International Journal for Modern Trends in Science and Technology, 6(12): 141-146, 2020

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST061227
Available online at: http://www.ijmtst.com/vol6issue12.html

Video Analytics Solution for COVID-19

Prerna Gupta

Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India

To Cite this Article
Prerna Gupta, Dr. Bhoomi Gupta and Vandana Choudhary, "Video Analytics Solution for COVID-19", *International Journal for Modern Trends in Science and Technology*, 6(12): 141-146, 2020.

Article Info
Received on 08-November-2020, Revised on 25-November-2020, Accepted on 30-November-2020, Published on 03-December-2020.

ABSTRACT

Amid the global crisis of the Corona virus pandemic, new demands have emerged in the market which uses Video Analytics for finding solutions to halt the transmission of the Virus. The COVID - 19 pandemic is devastating mankind irrespective of caste, creed, gender, and religion. Until a vaccine is discovered, we should do our bit to constrain the expanse of the corona virus. Using a face mask can undoubtedly help in managing the spread of the virus. The face mask detector, a video analytic solution uses MobileNetV2 model, deep learning techniques to successfully test whether a person is wearing a face mask or not. The face mask identifier is least complex in structure and gives quick results and hence can be used in CCTV footage to detect whether a person is wearing a mask perfectly so that he does not pose any danger to others. Mass screening is possible with video analytics and hence can be used in crowded places like Airports, Hospitals Entrance Exam Centers, Schools and Colleges.

KEYWORDS: *deep learning, facemask, convolutional neural network, mobilenetv2, opencv*

I. Introduction

With the occurrence of severe acute respiratory syndrome coronavirus (SARS-CoV-2), the virus that started from Wuhan, China from December 2019, has been spread throughout the world. The transmission of the disease is air-borne which affects the people who breathe or come in contact with the infected droplets that remain infectious when suspended in air over a long period of time [1]. Due to airborne transmission, the use of face masks has become ubiquitous for limiting the spread of the virus.

The World Health Organization (WHO) reports suggest that the two main routes of transmission of the COVID-19 virus are respiratory droplets and physical contact or through the air after the aerosolization process[2]. Wearing a medical mask is one of the prevention measures that can limit the spread of certain respiratory viral diseases, including COVID-19. While a mask use has been made a priority, there is insufficient information in the public domain about the role masks play and the ability of the mask to protect the wearer from infectious particles [3].Video Analytics solutions ensure adherence to safety norms laid down by the WHO to tackle the spread of the COVID-19.

The use of masks in society is still a challenge. There are several strategies in which people using face mask in certain place and case[4]. Several reasons are the people's health condition, misinformation and misinterpretation, politics, beliefs, mental health conditions and herd immunity. Another challenge is the limited authorities' personnel which resulted the monitoring of masks usage becomes less and less effective. To overcome such problem of ineffective monitoring, this paper proposes a method to use video analytics to detect face mask through image that can be produced by cameras or image files. To detect, we use classification method called as MobileNetV2, in deciding whether a face image wears a mask or not.

Video analytics makes surveillance systems more efficient, overcome the challenge of limited authorities' personnel for monitoring, reduces the workload on security and management staff, and helps capture the full value of security video by making the camera system more intelligent in its work. Video analytics through CCTV cameras enhances visibility and brings in transparency for organizations to act faster. These systems can be deployed in a very short time and support integration to almost all systems enabled by the web.
Some of the video analytics solutions for COVID-19 are:
¬ Crowd Density Detection
¬ Social Distance Monitoring
¬ IR(Thermal Sensors) cameras for temperature detection
¬ Personal protective equipment (PPE) detection and alerts
¬ Facial recognition with masks
¬ Face Mask Detection

The model is built for detection of face masks in the following stages:
1. Training custom face mask detector.
2. Implementation to:
    - Detect COVID-19 face masks in images
    - Detect face masks in real-time video streams

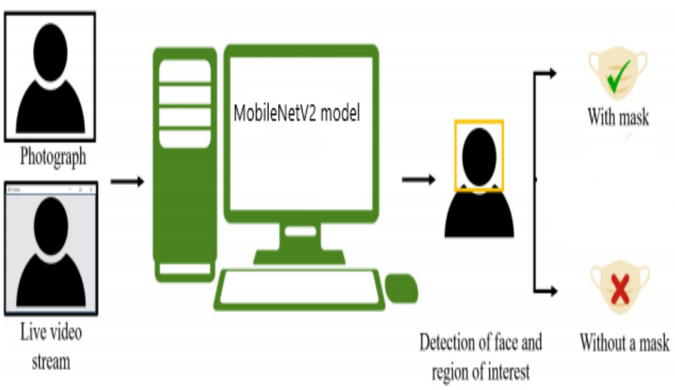


Fig. 1: Image or live video stream is given as an input to the model.

