

## A Method for Detecting Face Masks Using Deep Learning

Saksham Singh<sup>1</sup>, Diwakar Chaubey<sup>2</sup>

<sup>1</sup>C.S.E, Amity University, Noida, India

<sup>2</sup>C.S.E Amity University, Noida, India

### ABSTRACT

*We all are familiar with this situation and surroundings of what has happened throughout the calendar year 2019, 2020 and continuing in the calendar year 2021. It is a very unpredictable, unknown, and unwanted situation, which we are facing because of the corona and so called COVID-19 virus. The whole world is affected by the terror of this virus. Neither is true, for a city, state, country, or a continent, in fact it is the case of whole world wherever there is existence of human beings and life, there is this danger and terror of COVID-19 virus. There are some medications available for this virus in different countries including India also, such as vaccine but till now there is no medication or vaccine, which ensures 100 percent guarantee after using that specific medicine there will be no further probability of hitting the virus COVID-19. As such, to stop the spread of Corona virus, there have been two important measures, which we can take to stop the Corona virus from spreading, these are "Mask" and "Sanitizer". In this paper, we are covering and construct one of previous mentioned measure, that is covered face mask detection. We will be implementing this work in two steps. First step is to train a model using CNN (Convolutional Neural Network) on a self-made dataset using TensorFlow and Keras & finally implementing it in real-time using OpenCV using our device webcam. This will prove to be a very effective step towards the automation, which will help us in moving further towards the human less system.*

**Keywords:** Convolutional Neural Network, Face mask, Keras, OpenCV, TensorFlow.

### 1. INTRODUCTION

The imperative of placing mask on the face is increasing as a result of ongoing pandemic Corona virus across the globe. Individuals carried face masks to protect themselves from the pollutants, which are present in our immediate vicinity air, for example sulphur dioxide, Lead, Nitrogen, Carbon monoxide etc. It is observed some other people wear face mask to conceal their sentiments from the public. It has been proven scientifically that wearing face mask precludes the COVID-19 virus propagation. Furthermore, it is mandatory to wear mask on the face at public places such as public transportation, the regional metro, and in fact in most of the states of India, it is compulsory to wear mask even if one or more persons is travelling from their private vehicle also. COVID-19 is the latest epidemic which has affected the human beings, so deadly that even WHO (World Health Organization) has stated this as a worldwide epidemic. The latest reports claim that this pandemic has hit over 130 million people out of which over 3 million people lost their lives and it has affected over 200 countries around the globe.

We provide a model for detecting face masks focused on deep learning, machine learning, IOT (Internet of Things), and artificial intelligence in this paper. The proposed model can be used in conjunction with cameras which are being used for the surveillance at different public places and at other places. In fact our model can be implemented anywhere, where there is a setup of CCTV(Closed circuit television) cameras, we can implement it to the places where there is no CCTV cameras also, but we will be requiring a external camera for this scenario and little bit of hardware setup also needed.

This paper has 6 different sections. First section is discussing about introduction section. In the second section we define the problem statement. Third section section is discussing about related work in this area. The dataset is discussed in the the fourth section. Fifth section of this paper Provides information on the methods used.which we are using. The fifth section has some sub-sections also out of which preprocessing is the first sub-section in which we have our method for the pre-processing images and the dataset. Second sub-section is about data augmentation, where our discussion took place the various data augmentation techniques in detail. Then third sub-section has been about various attributes. And the sixth section is the results section here we have discussed all the outcomes of our physical implementation of our model. And the final section is the conclusion section where we have written our final thoughts about the model and implementation.

## 2. PROBLEM STATEMENT

Currently it is observed in the preceding section COVID-19 is a pandemic that has spread throughout the world affected over 130 million population all over the world. Now, the issue that society now faces is that many individuals do not use face masks or some of them wear also but not properly. So, our job is intended to automate system to detect and track down the individual who is not wearing or improperly wearing the face mask. It should be done to do this as to avoid physical contact with the person to person to avoid the spread of COVID-19.

