Face Mask Detection Using VGG16 Model Architecture

Yash Soni - 001145775 - ys6200w

Abstract

Strict policies have been formed in various countries that makes wearing face mask mandatory in public places to tackle COVID-19. The rising importance of wearing a face mask has urged governing bodies to go 'an extra mile' to enforce the rules. This paper proposes an approach that uses facial recognition system to determine if an individual is wearing a face mask covering. The system uses network of VGG16 model architecture to create a machine learning model that predicts if an individual in image is wearing a face mask. Facial recognition system has diverse real-world applications such as, security and is capable to solve complex problems. The proposed model provided accuracy of 98% on validation data. Integrating the system with live street cameras can help authorities monitor the enforcement of rules.

1. Introduction

This paper reports the development of a machine learning model that uses image recognition system to predict if an individual in the input face image is wearing a face mask or not. The further sections include Method that describes architecture of the model used, Experiment section describes the experiment settings, evaluation, and discusses results. Finally, Conclusion section describes about effectiveness of the model in real-world applications.

1.1. Background

Pioneer institution World Health Organization (WHO) and a range of studies related to infectious disease such as COVID-19 have established the fact that wearing a face mask can help reducing the spread of infection. (Cheng et al. 2020; Howard 2020; WHO 2021) According to a study on understanding measures to fight COVID-19 published by researchers at the University of Edinburgh, a face covering over nose and mouth can help cutting the risk of coronavirus by limiting the distance travelled by exhaled breath by 90%. (Godoy et al. 2020)

Prior to pandemic, very few people preferred to wear mask in public but due to above mentioned fact, the importance and demand of face mask supplies have increased from the beginning of the year 2020 when COVID-19 pandemic was at its initial stage. Federal and local governing authorities have formed SOPs (standard operating procedures) in major countries and regions that makes wearing face mask mandatory in public places like shopping malls, theatres, schools and colleges, etc.

Concerns have been raised regarding the enforcement of COVID-19 rules and regulations. Recently, a news article was published in which, concerns were raised for enforcing mask requirements in public transport in England. (Guardian 2021) UK supermarkets are calling for police to enforce the same. (WEF 2021) However, it seems impractical to monitor the enforcement of mandatory face mask rules using traditional means.

1.2. Proposed Solution

Artificial Intelligence techniques could be used to enhance the enforcement of COVID-19 rules. They are capable to solve very complex problems. These are being applied in a wide range of fields. The most popular area of artificial intelligence is machine learning.

With that been said, in this paper, I propose using a machine learning model that uses facial recognition system to detect if an individual in the image is wearing a face mask or not. Facial recognition system is a modern computer technology that recognizes a face by mapping its features. It is capable to solve complex problems and has various applications like bio-metric identification, airport security, and many more. By integrating the model with live street cameras, it could be possible to enhance the monitoring of COVID-19 rules enforcement.

1.3. Literature Review

A remarkable progress has been made in the area of image recognition system due to largely available image datasets.(Shin et al. 2016) The most widely used image recognition algorithm is Convolutional Neural Network (CNN). This deep learning algorithm can classify input images by assigning weights to numerous aspect in the image. Compared to any other classifier algorithm, the pre-processing of data required is very low. (Saha 2021)

Moreover, pre-trained CNN network models are available to solve a particular problem using deep learning. Among them is the VGG16 network model that consists of 16 layers and 138 million parameters that are trained using ImageNet dataset of 14 million images of 1000 classes with accuracy of 92.7% (Neurohive 2018) This network was mainly design to classify images. It is popular due to its simplicity and easy implementations. In this paper, I have used VGG16 network to train the machine learning model mainly due to the fact that even with limited training data, the model can learn efficiently from layers and weights in VGG16.

