Turkish Journal of Computer and Mathematics Education *Vol.12 No.6 (2021), 2841-2848*
Research Article

A Comparative Analysis of Variant Deep Learning Models for COVID-19 Protective Face Mask Detection

RohanKatari[a], SreekarKaza[a], B RamyaSree[b], V Divyavani[b], Mohammed AbubakarJ[a]

[a]UG Student, Department of CSE, Institute of Aeronautical Engineering, JNTU(H), Hyderabad, India
[b]Assistant Professor, Department of CSE, Institute of Aeronautical Engineering, JNTU(H), Hyderabad, India

Abstract: The world is in the midst of a paramount pandemic owing to the rapid dissemination of coronavirus disease (COVID-19) brought about by the spread of the virus 'SARS-CoV-2'. It is mainly transmitted among persons through airborne diffusion of droplets containing the virus produced by an infected person sneezing or coughing without covering their face. The World Health Organization (WHO) has issued numerous guidelines which state that the spread of this disease can be limited by people shielding their faces with protective face masks when in public or in crowded areas. As a precautionary measure, many nations have implemented obligations for face mask usage in public spaces. But manual monitoring of huge crowds in public spaces for face masks is laborious. Hence, this requires the development of an automated face mask detection system using deep learning models and related technologies. The detection system should be viable and deployable in real-time, predicting the result accurately so as to be used by monitoring bodies to ensure that the face mask guidelines are followed by the public thereby preventing the disease transmission. In this paper we aim to perform a comparative analysis of various sophisticated image classifiers based on deep learning, in terms of vital metrics of performance to identify the effective deep learning based model for face mask detection.

Keywords: Face Mask Detection, COVID-19, Deep Learning, Transfer Learning, MobileNet, ResNet50, Inception, VGG16, Xception

1. Introduction

The outbreak of COVID-19 has created a catastrophic situation around the world. According to the latest COVID-19 Epidemiological Update report [1] published by the World Health Organization (WHO) over 146 million people got infected and over 3 million have died with the coronavirus disease 2019 (COVID-19).Researchers of various fields have been working to develop intelligent systems using digital technologies which aid in monitoring and control of the spread of the disease. S. Tuli, et al. [2] have devised a system for predicting development of the outbreak using machine learning and cloud computing so as to develop policies and strategies for managing its propagation and to efficiently monitor disease. The system presented in [3] would use an Internet of Things (IoT) framework to gather users' data on the disease symptoms in real-time to detect suspected COVID-19 cases in the earlier stages, to track the treatment response of patients recovered and to comprehend virus behaviour by collecting and analysing relevant data. An Artificial Intelligence (AI) based tool was used by Y. Ke, et al., in [4], to identify the potential of marketed medicines for treating the COVID-19 disease.

Following the COVID-19 outbreak, the WHO has released a number of precautionary recommendations to combat coronavirus transmission. Some of the most significant guidelines are practicing social/physical distancing, sanitization of hands, wearing face masks, avoiding crowds, etc. Even though few global pharmaceutical companies have successfully developed COVID-19 vaccines which were approved by the WHO and medical authorities, more and more COVID-19 cases are emerging everyday worldwide. This is because of the disease transmission caused among people by not following the protective measures and guidelines mentioned earlier. The use of face masks plays a significant role to help reduce the person to person transmission of the virus [5] and this helps in breaking the chain. Governments of the majority of countries have mandated the use of face masks in public places to control community spread of the virus. However, manually inspecting huge crowds in public spaces for face masks is an arduous process. Thus, the implementation of an automatic face mask detection system is needed to aid this process.

The idea of a face mask detection system is to identify whether or not a person is wearing a face mask using data in the form of images, recorded videos or a live stream video. An efficient system is developed by selection of a Convolutional Neural Network (CNN) which is optimal in terms of computational intensity without compromising the performance of the system. In this paper we perform face mask detection using various deep learning networks to compare performance and analyse their efficiency. This paper is further structured as follows: section II shows the related work and literature survey, section III discusses the examined methodology, section IV presents the results and section V gives the conclusion.