The system should be automated the system fully without any need of manpower to help the working of government officials and by doing so will strengthen our economy system. The implementation cost of our model is zero where any Closed circuit television (CCTV) is already installed set, and it is installable, the setup where there is no CCTV cameras with cost as minimum as 2k-3k Rupees.

## 3. RELATED WORK

Generally, the aim of most journals and academic articles is to concentrate on recognising a person's identity while wearing a face mask. But here our case is different, It has been tried to ascertain if the person's face behind the mask is matching our database or not, it is being tried to find out which person is not wearing face mask. Because, it has been scientifically shown that using a mask prevents and stops the spread of the COVID-19 virus. In early 1994, The use of Neural Networks to detect faces was first proposed in 1994. Here, a CNN (Convolutional Neural Network) is often used to ascertain whether or not a particular pixel is part of the face or not, and then another CNN is used to determine the position of the face. Following that, a large number of researchers focused on the model, advancing it to the next level. There has been a lot of testing done in the last three years on deep learning, a subject that has the use of Computer Vision technologies, including facial recognition, has been greatly promoted. Li et al. [12] suggested a cascaded CNN model for fast fce detection, that is a multi - scale network

capable of removing contextual areas in low res and thoroughly testing problematic selections in the resulting high pixel image. Ranjan et al. [3] suggested a normalization-based Deep convolutional neural network extracts model characteristics. Wu et al. [8], the authors have categorized the face mask wearing in 3 classifications. Right face covered mask wear, wrong face covered mask wear, and no face covered mask wear are the three types. It is a reliable algorithm that achieved a research accuracy of 98.70 percent. Qin et al. [9] used PCA (Principal Component Analysis) to analyze to identify the individual using covered masked and uncovered mask face recognition info. And as result, wearing masks has a major impact on the performance of facial recognition using PCA. When an individual is wearing a mask, the accuracy rate drops below 70%. Ejaz et al.[10] used PCA as well, proposing a process in which glasses are separated from the face and the face is restored using PCA restoration.

Park et al. [11] used the YOLOv3 algorithm. The YOLOv3 algorithm's backbone is Darknet-53. They were able to reach a 93.9 percent accuracy rate. CelebA and the WIDER FACE datasets were used to train the algorithm. Nizam et al proposed a new GAN- driven model that is capable of automatically erase the mask, rebuild the region of the face, and restore the area via the completion of missing hole. This model generates a full picture of the face that appears authentic and normal. The writers suggested a method for determining whether or not a face mask could be worn inside the operating room in [13]. The objective was to significantly reducing the amount of false positive face detections for left detections in masks as much as conceivable in order to alert medical personnel who may not use a surgical mask in the surgery room. The model's accuracy was estimated to be about 95%. MRGAN is an active approach suggested by Rodriguez et al [14]. This strategy entails creating a picture in which a portion of the covered mask is obscured by the microphone. It eliminates the microphone and makes use of a Generative Adversarial Network in order to recreate the missing face region. Deep learning true face emotions detection and recognition was used by [15]. VGG-16 was used to distinguish various expressions of the face. The system was constructed using the KDEF dataset and reached a precision of about 88%.

The layout and orientations of several distinct faces were simulated using DPM in order to perform categorization based on Deformable Part Models. In 2006, Ramanan proposed a Randomized forest tree approach for describing face masks, which correctly predicts face structures and facial poses. One of the renowned experts, Zhang et al [33], developed a face mask detector based on DPM employing about 30,000 faces split into covered masks and without the usage of masks, the task had a remarkable precision of 97.14 percent. There was a development further in face mask detector models. Face mask detection models based on DPM will typically attain magnificent accuracies, not as a result of the use of DPM, they must be tolerant of the exceptionally high computational cost. The identification of faces has been maintained to be used in use since the publication of the initial real-time face detection system, Viola-Jones [17], in 2004. The well-known Viola- Jones can perform real-time detection using the Haar function and cascaded configuration, but it has some disadvantages, including a large feature size and a poor recognition rate in complex scenarios. In the meantime, several systems for COVID19 in networked smart cities have been created. [18] launched the BlueDot and HealthMap services. The BlueDot methodology was first used to detect a cluster of uncommon neumonia at Wuhan, that finally resulted in the pandemic being declared. Additionally, the infection was believed to travel through Wuhan to Hong Kong, Taepei, Bankok and Singapore. Using big data, and artificial intelligence the San Francisco-based