1.4. Dataset

For this paper, I have used a publicly available image dataset from Kaggle that consists of total 11,792 images. The data is distributed into three categories: train, test, and validation. Each categories consists of two types of images that are evenly distributed: Images of face with a mask and Images of face without a mask. Training data contains 10000 images, Testing data contains 992 images, and Validation data contains 800 images. It could be accessed from: https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset

1.5. Summary

The trained model provided 98.75% accuracy against validation data while providing 98.89% on test data. The loss function used during the training is categorical cross-entropy. In order to diversify the training data, data augmentation techniques were also applied.

2. Method

The problem identified in the Introduction section falls under the category of multi-class classification problems as the task is to classify input images into two classes: with mask and without mask. To solve classification problems, diverse supervised and unsupervised classifier algorithms could be identified. (Sekeroglu, Hasan, and Abdullah 2019)

2.1. Convolutional Neural Network

For image classification, the most widely used method is Convolutional Neural Network (CNN), also known as ConvNet. (Saha 2021) It mainly includes three type of layers: Convolutional, Pooling, and Fully-Connected. The Convolutional layer is the first layer of a CNN network which could be followed by another Convolutional layer or Pooling layer. The Pooling layer reduces the number of parameters while a Fully-Connected layer is finally a output layer. (IBM 2021) There are various CNN architectures developed that makes implementing CNN simple and easy for example, ResNet, VGGNet, etc. I have used VGG16 in this paper which is briefly explained in the next section.

2.2. Model Architecture

As mentioned in previous sections, I have implemented VGG16 architecture model to solve the image classification problem. The model contains a total of 21 layers but only 16 layers contain learnable parameters. There are 13 convolutional layers, 5 pooling layers, and 3 fully-connected layers. The architecture of VGG16 model could be more clearly understood using following figure.

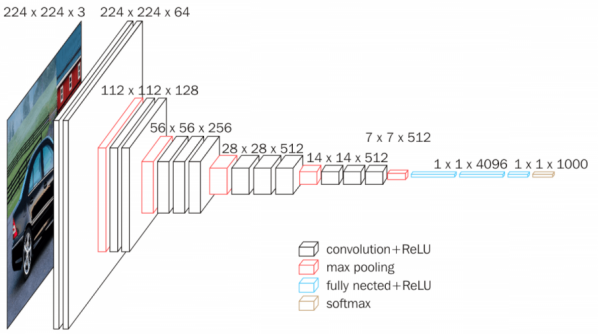

Figure 1. VGG16 Architecture (Neurohive 2018) depicting various components like convolutional layers, pooling layers, fully-connected layer, and the sizes of filters.

As could be seen in Figure 1, the model takes an input image of size 224x224 with 3 channels. The first two convolutional layers has 64 filters which is followed by a pooling layer. The next two convolutional layers has 128 filters before a pooling layer. Likewise, it goes on for two iterations with two sets of three convolutional layers with 256 and 512 filters respectively. Finally, there are three fully connected dense layers which are mainly responsible for classification task. The first two connected layers have 4096 channels each while the third layer provides the output. As there are two possible outcome of our model, probability of with mask and without mask, 2 channels are defined in the third output layer.

2.3. Activation functions

An activation function defines how the weighted sum of an input is transformed into an output. It plays major role in training the model. The rectified linear activation function, also known as ReLU is used in intermediate layers of the model. It is known to enhance the performance of a neural network. If the input is positive, it will pass the same as output. But if the input is negative, it will pass as zero. (Brownlee 2019) The equation of ReLU function is provided below.

$$ReLU = max(0, x) \tag{1}$$

x = Input

As the problem to be solved is multi-class classification, Softmax activation function is used in the output layer. It provides probability from each node in the output layer where the sum of these values will be equal to 1.0. The equation of Softmax function is provided below.

$$softmax(z_i) = \frac{exp(z_i)}{\sum_j exp(z_j)} \tag{2}$$

In the above equation, z is the value of neuron from output layer while exponential is a non-linear function.

