APPLICATION OF DEEP NEURAL NETWORK TO PROSTATE CANCER DIAGNOSIS

By

Feifan Zhang

Senior Thesis in Computer Engineering

University of Illinois at Urbana-Champaign

Advisor: Prof. Gabriel Popescu

May 2019

Abstract

Deep neural network (DNN) has been widely used in biomedical fields to help in understanding diseases. A well-trained DNN can learn inconspicuous features and give a better overall diagnosis than a human. Prostate cancer, one of the most dangerous cancers, has a 5-year survival rate of around 70% when diagnosed in Stage II but this rate decrements to 40% when diagnosed in Stage III. Clearly, we need to find a way help diagnose prostate cancer as early as possible.

In order to build such a diagnosis network, we make use of the distinctive gland distribution in Prostate cells, training a U-Net (one kind of deep convolutional neural network) to do the cell semantic segmentation. This segmentation gives us a 3-label output image which labels the areas of gland, stroma and background in the input image. Based on the segmentation result, the U-net is trained to learn and output the prediction stage of prostate cancer data. This project uses Tensorflow library in Python to build and train the U-net.

Subject Keywords: Prostate Cancer; Semantic Segmentation; U-Net; Gleason Score

Acknowledgments

This senior thesis is supported by Prof. Gabriel Popescu and my graduate advisor Mr. Mikhail Kandel. Prof. Popescu proposes the idea of this project and has directed me during research. Mr. Kandel shares his experience with me, which helps me a lot in setting up the lab environment and using techniques in this research project. I thank them for their sincere and helpful suggestion during this project.

I also thank pathologists from University of Illinois at Chicago for their prostate cancer sample dataset. This is the core of my project. All the members in Prof. Popescu's group also give me a hand during these months which I really appreciate.

Contents

1. Introduction

Prostate cancer is the second most common cancer among men. Every year there are about 174,650 new prostate cancer cases in the United States and about 31,620 patients die from the disease [1]. About 1 man in 9 will be diagnosed with prostate cancer during his lifetime and about 1 man in 41 will die of prostate cancer. Prostate screening is the initial step before prostate cancer diagnosis, consisting of two tests: digital rectal exam (DRE) and prostate-specific antigen (PSA) test [2]. In DRE, the doctor inserts a finger to patient's rectum and tries to find any abnormality of gland tissues. PSA test is a blood sample analysis which observes the potential abnormal level of PSA. The next step is prostate cancer diagnosis which can be done in three ways: ultrasound, collecting sample of prostate biopsy issue [3] and MRI fusion. In our project, we use samples of prostate issue as our dataset, and we will focus on this technique in this thesis. Normal prostate cells have well-organized and differentiated glands [4], [5]. As the prostate cancer cells grow, glands become poorly differentiated and we can use Gleason scores to name the different stages of this abnormal appearance [6]. In our project, we try to build one kind of deep neural network to learn this feature of prostate cancer development. U-Net is a convolutional neural network which is developed for biomedical images [7]. We choose U-Net as the basic structure of our network because it fits better for cell image training [8]. Given a prostate cancer sample image, this U-Net outputs the expected Gleason score. In order to better observe how this network learns the features of the input dataset, we output the semantic segmentation result which is generated by the U-Net during learning.

1.1 Prostate Cancer and Gleason Score.

Gleason score is a grading system for the aggressiveness of prostate cancer [9]. The Gleason score ranges from 1 to 5, with higher score meaning higher aggressiveness. Usually cancer has a Gleason Score of 3 and higher. For each tissue sample, 2 grades are given. The first is the score for the largest abnormal

gland region while the second one is the score for the second largest gland region in the sample input [10]. Score 2 prostate cell is benign and healthy while Score 5 is the most dangerous and has only 30% 5-year survival rate [11]. Given Gleason scores of patients, doctors can give specific treatments to patients, raising their survival rates. However, doctor calculation of Gleason scores is inefficient and sometimes imprecise, it would be helpful to build a machine learning detection network to do this job. In this project, we focus on data samples of Score 2, 3, 4 and 5 since they are the most common cases among patients.

