成都理工大學中英合作办学
CDUT Sino-British Collaborative Education

OXFORD BROOKES UNIVERSITY

Deep learning Application for the recognition of Breast Cancer

Oxford Brookes University- CDUT
Supervised by Dr. Grace Ugochi Nneji
Joseph Zhang

Abstract

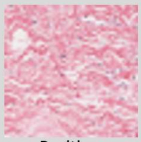
Invasive ductal carcinoma (IDC) is the common type of the breast cancer and the breast cancer is a danger disease of female. The core needle can provide the detect result of between IDC and benign lesion in the non-operative way of breast cancer. The process can use the deep learning model to recognition the positive character in core needle image to help the pathologists to analysis the disease. With the robust performance of the deep learning in the recognition of medical images, this project has inculcate the use of convolutional neural network for the recognition of breast cancer. In this project, the image recognition model uses AlexNet model to improve the recognition result. The dataset used in this work 42063 images with a split ratio of 82% for training and 18% for testing. More so, the project model achieves a performance of 81.9% and 0.326 in accuracy and loss. Additionally, this project designs a GUI to input the image of IDC and identify the result correctly which can be introduced to medical practitioners as this will save time and resources for them.

Introduction

The breast cancer is a serious and famous disease of female. According to Nneji et al [1] although the breast cancer own several check way by using mammography and others, the cure result of breast cancer will be raised when the tumor is in the early stage and biopsy is the most authoritative way to check breast cancer. Of course, the biopsy need the medical officer use the microscope to check the histological slides and define the area cancer situation. In fact, the identity of the Invasive ductal carcinoma (IDC) is the key of detect the breast cancer. As for the meaning of the IDC, Kanavati and Tsuneki [2] stated that and IDC is a group of the tumor that lack of the enough character to classification as a specific histopathological type and IDC own many histopathological characteristics such as the diffuse-sheets, well-defined nests, cords or single cell in the checking by using microscope. In fact, Kanavati and Tsuneki [2] also noticed that the deep learning is own the widely used in the computational histopathology algorithms to detect the core needle biopsy specimens and deep learning is exist the example in breast histopathology Classification including the cancer cell classification and segmentation and predict the cancer patient core result and prognosis. In other words, the deep learn own the positive effect in the breast cancer recognition.

Methodology

The model is made by the CNN to recognize the image. In fact, the CNN model will build in the Pycharm by using the Pytorch and the dataset will be used to train the model and figure out the loss. Of course, when model is convergence the test dataset will check the loss and confirm overfitting. Besides, the loss chart will be draw to analysis the model train result. As for the specific model of the CNN, the program will use the AlexNet model. In fact the AlexNet network in the report is following the structure of the Krizhevsky and Sutskever and Hinton [3] research report. In fact, the report using one GPU to train the network to save the time and decrease the calculation quantity in training. Then the dataset is a Breast Histopathology Images from the Kaggle and the dataset is a combine with Invasive Ductal Carcinoma (IDC) images. The dataset is made by 198,738 IDC negative image patches and 78,786 IDC positive image patches. Of course, the report use the negative IDC image is 16,247 and the positive one is 18,276 in the dataset for train and use 4,027 images in negative and 3,513 images in positive for validate. The 34,523 images can escape the over fit in the train process by the dataset is too small and the class of the dataset can keep balance.


Positive

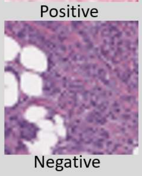

Negative

Figure 1. The dataset image

Model Design

The network is made by 8 layers including 5 convolution layers and three fully connect layers. The network using drop out function in fully connect layer to flexible the network

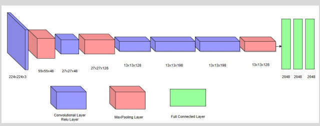


Figure 2. The CNN structure

GUI Design

GUI is using the Tkinter model to implement the GUI for input the IDC images to the model and recognition the correct result of image and show it for the user.

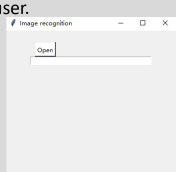


Figure 3. The GUI windows

Train Result

Model get the accuracy and loss of the model train result. The result of the train is get the 83.1% accuracy and 0.325 validate loss.

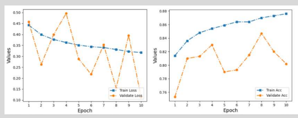


Figure 4. The curve of acc and loss

GUI Result

GUI is shown the image, which the user input and the IDC recognition of the image is and shown in the GUI textbox line.

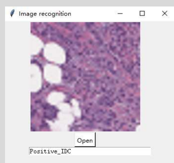


Figure 5. The GUI result

Discussion

The IDC image detection model is using the AlexNet model to implement the image recognition functions. In the model training process, the model accuracy is achieve the 83.1% and 0.325 loss. Then the application function is implement in the GUI and the GUI is using the Tkinter frame to implement the image input and recognition result.

In fact, the model can get many point of the future work, the model should add the transposition layer to add the sensitive of the character graph of the model. Of course, the next train plan should use the other image of the dataset for meet the more requirement.

Reference

[1] G. U. Nneji et al., “A Super-Resolution Generative Adversarial Network with Siamese CNN Based on Low Quality for Breast Cancer Identification”, 2021 the 4th International Conference on Pattern Recognition and Artificial Intelligence, doi: 10.1109/PRAI53619.2021.9551033.

[2] F. Kanavati and M. Tsuneki, “Breast Invasive Ductal Carcinoma Classification on Whole Slide Images with Weakly-Supervised and Transfer Learning”, doi: 10.3390/cancers13215368.

[3] Y. Hou, “Breast cancer pathological image classification based on deep learning”, Journal of X-Ray Science and Technology 28(7):1-12, doi: 10.3233/XST-200658