Chest X-Ray Pneumonia Detection

You-Pin, Chen (310553043), Evelyne Calista(409351006)

March 25, 2022

1 Introduction

Pneumonia is an inflammatory condition of the lung affecting primarily the small air sacs known as alveoli. Symptoms typically include some combination of productive or dry cough, chest pain, fever and difficulty breathing. Diagnosis is often based on symptoms and physical examination. Chest X-ray, blood tests, and culture of the sputum may help confirm the diagnosis. Given a dataset of chest x-ray images as seen in Figure 1, our goal is to develop algorithms to automatically detect the pneumonia. There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal). Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care. For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert. The code is available in this github [link].

Figure 1: Chest X-Ray Image

2 Experiment Setups

2.1 ResNet

ResNet is a network architecture using residual block to solve the issue of gradient dispersion and accuracy degradation (training set) in deep networks. Therefore the network can get more deeper while still controlling the accuracy and speed. The gradient dispersion is happened as the number of layers is increasing, the gradient in backpropagation in the network become unstable. While as the layer is increased, the accuracy will reach into saturation at one point then degrade and this issue is not caused by overfitting. In order to solve the degradation issue, the shallower the network will perform better. By using this idea, ResNet skip the training of few layers by using skip connections or residual connections. Figure 2 showing the ResNet architecture and Figure 3 showing the single residual block used in ResNet. The layers in a residual network are learning the residual ($R(x)$), hence the name is Residual Block [1].

$$R(x) = Output|Input = H(x)|x \tag{2.1.1}$$

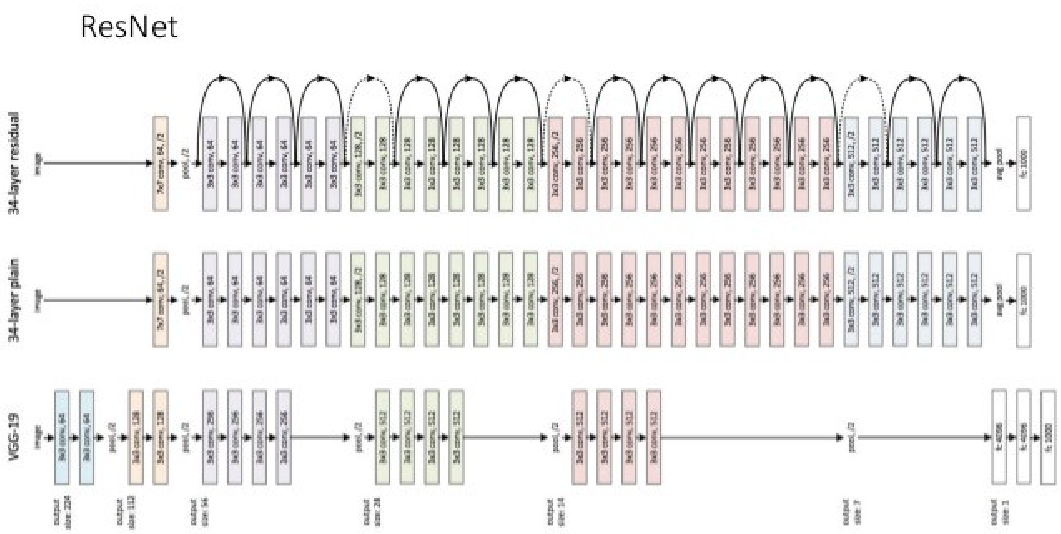


Figure 2: ResNet architecture

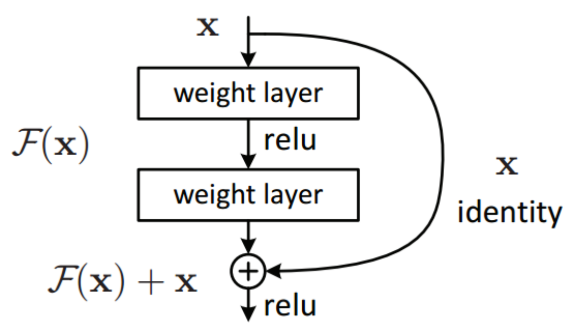


Figure 3: ResNet single residual block

2.2 Transfer learning

Transfer learning (TL) is a research problem in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks. This area of research bears some relation to the long history of psychological literature on transfer of learning, although practical ties between the two fields are limited. From the practical standpoint, reusing or transferring information from previously learned tasks for the learning of new tasks has the potential to significantly improve the sample efficiency of a reinforcement learning agent.

