

# The Study of Convid-19 Recognition and Performance Optimization Based on AlexNet

Yangyu Huang\*

Dept. of Mechanical Engineering, North Carolina State University, Oval Drive, Raleigh, NC, USA,  
27606

\* Corresponding author: yhuang54@ncsu.edu

**Abstract.** Corona Virus Disease 19 (COVID-19), a severe disease that killed millions of human's lives. It has been found a quick and highly efficient way to classify whether a people get COVID-19 or not. Computer Vision (CV) is used to scan the chest X-ray image and help identify patient's condition. In this paper, one of deep learning algorithm - AlexNet provides several hyper-parameters to increase the system. Four main hyper-parameters are modified to increase the accuracy: size of input data, epochs, learning rate and Batch Normalization (BN) layer. In the first experiment, compare the accuracy of two input data size. Resizing is a crucial process in this test. Since the default image size of AlexNet is 256, it is required to change the size in both training and test parts. Then, three learning rate are tested to find which one is the steadiest one. ReduceLROnPlateau (RLROP) is one of the best learning rate methods to AlexNet. It can modify the number every time if the accuracy doesn't increase. BN layer gives a surprised result to the classification report: it has a negative effect to the system and has been denied quickly. During the two experiments, the value of epochs is constant. So, it is easy to find the best epoch number. The consequence of classification report shows that a smaller image size without BN layers gives a highest accuracy and has a highly efficient and steady system.

**Keywords:** COVID-19, computer vision, AlexNet, Batch Normalization, learning rate.

## 1. Introduction

In last three years, Corona Virus Disease 19 (COVID-19) had a severe effect on people's live. It became increasingly crucial for people to understand whether they got COVID-19 or not. At the beginning, doctor and nurse did nuclei acid test for the patients and check the answer one by one. It took a long time to achieve the outcome, and it is not very efficient. Only a limited number of people had the opportunity to take the test. In another side, doctor and nurse did not have any time to take a rest. It is essential to find a more accurate and high efficiency way to identify patient's condition. In this case, computer vision was introduced into this field. Because there are a huge lung X-ray images as dataset, scientist can test computer got learn how to classify different X-ray image to tell people whether they got COVID-19 or not. In the aspect of machine learning, Convolutional Neural Network (CNN) has a high accuracy to distinguish between different images. In this paper, AlexNet- one of CNN was used to classify the image.

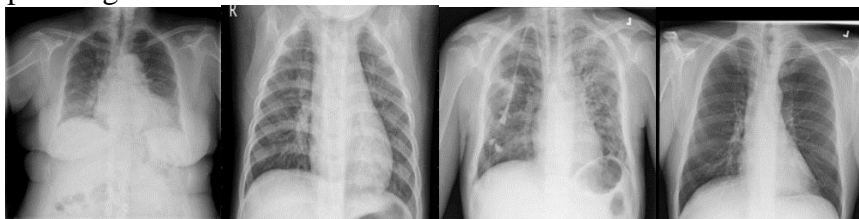
In recent research, deep learning algorithm was used to increase the efficiency of COVID-19 classification. It might be true that compare different algorithm can help find the highest accuracy, focus on one structure can find the best factor value. For example, there is an article that discussed generative adversarial networks (GANs) and compared different structures to find the best method [1]. It compared structures like VGG-16, ResNet-50, DenseNet-192 and DenseNet-121 to find which one has a useful result. In this paper, only one CNN structure was discovered to find the highest accuracy and classification efficiency. Tuning several hyper parameters to achieve a better process but not use several structure and compare their accuracy. In past research, CXR and CT image were most used to be the data source. However, the different composition of CXR and CT image may affect the machine to do the classification. In [2], the author used CXR and CT image as the input data to classify and find a high accuracy. It gave a perfect result but the report is not pure. The different image labels mislead the machine to do the classification.

This paper focuses on AlexNet structure and tunes more hyper parameters to try to obtain a better performance. There are four groups of hyper parameters were discussed in this project: size of input data, epochs, learning rate and loss function. This study would only look at one structure and find the highest accuracy in this structure. Compared to other article which introduce several algorithms, the effect of different hyper parameter can be told through the paper. For example: the increase of learning rate number will decrease the precision. The smaller picture size reflects a greater accuracy.