HealthMap service identified people with a cough, the initial sign of COVID-19. Haleboua [19] introduces a report on the use of facemasks to inhibit the development of COVID-19. The research discovered that masks that are correctly fitted efficiently prevent the spread of particles released while sneezing or coughing. Masks which aren't properly fitted can retain airborne virus and pollutants. Jones and Allam [20] suggested a network concept for smart cities focuses regarding the exchange of data during the COVID-19 epidemic. The latest project examined the economic and national security effects of evidence on urban health. The device collects data from different positions in the city using cameras, trackers, and laboratories. Jiang et al. [22] propose the usage of the RetinaFaceMask paradigm in combination with just one cross-class entity removal algorithm for detecting face masks. The advanced model has a two-stage detector comprised of a hopfield neural network, resulting in slightly enhanced recall and precision over the standard result. They used transfer learning has been a well technique for deep learning method, to address the dataset scarcity issue. Gupta et al. [23] suggested a proposal for enforcing social distance by the use of smart cities and Intelligent Transportation Systems (ITS) is activated in the case of a COVID-19 pandemic. Their model entailed the placement of sensors across the city to track the location of items in real time and included a data-sharing network. Won Sonn and Lee [24] demonstrate the significant contribution of a smart city to coronavirus management in South Korea. A time-space cartographer accelerated the city's touch monitoring, which included patient travel, sales background, mobile phone use, and position. On CCTV cameras installed in residential building hallways, real-time surveillance has been conducted. Singh et al. [25] focused on the function of IoT in combating COVID-19. The established structure places a premium on interconnected devices or operations that allow for the tracking of patients as well as suspicious events. A well-informed community is developed utilising interconnected instruments to greatly classify clusters. Sonn et al. [26] created a unique epidemic surveillance concept in a smart city that does not need lockout. Patients were interviewed and their previous movements were tracked. They argue that certain patients attempted to hide their previous mobility, but a real-time monitoring device discovered the precise details. Jaiswal et al. [27] suggested a method for hazard mitigation in COVID19. Their suggested methodology took advantage of the position of equipment to discover persons who were infected. Drone and bots were either utilised to give emergency treatment to afflicted persons. Wang et al. [28] discussed the expansion of smarter cities in China through COVID19 and the management of pandemics in the country. The uninterrupted delivery of critical resources and the contactless logistic delivery of devices to civilization prepared the path for coronavirus to be contained. The concepts of real-time image reflections and ITS were applied to stymie vehicle travel during the pandemic. Additionally, driverless cars were used to track the situations as they unfolded in the region. Recently, some scientists and investigators have focused on colorless face photographs [29]. While some relied exclusively on pattern matching algorithms with previous information of the reference image, others used AdaBoost [30], a strong predictor for testing and training. Following it, the Jones-Voila Analyzer was developed, a significant technical advancement. Following it, the Jones-Voila Analyzer was developed, a significant technical advancement that allowed real-time face detection. It encountered a variety of difficulties, including the face's direction and visibility, which made it difficult to intercept. Thus, it was unable to work in dull and dim sun. As a consequence, researchers started searching for a new paradigm which is able to identify both faces and facial masks. Numerous face detector datasets are being created over time to aid in the comprehension of face mask

