A Novel Computer-Aided Lung Cancer Detection Method Based on Transfer Learning from GoogLeNet and Median Intensity Projections

Tiantian Fang
The Hill School
Pottstown, PA, USA
e-mail: Hzhc_wjdi@163.com

Abstract—**In this research, a fast, accurate, and stable system of lung cancer detection based on novel deep learning techniques is proposed. A convolutional neural network (CNN) structure akin to that of GoogLeNet was built using a transfer learning approach. In contrast to previous studies, Median Intensity Projection (MIP) was employed to include multi-view features of three-dimensional computed tomography (CT) scans. The system was evaluated on the LIDC-IDRI public dataset of lung nodule images and 100-fold data augmentation was performed to ensure training efficiency. The trained system produced 81% accuracy, 84% sensitivity, and 78% specificity after 300 epochs, better than other available programs. In addition, a t-based confidence interval for the population mean of the validation accuracies verified that the proposed system would produce consistent results for multiple trials. Subsequently, a controlled variable experiment was performed to elucidate the net effects of two core factors of the system - fine-tuned GoogLeNet and MIPs - on its detection accuracy. Four treatment groups were set by training and testing fine-tuned GoogLeNet and Alexnet on MIPs and common 2D CT scans, respectively. It was noteworthy that MIPs improved the network's accuracy by 12.3%, and GoogLeNet outperformed Alexnet by 2%. Lastly, remote access to the GPU-based system was enabled through a web server, which allows long-distance management of the system and its future transition into a practical tool.**

Keywords-lung cancer detection; deep learning; transfer learning; GoogLeNet; median intensity projection (MIP)

I. Introduction

Lung cancer is the most fatal type among all cancer-related diseases nationwide, and has a five-year survival rate of 17.7% [1]. Any accurate and automatic determination of lung nodule malignancy can aid in expediting early diagnoses leading to improved treatment and higher survival rates. In addition, such innovation can save radiologists significant manual labor and valuable time.

Recently, researchers introduced computer deep learning into lung cancer detection, which promises to improve speed and accuracy in determining lung nodule malignancy. Computers are trained to study distinctive features of malignant lung nodules via processing lung nodule computed tomography (CT) scans. In previous research, two-dimensional CT images only incorporated one view of lung nodules; failure to include multi-view information in CT scans might lead to incomplete understanding of important lung nodule features [2]. In addition, data were processed via shallow task-specific convolutional neural networks (CNNs) or outdated pre-trained networks in visual imagery analysis: the former often has limited feature detection capability due to limited training, and the latter already has more efficient substitutes [3, 4]. Furthermore, none of these methods has been applied to clinical trials. The inadequacy of current models gives rise to the research of a fast, accurate, and stable system of deep learning-based lung cancer detection. In contrast to previous methods, a GoogLeNet-based [5] CNN using Median Intensity Projections (MIPs) processed from original CT scans was trained and developed. Access of remote computers to the GPU-based detection system was also enabled, which would ease the system's transition into a practical tool.

Contributions:

- A GoogLeNet-based CNN trained with MIPs is developed to accurately extract information from CT scans.
- Transfer learning method is employed to enable fast convergence and high accuracy.
- 100-fold data augmentation is performed to produce sufficient training data.
- Access to the GPU-based detection system from distanced hosts is enabled, which allows the system's future transition into a practical tool.

II. Materials

The proposed system was evaluated on the LIDC-IDRI dataset from Lung Image Database Consortium (LIDC) [6], one of the largest public dataset of low-dose lung nodule CT scans. The dataset comprises 1018 scans with slice thickness varying from 0.45 mm to 5.0 mm. Lung nodules with diameters greater than or equal to 3 mm were annotated by four expert radiologists. In order to integrally present all eligible nodules, scans with slice thickness greater than 3 mm were excluded. Additionally, nodules annotated by at least three radiologists were sampled; each nodule's malignancy was rated from 1 to 5 where 1 represents low malignancy and 5 high malignancy. Nodules with an average score of 3 were discarded due to uncertainty in radiologists' determination. There were 1,100 nodules satisfying the criteria, consisting of 450 benign and 650 malignant for classification. Among the images, 80% were divided into the training set, and the rest were used for validation.

