

Cancer Recognition Based on Deep Residual Network

Xinlin Yang, Guanyi Li, Xiaotong Li, Mei Li*

School of Information Engineering, China University of Geosciences, Beijing, China

Abstract

Cancer has become a great threat to people's health due to its high mortality rate. According to the global statistics in 2018, the incidence and mortality of male lung cancer account for the first place in malignant tumors. The recognition of early cancer by medical imaging technology can reduce the mortality of cancer. In image recognition, many image features that are difficult to be detected by human eyes can be found with the help of deep neural network. The application of deep residual network structure to medical image recognition can improve the accuracy of recognition and alleviate the problem of gradient disappearance caused by increasing depth in deep neural network. This paper USES RESnet-50 in the deep residual network to identify the categories of pulmonary nodules. According to the label information on the picture, label the picture of pulmonary nodules to be identified by type. The model was trained through the LIDC-IDRI data set, and the samples on the data set were tested using the trained RESnet-50 network architecture. The final identification accuracy was stable at 0.93. By comparing the recognition accuracy of other networks, the architecture can improve the recognition accuracy. The technology will reduce the work intensity of doctors to reduce the rate of misdiagnosis, and realize the early detection and treatment of cancer.

Keywords

Deep Learning, Convolutional Neural Networks, ResNet-50, Residual Network, Lung Nodules, Cancer Recognition

Received: October 13, 2020 / Accepted: October 28, 2020 / Published online: November 6, 2020

@ 2020 The Authors. Published by American Institute of Science. This Open Access article is under the CC BY license.

<http://creativecommons.org/licenses/by/4.0/>

1. Introduction

For a long time, lung cancer is the cancer with the fastest increase in morbidity and mortality, and the greatest threat to the health and life of the population [1]. Regrettably, there is currently no cure for lung cancer. The effective way is early detection and treatment. Because some lung nodules are very small and difficult to identify, it requires a doctor's rich experience to accurately determine the disease. However, experienced doctors are limited. After reviewing a large number of patient pictures, misjudgments caused by fatigue often occur. This will prevent patients from receiving effective treatment and increase their risk of illness. Using deep learning to process medical images is one of the most active areas of machine learning, especially deep learning has

been widely used in cancer recognition [2].

Deep learning has developed rapidly in the field of image recognition, from LeNet-5 to ZFNet to AlexNet. Subsequently, many advanced algorithms such as VGG and GoogleNet appeared one after another [3]. In 2015, He Kaiming and others proposed ResNet-34, ResNet-50, and ResNet-101. Now ResNet (deep residual network) can reach thousands of layers, and both speed and accuracy have been greatly improved.

The research content of this article is to use ResNet-50 to identify the types of lung nodules in the LIDC-IDRC dataset, and to analyze the pros and cons of the recognition effect by

* Corresponding author
E-mail address: maggieli@cugb.edu.cn (Mei Li)

comparing with LeNet-5 and VGG-16.

2. Material and Method

2.1. Convolutional Neural Network and Residual Network

In traditional neural networks, such as multilayer perceptrons (MLP), the input is usually a feature vector that requires manual design of feature values, but the manually found



Figure 1. Basic architecture of convolutional neural network.

There are five hierarchical structures in CNN network: input layer, convolutional layer, Pooling layer, fully-connect layer and the output layer [4].

Just like the traditional neural network, the model needs to preprocess the input, and the common preprocessing methods include: mean removal, normalization, PCA/SVD dimension reduction, etc

The convolution layer slides the convolution kernel on the image, multiplies and sums the corresponding positions, and combines the three structural ideas of local receptive field, weight sharing (or weight copy), and time or space down sampling to obtain a certain Degree of displacement, scale, and deformation invariance [5].

The role of the pooling layer is to downsample the data, which is mainly used for feature dimensionality reduction, Compress data and number of parameters, reducing overfitting, and improving the fault tolerance of the model [6]. There are two major ways Max Pooling and Average Pooling. Maximum pooling is to define a spatial neighborhood (for example, a 2*2 window), extract the largest element from the modified feature map in the window, and average pooling to define a spatial neighborhood (for example, a 2*2 window), Calculate the average value from the corrected characteristic map in the window. This paper uses a maximum pooling, because it features more prominent, so the effect is more good.

