iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56652

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

Real Time Face Mask Detection Using Inception-V3

Swarali Degaonkar[1], Aarti Agarkar[2]
Computer Department, Marathwada Mitra Mandal's College of Engineering

Abstract: This project was developed for detecting people's faces and segregating them into two classes, masks and without masks is done with the help of image processing in real-time and deep learning. The proposed model is built by fine-tuning the pre-trained deep learning model, InceptionV3.The model was trained on a WIDER dataset of 8,262 images. The model performed greatly achieving an accuracy of 99% overall.
Keywords: Deep Learning. Transfer Learning. Face Detection. Convolutional neural network. InceptionV3

I. INTRODUCTION

It's been 2 years since the Covid pandemic hit us which forced us to stay home. The corona virus or the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a virus which attacks the person's respiratory system, more importantly affecting the lungs of the human body. Globally, as of 11 April 2022, there have been 497,057,239 confirmed cases of COVID-19, including 6,179,104 deaths, reported to WHO. The virus spreads like an infectious disease that is released from an infected person when they speak, sneeze, or coughs by respiratory droplets. This spreads quickly through close contact with anyone infected, or by touching objects or surfaces affected with a virus. Hence WHO recommended mandatory usage of masks and maintaining a social distance of 6ft at all times.

So, wearing masks definitely reduces your contact with viruses almost to 80%. Since it has now been declared that the virus is actually airborne it is recommended to wear masks to avoid any contact and to use sanitizing features, also by wearing n95 masks which have an efficiency of 95%.

Since India has such a huge population it's hard to monitor each person, if he/she is following the rules or therefore, face mask detection monitoring has become a crucial computer vision and deep learning task to help the global society. The pandemic of COVID-19 forced each sector to work on it. As a result, Deep learning and Machine learning helped the scientists, researchers in analysing and evaluating the vast data forecast of the distribution of COVID-19. This has also helped in predicting an early warning mechanism for the upcoming 2nd wave of the pandemic, and to classify vulnerable populations. The provision of healthcare needs funding, and in order to better understand infection rates and to trace and quickly detect infections. People are forced by laws to wear face masks in public in many countries. These rules and laws were developed as an action to the exponential growth in cases and deaths in many areas. However, the process of monitoring large groups of people is becoming more difficult. The monitoring process involves the detection of anyone who is not wearing a face mask. Hence, we introduce a mask face detection model that is based on computer vision and deep learning.

II. LITERATURE REVIEW

Generally, most of the publication focus is on face construction and identity recognition when wearing face masks. In this research our focus is on recognizing the people who are not wearing face masks to help in decreasing the transmission and spreading of the COVID-19. Researchers and scientists have proved that wearing face masks help in minimizing the spreading rate of COVID-19. Deep learning technique has been useful for big data analysis and has its applications in computer vision, pattern and speech recognition, etc. Liu's et al [7]. However, face detection is more challenging because of some unstable characteristics, for example, glasses and beard will impact the detecting effectiveness. Moreover, different kinds and angles of lighting will make detecting faces generate uneven brightness on the face, which will have influence on the detection process. To overcome these problems, the system used Adaboost algorithm [13] implemented using Haar classifiers for face detection and PCA algorithm [11] for face recognition implemented using face recognizer function of OpenCV. Since 2004, face detection is performed fairly easily and reliably with Intel's open-source framework called OpenCV [13]. This framework has an in- built Face Detector that works in roughly 90-95% of clear photos of a person looking forward at the camera.