## 2. Method

### 2.1. Dataset description and preprocessing

In this project, COVID-19 Radiography Database which provided by the database in Kaggle was used. There are 4 categories in this experiment, namely: COVID-19 positive case, normal, lung opacity and viral pneumonia. This chest X-ray (CXR) image database was created to classify COVID-19 positive cases along with normal and viral pneumonia images [3]. During the first collection, it gathered 219 COVID-19, 1341 normal, and 1345 viral pneumonia CXR image. Then the database was updated and 1200 COVID-19 CXR was collected. In the second update, a new label— lung opacity (Non-COVID lung infection) was added into the database [3]. There were 3616 COVID-19 positive cases, 10192 normal, 6012 lung opacity and 1345 viral pneumonia CXR. The total in sample number is 21165 and the size of image is 299x299. The color of CXR image is black and white. Figure 1 presents the sample images.



**Figure 1.** From left to right side, the images show the chest X-ray of COVID-19, Viral pneumonia, Lung opacity and Normal patients.

### 2.2. Proposed approach

The Deep Learning algorithm used in this paper is CNN, a convolutional calculation and feedforward neural network which is widely employed in many tasks e.g. automatic driving and medical images analysis [4, 5]. It has the ability of representation learning and use different layers to shift-invariant classify the input data [6]. CNN contains three main layers: convolutional layer, pooling layer and fully-connected layer. There are several convolutional kernels in a convolutional layer. The elements that compose a kernel correspond to a weight coefficient and a bias vector. The convolutional layer will scan the input data and add deviation value to different features. The pooling layer used in this project is max pooling layer. Max pooling layer helps pick the highest value in each part and decrease the size of image. Normally, the size of pooling layer is 2x2. This means the highest value in each 2x2 matrix will be extracted to pass the max pooling layer and constitute a smaller data. The fully-connected layer is always added to the last part of the CNN model. It will only launch signal to other fully-connected layer and spread the features figure. In this project, a new version of AlexNet structure was used to do the calculations. Originally, an AlexNet structure has thirteen layers: five convolutional layers, three fully connected layer, three max pooling layer and two dropout layers [7]. The modified AlexNet structure has fifteen layers total. There are five convolutional layers, three fully connected layers, three max pooling layers, two dropout layers and two batch normalization layers. The first convolutional a layer has a size of 11x11, and a max pooling layer and a BN layer were followed with. The second part consist of one convolutional layer, one max pooling layer and one BN layer. The size of convolutional layer is 5x5. After that, there are three convolutional layers with a size of 3x3. The input image was shrunk to the size of 256x256 and pass to convolutional layer.

BN layer changed the color range from 0~256 to 0~1. A flatten layer followed to reshape the image. The image looked like a 1D number now. This step made the machine to easily do the classification. Three dense layers were added to the system. Two of them were used to activate Rectified Linear Unit (ReLU) function [8], and the other one used to activate softmax function [9].

### 2.3. Implementation details

The learning rate method used in this project is ReduceLRonPlateau (RLROP) [10]. The initial learning rate was 0.001. In every three classifications, the learning rate would decrease 0.5 percent if the accuracy does not increase. This method could modify the learning rate all the time and gave a better accuracy. Tensorflow and Keras framework was used in this project. Batch size is 32 and epoch number is 30. Adam optimizer was applied into the system.

The cross-entropy loss function can support each neuron to use different features to classify different categories. In addition, the cross entropy has the ability to slow down learning rates' decrease rate, which means neuron's learning ability has been increase with the use of cross entropy. It is important that the initial input data has a giant difference, the more different characters it has, the easier for the neuron to classify it and learn to how to distinguish them. This will create a higher accuracy in the test part.

$$C = (-\frac{1}{n}) \sum_x [y \ln(\alpha) + (1 - y) \ln(1 - \alpha)] \tag{1}$$

### 3. Result and discussion

**Table 1.** Classification report of all test sets

Test Sets	Precision	Recall	F1-Score	Support	Accuracy
256x256 with BN	0.2	0.365	0.235	2120	0.48
256x256 Without BN	0.885	0.86	0.87	2120	0.87
128x128 With BN	0.905	0.895	0.905	2120	0.89
128x128 Without BN	0.915	0.90	0.905	2120	0.90
256x256 Learning rate 0.05	0.775	0.81	0.825	2120	0.835
128x128 Learning rate 0.05	0.79	0.85	0.85	2120	0.845