2.3 Weighted Cross Entropy Loss

The train data suffers from overfitting where pneumonia class 3,875 and normal class 1,341. To overcome this weighted cross entropy loss is applied. The concept is to weigh the loss computed for different samples differently based on whether they belong to the majority or the minority classes.

$$LogH(p,q) = -\omega_i \sum_{i=1}^{n} p(x_i) Log_e(q(x_i)) \tag{2.3.1}$$

where ω_i represents the weight of the loss function when the actual label of the current data is.

2.4 Augmentation

Data augmentation is the process of modifying, or augmenting a dataset with additional data. Its use in machine learning and deep learning algorithms helps improve their performance. The prediction accuracy of the models is largely reliant on the amount and the diversity of data available during training. Oftentimes, when working on specific complex tasks such as in medical images, it is hard to get large dataset. Though transfer learning techniques could be used to great effect, the challenges involved in making a pre-trained model to work for specific tasks are tough. Another way to deal with the problem of limited data is to apply different transformations on the available data to synthesize new data using data augmentation. The augmentation variation include flipping, rotation, zooming, scaling, cropping, translating, adding Gaussian noise.

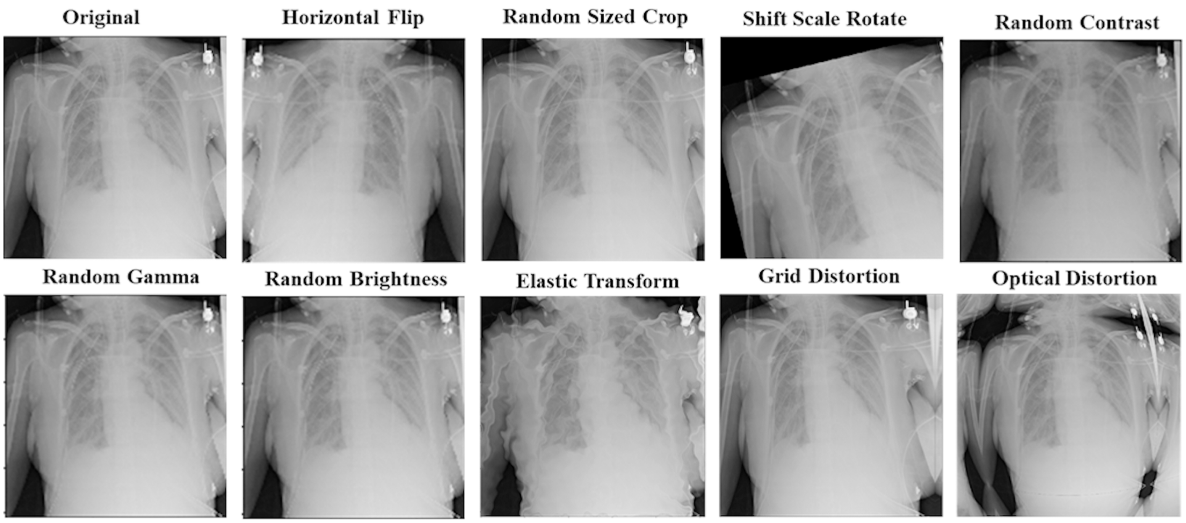


Figure 4: Example of augmented images by rotating, flipping, contrast...etc.

2.5 Datasets

There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal). Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou.