III. METHOD

A. Transfer Learning Based on GoogLeNet

In recent years, deep CNNs have witnessed tremendous success in the field of visual imagery analysis. The ILSVRC Imagenet competition is one of the largest and most competitive computer vision challenges worldwide. Google's entry into Imagenet 2014 - GoogLeNet - achieved 93.33% accuracy in the classification of daily objects [7]. Inspired by its results, a similar CNN structure was built using a transfer learning approach. In particular, fine-tuning, a sub-division of transfer learning, was employed due to the small amount of labelled target data. Furthermore, in order to avoid over-fitting, a relatively conservative learning strategy was applied. Most of the parameters in the CNNs of both the target and the source were designed to be as similar as possible, and fine-tuned only parameters in some layers, while most others untreated.

The classification tasks of daily objects and that of lung nodules are very similar in terms of some low-level features, such as border and color information. Therefore, using pre-trained parameters enables my network to go straight into high-level feature detection, which greatly saves its computational budget. The functionality of GoogLeNet's powerful feature detection is also inherited by the proposed network to accurately detect lung cancer.

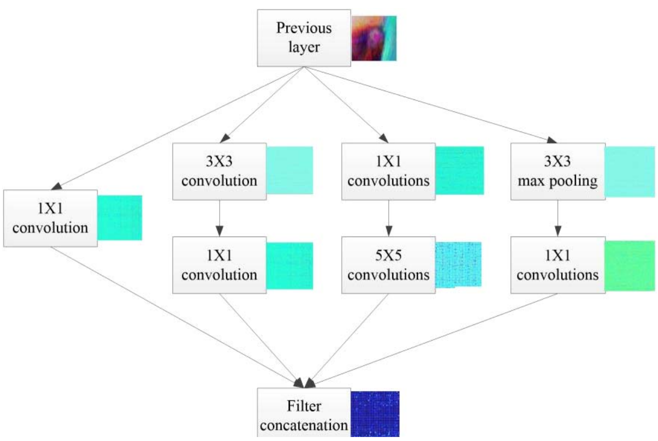

Figure 1. A demonstration of the inception module.

The proposed network consists of 22 layers in total, including 9 inception modules [5] and 3 softmax layers. In particular, inception modules are embedded to increase the network depth and width by processing data through several convolutions simultaneously.

One module comprises four 1*1 convolutions, one 3*3 convolution, one 5*5 convolution, and one 3*3 max pooling. Data from the previous layer are divided into four batches, and again concatenated into one batch upon exiting the inception module. In sharp contrast to conventional networks that process data through convolutional layers consecutively, inception modules not only shorten the network's training time but also increase the utilization of computing resources.

B. Median Intensity Projections

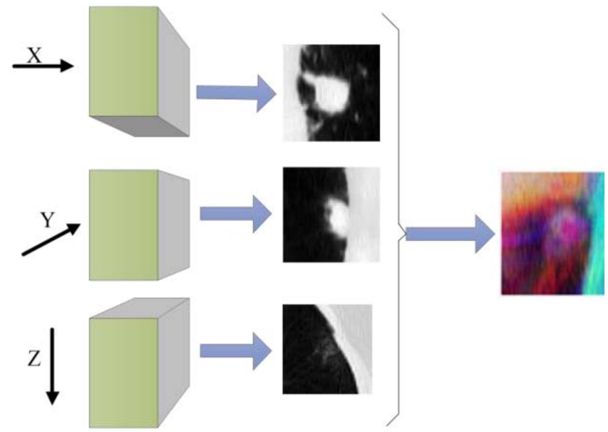

Figure 2. An overview of MIP.

