RESEARCH ARTICLE

SPECIAL ISSUE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
Vol. 05, Issue 05S May
www.rspsciencehub.com

Check for updates

http://dx.doi.org/10.47392/irjash.2023.S010

International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER) 2023

Face Mask Detection using MobilenetV2 of Convolution Neural Network

S V Hemanth[1], *K Sanjeevaiah*[2], *Dharmendra Kumar Roy*[1], *K Nagaraju*[3]

[1]Associate Professor, Department of Computer Science and Engineering, Hyderabad Institute of Technology and Management, Telangana, India

[2]Assistant Professor, Department of Computer Science and Engineering, Hyderabad Institute of Technology and Management, Telangana, India

[3]Assistant Professor, Department of Computer Technology, Kavikulguru Institute of Technology and Science, Ramtek, Nagpur, Maharashtra, India

Email: hemanth.sandepaga@gmail.com

Article History

Received: 24 February 2023
Accepted: 6 March 2023

***Keywords*:**

MobileNet V2;
Kaggle;
image;
respiratory

Abstract

In light of the fact that the COVID-19 pandemic is spreading swiftly over the earth, it is crucial to create new technologies to study and resist the disease. Face masks and gloves are required for protection against the coronavirus, and scientists and doctors have advised everyone to wear the mask for the whole day. Thus, various procedures are accessible to different people wearing face masks. Masks are advised as a basic barrier to stop respirational beads from receiving into the air and against other people once someone is found to be using cover hacks. Additionally, this is known as source governance. This article is based on the current understanding of respiratory beads' function in the escalation of COVID-19 infection. In this problem, the face mask procedure was built using MobileNetV2. Compared to the current system, Mobile Net V2 can be used to identify face masks among individuals with greater accuracy. The input data file contains 500 images taken from the Kaggle face mask Detection Dataset. A scene with a mix of people donning masks and without mask. The output is a segmented picture of the same. Later, this process is improved by using a webcam to capture real-time video. The video is then segmented into the frame and resized as required, and the result is a video-segmented image. The model was then run to determine whether or not individuals were wearing masks after performing the pre-processing function. An accuracy of 80 was used to acquire the results.

1. Introduction

In line with the rapid growth in the global population, COVID-19 cases have also risen vertically. There are 11 million COVID-19 cases that are currently active globally. By the end of 2021, this figure is expected to surpass 50 million, with 1 in every 6 Indians being affected by COVID-19. Implementing the recommended task is essential for the general good of various large organizations and businesses attempting to prevent the dissemination of COVID-19 while preserving their efficiency. In this COVID-19 circumstance. There should only be a certain number of people in each location, and they should always maintain an equal social distance. In this exercise, deep learning tools will be used to determine whether or not someone is wearing a mask.

There are nearly 2,000 images total in each of the two folders that make up the dataset, which have names that include and exclude masks. Face masks are the most important form of protection after the global COVID-19 pandemic, which has created a pressing need for them. The main goal of the initiative is to identify human faces wearing face masks in images and live-streamed videos. The facial detector model was created using a deep learning method. Finally, the fundamental ideas of transfer learning in neural networks output whether or not a facial mask is present in an image or video stream. According to experimental findings, this paradigm works well. A facial mask detection tool can be used to confirm this. Face mask recognition is the method of determining whether or not somebody is wearing a disguise. Due to the fact that identifying a face is the first step in figuring out whether a mask is present, the approach is divided into two parts: identifying faces and identifying masks on people's faces. Face recognition is one use of object identification that has many uses in security, biometrics, law enforcement, and other areas. It is possible to distinguish between a visage with a mask on and one without one using a facial mask detector model. The implementation of the algorithm is on live video streams, images, and videos.

