Lung Nodule Classification using A Novel Two-stage Convolutional Neural Networks Structure'

Yang An[1],Tianren Hu[1] ,Jiaqi Wang[1], Juan Lyu[2], Sunetra Banerjee[1], Sai Ho Ling[1]

Abstract— Lung cancer is one of the most fatal cancers in the world. If the lung cancer can be diagnosed at an early stage, the survival rate of patients post treatment increases dramatically. Computed Tomography (CT) diagram is an effective tool to detect lung cancer. In this paper, we proposed a novel two-stage convolution neural network (2S-CNN) to classify the lung CT images. The structure is composed of two CNNs. The first CNN is a basic CNN, whose function is to refine the input CT images to extract the ambiguous CT images. The output of first CNN is fed into another inception CNN, a simplified version of GoogLeNet, to enhance the better recognition on complex CT images. The experimental results show that our 2S-CNN structure has achieved an accuracy of 89.6%.

I. INTRODUCTION

According to 2018 Global Cancer Statistics Report, there will be 18.1 million new cancer cases and 9.6 million cancer deaths worldwide. Lung cancer is the most common type of cancer (11.6%) and the leading cause of cancer deaths (18.4%) for the overall population. Thus, an effective early detection of lung cancer plays an important role to increase the survival rate of patients.

Computed Tomography (CT) is a versatile medical imaging technology and finds its application in detection of a variety of diseases. It is typically characterized by a quick scanning time and captures clear images. However, the inference from a CT diagram can vary based on the assessing doctor's experience and capability. This can be overcome by using Computer Aided Diagnosis (CAD), a more effective and accurate way to classify the CT diagram, which reduces the chances of errors due to human subjectivity and facilitates automation.

Recently, convolutional neural network has gained popularity as an effective technique to classify images. Much work has been done in this front, [1], wherein the basic CNN structure has been applied to lung cancer CT images segmentation. Although it did not classify the types of the lung cancer, the CNN with u-net structure effectively extracted feature points from the CT images. Secondly, Rotem et al. [2] proposed a multi-structure CNN. The first CNN structure is trained by back-forward propagation using segments of a CT images. The output is then re-fed into the same CNN for successive iterations. This process is repeated until the presence or absence of a lung nodule in the CT image is sufficiently established. However, this technique only achieved 78.9% of sensitivity which was lower than other methods. Thirdly, Wei Shen et al. [3] proposed a multi-crop convolutional neural network (MC-CNN) to classify the lung CT images. This method crops the features two times in each pooling layer, cutting the original picture into a relatively small picture. The output size was further reduced by employing max-pooling methods. Although the method is more advanced and has lesser computational requirements, there is a high chance of losing information in the cropping process.

Needless to say, the structure of a CNN is one of the governing factors which determines the accuracy of the outcome in an image classification exercise. This paper proposes a novel two-stage CNN (2S-CNN) structure to enhance the classification performance of the lung cancer CT images compared to the earlier attempts. Throughout our experiments, we found that our proposed CNN structure did not lose much image's information and also had a good recognition effect for complex medical images.

The structure of our proposed 2S-CNN consists of two CNNs. The first CNN classifies the initial input CT images. For a single CNN to achieve good results, CNN must be capable enough to classify difficult images as well. In most cases, a single CNN is not very capable to give the required performance. Therefore, we developed a second inception CNN. The second CNN is a simplified version based on GoogLeNet [4]. GoogLeNet is known to have a good image recognition ability with an inception structure which can be used to better identified even complicated images. In our proposed 2S-CNN method, we used the output of the unrecognized CT images from first CNN as an input to an inception CNN (second CNN) to classify. With this proposed structure, the classification accuracy is improved significantly.