Viola and Jones devised an algorithm, called Haar Classifiers, to rapidly detect any object, including human faces, using Adaboost [14] classifier cascades that are based on Haar-like features and not pixels. [10] OpenCV uses Viola Jones method published in 2001, to detect faces using 4 key concepts: 1) Simple rectangular features called Haar features, 2) An integral image for rapid face detection, 3) The Adaboost machine learning method, 4) A cascaded classifier to combine many classifiers efficiently. [10]

In machine learning, a Convolutional Neural Network (CNN) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multi-layer perceptron designed to require minimal prepossessing [5]. These biologically inspired computational models are able to far exceed the performance of previous forms of artificial intelligence in common machine learning tasks. These CNN models are designed to make classification like the human brain. Convolutional Neural Network (CNN) is a structured deep learning process that plays a ground-breaking push for a variety of applications focusing on computer vision and image-based applications. The fields in which CNN is prevalently used are facial recognition, object recognition, classification of images. [3] As for the convolution process portrayed in fig. 1, it begins through the extracted input image along with its features using a filter of 3x3 across with a stride of 1 as mentioned to be convolution (Conv). The resulting output from the Conv process produces a featured map through the dot product of the preceding Conv layer. Each featured map keeps precise details of the original image to establish a specific input and it will be down-sampled through the ReLU method to go on other values intact and downgrade negative values to zero values. Additional down-sampling procedure following each Conv named as max-pooling decreases the values into half of its original value by simply selecting the max values only from the matrix of kernel. Providing the primary clues in identifying a precise image for flexible handling of resources is the work of the pooling layer. The pooled-features are distributed and flattened in the fully connected layers (FC layers) that translate the activation from one or zero values. Then, the SoftMax-activation function produces probabilities through its neural networks in classifying input data. [4] A common prescription to a computer vision problem is to first train an image classification model with the ImageNet Challenge data set, and then transfer this model's knowledge to a distinct task. It allows model creation with significantly reduced training data and time by modifying existing rich deep learning models.

The concept has a name: Transfer Learning. The common practice is to truncate the last layer (SoftMax layer) of the pre-trained network and replace it with our new SoftMax layer that are relevant to our own problem. Essentially, instead of starting the learning process from a (often randomly initialized) blank sheet, we start from patterns that have been learned to solve a different task Two common approach may be used: develop model approach and pre-trained model approach. We chose the second approach which consists of: - Select Source Model. A pre-trained source model is chosen from available models. Many research institutions release models on large and challenging datasets that may be included in the pool of candidate models from which to choose from. We used the Inception-v3 model. -Reuse Model. The model pre-trained model can then be used as the starting point for a model on the second task of interest. -Tune Model. Optionally, the model may need to be adapted or refined on the input-output pair data available for the task of interest. [8] Inception-v3 is one of the pretrained models on the TensorFlow. It is a rethinking for the initial structure of computer vision after Inception-V1, Inception-v2 in 2015. The Inception-v3 model is trained on the ImageNet datasets, containing the information that can identify 1000 classes in ImageNet, the error rate of the top-5 is 3.5%. TensorFlow also provides detailed tutorials for us to retrain Inception's final Layer for new categories using transfer learning. Inception-v3 network model is a deep neural network, it is very difficult for us to train it directly with a low configured computer, it takes at least a few days to train it. TensorFlow provides tutorials for us to retrain. Inception's final Layer for new categories using transfer learning. We use the transfer learning method which keeps the parameters of the previous layer and removes the last layer of the Inception-v3 [9] model, then retrains the last layer. The number of output nodes in the last layer is equal to the number of categories in the dataset. For example, the ImageNet dataset has 1000 classes, so the last layer has 1000 output nodes in the original Inception-v3 model.

III. PROPOSED METHODOLOGY

TensorFlow is a GPU based platform and hence needs a GPU backed run-time to work faster. Even though it does work on CPU, the running time is remarkably slow for training hundreds of epochs. Hence to overcome this we will be using Google Collab an initiative by Google to give free cloud-based GPU runtime. Deep learning models need a huge dataset to train on and accessing the dataset and training the model takes a lot of speed and accurate runtime which is available on Google Collab. The model is trained on google Collab environment which has great speed of execution. It becomes very easy to build models and train them with such fast GPU runtime. The pre-trained chosen was InceptionV3.Inception v3 is a widely-used image recognition model that has been shown to attain greater than 78.1% accuracy on the ImageNet dataset. The model is the culmination of many ideas developed by multiple researchers over the years.

