

Pathological Image Classification of Cancer Cell Subtypes Based on Deep Learning

Kai Chen^a, Shiyi Tang^b, Quansheng Qin^c, Lianjiang Hu^d and Jing Ye^e

College of Computer Science and Technology, Jilin University, Changchun 130012, China

^achenkai2116@mails.jlu.edu.cn, ^btangsy2116@mails.jlu.edu.cn, ^ctanqs2116@mails.jlu.edu.cn, ^dhulj2116@mails.jlu.edu.cn, ^eyejing2116@mails.jlu.edu.cn

Keywords: Machine Learning, Deep Learning, Convolutional Neural Network, TensorFlow, Cancer Pathology Tissue Image, Digital Image Analysis

Abstract: Due to the accumulation of rich digital pathology images, the need for analysis by machine learning has increased. However, digital pathology image analysis is not the same as general machine learning image recognition. In this paper, we introduce the use of deep network algorithms to analyze pathological images, as well as special problems and possible solutions.

1. The background introduction

In recent years, lung cancer has become one of the highest morbidity and mortality in human society, and with the rapid development of medical technology and the emergence of digital scanners, there has been a sharp increase in the number of digital high-resolution pathological images of lung cancer. Although the pathologist can obtain information from this vast amount of data and make the next diagnosis and assessment of the patient's condition, it will also take a lot of time and labor costs at the same time. In addition, the judgment of the pathological picture, to a great extent, has the subjective consciousness of pathologists, which depends on the accumulation of many years of clinical experience and professional knowledge of pathologists. Different doctors may have different judgments. Therefore, how to quickly and accurately analyze the pathological image becomes an urgent problem to be solved at present.

Machine learning is a discipline dedicated to researching how to use data to improve the performance of the system itself [1]. With the development of technology in recent years, human beings have the ability to collect, store, transmit and analyze information. Development has accumulated a huge amount of data in various fields, and machine learning requires a large amount of data as a support. It has just adapted to the trend of the times and has become the new darling of the times. It has received attention from all walks of life. Therefore, machine learning applications will be applied. The processing of medical pathology images will be a new choice in medical pathology image analysis.

2. Related work

Since the development of machine learning, the Convolution Neural Networks (CNN) is the most widely used in the field of image recognition. The reason why the traditional neural network has withdrawn from the historical stage is that the calculated parameters of the fully connected hidden layer in the middle of its neural network will exponentially increase with the increase of the number of network layers, which will greatly increase the time and cost of the operation and lead to network over-fitting. David H. Hubel and Torsten Wiesel experimented on cats in 1958 and in a paper published in 1962 [2], they pointed out that many neurons in the visual cortex have a small local acceptance domain and they responds only to a limited area of the visual field. The local acceptance domains of different neurons may overlap, putting together to form the entire field of vision. At the same time, it is also suggested that some neurons with large receptive domains only respond to

higher-level patterns made up of lower-level patterns, which provides a strong theoretical basis for convolutional neural networks.

In 1998, LeCun and his partners applied the Back Propagation Algorithm (BP) to the multi-layer neural network and proposed LeNet-5 [3]. The Convolutional Neural Network was formally proposed and used for handwritten digit recognition. In LeNet-5, some nodes of the previous layer network will be used as the unit sub-nodes of the network of the latter layer after the convolution kernel operation. After multiple convolutions and pooling, the parameters are greatly reduced. At the end of the network, two layers of full connection layer are used to classify. Finally, the function of image recognition is realized. In the ILSVRC 2012 (ImageNet Large-Scale Visual Identity Challenge) in 2012, Alex Krizhevsky led his team to use AlexNet [4] they proposed to win the champion of the year with a 15.4% turnover rate. The network had a number of innovative improvements, such as the introduction of Dropout and data enhancement to prevent network overfitting, using ReLu instead of the traditional tanh as an excitation function to introduce nonlinearity and sparsity into the network, becoming a milestone in the development of Convolution Neural Network. In 2013, New York University's Matthew Zeiler and Rob Fergus proposed ZFNet [5] based on AlexNet and won the 2013 ILSVRC championship. ZFNet shortened the step size of the convolution kernel on the basis of AlexNet to prevent extracted features from being blurred. It used the Transport Neural Network, visualized the feature map, and demonstrated that the number of layers of the network of convolutional nerves is directly proportional to its performance. Subsequent VGGNet (VGG16 and VGG19), jointly proposed by the University of Oxford Computer Vision Group and Google DeepMind, made improvements on the basis of AlexNet.

