# Research on the Accurate Detection System Based on the Convolutional Neural Network

**Xiangli Kong**

*Cannon School, Concord, NC, USA*
*kxlleo@icloud.com*

**ABSTRACT.** *The purpose of this project is trying to detect tumors by computer based on deep learning techniques when a picture of a tumor is shown. In this research, a fast and accurate colon cancer detection is proposed, which means this research can dramatically increase the speed of diagnosis and can also improve the accuracy of confirming a diagnosis. During the experiment, a Convolutional Neural Network (CNN) structure akin to that of VGG Net and ResNet was built. A GPU computer with two 2080 Ti GPUs was used for training. The result of training produced 94% accuracy with a loss lower than 10%. Respectively, this result improved over 10% of accuracy compared to the detection by human eyes. Lastly, this program can be used by any computer to predict the tumor, which allows it transits to a practical tool in the future.*

**KEYWORDS:** *Accurate, Convolutional Neural Network(CNN)*

## 1. Introduction

Colorectal Cancer is the third most common cancer in the US. Approximately 50,630 people died because of colon cancer in 2018, and there were 140,250 new cases of colon cancer in 2018 [1]. A more efficient and accurate detection of colon cancer can lead to improving better treatment and higher survival rates. Moreover, such innovation can save an amount of valuable time, which means saving countless lives.

In the present, researchers introduced a technique of deep learning area into colon cancer diagnosis, which promises to improve accuracy and speed in determining tumor. Computers are trained to identify the features of tumors and distinguish tumors from other organizations via processing tumor images taken by colonoscopy. In addition, images are processed through Convolutional Neural Network (CNN); this research will use VGG Net, as well as ResNet, to train the data. Furthermore, neither of them were used in clinical trials, which provides an

excellent chance to the research of a fast, accurate deep learning-based tumor detection.

The system solves the problem of low efficiency and accuracy to detect tumors by human.Create the MiniVGGNet based on VGGNet which can highly classifar the colon tumor and other normal parts.

## 2. Method

### *2.1 Deep Learning based on Convolutional Neural Network(CNN)*

Recently, CNNs have shown tremendous success in fields of visual image analysis [2]. VGGNet, ResNet, and GoogleNet are remarkable neural networks. VGGNet was the runner-up at the ILSVRC 2014 competition, which is one of the most competitive and largest computer vision challenges worldwide. VGGNet consists of 16 convolutional layers and is very appealing because of its very uniform architecture [3]. The layers are listed as follows:

| ConvNet Configuration | | | | | |
|---|---|---|---|---|---|
| A | A-LRN | B | C | D | E |
| 11 weight layers | 11 weight layers | 13 weight layers | 16 weight layers | 16 weight layers | 19 weight layers |
| input (224 × 224 RGB image) | | | | | |
| conv3-64 | conv3-64 **LRN** | conv3-64 **conv3-64** | conv3-64 conv3-64 | conv3-64 conv3-64 | conv3-64 conv3-64 |
| maxpool | | | | | |
| conv3-128 | conv3-128 | conv3-128 **conv3-128** | conv3-128 conv3-128 | conv3-128 conv3-128 | conv3-128 conv3-128 |
| maxpool | | | | | |
| conv3-256 conv3-256 | conv3-256 conv3-256 | conv3-256 conv3-256 | conv3-256 conv3-256 **conv1-256** | conv3-256 conv3-256 **conv3-256** | conv3-256 conv3-256 conv3-256 **conv3-256** |
| maxpool | | | | | |
| conv3-512 conv3-512 | conv3-512 conv3-512 | conv3-512 conv3-512 | conv3-512 conv3-512 **conv1-512** | conv3-512 conv3-512 **conv3-512** | conv3-512 conv3-512 conv3-512 **conv3-512** |
| maxpool | | | | | |
| conv3-512 conv3-512 | conv3-512 conv3-512 | conv3-512 conv3-512 | conv3-512 conv3-512 **conv1-512** | conv3-512 conv3-512 **conv3-512** | conv3-512 conv3-512 conv3-512 **conv3-512** |
| maxpool | | | | | |
| FC-4096 | | | | | |
| FC-4096 | | | | | |
| FC-1000 | | | | | |
| soft-max | | | | | |

*Figure. 1 Layers of VGGNet*

GoogleNet was also introduced at the ILSVRC 2014, and GoogleNet was the winner of the competition. Their architecture consisted of a 22 layer deep CNN but reduced the number of parameters from 60 million to 4 million [3].