II. Methodology

Pre-processing: Augmentation

Pre-processing is a first step in improving image information such as elimination distortion so that it can be used to process information better. In this study, we use several types of pre-processing, namely augmentation, which aims to avoid overfitting so that if the device faced with certain problem with micro differences encountered, the program can still make predictions correctly. Augmentation is a process of increasing variability using training data sets. The training data is then transformed randomly involving random rotation, image resizing, shifting, and so on. The augmentation process in Tensorflow itself replaces the training data with new training data that has been transformed according to the wishes of the users themselves. The augmentation process itself in Tensorflow does not add to training data but transforms it. There are several augmentation parameters that have been added in this program: rotation, flip, and image shift from the original point (range shift). These augmentation parameters are needed because the variations in image capture using the camera are quite varied: face tilt, camera flip direction, and images that move at a position not at a certain point. We have opted for a geometric method for image augmentation in this paper. Luke Taylor et al.[5] used the Generic Data Augmentation to improve deep learning.

Image Classification Using MobileNetV2

The system has been worked with the following classifier: MobilenetV2: It is a state of the art model for mobile visual recognition including classification, object detection and semantic segmentation. This classifier uses Depth wise Separable Convolution which is introduced to dramatically reduce the complexity cost and model size of the network, and hence is suitable to Mobile devices, or devices that have low computational power. In MobileNetV2, another best module that is introduced is inverted residual structure. Non-linearity in narrow layers is deleted. Keeping MobileNetV2 as backbone for feature extraction, best performances are achieved for object detection and semantic segmentation.
When compared to MobileNetV1 and ShuffleNet 1.5, the use of MobileNetV2 has a better performance in terms of the computational load and has an effect on the shorter iteration time [6].

MobileNetV2 is an architecture for describing neural networks using a pre-trained model. This pre-trained model is a training model in the form of an artificial neural network whose parameters are already owned and ready for adding layers or what is commonly called fine-tuning.

Face Recognition Using Caffe Model

Caffe Model is used to get features of image. These features are obtained using previously trained training. Some of the abilities used include object classification, studying the semantic features of an image, and object detection. Caffe provides a reference model for solving visual problems. In this study, the Caffe Model is used to perform a face selection in the image produced by the input device. This needs to be done so that the prediction process that has been carried out at the training data stage produces the expected value. To use Caffe Model, it is necessary to use several files that OpenCV requires: a file with the extension .prototxt containing the neural network configuration, and the file which contains the weight values of the previously trained model.

Proposed Model

The classification model is built using a *Convolutional Neural Network (CNN)* for object classification [7] as shown in Fig 3 as the different layers of CNN for the task of object detection. The structure of the CNN consists of the following layers:

Table 1- Convolutional Neural Network for object classification

INPUT 224 ✖224 image
Layer 1 3✖3 Conv2d, Activation= Rectified Linear Unit (ReLU) Pooling = Avgpool2d(7,7)
Layer 2 3✖3 Conv2d, Activation= Rectified Linear Unit (ReLU) Pooling Avgpool2d(7,7)
Layer 3 Flatten Layer
Layer 4 Dropout Layer with a dropout rate of 0.5
Layer 5 Dense Layer with activation Relu
Layer 6 Dense Layer with activation Softmax

In Layer 1, Conv2d is used to scan images which do not have a moving frame like a video. It is used to create a kernel that is wind with layers input which helps produce a vector of outputs. These vectored outputs are the input to the further layers in the network. It is supported by the Rectified Linear Unit (ReLU) activation function. ReLU is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. Average pooling[8] layer is used for operation for 2D spatial data. It down samples the input representation by taking the average value over the window defined by pool size for each dimension along the features axis. Flatten Layer[9] is used to flatten and convert the data into a 1-dimensional array for inputting it to the next layer. We flatten the output of the convolutional layers to create a single long feature vector.