After several times of convolution and pooling, is full connection layer. Before the fully connected layer, if the number of neurons is too large and the learning ability is strong, overfitting may occur. Therefore, the dropout operation can be introduced to randomly delete some neurons in the neural network. [14] such as local normalization (LRN) and data enhancement can also be performed to increase robustness. Each neuron in the fully connected layer is fully connected with all the neurons in the previous layer. It ACTS as a classifier in the

features are not accurate, and sometimes the selected features are biased or even incorrect to the training samples.

If the entire image is used as a feature to train the neural network, although there will be no loss of information, the amount of data is huge, especially when we recognize more complex images, we need enough depth to abstract it.

Convolutional Neural Network (CNN) solves the problem of finding features in neural networks. Figure 1 is a convolutional neural network in the basic structure.

whole convolutional neural network, which can be understood as a simple multi-classification neural network (e.g., BP neural network) through Softmax function to get the final output. So far, the entire model training is completed

In theory, The deeper the network, the more able to extract features, the higher semantic level of abstraction, and the effect should be better, but in fact, simply increasing the depth of the network does not enhance the effect, it will be degenerate, In the traditional convolutional network or fully connected network, information loss is more or less a problem. At the same time, there is the problem of gradient disappearance or gradient explosion, which makes the deep network difficult to train [7]. ResNet addresses these issues to a certain extent. The integrity of the information is protected by detouring the input information directly to the output. The whole network only needs to learn that part of the difference between input and output, which simplifies the learning goal and difficulty, and makes the network can be deep. ResNet is currently in the thousands of level [8].

2.2. Residual Learning Module

The basic unit of ResNet is the residual learning module. In order to solve the difficulty of information loss parameter training, the residual network is designed to fit the residual function between input and output, that is $F(x)=H(x)-x$. Obviously, It is easier to fit $F(X)$ than $H(X)$. [15] Inspired by Highway networks, the designer designed a shortcut connection or "skip connection", which connects another path between input and output to achieve a jump connection at the network layer. In this way, as the network depth increases, information will not be lost, and the convolutional layer only needs to learn the residual between the input and the output.

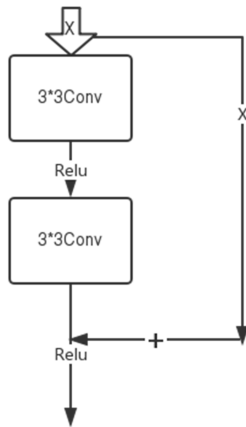


Figure 2. Residual module 1.

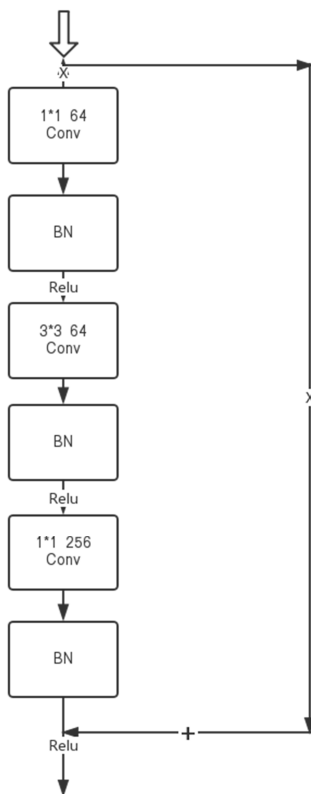


Figure 3. Residual module 2.

The residual module has two structures, the first structure is made by two 3×3 convolution layers and the Relu activation function with shortcut connection between them, Finally, output to the Relu activation function. The process shown in Figure 2.

The second type consists of two 1×1 convolutional layers, a 3×3 convolutional layer sandwiched between two 1×1 convolutional layers, the Batch Normalization (BN) layers and the activation function Relu between them, and finally output, The process shown in Figure 3.

The first structure is to form the residual unit of ResNet-34, the second structure is to form the residual unit of ResNet-50/101/152, so this article uses the second structure of residual unit. Among them, the 1×1 convolutional layer has four functions: 1) Reduce the amount of parameters, thereby greatly reducing the amount of calculation. 2) Flexible realization of dimension upgrade and reduction. 3) Realize cross-channel interaction and information integration. 4) The activation function is increased by adding a 1×1 convolutional layer, which enhances the nonlinear expression of the network. Regularization is a way of regression, which constrains, adjusts or reduces the coefficients towards the direction of zero estimation. In other words, regularization can reduce model complexity and instability during the learning process, thereby avoiding the risk of overfitting.

