I-Project: Face Mask Detection

YuJia Cao

June 27, 2021

Abstract

Due to the impact of the COVID-19, the need of image recognition for the mask are growing rapidly. Undoubtedly, using machines to check the wearing of mask can greatly save the labor cost. Through the experiment, it had been found out that having 4 layers is an rational choice. After comparing different existing models and their performance on this specific task, the neural network ”DenseNet” was discovered to has the best accuracy and lowest loss. After that, the research specified on to the ”DenseNet” network, by changing its learning rate and activation function, the model has successfully reach the accuracy of 99.5%, performing the image classification work just as intended.

1 Introduction

Under the situation of the COVID-19, people need to wear masks everywhere they go. Checking whether people had properly wear the mask or not is an important work to be down to ensure health security. To release the pressure of hiring men to do this simple checking work, picture recognition technique can be used to detect the mask on faces.

In the work, five different neural network models were used to achieve face mask detection, and compared based on their performance. The first part of the work is to found out the more suitable number of layers for this recognition work. The second part is using five existing models and the trained parameters and comparing the most form of neural network. The third part is specify on the most suitable model and find the proper parameter value for this face-mask detection task.

Figure 1: People wearing masks whenever they go.

2 Research Setups

2.1 Testing environment

Hardware:

- MacBook Pro (16-inch, 2019)
- CPU: 2.6 GHz Intel Core i7
- Storage: 16 GB 2667 MHz DDR4
- GPU:AMD Radeon Pro 5300M 4 GB; Intel UHD Graphics 630 1536 MB

Software:

- Language: Python 3.9
- Shell: Pycharm CE
- Requirements: Keras, Tensorflow, Matplotlib, Pandas, Numpy

2.2 Training & Testing Dataset

The dataset consisting 10003 images for training, 803 images for validation, and 995 images for testing. Numbers for images with mask and without mask are approximately equal.

Figure 2 and 3 are sample images from the test and training dataset, showing that it both consist of image of face, not including other information.

2.3 Image Augmentation

To better train the model made it better adapted to nonstandard images, images have been augment using various techniques. By using image data generator, with techniques such as re-scale, rotation, shifts, zoom, flip etc. (Figure 3)

2.4 Variables and Parameters

Model variables

- Kernel Size: An integer or tuple/list of 2 integers, specifying the height and width of the 2D convolution window.
- Activation function: A node defines the output of that node given an input or set of inputs.
- Optimizer: An automatic optimizing algorithm which will adjusting the weights in the neural network model.

Model evaluation dimensions

- Loss function: A function to map values of variables to a simple real number to represent the penalty for failing to achieve a desired value.
- Accuracy: The percent of images being correctly classified.
- Validation loss: The loss of the classification process when dealing with testing images.
- Validation accuracy: The accuracy of the classification when dealing with testing images.
- Time per step: The time used in calculating process per step in an epoch.

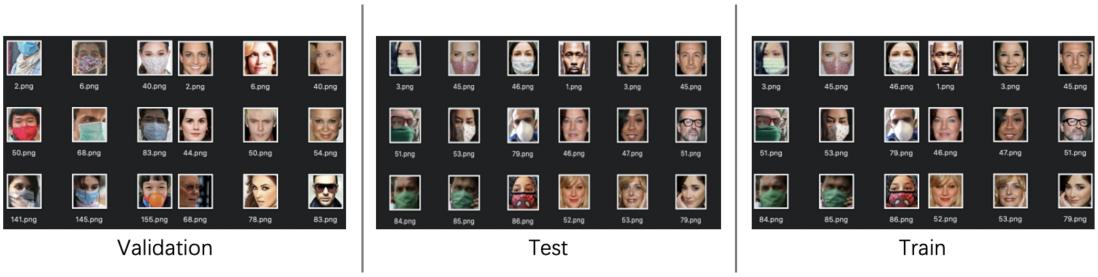

Figure 2: Sample images in the dataset.