Dropout[10] is a technique where some neurons are dropped or ignored during training in order to avoid repeated feature training. They are “dropped-out” randomly. Dense layer represents a fully connected layer, which means that all the individual neurons in a single layer are connected to the neurons present in the next layer.

Face mask detection has been accomplished by adopting Deep Learning techniques. We have designed our project into two phases: training face mask detector and implementing face mask detector. Fig 2 depicts the training and detection phase of our face mask detector model. The dataset is loaded for the model to be trained and the model is serialized in the training phase. Further, the trained model is loaded, the faces are detected in images and video streams and then the region of interest (ROI) is extracted. Finally, the face mask detector is applied and the images or faces in the video streams are classified as with a mask, improperly worn mask, without a mask.

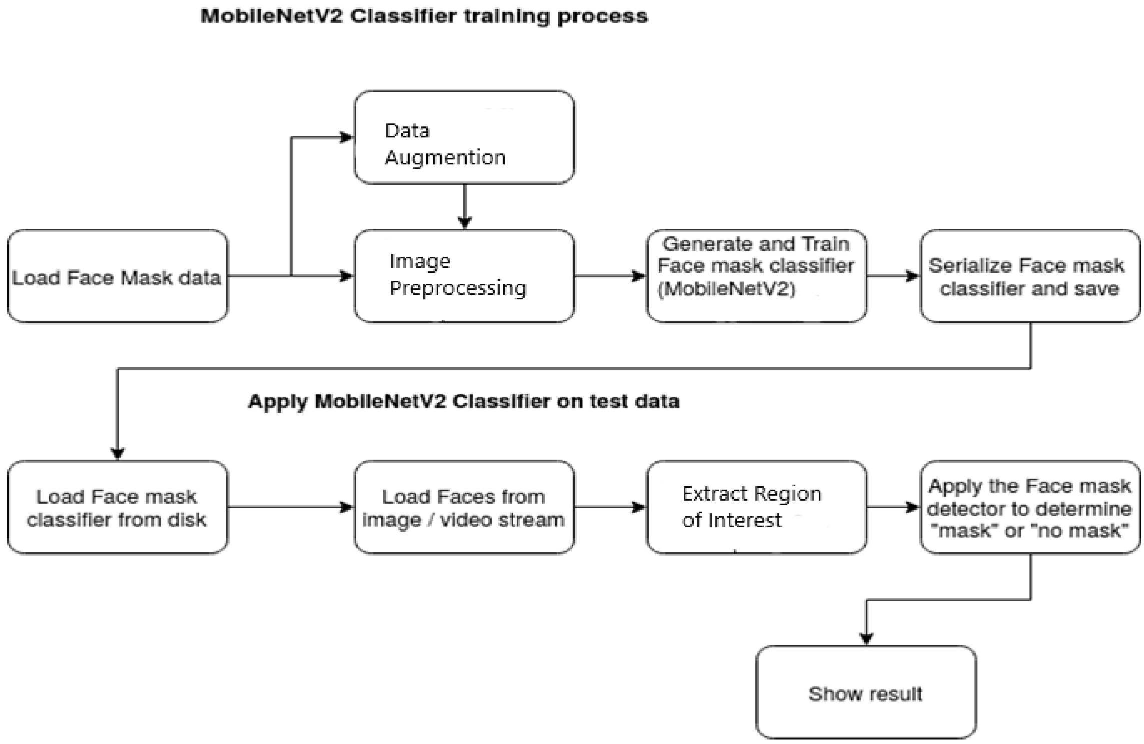


Fig. 2: Training and detection are the two phases of our face mask detector model. The dataset is loaded for the model to be trained and the model is serialized in the training phase. The trained model is loaded, the faces are detected in images and video streams and then the region of interest(ROI) is extracted. Finally, the face mask detector is applied and the images or faces in the video streams are classified as with a mask or without a mask.

III. RESULTS

In this section, the performance of the model is analyzed with respect to the accuracy detection of the image with the mask and without the mask. Further mask with the different geometric shapes are augmented on the detected face without the mask.