1.2 U-net

Convolutional neural network is a class of deep neural network [4]. It consists of convolutional layers, subsampling layers and fully connected layers based on different designs. With convolution used in the data transmission between layers, it can better collect features of one pixel and its neighbor pixels. U-Net, a class of convolutional neural network, is designed for machine learning in biomedical imaging especially for image segmentation. U-Net is a symmetric network with the same number of upsampling layers and downsampling layers, which gives it a U-shaped structure that is well suited to concatenating input image features during downsampling and recovering more precisely during upsampling layers. Our trained network has 7 upsampling and 7 downsampling layers. Compared to training normal machine learning images, analyzing biomedical images requires higher precision since microscope images have higher density of features. And U-Net performs better in these cases. Since our project aims to output prediction of Gleason scores of input images, we add three fully-connected layers after the normal U-Net output layer. Each neuron in a fully-connected layer is generated from all elements from the previous layer, which reduces the possibility of losing important features. So given an image sample of prostate cancer cell, the revised U-Net outputs a predicted Gleason score, which helps doctors treat patients.

1.3 Semantic Segmentation

Semantic segmentation is a kind of segmentation which is widely used in biomedical imaging. It describes the process of associating each pixel of an image with a class label. In this project, semantic segmentation is used to assign three labels to input image pixels: gland, microscope background and stroma. As mentioned in the previous section, Gleason scores of prostate cells are associated with gland distribution and connection. Healthy prostate cells contain clearly differentiated glands, indicated by low Gleason score, while higher score cells have poorly differentiated glands, indicated by high Gleason score. Doing semantic segmentation could visualize the gland distribution, and the network can learn the features of the distribution to make predictions. The result of semantic segmentation is designed as one of our U-Net outputs, and it gives us a clear visualization during training and prediction process.

2. Literature Review

2.1 Data preprocessing

Given a dataset consisting of prostate cancer cells with different Gleason scores labeled, we have one potential problem: Imbalance of numbers in the data of different groups. Gleason Score 2 means benign cell; as a result, the number of Score 2 samples is much less than that of others in this dataset consisting of cancer patients. However, there exists some techniques that could solve this potential issue. Cheng et al. [12] have trained a rotation-invariant CNN for object detection, meaning that the angle of input images would not affect its output result. This technique could also be applied to our project by randomly rotating images in Score 2 label. That gives us multiple times more input data than the initial one. What's more, Japkowicz [13] gives comparison results for several strategies to solve imbalanced datasets. A strategy called over-sampling is introduced which randomly resamples the small dataset until it has as many samples as other labels. This strategy is proved to be effective in this thesis and gives us a new solution for the imbalance dataset issue. With these two techniques being adopted, the accuracy and reliability of the Score 2 result approaches that of the rest labels.

2.2 Features of Prostate cells

Like the other cancer cells, prostate cancer cells grow gradually and thus can be staged. In each stage, prostate cancer cells show different features among which the gland distribution is the most important one according to Engelbrecht et al. [14]. They claim that the central gland of prostate cancer cells will grow toward the peripheral zone of a speed affected by the patients age, Gleason score of prostate cancer and tumor stage. What they do is first keep environmental parameters the same and then use magnetic resonance imaging (MRI) to find the optimal result of AUC (area under the curve) in all the different cases. While we use sample collecting in our project instead of MRI, this result still holds true since the mode of observation does not affect the growth of prostate cancer cells.

Oto et al. [15] also support this claim. They analyze the different features of benign prostate hyperplasia (an intermediate stage of prostate cancer) and central gland cancer. With MRI being used, apparent diffusion coefficient (ADC) is calculated. Then receiver operating characteristics (ROC) is performed for differentiation between hyperplasia cells and cancer cells. They conclude that ADC differs significantly between hyperplasia cells and cancer cells, which means that they have different rate of diffusion and gland distribution. Based on these experimental results, we can make use of these gland features in our diagnosis project.

