Noname manuscript No.
(will be inserted by the editor)

Real-time face mask identification using Facemasknet deep learning network

Madhura Inamdar · Ninad Mehendale *

Received: date / Accepted: date

Abstract The COVID - 19 pandemic is devastating mankind irrespective of caste, creed, gender, and religion. Until a vaccine is discovered, we should do our bit to constrain the expanse of the coronavirus. Using a face mask can undoubtedly help in managing the spread of the virus. COVID - 19 face mask detector uses or owns Facemasknet, deep learning techniques to successfully test whether a person is with wearing a face mask or not. The manuscript presents three-class classification namely person is wearing a mask, or improperly worn masks or no mask detected. Using our deep learning method called Facemasknet, we got an accuracy of 98.6 %. The Facemasknet can work with still images and also works with a live video stream. Cases in which the mask is improperly worn are when the nose and mouth are partially covered. Our face mask identifier is least complex in structure and gives quick results and hence can be used in CCTV footages to detect whether a person is wearing a mask perfectly so that he does not pose any danger to others. Mass screening is possible and hence can be used in crowded places like railway stations, bus stops, markets, streets, mall entrances, schools, colleges, etc. By monitoring the placement of the face mask on the face, we can make sure that an individual wears it the right way and helps to curb the scope of the virus.

Keywords face mask · identification · deep learning

1 Introduction

It is believed that novel coronavirus has been originated from bats in Wuhan, China on the 17 th of November 2019 and spread from one country to another in no time. Deadly coronaviruses have previously induced respiratory infections specifically Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). The prevalent symptoms of COVID - 19 are fever, tiredness, dry cough, anosmia, sore throat, headache, etc. Its arrival has stopped the world due to its severity and adverse effects on humans. For a person having mild symptoms, it takes a fortnight for getting recovered. The recovery period for patients having critical symptoms depends on the severity. It is advisable for a person to stay quarantined or to be in self-isolation if affected by coronavirus. Reverse transcription-polymerase chain reaction (RTPCR) is a standard method that is currently being implemented to detect the presence of the virus in an individual's body.

Putting on a face mask can restrict the spread of the virus. In many cases, coronavirus can be asymptomatic too. As it is rightly said, prevention is better than cure, one should wear a face mask while coming in contact with people. By doing this, an individual ensures his safety, another person's safety, and in this way helps to curb the spread of the disease. The World Health Organization (WHO) as well as the Centers for Disease Control and Prevention (CDC) has suggested the use of face masks for decreasing the spread of the virus. Various

* Corresponding author
N. Mehendale
B-412, K. J. Somaiya College of Engineering, Mumbai, India
Tel.: +91-9820805405
E-mail: ninad@somaiya.edu

