Application of Deep Learning in Liver Pathological Image Diagnosis

Yan Wang[1,a], Jin Liu[1] and Bing Yu[1]

[1] *Institute of Computer Application, China Academy of Engineering Physics, Mianyang 621900, China*

Abstract. Hepatopathy is a kind of disease with high incidence, so the progress in the field of liver disease research is highly valued. Medical image, as an important part of medical diagnosis, provides an important basis for doctors to make correct diagnosis. Pathological images play a significant role in clinical application because they can facilitate doctors to clearly observe the degree of lesions and make accurate judgments. As an important component of computer vision, deep learning has been paid more and more attention by researchers. The application of computer aided technology in medical image detection has become an important application of computer vision. In view of this situation, an automatic diagnosis method of liver pathological images based on deep learning method is proposed. We analyze the image features, then design and train the classification model. The final results confirmed that this method can effectively classify the pathological images of liver and has a high accuracy rate.

1 Introduction

In recent years, with the improvement of people's living standards, people pay more and more attention to health, and health has become one of the most concerned topics in today's society. A pathological image of the liver is a sample obtained by a liver biopsy, the image obtained by observing the sample through a microscope. It is widely used in the differential diagnosis of various liver diseases in the medical field. For example, it is often used to clarify the nature and cause of jaundice, to determine the degree and activity of liver injury, to provide a basis for the diagnosis of various viral hepatitis, and to detect early static or compensatory cirrhosis in a timely manner. In particular, the severity of liver fibrosis is an internationally recognized standard. As its accuracy rate can reach more than 80%, it is considered as a better inspection method at present.

During the detection process, percutaneous puncture was conducted under the positioning and guidance of b-ultrasound and CT. The samples were taken and the liver tissue and cell morphology were observed under a microscope. With the increasing number of patients and higher requirements for accurate diagnosis of diseases, Pathological examination, especially the analysis of its microscopic images, has multiplied. As a result, the workload of pathologic laboratory experts increased, and more personnel and equipment were needed to cope with the larger needs of histopathological analysis. A lot of general pathologic image feature recognition and diagnosis takes a lot of doctors' energy to make a diagnosis[1]. In addition, traditional classification methods are limited by physicians' experience, subjective differences in judgment criteria and non-linear features of lesions. The accuracy of test results will be affected to some extent. If some pathological image analysis and diagnosis work can be processed by computer aided system, it is of great significance to improve the current environment of shortage of medical resources[2].

The popular artificial intelligence technology in the field of computer in recent years, especially the deep learning algorithm has been widely used in the field of image processing. There are many applications in the field of medical imaging. In order to eliminate the influence of human factors on accuracy in traditional methods[3]. A classification method of liver biopsy pathological images based on convolution neural network was proposed in this paper. Use the computer to make basic judgments about the image. It is hoped that it can be used as an auxiliary means to help doctors make judgments. On the other hand, it can also reduce the workload of doctors and save the cost of testing for patients[4,5].

The samples used in this study were mainly divided into three categories - healthy, mild lesions and severe lesions. As depicted in Fig.1

| **Healthy** | **Mild lesions** | **Severe lesions** |
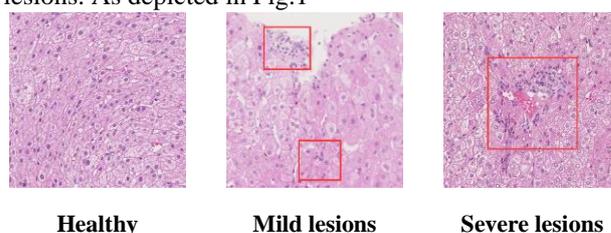
Figure 1 : Pathological Image of Liver

The characteristics of the lesion area vary with the degree of lesion. Therefore, it is difficult to design a feature extraction algorithm that can distinguish between

---

a Corresponding author: wangyanlkj@sina.com

the lesion area and the background area. Deep learning is a good tool for feature extraction, training and classification. As an important part of deep learning, convolutional neural network is a kind of artificial neural network widely used in the world in recent years. This paper conducts experiments based on this artificial neural network.