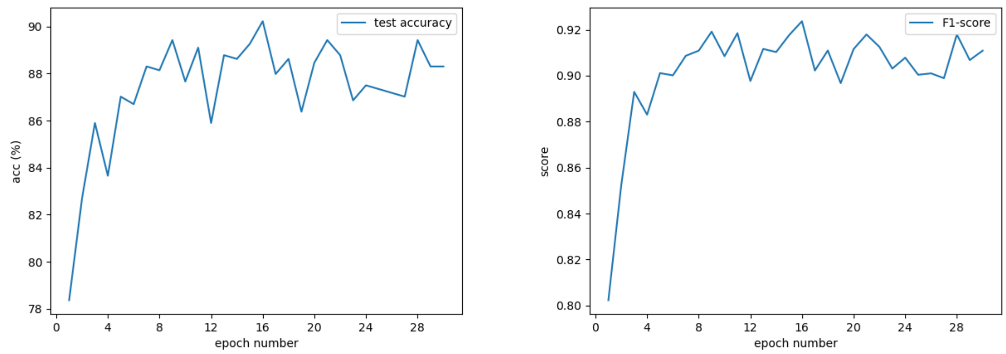
3 Experimental Results

3.1 Model Structure

Model structure affects the result directly. In our experiment, We applied ResNet18 compared with ResNet50 in Table 1. The distinction between them is on the number of hidden layer. Although, in general, more complex model lead to the best performance, but sometimes are not exactly. Since our objective is minimize cost in loss function, sizeable model will that our question make complex. Instead, lite model can be as enough powerful to settle the task. All experiments apply Cross Entropy as loss function and Adam as optimizer, learning rate is 3e05 with 30 epochs. The model do not use the pre-trained weights, the best result (Figure 5) came from ResNet18 model with accuracy 90.22%.

Model	Loss	F1-score	Acc. (%)
ResNet50	0.0115	0.8645	0.83
ResNet18	0.0045	0.9237	0.90

Table 1: Performance comparison with 2 types of ResNet structure.



(a) Accuracy curve.

(b) F1-score curve.

Figure 5: ResNet18 without pre-trained weights.

3.2 Learning Rates

Learning rate is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum of a loss function. Since it influences to what extent newly acquired information overrides old information, it metaphorically represents the speed at which a machine learning model "learns". Since we obtained the best result in previous experiment, we attempted to apply different learning rate trying to see if we can find the accurately global minimum. However, in Table 2, the results were not significant.

Model	lr	F1-score	Acc. (%)
ResNet18	1e05	0.8715	0.86
ResNet18	2e05	0.9187	0.89
ResNet18	3e05	0.9237	0.90