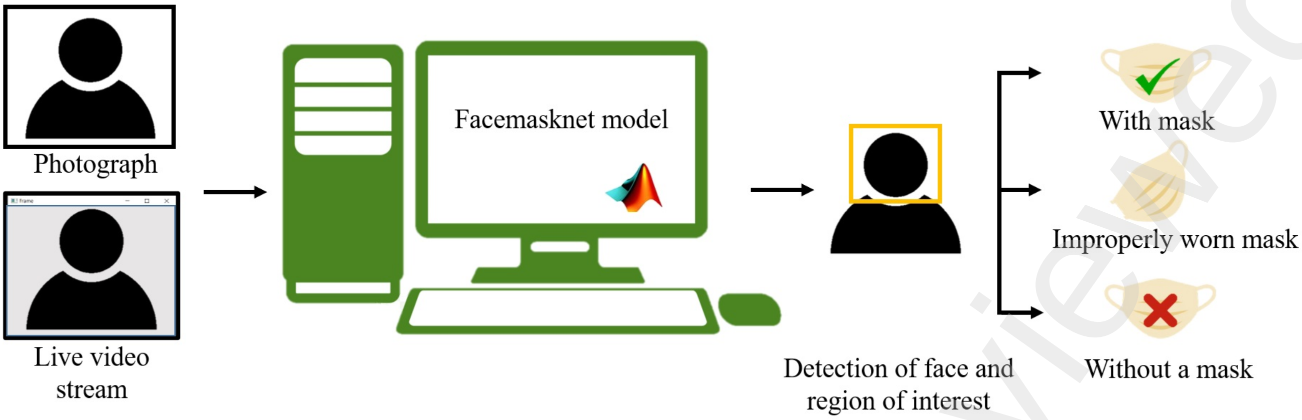

Fig. 1 Image or live video stream is given as an input to the Facemasknet model. Classification of an image takes place as with a mask, improperly worn mask, without a mask.

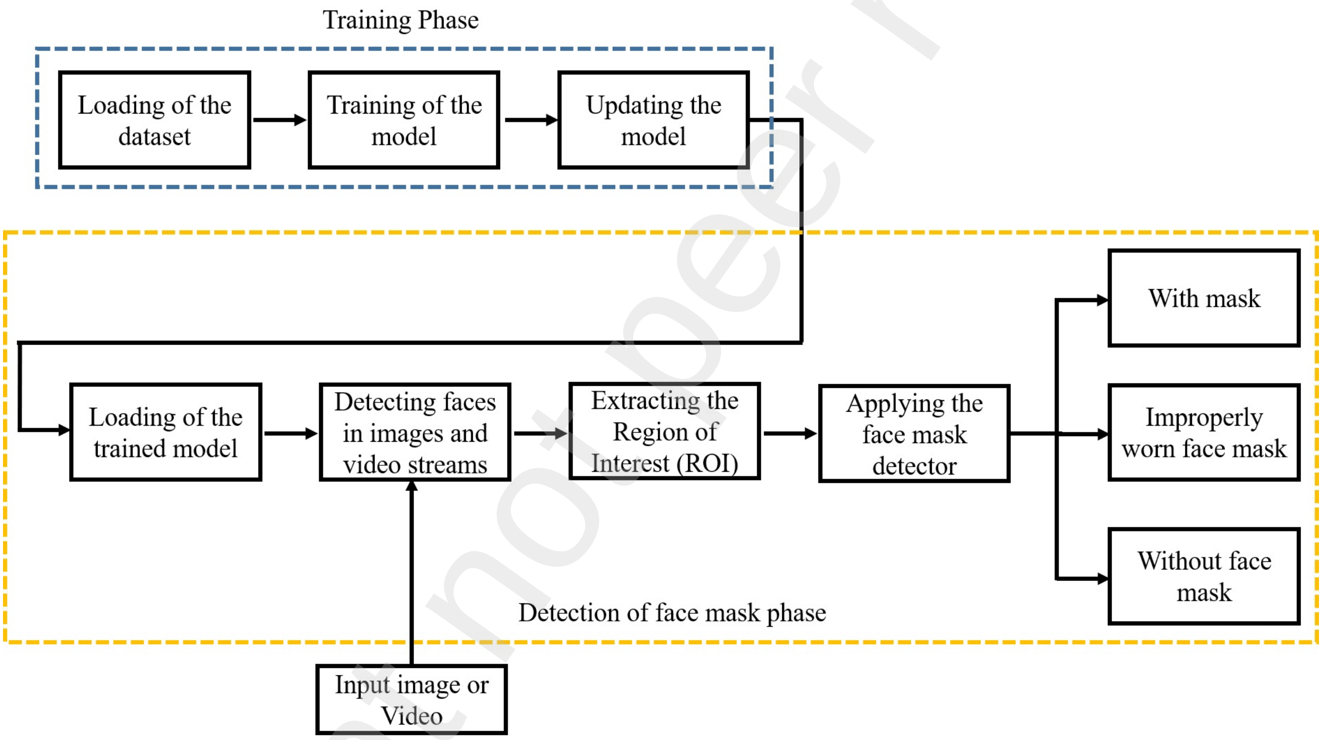

Fig. 2 Training and detection are the two phases of our face mask detector model. The dataset is loaded for the model to be trained and the model is serialized in the training phase. The trained model is loaded, the faces are detected in images and video streams and then the region of interest (ROI)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