2. Related Work

2.1. Problem Definition

Face mask detection is the process of determining whether or not a person is wearing a mask and where their face is located. It is a combination of general object recognition, which is used to identify different categories of objects (the mask in this case), and face detection, which is used to identify the face, in a digital image. The aforementioned systems have an assortment of real world applications such as surveillance, autonomous vehicles, robot vision and activity, etc.

Object recognition is primarily concerned with locating and classifying objects in images. Traditional algorithms (non-neural approaches) such as Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG) can be used for such tasks, but they depend largely on feature engineering. Neural networks (in deep learning) can outshine the traditional algorithms without the need of handcrafted features. There are two groups of such object detection algorithms [6]:

- Two-stage detectors: In stage one, a Region Proposal Network (RPN) will identify regions of interest. In stage two, a classifier processes the region candidates identified. Examples are Faster R-CNN (Region-based Convolutional Neural Networks), Mask R-CNN.

- One-stage detectors: They directly run detection over the possible locations (a dense sample) and exclude the RPN stage. Examples are You Only Look Once (YOLO), Single Shot MultiBox Detector (SSD), etc.

2.2. Literature Survey

The following is the literature survey of various research works related to the development of face mask detection systems.

The proposed system by A. Das, et al. [7] entails a cascaded classifier and a CNN (pre-trained) containing dense neuron layers connected with two 2D convolution layers; it gave an accuracy of 94.58% on a dataset obtained from Kaggle [8]. A. Negi, et al. [9] built a custom image based CNN model with Haar cascade classifier for face mask detection and used Keras-Surgeon for model pruning, to reduce the model size so that it can be implemented on embedded systems, and attained 98.9% validation accuracy. In [10] M. R. Bhuiyan, et al. have used the YOLOv3 network, a one-stage object detector, for detection of face masks and obtained a mean average precision (mAP) of 0.96. G. Draughon, et al. [11] have presented a framework to detect and track people and face mask usage by them, using deep learning and Computer Vision. The face mask detector employed in the framework was a CNN-based binary classifier with residual network architecture; it gave a classification accuracy of 96%. M. Jiang, et al. in [12] have proposed the use of the one-stage detector, RetinaFaceMask. Higher-order semantic information is fused with several feature maps using the feature pyramid network comprised in the detector. RetinaFaceMask with ResNet was used to detect face masks with a precision of 93.4%. In [13] J. Zhang, et al. have developed a sophisticated framework named Context-Attention R-CNN which gave an mAP of 84.1% on a new practical dataset they created covering various conditions to achieve detection of fine-grained wearing condition of face masks. B. Wang, et al. in [14] devised a hybrid (deep) transfer learning and broad learning system for detecting face masks. It contains two stages: pre-detection (implemented using Faster-RCNN) followed by verification, the system gave a 94.84% precision score.

3. Methodology

As stated earlier, the aim of this paper is to compare various neural network architectures and evaluate their performance for face mask detection. A Convolutional Neural Network (CNN) convolves the input images or feature maps with convolution kernels in order to extract higher-level features. Thus it is an effective tool for Computer Vision tasks like classification of images, detection of objects, identifying patterns, etc. The neural network architectures studied, evaluated and compared in this paper are VGG16, InceptionV3, ResNet50V2, MobileNetV2 and Xception. Image classification is better achieved by the transfer learning of these models.

3.1. Understanding Transfer Learning

One of the most widely used methods for computer vision activities such as classifying and segmentation is transfer learning. In this method weights or information gained from solving a problem are shared to solve other problems that are similar to it. When the application areas are closely affiliated, transfer learning is beneficial to decrease training time. Transfer learning can be accomplished in one of the two following ways.

- The first is by employing a pre-trained model. It means it is a model that has been trained beforehand, on an extensive standard dataset such as ImageNet, MNIST, CIFAR, etc. ResNet, DenseNet, MobileNet, etc. are examples of this kind.
- The second way to achieve transfer learning is by using a custom output layer to perform classification which uses the features extracted by the pre-trained model (without its output layer).

