Resolution versus Precision in X-ray detection of Pneumonia

Zanchen Li
Department of Biomedical Engineering
Duke University
zl195@duke.edu

Abstract

In this paper we try to explore the relationship between resolutions and precisions in X-ray detection. I design a 3 convolution layers network and run different sized images through the network and found that while recall rate of the algorithm increases with the size of the image, the precision decreases with the size of the image. I also try a different method, fourier transform and multiply with low pass filter, to blur the image and achivevd a similar result.

1 Instructions

Pneumonia is a very common disease that can be caused by bacteria or viruses. It is said to have affected "approximately 450 million people globally (7 percent of the population) and results in about 4 million deaths per year"[1]. The diagnosis of Pneumonia is typically based on chest X-rays[2]. However, the diagnosis of Pneumonia can sometimes be difficult due to the ambiguity of the picture and differences between different patients as shown in Figure 1.

normal and pneumonia.png

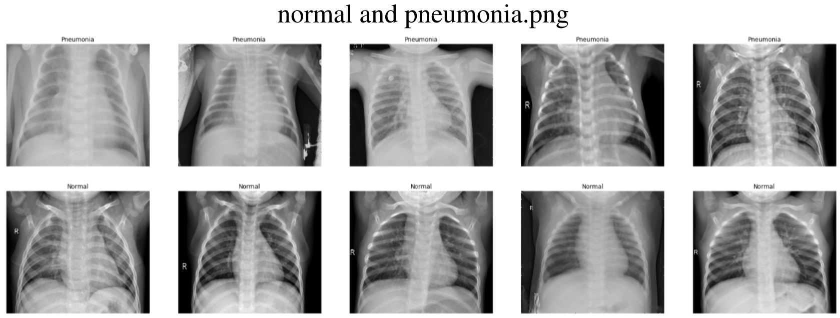


Figure 1: X-rays of normal and pneumonia patients

Consequently machine learning techniques in imaging processing has been used to assist physicians and doctors at diagnosis of the disease. The sizes of X-rays pictures are usually very big so processing them using Convolution Neural Network(CNN) can take a long time. Here I have experimented with different methods to preprocess pictures before feeding them to CNN for classification. And found that while large image size does lead to better recall rate, it also leads to lower precision.

2 Methods

The data used for the classification is acquired at [5]

`https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia`

It contains 5217 annotated training samples of normal, bacterial pneumonia and viral pneumonia(bacterial and viral pneumonia are not differentiated by the annotation).

The classification of the X-rays pictures is done using both keras and tensorflow, as others have done previously[3]. This paper's algorithm learns from other Kaggle CNN algorithms.

For the keras implementation of the problem, the data is augmented by shearing, zooming and flipping horizontally. The training, validation and test batches are generated using flow data generator. The network is consisted of three convolution layers with 32, 32 and 64 channels, each followed by a max pooling layer. At the end is two dense layers with a dropout rate of 0.5. The training takes 20 epochs. The loss function is binary cross-entropy and the optimizer used here is rmsprop optimizer. The experiment done on the keras implementation of the classification program is to change the image size when getting it from the directory, by inputting different sized images the networks will be able to learn different amount of features from the images. As shown in Figure 2 is a image resized to 50x50 during data generation.

image.png

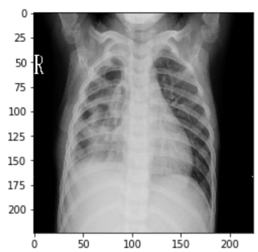


Figure 2: X-rays image of 1040*1040 resized to 50*50

For the tensorflow implementation the network used is identical to the one in keras. However, the loss function utilized here is sigmoid cross entropy. Batch size and optimizers are the same as the keras implementation of the classification program. For the tensorflow implementation, the decrease of resolution is done by applying low pass filter after fourier transform of the X-ray picture. As shown in Figure 3 is a image after a low pass filter with a diameter that is one twentieth of the width of the image. The fourier transformed image is then fetched into the network for classification.

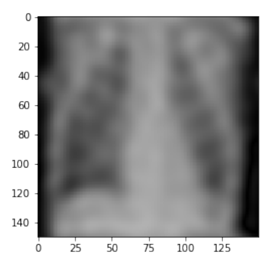
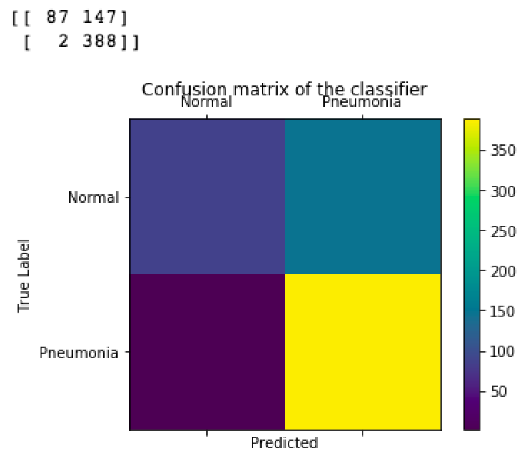


Figure 3: X-rays image after low pass filter

3 Results

The classification model achieves fairly good accuracy when the input image is 150*150(the size that is popular among Kaggle kernels) as shown in the confusion matrix in Figure 4.