types of face masks are surgical face masks, N95 masks, and cloth masks. The primary purpose of a face mask is to trap the droplets that are discharged from a person's mouth while communicating, sneezing, or coughing. For a face mask to be effective, it should cover the mouth and the nose to trap the water droplets. The face mask should not be brought down to the chin. If done so, the face mask will get contaminated as it comes in association with the exposed chin. When the face mask is placed back on the face, the mouth and nose will get infected by the virus, bacteria, or germs. Hence, the face mask should be completely removed in case of eating, drinking, or doing other such activities that involve the removal of a face mask.

Our Facemasknet model aims at detecting whether a person is wearing a mask or not. Along with this, improperly worn masks will also be detected. The results are tested using live a video stream or when an image

	Name	Type	Activations	Learnables
1	data 227x227x3 images with 'zerocenter' normalization	Image Input	227×227×3	-
2	conv1 96 11x11x3 convolutions with stride [4 4] and padding [0 0 0 0]	Convolution	55×55×96	Weights 11×11×3×96 Bias 1×1×96
3	relu1 ReLU	ReLU	55×55×96	-
4	norm1 cross channel normalization with 5 channels per element	Cross Channel Nor...	55×55×96	-
5	pool1 3x3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	27×27×96	-
6	fc 10 fully connected layer	Fully Connected	1×1×10	Weights 10×69984 Bias 10×1
7	prob softmax	Softmax	1×1×10	-
8	classoutput crossentropyex	Classification Output	-	-

Fig. 3 (a)The accuracy versus iteration plot. Fig 4 (b) and Fig 4(c) show the architecture of the Facemasknet model.

has been given as an input to the face mask detector as illustrated in Fig 1. The face gets detected and the region of interest is extracted. Then the input is classified as with a mask, improperly worn mask, and without a mask. Cases in which the mask is improperly worn are when the nose and mouth are partially covered. By monitoring the placement of the face mask on the face, we can make sure that an individual wears it the right way and helps to restrict the scope of the virus. Our face mask identifier is least complex in structure and gives quick results and hence can be used in CCTV footages to detect whether a person is wearing a mask perfectly so that he does not pose any danger to others. Mass screening is possible and hence can be used in crowded places like railway stations, bus stops, markets, streets, mall entrances, schools, colleges, etc. By monitoring the placement of the face mask on the face, we can make sure that an individual wears it the right way and helps to curb the scope of the virus. Previously, binary classification has been performed to detect face masks. We are the first one to performs three-class classification and hence is uncommon.

2 Literature review

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.* [6] work focuses on some commonly implemented deep learning architectures and their applications. The autoencoder, the convolutional neural network, Boltzmann machine, the deep belief networks are the networks that are presented in detail. Deep learning can be used in unsupervised learning algorithms to process the unlabeled data. A CNN model for speedy face detection has been introduced by Li *et al.* [7] that evaluates low resolution an input image and discards non-face sections and accurately processes the regions that are at a greater resolution for precise detection. Calibration nets are used to stimulate detection. The advantage of this model is that it is fast and achieves 14 FPS in case of standard VGA images on the CPU and can be quickened to 100 FPS on GPU.A face detection system called Deep Dense Face Detector (DDFD) was proposed by Farfade *et al.* [8] which we considered the problem of multi-view face detection. The proposed method is least complex and it does not demand segmentation, bounding-box regression, or SVM classifiers and can recognize faces at numerous angles. A novel data augmentation approach for