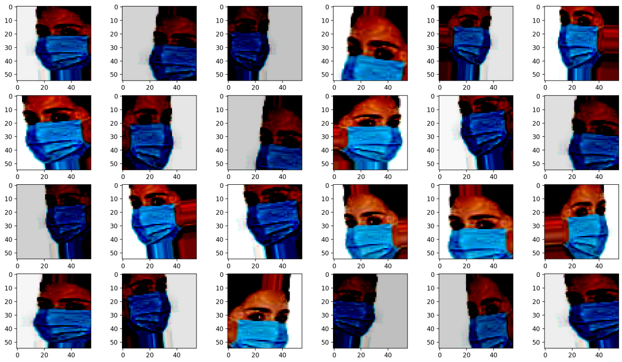

Figure 3: A sample of image augmentation.

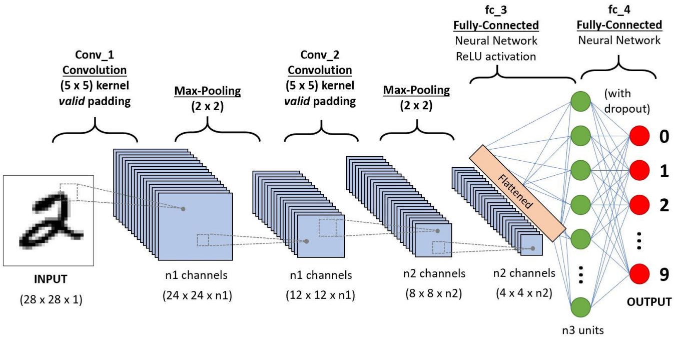

Figure 4: A Sample convolutional neural network.

2.5 Technological Terms

- CNN: Convolutional neural network.
- 2D Convolution Layer: A layer which creates a convolution kernel that is wind with layers input which helps produce a tensor of outputs.
- Batch normalisation: A mechanism that is used to improve efficiency of neural networks, stabilising the distributions of hidden layer inputs and thus improving the training speed.
- Max pooling: Down-samples the input along its spatial dimensions (height and width) by taking the maximum value over an input window for each channel of the input.
- Dropout: A Simple Way to Prevent Neural Networks from over-fitting, ignoring units during the training phase of certain set of neurons which is chosen at random.
- Flatten layer: A layer which collapses the spatial dimensions of the input into the channel dimension.
- Dense layer: A fully connected layer, meaning all the neurons in a layer are connected to those in the next layer.
- Epoch: Indicates the number of passes of the entire training dataset the machine learning algorithm has completed.
- Step: The number of batch iterations before a training epoch is considered finished.

Layer(type)	Output Shape	Numbers of Parameters
conv2d(Conv2D)	(None, 126, 126, 32)	896
batch_normalization(BatchNo)	(None, 126, 126, 32)	128
max_pooling2d(MaxPooling2D)	(None, 63, 63, 32)	0
dropout(Dropout)	(None, 63, 63, 32)	0
flatten(Flatten)	(None, 127008)	0
dense(Dense)	(None, 2)	254018

Table 1: The structure of a single layer neural network

Types of model	Accuracy	Validation Accuracy
1-layer	0.9825	0.9875
2-layers	0.9840	0.9844
3-layers	0.9866	0.9871
4-layers	0.9868	0.9893
5-layers	0.9895	0.9884

Table 2: The structure of a single layer neural network

3 Model

3.1 Impact of Layers

As shown in the Table 1,[KS15] the 1-layer neural network consist of different layers, which functioned together and do the image recognition work. The activation function in the conv2d layer is ”relu”, and the activation function in the dense layer is ”sigmoid”.

[DV17]

In the experiment, I construct five different neural network model from the one consist of 1 layer to 5 layers. All models go through 10 epochs and each with 208 steps.

The result shows that all five basic model reach the accuracy of 98%.(Table 2) With the increase in the number of layers, the time spend in each step had increase, that effect the increase in total time for training too. (Figure 5) The model performance on both loss and accuracy had improve along with the increase in the number of layers presented. However, when the layer number reach 4, the improvements become insignificant. (Figure 6)

3.2 Comparison between existing models

Similar to the previous steps, I used the trained model and detected its performance on the face-mask detection task.