*Figure. 2 Introduction to GoogleNet*

ResNet was introduced at the ILSVRC 2015, which featured an architecture with "skip connections" and featured heavy batch normalization. It is successful because with such a skip connection, it can train a neural network with 152 layers while still have a low complexity [4]. ResNet works are shown as follows:
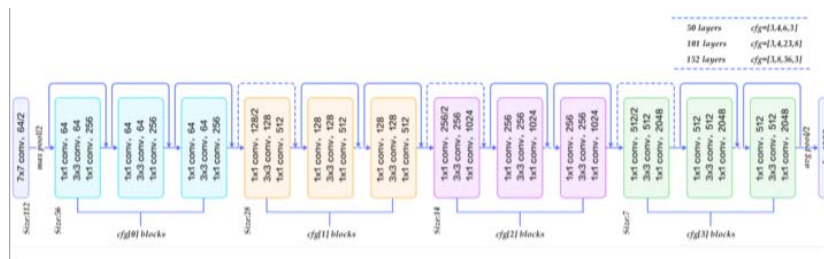


*Figure. 3 Introduction to ResNet*

Because of the size of the dataset, I made few changes to VGGNet. I decrease the convolutional layers from 16 layers to 11 layers for better fitting. I name it Mini VGGNet the changing of the VGGNet is shown below:

```
Layer (type)                   Output Shape         Param #    Connected to
==================================================================================================
conv2d_47_input (InputLayer)   (None, 64, 64, 3)    0

lambda_11 (Lambda)             (None, 64, 64, 3)    0          conv2d_47_input[0][0]

lambda_12 (Lambda)             (None, 64, 64, 3)    0          conv2d_47_input[0][0]

sequential_2 (Sequential)      (None, 8)            4628488    lambda_11[0][0]
                                                               lambda_12[0][0]

activation_16 (Concatenate)    (None, 8)            0          sequential_2[1][0]
                                                               sequential_2[2][0]
==================================================================================================
Total params: 4,628,488
Trainable params: 4,626,376
Non-trainable params: 2,112
```

*Figure. 4 MiniVGGNet*

### 2.2 Import Images to Data

The images in the dataset were taken by colonoscope. A colonoscope is a long, flexible, tubular instrument about 1/2-inch in diameter that transmits an image of the lining of the colon so the doctor can examine it for any abnormalities.
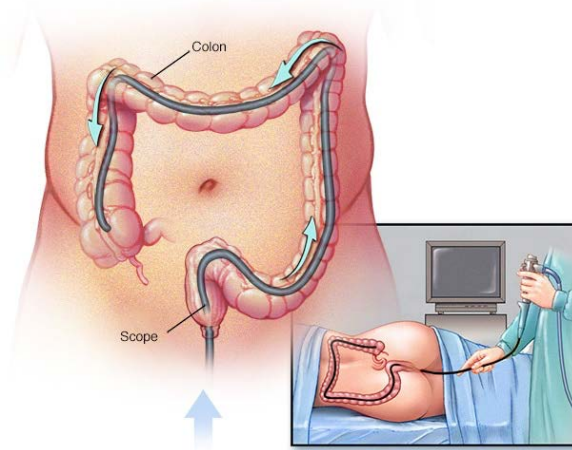
*Figure. 5 Colonoscopy*

After taking images by a colonoscope, the images can import to the computer. Then the CNN can process images. The computer sees the input images as an array of pixels. It takes an image and classifies it under certain categories to do image recognition. Each image will pass through a series of convolutional layers. The figure below is a complete flow of CNN to process an input image and classifies the objects based on values.
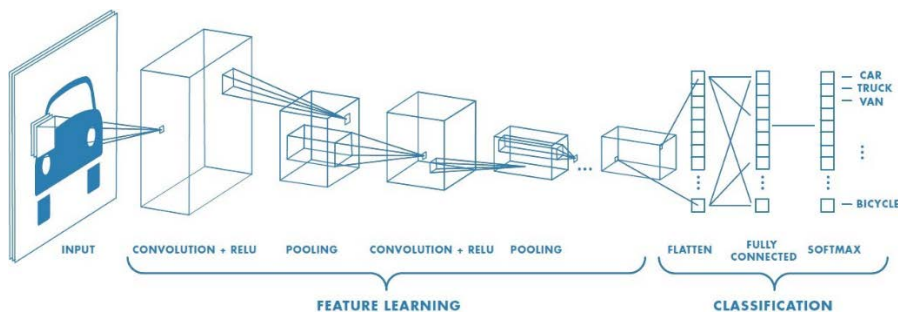
*Figure. 6 Neural Network with Many Convolutional Layers*

Convolution is the first stage of CNN, which extracts features from an input image. Convolution uses small squares to learn features and preserves the relationship between pixels.

### 2.3 Mathematical Description

CNNs are built based on the mathematical expression. The mathematical expression of CNN is integrating two functions to a specific output. Here is how convolution works:

S(i,j) = (I*K)(I,j)

$$= \sum_m \sum_n I(m,n)K(i-m,j-n)$$

$$= \sum_m \sum_n I(i-m,j-n)K(m,n)$$

The primary way to improve data accuracy is to prevent overfitting, which is a phenomenon where a machine learning model models the training data too well but fails to perform well on the testing data. It typically shows the decadence of accuracy after reaching a maximum accuracy in one epoch.