What we used is GoogleLeNet [6], proposed by Google in 2014, which won the ILSVRC championship with an accuracy of 93.3%. At the same time, VGGNet only won the single title of "Classification and Positioning" of the year. GoogleLeNet is better than VGGNet because a 1*1 convolution layer is added behind the convolutional layer in its core Inception unit, and the last fully concatenated layer of the previous Convolutional Neural Network is replaced with a global average pool. It reduces the number of parameters greatly, thereby increasing the depth of the network and preventing network overfitting.

TensorFlow, Google's open-source deep learning framework in 2015, includes Google's exploration of artificial intelligence and successful commercial applications over the past 10 years. Its predecessor was the DistBelief developed by Google's brain team. Tensorflow does some time-consuming matrix operations outside of the Python language by constructing a computational graph, avoiding additional operational overhead. These features make TensorFlow faster and more efficient than previous software platforms for deep learning, greatly improving productivity.

3. The experiment steps

3.1 Data processing and data enhancement

Our raw data is an image of two lung cancer subtypes (adenocarcinoma and squamous cell carcinoma) from the TMA database. This kind of pathological image pixel even reaches hundreds of millions of pixels (Fig.1), which causes the general image processing library to be unable to process the image, and it is impossible to directly put it into the neural network for training data. So in order to train as training data, we used Python's third-party library openslide as a tool for image processing, split the path image file of special SVS format into small images of 512*512 and converted it to jpg format.

However, because a complete pathological cell image is not only dyed cells, but also not-dyed in the picture shown as white tissue fluid. These images will cause a large error in the feature extraction of our training process. So we also need to process the picture, eliminate the white tissue fluid picture. We are based on A. R. Smith presented the conical color space model (Hue, Saturation, Value, HSV) in 1978, using OpenCV to obtain the HSV value of each picture. And set the HSV color range of the cell fluid portion, calculate the white area of each picture. Then remove the picture

containing a large amount of cell fluid with little cell content. Finally we Package the remaining images into TFRecord. So that not only can each small image be analyzed and feature extracted independently, but also we can increase the number of samples, making network training more formal.

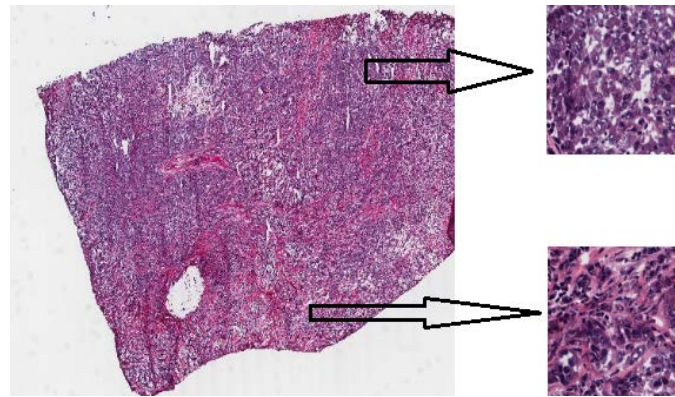


Figure 1. Cutting the SVS image into several small 512*512 images

In the field of pathological image processing, the lack of labeled pathological images is also a major difficulty, which will lead to too little training data and lead to training failure. Therefore, we used the following measures for data enhancement to ensure that the training samples are sufficient in quantity. Firstly, by dividing the high-resolution SVS file, we can increase the number of samples. And also, we can rotate the different angles of the segmented small image and mirror the folded image to increase the number of training samples. We have further increased the amount of our training data by means of pixel movement. However, due to the particularity of the medical image we deal with, we cannot use the operation of saturation, brightness, contrast, sharpening and other color jitters, nor can we use the operation of image deformation, image denoising, and so on. Because these operations will lead to the loss of important features of the image, and thus lead to the failure of network training.

