

Study on the Depth of Convolutional Neural Network for Pneumonia Detection

Yunshan Nicole Qu *

College of Civil Engineering, Tongji University, 1239 Siping Road, Shanghai, 200092, China

* Corresponding author's email: 1950491@tongji.edu.cn

Abstract. Pneumonia is a dangerous but common disease to human. Due to the similar symptom between cold and pneumonia, it is essential to have an effective method to diagnose pneumonia. In this paper, a prevalent concept called computer vision was introduced to help detect pneumonia based on chest X-ray images. More specifically, a Convolutional Neural Network (CNN) was proposed to classify the images into normal class or pneumonia class. The image dataset was preprocessed by unifying, normalizing, shuffling and augmentation. To finally achieve a model with relatively reasonable and accurate prediction, four CNN models were built and compared. To keep them comparable and have certain similarity, a small architecture with one convolutional layer, one max pooling layer and one batch normalization layer was defined and piled up repeatedly. The channel numbers in the convolutional layer were increased by proportion. These four models were eventually evaluated on their accuracies both on training set and testing set and F1 scores on two categories. Experimental results indicated that the model with four convolutional layers had the best performance which achieved an accuracy of 0.8606 on test set. Also, the model had F1 scores of 0.86 on normal class and 0.88 on pneumonia class.

Keywords: Pneumonia Detection; Convolutional Neural Network; CNN.

1. Introduction

Pneumonia is a kind of respiratory inflammation happening in lung. It can result in cough and fever which is similar to the symptom of cold and flu. However, pneumonia can be extremely dangerous, especially to children, that it is responsible for 14% of under-five-year-old children's death in 2019 [1]. Early disease detection will result in less suffering and an easier recovery for the patient. Fortunately, this disease can be chiefly confirmed and distinguished based on chest X-rays or CTs. But the issue is that the diagnose on chest X-rays these days is mainly given out by the radiologists based on their medical knowledge and experience, which can be somewhat subjective.

In the past few decades, researchers called attention to computer vision that requesting computers to understand and analysis images like human do [2, 3]. When dealing with the image classification problem, distinguishing and telling the contents of the images, many typical classification algorithms are involved, e.g., random forest, decision tree, support vector machine (SVM). However, these conventional machine learning algorithms require feature extracting which can be ambiguous and omit some essential information. They might also easily come to a plateau when dealing with massive amount of data. To be specific, their accuracy keeps stable in a medium level and is tough to be further improved. Since images consist of pixels and normally have three channels when in RGB format, even one single image can involve lots of information. Now deep learning comes with a solution. Neural network is built with multiple layers for computation and has a complicated structure that unexpectedly improved the accuracy of various image processing [4]. It has massive parameters and can automatically extract features. Convolutional neural network (CNN) with its unique layers doing filtering and pooling specializes in dealing with continuous dataset like images.

Now CNN is widely applied to raise the credibility and accuracy of diagnosing pneumonia. In [5], the authors developed a Deep Convolutional Neural Network (DCNN) model to predict pneumonia and achieved an accuracy of 84%. Their research laid an emphasis on comparing DCNN with the conventional classifiers like SVM and random forest to show the advantage of DCNN. And similarly, in [6], the authors compared different deep learning model like CNN, Multi-layer Perceptron (MLP)

and Recurrent Neural Networks (RCNN) on the same pneumonia image dataset to figure out the most appropriate model. In [7], a CNN named Depth-Wise Convolution with Attention neural network (DWA) was proposed to detect pneumonia and reached a testing accuracy of 96%. In [8], a series of techniques, like single-shot detectors, squeeze-and-extinction deep convolution neural networks, augmentations and multi-task learning, were involved to figure out a high accuracy on pneumonia diagnose. However, instead of comparing CNN to other algorithms and practicing various techniques, this work would like to focus on the CNN model itself and have in-depth research on the correlation between the depth and the performance of the model.

In this regard, this work would like to design a CNN model to diagnose pneumonia with high accuracy. To achieve the best practice, the model will be adjusted by altering the depth of the network. To be specific, a small architecture with one convolutional layer, one batch normalization layer and one max pooling layer was defined. Then pile this small architecture up repeatedly with different times and get big architectures with different depths. Finally connect each big architecture to several full connection layers. The model will eventually predict and give out the possibility of having pneumonia. Models with various depths will be trained on one same dataset and the accuracy of each model will be recorded to obtain the correlation between the depth and the accuracy. The most accurate model will be accepted.

2. Method

2.1 Dataset Description and Preprocessing

The dataset used in this research consisted of a series of chest X-ray images and was downloaded from Kaggle [9]. The images are divided into three parts, namely training set, validation set and testing set, and then saved into three folders. Each folder contains two subfolders for images from two different categories: normal and pneumonia. There are 1,341 normal ones and 3,875 pneumonia ones in training set, 8 normal ones and 8 pneumonia ones in validation set and 234 normal ones and 390 pneumonia ones in test sets. The size of one single image is not unified, but all of them are grayscale images in JPEG format.

