Patch-Based Convolutional Neural Network for Breast Cancer Detection

Haoyuan Chen, Jialiang He, Jianxia Lin and Linrui Li
Department of Computer Science and Engineering
Hong Kong University of Science and Technology, Hong Kong

1. Introduction

In this project, we are required to classify the x-ray images into normal and abnormal class to help detect the Breast Cancer, the most common cancer in women worldwide.

- **Problem description**
  Our target is to build a model to conduct a binary classification on x-ray images. However, for medical x-ray images, they have much higher resolution than normal ones, so it is hard to store and handle such big feature mappings in the GPU. Furthermore, the common solutions of resizing or compressing the images are not recommended because those operations may lead to loss of information details.

- **Solutions**
  To solve the problem, we divided the images into several patches, and then turned it into a multi-instance learning task. In the multi-instance task, we can make use of all information of the given images. For the labels of each patch, we assumed that the image-level label can represent every patch from the same image. Then we use patches and its labels as input to build a CNN (Convolutional Neural Network) model.

2. Proposed algorithm

Generally, we built a two-step neural network model to divide the problem into two parts. For each part, we proposed a specific solution. First, we proposed a patch-based CNN algorithm, which is implemented by cutting each x-ray image into several patches. However, the patches we got may make little contributions to the image-level labels, so we built a convolutional layer to extract discriminative patches. For the second part, we put the picked patches into another convolutional network to predict whether the image is normal or not. The link of these two networks is based on Expectation Maximization concept. The structure of these two networks and their connection is given as below. Actually, they compose a large network which has two convolutional networks linked by a condition.

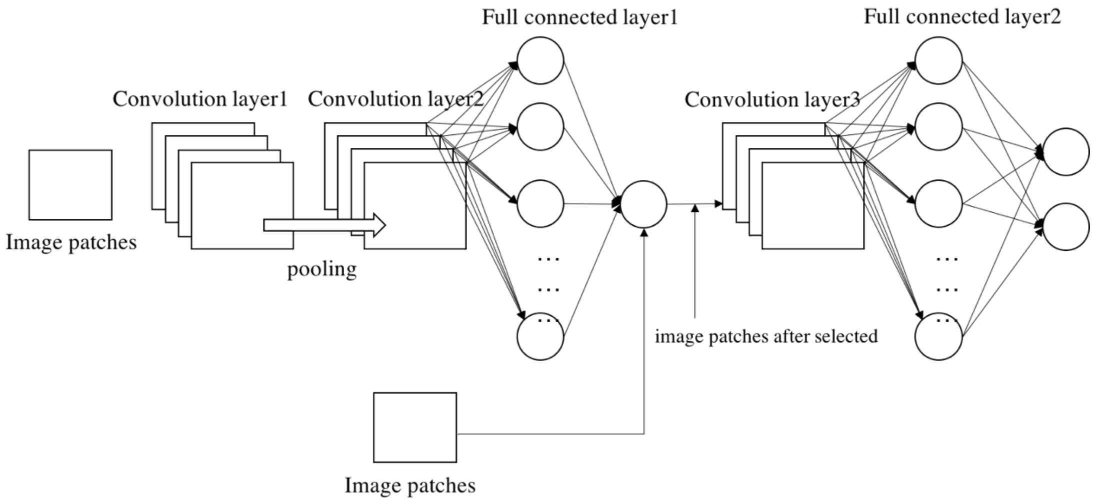

Figure1: structure of the whole network

- **Convolutional network for patch extraction**

  This part targets the extraction of discriminative patches. This is a patch-based method, unlike the traditional image classification which is usually based on the whole image. The reason why we need this part is that the resolution of our original images is too lager to use CNN network to train the model. In this part, we use patch-based CNN network to extract the discriminative patches which will be the input of next convolutional network.

  The input of this part is the image patches and output is probability of discrimination for each patch. If the output of a patch is larger than a threshold, then the patch will become the input of next convolutional network.

  The structure of this network below:

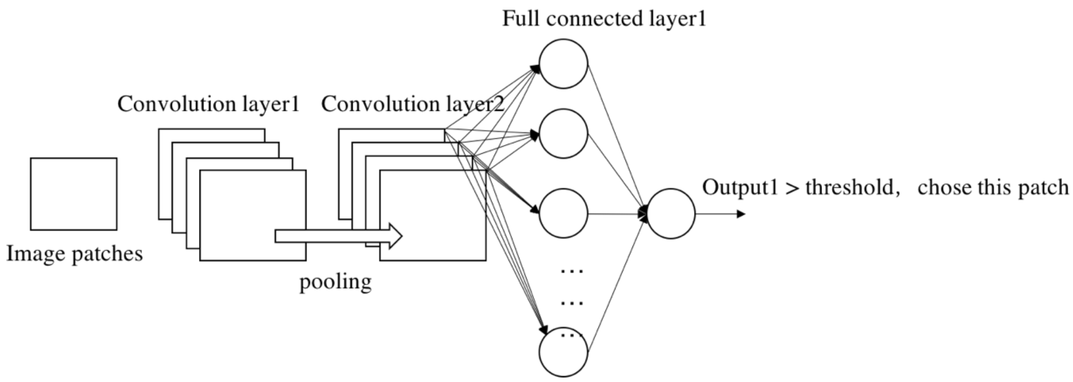

Figure2: structure of first network

- **Convolutional network for classification**

  This part targets the classification of the discriminative patches extracted from the first convolutional network. If the output of first convolutional network is larger than threshold, the network will extract the corresponding patch and become the input of this network. The output of the second network is a probability of belonging to the given class. it is a supervised network. The label is the class of the image which contain the input patch.