2. Literature survey

In the past, patterns were recognized from close features' centers, contours, and lines to conduct facial recognition patterns. These techniques are used to locate binary examples nearby. These methods require very little computational work and are very effective for processing grayscale pictures (T, Pietikainen, and Maenpaa). The regression function will be fitted to the dataset by the regression-based classification algorithm AdaBoost. During the rollback to optimization results, some missing objects from the initial data awaited correction (Iscide). The temporal object model proposed by the Viola Jones Detector is used in real-time to identify various classes of features. It evaluates any image with contours, lines, and four rectangular features using a basic window size of 24x24. Harrlike features resemble dials that can be used to verify whether a feature is present in an image (P and Jones). Although it performs poorly when the image is in various orientations, this model does not function when the brightness of the image changes. Classification issues are the primary application for the convolutional network. There are various CNN designs, such as the VGG16. This architecture consists of two convolutional layers with an input core size of 224 (64.3x3), a maximum convolutional group of size 2x2, two additional convolutional layers, a maximum group, three additional convolutional layers with the maximum group, and three completely connected layers after that. The FC final has a soft maximum. When compared to AlexNet, this design performs well (Viola and Jones Szegedy et al.). There are roughly 22 layers with convolution and maximum aggregation in the Google Net architecture, which essentially uses the bootstrapping method to create small convolutional layers to reduce the number of parameters. Alex Net has 4 million characteristics (He et al.). In this study, deep neural networks with 152 layers are used to train deeper models, which are 8 times more complex than VGGnet. On the COCO dataset, this method produces comparatively better object detection results (Priya et al.). In this study, ventricle segmentation is carried out using UNet and SEnet. In this model, the weights are sorted so that more weights are given to essential features and fewer weights to less significant ones. (Fu and Qu). The Support Vector Machine will create equations to create the row and classify the objects based on the values mapped to that row in order to conduct classification on the objects. In this article, the semantic segmentation method is used for mask detection; for training, the VGG network was used, and the FCN was used for semantic segmentation of the available faces in images (Kumar et al.). Experiments were conducted on various human analytic datasets, and more accurate results were attained. Processing of medical images was done in this study. They scanned human brains and received training in the highly effective use of FCN to detect tumours. In this work, we used 3D segmentation to detect tumours rather than 2D segmentation [twelfth]. Lakshmi Ramani, Tumuluru, et al. (Malathi et al.), used the CNN model, which is useful in security-related applications, for human face recognition. In this study, they built a face model from various facial features, including the mouth, nostrils, and eyes, and used it to identify differences between faces. Malathi, J. et al (Krishnaveni, Bhavani, and Lakshmi). Its primary emphasis is on spotting fake images that are used in various contexts, including public media and further settings where advertising is necessary. In this piece, various methods for identifying a fake image's features are suggested. These methods include copy displacement and image concatenation attacks, which can be countered by using correlation analysis to identify dual features. Skeletor, et al (Satapathy et al.) suggested a model to determine the quality of iron ore by analyzing material samples' properties in the mining sector. The ore's grade must be evaluated carefully. SVR encourages the use of vector regression for live ore grade measurements. SFFS is a model created using SVR in which a family's 280 features are extracted for object identification. The study of objects has grown to be a crucial component of picture analysis. There are various methods for image processing (Satapathy et al.). In this article, the author introduces a wavelet-based digital network for feature extraction and learning that is efficient for object detection. A license plate detection model was suggested by Satapathy, Sandeep Kumar, et al., et al. (Pathak, Bairagi, and Srinivasu) as a crucial problem to assist police in prosecuting many criminal cases. In order to collect comprehensive information about the owner, the authors used an OCR-based method to identify characters on license plates. These characters were then stored and processed using a client-server model. The accuracy has been enhanced by using an entropy-based CNN, which performs well in the dim lighting present. In order for common people to recognise spice plants, the discovery of medicinal plants becomes a crucial problem (Kishore et al.). The authors of this article have suggested a model using CNN. It was able to identify medicinal plants more precisely after being taught pictures of the leaves of medicinal plants. One of the key studies receiving a lot of focus right now is the detection of human posture (Ravi et al.). The author of this article has put forth a model that, based on a person's pose, can identify a traditional dance. In order to successfully identify the traditional dance, they combined CNN with a variety of traditional dance movements that were trained and taught a pattern. By training a CNN model that can recognise cues in the actual film, as described by Ravi, Sunitha, and colleagues , sinusoidal language detection was carried out. Four driverless vehicles can also benefit greatly from this. Even sign language can benefit from computer translation. The Common Angle Displacement method was used by CNN to improve its capacity to record 3D motion sign language in real-time, which is currently applicable to various uses . In the mining industry, Patel, Ashok Kumar, et al. suggested a model to determine the quality of iron ore by identifying the traits of a material sample the ore's grade must be evaluated carefully. Online measurements of ore quality were made using support vector regression, or SVR. They extract 280 features for object recognition during this process; SFFS is a developed model made with SVR.