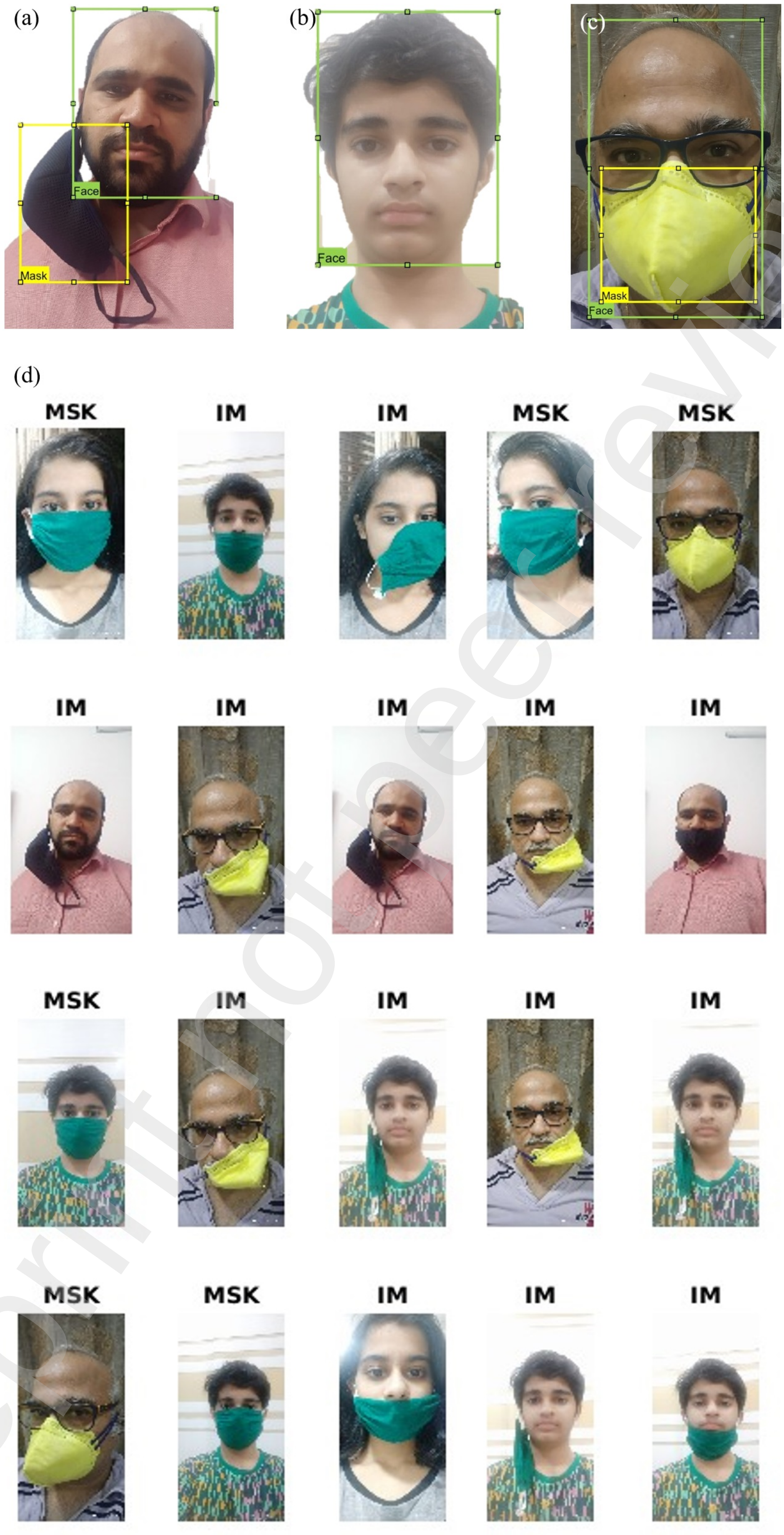

Fig. 4 Results obtained when the facemask detector was tested on various images. The green and yellow red-colored rectangular frames in (a), (b), and (c)respectively represent the detected face and facemask. (d) shows the sample dataset which was used to train the model.

