

Lung Cancer Classification Using 3D-CNN with a Scheduled Learning Strategy

Yadi Li¹, Yin Tian² and Bao Ge^{1,*}

¹School of Physics & Information Technology, Shaanxi Normal University, Xi'an, Shaanxi, China

²College of Information & Communication, National University of Defense Technology, Xi'an, Shaanxi, China

*Corresponding author

Abstract—Lung cancer is one of the most common forms of cancer resulting in over a million deaths per year worldwide. In order to classify the lung CT images, this paper presents a classification method using 3D-CNN with a scheduled learning strategy. To get compact and uniform data for training and feature extracting, the input should be unified into 100×100×20 dimension. We construct a 3D-CNN model where a scheduled learning strategy method is proposed in the process of network training. This method is shown to out-perform the state-of-the-art approaches by experiments conducted on the datasets of lung CT scans in Kaggle.

Keywords—3D-CNN; scheduled learning; lung cancer

I. INTRODUCTION

As the leading cause of cancer-related mortalities, lung cancer is responsible for approximately 1.38 million deaths annually worldwide [1]. Although target therapeutics and various chemotherapy regimens are available, locally advanced lung cancer carries a very poor prognosis, with a mean survival time of less than 12 months. Thus, early detection of a lung lesion to increase the likelihood of survival rate is important. Chest computed tomography (CT) scan, has been widely accepted for detection of lung tumors [2].

A traditional CAD (Computer-aided diagnosis) scheme for classify the lung cancer commonly includes the preprocessing of image, segmentation, feature extraction [3], and then a classification task for differentiation of a tumor. Its performance relies heavily on the intermediate steps [4]. The recent advent of deep learning techniques has the ability to automatically uncover features from the training images and perform the classification task, it can be seen as an integrated tool to recognize the lung cancer without a complicated pipeline of image processing and pattern recognition steps in the traditional methods [5]. One of most prevailing deep learning method for classification is CNN(Convolutional Neural Network)[6], while the popular CNN models like VGG(Visual Geometry Group)[7], are two dimensional, it means that if we want to use these models on three dimensional (3D) images, like CT images, we should treat the third dimension as a channel of CNN and sum up the convolution sum of all channels with a 2D filter, each channel corresponding a 2D slice of 3D image, this will weaken the features of the third dimension. So, three dimensional CNN (3D-CNN) is naturally more appropriate for the three dimensional images, and can extract the 3D feature via a 3D filter. Recently, there existed research work made a good

attempt that applying the 3D-CNN to identify lung CT images [8], while its network structure and learning technique was preliminary and coarse. In this paper, we presented an improved 3D-CNN model with a scheduled learning strategy to differentiate the normal and cancer lung images.

II. MATERIALS AND METHOD

A. Overview

Our framework of the proposed approach is presented in Figure 1. First of all, we preprocessed all the lung CT images, and then spliced them together to be one entire matrix in a scheduled order, which was then fed into the 3D-CNN model as a train set, finally the trained 3D-CNN was used to classify the lung CT images.

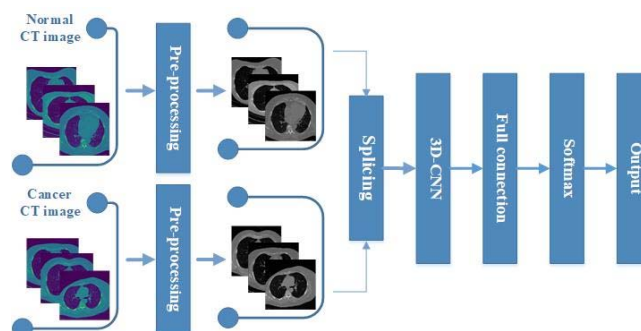


FIGURE 1. FRAMEWORK OF THE PROPOSED APPROACH

Datasets: In this paper, the datasets are from Data Science Bowl 2017 which is an annual data science competition hosted by Kaggle. The datasets include over a thousand low-dose CT images from high-risk patients in DICOM format. Each image has a variable number of 2D slices. There are lung CT scans of 1,397 subjects, of which 1,036 are normal and the rest are cancer patients.

Pre-processing: The size of original images is 512×512. Due to a variable number of slices in one scan and the limitation of the fully connected layer in the convolutional neural network, the pixel size and the number of input images should be normalized. Firstly, the pixel size of slices is resized into a standard 100×100 pixels via size normalization, and then we sorted the slices according to the Image Position attribute which describes the actual image position in the scan. Secondly, one scan may have hundreds of slices, we divided all the slices into a certain number of chunks and average them within

a chunk, we set the number to be 20 fixedly, and the slices were finally a centimeter thick or so. The pre-processing results were gray scale images with a dimension of 100×100×20. Take a cancer patient for example, the original and pre-processed image are shown in Figure 2:

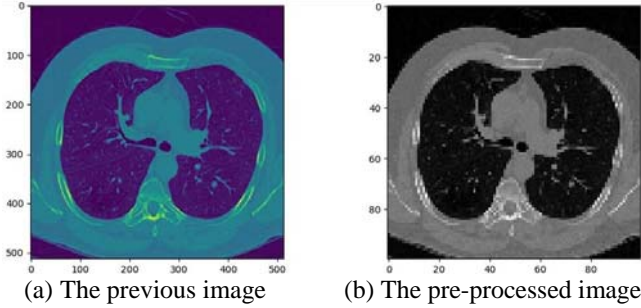


FIGURE II. THE ORIGINAL AND PRE-PROCESSED IMAGE

B. 3D-CNN Model

3D convolutional layer: 3D convolution itself is achieved by convolving a 3D kernel to the cuboid formed by stacking multiple contiguous frames together, in which the main factors to be considered in this layer are filter sizes, strides, padding options, and the number of filters. The equation used in this stage is Eq. 1:

$$O_n^j(x, y, z) = \sigma \left(b_{jn} + \sum_n \sum_{p=0}^{p_j-1} \sum_{q=0}^{q_j-1} \sum_{r=0}^{r_j-1} w_{mn}^j(p, q, r) \cdot I_{im}(x+p, y+q, z+r) \right) \quad (1)$$

where O is the output of the j -th layer, I is the input of the i -th layer to the j -th layer, σ is the activation function, w_{mn}^j is the weight between the m -th feature map of the i -th layer and the n -th of the j -th layer, b_{jn} is the shared bias of the cube.

3D maxpooling layer: this layer is operated in order to simplify the computational complexity and extract the main features. The equation used in this stage is Eq. 2:

$$O(x, y, z) = \max_{(0 \leq i \leq s_1, 0 \leq j \leq s_2, 0 \leq k \leq s_3)} (I_{x \times p + i, y \times q + j, z \times r + k}) \quad (2)$$

ReLU(rectified linear unit)layer: the ReLU[9] layer is then applied as the activation function which used to speed up training instead of using tanh units. Eq. 3 is the method used to obtain the output for a given input in a ReLU: $f(x) = \max(0, x)$.

Dropout layer: the dropout [10] is used as the main regularization approach to reduce overfitting. The basic idea of dropout is that one part of neurons is discarded by probability p and the others are reserved by the probability $q = 1-p$, while the output of the discarded neurons is set to zero.

Softmax layer: softmax function can be regarded as an activation function or a link function, each value resulting from this layer indicates the probability of an image belonging to a certain class, and the sum of the values becomes 1. The equation used in this stage is Eq. 3.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (3)$$

where, $j=1, 2, \dots, K$.

Cross-entropy: In this paper, the calculation of loss is mainly based on the Eq. 4.

$$\text{LogLoss} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (4)$$

where n is the number of data in the test set, \hat{y}_i is the prediction probability of one cancer patient, y_i is the label of each image.

C. Improved 3D-CNN with a Scheduled Learning Strategy

Before feeding the input data into the 3D-CNN model, we re-organized the training samples with a scheduled input order. The reason lies in that we think that one can guide training and remarkably increase the speed and quality at which learning can occur by choosing which examples to present and in which order to present them to the learning system [11]. For example, child always studies addition and subtraction firstly, then move to multiplication in the course of studying math. So, we let the 3D-CNN learn in the same manner, i.e., we re-organized the input data, and put the all normal CT images together and put the cancer data together, then spliced them into an overall training data, finally the 3D-CNN learned and adjusted itself in the scheduled input order of samples.

The 3D-CNN includes a series of layers, e.g., convolution, pooling and fully connected layers, which are linked serially. The first layer (C1) consists of 32 feature maps produced by 5×5×5 spatial convolutional kernels, the total number of variables are 32 × 100 × 100 × 20. The first max pooling layer(S1) with strides 2×2×2 was then applied, the total amount of variables were 32 × 50 × 50 × 10. The second convolution layer (C2) uses 64 feature maps with 5×5×5 kernels followed by the 3D max pooling layer (S2) with strides 2×2×2, the amounts of variables were 64 × 50 × 50 × 10, 64 × 25 × 25 × 5, respectively. The third layer(C3) was composed of 128 feature maps with 5×5×5 kernels followed by the 3D max pooling layer(S3) with strides 2×2×2, the amounts of variables were 128 × 25 × 25 × 5, 128 × 13×13 × 3 = 64896 respectively, in the end, a fully connected layer with a size of 1024 was used to get features, which were then passed to the softmax layer, the final classification of normal and cancer images was performed based on the output of the softmax layer. The suggested structure of the 3DCNN is presented in Figure 3.