There are two residual learning units in ResNet-50, one is ID-Block and the other is Conv-Block. Their main flow is composed of three convolutional layers and regularization and activation layers. The difference lies in the convolution step size. The convolution step size of ID-Block is 1, and the convolution step size of Conv-Block is 2 [9].

2.3. ResNet-50

ResNet-50 contains 49 convolutional layers and 1 fully connected layer. There are 5 stages in total. Stage1 is preprocessed by a 7×7 convolutional layer and a 3×3 pooling layer; the remaining stages are shown in Table 1

Table 1. Stage collect.

Stage	Conv-Block	Output channel	Convolution step size	ID-Block	Output channel	Convolution step size
Stage2	1	[64,256]	2	2	[64,256]	1
Stage3	1	[128,512]	2	3	[128,512]	1
Stage4	1	[256,1024]	2	5	[256,1024]	1
Stage5	1	[512,2048]	2	2	[512,2048]	1

Among them, IDBLOCKx2 in the second to fifth stages represents two residual units with no change in size, and CONVBLOCK represents a residual unit with added scale,

each residual unit contains 3 convolutional layers, so there is $1 + 3 \times (3 + 4 + 6 + 3) = 49$ convolutional layers, the structure is shown in Figure 4.

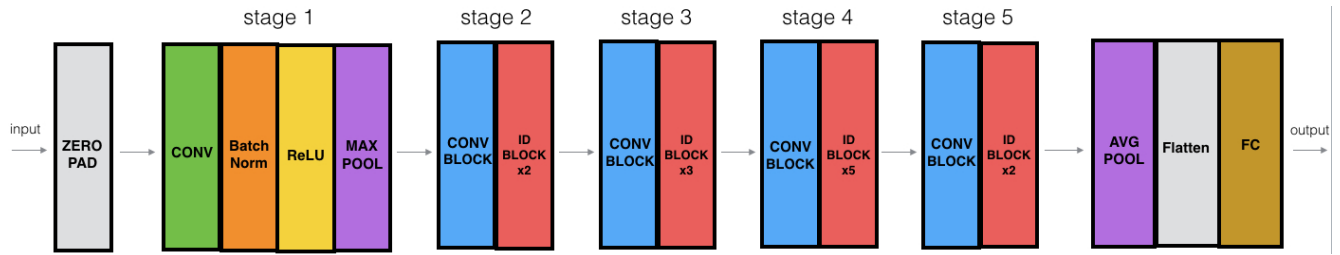


Figure 4. ResNet-50 structure diagram.

2.4. Back Propagation

In this algorithm, we choose the cross-entropy loss function as the objective function, that is,

loss=

`tf.reduce_mean(-tf.reduce_sum(y_data*tf.log(dense),axis=1)), tf.reduce_sum()` is the calculation tensor Specify the sum in the axial direction, that is, the sum of dimension reduction. `tf.reduce_mean()` is to calculate the average value on the specified axis of the tensor, that is, to reduce the dimensionality and average.

Use Adam instead of gradient descent as the optimization algorithm for back propagation, that is, `train_step=tf.train.AdamOptimizer(0.001).minimize(loss)`, and the learning rate is 0.001.

`accuracy=tf.reduce_mean(tf.cast(tf.equal(tf.argmax(y_data, 1),tf.argmax(dense,1)), dtype=tf.float32))` is measured the effect of network learning. Among them, `tf.argmax()` means to return the subscript of the maximum value. `tf.equal()` is used to judge whether two numbers are equal, if yes, it returns True, otherwise it returns False. `tf.cast()` forces True and False into float form, and finally, calculate the average of this round of errors, that is, the degree of agreement between the predicted value and the true value after training.

3. Experiment and Result

3.1. Data Set

The LIDC-IDRI (The Lung Image Database Consortium) data set consists of lung medical image files (such as CT) and corresponding diagnostic nodule annotations. LIDC-IDRI was collected by the National Cancer Institute (National Cancer Institute) in order to study early lung cancer recognition.