Using a trained model can reduced the time used in training process, and they have steady performance on all sorts of classification tasks.

Optimizer used in all networks are ”adam” and the loss are measured using the standard of ”binary_crossentropy”. All models go through 10 epoch and each with 208 steps.

As shown in the Figure 7, DenseNet has the best performance compares with all other neural networks. It reached the validation accuracy of 0.9923.

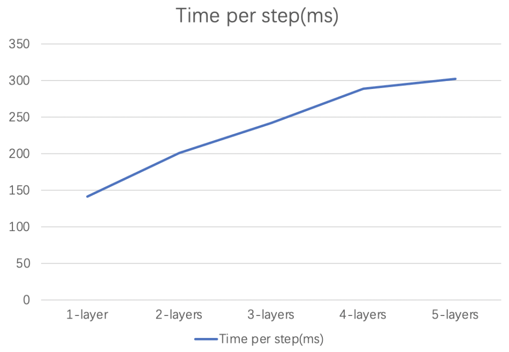

Figure 5: Average time used per step in each model.

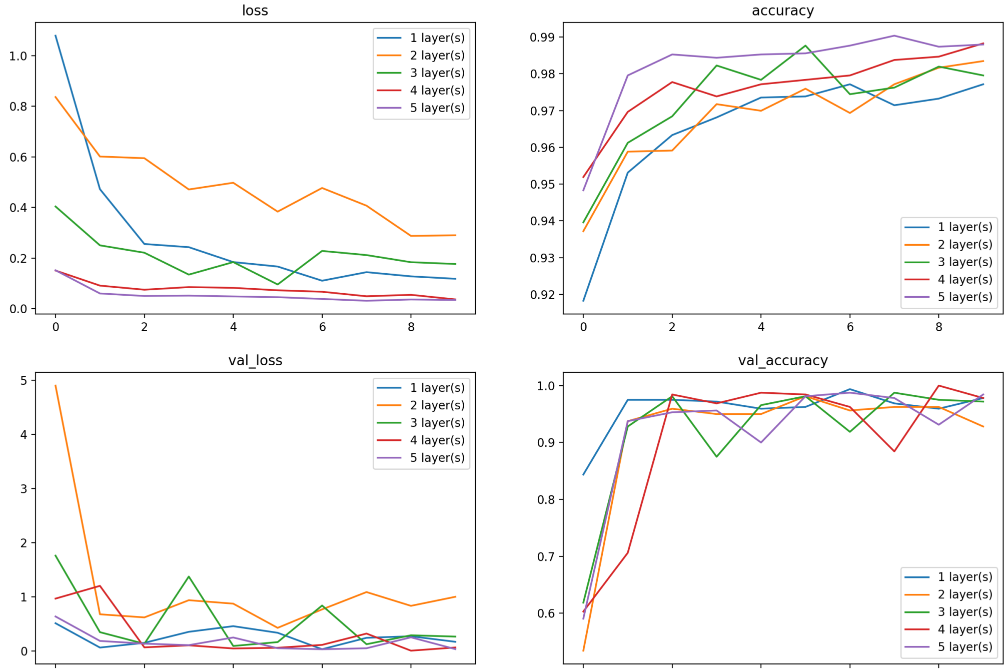

Figure 6: Performance of different models

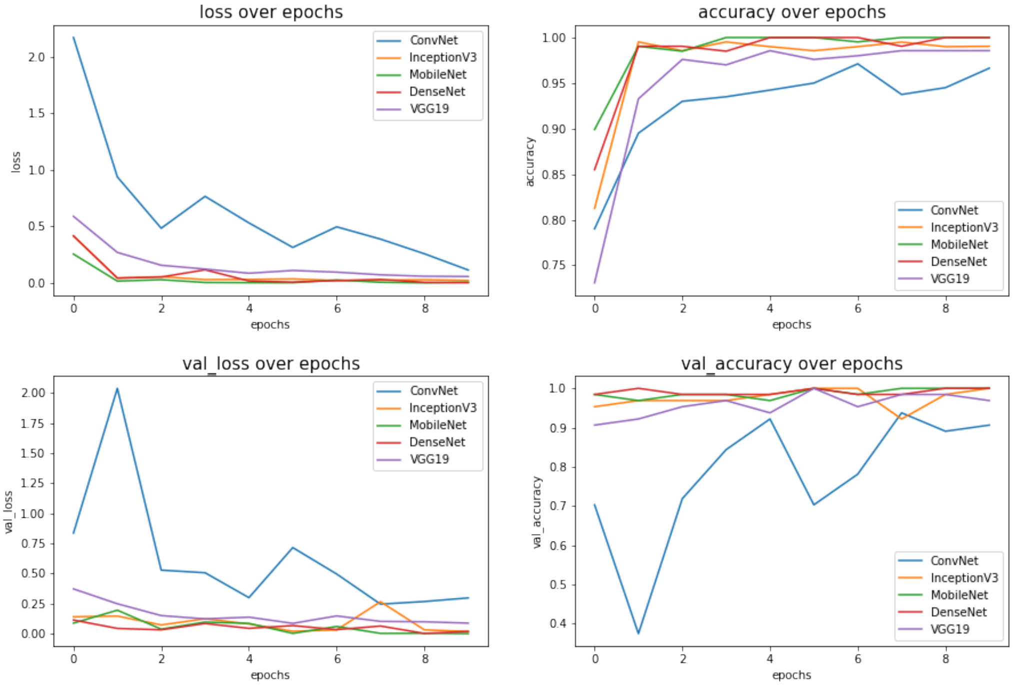

Figure 7: Performance of different neural networks

Layer(type)	Output Shape
Convolution	(126, 126)
Pooling	(56, 56)
Dense Block	(56, 56)
Dense Block	(28, 28)
Dense Block	(14,14)
Dense Block	(7, 7)
Classification Layer	(1,1)

Table 3: The structure of a single layer neural network

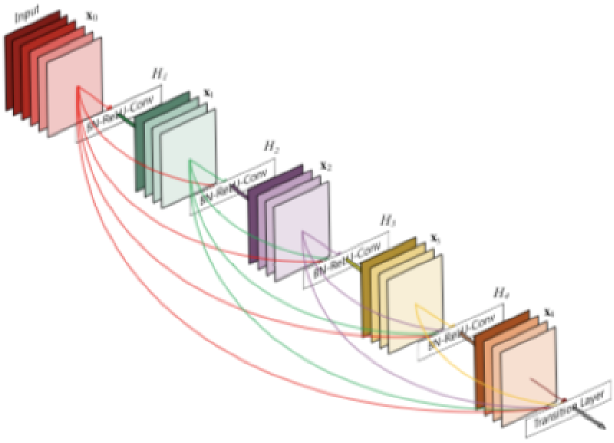

Figure 8: Structure of DenseNet.

The DenseNet's structure is in Table 3, which shows DenseNet is basically constructed with difference dense layer which gradually smaller the size of the input and thus extract the characteristics of the image and do the classification work.

3.3 Application of DenseNet

DenseNet has an uncommon feature comparing to the traditional CNNs. DenseNets do not sum the output feature maps of the layer with the incoming feature maps but concatenate them. For traditional CNNs, if existing L layers, there exist L connection. But for DenseNet, it has $\frac{L(L+1)}{2}$ connections in the model. (Figure 8)

Figure 9 shows the structure of DenseNet. This special kind of structure lowers the size of output and made the optimization process easier. Each layer has direct access to the gradients from the loss function and the original input signal, leading to an implicit deep supervision.[Rui18]

The equation of the structure can be summarized as this:

$$x_l = H_l([x_1, x_2, ..., x_{l-1}])$$

[GvdM18] Noted that the $[x_1, x_2, ..., x_{l-1}]$ in the formula represents the concentration of the out put feature maps from the first layer to the $l-1$layer.