It is based on the original paper:" Rethinking the Inception Architecture for Computer Vision" by Szeged, et. al. [9] The whole process was divided into 2 phases:

1) *Model was trained for static images:* The datasets were uploaded on Google Collab notebook for accessing. The dataset went through the necessary image augmentations for the CNN model to be accepted. The image size was set to 160x160 px. The model was also made to undergo dimension reductions for the optimal training and to retain computational speed. The model was successfully able to classify the images as" Mask" or" no Mask".
2) *Model was trained for Real-Time video input:* For this phase we used the concept of Computer Vision and the OpenCV library for manipulating the real time video input frames. For more faster classification real time, Haar cascade classifier is used to detect face. Haar Cascade accepts the images in Black and white, hence with the help of OpenCV library, the captured RGB frames are converted to gray images. These frames are then augmented according to the model's parameters and sent for prediction in the model. According to the prediction a box is drawn around the detected face indicating wearing Mask or Not. Finally, the Realtime captured video footage is saved as an output file with the predictions for further classification. Let's see the working in detail:

A. Dataset

Deep learning models learn extensively from the datasets. So, it is recommended that large datasets should be used to train the model the for the better accuracy. For this project we have used the WIDER dataset this dataset contains two folders named mask and no-mask and contains a total image of 8,262 items. WIDER FACE dataset is a face detection benchmark dataset, of which images are selected from the publicly available WIDER dataset. We choose 32,203 images and label 393,703 faces with a high degree of variability in scale, pose and occlusion as depicted in the sample images. WIDER FACE dataset is organized based on 61 event classes. For each event class, we randomly select 40%/10%/50% data as training, validation and testing sets. These datasets have images which are augmented to have only the specific features of the face and have similar simulated images of people wearing mask, this helps is retaining the features while passing through each layer of the CNN model and reduces the burden of image augmentation processes.

B. Image Augmentation and Pre-processing

Image augmentation is a technique used to increase the size of the training dataset by artificially modifying images in the dataset. The input images and the images in the dataset are in a pixel range of 0 - 255, CNN models usually cannot take such a high pixel range. Hence the first image augmentation operation is to normalize the pixel range 0 - 1. This results in feature retention while the images pass through the multiple layers of the model. In this research, the training images are augmented with eight distinct operations namely resizing, shuffling, flipping horizontally, rotating, zooming. Then the images are resized to 160x160 px for every batch.

C. Feature Extraction in InceptionV3

Deep neural networks are used for image classification because of their better performance than other algorithms. But training a deep neural network is expensive because it requires high computational power and other resources, and it is time consuming. In order to make the network to train faster and cost-effective, deep learning-based transfer learning is evolved. Transfer learning allows to transfer the trained knowledge of the neural network in terms of parametric weights to the new model. Transfer learning boosts the performance of the new model even when it is trained on a small dataset. Creating model from scratch takes up too much computational speed and time, hence here comes the of concept of using Transfer Learning, where we use a pretrained model which is used to classify more than 1000 classes to train on our own specific dataset. This reduces the training time and the overall computational speed for the execution.

In this paper, a transfer learning-based approach is proposed that utilizes the InceptionV3 pre-trained model for classifying the people who are not wearing face mask. Inception-v3 is a convolutional neural network that is 48 layers deep. Since the model has been trained on the ImageNet dataset, they have provided some pre-trained weights such as" image net" have been provided. Google Net devised a module called inception module that approximates a sparse CNN with a normal dense construction. Since only a small number of neurons are effective, the width/number of the convolutional filters of a particular kernel size is kept small. Also, it uses convolutions of different sizes to capture details at varied scales (5X5, 3X3, 1X1). In addition to this, it has a bottleneck layer which helps in the massive reduction of the computation requirement. For this project we have removed the last layer of the model and stacked it with a sequential model of 3 layers each with some normalization features which marginalized dimension reduction.