2.3 Machine Learning idea in Biomedical image analysis

T.H. Nguyen et al. [16] introduce an approach to do the automatic diagnosis of input prostate cancer, which has a high similarity with what we aim to do. Nguyen combines the idea of machine learning with prostate cancer image analysis. Aiming to extract features from input images [17], they use K-means clustering as a machine learning tool. K-means clustering collects and labels each pixel according to K-means which are initially set. Compared to the deep neural network, K-means puts more focus on each pixel itself than its neighbor pixels, so their performance is a little different [18]. However, the idea of training a network to learn features is the same. The combined use of random forest (RF) classifier and K-means clustering gives out accurate result for most Gleason score groups. Almost all score groups have an AUC value higher than 0.9. That result proves that adopting machine learning in biomedical image analysis is not impossible. The ability of the network to learn trivial features of images enables it to learn most characteristics of input and make precise prediction.

Similarly, machine learning, more specifically deep learning, has been proved successful in biomedical image classification. Siriwukunwattana et al. [19] apply convolutional neural network (CNN) to colon cancer nuclei classification. Giving input dataset containing more than 20,000 annotated nuclei belonging to four labels, the CNN produces a higher F1 score than other traditional classification

solutions. Image segmentation is closely related to image classification since both of them collect trivial features, which gives us confidence adopting deep neural network in our project.

2.4 Performance of U-Net

Of the many neural networks in machine learning, we need to find the optimal one for our project. Different from normal neural networks, the neural network we build should output an image the same size as the input image. At the same time, since our input cell data are 10000 x 10000 pixels, the neural network must process large images without losing any details. Li et al. [20] claim that U-Net would satisfy our requirement. In their Liver tumor segmentation, they compare the performance of normal fully connected neural (FCN) and well-built U-Net. Their conclusion is that U-Net not only saves computational costs but also extracts precious spatial features at the same time [21]. Their input data is in 3D which is extremely big. Using 3D FCN requires huge computational cost which is unrealistic and inefficient [22]. 2D U-Net could greatly reduce the computational cost while still collecting important spatial features. In our project, we have similar trouble dealing with large size input, for which U-Net would be a good choice.

3. Description of Research Methods

3.1 Input dataset

This project focuses on the Prostate Cancer cells. Each prostate cell image contains gland, stroma and background. These images are generated by microscopes, and glass of microscope are also included which is just the label background [23]. Gland is one of the components of prostate cell which has a circle-like round shape and clear linear boundary in healthy situation. Stroma is the tissue between glands with no fix shapes.

Figure 1 An input cell image with size of 10000 x 10000 pixel is cut into 4 patches

Our prostate cancer cell dataset contains 10 samples of Score 2, 90 samples of Score 3 and 61 samples of Score 4 which in total is 161 samples. Each image has size of 10000 x 10000 pixels which requires heavy computational cost. So we first cut each image into four patches, making each patch has size of 2500 x 2500 pixels. Compared to 10000 pixels, this is a more decent size for the network to deal with. With each image being cut into 4 patches, currently the dataset has a larger size of 161 x 4 = 644 input patches. The detailed sample is notated in Figure 1.

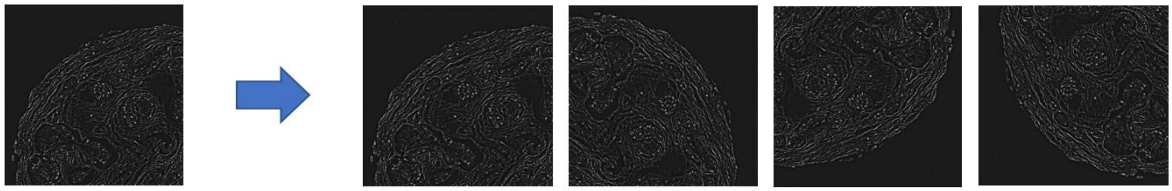

Figure 2 a patch of cell image is rotated for data preprocessing

As we mentioned in Section 2.1, we do the data rotation as the data preprocessing to strengthen the ability of neural network during training. Figure 2 is one example of such rotation. Each cut patch would be rotated and generate 4 output images with different angles being rotated. By doing rotation toward patches, we could obtain more meaningful input patches for the network, which is 644 x 4 = 2576 patches in total.