II. METHODOLOGY

The structure of proposed 2S-CNN, is presented in Fig. 1, composes of two CNN. The first one is a conventional CNN (Fig 1a) and the second one is an inception CNN (Fig 1b). At the beginning, the unprocessed medical images are provided as input into the first CNN. In the convolution layer of CNN, there are many random local

[1]Y. An, T. Hu, J. Wang, S. Banerjee, and S.H. Ling are with the School of Biomedical Engineering, University of Technology Sydney, NSW, Australia Yang.An-1@student.uts.edu.au, Tianren.Hu-1@student.uts.edu.au, Jiaqi.Wang-5@student.uts.edu.au, Sunetra.Banerjee@student.uts.edu.au, Steve.Ling@uts.edu.au

[2]Juan Lyu is with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, 150001,China Juanlyu91@gmail.com

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

reception fields or filters with shared weights and biases. The process of updating can be expressed below,

$$a_{out} = \sigma(b + \sum_{l=0}^{l} \sum_{m=0}^{m} w_{l,m} a_{j+l,k+m}) \quad (1)$$

$$a_n = \sigma(b + w * a_{n-1}) \quad (2)$$

where l and m denote the row and column of image. k and j are times. w is weight and b is bias. a is the output from the last calculation. $\sigma()$ is activation function. We used rectified linear units (ReLU) as the activation function because it can avoid over-fitting when updating the network.

Then the output will be input into the pooling layer whose function is to compress the size of images. In this network, we adopt max pooling layer.

The updating method is stochastic gradient decent (SGD) algorithm. The result, obtained through a softmax classifier, includes two values which represent the probability of each kind of nodules. It can be defined as

$$a_q^L = \frac{e^{z_q^L}}{\sum_p e^{z_p^L}} \quad (3)$$

where p is the total number of neurons and q is the$q$-th neuron. For example, if the result is (0.4,0.6), it means that there is a 0.4 probability of the nodule to be benign and 0.6 probability to be malignant. However, there may be results wherein both the two values are similar such as 0.51 and 0.49. In such cases, the network might figure that the nodule is malignant but, in all likelihood, the result is inconclusive. Hence, we decided to define the two values in output as Value1 ($V1$) and Value2 ($V2$) and assign a new parameter, D, is the absolute difference between $V1$ and $V2$. It can be represented as follows:

$$D = |V1 - V2|. \quad (4)$$

After multiple experiments, we found that for values of $D > 0.60$, the network could classify an image correctly and for values of $D < 0.35$, the network was not able to classify an image correctly. Therefore, we defined the threshold limits of D as less than 0.35 for uncertain result and greater than 0.60 for certain result. Then, we took out of these uncertain results ($D < 0.35$) and re-classified their corresponding images. In doing so, we realised that if the same CNN structure or network was re-used, to classify the uncertain images again, the results were not satisfactory. To overcome this, we decided to work on creating another structure, an inception structure, specifically to handle the uncertain images. The initial structure of the inception CNN is shown in Fig. 1(b). In Fig.1, BN is Batch Normalization,avpool is Average Pooling, concat is Concat Layer, drop is Dropout layer and fc is Fully Connected Layer.

One of the well-known ways to improve the performance of deep convolutional networks is to increase

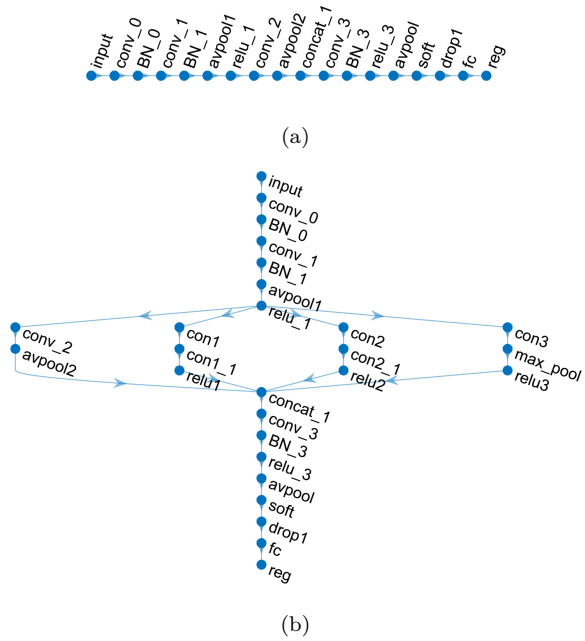

Fig. 1. The structure of 2S-CNN. (a) first CNN, (b) inception CNN (second network).