recognition systems. Previously collected data sets built using photographs gathered in supervised settings, while some recent datasets use web images such as WiderFace [31], annotations for current faces in these databases are given in comparing to previous versions. Large databases are quite essential for creating superior testing & training data and performing true implementations in a significantly more direct approach. This demands the development of a large number of computer vision algorithms capable of directly recognizing pictures and masks coverings from human intervention. When anyone is contagious with a cold or fever, infectious particles comprising bacterial (droplets bacteria and viruses are emitted [32]. However, prior to the latest H1N1 and H3N2 influenza viruses exposing their medical implications and global consequences, the two most recent flu pandemics in 2003 and 2009 received no press attention. a number of pandemic pathogenicity outbreaks, during which several efforts were made to prevent or minimize airborne pathogens, as well as the 5 to 10-micron range (resolution limit) for pathogens is an extremely tiny range, much smaller than the 1 to 2 micron (v-limit) range, and includes many cases of respiratory viruses; unfortunately, this can occur not only there, but often elsewhere. This is distinct from all other forms of airborne diseases, which are distributed by airborne particles, since airborne pathogens must be contained inside droplets to be transported and spread, and this mode of transmission will travel up to 1 m through extended stretches of time in the air of time. COVID19 was initially believed to be distributed by the pulmonary system, however it has since been determined that it is mainly distributed via airborne transmission. Controlling airborne pathogens requires a variety of approaches, since diverse parties have differing views about the effectiveness of specific strategies. Numerous studies have been undertaken. on droplet wind flow and dispersal in order to get a greater understanding of how to enhance droplet distribution [33].

#### 4. DATASET

There are a lot of datasets available online but there are very few datasets which have everything correct with it. Some dataset is created with doesn't represent real picture accurately, some dataset are very noise or some with the wrong labels. So, choosing the correct dataset was a difficult task for us. But, we have managed to make a dataset which has it's resources from multiple online and offline sources. Such as few images were taken from Kaggle, few from other online resources and around 200 images were clicked by us manually. Finding our dataset was difficult because we just don't need a face mask detector in fact we need a 180 degree face mask detector for which we need images with side view also but that was not available online anywhere. So, we had to create our own dataset in which we included 6 views of a single person which from left, from right side and from front side which includes both with and without mask. Our collection has 1400 photos in total, including 700 photos with mask and same amount of images without mask. To apply our algorithm, first we need a picture without a face mask, from here we will apply face detection and once face is detected in the frame then we will be detecting facial landmarks such as eyes, nose, mouth, jawline etc. Now once we detect the facial landmarks it becomes very easy for us to locate at what position the mask will be placed and then the mask will be placed at the right position using the facial landmarks. So, it is a very important step to detect the facial landmarks to proceed further for which we need a very clear image without any distortions.

## 5. METHODOLOGY

To find out whether a face is covered with the face mask in right manner, the first step is to construct the system with the appropriate dataset. The dataset's specifics have already been addressed in the third section. Once the classifier has been trained, an exact face detection model must be used to recognize the faces in order our framework can infer whether or not the individual has worn a mask. Our key goal is to improve the model's accuracy without making it too resource-intensive. The DNN module from OpenCV is being used to complete this mission. It contains an objective detection model that is built on the ResNet-10 concept and architecture basis. This method aids in the real-time detection of faces, even on small devices like the Raspberry Pi. To ascertain whether the individual is putting on a mask on the face, the classification algorithm employs a qualified model called MobileNetV2. The method used in this case is seen in Figure 2.

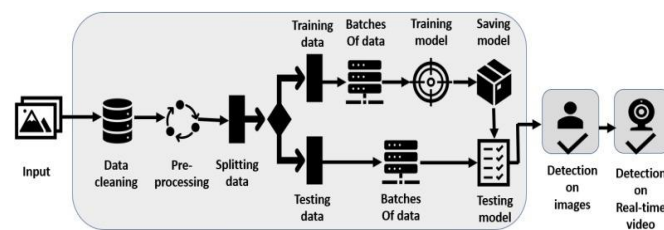


Figure 1: Flow Diagram

Nagrath P, Jain R, Madan A, Arora R, Kataria P, Hemanth J. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustain Cities Soc. 2021;66:102692. doi:10.1016/j.scs.2020.102692.

We can see the flow of our model how it will work in the above figure 1. First an image will be taken as an input in one frame. Then next comes the phase of data cleaning where we try to remove the noise from the data. Then we will apply different pre-processing techniques and then we will split the data in the training and testing batches. After that, the network is generated using the provided data. Finally we will execute our model on static images. And at the end we will test it in real-time using OpenCV.