2.1. *Existing System*

The existing system manages face recognition or individual identification. In the happened system, the machine was unable to identify anyone who was not wearing a mask.

2.2. *Drawbacks of Existing System:*

The CNN used in existing systems is slow. The existing system does not detect faces from all directions. The happened system does not detect multiple faces.

3. Proposed System:

The shortcomings of the current system have been addressed in the suggested system through evolution. The suggested system creates classification and predictive models that can be taken into consideration for precise classification grouping and prediction of Face masks on a person's face. Both datasets of people donning masks and those who aren't can be used to train this system. The system can regulate whether or not a person is wearing mask after training the model. It can also view the webcam and foretell the outcome Using MobileNetV2, the system will concentrate on boosting detection probability. This system can also recognize individuals who are not donning masks.

3.1. *System Architecture*

Data Augmentation, Data Visualization, Splitting the data, labeling the data, importing the face detection, and Detecting the face with and without a mask.

3.1.1. Data Visualization

Let's start by making a visual representation of the overall dataset image count across both categories. This shows that 500 images are classified as ”yes” categories and 500 images are classified as ”no” categories.

3.1.2. Data Augmentation :

The next stage involves expanding the dataset used for training to include more images.

All of the images in the collection are rotated and flipped in this data augmentation phase.

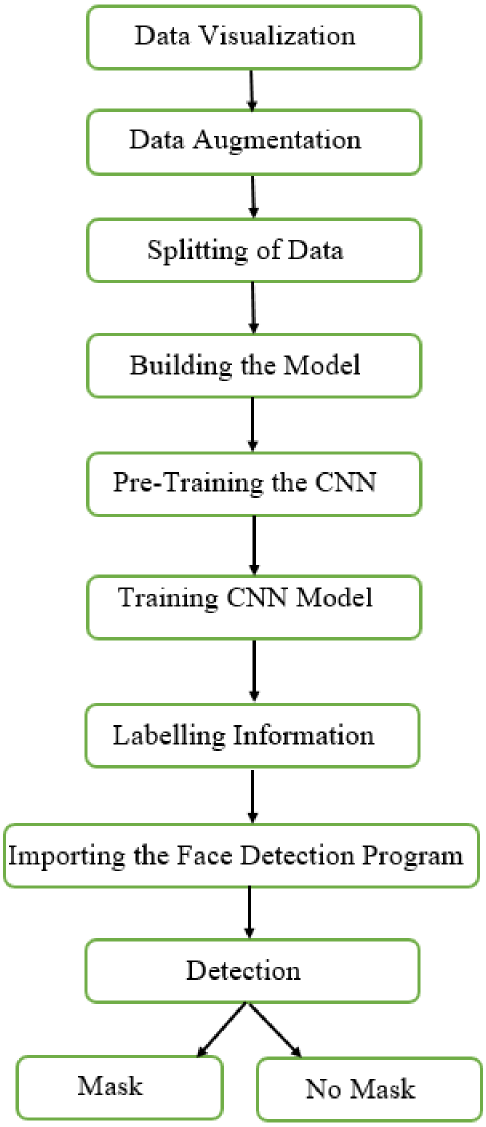


FIGURE 1. System Architecture

3.1.3. Splitting the data

This phase, the data is split into two sets in this phase: the training set, which includes the images needed to teach the CNN model, and the test set, which is used to evaluate the model. Construction of the Model After that, the sequential CNN model is constructed using multiple layers, including Conv2D, MaxPooling2D, Flatten, Dropout, and Dense.

3.1.4. Pre-Training the CNN model

Let's construct the model first, and then in the following step, create the ”train generator” and ”validation generator” to fit them into the model.

3.1.5. Training the CNN model

In this phase, the images from the sets of practice and test images are added to the sequential model, which is where the keras library is used to build the model. I trained the algorithm for 30 iterations. However, More iterations can be used in this project to improve the accuracy without running the danger of overfitting.

Evidence of the Truth After the model is created, two chances for these results are labeled. ”0” denotes ”without query,” and ”1” denotes ”with a mask.” The RGB numbers are also being used by me to modify the color of the boundary rectangle. The importation of the Face Recognition Software The test instance will then be checked to see if it is using a face mask through the computer's capturing device. For this, face detection must first be used. To recognize the face features in this, I'm using CNN.