the size of both depth and width. The downside is that the large-sized networks require more parameters, which may result in over-fitting when the data set is not large enough. We realised the solution for this problem is to change the fully connected layer into a sparse link layer. The disadvantage of having a non-uniform sparse network is that it is not very efficient. To overcome this, we decided to combine multiple sparse matrices into more denser sub-matrix through an inception architecture. The main objective of the inception architecture is to make fit some existing dense components closer to (and cover if possible) the best local sparse structure in a convolutional network [5]. Using 1×1 convolution structure reduces the number of channels and enhances the non-linearity of the network. The horizontal convolution kernel arrangement makes multiple convolution kernels, of different sizes, obtain more information from different parts of the image. The inception structure is specifically created to classify the images which are difficult to be recognized, i.e. the uncertain results from the first network. However, application of the inspection structure did not enhance the accuracy of the results. The accuracy was still 85%, same as the result obtained from traditional structure of convolutional neural network, when it was used to classify the lung nodules images.

In the experiments, another problem emerged. Although the uncertain results from the first CNN (network 1) were removed, the accuracy reached only 95%. This meant that there were still lots of potential to further improve the network. We propose another idea to solve this problem called as "*voting*". To enumerate,

we created five networks with similar structure but different parameters. For the same feature, although the different results could be generated, their values were observed to be quite indistinguishable. For example, if the five networks could get the distinct values such as (0.4,0.6), (0.3,0.7), (0.45,0.55), (0.25,0.75), (0.28,0.72), the resulting image can be ascertained to be malignant nodule with a reasonable accuracy. But if the values have lesser variations such as (0.6,0.4), (0.3,0.7), (0.45,0.55), (0.25,0.75), (0.28,0.72), we cannot conclusively determine if the nodule is malignant or benign because there is one result toward benign nodule and other four results toward to malignant nodule. To overcome this, we use the method "*voting*". The final "voted" result is obtained from the results obtained by the earlier networks. "*Voting*" can help to minimize the errors. After multiple iterations, we achieved an accuracy of 97%, an improvement from result of the first CNN (network 1) of 95% calculated after taking out the uncertain results.

The whole structure of the classification system is presented in Fig. 2. Firstly, the images was fed into 5 different networks (from network A to network E) simultaneously. These 5 networks have similar structures (the structure of 2S-CNN in Fig 1(a)). From the 5 results obtained, voting method was applied. As seen in the figure, Network 1 consists of 5 voted networks. The output from Network 1 was made up of 2 types of results: (a) The uncertain results which were segregated and provided as an input to network 2 for further classifications and (b) the remaining results were designated the outcome 1. The structure of network 2 is presented in Fig. 1(b). The results got from network 2 were taken as outcome 2. Finally, both outcome 1 and outcome 2 were combined to determine the final classification outcome.

We have used data set of about 300 data points to test the network and 700 data points to train the network. From the outcomes, we saw that there are 54 outcomes whose D is below 0.35. These 54 outcomes were classified as uncertain results. If these uncertain outcomes were left as it is, the accuracy would be 246/300=82%. Similarly, we can make sure that most of uncertain results can be used to classify the nodules incorrectly. From this experiment, only 10 uncertain results could be used to correctly classify the nodules. We took 54 uncertain results out from 300 data and used the remaining 246 data to test the network. Therefore, the accuracy achieved from network 1 was 237/246=96.3%, which meant the number of correct classification was 237 and the error was 9. After that, we input the left 54 data into network 2. We hypothesised that 27 results of classifying the nodules were correct, that is an accuracy of 27/54=50%. Therefore, combining these two results meant that there were a total of 27+237=264 correct results. The final accuracy is 264/300=88%. On the other hand, if we had used network 1, the accuracy is only (237+10)/300=82%. This experiment significantly improved the accuracy.