Noticeably, it is impractical to input three-dimensional CT scans into the GoogLeNet-based network which processes only three channels. To preserve the integrity of three-dimensional information, Multi-view CT scans were combined into one image using MIP. CT scans with distances of coordinates smaller than 5 were classified into one group to represent the same nodule. The images of three distinct.views across the x, y, and z axes of the same nodule were input into one three-dimensional matrix, where their MIPs were calculated according to Formula 1.

$$\begin{aligned} \varphi(y,z) &= \underset{x}{med}[I(x,y,z)], \\ \varphi(y,z) &= \underset{y}{med}[I(x,y,z)], \\ \varphi(y,z) &= \underset{z}{med}[I(x,y,z)], \end{aligned} \tag{1}$$

The median projected images generated from all three axes were then concatenated into three respective channels of one synthesized RGB image. Finally, bilinear interpolation was applied to enhance the image. This method makes good use of CT scans' monochromatic characteristic to include multi-view information in one image, and saves the time and energy of building another task-specific three-dimensional CNN that might lack sufficient training.

C. Data Augmentation

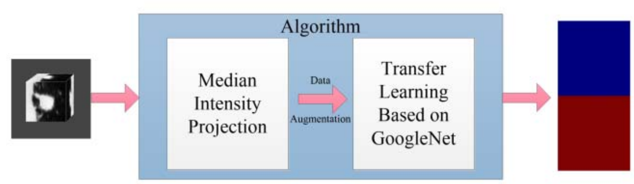

Figure 3. An overview of the proposed system. First, MIP images are generated, on which data augmentation is performed. The proposed CNN is trained with these images. Second, the determination of lung nodule malignancy is outputted.

Due to insufficient example availability, additional training samples were generated from the input images. Saltpepper noise, Gaussian noise, random rotation and median filter were added to the input images, which augmented the training set by 50 times. Prior to being input into the proposed CNN, data were automatically mirrored, which, once again, doubled the set's size. Data augmentation enabled images of both benign and malignant nodules to be sufficient for network training purposes. The CNN was trained using the Interactive Deep Learning GPU Training System(DIGITS) on Ubuntu 14.05 and accelerated by a GPU GTX1080.

D. Remote Access

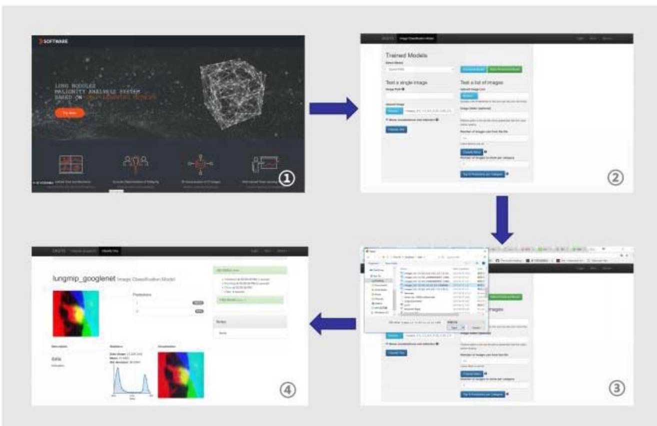

Figure 4. An overview of access to the proposed system from distanced hosts.

In order to allow long-distance management of the GPU-based automatic detection system and its future transition into a practical tool, remote access to the system was enabled through a web server. A link connecting to the server was embedded into a website through which users could access the system from remote computers. By clicking on the URL on the website, users will be redirected to the webpage for testing of the proposed system on DIGITS. Users will be able to upload eligible CT images and obtain malignancy determination within seconds.

IV. Data Analysis

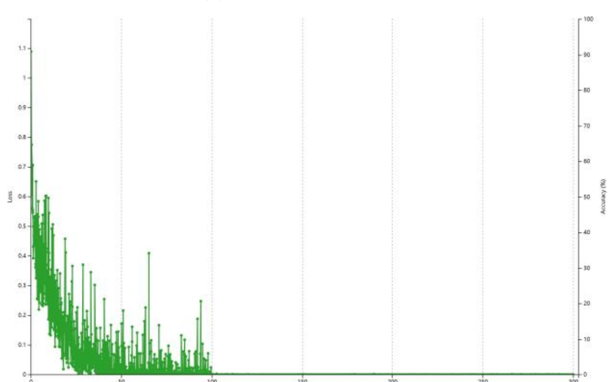

Figure 5. Loss curve converges around the x-axis after 100 epochs.