3.2 Network and its environment

3.2.1 Environment Setting

Consider the large size of input dataset, we need to find an environment could effectively deal with them. Tensorflow, in this case, becomes a good option for our project. Tensorflow is a Machine Learning system which could easily operate large scale data. Data in Tensorflow flow between nodes as data transmission, makes calculation and transmission more efficient. It also offers TensorBoard which records the dataflow and performance of network locally. Another important advantages of using Tensorflow is its compatibility with GPU computing. Hwu [24] claims that GPU, compared to CPU, greatly accelerate data transmission in computer hardware structure since it could arrange data parallelly. Same could be applied to DNN implemented in Tensorflow. With parallel computing at the same time using GPU, time of forward propagation and backward propagation would be exponentially shortened. Our dataset consists of 10 Terabyte Prostate Cancer cell data, which only takes 8 hours for a complete

training using GPU computing. Fast computing gives us more time to optimize the neural network and adjust parameters of network, which potentially improve our training proficiency. We install package Tensorflow-GPU in python. This module changes Tensorflow from default CPU-calculating mode to GPU-calculating mode. The computer in our laboratory used to do the training has 2 GTX-1070 GPU along with 32G RAM.

3.2.2 Network setting

Setting up a U-Net involves parameters of layers, depth of overall structure and forward and backward propagation settings. In this network, there are 3 different kinds of layers: upsampling layer, downsampling layer and fully-connected layer. Basically, upsampling layer share the similar size as downsampling one, consisting of pooling layer, filters and convolutional layers. Filters and pooling layer have sizes of 5 x 5 and 2 x 2 respectively. And the shape of convolutional layers varies according to the output of previous layers. For fully connected layers, the output of them is of size 3, representing the three possible Gleason Scores in this project. We build 7 downsampling layers first, following by 7 upsampling ones and finally we have 3 fully connected layers. Parameters are all randomly initiated values before the first epoch of training.

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta)$$
$$\theta = \theta - v_t \tag{3}$$

During backward propagation of training, we pick F1-loss as the loss function, adopt L2-regularization between each layer and use momentum optimizer (Equation 3) in Tensorflow as the optimizer [25]. F1-loss is not a common loss function but useful in our project since it solves the imbalance between different Gleason Scores. By adding weights when calling F1-loss in Tensorflow, Score 2 samples would

be marked more important though number of them is less than others. They are all initiated and fixed when the network is set up. With Tensorflow being used, we have no need to implement these functions by ourselves but call pre-set functions in Tensorflow library.

4. Description of Research Results

4.1 Semantic Segmentation Result

After the input image being processed by 7 downsampling layers and upsampling layers respectively, we now have the output image which has the same size of input image. Before doing the further prediction using fully connected layers, the neural network would provide the intermediate semantic segmentation result for visualization. This works as a checkpoint for us to observe whether the neural network is indeed learning some patterns from the input image.