Method	Model	Accuracy
Qin *et al.* [1]	SRCNet	98.70 %
Li *et al.* [2]	FSA-Net , HGL method	Front accuracy 74.97 % , Side accuracy 75.08 %
Khandelwal *et al.* [3]	MobileNetV2	97.6 %
Jiang *et al.* [4]	ResNet and MobileNet	–
Ristea *et al.* [5]	ResNet	74.6 %
Proposed method	Facemasknet	98.6 %

Table 1 A comparison table with existing literature, based on methods used and accuracy obtained.

mask detection from speech was proposed by Ristea *et al.* [5] that could be used for communication amongst surgeons, used in forensic fields or infectious diseases like coronavirus. They have used multiple ResNet models and have trained Generative Adversarial Networks (GANs) with cycle consistency to build their project that could do binary classification. In their future work, they would be focusing on multiclass problems. Wang *et al.* [9] has made executing face mask related projects an obvious task by providing three samples of masked face datasets, which comprise of Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD).

Previously, Khandelwal *et al.* [3] had stated in his work about a deep learning model that binarizes an image as a mask is used or not mask. 380 images had a mask and 460 images had no mask and these images were used in the training of the MobileNetV2 model. The AUROC of the model was 97.6 %. A few limitations were observed in the model. Those remarks were: it could not correctly classify partially hidden faces and the model is not able to detect faces if the camera height is greater than 10 feet. A face mask-wearing classification system with the incorporation of image super-resolution using classification network (SRCNet), was made by Qin *et al.* [1]. It quantified mask, no mask, and incorrectly worn masks, based on 2D facial pictures. Image pre-processing, face detection and crop, image super-resolution, and face mask-wearing conditions identification formed the backbone of the algorithm. The training dataset comprised of 3835 images that included 671 images without a facemask, 134 images of incorrect face mask-wearing, and 3030 images of correct facemask-wearing. SRCNet gave and accuracy of 98.70 % accuracy. A Retina face mask has been proposed by Jiang *et al.* [4] which is a high-accuracy and efficient face mask detector. The models used are ResNet and MobileNet. Transfer learning was used to extract robust characteristics trained on a large dataset of 7959 images. Li *et al.* [2] worked on developing a HGL method for head pose classification with masks, using color texture analysis of pictures and line portraits. Front accuracy of 93.64 % was achieved along with a side accuracy of 87.17 %. The aforementioned project recognizes between wearing a face mask and not wearing a face mask. Matthias *et al.* [10] has done a face mask recognition project that focusses on capturing real-time images indicating whether a person has put on a face mask or not. The dataset was used for training purposes to detect the main facial features (eyes, mouth, and nose) and for applying the decision-making algorithm. Putting on glasses showed no negative effect. Rigid masks gave better results whereas incorrect detections can occur due to illumination, and to objects that are noticeable out of the face.

3 Methodology

Face mask detection has been accomplished by adopting Deep Learning techniques. We have designed our project into two phases: training face mask detector and implementing face mask detector. Fig 2 depicts the training and detection phase of our face mask detector model. The dataset is loaded for the model to be trained and the model is serialized in the training phase. Further, the trained model is loaded, the faces are detected in images and video streams and then the region of interest (ROI) is extracted. Finally, the face mask detector is applied and the images or faces in the video streams are classified as with a mask, improperly worn mask, without a mask. The green and yellow rectangular frame individually interpret the detected face and mask. The dataset consisted of 15 images of improperly worn masks, 10 masked images, and 10 images without a mask. We have used Matlab programming to build our facemask detector model. To train the model, we have used Facemasknet architecture. The initial learning rate is 1e-4 and the number of training epochs is 20. The data is then pre-processed. The images are resized to 227 x 227 x 3 pixels intensities in the input image. After this, the model was compiled to be trained and then the model was evaluated on the test set. The accuracy and iteration curves were plotted. After the model was trained, an image was loaded as an input to distinguish whether a person is wearing a mask or not or wearing it improperly. The input image is then loaded and pre-processed. To localize wherein the image all faces are, face detection takes place and also the region of interest (ROI). The green and yellow rectangular frame respectively represent the detected face and mask. Next, we recognized face masks in real-time video streams. Face detection takes place the same way as discussed earlier. A particular frame from the stream is grabbed and resized. After that detection of face mask takes place. The results are displayed on the screen after post-processing.