The proposed CNN was trained for approximately 100 epochs as the loss curve converged around the x-axis. Accuracy, sensitivity, and specificity of the network were recorded after epochs 30 and 300 respectively. The total training time was 14 hours and 9 minutes. Compared to other existing methods for deep learning-based lung cancer detection which utilized automated classification of lung cancer types from cytological images using deep convolutional neural networks [8], tri-layered Stacked Denoising Auto Encoder (SDAE) [2], and CNN with Least Absolute Shrinkage and Selection Operator (LASSO) [3], the proposed system produced higher accuracy after trained for only 300 epochs; MIPs and fine-tuned GoogLeNet both accounted for the system's high efficiency and its economical use of computational resources.

TABLE I. Comparison of Accuracies

Method	Accuracy
Deep CNN	71.1%
SDAE	79.29%
CNN + LASSO	79.56%
Proposed system	**81%**

TABLE II. Accuracies of the Proposed System

Accuracy	Epoch	
	Epoch #30	*Epoch #300*
Top-1 accuracy	82%	81%
Top-5 accuracy	100%	100%

The top-1 accuracies obtained at epochs #30 and #300 only showed 1.0% difference and the top-5 accuracies were both 100%, demonstrating no notable appearance of over-fitting or under-fitting. The CNN's consistent performance over a long period of time proved the high stability of the proposed method.

TABLE III. Comparison of Sensitivity and Specificity

Method	Output	
	Sensitivity	*Specificity*
Vanilla 3D CNN	59.3%	76.1%
Linear	65.2%	67.2%
GoogLeNet-based deep 3D CNN	77%	74.1%
Proposed system	84.0%	78.0%

In addition, unlike other research that employed regression algorithms to treat data, a feature training classifer - softmax - was used at the bottom of the CNN, which output not only accuracy but also medical indices such as sensitivity and specificity. As used for detection of medical images, these indices are important in determining whether the method is applicable in clinical settings. The proposed system achieved higher sensitivity and specificity than GoogLeNet-based deep three-dimensional CNN model, Vanilla three-dimensional CNN, and Linear model [9] as shown in Table 2. In other words, it produced higher accuracy in identifying sick people as sick and healthy people as healthy.

Following the completion of network training, the validation set at epoch 300 consisted of 100 lung nodule MIPs was, again, processed through the finalized CNN model (the true malignancy rating was labeled). As expected, the CNN obtained 81% top-1 accuracy, the same as that of the last training epoch. Each scan was given a possibility of malignancy, and the nodule on the scan was determined to be malignant if the possibility was higher than 0.5. The expected possibility for malignant nodules was 1, and that for benign nodules was 0. Then an accurate determination can be given by Formula 2.

$$accurate: -0.5 < expected - observed < 0.5 \quad (2)$$

If the value of observed possibility subtracted from the expected possibility is between -0.5 and 0.5, then the nodule is accurately classified, and vice versa. A t-based confidence interval was calculated to find a 95% confidence interval estimate for the population mean of this value in all possible trials. The sample size was 100. The following conditions were satisfied: 1) the sample was randomly selected; 2) the sample size was larger than 30, which means that the sampling distribution of sample mean was approximately normal. Since all conditions were met, the t-interval procedure was chosen. The 95% t-based confidence interval for the true absolute value for all possible trials was as follows:

$$\begin{aligned} & x \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = x \pm t_{0.025}(100-1)\frac{s}{\sqrt{n}} \\ & = -.040027 \pm 1.984\frac{.4270274561}{\sqrt{100}} \\ & x = -.040027 \\ & s = .4270274561 \\ & n = 100 \\ & SE_{\bar{x}} = \frac{s}{\sqrt{n}} = .04270274561 \end{aligned} \quad (3)$$

The 95% confidence interval was (-0.1248, 0.0447). Therefore, we were 95% confident that the true value of the population mean of observed possibility subtracted from the expected possibility fell between -0.1248 and 0.0447. Since the interval was constrained within the critical values and had a small standard error of the mean (SEM), we were convinced that the proposed method would produce consistent results in multiple rounds of testing.