2.4. Loss Function

The loss functions are responsible to minimize the error. It evaluates how well the algorithm models the dataset. (Algorithmia 2018) For my multi-class classification model, I have used Categorical Cross Entropy loss function which calculates cross entropy loss between labels and predictions. For better modelling, the value of loss function should be near to zero. The equation of categorical cross entropy funtion is provided below.

$$Loss = -\sum_{i}^{n} y_i \cdot log(\hat{y_i}) \tag{3}$$

In the above equation, n represents the size of output, y_i represents the true value, and $\hat{y_i}$ represents the predicted value.

2.5. Training Environment

The model was trained in Google Colab notebook environment with GPU enabled processor. The python packages that were used includes numpy, pandas, tensorflow, keras, sklearn, and matplotlib.

3. Experiments

As mentioned previously, I have used a publicly available image dataset from Kaggle that contains 10,020 face images for training, 800 face images for validation and 992 face images for testing. Images are of 2 classes: with mask and without mask. The available data was already clean so no data-cleaning techniques were required. However, data-augmentation techniques were applied to diversify the dataset, like vertical flip, horizontal flip, zoom, shuffle and re scale.

3.1. Experimental settings

The model consists a total of 23 layers, out of which 21 layers are of VGG16 model and two dense layers were defined for output. The number of parameters involved in the layers could be known from the table below.

Table 1. Sequence of layers in the proposed model along with corresponding number of parameters

Layer	Parameters
VGG16 (21 Layers)	14714688
Dense Layer 1	1605696
Dense Layer 2	130

From the Table 1, it could be understood that there are total 16,320,514 out of which, only 1,605,826 are trainable as parameters in the VGG16 models were made non-trainable as they are already trained to classify the images by learning features.

The task of layers in the neural network model is to classify the inputs into two categories by learning features from the same. In the process, to help convolutional layers detect spatial features, kernel plays an important role. The kernels performs a dot product on vectors of image pixel values. The size of a kernel in model is 3x3. The optimizer used in the VGG16 model is Adam.

3.2. Evaluation criteria

Accuracy generally describes how the model performs on all the classes. To drive the training of model, I chose 'accuracy' as metrics as the aim of the model is to provide accurate predictions on all the classes. If we intuit confusion matrix, the equation of accuracy can described as follows.

$$accuracy = \frac{T_{pos} + T_{neg}}{T_{pos} + T_{neg} + F_{pos} + F_{neg}} \tag{4}$$

In the above equation, T_{pos} = True Positive, T_{neg} = True Negative, F_{pos} = False Positive, and F_{neg} = False Negative.

Moreover, various evaluation metrics are also applied to evaluate the trained model such as ROC-AUC Score, Precision, Recall, and F1 Score. The results are discussed in the next section.

3.3. Results

In this subsection, I am going to describe various results of evaluation metrics. The accuracy is measured while training the model with training and validation data. The highest validation accuracy obtained while training is 99% with loss of 3.77%. When evaluated the trained model on test data,

the data which it has never seen, it provided accuracy of 98.79% with loss of 2.61%. The following figure represents history of the trained model.

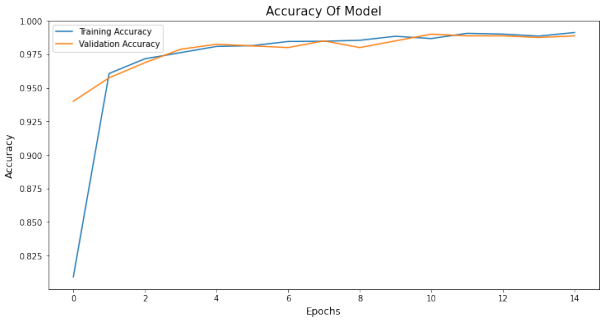

Figure 2. Plot of accuracy during training of model. There are total 15 epochs and the accuracy ranges from 0.8 to 1.0. Plot contains training accuracy and validation accuracy.