**Table 2.** Classification report of 128x128 input image size without BN layer

Labels	Precision	Recall	F1-Score	Support
Lung-Opacity	0.88	0.88	0.88	530
Normal	0.89	0.93	0.91	530
Viral-pneumonia	0.98	0.94	0.96	530
Lung-Opacity	0.88	0.88	0.88	530

The result showed that the highest accuracy can achieve to 90 percent. This meant 1908 chest X-ray image can be correctly classified in each test. The classification report showed that batch size=32, epoch=30, learning rate=0.001 and image size=128. In Table 1, the test with 256x256 input image size with BN layer has the lowest value in precision, recall number, F1-score and accuracy. The result with 128x128 input image performed better than the result of 128x128 ones'. Learning rate of 0.05 decrease all the values. In Table 2, the viral-pneumonia test set has the highest accuracy, recall number and F1-score.

From the Table 1, there were three learning rate number that were used in experiment: 0.001, 0.000142 and 0.05. However, it has limited effect to the accuracy. The reason is the learning rate was modified in each epoch. RLROP would decrease the value of learning rate when there was no increase in accuracy. BN layer is another point that might affected result. The test told that BN layer had a negative effect to the classification, which means experiment with two BN layer had a lower accuracy compared to the one that did not use BN layer. It seems that the dataset does not match to the normalization method. So the two BN layer actually did no work in this AlexNet structure. The surprised effect is the resize of image size. Normally, a bigger size of input image can help machine scan the whole frame better and create a higher classification report. Combine to the use of BN layer, it should perform the highest experiment result in this project. But, the data in Table 1 revealed an impressed answer: the input image size of 128x128 without BN layer has the highest accuracy in all the test. The reason leads to this result is because the resize images are easier for the machine to classify and find the features in these four categories. In another way, the velocity of each neuron to learn to how to classify the different chest X-ray images in bigger than the velocity with normal size images.

#### 4. Conclusion

In this project, the number of learning rate, batch size, input image size and BN layer has been changed to help find the best classification report. The experiment result beyond the expectation that a normal input image size with two BN layer should give the best result. After plenty of experiments, the input image size of 128x128 without BN layer has a better performance than the one with 256x256 input image size with two BN layers. A small adjustment to a hyper-parameter will give rise to a big change to the consequence. In the future, more hyper-parameters can be modified to improve the experiment value. Data augmentation is a good way to get some similar but useful dataset. The input image size is another point to find some unexpected answer.

#### References

- [1] Zunair, H., "Synthesis of COVID-19 chest X-rays using unpaired image-to-image translation." *Social network analysis and mining* 11.1 (2021): 1-12.
- [2] Aggarwal, P., et al. "COVID-19 image classification using deep learning: Advances, challenges and opportunities." *Computers in Biology and Medicine* (2022): 105350.
- [3] Kaggle. "COVID-19 Radiography Database." 2022, <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>.
- [4] Wen, S., et al. "Hybrid CNN-LSTM Architecture for LiDAR Point Clouds Semantic Segmentation." *IEEE Robotics and Automation Letters* 7.3 (2022): 5811-5818.
- [5] Yu, Q., et al. "Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training." *Biomedical Signal Processing and Control* 72 (2022): 103323.
- [6] O'Shea, K., et al. "An introduction to convolutional neural networks." *arXiv preprint arXiv:1511.08458* (2015).
- [7] Krizhevsky, A., "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

- [8] Agarap, A.F., et al. "Deep learning using rectified linear units (relu)." arXiv preprint arXiv:1803.08375 (2018).
- [9] Sharma, S., et al. "Activation functions in neural networks." towards data science 6.12 (2017): 310-316.
- [10] Al-Kababji, A., et al. "Scheduling Techniques for Liver Segmentation: ReduceLRonPlateau Vs OneCycleLR." arXiv preprint arXiv:2202.06373 (2022).