In terms of data preprocessing, it consists of four parts. First, the size of each image is converted to 224×224 in that the neural network needs a uniform input image size. Second, the matrix of each image is divided by 255 to be normalized in order to increase the speed of convergence and can raise the accuracy. Third, the images should be shuffled. Initially the dataset is split, classified into two categories and saved into different subfolders. When using the `os.listdir` function to load the images, all images are in sequence. To be specific, all normal images are loaded first and then turn to the pneumonia ones. This may cause overfitting that the model focuses too much on the normal ones in the first stage and then too much on the other class later. Finally, again, to avoid overfitting, data augmentation is applied. Images are rotated 30 degrees, shifted on width and height both in a range of 0.2, flipped horizontally, zoomed about 10% and brightened in a range from 0.4 to 1.5.

Figure 1 displays the first four images in the validation set. The first row shows the original ones, and the second row shows the images after resizing, normalization, randomization and augmentation.

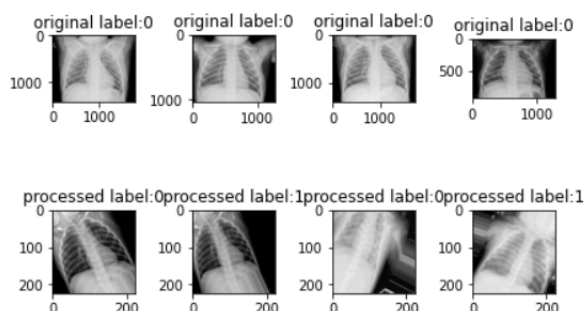


Figure 1. Examples of the data.

2.2 Proposed Model

The model discussed in this paper was CNN, which is now a representative of neural network expert in analyzing images. Except dense layer, it has unique architectures including convolutional layer, pooling layer and flatten layer to deal with grid-like-structure data. First, the convolutional layer uses a small matrix called filter to scan and move around the original picture by a certain stride. Then it does matrix multiplication and records the sum of the dot product between the filter and the portion of the image that the filter is hovering. If the sum is large, it means that the filter is similar to the image. By this, the layer can extract the feature of the image. Second, pooling layer can further reduce the size of the data and do denoising. Like convolutional layer, a kernel scans the image. The maximum value or the average value in the area covered by the kernel is saved. Third, flatten layer converts the image matrices into vectors and sends them to dense layer. Finally, dense layer does classification according to the features it receives.

To find out how depth of a model would affect the whole performance, models with similar architecture but different depths were trained and evaluated. In this work, one convolutional layer plus one max pooling layer plus one batch normalization layer was defined as one small architecture. The whole structure of the models was like: several small architectures piled up, one flattens layer, one dropout layer (the rate was 0.1), one dense layer with 128 output units, one relu activation function layer, one dropout layer (the rate was 0.2), one dense layer with 1 output unit and finally sigmoid activation function layer. The first model had two small architectures piled up. In models after that, each time added two more small architectures. That is, these models had 2, 4, 6... small architectures. The channel number in each convolutional layer increased in proportion, e.g., 32, 64, 128, 256... In all convolutional layers, filter size was 3×3, stride was 1, Same Padding was used, and activation function was relu. In all max pooling layer, kernel size was 2×2, stride was 2 and Same Padding was used.

2.3 Implementation Details

The implementation was based on Keras API using Tensorflow2 as its backend. The models were trained with GPU (Tesla T4) on Google Colab to accelerate the training process. Batch size was 8 and epoch number was 10. Since this is a binary classification, binary crossentropy was used as the loss function (Equation 1). Adaptive Moment Estimation (Adam) was chosen as the optimizer, which could adjust the learning rate for the parameters to rapidly minimize the loss function. In Adam, the initial learning rate was 0.001, β_1 was 0.9, β_2 was 0.999 and epsilon was 1e-7. Although Adam itself is adaptive to learning rate, applying an extra learning rate decay schedule can further improve the performance [10]. The learning rate was set to reduce by 20% if there was no improvement in validation set accuracy in 2 epochs. The minimum of learning rate was 1e-6. Metrics were accuracy (Equation 2) and F1 score (Equation 3,4,5), which are both common statistical measures in evaluation. Accuracy was computed each epoch in the training process to show how well the model was learning. F1 score finally reflected the performance of the model on test set, which ranges from 0 to 1. If F1 score is close to 1, the model performs well.

$$Loss = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log y_i + (1 - y_i) \cdot \log(1 - y_i) \quad (1)$$

$$Accuracy = \frac{\text{number of correct classification}}{N} \quad (2)$$

$$precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3)$$

$$recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (4)$$

$$F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

3. Results and Discussion

In this work, pneumonia was predicted using four CNN models of varying depths that were trained on the same chest X-ray dataset. They consisted of 2, 4, 6, 8 convolution layers correspondingly. Their accuracies on both training set and testing set were shown in Table 1. According to Table 1, all models had a high accuracy on training set, while the accuracy on testing set rose at first and then dropped. Model 2 and 3 both had a relatively high accuracy on training set and testing set.