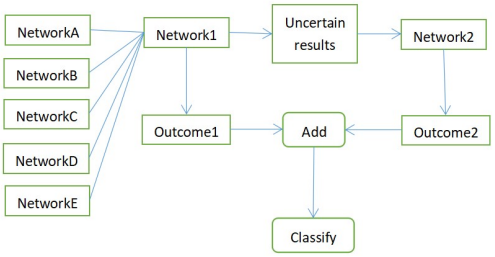

Fig. 2. The overall structure of the concept.

III. Experiments and Results

A. Database

The training data source for this project is from LIDC-IDRI, which is composed of chest medical image files (such as CT, X-ray) and corresponding diagnostic result lesions. The data was collected by the National Cancer Institute for the public study of early cancer detection.

Our experimental database included 1000 images in total. For each of the images in the example, a two-stage diagnostic label was used by four experienced chest radiologists. In the first phase, each physician independently diagnosed and labeled the patient's location into three categories: 1) >=3mm nodules, 2) <3mm nodules, 3) >=3mm non-nodules. In the following second phase, the physicians independently reviewed the annotations of the other three physicians and gave their final diagnosis. Such a two-stage annotation is helpful for marking all results as completely as possible while avoiding forced consensus.

B. Result and Discussion

For our experiment, the total number of training data used was 700 from a data base of 1000 data points, the rest 300 data-points was used for testing. The size of CT image is 28x28. In order to improve the performance, we changed the CNN structure as well including learning rate, the number of filters, the number of Convolution layers, with or without ReLU function.

We initially set up the parameters including the batch size as 128, the number of epochs as 40, using softmax as the classifier.

The overall classification results are tabulated in Table I. In this Table, *C* in structure column represents a Convolutional layer. *P* represents Pooling layer. *F* represents Fully-connected layer and *S* represents Softmax function. To optimize the structure of first CNN, firstly, we need to find the optimal number of convolutional layer. We had tried the number of convolutional layer from 1 to 5 layers and found that the best number of convolutional layers is four. Thus, the optimized structure of first CNN is CCCCPFS where there has 4 convolutioanl layers followed by single pooling layer, one fully-connected and Softmax function. Secondly, we

need to find the appropriate learning rate to achieve best results, according to the results of the Table I , the most appropriate learning rate for the network is 0.025. The reason could be that the higher or lower initial learning rate could lead to overfitting or information loss, which could cause the errors occur. After to choose the appropriate learning rate, we changed the number of filters for each convolutional layer.

As shown in the Table, when the number of filter is 10, 20, 30, 40 for each Convolution layer respectively, the system has the best classification accuracy. The function of filter is to get the features of the images. Increasing the number of filters could improve the accuracy. However, large number of filters could generate many repeated features, which could cause over-fitting.

Then, we worked towards optimizing the inception structure (second CNN). After multiple attempts, 68 images were found to be difficult for classification. Therefore, we segregated them as the difficult data-set. Then, we also use 700 images to train the inception CNN and used these 68 images to test the network. Finally, the whole structure of inception structure could be optimized and is shown in Fig. 3.

Apart from these, we also changed the existence of ReLU function in our system. After we added the ReLU function, the accuracy is improved from 77.6% to 85.8%. The reason is that ReLU can make the nonlinearity mapping of the convolutional layer output. Finally, we combined the first CNN and the second inception CNN to be a two-stage convolutional neural network structure. We found that the proposed 2S-CNN achieved a best classification accuracy which is 89.6%.

TABLE I
The overall classification results.