Figure 2: Extract face ROI with OpenCV and Numpy

## 5.1 Pre-Processing

The term "pre-processing" refers to the which is related with data based problems such as data table, but in images also there is pre-processing. In images basically there are two types of pre-processing, which area analogue and digital. In this research paper, we will talk about digital image processing which is being used in the model we use. The term "digital image processing" refers to the process of analysing data images through computer methods.



Figure 2: Detecting facial landmarks

Opposite to analogue image processing, digital image processing has many advantages. It is far more capable of supporting a significantly bigger number of algorithms. The purpose of digital image processing is to improve the quality of images by using various strategies such as eliminating unnecessary distortions and maximizing a few key image features so that artificial intelligence and computer vision can make greater use of it. An image can simply be described as an two-dimensional elements of similar type (array) of numbers(pixels) ranging between 0 to 255.  $F(x,y)$  defines it's mathematical function here y and x are coordinates of vertical and horizontal axis respectively.

## 5.2 Data augmentation

Augmentation of data is a very important part of our model and it is required because it gives a versatility to the dataset. Here we rotate, flip, shear, slice, zoom etc to improve the amount of data feed to the model. It gives us a confidence that no format of the same picture is missed all the forms are fed to the model correctly.

## 5.3 Attributes

We have different attributes to identify whether the face is present in the frame or it is not present in the frame by which we can identify and pass the obtained result to our model so that it can do the remaining work of finding whether the face is covered with the mask or not covered with the mask. Following are the attributes which we will detect to find whether face is present or not present in the frame:

- Eyes
- Nose

- Eyebrows
- Jaws
- Lips

## 6. FINDINGS

The phrase "image classification" involves the procedure through which a photograph is assigned a class label. Localization is the process of creating a classification model centred on an object in a photograph. The more difficult process of item recognition combines these two tasks by creating a classification model around the whole desired area in the image and giving it a target class. Consider adjusting simple logical or micro techniques to generate the object's bounding areas. We just change the output labels to instruct our model on the entity's class and position in the image. We augment the output layer with four more numbers, including the object's centroid position and the breadth and length proportions of the feature vector in the image. This overcomes the problem of features extraction. Now, instead of just one item, we want our system to be capable of categorising and localising all items included inside a snapshot. As a result, the approach is to crop the image into many images and then execute CNN on each compressed image.

To begin, create a window that is much smaller in size than the original picture. Crop it and send it to CNN for prediction making. Continue sliding the window open and sending the clipped photographs to CNN. After resizing the picture to fits within this scale factor, perform every one of the previous steps for a slightly bigger scale factor. Again, provide CNN with cropped images and allow it to make predictions. Finally, you'll get a collection of cropped regions that contain an object, along with information about the object's subclass and bounding box.. Object detection with sliding windows is the term used to describe this approach. The YOLO algorithm makes use of this concept to identify objects. YOLOv3 employs consecutive 3×3 and 1×1 convolutional layers and also includes some shortcut connections. It is composed of 53 convolutional layers.

The quickest R-CNN is the most often used R-CNN variation. These networks are often composed of the following components: a) A area suggestion methodology for producing "bounding boxes" or areas of likely things in the image. b) A step in which these objects' features are generated, often utilising a CNN. c) A classifying layer that identifies the category which that item belongs and d) A regression network that increases the accuracy of the item's feature vector . Fast R-CNN suggested areas are discovered using a Hardware selective search methodology that takes around two seconds per image and is totally Hardware. The Faster R-CNN research solves this problem by employing the Region Proposal Network to generate region ideas (RPN). This decreases the time required for feature map from two seconds to ten milliseconds every image and permits the area propose phase to swap layers with future detecting phases, resulting in an improvement in total image classification.

## 7. RESULTS

In this section we have discussed different techniques available for face covered mask observance and in this results part we will summarize all the things and pack in a nutshell.

Technique	Size	Time Consumption(ms)	Average Precision	Average Precision 50	Average Precision 75
YOLOv3	400	29	49.5 %	93.6 %	71.1 %
SE-YOLOv3	700	36	56.2 %	94.1 %	72.3 %
SSD	450	38	55.4 %	92.5 %	79.8%
RetinaNet	628	40	72.4 %	97.2 %	80.5 %
EfficientDet	828	90	66.9 %	92.3 %	69.2 %

**Table 1: Comparison of different techniques**

Only candidates that cover more than 50% of the area when compared to the ground reality are included for AP50, yet only contenders that cover more than 75% of the area when compared to the real data are included for AP75. Then we can say average precision of 50% covered area will be the greatest value and then average precision 75 will be lesser than that of average precision 50 and average precision value will be the lowest as observed in the table 1 above.