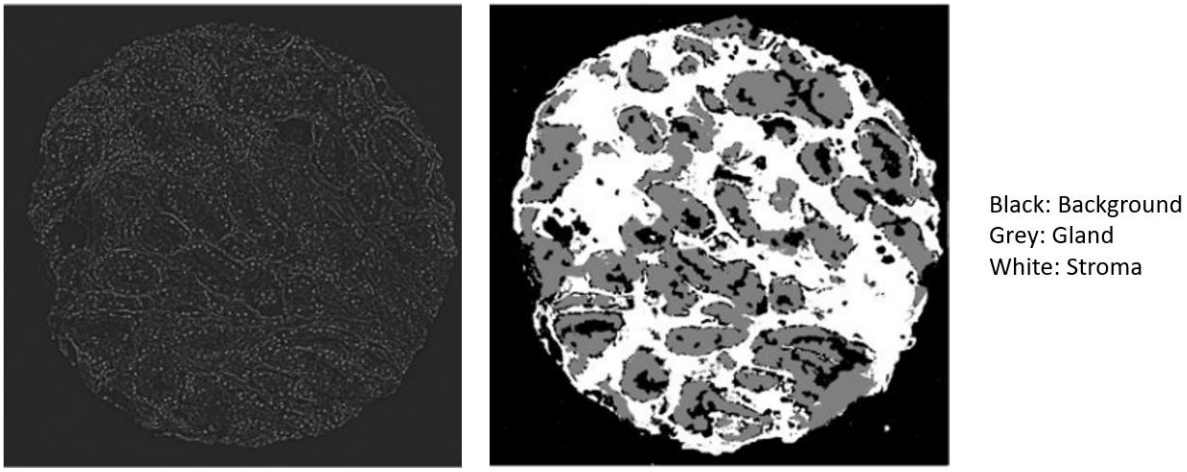

Figure 3 A) is the original input of network. B) is the output semantic segmentation result from the network.

Figure 3 gives us the output of a sample from Gleason Score 3 sample. We compare it with the original manual label to compute the accuracy of segmentation given by the network. The outer black area which is microscope background has nearly 85% precision while the inner area of gland and stroma perform at around 65%, giving us an overall accuracy of 68%. As the Gleason Score of sample increases, its accuracy of inner gland and stroma continues to slightly decrease. This is due to the worse gland distribution in samples with high Gleason Scores. Given gland and stroma with unclear boundaries, it is

much more difficult for a neural network to generalize the pattern of gland distribution. As a result, the accuracy around boundaries are less than the average overall accuracy. In order to solve this issue, we adjust the weights of input patches around boundaries, telling the network that these areas are more important than other places. With this technique, the final precision for Score 2 is 67%, Score 3 is 68% and Score 4 is 65%.

4.2 result of score prediction network

After we get the semantic segmentation network, it gives us a general view of the gland distribution in the Prostate Cancer cell, which is the critical signal of the stage of cancer. As a result, we use fully connected layers to fully make use of all pixels in the segmentation output, predicting the Gleason Score of input image. All the parameters in these fully connected layers are initialized randomly and are changed during training by backward propagation.

Table 1 Confusion matrix of prediction result from U-Net

	Predict Score 2 samples	Predict Score 3 samples	Predict Score 4 samples
Actual Score 2 samples	93	58	9
Actual Score 3 samples	178	963	342
Actual Score 4 samples	33	258	642

Table 1 contains the prediction of total 2576 patches as we know in previous section. From this table, we are able to calculate accuracy, false-positive rate and false-negative rate for all three Gleason Scores. The overall accuracy for correct labeling is around 64%, which is slightly lower than the accuracy of previous semantic segmentation output. Two factors contribute to this result. The first is fully connected layers themselves fail to learn part of the features from segmentation. Such error is unavoidable in Deep Learning which is still acceptable. The second is since it uses segmentation given by the network, errors

in segmentation would potentially affect prediction ability of the network. That means, by optimizing segmentation results prediction results could also be improved.

5. Conclusion

In this project, we applied the theory of Deep Learning to the field of Prostate Cancer. Manually built layer and layer by us, U-Net keeps high accuracy and fits Biomedical images among all the other Deep Convolutional Neural Networks. Segmentation output from U-Net keeps and learn features from the input images. Based on this fact, Gleason Score prediction could be done from these segmentation results. The overall training process could be further optimized by adopting GPU computing and use other tricks such as mini-batch during training. For now, the accuracies still not reach our project goal. We are still seeking other techniques to improve the performance. Making more labeled data is also another feasible way since more different data would make network learn better. Biomedical area, especially disease detection would be a high-demanding field in the future. And we hope by doing this project we could make a small contribute to its progress.

References

[1] American Cancer Society, web page, Available at: https://www.cancer.org/cancer/prostate-cancer.html.

[2] Mayo Clinic, web page, Available at: https://www.mayoclinic.org/diseases-conditions/prostate-cancer/diagnosis-treatment/drc-20353093.

[3] G. Popescu, T. Ikeda, R.R. Dasari, M.S. Feld, “Diffraction phase microscopy for quantifying cell structure and dynamics”, Optics Letters, Vol. 31, pp.775-777, 2006.