3.2. Neural Networks Evaluated

The following neural network architectures (pre-trained models) have been selected for evaluation. These models are pre-trained on the standard ImageNet dataset and can perform a 1000 class classification on any colour image given as input.

3.2.1. VGG16

The VGG-16, developed by the Visual Graphics Group (VGG) at the University of Oxford [15], is a popular pre-trained model for classification of images. Here '16' in VGG16 denotes the number of weighted layers in the network. VGG16 comprises of 13 convolutional, 3 dense and 5 pooling layers. VGG uses smaller filters with more depth because of fewer parameters and stacks more of them rather than using larger filters.

3.2.2. InceptionV3

Inception orGoogleNet [16] developed by Google was the winner ofILSVR competition in 2014. This network uses inception modules. In a naïve version, in each inception module,convolution on an input is performed with filters of three sizes p x p (where p=1,3,5) along with max pooling and the outputs are joined in sequence and sent to the following inception module. The architecture contains 9 such modules and it is 22 layers deep. InceptionV3, which is 48 layers deep, is the improved version of InceptionV2. InceptionV3 additionally used greater factorization, RMSProp Optimizer and normalization of batches.

3.2.3. ResNet50V2

It is a Residual Neural Network with a depth of 50 layers [17]. It uses residual learning concept to overcome accuracy saturation caused by growing depth of the network. Here, rather than learning features, the network seeks to learn a residual. Residual is the difference obtained by subtraction of a feature learned from a given layer input. For this, shortcut connections are used i.e. directly connecting input of layer x to another layer (x+n). Contrary to stacking of layers like in VGG16 or Inception networks, ResNet modifies underlying mapping of the layers.

3.2.4. MobileNetV2

MobileNetV2 expands on MobileNetV1's (developed by Google in 2017) concepts [18,19] by using depthwise separable convolution as effective building blocks and adds two additional architectural features- linear bottlenecks between layers, and shortcut links between bottlenecks. Depthwise separable convolution is a combination of channel-wise (p×p, p=number of channels) spatial convolution called depthwise convolution followed by a convolution of size 1x1, called pointwise convolution. This architecture has 53 layers.

3.2.5. Xception

Xception (Extreme version of Inception) [20] is an Inception architecture extension in which depthwise separable convolutions replace the standard Inception modules of Inception network. Xception architecture has 36 convolutional layers establishing the its feature extraction foundation. It uses refashioned depthwise separable convolution i.e, the pointwise convolution (1x1) is performed first and is followed by channel-wise spatial convolution (nxn).

3.3. Proposed Methodology

The diagrammatic representation of the proposed model is depicted in Figure 1 and steps for the process flow are as follows:

Step 1: The dataset (samples shown in Figure 2) taken is a publicly available dataset [21]. It consists of 3850 real-time images in which 1920 images contain a face mask and 1930 images are without a face mask. The dataset is split in the ratio 80:20 for training and testing.

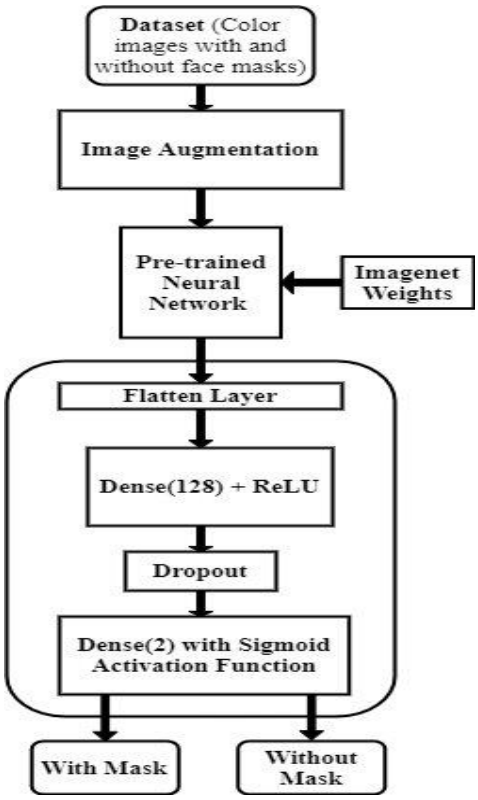

Figure 1. Proposed Methodology

Figure 2. Sample images from dataset (with mask on the left and without mask on the right)