As we can observe from table 1 above highest precision is for RetinaNet which when combined with MobilkeNetV2 will produce the best results in terms of accuracy but if we talk about time consumption SE-YOLOv3 is the fastest. Now if talk about least precision which here is EfficientDet which has an precision of around 92.3 percent and in fact it is the most time consuming method also where it consumed 90 milli second.

## 8. CONCLUSION

To further develop our covered f mask project, it is necessary to compile authentic photographs of persons wearing masks (rather than photoshopped images). While our fictitious dataset performed beautifully in this case, there is no feasible substitute for the true object. Second, to amass images of looks that may "confound" our classifier into believing the subject is wearing a cover while, in fact, they are not completely qualified models, including shirts folded over their faces, handkerchiefs over their mouths, and so forth. These are cases when our face cover marker is confused as a face veil, and so it should be evaluated for developing a dedicated two entity identification rather than a simple image classification. The argument in favour of this method is that by definition, a face mask darkens the skin tone of the face. When the whole picture is darkened, the face cannot be identified, and so the face cover finder is disabled. To circumvent this problem, we need create a two-class object finder consisting of a class with covered face masks and a category lacking covering face masks..

## REFERENCES

- [1] H. Qin, J. Yan, X. Li and X. Hu, "Joint Training of Cascaded CNN for Face Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3456-3465, doi: 10.1109/CVPR.2016.376.
- [2] Yang, B., Yan, J., Lei, Z., & Li, S. Z. (2017). Convolutional channel features. In *Proceedings of the IEEE international conference on computer vision* (pp. 82-90).
- [3] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Transactions on Pattern*

- Analysis and Machine Intelligence, vol. 41, no. 1, pp. 121–135, 2017.
- [4] H. Li, Z. Lin, X. Shen, J. Brandt and G. Hua, "A convolutional neural network cascade for face detection," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5325–5334, doi: 10.1109/CVPR.2015.7299170.
- [5] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 650–657, Washington, DC, USA, June 2017
- [6] Li, L. Q., Huang, T., Wang, Y. Q., Wang, Z. P., Liang, Y., Huang, T. B., ... & Wang, Y. (2020). COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. *Journal of medical virology*, 92(6), 577-583.
- [7] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: an improved faster RCNN approach," *Neurocomputing*, vol. 299, no. 1, pp. 42–50, 2018.
- [8] W. Wu, Y. Yin, X. Wang, and D. Xu, "Face detection with different scales based on faster R-CNN," *IEEE Transactions on Cybernetics*, vol. 49, no. 11, pp. 4017–4028, 2019.
- [9] QIN and D. Li, Identifying facemask-wearing condition using image super resolution with classification network to prevent COVID-19, May 2020, doi: 10.21203/rs.3.rs-28668/v1.
- [10] M.S. Ejaz, M.R. Islam, M. Sifatullah, A. Sarker, Implementation of principal component analysis on masked and non-masked face recognition, in: *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019.
- [11] Jeong-Seon Park, You Hwa Oh, Sang Chul Ahn, and Seong-Whan Lee, Glasses removal from facial image using recursive error compensation, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (5) (2017) 805–811, doi: 10.1109/TPAMI.2005.103.
- [12] C. Li, R. Wang, J. Li, L. Fei, Face detection based on YOLOv3, in: *Recent Trends in Intelligent Computing, Communication and Devices*, Singapore, 2020, pp. 277–284, doi: 10.1007/978-981-13-9406-5\_3.
- [13] N. Ud Din, K. Javed, S. Bae, J. Yi, A novel GAN-based network for unmasking of masked face, *IEEE Access* 8 (2020) 44276–44287, <https://doi.org/10.1109/ACCESS.2020.2977386>.
- [14] A. Nieto-Rodríguez, M. Mucientes, V.M. Brea, System for medical mask detection in the operating room through facial
- [15] Attributes, *Pattern Recogn Image Anal. Cham* (2015) 138–145, [https://doi.org/10.1007/978-3-319-19390-8\\_16](https://doi.org/10.1007/978-3-319-19390-8_16).
- [16] M.K.J. Khan, N. Ud Din, S. Bae, J. Yi, Interactive removal of microphone object in facial images, *Electronics* 8 (10) (2019) , Art. no. 10, doi: 10.3390/electronics8101115.
- [17] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2014.
- [18] G. Halegoua, "Smart City Technologies", *Smart Cities*, 2020.
- [19] L. P. Garcia, "Uso de máscara facial para limitar a transmissão da COVID-19", *Epidemiol. e Serv. saude Rev. do Sist. Unico Saude do Bras.*, vol. 29, no. 2, pp. e2020023, 2020.