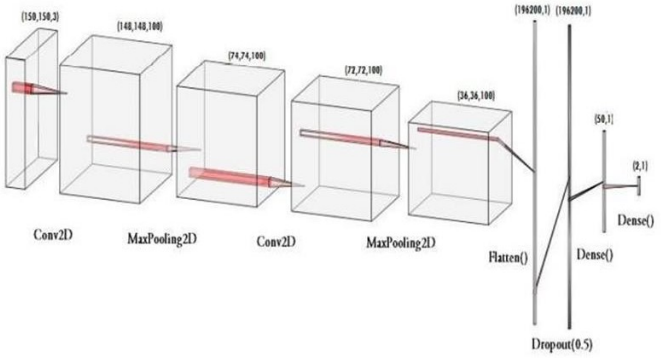


FIGURE 2. Building Model

4. Results and Discussion:

4.1. MobileNetv2:

A Convolutional Neural Network model called MobileNetV2 focus to function well on mobile devices. The bottleneck layers are linked by residual connections, and it is constructed on an inverted residual structure. Lightweight depth-wise convolutions are used in the intermediary expansion layer as a source of non-linearity to filter features. A 32-filter initial fully convolution layer and 19 additional bottleneck layers are both included in the MobileNetV2 architecture. a created model that processes images using MobileNetV2.

There are 53 levels in this CNN. 1000 categories can be used to categorize pictures. It employs depth-wise separable convolutions as the basic structural unit. It is a cutting-edge network that combines, object identification, classification, and semantic segmentation for mobile visual recognition. Launched as a component of the Tensor Flow-Slim Image is MobileNetV2.

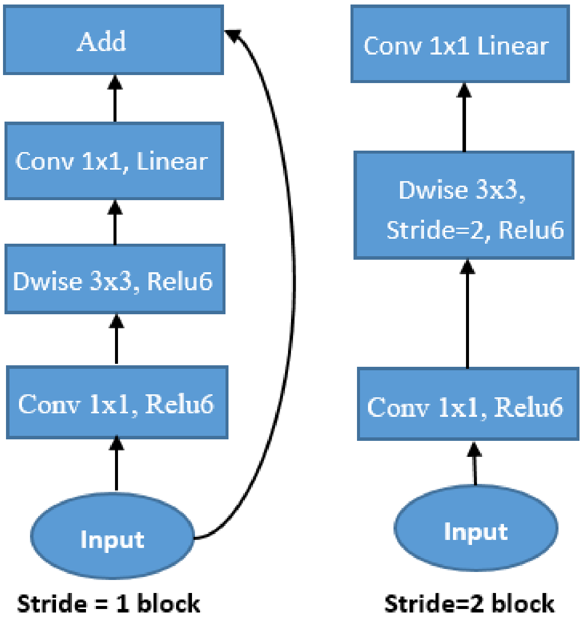


FIGURE 3. Comparision of MobilenetV1 and MobileNetV2

The fundamental idea behind MobilNetV2 is that while the inner layers contain the model's capacity to develop from low-level concepts to high-level concepts, the bottleneck layers encode the model's input and output. It is the conversion of pixels to picture categories in this instance. It utilizes two times fewer operations and requires fewer parameters than MobileNetV1, making it faster. It has better accuracy as a result of the same latency value.

4.2. Representation in Digitized form image:

Representation of digitized form image is shown in figure 6 and figure 7 stepwise.

4.3. Convolution Neural Network:

Yannlecun introduced a convolution neural network in 1988, which is a unique architecture of an artificial neural network. Image classification is

one of the most widely used applications for building. CNNs are widely used in natural language processing, recommender systems, and image and video identification.

The fundamental idea, though, is the same and can be used for any other use case! Similar to neural networks, CNNs are composed of neurons with trainable biases and weights. Each neuron takes in a number of inputs, weights them, and then directs the outcome through an initiation function to give an output. All the techniques that were created for neural networks still hold true for function. More precisely, the result is created following the image's passage through several convolutions, pooling, nonlinear, and fully connected layers.