3.1 Details of Facemasknet

We have developed a Facemasknet model which is of 8 layers. The accuracy versus iteration plot is depicted in Fig 4 (a). Along with it, (b) and (c) part of the figure show the architecture of the FacemaskNet model. The input layer takes images of size 227 x 227 x 3. The 2D convolution layer is used to take as it takes input which is three-dimensional i.e it has red, blue, and green pixels. The ReLu layer (Rectified Linear Unit) is used to activate functions used for the output in convolutional neural networks. Faster training and better accuracy is achieved due to the norm1 layer. With the help of maxpooling layer, images are resized. The input is multiplied by a weight matrix and then added to a bias vector using the fully connected layer. After using the softmax layer, the required classification is obtained which identifies a face and a mask.

4 Results and Discussions

In all, 35 images were included in the dataset. Out of these cumulative images, 10 had photographs of individuals wearing a mask, 15 pictures of improperly worn masks, and 10 pictures involved a person's face without any mask put upon their faces. This dataset was used to train our Facemasknet model which resulted in an accuracy of 98.6 % in identifying face-masked and without face-masked photographs. The results obtained when the face mask detector was tested on various images is seen in fig 3. The green and yellow red-colored rectangular frames in (a), (b), and (c)respectively represent the detected face and facemask. (d) shows the sample dataset which was used to train the model. Despite having limited training data, our face mask detector model works adequately. In the table shown in fig 5 , a comparison of researchers who have done similar work is executed. A comparison is also produced of the method used and the accuracy obtained. The maximum and minimum accuracy obtained was 98.7 % and 74.97 % respectively.

5 Conclusions

Our work distinguishes face masks from images and live video streams. On training the model using Facemasknet, we got an accuracy of 98.6 %. This classifier was then implemented to images and live video streams. The faces were recognized in images and videos and these faces were extracted. Then, our face mask classifier was applied to achieve the required results. The green and yellow rectangular frame respectively represent the detected face and mask. Our face mask identifier is least complicated in structure and gives instantaneous results and hence can be used in CCTV footages to identify whether a person is wearing a mask correctly so that he does not pose any hazard to others. We are the first ones to perform a three-class classification and hence is uncommon. Mass screening is possible and hence can be used in crowded places like railway stations, bus stops, markets, streets, mall entrances, schools, colleges, etc. By monitoring the placement of the face mask on the face, we can make sure that an individual wears it the right way and helps to curb the scope of the virus.

Acknowledgements Authors would like to thank all colleagues from Ninad's Research Lab.

Compliance with Ethical Standards

Conflicts of interest

Authors M. Inamdar and N. Mehendale, declare that he has no conflict of interest.

Involvement of human participant and animals

This article does not contain any studies with animals or Humans performed by any of the authors. All the necessary permissions were obtained from the Institute Ethical Committee and concerned authorities.

Information about informed consent

No informed consent was required as the studies does not involve any human participant.

Funding information

No funding was involved in the present work.

References

1. B. QIN, D. LI, Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19 (2020)
2. S. Li, X. Ning, L. Yu, L. Zhang, X. Dong, Y. Shi, W. He, in *2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)* (IEEE, 2020), pp. 1–5
3. P. Khandelwal, A. Khandelwal, S. Agarwal, Using computer vision to enhance safety of workforce in manufacturing in a post covid world, arXiv preprint arXiv:2005.05287 (2020)
4. M. Jiang, X. Fan, Retinamask: A face mask detector, arXiv preprint arXiv:2005.03950 (2020)
5. N.C. Ristea, R.T. Ionescu, Are you wearing a mask? improving mask detection from speech using augmentation by cycle-consistent gans, arXiv preprint arXiv:2006.10147 (2020)
6. W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing **234**, 11 (2017)
7. H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 5325–5334
8. S.S. Farfade, M.J. Saberian, L.J. Li, in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval* (2015), pp. 643–650
9. Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, et al., Masked face recognition dataset and application, arXiv preprint arXiv:2003.09093 (2020)
10. D. Matthias, M. Chidozie, Face mask detection application and dataset