One of the ways to prevent overfitting is using regularization, which is the process of regularizing the parameters that constrain, regularizes, or shrinks the coefficient estimates towards zero [5]. The overfitting occurs because the graph of the curve has a larger coefficient, the regularization is used to add weight so the curve can be smoother. The mathematical expression is shown below:

$$J = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(X^{(i)}) - y^{(i)})^2$$

The regularization can not only minimize the cost function but also restrict the parameter not to become too large. The term is shown below:

$$J = \frac{1}{2m}\sum_{i=1}^{m}(h\theta(X^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

The other way to prevent overfitting is using early stopping, which guides as to how many iterations can be ran before the learner begins to over-fit. When the accuracy begins to decrease dramatically, early stopping will stop training and report the highest accuracy data. The mathematical term of early stopping is:

$$f_\rho(x) = \int_X^Y y d\rho(y|x), x \in X$$

The other significant way to improve the accuracy is using data augmentation. Data augmentation is used for enhance the quality of the dataset by flipping the

images. One image can turn out to be four images by flipping 90 degrees each time. It can essentially improve the abundance of the dataset.
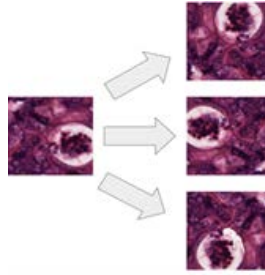


*Figure. 7 Data augmentation*

### 2.4 Overall Process

The overall process of this project is first importing the dataset into neural networks. Secondly, the networks will train the computer by using the dataset. After training, test images will be imported, and the prediction will be shown.
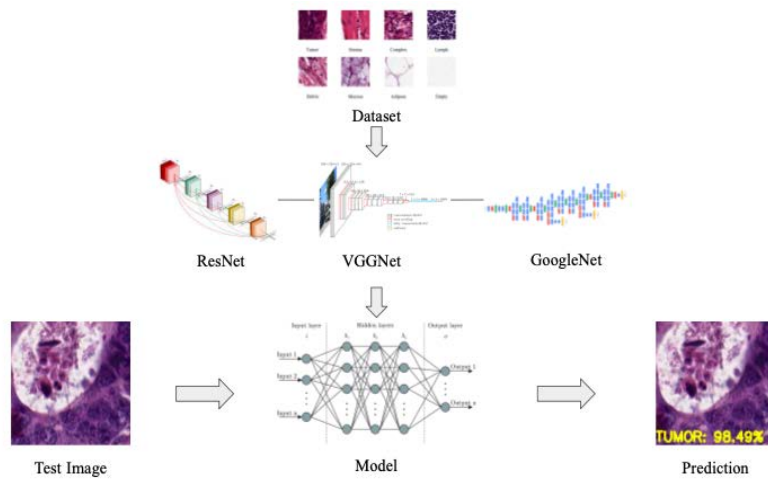


*Figure. 8 Overall Process*

### 2.5 Optimization

Using different optimizers can change the results. SGD, stochastic gradient descent, is one of the common optimizations used in CNN [6]. This project uses SGD for a better result with smoother graphs.
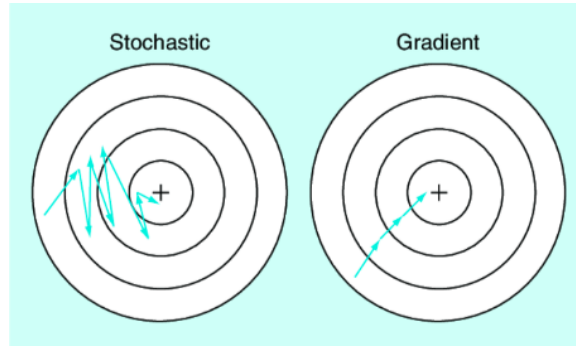
*Figure. 9 Stochastic Gradient Descent*

The other optimizer this project chooses is AdaDelta; it works by the equation shown below [7]:

$$\Delta x_t = \frac{\eta}{\sqrt{\sum_{i=1}^{t} g_i^2}} * g_t$$

$$x_t = x_{t-1} - \Delta x_t$$

After applying both optimizers, the AdaDelta shows a better result, so we choose AdaDelta for better effects.


## 3. Data Analysis

Various neural networks, including VGGNet, ResNet, and GoogleNet, are tested in the project to figure out one network that can provide the best performance.