TABLE IV. COMPARISON OF ACCURACIES FROM THE EXPERIMENT

Network	Data	
	Common 2D CTs	***MIP***
Fine-tuned Alexnet	67.6%	79%
Fine-tuned GoogLeNet	68.7%	81%

In order to elucidate the net effect of MIPs and fine-tuned GoogLeNet in lung cancer detection, and to build a solid foundation for future research, a controlled variable experiment was performed during the data analysis phase. Ceteris paribus, MIPs and common two-dimensional CT images were used to train fine-tuned GoogLeNet and Alexnet [10], respectively, forming four treatment groups. Alexnet comprises 8 layers in total, including 5 convolutional layers and 3 fully connected layers. It has less complexity than GoogLeNet due to lack of network depth and inception modules. Accuracy was recorded for each treatment group after epoch #300.

Both trained on MIPs, the accuracy of fine-tuned GoogLeNet was 2.0% higher than that of Alexnet, which verified that the complex inception module and numerous convolutions achieved better efficacy than the simple, linear structure of Alexnet. Moreover, when GoogLeNet was trained on MIPs and two-dimensional images, respectively, the former produced a 12.3% higher accuracy, showing that MIPs allowed the proposed CNN to study more lung nodule features and perform better classification. Here it is most important to note that the GoogLeNet trained on MIPs had the highest accuracy among all four treatment groups, which proves that the combined use of GoogLeNet and MIP has been successful.

V. DISCUSSION AND CONCLUSION

Unlike previous studies that trained simple task-specific CNNs or outdated pre-trained networks, in this research, GoogLeNet- one of the most advanced CNNs in visual imagery analysis- was fine-tuned and applied to the field of determining lung cancer malignancy. This significantly increased the proposed system's convergence rate and its accuracy, sensitivity, and specificity. In other words, less training time produced better output results. In addition, MIPs were applied to concatenate multi-view information of three-dimensional CT scans into RGB images that are compatible with the fine-tuned GoogLeNet. MIPs enabled the proposed system to integrally learn features of malignant and benign lung nodules during the training, and obtain high accuracy when tested on the validation sets.

As shown in the data analysis, the proposed system outperformed many available programs of lung cancer detection in accuracy, sensitivity, and specificity. Furthermore, the t-interval constructed based on validation accuracy indicated that the system would produce consistent results in multiple rounds of testing. The net effects of the system's two core factors - fine-tuned GoogLeNet and MIPs - were revealed via a controlled variable experiment; this opened up space for future studies of these two factors, which have the potential of achieving even higher accuracy in lung cancer detection.

Finally, remote access to the GPU-based system was enabled through a web server; this allowed long-distance management of the detection system and its transition into a practical tool. In the future, testing data can also be collected through this channel for adjustment of the system.

REFERENCES

[1] U.S. National Institutes of Health. National Cancer Institute. SEER Cancer Statistics Review, 1975-2013.

[2] Tourassi, G. D., Armato, S. G., Sun, W., Zheng, B., & Qian, W. (2016). Computer aided lung cancer diagnosis with deep learning algorithms. Proceedings of SPIE.

[3] Hussein, Sarfaraz & Gillies, Robert & Cao, Kunlin & Song, Qi & Bagci, Ulas (2017). TumorNet: Lung Nodule Characterization Using Multi-View Convolutional Neural Network with Gaussian Process.

[4] Hussein, Sarfaraz & Cao, Kunlin & Song, Qi & Bagci, Ulas (2017). Risk Stratification of Lung Nodules Using 3D CNN-Based Multi-task Learning. 249-260. 10.1007/978-3-319-59050-9_20.

[5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions, computer vision and pattern recognition.

[6] Clark, K. W., Vendt, B. A., Smith, K. E., Freymann, J., Kirby, J., Koppel, P., ... & Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. Journal of Digital Imaging, 26(6), 1045-1057.

[7] Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.

[8] Teramoto, Atsushi & Tsukamoto, Tetsuya & Kiriyama, Yuka & Fujita, Hiroshi (2017). Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks. BioMed Research International. 2017. 1-6. 10.1155/2017/4067832.

[9] Chon, A., Balachandar, N., Lu, P., Deep Convolutional Neural Networks for Lung Cancer Detection. (2017).

[10] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. neural information processing systems.