Step 2: Image augmentation of the dataset is performed. It is a method to enlarge the training dataset by altering images in the dataset artificially using the operations rotation, zoom, horizontal flipping, shearing, height and width shift, etc.

Step 3: Images of size (224, 224, 3) (obtained after step 1) are given as input to the transfer learning model pre-trained using ImageNet weights.

Step 4: The actual output layer of the pre-trained model is substituted with the subsequent set of layers- a flattening layer (to convert the data from pre-trained model into a 1-dimensional array for inputting it to the next layer), continued by a dense layer (with 128 neurons) with Rectified Linear Unit (ReLU) activation function and a 0.5 dropoutrate (dropout layer helps prevent the model from overfitting). The feature map from step 3 passes through these layers.

Step 5: The output from step 4 is sent to a dense layer with sigmoid activation function and two neurons. This layer classifies whether or not the person in the image is wearing a mask.

4. Results and Analysis

This experiment of evaluation of deep learning models for face mask detection is implemented on Google Colaboratory (Colab Notebook) that runs on the cloud. The proposed methodology was implemented using Python and TensorFlow, the model training and tests are performed on a TESLA K80 GPU by NVIDIA.

The deep learning models used in this experiment are VGG16, InceptionV3, ResNet50V2, MobileNetV2 and Xception and the nature of classification results of the testing data is divided into the following four categories used for performance calculation:

1) *True Positive (TP):* Predicted that face mask is present and the original result is the same.
2) *False Positive(FP):* Predicted that mask is present but original result is no mask present.
3) *True Negative(TN):* Predicted that face mask is not present and the original result is the same.
4) *False Negative(FN):* Predicted that mask is not present but original result is mask is present.

4.1. Confusion Matrix

It is a square matrix of size 'p', used to evaluate the efficiency of a classification model (where p = number of target categories). It compares the factual goal values with the model's predictions. Here p=2 as target groups are with and without masks. Table 1 shows the confusion matrices obtained for each model evaluated.

Table 1. Confusion Matrices of Various Models

Model	Confusion Matrix
VGG16	[[360 24] [35 351]]
InceptionV3	[[382 2] [6 380]]
ResNet50V2	[[379 5] [7 379]]
MobileNetV2	[[382 2] [17 369]]
Xception	[[381 3] [6 380]]

4.2.Performance Report

There are several measures that can be used to demonstrate the performance of a classification model. The four central metrics considered here are precision, recall, f1-score and accuracy.

1) *Precision:* It is the number of positives that are true to all positives (false and true).
2) *Recall:* It is the number of positives that are true to all positive class instances, from a dataset.
3) *F1-Score:* It is the harmonic mean of recall and precision.
4) *Accuracy:* It is the ratio of rightly predicted samples to all of the samples in the dataset.

The classification report i.e. the precision, recall, f1-score and accuracy values are shown in Table 2. In the table, 0 signifies without mask and 1 signifies with mask.

Table 2. Performance Report of Various Models

Model	Classification Performance						
	Precision		Recall		F1-Score		Accuracy
	0	1	0	1	0	1	
VGG16	93.6	91.1	90.9	93.7	92.2	92.4	92.3
InceptionV3	99.4	98.4	98.4	99.4	98.9	98.9	98.9
ResNet50V2	98.6	98.1	98.1	98.6	98.4	98.4	98.4

MobileNetV2	99.4	95.7	95.5	99.4	97.4	97.5	97.5
Xception	99.2	98.4	98.4	99.2	98.8	98.8	98.8

4.3.Loss and Accuracy Graphs

Loss signifies how well or bad a model performs after each optimization iteration. The relation between accuracy and loss incurred during training is used to understand the nature and size of errors that a model has made. Figure 3 shows the training loss, validation loss, training accuracy and validation accuracy measured for various models.

