

Face Mask Detection System using Mobilenetv2



Mayank Arora, Sarthak Garg, Srivani A

Abstract: In this pandemic, it is getting more and more difficult to keep a track of people who are wearing masks regularly or not. It cannot solely depend on human efforts to take care of this task and therefore there is a need to develop software that can automatically detect whether any given person is wearing a mask or not. Face Detection has evolved as a really popular problem in image processing and computer vision. Many new algorithms are being devised using convolutional architectures to form the algorithm as accurately as possible. These convolutional architectures have made it possible to extract even the pixel details. Training is performed through Fully Convolutional Neural Networks to semantically segment out the faces present in that image. Feature detection and feature extraction techniques help us identify whether a person is wearing a mask or not. The face mask detector will use a dataset of morphed masked images. Therefore, the created model will be accurate and it will also be computationally efficient and easily deployable in embedded systems since the MobileNetV2 architecture will be incorporated (Raspberry Pi, Google Coral, etc.). This framework can also be used in real-time applications that, due to the outbreak of Covid-19, require face-mask detection for safety purposes. This project can be merged with embedded application systems at airports, train stations, workplaces, schools, and public places to ensure compliance with the guidelines for public safety. The above topic is very prominent in recent times as the identification process will not only help us classify individuals but also will reduce the workforce required to do the same exponentially.

Keywords: Face Mask Detection, Neural Networks, MobileNetV2, Image Data generator, OpenCV

I. INTRODUCTION

Due to the Covid-19 pandemic, which requires people to wear face masks, maintain social distance, and washes their hands with hand sanitizers, [1] Face Mask detection is the need of the hour. Although other social distancing and sanitization issues have been discussed in the past, face mask identification has yet to be properly addressed. Wearing a mask during the whole pandemic is a crucial preventive measure, and it is particularly important in times when maintaining social distance is difficult. A mask is required for

everyone who is at risk of severe illness from COVID-19 diseases, particularly those who are at higher risk. As a result, lowering the possibility of transmission of this deadly virus from an infected person to a healthy individual will greatly reduce the spread of the virus and the prevalence of the infection. Training models through CNN and by incorporating convolution operators yields a generic output. By the means of this project, we are trying to add a layer of efficiency that will generate better output. We will be incorporating MobileNetV2 for the same. MobileNetV2 has a faster rate of processing as compared to CNN. It uses few parameters as well, which simplifies our entire process altogether. Apart from novel MobileNetV2, we will also be using an image data generator to augment our dataset. IDG will ensure that optimal output is generated because of the presence of enough images in the training set. We will also be using "Relu" activation function because of the presence of non-linear test cases in our database, i.e. images. "Relu" activation function is best suited for images theoretically. In the output layer, we will be including dual layers to identify whether or not a user is wearing a mask.

ABOUT CORE TECHNOLOGIES

A. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a form of deep, feed-forward artificial neural network used to analyze visual imagery. These networks' architecture was loosely influenced by biological neurons that interact with one another and produce outputs based on inputs. Although work on CNNs began in the early 1980s, they have only recently gained popularity as a result of recent technological advances and computational capabilities that allow the processing of large quantities of data and the training of sophisticated algorithms in a reasonable amount of time.

B. MobileNetV2

MobileNetV2 is a major advancement over MobileNetV1 in terms of classification, object identification, and semantic segmentation for mobile visual recognition. MobileNetV2 is available as part of the TensorFlow-Slim Image Classification Library, or one can use Colaboratory to get started with it right away [2]. It can also download the notebook and use Jupyter to explore it locally. MobileNetV2 is also available as TF-Hub modules, with pre-trained checkpoints available on github. MobileNetV2 expands on the concepts of MobileNetV1 [1], using depthwise separable convolution as a cost-effective building block. However, V2 incorporates two new architectural features: 1) linear bottlenecks between layers, and 2) shortcut connections between bottlenecks.

C. Image Data Generator

The Image Data Generator class in Keras does not perform an "additive" process.

Manuscript received on April 05, 2021.
Revised Manuscript received on April 13, 2021.
Manuscript published on April 30, 2021.

* Correspondence Author

Mayank Arora*, UG Student, VIT – Vellore, Tamil Nadu, India.

mayank.arora2017@vitstudent.ac.in

Sarthak Garg, UG Student, VIT – Vellore, Tamil Nadu, India.

sarthak.garg2017@vitstudent.ac.in

Srivani A, Assistant Professor (Sr.), VIT – Vellore, School of Computer Science and Engineering (SCOPE), VIT, Vellore, Tamil Nadu, India. asrivani@vit.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

It doesn't take the original data, transforms it randomly, and then returns both the original and transformed data[3]. Instead, the ImageDataGenerator takes the original data, transforms it randomly, and then returns the transformed data. Data augmentation refers to a collection of techniques for creating "new" training samples from existing ones by introducing random jitters and perturbations (but at the same time ensuring that the class labels of the data are not changed). When we use data augmentation, we want to improve the model's generalizability.