In Figure 2, x-axis represents the number of epochs while y-axis represents the obtained accuracy. The blue line represents accuracy on training data while orange line represents accuracy on validation data. The results of other evaluation metrics that were performed using trained data could be referred from following table.

Table 2. Results of various evaluation metrics on trained model, performed with testing data

Test	Score
ROC-AUC	0.98
Precision	0.99
Recall	0.98
F1	0.98

The interpretation of test results shown in Table 2 are discussed in further section. Confusion matrix was also used to evaluate model on test data. Out of 992 predictions, it provided 478 True Positives, 503 True Negatives, 5 False Positives, and 6 False Negatives.

3.4. Discussion

Due to a large series of trained layers, the VGG16 network made it easy for the model to learn features from the input images. The test results show that the model has performed quite well in classifying the face images with mask and without mask. Moreover, data augmentation techniques made the dataset more diverse. However, it could be argued that due to model trained on already cleaned dataset, it performed efficiently on various evaluation tests. Also, we cannot ignore 138 million trained parameters of VGG16 in contributing to the performance.

4. Conclusion

COVID-19 is still a major medical threat throughout the globe. With new variants being discovered in a period of time, it is highly recommended by the medical experts to wear a face mask. When it comes to enforcement of rules related to it, I believe machine learning can help solve this problem. The neural network model developed using VGG16 proved to be efficient by obtaining 98.79% accuracy in classifying image with face mask and without face mask. By embedding the model with street cameras, it is possible to monitor the enforcement of mandatory face mask rule. The future work will include an additional class where it will identify if the mask worn covers nose and mouth as well. At last, the report concludes that the model is effective in solving the problem.

References

Algorithmia (2018). *Introduction to Loss Functions*. URL: https://algorithmia.com/blog/introduction-to-loss-functions.

Brownlee, Jason (2019). *A Gentle Introduction to the Rectified Linear Unit (ReLU)*. URL: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/.

Cheng, Vincent Chi-Chung et al. (2020). "The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2". In: *Journal of Infection* 81.1, pp. 107–114.

Godoy, Laura R Garcia et al. (2020). "Facial protection for healthcare workers during pandemics: a scoping review". In: *BMJ global health* 5.5, e002553.

Guardian (2021). *Concerns over masks enforcement on public transport in England*. URL: https://www.theguardian.com/politics/2021/nov/29/concerns-over-masks-enforcement-on-public-transport-in-england-mandatory-face-coverings-police.

Howard, Matt C (2020). "Understanding face mask use to prevent coronavirus and other illnesses: Development of a multidimensional face mask perceptions scale". In: *British journal of health psychology* 25.4, pp. 912–924.

IBM (2021). *What are Convolutional Neural Networks? — IBM*. URL: https://www.ibm.com/topics/convolutional-neural-networks.

Neurohive (2018). *VGG16 - Convolutional Network for Classification and Detection*. URL: https://neurohive.io/en/popular-networks/vgg16/.

Saha, Sumit (2021). *A Comprehensive Guide to Convolutional Neural Networks—the ELI5 way*. URL: https://towardsdatascience.com/a-comprehen

sive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

Sekeroglu, Boran, Shakar Sherwan Hasan, and Saman Mirza Abdullah (2019). "Comparison of machine learning algorithms for classification problems". In: *Science and Information Conference*. Springer, pp. 491–499.

Shin, Hoo-Chang et al. (2016). "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning". In: *IEEE Transactions on Medical Imaging* 35.5, pp. 1285–1298. DOI: 10.1109/TMI.2016.2528162.

WEF (2021). *COVID-19: UK supermarkets call for police to enforce mask wearing*. URL: https://www.weforum.org/agenda/2021/01/covid-19-spread-supermarkets-minister-vaccine-masks-safety/.

WHO (2021). *Coronavirus disease (COVID-19): Masks*. URL: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-masks.