International Core Journal of Engineering Volume 8 Issue 6, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0073

Researsh on the Application of Machine Learing in Cancer Assisted Diagnosis

Zijun Guo, Zhen Chen, and Zongru Yang

North China University of Science and Technology, Guangzhou 063000, China

Abstract

In recent years, with the development of image processing and artificial intelligence technology, the application of medical image big data-based analysis methods to assist doctors in decision-making and solve difficult problems in clinical practice has become a research hotspot. This project intends to establish an intelligent recognition system of lung cancer based on the image data of lung cancer patients by using convolutional neural network method. Using Python as a tool, combined with the collected medical image data, the feature extraction of lung cancer image is realized by using deep learning framework, and the unknown image is recognized by clustering learning to achieve higher classification accuracy. Finally,a software is developed to assist doctors in accurate diagnosis and improve the accuracy of diagnosis.

Keywords

Image Processing; Convolutional Neural Network; Python; Lung Cancer; Clustering Learning.

1. Introduction

Due to the uncertainty of pathogenesis and occult disease, early lung cancer is difficult to be detected. About 75% of patients have advanced lung cancer when they are treated, so the 5-year survival rate is lower than 20% [1]. This is very inconsistent with the social advocacy and medical advice of "early detection, early diagnosis, early treatment", therefore, to improve the survival rate of patients must pay attention to early diagnosis and standardized treatment. Lung CT[2] has high resolution in lung tomography plain scan (LUNG CT), and can observe the range, size and relationship between lesions and surrounding tissues from multiple angles and in multiple directions. In recent years, it has become an important method widely used in the early detection of lung cancer [3]. Detection of cancer regions in lung cancer pathological images is the basis for assisting doctors in pathological diagnosis, and its goal is to automatically detect and locate cancer regions in pathological section images by using computer technology [4]. Compared with traditional pathological image analysis methods, pathological image analysis based on deep learning has achieved better results in recent years, which can identify cancerous sections in lung biopsy and provide rough location of cancer region.

In recent years, deep convolutional neural network has achieved great success in the field of image recognition [5], and has been gradually introduced into the fields of medical image recognition and segmentation. For lung cancer images, the image processing algorithm is based on Python, while combining OpenCV open source library, using Numpy and SCIPY image recognition library for image preprocessing image segmentation and a series of processing processes. The large-capacity volume set network is applied to the bottom-up regional proposal for the localization and segmentation of image objects [6]. The image processing technologies applied include image enhancement, image denoising, image segmentation, morphological processing and inversion. Therefore, an improved deep convolutional neural network model was used to achieve automatic classification of lung cancer pathological images. At the same time, the learning method of data

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0073

enhancement and transfer is used to effectively avoid the over-fitting problem that is prone to occur when the deep learning model is limited by sample size. [7].

2. Image Preprocessing

2.1 First, the CT Images in DCM Format Are Converted

The image information of the original data set is stored in DCM format, but most of the images we input into the network as training data are in JPG or PNG format. Therefore, in order to facilitate the later training, we first need to convert the original image into JPG format or PNG format for storage, which is stored in JPG format here. Use Python code to iterate through all files in the folder, separating their file names and suffix. DCM into the list, and then loop through the list to convert all DCM files into JPG files.

2.2 Unified Picture Scale

In target detection, images in convolutional network training usually require fixed size images (such as 416*416). In order to facilitate subsequent image cutting operations, image scale needs to be unified, and the unified image scale only needs to be scaled to the specified scale through the resize() method in OpenCV.

3. Image Cutting, Cutting Pulmonary Nodules

3.1 Processing Data

Firstly, liDC data set was downloaded, and the PYLIDC library was used to process the LIDC data set. Then, Python code was used to cut all nodules with a probability of malignancies greater than or equal to 3 in the data set. After that, the data is cleaned, and the image is rotated clockwise every 10° to increase the amount of data. After that, the function method in OpenCV in Python is used to corrode the image. After that, the image is converted into digital matrix into NPARray and put into the convolutional neural network.

3.2 Image after Cutting

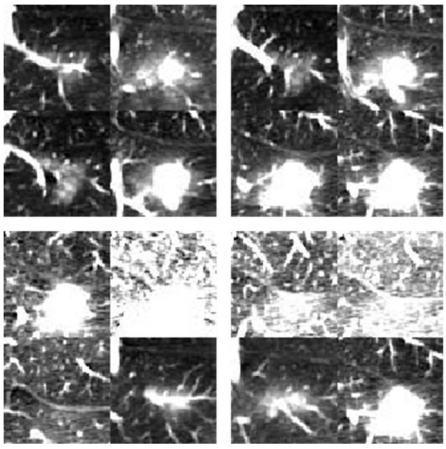

Fig.1 Image after lung node and lobe cutting

4. Gan Adversarial Network is Enhanced

GAN is to train a generation model whose goal is to generate a distribution that is consistent with the distribution of real data. The distribution of real data is often unknown. Therefore, a classifier is needed to generate decision boundaries to fit possible distribution shapes with the generator. This is not technically complete, so you need a lot of tricks to make sure you can train a working model.

The most important discovery is that data augmentation for both real data and generated data can maximize the use of data augmentation to improve the model capacity. Meanwhile, it is found that

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0073

spatial data augmentation is more helpful to the model than only image brightness, color and Noise Instance augmentation.