In LIDC-IDRI, a total of 1018 research examples are included. For the images in each example, four experienced lung radiologists performed a two-stage diagnosis and annotation. In the first stage, each physician independently diagnosed and marked the location of the nodules. There are three types of marks: 1. Nodules greater than or equal to 3mm; 2. Nodules less than 3mm; 3. Non-nodules greater than or equal to 3mm.

In the second stage, each doctor independently reviewed the labels of the other three doctors and gave their final diagnosis results. Two-stage labeling can ensure that all results are labeled as complete as possible [10], as shown in Table 2.

Table 2. Lung image data information.

Lung nodule dataset	Updated 5/22/2012
Data size	123GB
	CT (computed tomography)
	234897sheets
Image type	DX
	CR (computed radiography)
Number of pictures	243672sheets
Number of patients	1008
Number of series	1010 CT
	280 CR/DX
Number of studies	1205

Dicom file is a standard format for medical images. [11] In addition to the image matrix information, the file also includes some other tags, such as Patient ID, Patient-Name, Study Instance ID (used to uniquely identify each dcm file slice), Image Position (the x, y, z coordinates of the upper left corner of the image in the spatial coordinate system, in millimeters), etc.

Each case has 133 slices, each with a size of 512*512, which is a matrix of 133*512*512, which are sequentially stored in a binary file, and each pixel has a size of 2 bytes (corresponding to the short type). The corresponding structure of each case folder:

LIDC-IDRI-XXX/Study Instance UID/Series Instance UID/* .dcm,* .xml

XXX: from 0000 to 1018;

Study Instance UID: the examination serial number corresponding to each case;

Series Instance ID: the serial instance number corresponding to different checks;

.dcm/ .xml: correspond to each dcm slice and xml annotation file respectively.

Here I choose to use Python to process data in Dicom format, and install the pydicom module through pip or anaconda, which is a library for python specifically to process dicom format files. Through the function `dicom.read_file()`, you can directly get all the information of the dicom file.

3.2. Software and Hardware Environment Installation

Since the processing object of deep learning is tensor, the processing process involves a large number of matrix operations, and GPU (graphics processing unit) is particularly good at a large number of matrix parallel operations. The processing speed of GPU is much faster than CPU, so in deep learning tasks, a good independent graphics card is crucial. In this study, the hardware configuration is shown in Table 3, and the software configuration is shown in Table 4.

Table 3. Hardware configuration.

Configuration object	Model and brand
CPU	Intel i5-5200 2.2G HZ
RAM	4GB
GPU	NVIDIA GeForce 940M
computer	ASUS VM590L

Table 4. Software configuration.

Deep learning framework	tensorflow
Integrated compiler	pycharm
Programming language	python
Python version	Python 3.6
Tensorflow-gpu version	Tensorflow-gpu 1.40
CUDA version	CUDA8.0
CUDNN version	CUDNN6.0
Anaconda version	Anaconda 3.0

Anaconda 3 contains python and integrates various third-party libraries, which is very convenient for the management of installation packages. What needs to be explained here is that anaconda defaults to the latest python version. I chose tensorflow because tensorflow is the mainstream deep learning framework, and there are abundant resources on the Internet to learn from.

CUDA is a parallel computing platform and programming model developed by Nvidia for general-purpose computing on its own GPU (graphics processing unit). [13] CUDA enables developers to use the power of GPU to accelerate the parallelization of computationally intensive applications. The part

The full name of cuDNN is NVIDIA CUDA Deep Neural Network library, which is a GPU-based acceleration library designed by NVIDIA specifically for basic operations in Deep Neural Networks. cuDNN provides highly optimized implementations of standard processes in deep neural networks, such as forward and backward processes of convolution, pooling, normalization, and activation layers. cuDNN is just one of the acceleration libraries in the NVIDIA deep neural network software development kit.

Before installing tensorflow-gpu, you need to install CUDA and cuDNN. It is worth noting that the installed CUDA needs to match the GPU, cuDNN needs to match CUDA, and tensorflow needs to match CUDA and cuDNN. If you have

installed CUDA and the corresponding cuDNN and pip, you only need to enter on the command line: `pip install tensorflow-gpu==1.40` and wait. After installation, you can enter on the command line: `import tensorflow` to detect tensorflow- Whether the gpu is installed successfully.