Convolutional neural network is inspired by the physiological study of cat's visual cortex. And then somebody came up with a convolutional neural network. Yann Lecun was the first to use it successfully in MNIST handwritten numeral recognition task[6], and he has maintained a dominant position on the issue. Therefore, convolutional neural network began to be known to the public. In recent years, convolutional neural network has begun to play its advantages in many fields. For example, medical image recognition and classification[7], speech recognition[8], face recognition[9], behavior judgment[10], emotion prediction[11], etc.

The goal of this study was to classify the pathological images of liver puncture automatically. An experimental method based on convolution neural network is designed. The required classifier is obtained by training the data. The specific arrangement of this article is as follows: The second part introduces experimental algorithm, neural network model and experimental data processing. The third section introduces the experimental results and the analysis of the results. The fourth section is the summary of the experiment.

2 Experiment

2.1 Whitening

The purpose of whitening is to remove the correlation between data. This is a preprocessing step for many algorithms[12]. For example, when training image data, Since there is a correlation between adjacent pixel values in the image, a lot of information is redundant. At this point, the de-correlative operation can adopt the whitening operation. Data whitening must satisfy two conditions: The first is that the correlation between different features is minimal, close to 0. The second is that the variance of all the features is equal. Common whitening operations are PCA whitening and ZCA whitening. ZCA whitening was used in this experiment.

ZCA transformation is obtained by removing dimension reduction from PCA transformation of data. Because you have to pick all the features, and the features are independent of each other.

The formula of ZCA whitening is formula 1:

$$x_{ZCAwhite} = U \frac{x_{rot,i}}{\sqrt{\lambda_i}} \tag{1}$$

In the formula, $x_{rot,i}$ is the i th dimensional eigenvector. In practice, when PCA whitening or ZCA whitening is needed, Sometimes some eigenvalues λ_i

are close to zero in value, So it's going to be divided by something that's close to 0 ,When it divide by $\sqrt{\lambda_i}$. This can cause data overflow or numerical instability. In practice, this process is implemented with a small amount of regularization. And the way you do that is you add a very small constant ε to the eigenvalue before you take the square root and the inverse. So we get formula 2:

$$x_{ZCAwhite} = U \frac{x_{rot,i}}{\sqrt{\lambda_i + \varepsilon}} \tag{2}$$

2.2 Network Design

We use convolutional neural network in this experiment. The convolution layer and the subsample layer appear alternately in the network structure. In the convolution layer, the samples are processed by convolution computation of the convolution kernel to form the characteristic graph. In the subsample layer, the sample data will carry out the subsample operation according to the subsample operator.

Our network structure is inspired by the Alexnet structure for image detection[13]. And modified according to this structure. After modification, it can be more suitable for the experimental data. The specific structure of the network is shown in figure 2.

In the convolution layer of the network model in this paper. The activation function similar to ReLU is adopted[14,15]. The activation function is simple but it works really well. This is because there are usually only a few target-related features in the training depth classification model. Therefore, the sparse model can better mine relevant features and fit the training data.

The activation function is as follows:

$$\phi(x) = \begin{cases} x & if \quad x > 0 \\ 0.1x & if \quad x \le 0 \end{cases} \tag{3}$$

Compared with other activation functions, The activation function has the following advantages in this experiment:

For linear functions, The activation function is more expressive, especially in deep network. For the nonlinear function, as the Gradient of the non-negative interval is constant, the Vanishing Gradient Problem does not exist, and the convergence speed of the model is maintained in a stable state.