5. Feature Extraction of Convolutional Neural Network

In convolutional neural networks, the first step is to extract features with convolution kernels. These initialized convolution kernels will be updated again and again during iteration in the process of back propagation, so as to approach our real solutions infinitely. The essence of image matrix to solve, but to initialize a fits the characteristics of a certain distribution vector set, then update the infinite in back propagation feature set, can make it an approximation to the conceptual feature vector in the math, so that we can extract the feature vector of numerical method to simulate the matrix. This study on X-ray imaging image size is too large and only a small fraction of lesions of the image, and can affect the characteristics of the surrounding structures, using the image threshold segmentation technology background and lung tissue segmentation, and redefine size, make the image to the need of the model and size, and reduces the complexity of the model. The degree of membership of each tumor in these models.

Is calculated and treated as a new feature in the training model. Convolutional neural network (CNN) is often used for feature extraction of images. It mainly extracts feature values from images through self-set convolution kernel and uses K-means algorithm to identify benign and malignant tumor patterns respectively, and then classifies them according to the calculation results of their feature values. Thus the accuracy of the model is obtained.

5.1 Loss Function Image

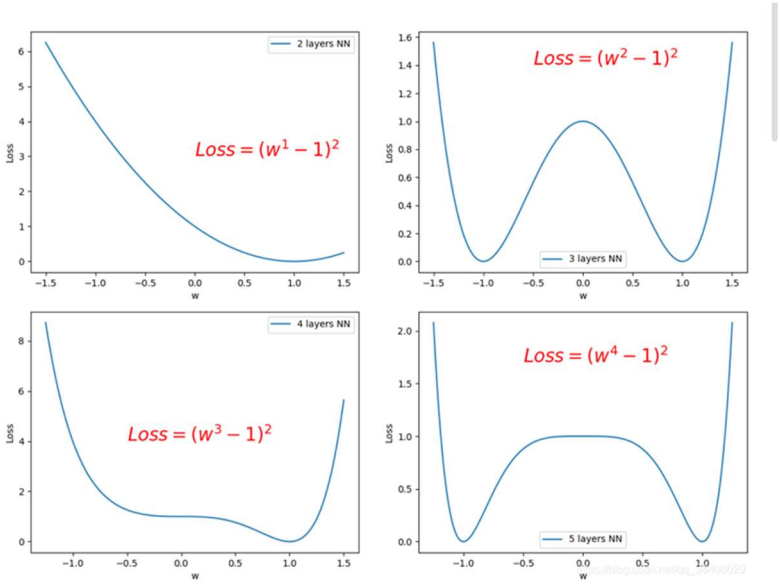

Fig.2 Loss function image

5.2 Model accuracy

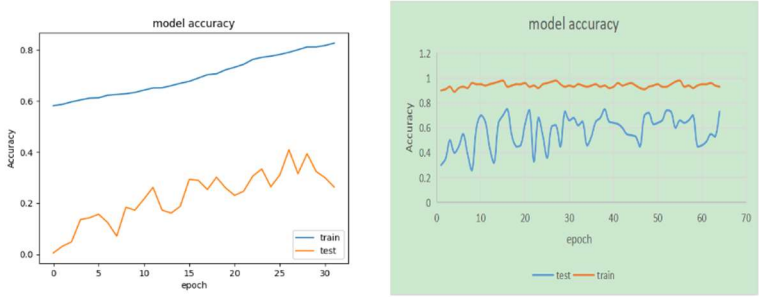

Fig.3 Model accuracy

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0073

References

[1] Chen J.X,Hou Y.J,Qi H,Zhang C.Q.Value of CT energy spectrum imaging in differential diagnosis of primary lung cancer and lung metastasis.(Journal of medical,China 2016.26(04)) imaging,p.626-629.

[2] Liu J.B Study on computer aided system for CT diagnosis of pulmonary nodules[D](capital medical university,China 2017).

[3] Li M.Y Preliminary study of CT energy spectrum imaging in the diagnosis of benign and malignant pulmonary masses [D].(Shandong University,2013).

[4] Wu G.Y,Wu J.L,Research progress of computer aided detection and diagnosis of pulmonary nodules based on CT[J].(Chinese medical imaging technology,2018.34(07))p.1114-1117.

[5] Liu K,Research on artificial intelligence for cancer screening and diagnosis based on Neural Network[D]. (Beijing University of Posts and Telecommunications 2018).

[6] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation, " in IEEE Transactions on Pattern Analysis andMachine Intelligence, vol.38, no.1,pp.142-158, 1 Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.

[7] He X.Y, Han Z.Y, Wei B.Z. Automatic classification of breast cancer pathological images based on deep learning[J]. ("Computer engineering and Application "The only official website, 2018, 54(12)) p.121-125.

[8] Zhang S,Gong Y.H,Wang J.J.Development of deep convolution neural network and its application in the field of computer vision[J].(Journal of Computer Science,2019,42(03))p.453-482.

[9] Dou Q, Chen H , Jin Y , et al. 3D Deeply Supervised Network for Automatic Liver Segmentation from CT Volumes[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2016.

[10] Liu Y.J,Xie H.W,Liu A.Y,Zhang H,Qiang Y.Classification of benign and malignant pulmonary nodules based on improved convolutional neural network[J]. (Computer engineering and design,2019,40(07))p. 2013-2018.

[11] Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms[J] . Bichen Zheng,Sang Won Yoon,Sarah S. Lam. Expert Systems With Applications . 2014 (4).