[4] Z. Wang, “Spatial light interference microscopy (SLIM)”, Optics Express, Vol. 19, pp. 1016-1026, Jan 2011.

[5] H. Majeed et al., “Breast cancer diagnosis using spatial light interference microscopy”, Journal of Biomedical Optics, Vol. 20, 2015.

[6] G. Popescu, “Quantitative phase imaging of cells and tissues”, McGraw Hill Professional, Mar 2011.

[7] Y. Park, C. Depeursinge, G. Popescu, “Quantitative Phase Imaging in Biomedicine”, Natural Photonics, Vol. 12, No. 10, p. 578, 2018.

[8] Deep Learning, web page, Available at: http://www.deeplearning.net/tutorial/unet.html.

[9] K. Martin, web page, Available at: https://www.prostateconditions.org/about-prostate-conditions/prostate-cancer/newly-diagnosed/gleason-score.

[10] T.H. Nguyen, M.E. Kandel, M. Rubessa, M.B. Wheeler, G. Popescu, “Gradient light interference microscopy for 3D imaging of unlabeled specimens”, Nature Communication, Vol. 8, p. 210, 2017.

[11] M.E. Kandel et al., “Label-free tissue scanner for colorectal cancer screening”, Journal of Biomedical Optics, Vol. 22, 2017.

[12] G. Cheng, P. Zhou, J. Han, “Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images”, IEEE Transactions on Geoscience and Remote Sensing, vol.54, Dec 2016.

[13] N. Japkowicz, “Learning from Imbalanced Data Sets: A Comparison of Various Strategies”, AAAI workshop on learning from imbalanced data sets, 2000.

[14] Marc R. Engelbrecht et al., “Discrimination of Prostate Cancer from Normal Peripheral Zone and Central Gland Tissue by Using Dynamic Contrast-enhanced MR Imaging”, Radiology, Vol.229, No. 1, Oct 2003.

[15] A. Oto et al., “Prostate Cancer: Differentiation of Central Gland Cancer from Benign Prostatic Hyperplasia by Using Diffusion-weighted and Dynamic Contrast-enhanced MR Imaging”, Radiology, Dec 2010.

[16] T.H. Nguyen et al., “Automatic Gleason grading of prostate cancer using quantitative phase imaging and machine learning”, Journal of Biomedical Optics, Vol. 22, No. 3, Mar 2017.

[17] C. Hu, G. Popescu, “Quantitative Phase Imaging (QPI) in Neuroscience”, IEEE Journal of Selected Topics in Quantum Electronics, 2019.

[18] M. Takabayashi, H. Majeed, A. Kajdacsy-Balla, G. Popescu, “Tissue spatial correlation as cancer marker”, Journal of Biomedical Optics, Vol. 24, 2019.

[19] K. Siriwukunwattana, S.E.A Raza, Y. Tsang, D.R.J. Snead, I.A. Cree, N.M. Rajpoot, “Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images.,” IEEE Transactions on Medical Imaging, 35(5), pp. 1196-1206.

[20] X. Li, H. Chen, X. Qi, Q. Dou, C. Fu, P. Heng, “H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes”, IEEE Transactions on Medical Imaging, Vol. 37, Dec 2018.

[21] M.E. Kandel et al., “Three-dimensional intracellular transport in neuron bodies and neurites investigated by label-free dispersion-relation phase spectroscopy”, Cytometry Part A, Vol. 91, 2017.

[22] C. Hu, S. Zhu, L. Gao, G. Popescu, “Endoscopic Diffraction Phase Microscopy”, Optics Letters, Vol. 43, 2018.

[23] M. Takabayashi , H. Majeed , A. Kajdacsy-Balla , G. Popescu, *“*Disorder strength measured by quantitative phase imaging as intrinsic cancer marker in fixed tissue biopsies*.”*, PLOS ONE, Vol. 13, 2018*.*

[24] W. Hwu, “GPU as part of the PC Architecture,” class notes for ECE 408, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 2017.

[25] Tensorflow Deep Learning Library, web page, Available at: http://tflearn.org/optimizers.