  In this network, after the get the output of every iteration, the network will adjust parameters of both networks according to the loss function using gradient descent. The loss function is cross-entropy loss.

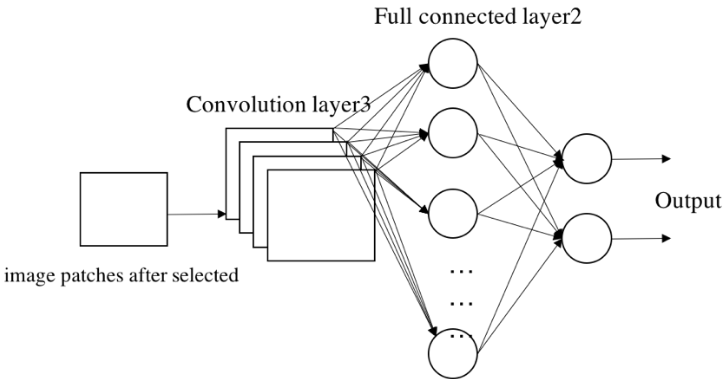

Figure3: structure of second network

The whole process is based on Expectation Maximization (EM) algorithm. We initially input all the image patches into the second network, that is: we assume all the patches is discriminative at first.

After the first iteration, we get the loss of the second network, and adjust the parameters of first network, then input all the image patches into first network and get the discrimination of each patch, and adjust the input of the second network. Repeat these steps again and again until convergence.

3. Experiment

The main procedure of our task experiment is shown below in detail.

- **Data processing**

  The dataset contains 410 x-rays images of breasts with normal and abnormal class. The resolution of each image is around 3000 pixels times 4000 pixels. The size of images is too large to handle in a machine, so we decided to cut those images into patches that the resolution of each patch is 100 pixels times 100 pixels. After cutting the images, it can be noticed that there existed several patches filling with all black color and contained no information. So we discarded those useless patches before we built models.

- **Modeling**

  The input of the first convolutional network is all the patches except for the discarded black ones. After that, each time we picked 60% of the output of convolutional layers to be the input of the next classification network.

  The loss function of our two-step model is L2-norm, which measures the accuracy of binary classification. Through several times of feed-forward and back-propagation, the output of the extracting convolutional layer is the degree of importance of each patch. Then we picked the top 60% of those patches to do the classification. We use softmax to be the activation function and finally the neural network gave out the probabilities of each patch to be abnormal.

- **Results**

Since the problem is a binary classification, we choose to use the metrics of accuracy, precision, recall and f1-score to evaluate the model performance in the validation set. The formulas of those metrics are provided below.

$$Accuracy = \frac{Y_{correct}}{Y_{total}}, Precision = \frac{tp}{tp + fn}, Recall = \frac{tp}{tp + fp},$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}.$$

The validation set contains 40 x-ray images. Put the validation images into our trained model, we get the predicted label of those images. The generalized confusion matrix is shown below.

Truth / **Predicted**	Positive (1)	Negative (0)
Positive (1)	2	1
Negative (0)	10	27

Table1: Confusion matrix of classification results

To sum up, we have the performance of the model in the validation set:

Accuracy = 29/40 = 0.725,

Precision = 2/3 = 0.666,

Recall = 2/12 = 0.167,

Precision = 0.267.

From the classification results we got above, we can see that our model has a good performance in predicting the normal class (0 class). But for the abnormal class, the model did a poor job since the recall is quite low, which means that the model tends to classify the patients with breast cancer into normal person. This result will lead to a severe problem that it will delay treatments for real patients. Therefore, we should improve the model later on by thinking more highly of the recall.

4. Conclusion

In this task, we studied how to deal with medical x-ray images by cutting them into patches and then use those patches to build the classification model. For the problem of labelling those patches, we estimate the label of each patch by their corresponding image-level label, and use convolutional neural network to pick the discriminative patches. After the process, we can give out a prediction of a test image through voting the prediction from its patches.

We modelled deep networks to extract discriminative patches to do the next step of classification, the accuracy is good, but the result is poor in the abnormal class. In conclusion, the performance will get worse when we sacrifice to pursue the disposable solution. Therefore, the structure and parameters of our neural network should be trained more careful. In the future, we may improve our model by adjusting the model structure and do more image processing.