Dimension reductions creates a bottleneck effect which immensely increases the computational speed through the layers, hence to retain the image features we must stack the model with some output layers like Flatten, Dense, Dropout, Batch Normalization and the activation layers of Relu. The Model is able to classify images perfectly in mask and no mask, getting an accuracy on 99.9%.
To redeem the features which could be lost during dimension reductions, we have added the following layers:

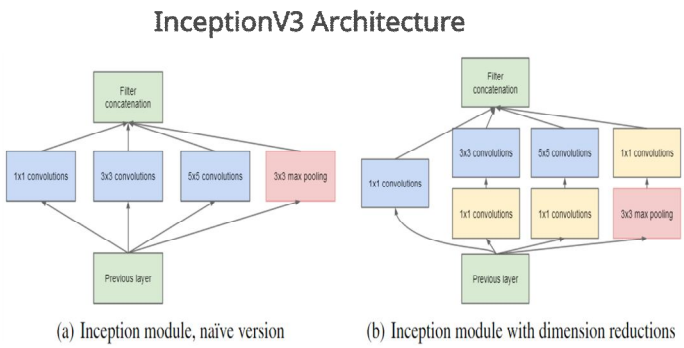


Fig -1: InceptionV3 Architecture

D. Haar Cascade Classifier for Face Detection

So, what is Haar Cascade? It is an Object Detection Algorithm used to identify faces in an image or a real time video. The algorithm uses edge or line detection features proposed by Viola and Jones in their research paper. First published in their 2001 paper [13], Rapid Object Detection using a Boosted Cascade of Simple Features, this original work has become one of the most cited papers in computer vision literature. The algorithm is given a lot of positive images consisting of faces, and a lot of negative images not consisting of any face to train on them. The repository has the models stored in XML files, and can be read with the OpenCV methods. Haar cascade is one of the most efficient classifiers for face detecting in real time. It returns the coordinates of the image along with its height and width. It the whole frame as an input and detects the face from it. One of the conditions for passing parameters to the Haar cascade classifier is that it accepts grey scale images, hence we first convert the captured frame into grey scale image and then activate the classifier.

E. Algorithms

The proposed system focuses on how to identify a person wearing a mask on the image/video stream. Help with computer vision and deep learning algorithms by using OpenCV, Keras library. Method: 1. Train deep learning model (InceptionV3). 2. Apply a face mask detector over images/live video streams.

1) Step 1. Data visualization

In the first step, let us visualize the total number of images in the dataset in these two categories. We can see that there are 8,262 images in the dataset. The dataset is divided into Training, Validation and Test, which are further divided into two folders of Masked images and Non mask images.

2) Step 2. Data Augmentation:

In the next step, we expand the data set to include a larger number of images for training. In this step of data expansion, we rotate and flip each image in the data set.

3) Step 3. Splitting the data

The data is already split into train, test and validation and hence their path is allocated for google collab to access them.

4) Step 4. Building the Model:

We have selected the pre-trained model InceptionV3, we have used the pre-trained weights of 'imagnets' for the model. The last layer is then removed to create a bottleneck effect, we then stack a sequential model over it with layers like Flatten, Dense ReLU, Dropout with decreasing value till we connect the final output layer with the sigmoid activation function.

Finally, the model is compiled using the RMSprop compiler with 'binary cross entropy' since we have 2 classes to classify.

5) Step 5. Haar Cascade

Once the model is trained on static images, we can use one real time video input as well. With the help of OpenCV library we can access the available camera for video input. We use the OpenCV library to run an infinite loop to use our webcam, where the cascade classifier is used to detect faces. The code webcam =cv2.videoCapture (0) indicates the usage of the webcam. Once the frame is captured with the help of OpenCV, we convert them to grey scale images for the classifier to detect faces among the frames. Once the face is detected the image is sent for more preprocessing for it to pass the layers of the model.

6) Step 6. Classification

Finally, the model does the classification of the images as "Mask" or "No Mask", and we draw boxes around the detected faces according to the prediction, with help of the OpenCV library. Finally, the predicted output is stored as an output file for further classification.

IV. OUTPUT

The model works perfectly and is able to classify the images both as static and as a live input stream. The model gave a great accuracy of 99.9% on the training set as well. Hence the live frame is captured with the help of OpenCV library which then converts the frame to a greyscale image for the Haar Cascade Classifier. The classifier then returns the face detected image which is then augmented according to the model's parameters for prediction. The model then classifies the image accordingly as Mask or No Mask.