- [20] Z. Allam and D. S. Jones, "On the Coronavirus (COVID-19) Outbreak and the Smart City Network: Universal Data Sharing Standards Coupled with Artificial Intelligence (AI) to Benefit Urban Health Monitoring and Management", *Healthcare*, vol. 8, no. 1, pp. 46, 2020.
- [21] M. Jiang, X. Fan and H. Yan, "RetinaMask: A Face Mask detector", 2020, [online] Available: <http://arxiv.org/abs/2005.03950>.
- [22] M. Gupta, M. Abdelsalam and S. Mittal, "Enabling and Enforcing Social Distancing Measures using Smart City and ITS Infrastructures: A COVID-19 Use Case", 2020, [online] Available: <https://arxiv.org/abs/2004.09246>.
- [23] J. Won Sonn and J. K. Lee, "The smart city as time-space cartographer in COVID-19 control: the South Korean strategy and democratic control of surveillance technology", *Eurasian Geogr. Econ.*, pp. 1-11, May. 2020.
- [24] R. P. Singh, M. Javaid, A. Haleem and R. Suman, "Internet of things (IoT) applications to fight against COVID-19 pandemic", *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 4, pp. 521-524, Jul. 2020.
- [25] J. W. Sonn, M. Kang and Y. Choi, "Smart city technologies for pandemic control without lockdown", *Int. J. Urban Sci.*, vol. 24, no. 2, pp. 149-151, 2020.
- [26] R. Jaiswal, A. Agarwal and R. NEGI, "Smart Solution for Reducing the COVID-19 Risk using Smart City Technology", *IET Smart Cities*, vol. 2, pp. 82-88, 2020.
- [27] X. Wang, X. Le and Q. Lu, "Analysis of China's Smart City Upgrade and Smart Logistics Development under the COVID-19 Epidemic", *J. Phys. Conf. Ser.*, vol. 1570, pp. 012066, 2020.
- [28] Nagrath P, Jain R, Madan A, Arora R, Kataria P, Hemanth J. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustain Cities Soc.* 2021;66:102692. doi:10.1016/j.scs.2020.102692.
- [29] Kaihan Lin, Huimin Zhao, Jujian Lv, Canyao Li, Xiaoyong Liu, Rongjun Chen, Ruoyan Zhao, "Face Detection and Segmentation Based on Improved Mask R-CNN", *Discrete Dynamics in Nature*.
- [30] K. Suresh, M. Palangappa and S. Bhuvan, "Face Mask Detection by using Optimistic Convolutional Neural Network," *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2021, pp. 1084-1089, doi: 10.1109/ICICT50816.2021.9358653.
- [31] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic.
- [32] J. Zhang, F. Han, Y. Chun and W. Chen, "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, 2021, doi: 10.1109/ACCESS.2021.3066538.
- [33] Huisi Wu, Xuheng Lu, Baiying Lei, Zhenkun Wen, Automated left ventricular segmentation from cardiac magnetic resonance images via adversarial learning with multi-stage pose estimation network and co-discriminator, *Medical Image Analysis*, Volume 8, 2021, 101891, ISSN 1361-8415, <https://doi.org/10.1016/j.media.2020.101891>.