CNNs because the entire network has a loss Patches from earlier levels are combined in every layer. Convolution Networks are multistage networks with stages that can be trained. The input and output of each phase are feature maps, which are collections of arrays. Each feature map at the output depicts a specific feature that was extracted from all input locations. A nonlinearity layer, a feature pooling layer, and a filter bank layer make up each step. A classification module follows one, two, or three of these steps with three layers each in a ConvNet. CNN's fundamental architecture, where S2 and S4 are pooled/sampled layers C1, and C3 are convolution layers. Filter: The filter bank's trainable filter (kernel) links the incoming feature map and the output feature map. Convolution layers operate on the input and then send the outcome to the following layer. The convolution simulates how a single neuron would react to everyday events.

4.4. Data Set Used:

One of the few available precautions for COVID-19 in the absence of immunization, masks are essential for safeguarding people's health against respiratory diseases. With this dataset, a model to detect individuals wearing masks, or not wearing them, can be created, this data collection includes 5000 images from the following two classes:

1. With mask;
2. Without mask

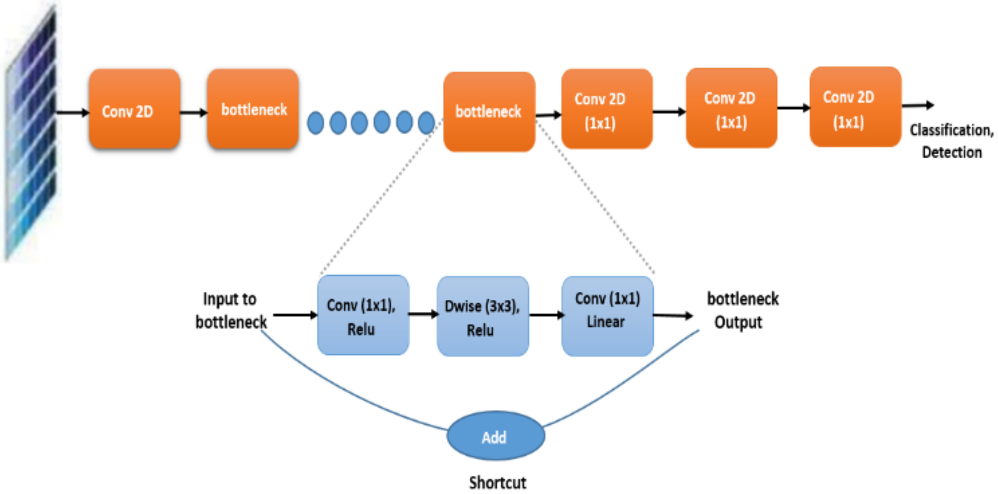


FIGURE 4. Architecture of MobileNet2

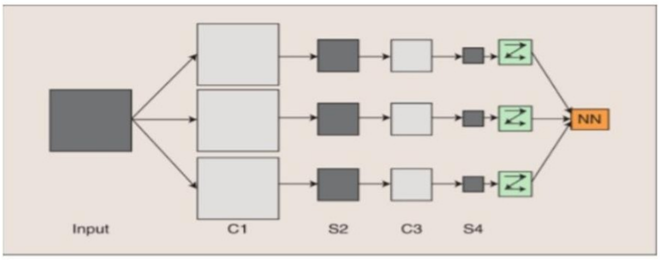
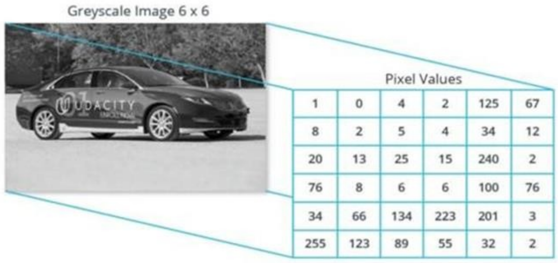


FIGURE 5. Convolution Neural Network



1	0	4	2	125	67
8	2	5	4	34	12
20	13	25	15	240	2
76	8	6	6	100	76
34	66	134	223	201	3
255	123	89	55	32	2