Its growing rate can be represent by:

$$k_l = k_0 + k \cdot (l-1)$$

[GvdM18] As we see in the figure 9, every layer has access to its preceding feature maps, and therefore, to the collective knowledge. Each layer is then adding a new information to this collective knowledge, in concrete k feature maps of information.

By using the learning rate that decay with the help of a polynomial function of a root (power of 0.4).The learning rate range will go through 10000 steps and the value will be between 0.01 and 0.00001.

The model goes through 5 epochs, each with 157 steps. The results are shown in the Figure 9.

The confusion matrix also visualized the actual performance of this model.[Kar21] (Figure 10)

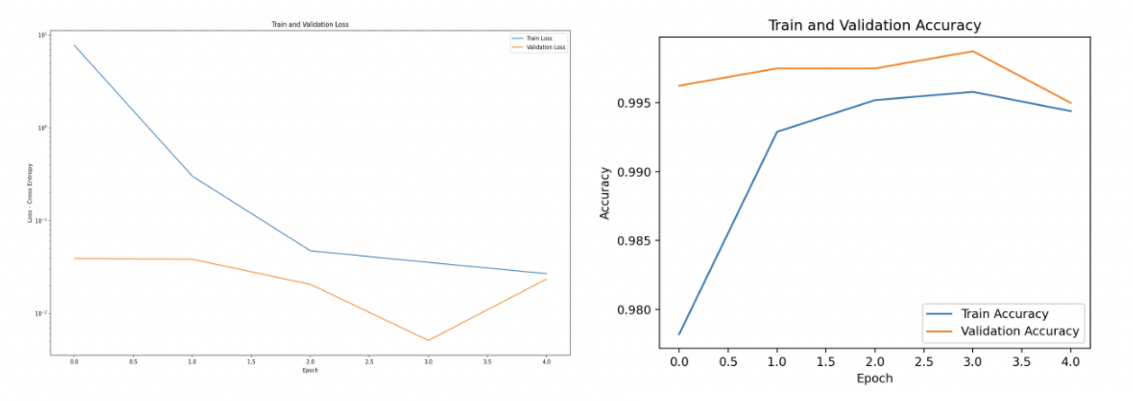

Figure 9: The performance of DenseNet.

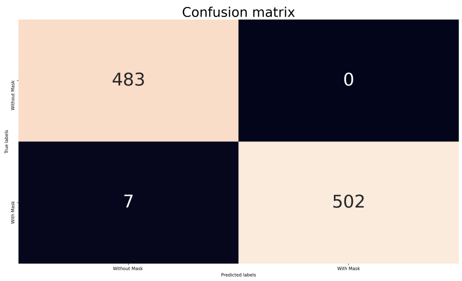

Figure 10: Confusion matrix for the results of DenseNet.

4 Summary and Future works

In the paper the impact of having different numbers of layers and different kinds of CNNs were being discussed. After that, a close look to DenseNet had been used and reached an accuracy of . Throughout the research project, various model evaluation dimensions had been used.

For future works, since it already fulfilled the task of classify face image with or without mask on it. Combining with the face detection technique (Opencv) , the face mask detection can be used in larger range of pictures, not only the one with face in it, but also picture of a crowd.

Also, further exploration on the raising of accuracy can be carry on. Finding a better model for greater accuracy, lower loss, and a simpler structure is what everyone is working at. Even if the facial recognition had been used widely in the society, further improvement can be done to improve performance to human's level.

References

[DV17] Dumoulin and Visin. Convolution and Pooling Arithmetic for Deep Learning. 2017.

[GvdM18] Z.Liu G.Huang and L. van der Maaten. Densely Connected Convolutional Network. 2018.

[Kar21] Shahaf Karp. Face Mask Detection with DenseNet 201. 2021.

[KS15] S.Ren K.He, X.Zhang and Jian Sun. Deep Residual Learning for Image Recognition. 2015.

[Rui18] Pablo Ruiz. Densenet for Image Net. 2018.