Before using tensorflow to process data, you need to install some third-party libraries:

cv2-python: is the interface between python and opencv, through which we can call some opencv library functions for image processing.

pydicom: a special python package for processing dicom images.

Pickle: Pickle has two uses in this article. One is to record the read XML tag information and store it as an apk file; the other is to store training parameter weights and network architecture.

3.3. Data Preparation

Image format conversion: The LIDC-IDRC data set is a medical image file in dicom format, but python cannot directly input dicom format image data, so we need to convert it to jpeg or png format.

Extract Label information: Each slice has nodule information marked by 4 experts, including the size of the nodule and the image coordinates of the nodule. [12] By reading the XML information, you can find the specific position of the nodule and prepare for the subsequent image cutting. By reading the XML file and recording its information, save it as an apk file, and then use `pickle.load(f)` to read the information of the apk file.

Image cutting: LIDC-IDRC data set has 124GB, but we are only interested in the position of the nodule, so using the nodule Label information extracted in the previous section, we can cut the nodule of the original image according to the provided nodule coordinate position, As shown in Figure 5:

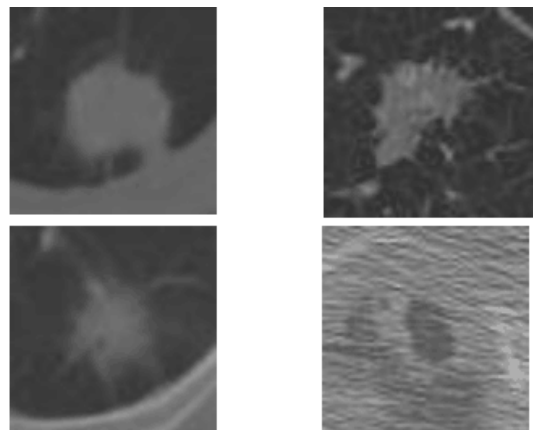


Figure 5. Image of cut out lung nodule.

3.4. Make Training-set and Test-set

We need to make the data in the form of training and test sets as input to the neural network, with the csv file made in the previous step, We first process the image in the first column to grayscale, and then call `numpy.array`, Convert the processed pixel matrix into a numpy array; then, input the corresponding label into a list, and finally encode it as one-hot.

One-hot is the encoding method for making the label of the data set. In this article, I need to encode 1, 2, 3; after one-hot encoding, label becomes a matrix of $\text{num} \times 3$, where num is the number of training samples. If the label of a sample image is '1', it is expressed as 1 0 0, if it is '2', it is expressed as 0 1 0, if it is '3', it is expressed as 0 0 1.

Due to computer memory limited, we can't input all the data into the memory for processing at one time. Instead, we set a batchsize artificially, and then create two placeholders for the input and the label to load the batch, that is, the batch is the training unit. Each training batchsize samples are loaded into the placeholder.

Note that, the batchsize should not be too large or too small. At the beginning, the larger the batchsize, the better the effect. However, when the batchsize reaches a threshold, as the batchsize increases, the effect begins to decrease, so you need to follow your own Try different batchsizes in different situations and find the one that works best.

Finally, we also need to reshape the array after `numpy.array()` to `(numsize, 4096)` where numsize is the number of samples of the training data. Since 32 pixel units are cut at the upper, lower, left, and right sides of the marked nodule, all The size of the cut nodule image is $64 \times 64 = 4096$, all elements in each row of the reshape matrix are the pixel matrix of a sample image, so the new matrix is $\text{numsize} \times 4096$.

An iteration is based on the unit of BatchSize training samples. After the loss is obtained by forward propagation, the weight parameter is updated by back propagation. When we run the entire training data once, it becomes an Epoch. In this experiment, about 100,000 pictures of lung nodules were cut, about 70,000 were used for training, and 30,000 were used for testing. In this training, the BatchSize is set to 16, and the number of training is 7,500. That is, the parameters have been updated for 7,500 iterations and 2 rounds of Epoch. When the number of training exceeds 5000 times and the accuracy exceeds 0.90, save the entire network architecture and network parameters and stop training to prepare for subsequent experimental tests.

By comparison of the results of different network architectures in different amounts of data, it was found in the 60,000 or less sheets of samples, with an increase in the

amount of data, the overall recognition accuracy rising trend; in the same sample size, Obviously, ResNet-50 has the best recognition effect.