3.2 Migration learning

In the image recognition training based on machine learning, we can completely retrain the training weight data of the whole neural network model, but more often, we use the fine-tuning model to use the existing training model parameters. Only part of the neural network parameters are trained, or only the last layer of neural network parameters are trained. In practice, because of the small amount of data, we often do not retrain a neural network. We usually use the migration learning method. First, we mature the neural network training on a large data set. Then we use this model as a fixed feature extractor, and use the weights trained on the big data set as initial weights of the layers in the new neural network to be applied to specific tasks.

Retraining the neural network parameters of the entire model can indeed achieve higher accuracy to a certain extent, but the time and data required are also geometrically multiplied. In contrast, the time spent on migration learning and the amount of data required are both greatly reduced. Experiments have shown that in the case of too small a data volume, the fine-tuning model can largely avoid over-fitting. At the same time, based on the characteristics of Convolutional Neural Network, the feature learning of the latter layers of CNN pays more attention to the high-level features, that is, semantic features. For different data sets, the semantic features of different data sets analyzed in the latter layers of the neural network are completely different. Therefore, fixing the weight of the neural network in the first few layers, only fine-tuning the parameters of the latter layers, can get good training results.

In general, image training data for deep learning is extremely abundant in general image recognition tasks. But in our lung tumor cell subtype classification task, the biggest problem is the lack of a large amount of training data for reliable experiments. Only a small amount of medical

image data is published for use by researchers. So, in our experiments, migration learning has better performance than fully training the entire neural network parameter weights.

3.3 Training and performance evaluation of different classifiers

The Inception Network is a network with excellent topology. It can perform convolution and pooling of input images in parallel, and all the output results are reorganized into a very deep feature graph. Inception's parallel network uses different size convolution layers (1*1, 3*3, 5*5), so that we can get different feature information of the input image at the same time. By processing this information, we can get better image features than the ordinary convolution network.

The Inception v3 Network is based on the Inception v1. Inception v3 comes from the paper "Rethinking the Inception Architecture for Computer Vision" [7]. The authors propose a series of correction methods that can increase the accuracy and reduce the computational complexity. The Inception v3 Network changes the 5*5 convolution kernel in the Inception v1 Network to two 3*3 convolution kernels to speed up the computation because a 5*5 convolution kernel operation cost is 2.78 times of a 3*3 convolution kernel's, so the superposition of two 3*3 convolution kernels will improve performance. The RMSProp optimizer, the Factorized 7x7 convolution, and the BatchNorm classifier are also used in the Inception v3 Network.

In the experiment, we selected images of two lung cancer subtypes (adenocarcinoma and squamous cell carcinoma) from the TMA database. We selected two hundred of adenocarcinoma and squamous cell carcinoma for training and randomly selected 100 images to make up a collection to test the accuracy of each network. Since the images of the test set were selected from different patients (tumor slides), the algorithm never trained for their entire slide. We think it is an inter-tumor trial set.

We used the normal CNN network, Inception v1 Network and Inception v3 Network to train the same set of data sets (Fig.2, Fig.3) and keep the same training times, learning rate, batch size and other conditions, but the final training results are quite different. Some networks even have poor training results. The results are shown in the figure (Fig.4).

In the end, we got 65% accuracy in the normal neural network, 75% accuracy in the Inception v3 Network, and 83% accuracy in the Inception v1 Network.