FIGURE 6. Representation in Digitized form image step 1

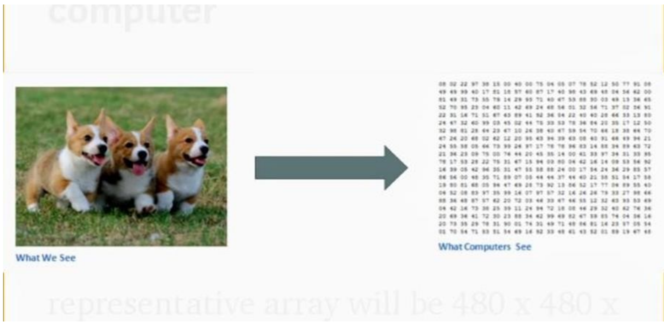


FIGURE 7. Representation in Digitized form image step 2

4.4.1. With Masks:

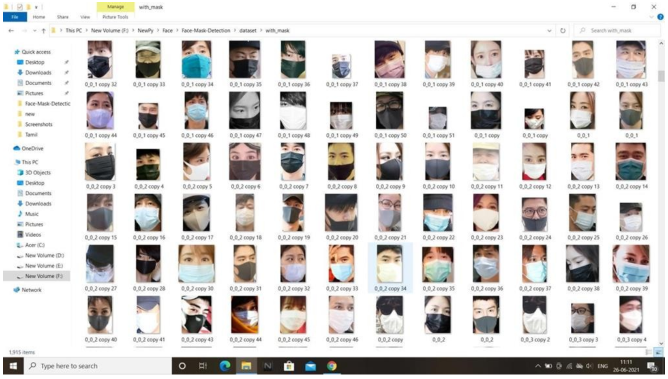

This PC › New Volume (F:) › Newfy › face › Face-Mask-Detection › dataset › with_mask
Type here to search

FIGURE 8. With Masks

4.4.2. Without Masks:

Images of typical human faces in various styles and angles are used to recognize human faces in input picture frames without the use of masks.

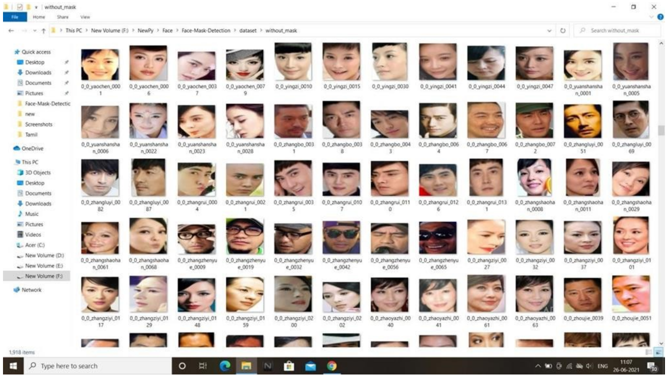

This PC › New Volume (F:) › Newfy › face › Face-Mask-Detection › dataset › without_mask
Type here to search

FIGURE 9. Output screen Without Masks

4.4.3. Output Screen:

The graph in figure 10 represents Training Loss and Accuracy. The output screen represents the face with mask is changed to face without mask as shown in figure 11 and figure 12.