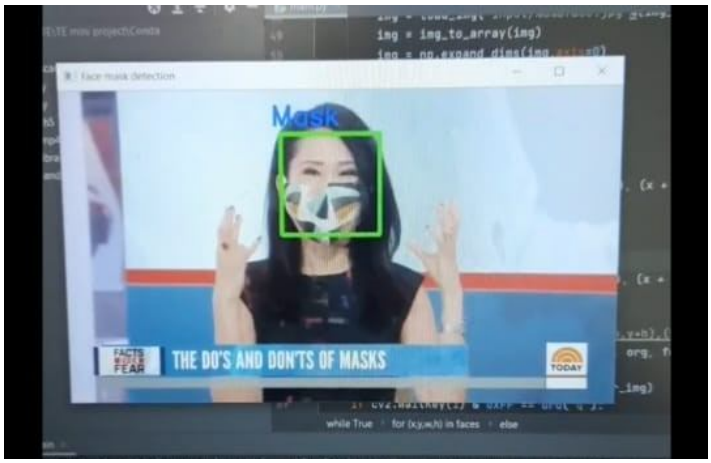


Fig -2: Output with Mask

Fig -3: Output with Without Mask

```
138/138 [==============================] - 11s 78ms/step - loss: 0.0198 - accuracy: 0.9957 - val_loss: 0.5053 - val_accuracy: 0.9068
Epoch 96/100
138/138 [==============================] - 11s 78ms/step - loss: 0.0197 - accuracy: 0.9957 - val_loss: 0.5274 - val_accuracy: 0.9068
Epoch 97/100
138/138 [==============================] - 11s 78ms/step - loss: 0.0198 - accuracy: 0.9959 - val_loss: 0.5056 - val_accuracy: 0.9052
Epoch 98/100
138/138 [==============================] - 11s 78ms/step - loss: 0.0195 - accuracy: 0.9959 - val_loss: 0.5331 - val_accuracy: 0.9062
Epoch 99/100
138/138 [==============================] - 11s 78ms/step - loss: 0.0195 - accuracy: 0.9957 - val_loss: 0.5462 - val_accuracy: 0.9041
Epoch 100/100
138/138 [==============================] - 11s 78ms/step - loss: 0.0191 - accuracy: 0.9961 - val_loss: 0.5447 - val_accuracy: 0.9057
```

Fig -4: Training Output with accuracy percentage.

V. CONCLUSION

The world is facing a huge health crisis because of pandemic COVID-19. The governments of various countries around the world are struggling to control the transmission of the coronavirus. According to the COVID-19 statistics published by many countries, it was noted that the transmission of the virus is more in crowded areas. Many research studies have proved that wearing a mask in public places will reduce the transmission rate of the virus. Therefore, the governments of various countries have made it mandatory to wear masks in public places and crowded areas. It is very difficult to monitor crowds at these places. So in this paper, we propose a deep learning model that detects persons who are not wearing a mask. This proposed deep learning model is built using transfer learning of InceptionV3. In this work image augmentation techniques are used to enhance the performance of the model as they increase the diversity of the training data. The proposed transfer learning model achieved accuracy and specificity of 99.0% on the WIDER dataset. For real time monitoring Haar cascade Classifier is used for the best face detection. The real time footage is also saved as an output file for further working phase. The same work can further be improved by employing large volumes of data and can also be extended to classify the type of mask, and implement a facial recognition system, deployed at various workplaces to support person identification while wearing the mask. All establishment/offices must employ measures to keep a check on the mandate of wearing mask always and every time when out of house. Instead of manual checking, use of technology can be resorted to which not only automates the task of detection but also prevents diversion of substantial manpower to keep a manual check on defaulters. With phase 4 of the future enhancement as described below, a miniature module would be able to perform the mask detection task and would be created as a total screening and monitoring device which could be operated from any remote location. This module will not only automate the entire process single handedly but also would be small, compact and easy to install.