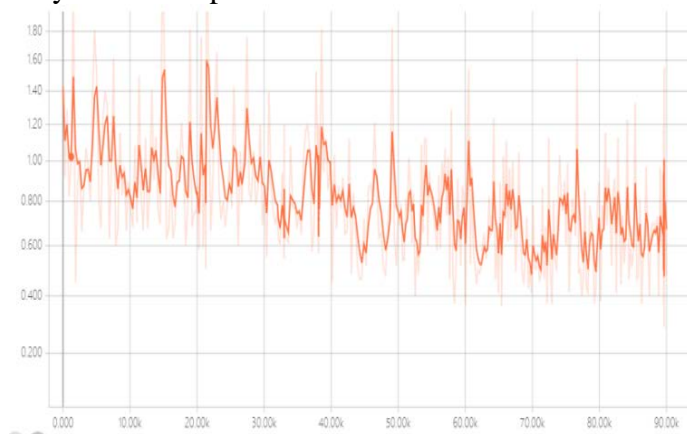


Figure 2. Inception V3 network total loss visualization image

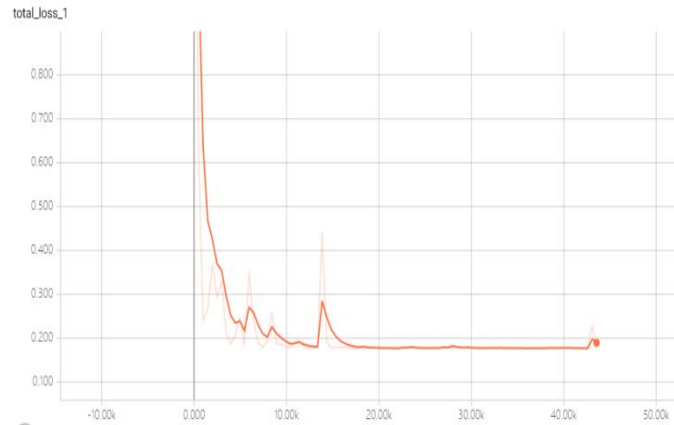


Figure 3. Inception V1 network total loss visualization image

	Least loss	Accuracy
Basic	0.7729	65%
Inception V1	0.197	83%
Inception V3	0.4931	75%

Figure 4. Experimental results

After many experiments, in the migration learning, Inception v1 Network has higher classification accuracy than Inception v3 Network in the environment with only a small amount of training data. After that, we increased the size of batch from 8 to 16. At the same time, we take a big-to-small approach to learning-rate during training, from 0.01 to 0.0001. We have found that the accuracy of training has also improved to some extent. In the end, we get an accuracy of 86%.

4. Summary

In this paper, we mainly introduce the application of machine learning in the classification of tumor cell pathology images. We applied the latest Convolutional Neural Network to the classification experiment of lung tumor cell subtypes. Through the test and comparison of different convolutional neural networks, the Convolutional Neural Network Inception v1 with the highest recognition accuracy was obtained. However, in the experiment, due to the high heterogeneity among different lung cancer patients and the lack of training data, and the final results of the experiment still need to be improved. In the future, we will also be committed to this research, in the hope that the use of tumor pathology image analysis methods based on machine learning can be popularized as soon as possible.

References

- [1] Zhihua Chou.2016.Machine Learning. Building A, xueyan building, Tsinghua University, Beijing: Tsinghua university press. 425pp.
- [2] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex [J]. The Journal of Physiology, 1962, 160 (1): 106-154. Peters J, Schaal S. Natural Actor-Critic [J]. Neurocomputing, 2008, 71 (7-9): 1180-1190.
- [3] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.
- [4] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet Classification with Deep Convolutional Neural Networks [C]. Neural information processing systems, 2012: 1097-1105.
- [5] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks [J]. European conference on computer vision, 2014: 818-833.

[6] Szegedy C, Liu W, Jia Y, et al. going deeper with convolutions [J]. Computer vision and pattern recognition, 2015: 1-9.

[7] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision [J]. Computer vision and pattern recognition, 2016: 2818-2826.