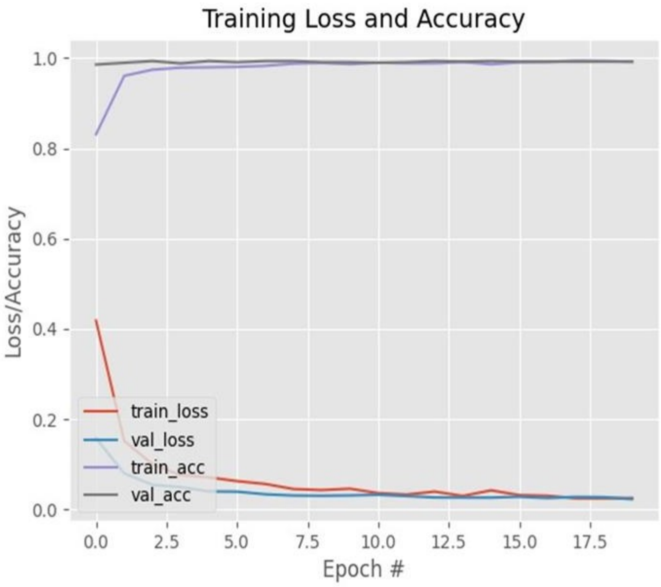


FIGURE 10. Training Loss and Accuracy

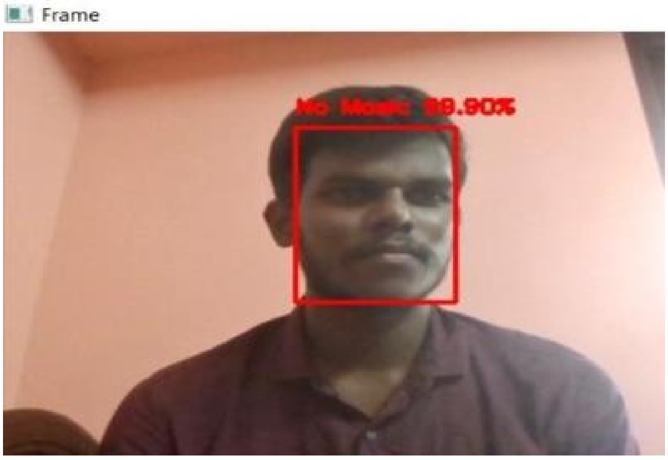


FIGURE 11. Output No Mask

4.4.4. Output:

Output is shown in figure 11 as the face is represented without mask.

4.4.5. Test Cases:

A test case is a group of procedures that should be ensured in order to evaluate a particular software feature. They are created for different situations so that testers can assess whether the program is operating correctly and delivering the desired results.

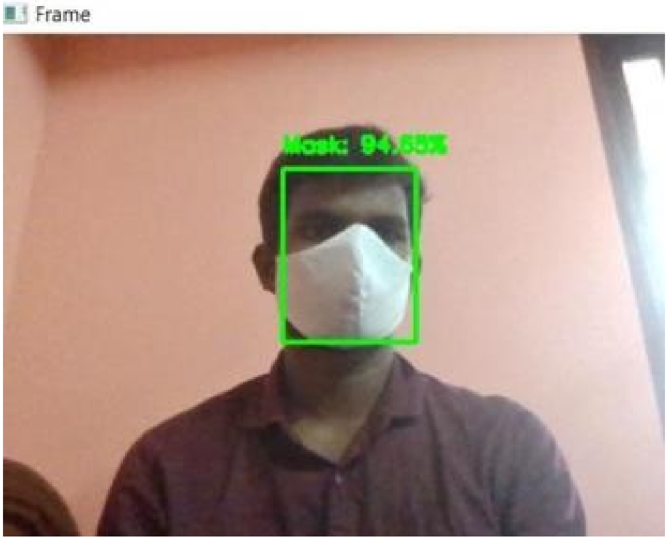


FIGURE 12. With Mask

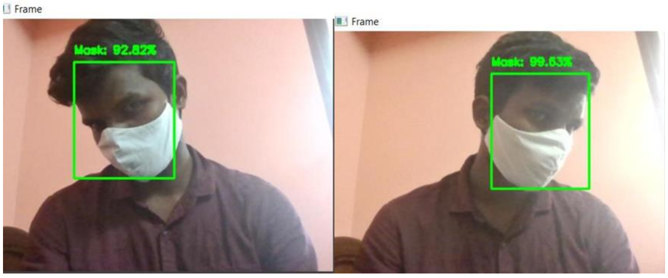


FIGURE 13. Inputs are given from different angles to test the model

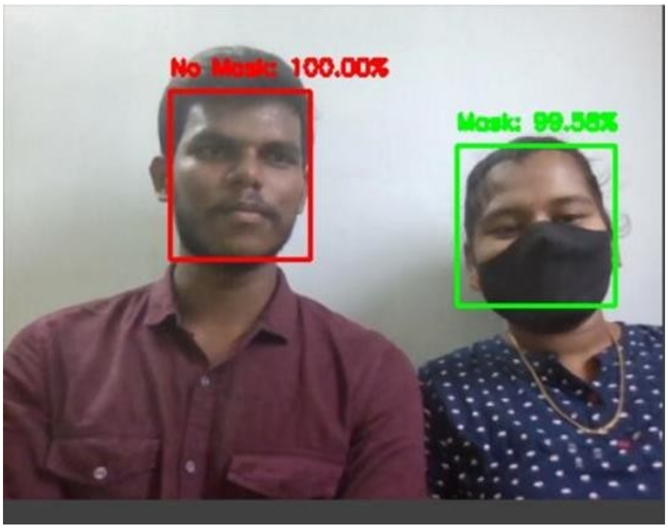


FIGURE 14. This is the test case where multiple human faces are detected by the mode

1. In this Test case, the inputs are given from different angles to test the model.
2. This is the test case where multiple human faces are detected by the mode.