D. Evaluation Metrics

Accuracy - The foremost intuitive indicator of success is accuracy, and it's merely a quantitative relation of properly foreseen observations to total observations. One can assume that our model is best if we've high exactitude. Therefore, to see the potency of your model, you have got to appear at alternative parameters. $Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$

Recall (Sensitivity) - Recall is that the quantitative relation of positive observations expected properly to all or any observations within the actual class-yes. $Recall = \frac{TP}{TP+FN}$

Precision - Precision is that the quantitative relation of the positive observations properly foreseen to the positive observations foreseen in total.

F1 Score - F1 Score is that the Precision and Recall weighted average. This score conjointly takes into thought each false positives and false negatives. it's not as simple to grasp intuitively as accuracy, however, F1 is usually a lot of helpful than accuracy, particularly if you have got associate degree uneven distribution of categories.

Accuracy functions well if there are unit equal prices for false positives and false negatives. If the expense of false positives and false negatives area unit somewhat completely different, each Precision and Recall area unit easier to appear at. $F1\ Score = \frac{2 * (Recall * Accuracy)}{(Recall + Accuracy)}$

II. LITERATURE SURVEY

This section provides an overview of the main human face recognition techniques that apply primarily to frontal faces, as well as the benefits and drawbacks of each approach. Eigenfaces (Eigen features), neural networks, dynamic relation architecture, hidden Markov model, geometrical feature matching, and prototype matching are among the methods considered.

The methods are evaluated based on the facial representations they employ [4]. The principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, are known as eigenfaces in mathematics. The eigenvectors are arranged to reflect varying amounts of difference among the faces. A linear combination of the eigenfaces will accurately represent each face. Only the "best" eigenvectors with the largest eigenvalues can be used to approximate it. The appeal of using neural networks may stem from the network's nonlinearity.

As a result, the feature extraction stage could be faster than the linear Karhunen-Loève methods. A single-layer adaptive network called WISARD, which includes a separate network for each stored person, was one of the first artificial neural network (ANN) techniques used for face recognition. For efficient identification, the method for constructing a neural network structure is critical. It is largely determined by the

intended application. Some of the other approaches use the pre-trained state-of-the-art deep learning model, InceptionV3. In this approach, the dataset used is Simulated Masked Face Dataset (SMFD).

Due to the limited availability of data, the model showed a high accuracy of 99% on the training data set but a fairly low accuracy on the test data set, hence it was overfitted, and hence not suitable [5]. It is also very important to know that how face recognition techniques have evolved and improved from the past. Various researchers and professionals have recently concentrated on gray-scale face images.

Although some were based purely on pattern recognition models, with no prior knowledge of the face model, others relied on AdaBoost, an excellent classifier for training. Real-time face detection became possible thanks to the Viola-Jones Detector, which offered a breakthrough in face detection technology [6]. It had several issues, including the orientation and brightness of the face, which made it difficult to intercept. So, in a nutshell, it didn't fit in dark or dim light. As a result, researchers began looking for a new alternative model capable of detecting faces as well as masks on the face. Many datasets for face recognition have been developed in the past to develop a perception of face mask detection models. WiderFace, IJB-A, MALF, and CelebA are examples of recent datasets that were generated by taking online photos. In these datasets, annotations are given for current faces, as opposed to earlier ones. Large datasets are far more needed for improved training and testing data, as well as for performing practical applications in a much more straightforward manner.

This necessitates a variety of deep learning algorithms that can read faces and masks directly from the user's data. Face Mask detection templates come in a variety of shapes and sizes. These can be classified into many groups. The Viola-Jones face detector, which was mentioned earlier in this section, was used to accept boosted cascades with easy haar features in boosting-based recognition. The Viola-Jones detector model was then used to build a Multiview face mask detector. In addition, decision tree algorithms were used to build a face mask detector model.

This category's face mask detectors were good at detecting masks [7]. Another study employs deep learning techniques to discern facial recognition and determine whether or not an individual is wearing a facemask. The dataset collected includes 25,000 images with a resolution of 224x224 pixels, and the qualified model performed with a 96 percent accuracy rate [8]. However, the test accuracy was also lower in this case. Although it was put into practice in real life, it could be improved.

Many datasets for face detection have been created in the past to shape an understanding of face mask detection models. Earlier datasets were made up of images taken in supervised environments, while more recent datasets are made up of images taken online [9]. The MobileNetV2 architecture classifier was an upgrade over the MobileNetV1 architecture classifier since

it used a three-layer convolutional layer as the initial layer, followed by 13 times the previous building blocks. MobileNetV2 architecture, on the other hand, consists of 17 3X3 convolutional layers in a row, as well as a 1×1 convolution, an average layer of max pooling, and a classification layer [10].