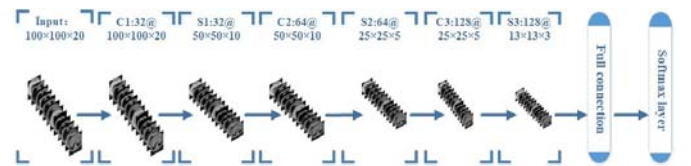


FIGURE III. THE STRUCTURE OF THE 3DCNN

III. RESULTS AND DISCUSSION

We applied the above 3D-CNN model with a scheduled learning strategy to the dataset, as a comparison, simultaneously we adopted the same 3D-CNN model but with

a random input order for training samples, then compared the performance of these two methods. There are 1,036 normal subjects and 361 cancer subjects in the dataset, we calculated the accuracy of classification using the 10-fold cross validation. Each time we selected 1/10 samples as validation data, the control group and the cancer group had 1/10 samples from themselves, respectively, that is, the number of training samples for the control group and the cancer group are 932 and 325, and the number of validation samples for them are 104 and 361, respectively. As a quantitative comparison, we computed the accuracy of classification and the loss according to the Equation 4. The result is shown in Table 1. We can see that the 3D-CNN model with a scheduled learning strategy has the higher accuracy which are 76% and higher than that (70%) of the original 3D-CNN model with random manner, after 10 epochs. In fact, the proposed method performed better at the first epoch. However, we can also see that the proposed method has higher loss than the original 3D-CNN model with random manner, perhaps the reason is that although our method can predict if the subject is normal (0) or cancer (1), but we are not very sure of that, i.e., we always did not give a more biased probability like 0.9, 0.1 etc.

TABLE I. THE COMPARISON OF THE PROPOSED METHOD AND THE SAME 3D-CNN MODEL BUT WITH A RANDOM INPUT ORDER

(a) The proposed method		
Epoch	Accuracy	Loss
Epoch1	0.61428	1.4247436
Epoch2	0.67857	0.6361115
Epoch3	0.69285	0.5992927
Epoch4	0.7	0.5781363
Epoch5	0.7	0.5715370
Epoch6	0.70714	0.5744853
Epoch7	0.70714	0.5709527
Epoch8	0.70714	0.5698619
Epoch9	0.70714	0.5684398
Epoch10	0.70714	0.5677213

(b) The same model but with a random input order		
Epoch	Accuracy	Loss
Epoch1	0.685714	1.469389
Epoch2	0.689914	1.0632085
Epoch3	0.592857	2.1511415
Epoch4	0.697857	0.927307
Epoch5	0.757143	1.2009408
Epoch6	0.757143	1.5676174
Epoch7	0.756613	0.8265165
Epoch8	0.759876	0.7198301
Epoch9	0.762257	0.8733496
Epoch10	0.764431	0.9331379

IV. CONCLUSION

In this paper, we proposed an improved 3D-CNN model with a scheduled learning strategy to classify the lung CT cancer images. The experimental results demonstrated that this method improved the accuracy of classification, but with the higher loss, we aimed to improve the sampling choice strategy and 3D-CNN model based on this initial learning method, and apply it into more datasets.

ACKNOWLEDGEMENT

This paper was supported by NSFC 61403243 and the Fundamental Research Funds for the Central Universities from China (No. GK201703017).

REFERENCES

- [1] T. Sun, J. Wang, X. Li, et al: Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set (2013). *Compute Methods Programs Biomed*, 111(2): 519-524.
- [2] K. L. Hua, C. H. Hsu, S. C. Hidayati, et al: Computer-aided classification of lung nodules on computed tomography images via deep learning technique (2015). *Onco Targets Ther*, 8: 2015-2022.
- [3] J. Z. Cheng, Y. H. Chou, C. S. Huang, et al: Computer-aided US Diagnosis of Breast Lesions by Using Cell-based Contour Grouping (2010). *Radiology*, 255(3): 746-754.
- [4] H. Mahersia, L. A. Gabralla and M. Zaroug: Lung Cancer Detection on CT Scan Images: A Review on the Analysis Techniques (2015). *International Journal of Advanced Research in Artificial Intelligence*, 4(4): 38-45.
- [5] T. W. Way, B. Sahiner, H. P. Chan, et al: Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features (2009). *Med Phys*, 36(7): 3086-3098.
- [6] Tran, L. Bourdev, R. Fergus, et al: Learning Spatiotemporal Features with 3d Convolutional Networks (2015). *Proceedings of the IEEE International Conference on Computer Vision*, 4489-4497.
- [7] K. He, X. Zhang, S. Ren, J. Sun: Deep residual learning for image recognition (2016). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.
- [8] Hussein, Sarfaraz, et al: Risk Stratification of Lung Nodules Using 3D CNN-Based Multi-task Learning (2017). *International Conference on Information Processing in Medical Imaging*, 249-260.
- [9] Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines(2010). *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814.
- [10] Krizhevsky, I. Sutskever, G. E. Hinton: Imagenet classification with deep convolutional neural networks (2012). In *Advances in Neural Information Processing Systems*, 1097-1105.
- [11] Bengio, Y., J. Louradour, R. Collobert, et al. (2009). Curriculum learning. *international conference on machine learning*, 41-48.