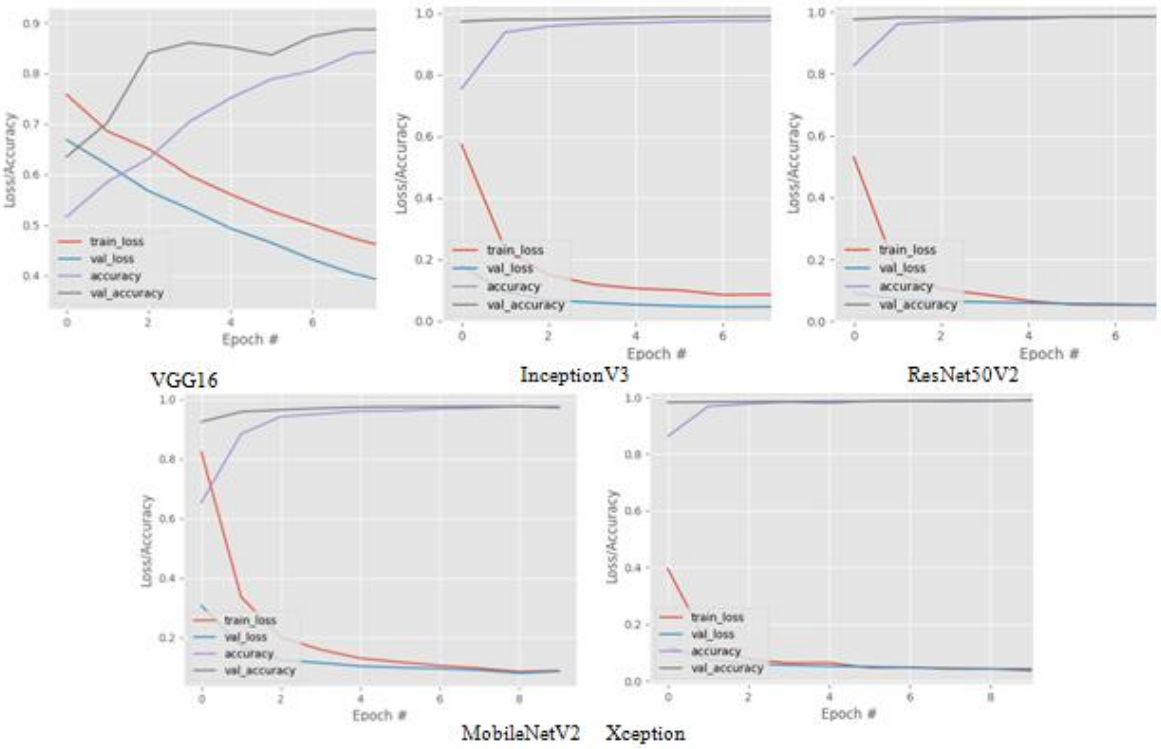

Figure 3. Loss and Accuracy Graphs

5. Conclusion

The epidemic caused by the COVID-19 disease has caused an upsurge of infected cases worldwide. The governments of several nations around the world have enforced obligatory use of protective face masks to prevent the transmission of the virus and prevent infections. But manual monitoring of public for use of face masks is an arduous process. Hence this propelled researchers for the development of automated face mask detection systems that can effectively identify whether or not people are using face masks. In this paper, we have implemented the pre-trained models VGG16, InceptionV3, ResNet50V2, MobileNetV2 and Xception for face mask detection and evaluated their performance using various metrics. The results show that InceptionV3 and Xception models have produced outstanding accuracies on the given dataset. Our future work will concentrate on detection of face masks using hybrid models.