Mask detection using Convolutional Neural Network (CNN)

The images of the detected face are analyzed for the presence and absence of the mask. The mask detection using CNN classifies the given image into with_mask and without_mask categories with an accuracy of **almost 99 %.**

Accuracy is defined as:

$$\frac{TP+FN}{TP+FN+FP+TN}$$

Where, True Positives (TP): a measure of outcomes in which the model is able to correctly classify the positive class.
True Negatives (TN): a measure of outcomes in which the model is able to correctly classify the negative class.
False Positives (FP): a measure of outcomes in which the model is able to incorrectly classify the negative class.
False Negatives (FN): a measure of outcomes in which the model is able to incorrectly classify the negative class.
Training accuracy is defined as the accuracy of a model on examples it was constructed on.
Validation accuracy is defined as the accuracy of a model on a new set of data. The graph of loss with respect to the number of epochs is obtained as shown in the Fig 3.

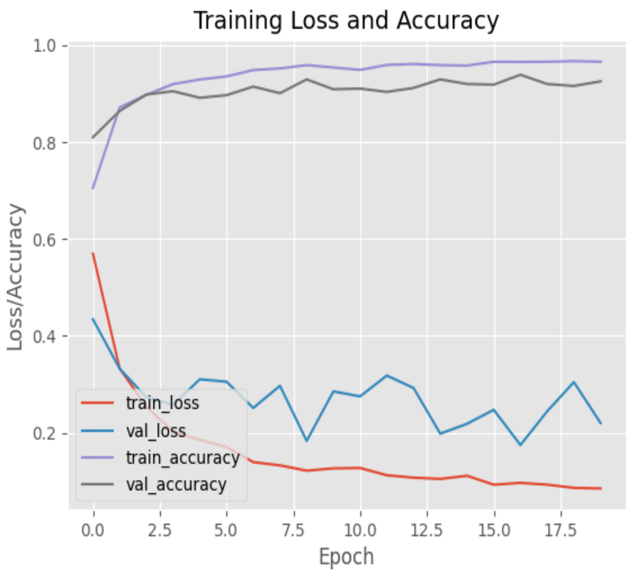


Fig3. Training accuracy vs. Validation accuracy

The model is trained for 20 Epochs. The plot represents the four attributes namely, train_loss, val_loss, train_accuracy and val accuracy obtained in the training for each Epoch. The plot shows almost 99% accuracy on the test set with little signs of overfitting with the validation loss lower than the training loss.

The degree of over fitting is measured by the gap between training and validation accuracy in Figure 3. The larger the gap, the higher the over fitting. The above graph shows that the validation accuracy increases with increase in the number of epochs in the model.

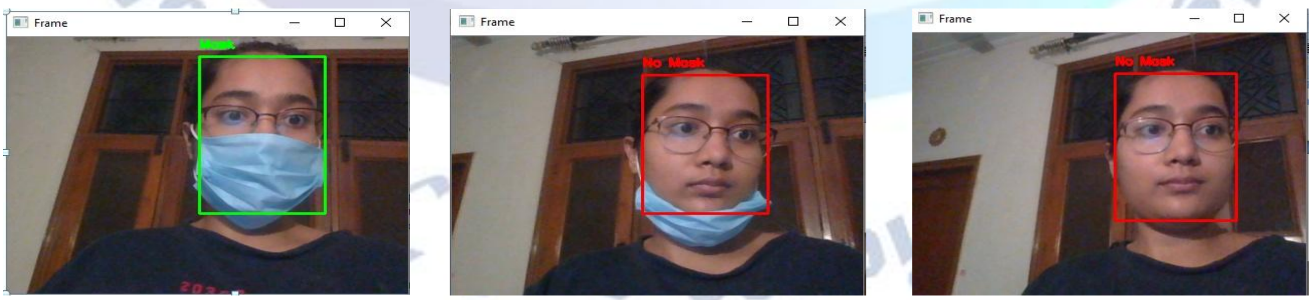
the classification without a mask). The model that has been made has been able to work properly

The predictions of whether a face mask is present or not is made using a predict_mask function which accepts three parameters: frame, facenet and masknet to predict whether a person is wearing facemask or not. The results in Fig.4 are obtained as:

Fig4 (a): Using Mask:
When the prediction Mask > withoutMask.
Fig 4 (b),(c): Partial Mask and No Mask
when the prediction Mask < withoutMask.