From the picture we can see that while the true positive (pneumonia-pneumonia) has the highest intensity, the false positive also has a very high intensity(second highest). This means that for the result more normal people are diagnosed as pneumonia than normal people that are cleared by X-ray. The reason behind the this is because there is a imbalance of normal and pneumonia population in the database, as shown in Figure 5[3]. This imbalance leads to not enough training of the normal cases in the network and thus a low accuracy. This will result in a relatively low precision compared to a high recall, but because in medical practice recall is always more important than precision(i.e you never want to send normal healthy patient to pneumonia treatment).



matrix.png

Figure 4: Confusion matrix of the model prediction vs true labels

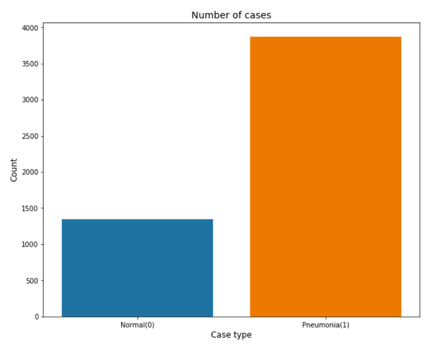


Figure 5: patient population statistics

Image sizes of 150*150,100*100,75*75 and 50*50 are tested in the experiment and their corresponding recall are shown below in Figure 6. As we can see, the recall is generally increasing with increasing image size. However the amount of increase is not significant, only 2 percent.

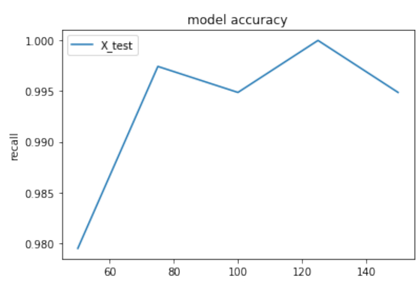


Figure 6: recall vs image size

The same images of size 150,100,75 and 50 are tested in the experiment and their corresponding precision are shown below in Figure 7. As we can see, the precision decreases as the number of image size increases. This is due to more normal cases are diagnosed as pneumonia at higher resolution.

The images resized to 50*50 are then fourier transformed and then dot multiplied by low pass filter of different diameters(1,1/2,1/4,1/8 length of the image width). Their recall and precisions are shown as below in Figure 8

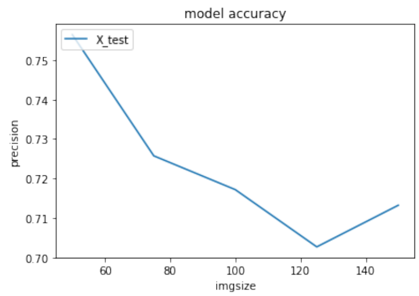
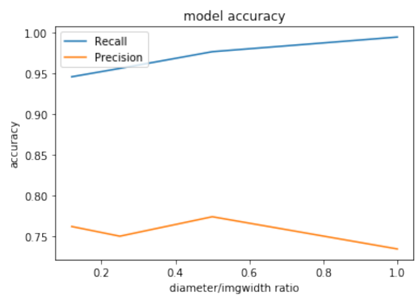


Figure 7: precision vs image size



transform stat.png

Figure 8: recall and precision plot of 50*50 image with different size lowpass filter

4 Discussion

The increase in recall as image size increases is expected, but precision decrease, unexpectedly, as image size increases. I think this might be due to the shallowness of the network. Only 3 convolution layers might not be able to extract a lot of features from a large image. However more layers and deeper networks would lead to excessive training time which the length of this class does not allow. I think in the future, the network should be redesigned to be able to extract more features from the images to increase accuracy on high resolution images.

Although recall is viewed as more valuable in medical field, the number of false positive in this model is still appalling. This is due to the lack of normal training cases compared to the abundance of pneumonia cases. Given the limitation of the data, I think I need to do more data augmentation, particularly for the normal cases, also normal training cases can be reused more often than pneumonia cases and some biases can be added to improve the precision to ultimately reach a balance with recall.

The other blur method using fourier transform achieves very similar result. I think this can be future improved by including a physical layer optimizing it.

References

[1] R. Lodha, SK. Kabra and RM. Pandey, " Antibiotics for community-acquired pneumonia in children," Cochrane Database of Systematic Reviews **6**, Art. No.:CD004874 (2013).

[2] S. Valrinder, A Satinder, "Pneumonia – Management in the Developing World",Paediatric Respiratory Reviews, **1** Volumn 12 52-59. (2011)

[3] Nain, "Beating everything with Depthwise Convolution", https://www.kaggle.com/aakashnain/beating-everything-with-depthwise-convolution#Augmentation

[4] V. Rohit, "Pneumonia Diagnosis using Lung's XRay", https://github.com/deadskull7/Pneumonia-Diagnosis-using-XRays-96-percent-Recall/blob/master/Pneumonia%20Diagnosis%20using%20Lung's%20XRay%20.ipynb

[5] Chakrabarti, S. (2019). F1000Prime recommendation of Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. F1000 - Post-publication Peer Review of the Biomedical Literature. doi:10.3410/f.732738622.793558610