4. Discussion

In summary, although the final recognition accuracy of ResNet-50 is the best among these three network structures, there is still much room for improvement in accuracy. First, because this experiment does not filter image after cutting, so do not rule out on the final results may have an impact; secondly, there is no training samples were further optimized and screening; furthermore, Compared with saving as a lossless format such as BMP or PNG, the cut image is saved as a JPEG format with a higher compression rate, which will result in compressed data volume, thereby affecting the accuracy of recognition.

5. Conclusion

In order to improve the recognition accuracy of lung cancer, this paper uses the deep residual neural network ResNet-50 to identify lung nodule image information. Comparing ResNet-50 with LeNet-5 structure and VGG-16 structure, the analysis shows that the deep residual network ResNet-50 achieves better cancer recognition effect, and the accuracy rate is stable at 0.93. This technology can be applied to the early diagnosis of lung cancer, reducing the work intensity of doctors, improving the accuracy of diagnosis, and realizing early detection and treatment of cancer.

In the experiment, it is found that the average effect of the experiment with a large sample size is better than that of the experiment with a small sample size. Therefore, you can try data enhancement, that is, sample rotation and other operations to expand the sample size; you can try to filter the image, or you can Start from the perspective of optimizing the sample, that is, select the positive sample with the farthest distance from the origin sample and the negative sample with the closest distance as the training sample, remove the samples with poor training effect, and improve the training efficiency; from the perspective of network structure, increasing the network width is the same Help improve the recognition accuracy.

Acknowledgements

This paper is supported by State Key Laboratory of Computer Architecture (ICT, CAS) under Grant No. CARCH 201806.

References

- [1] Zhang Chunyun, Zhai Yanping, Chen Lu, Tong Da, & Li Shiyue. (2016). Research on the correlation between human papillomavirus and non-small cell lung cancer in Guangdong. *International Journal of Respiration*, 36 (22), 1699-1702.
- [2] Zhang Tingting, Qu Ning, & Zheng Pu. (0). Application of machine learning in the diagnosis and treatment of thyroid tumors. *Chinese Journal of Cancer*, 27 (12).
- [3] Zhang Xiaoyu. (2019). The comparison of mixed reality target recognition technology and yolo algorithm. *Microprocessor*, 040 (005), 53-59.
- [4] Yang Xihui. (2017). Collaborative filtering model based on fusion convolutional neural network. *Software Guide*, 016 (012), 44-48.
- [5] Sun Yanfeng, Qi Guanglei, Hu Yongli, & Zhao Lu. (2015). Deep convolutional neural network recognition algorithm based on improved fisher criterion. *Journal of Beijing University of Technology* (06), 835-841.
- [6] Ma Chengxian, & You Yachen. (2019). Discussion on crack location identification and target detection methods of railway bridges. *Railway Survey*, 045 (005), 136-141.
- [7] Zhang Zhuping. (0). Research on high-accuracy pedestrian detection algorithms. (Doctoral dissertation).
- [8] Xie Bo. (0). Research on intrusion detection model based on ResNet. (Doctoral dissertation).
- [9] Wang Chenglong, Huang Yufeng, Zhuang Xuemin, & Xie Heng. (2018). Research on potato malformation detection method based on deep learning. *Journal of Huizhou University*, 38 (003), 61-66.
- [10] Zhao Pengfei. (0). Research on the detection and classification of small lung nodules based on convolutional neural network. (Doctoral dissertation).
- [11] Gu Jingping. (2017). Research on data Processing Technology based on CT image set surface Modeling (Master's thesis, Nanjing Normal University).
- [12] Wang Houhua. (2018). Study on the Detection Algorithm of pulmonary nodules in medical Images based on Deep learning (Master's thesis, Xidian University).
- [13] Lei Zhenyu, Wang Jianxiong, Lin Xiaosen, Xu Jia & He Li. (2013). An Introduction to Heterogeneous computing system based on GPU massive Parallel Computing. *Modern Cinema Technology* (12), 24-29.
- [14] Tang Xueting. (2009). Environment Recognition of Mobile Robots based on Neural Networks (Master's thesis, Xiamen University).
- [15] Wang Qihong. (2017). Research on single Low-illumination Image Enhancement Algorithm based on image fusion (Master's thesis, Xiamen University).