Applications of the face mask detection solution:

Airports**:** The Face Mask Detection System can be used at airports to detect travelers without masks. Face data of travelers can be captured in the system at the entrance. If a traveler is found to be without a face mask, their picture is sent to the airport authorities so that they could take quick action. If the person's face is already stored, like the face of an Airport worker, it can send the alert to the worker's phone directly.



(a) (b) (c)

Fig.4: Performance for Each Case of Face Mask (a) Using Mask; (b) Partial Mask; and (c) Without Mask

Performance of model

When it comes to the performance of the model, the system can detect facial images without a mask, with a mask, and partial mask (included in

Hospitals: Using Face Mask Detection System, Hospitals can monitor if their staff is wearing masks during their shift or not. If any health worker is found without a mask, they will receive a notification with a reminder to wear a mask. Also, if quarantine people who are required to wear a mask, the system can keep an eye and detect if the mask is present or not and send notification automatically or report to the authorities.

Entrance Exam Centers: Amid COVID-19, the Joint Entrance Examinations (JEE) Main for admission to the Indian Institutes of Technology (IIT) is being held with precautions. Face Mask detection will be one of the most valuable precautionary measure to be installed where safety of students teachers involved is maintained.

IV. CONCLUSION

This paper successfully proposes a model which is used to find video analytics solution for covid-19 that is the face mask detection system. This work distinguishes face masks from images and live video streams using Convolutional Neual Network. On training the model, I got an accuracy of approximately 99 %. MobileNetV2 model was used for image classification. This classifier was then implemented to images and live video streams. The faces were recognized in images and videos and these faces were extracted. Then, our face mask classifier was applied to achieve the required results. The green and red rectangular frame respectively represent that facemask is detected or not . The output is generated as in Fig. 4 showing the output of the project.

In future works, I plan to apprehend a dataset with more number of entries for the training and testing part. I would like to consider more classification techniques and compare their accuracies with the current solution. Along with that, I would like to propose more video analytics solutions like detection of PPE kits, social distance monitoring and crowd intensity detection. I would also like to add an additional feature of generating alerts via SMS/ mail in this model.

REFERENCES

[1] J. Howard et. al., Face Masks Against Covid-19: An Evidence Review, Medicine and Pharmacology, 2020.

[2] P. Bahl, C. Doolan, C. de Silva, A. A. Chughtai, L. Bourouiba, and C. R. MacIntyre, "Airborne or Droplet Precautions for Health Workers Treating Coronavirus Disease 2019?," J. Infect. Dis., no. Xx Xxxx, pp. 1–8, 2020, doi: 10.1093/infdis/jiaa189.

[3] Liping Yuan, Zhiyi Qu, Yufeng Zhao, Hongshuai Zhang, Qing Nian, "A Convolutional Neural Network based on Tensorflow for Face Recognition", 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 25-26 March 2017, Chongqing, China.

[4] J. Wang, L. Pan, S. Tang, J. S. Ji, and X. Shi, "Mask use during COVID-19: A risk adjusted strategy," Environ. Pollut., vol. 266, no. 7, p. 115099, 2020, doi: 10.1016/j.envpol.2020.115099.

[5] Taylor, L., & Nitschke, G. (2018). Improving Deep Learning with Generic Data Augmentation. *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1542-1547.

[6] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474

[7] Sharma, N., Jain, V., & Mishra, A. (2018). An Analysis Of Convolutional Neural Networks For Image Classification. *Procedia Computer Science, 132*, 377-384.

[8] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella and J. Schmidhuber, "Fast image scanning with deep max- pooling convolutional neural networks," *2013 IEEE International Conference on Image Processing*, Melbourne, VIC, 2013, pp. 4034-4038, doi: 10.1109/ICIP.2013.6738831.

[9] Culurciello, E., Jin, J., Dundar, A., & Bates, J. (2013). An Analysis of the Connections Between Layers of Deep Neural Networks. *ArXiv, abs/1306.0152*.

[10] "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov*; 15(56):1929−1958, 2014