In the first two fullconnected layers, we used linear activation functions. The softmax activation function is used in the final layer[16]. Softmax regression model is the extension of logistic regression model on multiple classification problems, in which the number of categories to be classified is greater than two and the categories are mutually exclusive. In this experiment, we need to output the categories of healthy, mild lesions and severe lesions.
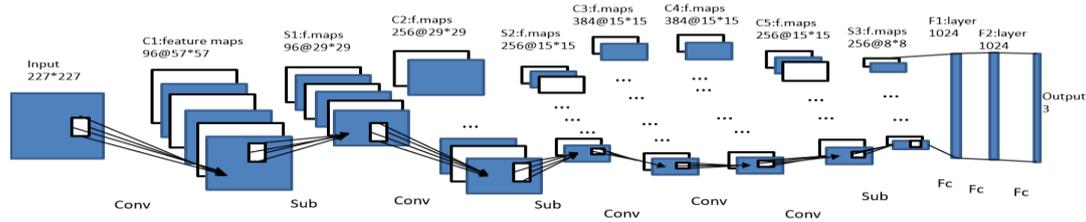
Figure 2: Network Structure

The formula of Softmax activation function is:

$$f(z_j) = \frac{e^{z_j}}{\sum_{i=1}^{n} e^{z_i}} \qquad (4)$$

you can see from the formula that if one z^j is bigger than the other z, This mapping component is going to approach 1, The other mapping components are close to 0，and all input data are normalized.

In order to prevent the over-fitting problem of convolutional neural network, We adopted the Dropout technique[17].

In general, for the same set of training data, after training with different neural networks. The average value of the output can reduce overfitting, and that's how Dropout works. It loses half of a hidden layer of neurons each time, It's the equivalent of training on different neural networks, This reduces the dependency between neurons. Each neuron can't depend on a few other neurons. The neural network can learn more robust features with other neurons.

2.3 Experimental Data

The dataset is the basis for classifying the parameters in the classifier. It has great influence on the final effect of classifier. The quality of datasets directly affects the experimental results. Currently, there is no standard open dataset in the field of liver puncture image processing. The goal of this experiment is to classify the pathological images of liver biopsy automatically. The dataset required for training should also be taken from patients undergoing liver biopsy in hospitals. It is very difficult to collect sample pictures.

In order to have a standard training set and test dataset. We investigated other areas of image processing and analyze their standard dataset. We found several common features of these datasets:

1. There's a lot of image in the dataset, Usually there are over 5000 pictures.
2. In the dataset, each class has a similar number of images.
3. They all set up test dataset that similar to the training set. It is used to detect the accuracy of the trained model.
4. The dataset used for training and samples from the dataset used for testing do not overlap. Therefore, the authenticity of the test results can be guaranteed.

Because of the specificity of the patient population, The original data were provided by the hospital. Among those who went to the hospital for tests, there were more healthy people than sick people. This led to a significant imbalance in the distribution of the number of samples in the three categories. The number of samples from healthy images was much higher than the number of samples from the other two categories.

Through analyzing the open datasets in other image processing areas. Compare it to our data, Rotate and invert our existing image, This produces a large number of pseudo-data samples that are available. Increase the number of available samples without affecting model training. This method is used to expand the number of images in the dataset. This method can also be used to balance the number of datas between several classes.

| **Original Data** | **Rotated image** |
| **Original Data** | **Inverted image** |
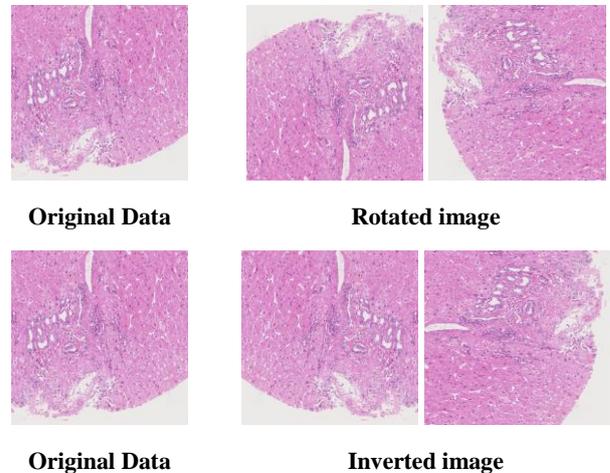
Figure 3 : Raw Data and Processed Data

Affected by the original dataset, The dataset used in this experiment still cannot reach the sample number of the open dataset. The original data set is processed to obtain the new data set. The training set used in this study was 1440 healthy samples, 1224 samples of mild lesions and 1168 samples of severe lesions. The training set contains 3,832 images in total. There were 208 healthy samples in the test set, 172 samples of mild lesions and 164 samples of severe lesions. The test set contains 544 images in total. And there is no overlap between the training set and the sample in the test set.