Structure	Number of Filters	Learning rate	ReLU	Accuracy (%)
CCCCPFS	10 20 30 40	0.025	0	77.6
CCCCPFS	10 20 30 40	0.5	0	75.8
CCCCPFS	10 20 30 40	0.01	0	74.9
CCCCPFS	20 40 60 80	0.025	0	72.8
CCCCPFS	7 14 21 21	0.025	0	75.3
CCCCPFS	10 20 30 40	0.025	1	85.8
CCCCPFS + inception nCNN (2S-CNN)	10 20 30 40	0.025	1	89.6

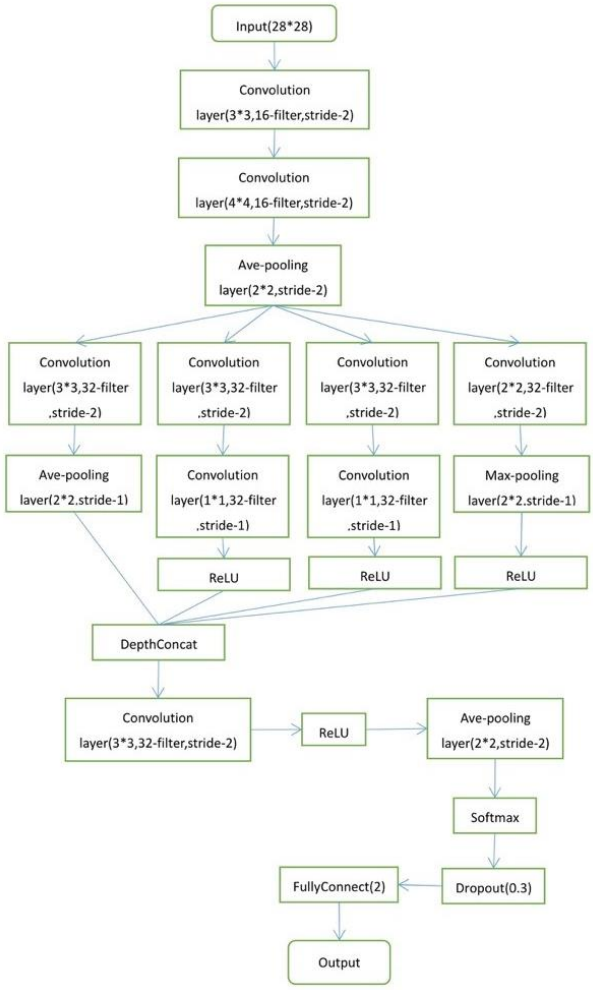

Fig. 3. The structure of inception CNN.

IV. Conclusions

In this paper, the lung nodules classification system by using a novel two-stage convolutional neural network is built and simulated. After testing and analyzing the experimental results of a number of CNN system with various parameters sets and different structures, we finally found that the best classification accuracy we have achieved is 89.6%. Also, we found that the multi-level CNN system could significantly improve the accuracy acheived by a one-layer system.

References

[1] B. A. Skourt, A. El Hassani, and A. Majda, "Lung ct image segmentation using deep neural networks," Procedia Computer Science, vol. 127, pp. 109–113, 2018.

[2] R. Golan, C. Jacob, and J. Denzinger, "Lung nodule detection in ct images using deep convolutional neural networks," in Neural Networks (IJCNN), 2016 International Joint Conference on. IEEE, 2016, pp. 243–250.

[3] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, and J. Tian, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," Pattern Recognition, vol. 61, pp. 663–673, 2017.

[4] Z. Zhong, L. Jin, and Z. Xie, "High performance offline handwritten chinese character recognition using googlenet and directional feature maps," in Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. IEEE, 2015, pp. 846–850.

[5] P. Tang, H. Wang, and S. Kwong, "G-ms2f: Googlenet based multi-stage feature fusion of deep cnn for scene recognition," Neurocomputing, vol. 225, pp. 188–197, 2017.