References

1. W. H. Organization et al., (2021, April). “COVID-19 Weekly Epidemiological Update” report, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

2. S. Tuli, S. Tuli, R. Tuli, S. Singh Gill, (2020). "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing", Internet of Things, Volume 11,100222,ISSN 2542-6605, https://doi.org/10.1016/j.iot.2020.100222.
3. M. Otoom, N. Otoum, M. A. Alzubaidi, Y. Etoom, R. Banihani, (2020). "An IoT-based framework for early identification and monitoring of COVID-19 cases", Biomedical Signal Processing and Control, Volume 62, 102149, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2020.102149.
4. Yi-Yu Ke, Tzu-Ting Peng, et al., (2020). "Artificial intelligence approach fighting COVID-19 with repurposing drugs", Biomedical Journal, Volume 43, Issue 4, 20, Pages 355-362, ISSN 2319-4170.
5. Schünemann, H. J., Akl, E. A., Chou, R., Chu, D. K., Loeb, M., Lotfi, T., Mustafa, R. A., Neumann, I., Saxinger, L., Sultan, S., & Mertz, D., (2020). "Use of facemasks during the COVID-19 pandemic", The LANCET Respiratory Medicine, SPOTLIGHT, Volume 9, Issue 10, p954-955.
6. P. Soviany and R. T. Ionescu, "Optimizing the Trade-Off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction," 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 2018, pp. 209-214.

A. Das, M. Wasif Ansari and R. Basak, (2020) "Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV," 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342585.

7. Chittampalli, Sai Prakash and J Sirisha Devi. (2019) "Nexus DNN for Speech and Speaker Recognition." International Journal of Engineering and Advanced Technology (IJEAT), Vol. 9, no. 2, pp. 2004-2007, www.ijeat.org/wp-content/uploads/papers/v9i2/B2963129219.pdf..

    A. Negi, P. Chauhan, K. Kumar and R. S. Rajput, (2020) "Face Mask Detection Classifier and Model Pruning with Keras-Surgeon," 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, pp. 1-6, doi: 10.1109/ICRAIE51050.2020.9358337.

8. M. R. Bhuiyan, S. A. Khushbu and M. S. Islam, (2020). "A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, pp. 1-5.
9. G. T. S. Draughon, P. Sun and J. P. Lynch, (2020). "Implementation of a Computer Vision Framework for Tracking and Visualizing Face Mask Usage in Urban Environments," 2020 IEEE International Smart Cities Conference (ISC2), Piscataway, NJ, USA, pp. 1-8, doi: 10.1109/ISC251055.2020.9239012.
10. Mingjie Jiang, Xinqi Fan, Hong Yan, (2020). "RetinaMask: A Face Mask detector", arXiv:2005.03950v2 [cs.CV].
11. J. Zhang, F. Han, Y. Chun and W. Chen, (2021). "A Novel Detection Framework About Conditions of Wearing Face Mask for Helping Control the Spread of COVID-19," in IEEE Access, vol. 9, pp. 42975-42984, doi: 10.1109/ACCESS.2021.3066538.

B. Wang, Y. Zhao and C. L. Philip Chen, (2021). "Hybrid Transfer Learning and Broad Learning System for Wearing Mask Detection In the COVID-19 Era," in IEEE Transactions on Instrumentation and Measurement, doi: 10.1109/TIM.2021.3069844.

12. Simonyan, Karen & Zisserman, Andrew, (2014) "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv:1409.1556 [cs.CV].

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, (2016). "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

13. K. He, X. Zhang, S. Ren and J. Sun, (2016). "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778, doi: 10.1109/CVPR.2016.90.

14. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, (2017), "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", arXiv:1704.04861 [cs.CV].
15. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, (2018). "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.
16. F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions, (2017). " 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.
17. Face-Mask-Detection., "Face mask detection," (2020), accessed 03 Apr 2021. [Online]. Available: https://github.com/chandrikadeb7/Face-Mask-Detection.