VI. FUTURE WORK

The model proposed is presently in phase I of operation and the demo video shown is a direct application of it. However, in future, following enhancements/ improvements can be superimposed to facilitate the ease of use and simplicity. These are described below: -

1) Phase-I: Current version uses Laptop with relevant software installed

As such, the requirement of such a loaded laptop with a web-cam is a must for execution of the software. Therefore, the laptop should be installed at the point of inspection itself and some skilled operator should be present to type and execute the relevant commands.

2) Phase-II: Replacement of laptop with a compact model with camera which could be easily integrated for video capture

The plan is to replace the laptop with a workable model/ device such as Raspberry Pi which needs to be installed at the site of inspection/ site. Raspberry Pi models are compact in size. This will reduce installation complexities by 90%. This device/ model will be attached to a laptop only for execution of commands with help of LAN cable. As such commands can be run by a remote operator also and the requirement of a laptop present at the site is negated.

3) Phase-III: Removal of laptop for all purposes. Execution only with help of model

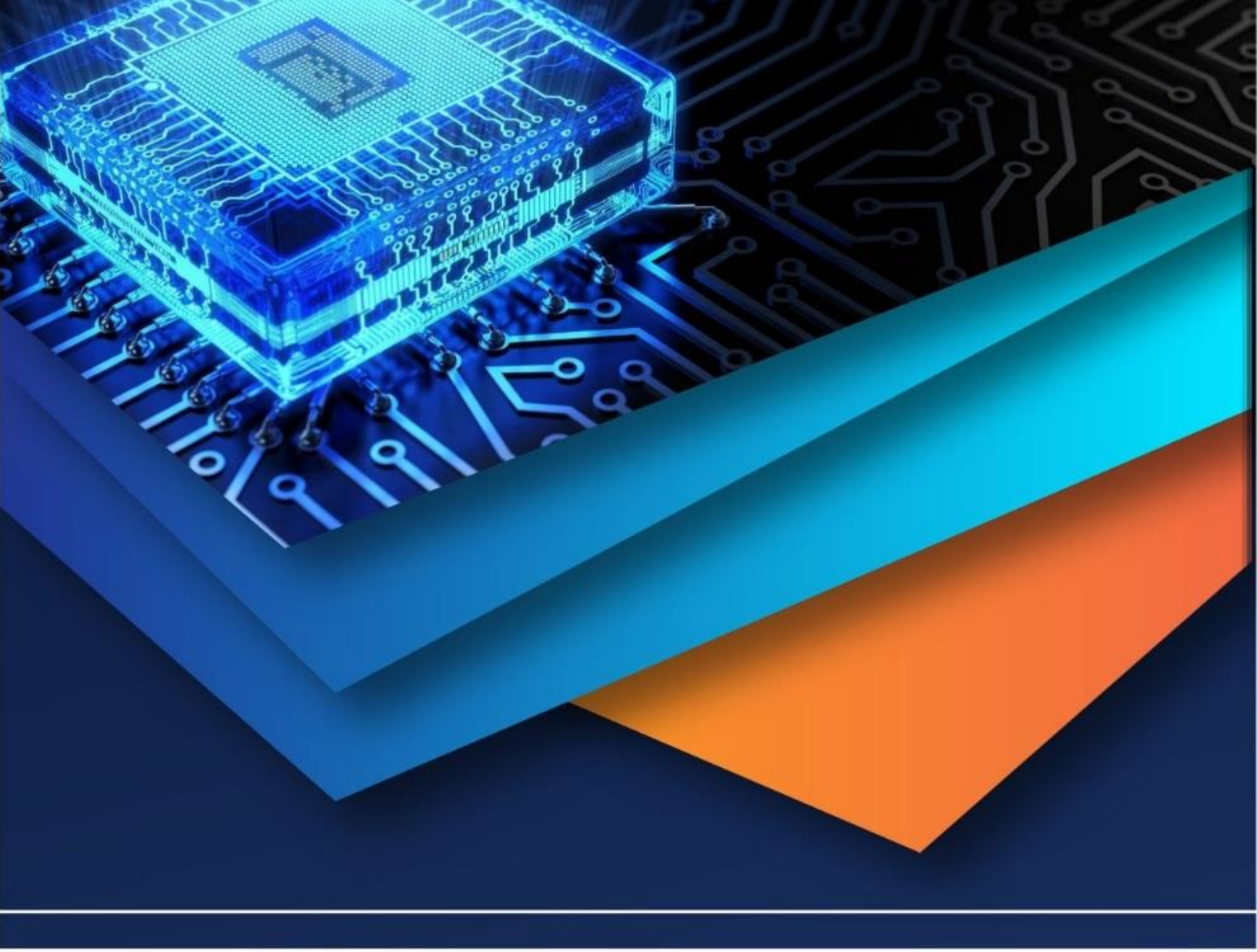
There will not be any constraint of the laptop for executing any command to run the model either at the site or at any remote location. The device which is comparable to the size of a credit card with attached camera and loaded with necessary algorithms/code would meet the requirements. There would not be any obligation of trained/ semi trained individuals to run the commands. A normal sentry/ operator should be able to operate the model.