Table 1. Accuracies of models with different depths

Model	Conv Layer Number	Conv Channels	Accuracy on training set	Accuracy on testing set
1	2	32, 64	0.8877	0.6715
2	4	32, 64, 128, 256	0.9237	0.8606
3	6	32, 64, 128, 256, 512, 1024	0.9423	0.8750
4	8	32, 64, 128, 256, 512, 1024, 2048, 4096	0.9482	0.8045

Also, the bar chart in Figure 2 depicted F1 scores for the two categories of the four models. In each model, F1 scores for the normal images and pneumonia images were similar to each other. Both model 2 and 3 had F1 scores which were close to 1 compared to model 1 and 4.

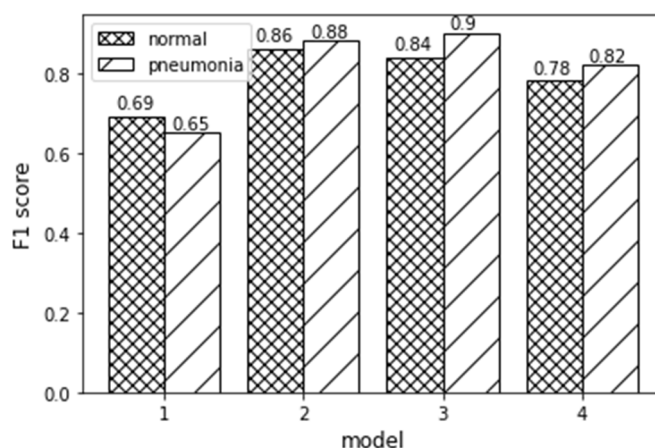


Figure 2. F1 scores of models with different depths

With the increase of the depth, a model would have more parameters. High accuracy on training set indicated that all four models had enough parameters to extract and save the features and information conveyed by the images. They all matched the training set well. Among these models, model 4 performed best.

When doing prediction, model 1 had a low accuracy. That is possibly because model 1 had the least scale in all four models. It could just fit the training set but not yet ready to deal with new images independently. Model 4 also had an obvious low accuracy because the weights in this model stucked to the training set too much that it was not generalized and was suffering from overfitting. Both model 2 and model 3 performed well which reflected that they just had the right number of parameters. This enabled them to learn the overall features of pneumonia and normal images and could further guess the categories of the images they have never encountered.

F1 score reflects the capability of a model to predict a certain category. Observing the statistical graph of F1 scores, each model had similar abilities on predicting the two categories. All of them did not come to the issue of unbalancing. This also proved that the class weight between the two categories set before training was reasonable given that the sample number of pneumonia was about three times larger than the normal ones. Depth of a model would not affect the balance and deteriorate

the performance. Model 2 and model 3 again had a nearly equivalent performance. Their F1 scores were all high. However, the gap between the F1 scores for pneumonia class and for normal class in model 2 was smaller than it was more stable and not biased. What's more, to achieve a similar result, model 3 had two more convolutional layers which led to higher computational cost. In this regard, model 2 had the best performance in this paper.

4. Conclusion

In the paper, CNNs with various configurations image classification were proposed to diagnose pneumonia based on chest X-rays. The CNN model was accepted to distinguish the images and help get a reasonable prediction. To obtain a model which could predict the disease most accurately, models with similar structure but varied depths were defined. Four experiments were conducted to discover the correlation between the depth and the performance of a model. Results indicated that the model with four convolutional layers, which was at a moderate depth, returned the best outcome. In the future, more parameters such as epoch and batch size can be adjusted to improve the model.

References

- [1] Pneumonia World health organization 2021 <https://www.who.int/news-room/fact-sheets/detail/pneumonia>.
- [2] Nirmala K et al. 2022 Investigations of CNN for Medical Image Analysis for Illness Prediction Computational Intelligence and Neuroscience.
- [3] Qiu Y et al. 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training Biomedical Signal Processing and Control 72 103323.
- [4] Chauhan N K and Singh K 2018 International Conference on Computing, Power and Communication Technologies (GUCON) A Review on Conventional Machine Learning vs Deep Learning p 347-352.
- [5] Jakhar K and Hooda N 2018 4th International Conference on Computing Communication and Automation (ICCCA) Big Data Deep Learning Framework using Keras: A Case Study of Pneumonia Prediction p 1-5.
- [6] Singh A Shalini S and Garg R 2021 11th International Conference on Cloud Computing Data Science & Engineering (Confluence) Classification of Pediatric Pneumonia Prediction Approaches p 709-712.
- [7] Wan S Hsu C Y Li J and Zhao M 2020 International Conference on Intelligent Computing, Automation and Systems (ICICAS) Depth-Wise Convolution with Attention Neural Network (DWA) for Pneumonia Detection p 136-140.
- [8] Gabruseva T Poplavskiy D and Kalinin A 2020 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops p 350-351.
- [9] Chest X-Ray Images (Pneumonia) Paul Mooney Mar. 2018 <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>.
- [10] Loshchilov I and Hutter F 2017 Decoupled weight decay regularization arXiv preprint arXiv:1711.05101.