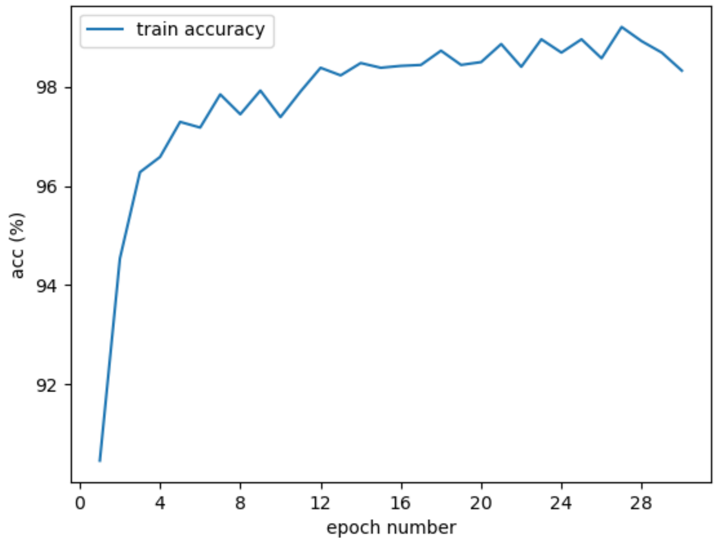
Table 2: Performance Comparison

3.3 Pre-trained

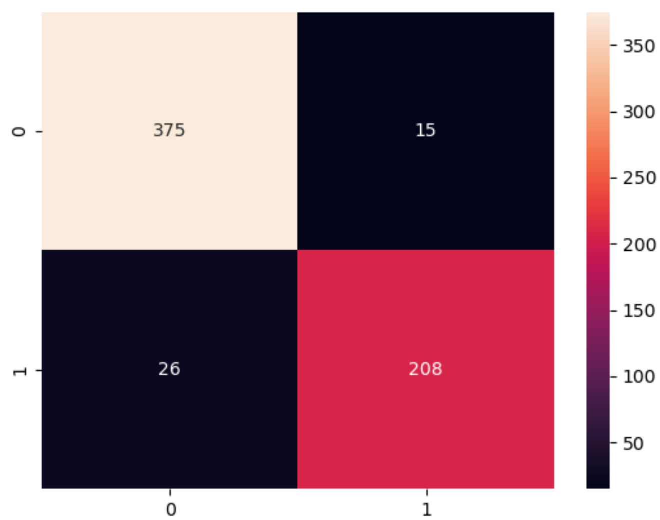
In previous works, we trained the model with our own chest X-ray data from scratch. The initialize weights were randomly created. However, we had mentioned the pros of transfer learning. Because of that, we tried to use pre-trained weights from ResNet which is trained using ImageNet datasets and fine tuning with our chest X-ray data. The outcome improved extremely huge. Figure 6 is the training process with pre-trained weights from ResNet50, the accuracy got 93.11% which break our best result.


```
Epoch:  16
Loss:  0.0011239242958142586
Training Acc. (%): 98.75%
F1-score: 0.9374
Test Acc. (%): 91.83%
100%|| 163/163 [01:22<00:00,  1.98it/s]
Epoch:  17
Loss:  0.0010272358189010386
Training Acc. (%): 98.75%
F1-score: 0.9040
Test Acc. (%): 86.86%
100%|| 163/163 [01:22<00:00,  1.97it/s]
Epoch:  18
Loss:  0.0008166820688992814
Training Acc. (%): 98.98%
F1-score: 0.9160
Test Acc. (%): 88.62%
100%|| 163/163 [01:24<00:00,  1.93it/s]
Epoch:  19
Loss:  0.0009389800065946474
Training Acc. (%): 98.91%
F1-score: 0.9456
Test Acc. (%): 93.11%
100%|| 163/163 [01:26<00:00,  1.88it/s]
```

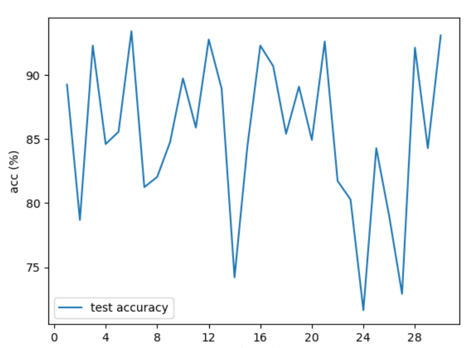
Figure 6: We got our best result with pre-trained weights.



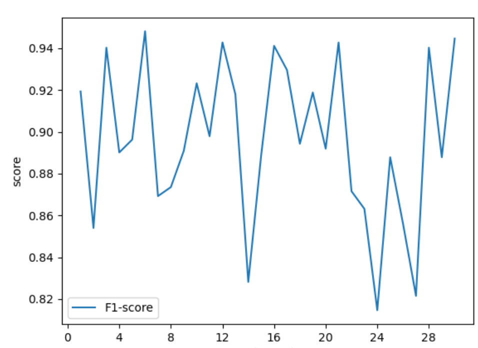
(a) Training accuracy curve.



(b) Confusion matrix.



(c) Testing Accuracy curve.



(d) F1-score curve.

Figure 7: Our best result with acc.(%): 93.11% with ResNet50, pre-trained model.

4 Discussion

4.1 Augmentation

Even though augmentation is useful to increase performance, however not the selection of augmentation shall be carefully chosen. Flipping, center crop, and random crop are not really useful in this case. For the flipping augmentation, it is bad idea since augmentation shall reflect the variations of data in real world, in this case in real world the position of lung and heart small percentage it would be flipped. For center crop, we thought that the center of image not have sufficient information of the disease feature.

4.2 Schedule learning rate

Schedule learning rate is the way to let us change the learning rate in the runtime. Since we had add exponential learning rate in our process, it did not seen the influence on minimize the loss. We guess that the main cause is because of the solution of our task are not that complex since we also got great accuracy (about 93%) without schedule learning rate.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.