0.9189	0.9527	0.8754
35	0.9896	0.9275	0.9638	0.9045
40	0.9812	0.9261	0.9589	0.9003

Table 4.2.2(b) Change batch size result

After change learning rate and batch size, we get the best score when learning rate is 0.1, and batch size is 35.

Average_dice = 0.9638.
Average_Jaccard = 0.9045.

4.2.3 RF+CNN method

In this chapter, we use CNN+RF method to do image segmentation.

We use pre-processing images and penultimate feature map to do feature extraction. Feature extraction is a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector. This approach is useful when image sizes are large and a reduced feature representation is required to quickly complete task.

Then, after feature extraction on pre-processing images and penultimate feature map, we use random forest to do classification.

The next figure shows the test groundtruth, amd prediction results.

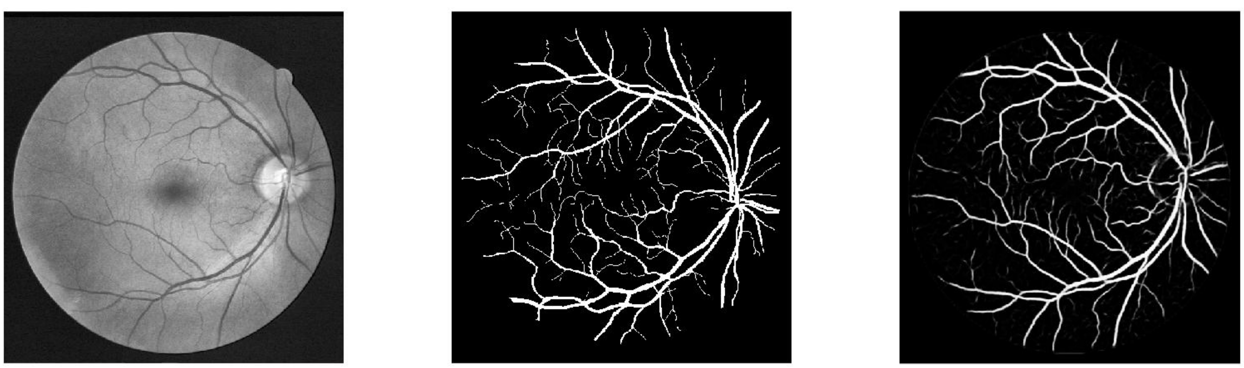


Figure 4.2.3(a) CNN+RF, Initial image, Groundtruth and prediction results

4.2.4 Compare three methods result

Now we compare the four results. First is the ground truth, second is RF prediction result, third is CNN prediction result, fourth is RF+CNN prediction results. We find CNN is better than RF in dealing with DRIVE segmentation dataset when using seperately, and the combination of CNN and RF can generate even better performance. Figure 4.2.4(a) below shows the four images, and Table 4.2.3(a) shows comparision among previous experiment, RF,CNN,RF+CNN results, we can find that RF+CNN performances better.

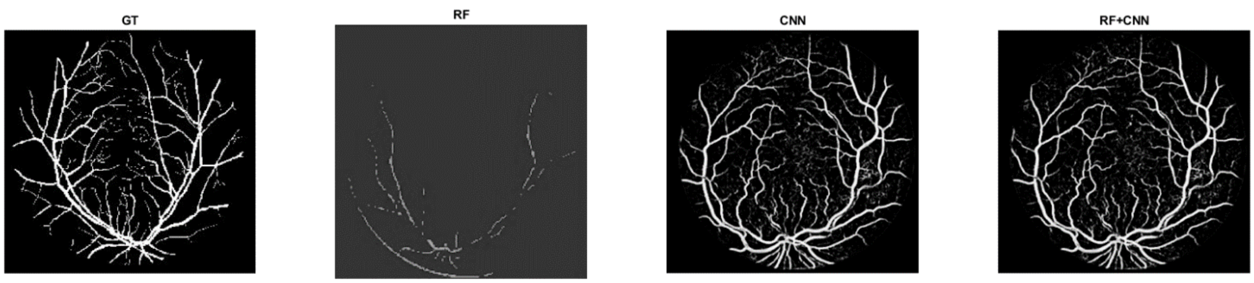


Figure 4.2.4(a) Ground truth, RF result, CNN result, RF+CNN result

	Previous Experiment		My RF Improved		My CNN Method		RF+CNN Method	
	training	testing	training	testing	training	testing	training	testing
Dice	0.9912	0.9725	0.9217	0.8912	0.9896	0.9638	**0.9912**	**0.9726**
Jaccard	0.9321	0.9116	0.8745	0.8461	0.9275	0.9045	**0.9314**	**0.9225**