5. Conclusion:

There must be action taken to slow the COVID-19 pandemic's spread. Using transfer learning techniques in neural networks, a facial mask detector was modeled. We used the dataset, which included 2000 masked face images and 2000 unmasked face images, to train, justify, and evaluate the blueprint. These images were pulled from a variety of sites, including Kaggle. From screenshots and real-time video feeds, the model was deduced. Metrics like recall, precision, and accuracy, as well as the MobileNetV2 architecture, were assessed in order to choose a basic model. A number of settings can make use of this face mask detector, containing airports, grocery shopping centers, and other busy areas, to keep an eye on the general populace and prevent the dissemination of disease by determining who is abiding by the rules and who is not. Recently, face shields became required to be worn in more than fifty nations worldwide. In public places such as offices, stores, supermarkets, and public transportation, people must conceal their faces. Software is frequently used by retail businesses to count the number of customers accessing their locations. They might also enjoy counting the number of times people view advertisements and digital displays. This software can be compared to any current USB, IP, and CCTV cameras to find individuals who aren't wearing masks. Web and desktop apps can use this live detection video feed to display notice messages for the operator. In the event that a mask is not being worn, software operators can still obtain a picture. Additionally, a sound alert can be installed to beep when somebody enters the room without a mask. Only individuals donning face masks are permitted entry thanks to a connection between this software and the entry gates.

References

Fu, Xiaomeng and Huiming Qu. "Research on Semantic Segmentation of High-resolution Remote Sensing Image Based on Full Convolutional Neural Network". *2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE)* (2018): 1–4. 10.1109/ISAPE.2018.8634106.

He, Kaiming, et al. "Deep Residual Learning for Image Recognition". *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016. 770–778.

Kishore, P V V, et al. "Indian Classical Dance Action Identification and Classification with Convolutional Neural Networks". *Advances in Multimedia* 2018 (2018): 1–10. 10 . 1155 / 2018 / 5141402.

Krishnaveni, G, B Lalitha Bhavani, and N V S K Vijaya Lakshmi. "RETRACTED: An enhanced approach for object detection using wavelet based neural network". *Journal of Physics: Conference Series* 1228.1 (2019): 012032–012032. 10.1088/1742-6596/1228/1/012032.

Kumar, Sanjay, et al. "A Deep Learning for Brain Tumor MRI Images Semantic Segmentation Using FCN". *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (2018): 1–4. 10.1109/CCAA.2018.8777675.

Malathi, J, et al. "Survey: Image forgery and its detection techniques". *Journal of Physics: Conference Series* 1228.1 (2019): 012036–012036. 10.1088/1742-6596/1228/1/012036.

P, Viola and M Jones. "Robust real-time face detection". *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001* (2001): 747–747. 10.1109/ICCV.2001.937709.

Pathak, Mrunal, Vinayak Bairagi, and N Srinivasu. "Multimodal Eye Biometric SystemBased on Contour Based E-CNN and Multi Algorithmic Feature Extraction Using SVBF Matching". *International Journal of Innovative Technology and Exploring Engineering* 9 (2019). 10.35940/ijitee.I7729.078919.

Priya, R Devi, et al. "MultiObjective Particle Swarm Optimization Based Preprocessing of Multi-Class Extremely Imbalanced Datasets". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 30.05 (2022): 735–755. 10.1142/S0218488522500209.

Ravi, Sunitha, et al. "Multi modal spatio temporal co-trained CNNs with single modal testing on RGB–D based sign language gesture recognition". *Journal of Computer Languages* 52 (2019): 88–102. 10.1016/j.cola.2019.04.002.

Satapathy, D, et al. “Deep learning based image recognition for vehicle number information”. *International Journal of Innovative Technology and Exploring Engineering* 8 (2019): 52–55.

Szegedy, Christian, et al. “Going deeper with convolutions”. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015. 1–9.

T, Ojala, M Pietikainen, and T Maenpaa. “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.7 (2002): 971–987. 10.1109/TPAMI.2002.1017623.

Viola, P and M Jones. “Rapid object detection using a boosted cascade of simple features”. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (2001). 10.1109/CVPR.2001.990517.

Embargo period: The article has no embargo period.

To cite this Article: Hemanth, S V, K Sanjeevaiah, Dharmendra Kumar Roy, and K Nagaraju. “Face Mask Detection using MobilenetV2 of Convolution Neural Network.” International Research Journal on Advanced Science Hub 05.05S May (2023): 72–80. http://dx.doi.org/10.47392/irjash.2023.S010