3 Results and Analysis

The experimental environment of this experiment is windows 7+tensorflow+python3.5. The convolution operation is accelerated by GPU. GPU model used is GTX1050Ti. In this experiment, Limited by equipment performance. I've only done 400 cycles. The learning rate is set to 0.001.

Batch training is adopted, Each batch contains 70 images, The training set is divided into 55 batches. The benefits of training this way are holistic compared to training one by one. It also increases machine utilization. This can significantly save training time compared to training as a whole. When all training samples are finished, one cycle is completed. Run this loop 400 times, a total of 22,000 batches.
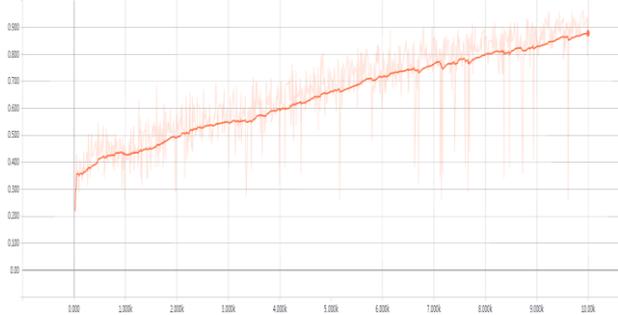
Figure 4 : Changes in Accuracy During Training

The figure 4 is the variation of the accuracy rate in the training diagram recorded by tensorboard. Because tensorboard can only record the training data of the first 10k batches. So we get figure 4.

According to the figure, the accuracy reached 88.47% when 10,000 training batches were completed. Finally, all 22,000 training batches were completed. The accuracy of testing with the test set was 91.54%.

Then, we analyzed the results to obtain the following data:

	Healthy	Mild lesions	Severe lesions
Right	198	152	148
Wrong	10	20	16
Accuracy	95.19%	88.37%	90.24%

Figure 5 : Results Analysis. The number of correct and incorrect classification results in each category.

As you can see from the figure above, The classification accuracy of "healthy" images was the highest among the three categories. The second is severe lesions. The lowest accuracy was mild lesions, only 88.37%. We found that 8 out of 10 misdiagnosed " healthy" data had been incorrectly identified as "mild lesions". 12 of the 16 images in the "severely damaged" category were considered as "slightly damaged. There are 12 images of "severe lesions" that were misdiagnosed as "mild lesions". A total of 20 images of mild lesions were misdiagnosed. Of these, 12 were considered healthy and the other 8 were considered severe lesions. For these misclassified images, we look for some data of the misclassified image running.

The results are as follows:

	Healthy	Mild lesions	Severe lesions
Probabilities	0.469707	0.527655	0.00264

Figure 6 : Misdiagnose "Healthy" as "Mild lesions"

	Healthy	Mild lesions	Severe lesions
Probabilities	0.509083	0.486777	0.00414

Figure 7 : Misdiagnose " Mild lesions" as " Healthy "

	Healthy	Mild lesions	Severe lesions
Probabilities	0.111326	0.433712	0.454962

Figure 8 : Misdiagnose " Mild lesions" as " Severe lesions"

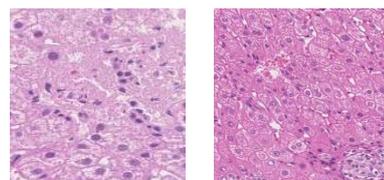|	Healthy	Mild lesions	Severe lesions
Probabilities	0.07233	0.484397	0.443274

Figure 9 : Misdiagnose " Severe lesions" as " Mild lesions"

"Healthy" images are mostly misdiagnosed as "mild lesions". The data above showed that the classifier mainly hesitates between health and mild lesions when diagnosing the health image. The misdiagnosis of severe lesions is similar. The classifier mainly hesitates between severe lesions and minor lesions. For mild lesions data, the classifier hesitates between correct diagnosis and misdiagnosis each time.