Table 4.2.3(a) Compare 3 results

Chapter 5 Conclusion and Recommendations

5.1 Conclusion

From the experiments that were conducted in this dissertation, for Nuclei test data, RF get 0.6613 dice score and 0.6173 Jaccard score, while CNN get 0.7524 dice score and 0.7341 Jaccard score, and RF+CNN get 0.7866 dice score and 0.7518 Jaccard score. The RF method splits the image and the effect is not very good. In contrast, the CNN method performs better when segmenting images. I think the main reason is that the selected CNN algorithm is the latest U-Net algorithm, and U-Net has been proved it performing good in many medical image segementation. When we combine RF and CNN method, using CNN to include the region and locally focusing on more accurate segmentation by Random Forest.

For DRIVE test data, RF get 0.8912 dice score and 0.8461 Jaccard score, while CNN get 0.9638 dice score and 0.9045 Jaccard score, and RF+CNN get 0.9726 dice score and 0.9225 Jaccard score. The RF method is not good, I think the main reason is that input data do not pre-processing well, and RF training structure is not so good. For CNN method, the previous result was good, but my CNN method was not that good, because of numbre of training epoches. CNN+RF method can improve a little, because CNN method is already good enough.

Another conclusion is that the preprocessing of medical image processing is very important. For example, in the processing of Digital Retinal Images for Vessel Extraction Dataset, if the data is not preprocessed, the training results obtained by RF and CNN are very poor. Because many interference factors directly affect the operability of input data, such as the occlusion of boundaries by liquids, light problems within the eye, and the sharpness of medical images.

In this disseration, RF and CNN are used to process two sets of medical data sets. The results of CNN experiments can get good accuracy as the prior art methods, thus laying the foundation for future work on these algorithms and similar technologies. The effect of RF algorithm is not very good, but it also provides a reference for similar segmentation methods.

5.2 Recommendations for further research

Medical image processing is a hot topic of research nowadays, and with the continuous advancement of medical level and algorithmic technology, intelligent medical diagnostic technology will certainly become more and more important in the future. It can assist doctors in making correct judgments, even on some issues, it may replace the doctor.

This disseration uses RF and CNN methods to do separation on two sets of medical image datasets, but it can also be used in more application areas, including tumor identification, lung cancer diagnosis.Nowdays, more and more medical image processing challenge competitions were held, which can collect a large number of medical datasets and different application areas, and the algorithm will be more robust.

There is also much improvment in the structure of the algorithm. For example, more than 50 kinds of awesome semantic segmentation methods are proposed in GitHub, including different CNN structures, DenseNet, SegNet, RF, RNN (Recurrent Neural Network), and MRF (Marcov Random Field). Methods, and some of these methods can be used in combination, which may also get better results[30].

References

[1] Criminisi, Antonio, and Jamie Shotton, eds. Decision forests for computer vision and medical image analysis. Springer Science & Business Media, 2013.

[2] Liu, Tianyi, et al. "Implementation of training convolutional neural networks." arXiv preprint arXiv:1506.01195 (2015).

[3] Becker, Carlos, et al. "Learning context cues for synapse segmentation." IEEE Transactions on Medical Imaging 32.10 (2013): 1864-1877.

[4] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[5] Coen, J. F., and F. De la Torre. "Automated Face Analysis for Affective Computing. The 508." (2014).

[6] Abdulkader, Sarah N., Ayman Atia, and Mostafa-Sami M. Mostafa. "Brain computer interfacing: Applications and challenges." Egyptian Informatics Journal 16.2 (2015): 213-230.