*Table 1 Result of Networks*

| Network | Validation Accuracy | Validation Loss |
|---|---|---|
| ResNet50 | 81% | 4.2 |
| ResNet101 | 84% | 8.5 |
| GoogleNet | 77% | 2.3 |
| VGGNet | 87% | 1.1 |
| MiniVGGNet | 92% | 0.2 |

After comparing the results from VGGNet, ResNet, and GoogleNet, VGGNet shows the best prediction. The reason for VGGNet shows the best result is that the project dataset is small. Comparing to the other two networks, VGGNet has fewer convolutional layers. It works better for a small amount of data. Networks like

ResNet and GoogleNet will likely show a prediction of high loss in a small dataset. More convolutional layers also mean there is a higher chance for overfitting to occur.
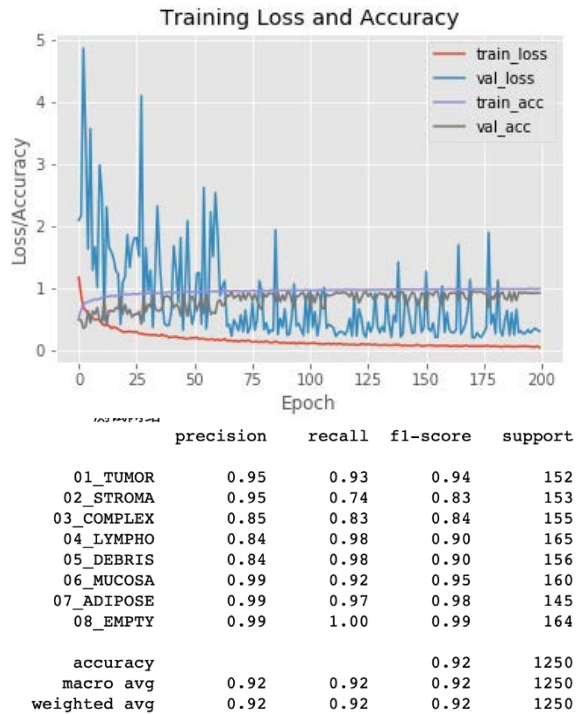


|  | precision | recall | f1-score | support |
|---|---|---|---|---|
| 01_TUMOR | 0.95 | 0.93 | 0.94 | 152 |
| 02_STROMA | 0.95 | 0.74 | 0.83 | 153 |
| 03_COMPLEX | 0.85 | 0.83 | 0.84 | 155 |
| 04_LYMPHO | 0.84 | 0.98 | 0.90 | 165 |
| 05_DEBRIS | 0.84 | 0.98 | 0.90 | 156 |
| 06_MUCOSA | 0.99 | 0.92 | 0.95 | 160 |
| 07_ADIPOSE | 0.99 | 0.97 | 0.98 | 145 |
| 08_EMPTY | 0.99 | 1.00 | 0.99 | 164 |
| accuracy |  |  | 0.92 | 1250 |
| macro avg | 0.92 | 0.92 | 0.92 | 1250 |
| weighted avg | 0.92 | 0.92 | 0.92 | 1250 |

*Figure. 10 Training and Test Results*

After understanding the reason why VGGNet wins the race, there are a few changes made to the project. I decrease the layers of VGGNet from 16 layers to 11 layers, named mini VGGNet. This mini VGGNet shows the perfect fitting to the dataset; the final result shows the highest accuracy of this project, which is 92 percent with a loss of 0.2.

## 4. Conclusion

In summary, this paper introduces a convolutional neural network that can classify tumors and other organizations from images. Differ from the traditional way of detecting tumors by human eyes; this network provides a more accurate and efficient way to diagnose colon cancer. Moreover, this project also makes improvements to the original networks to enhance performance. A VGGNet aided CNN is proved to be the best classifier of this project. Also, this network has a great prospect in the medical field.

**References**

[1] Cheng L., Eng C., Nieman L.Z., Kapadia A.S., Du X.L.Trends in colorectal cancer incidence by anatomic site and disease stage in the United States from 1976 to 2005. Am. J. Clin. Oncol. 2011; 34: 573–580. doi: 10.1097/COC. 0b013e3181fe41ed.

[2] He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 346–361. Springer, Heidelberg (2014).

[3] Deep Learning. MIT Press. Chapter 7. Chapter 9.

[4] Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning (2016).

[5] J. Moody. (1991). Note on Generalization, Regularization, and Architecture Selection in Nonlinear Learning Systems. In B.H. Juang, S.Y. Kung, and C.A. Kamm, editors, *Neural Networks for Signal Processing,* IEEE Press, Piscataway, NJ.

[6] BORDES. A., BOTTOU, L., and GALLINARI, P. (2009): SGD-QN: Careful Quasi-Newton Stochastic Gradient Descent. *Journal of Machine Learning Research, 10:1737-1754*. With Erratum (to appear).

[7] D. García-Lorenzo, S. Francis, S. Narayanan, D. L. Arnold, D. Louis Collins, "Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging," Med. *Image Anal.*, vol. 17, no. 1, pp. 1-18, 2013.