We re-looked at health images, mild lesions images, and severe lesions image. And what we found was that in the health images, some of the activity in the liver caused the cells to accumulate, The result is a local image similar to a mild lesions. This results in misjudgment of the classifier. And the misjudgment between mild and severe lesions, It is because the determination of lesions degree is mainly based on the physician's experience. There is no clear line yet. As a result, the difference between the mild lesions image and the health image, the mild lesions image and the severe lesions image in the data set is not obvious. This leads to misdiagnosis.
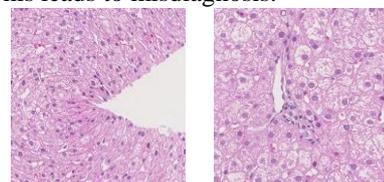
Figure 9 : Healthy Image

Figure 9 : Mild Lesions Image
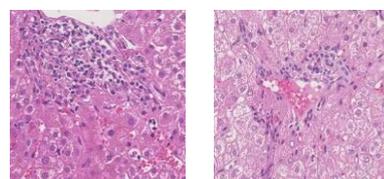
Figure 10 : Severe Lesions Image

You can see it in the image. The dividing line between mild lesions and the other two classes is not obvious. The classifier can only be distinguished according to the characteristics of the samples it learned during training. So that leads to misdiagnosis.

4 Conclusion

In this experiment and related work, A training method based on convolutional neural network is designed to obtain the classifier model. In the process of making datasets, a certain amount of false data is obtained by using various image transformation methods. According to the classification results, the diagnosis of the degree of image lesions is given. This classification of pathological images of liver biopsy achieves the desired effect in the available data.

Compared with traditional diagnostic methods that rely on the subjective experience of physicians. This experiment proves that convolutional neural network can be used to distinguish the degree of lesions, and the accuracy reached 91.54%. It is also proved that this algorithm can share part of pathological image analysis. In this way, the workload of pathologists can be reduced. Contribute to the current situation of medical tension.

References

1. A. Fogelberg, M. Ioffreda, K.F. Helm. J CUTAN MED SURG, **8(2)**:116-121(2004).
2. M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, B. Yener. IEEE REV. BIOMED. ENG, **2**:147-171(2009).
3. M.J. O'Brien, N.M. Keating, S. Elderiny, et al. AJCP, **114(5)**:712(2000).
4. D.C. Howlett, K.J. Drinkwater, D. Lawrence, et al. Radiology, **265(3)**:819,(2012).
5. D.C. Howlett, K.J. Drinkwater, L. David, et al. Radiology, **266(1)**:226-235 (2013).
6. Y. Lécun, L. Bottou, Y. Bengio, et al. Proceedings of the IEEE, **86(11)**:2278-2324(1998).
7. N. Hatipoglu, G. Bilgin. IPTA . IEEE, 1-6(2015).
8. W. Zhao, R. Chellappa, P.J. Phillips, et al. CSUR, **35(4)**:399-458(2003).
9. E. Park, X. Han, T.L. Berg, A.C. Berg. WACV, 1-8(2016).
10. W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, A. Stolcke. The Microsoft 2017 Conversational Speech Recognition System. (2017).
11. T.D. Wilson, D.T. Gilbert. CDPS, **14(3)**:131-134(2010).
12. A. Krizhevsky. *Learning Multiple Layers of Features from Tiny Images*. (2009).
13. A. Krizhevsky, I. Sutskever, G.E. Hinton. NIPS. 1097-1105(2012).
14. J. Redmon, S. Divvala, R. Girshick, A. Farhadi. CVPR. 779-788(2016).
15. M.J. Shafiee, B. Chywl, F. Li, A. Wong. Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video. (2017).
16. W. Liu, Y. Wen, Z. Yu, M. Yang. ICML, 507-516(2016).
17. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. JMLR, **15(1)**:1929-1958(2014).
18. J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, F. F. Li. CVPR. 248-255 (2009).
19. C. Gulcehre, M. Moczulski, M. Denil, Y. Bengio. ICML, 3059-3068 (2016).