[7] Y. J. Zhang, “An Overview of Image and Video Segmentation in the last 40 years”, Proceedings of the 6th International Symposium on Signal Processing and Its Applications, pp. 144-151, 2001.

[8] T. Lindeberg and M. X. Li, "Segmentation and classification of edges using minimum description length approximation and complementary junction cues", Computer Vision and Image Understanding, vol. 67, no.1, 1997.

[9] S. Saleh, N. V. Kalyankar and S. Khamitkar, “Image segmentation by using edge detection”, (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 03, 2010.

[10] M. R. Khokher, A. Ghafoor and A. M. Siddiqui, “Image segmentation using multilevel graph cuts and graph development using fuzzy rule-based system”, IET image processing, 2012.

[11] N. Senthilkumaran and R. Rajesh, “Edge Detection Techniques for Image Segmentation – A Survey of Soft Computing Approaches”, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.

[12] S. Angelina, L. Padma Suresh and S. H. Krishna Veni, “Image Segmentation Based On Genetic Algorithm for Region Growth and Region Merging”, International Conference on Computing, Electronics and Electrical Technologies (ICCEET), 2012.

[13] H. G. Kaganami and Z. Beij, "Region Based Detection versus Edge Detection", IEEE Transactions on Intelligent information hiding and multimedia signal processing, pp. 1217-1221, 2009.

[14] M. Yambal and H. Gupta, “Image Segmentation using Fuzzy C Means Clustering: A survey”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 7, July 2013.

[15] V. K. Dehariya and S. K. Shrivastava, R. C. Jain, “Clustering of Image Data Set Using K-Means and Fuzzy K- Means Algorithms”, International conference on CICN, pp. 386- 391, 2010.

[16] W. X. Kang, Q. Q. Yang and R. R. Liang, “The Comparative Research on Image Segmentation Algorithms”, IEEE Conference on ETCS, pp. 703-707, 2009.

[17] M. Yambal and H. Gupta, “Image Segmentation using Fuzzy C Means Clustering: A survey”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 7, July 2013.

[18] N. Senthilkumaran and R. Rajesh, “Edge Detection Techniques for Image Segmentation – A Survey of Soft Computing Approaches”, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.

[19] Xiong Minghui, et al. CTvirtual bronchoscopy; Imaging method and clinical application [J]. Chinese Medical Joumal, 2000, 113(11):1022- 1025

[20] Huang J, Zhang S, MetaxasD. Efficient MR image reconstruction for compressed MR imaging[J]. Medical Image Analysis, 2011, 15(5): 670-679.

[21] Li, Xiaoshuang, et al. "Voxelized Facial Reconstruction Using Deep Neural Network." Proceedings of Computer Graphics International 2018. ACM, 2018.

[22] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[23] Sundermeyer, Martin. Augmented Autoencoders for Object Orientation Estimation trained on synthetic RGB Images. Diss. Technische Universität München, 2017.

[24] Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest." R news 2.3 (2002): 18-22.

[25] Pal, Mahesh. "Random forest classifier for remote sensing classification." International Journal of Remote Sensing 26.1 (2005): 217-222.

[26] Caruana, Rich, Nikos Karampatziakis, and Ainur Yessenalina. "An empirical evaluation of supervised learning in high dimensions." Proceedings of the 25th international conference on Machine learning. ACM, 2008.

[27] Segal, Mark R. "Machine learning benchmarks and random forest regression." (2004).

[28] Zou, Kelly H., et al. "Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports." Academic radiology 11.2 (2004): 178-189.

[29] Martin, David R., Charless C. Fowlkes, and Jitendra Malik. "Learning to detect natural image boundaries using local brightness, color, and texture cues." IEEE transactions on pattern analysis and machine intelligence 26.5 (2004): 530-549.

[30] Yutake SASAI, Tsuguo OKAMOTO, kenji IMOV, Toru,Automatic Diagnosis of plant disease[J],Journal of TSAM,1999,61(2):119-126.

[31] Soltaninejad, Mohammadreza, et al. "Multimodal MRI brain tumor segmentation using random forests with features learned from fully convolutional neural network." *arXiv preprint arXiv:1704.08134* (2017).