III. EXPERIMENTAL RESULTS

Before conducting the experiment, an image data generator was used to augment the dataset. IDG ensures that multiple variations of any given image are taken into account while training the model. Post that, the entire input was passed through MobileNetV2 architecture which makes the entire model unique and novel. After that, the regular CNN process continued, ranging from feature enhancement, flattening, pooling, and finally displaying whether the person standing in front of the camera is wearing a mask or not. The above module was able to perform operations and run diagnostics at a faster rate as compared to CNN. The model used less number of parameters without compromising the accuracy of the model. "Relu" activation function was used in the hidden layers of CNN because of the presence of non-linear use cases (images). Soft-max activation was used in the output layer to predict whether or not a person is wearing the mask with the expected probability.

	Precision	Recall	F1-score	Support
With_mask	1.00	0.83	0.90	383
Without_mask	0.85	1.00	0.92	384

The below-mentioned table depicts the statistics post-experiment.

The above model showcased an accuracy of 96 percent. Precision, Recall, f1-score, support for with_mask attribute are 1.00, 0.83, 0.90, 383 respectively. The values obtained for without_mask attribute for the above-mentioned parameters are 0.85, 1.00, 0.92, and 384 respectively.

The following images showcase the working of our proposed model. The software captures the faces via a live-stream and tells the user whether a person is wearing a mask or not.

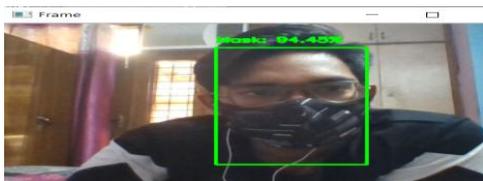


Figure 1: With Mask Demo

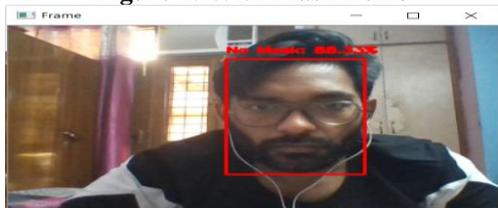


Figure 2: Without Mask Demo

IV. CONCLUSION

Both the training and implementation of the image dataset in the proposed face mask detection model will be split into two categories: people with masks and people without masks. The OpenCV deep neural network technique should produce positive results. The proposed method is special in that it uses

the MobilenetV2 image classifier to classify images, resulting in more accurate performance. Many previous studies had issues with their findings, although others were able to improve accuracy with their dataset. Since the dataset was gathered from a variety of other sources and the images used in the dataset were manually cleaned to improve the accuracy of the results, the issue of multiple incorrect predictions would be successfully eliminated from the model. Real-world implementations will be a much more difficult problem in the future. The new proposed model should, ideally, assist the concerned authorities in dealing with this massive pandemic that has spread across most of the globe.

REFERENCES

1. Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., & Hemanth, J. (2020). SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustainable cities and society, 102692.
2. Chen, Y., Hu, M., Hua, C., Zhai, G., Zhang, J., Li, Q., & Yang, S. X. (2020). Face Mask Assistant: Detection of Face Mask Service Stage Based on Mobile Phone. arXiv preprint arXiv:2010.06421.
3. Jiang, M., & Fan, X. (2020). RetinaMask: a face mask detector. arXiv preprint arXiv:2005.03950.
4. Militante, S. V., & Dionisio, N. V. (2020, August). Real-Time Facemask Recognition with Alarm System using Deep Learning. In 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC) (pp. 106-110). IEEE.
5. Chowdary, G. J., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020, December). Face mask detection using transfer learning of inceptionv3. In International Conference on Big Data Analytics (pp. 81-90). Springer, Cham.
6. Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2020). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement, 167, 108288.
7. Guo, J. M., Lin, C. C., Wu, M. F., Chang, C. H., & Lee, H. (2011). Complexity reduced face detection using probability-based face mask prefiltering and pixel-based hierarchical-feature adaboosting. IEEE Signal Processing Letters, 18(8), 447-450.
8. Sachdeva, R. (2020). Face Mask Detection System. Available at SSRN 3755508.
9. Pooja, S., & Preeti, S. (2021). Face Mask Detection Using AI. In Predictive and Preventive Measures for Covid-19 Pandemic (pp. 293-305). Springer, Singapore.
10. Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2020). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable Cities and Society, 102600.

AUTHORS PROFILE



Mayank Arora, UG Student, School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore. Core Interests: Machine Learning, Deep Learning, Natural Language Processing, Neural Networks Mail ID- mayank.arora2017@vitstudent.ac.in.



Sarthak Garg, UG Student, School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore. Core Interests: Machine Learning, Augmented and Virtual Reality, Automation Mail ID- sarthak.garg2017@vitstudent.ac.in.



Prof. Srivani A, Assistant Professor (Sr.) School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore. Core Interest: Image Processing, Machine Learning, IoT Mail ID - asrivani@vit.ac.in