4) Phase-IV: Integrating models with some other models for a total screening and monitoring

In future work, we will propose a deep learning framework by using pre-trained deep learning models to monitor the physical interaction (social distancing) between individuals in a real-time environment as a precautionary step against the spread of the COVID-19. The current ongoing system is graced with InceptionV3 classifier, one of the best systems which would be implemented along with the interface of the alarm alerting system in future and future generations

REFERENCES

[1] Dong Chen, Shaoqing Ren, Yichen Wei, Xudong Cao, and Jian Sun. Joint cascade face detection and alignment. In the European conference on computer vision, pages 109–122. Springer, 2014.
[2] Sachin Sudhakar Farfade, Mohammad J Saberian, and Li-Jia Li. Multi-view face detection using deep convolutional neural networks. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pages 643–650, 2015.
[3] Andreas Kamilaris and Francesc X Prenafeta-Bold´u. Deep learning in agriculture: A survey. Computers and electronics in agriculture, 147:70–90, 2018
[4] Prateek Khandelwal, Anuj Khandelwal, Snigdha Agarwal, Deep Thomas, Naveen Xavier, and Arun Raghuraman. Using computer vision to enhance the safety of the workforce in manufacturing in a post covid world. arXiv preprintarXiv:2005.05287, 2020
[5] Sanjay Kumar, Ashish Negi, JN Singh, and Himanshu Verma. A deep learning for brain tumor mri images semantic segmentation using fcn. In 2018 4th International Conference on Computing Communication and Automation (ICCCA), pages 1–4. IEEE, 2018.
[6] Narinder Singh Punn, Sanjay Kumar Sonbhadra, Sonali Agarwal, and Gaurav Rai. Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques. arXiv preprint arXiv:2005.01385, 2020.
[7] Bosheng Qin and Dongxiao Li. Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19. Sensors, 20(18):5236, 2020.
[8] MH Ramafiarisona and PA Randriamitantsoa. Using convolutional neural networks to perform content-based image retrieval systems.
[9] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2818–2826, 2016
[10] Prathamesh Timse, Pranav Aggarwal, Prakhar Sinha, and Neel Vora. Face recognition-based door lock system using opencv and c# with remote access and security features. International journal of engineering research and applications, 4(4):52–57, 2014.
[11] Matthew Turk and Alex Pentland. Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1):71–86, 1991.
[12] Isunuri B Venkateswarlu, Jagadeesh Kakarla, and Shree Prakash. Face mask detection using mobilenet and global pooling block. In 2020 IEEE 4th Conference on Information & Communication Technology (CICT), pages 1–5. IEEE,2020.]
[13] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, volume 1,
[14] Phillip Ian Wilson and John Fernandez. Facial feature detection using haar classifiers. Journal of Computing Sciences in Colleges, 21(4):127–133, 2006.
[15] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503, 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY