Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 5, Special Issue "COVID-19: Emerging Research", 2021

Real-Time Monitoring of COVID-19 SOP in Public Gathering Using Deep Learning Technique

Muhammad Haris Kaka Khel [1], Kushsairy Kadir [1*], Waleed Albattah [2], Sheroz Khan [3], MNMM Noor [4], Haidawati Nasir [4], Shabana Habib [2], Muhammad Islam [3], Akbar Khan [1]

[1] *Electrical Section, Universiti Kuala Lumpur British Malaysian Institute, 53100, Malaysia*

[2] *Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia*

[3] *Department of Electrical Engineering, College of Engineering and Information Technology, Onaizah Colleges, Qassim, Saudi Arabia*

[4] *Computer Engineering Section, Universiti Kuala Lumpur Malaysian Institute of Information Technology, Kuala Lumpur, 50250, Malaysia*

Abstract

Crowd management has attracted serious attention under the prevailing pandemic conditions of COVID-19, emphasizing that sick persons do not become a source of virus transmission. World Health Organization (WHO) guidelines include maintaining a safe distance and wearing a mask in gatherings as part of standard operating procedures (SOP), considered thus far the most effective preventive measures to protect against COVID-19. Several methods and strategies have been used to construct various face detection and social distance detection models. In this paper, a deep learning model is presented to detect people without masks and those not keeping a safe distance to contain the virus. It also counts individuals who violate the SOP. The proposed model employs the Single Shot Multi-box Detector as a feature extractor, followed by Spatial Pyramid Pooling (SPP) to integrate the extracted features to improve the model's detecting capabilities. The MobilenetV2 architecture as a framework for the classifier makes the model highly light, fast, and computationally efficient, allowing it to be employed in embedded devices to do real-time mask and social distance detection, which is the sole objective of this research. This paper's technique yields an accuracy score of 99% and reduces the loss to 0.04%.

Keywords:

COVID-19;
Social Distancing;
Crowd Management;
Hajj Umrah; Mask Detection;
Convolutional Neural Network.

Article History:

Received:	02	August	2021
Revised:	18	October	2021
Accepted:	27	October	2021
Published:	16	November	2021

1- Introduction

Gatherings of people referred to as crowds are part of our daily human experience. Everyday activities such as commuting to work via city inner transportation means or going to malls in retail environments, and social events such as visiting bars and restaurants or entertainment stand for music festivals, football matches, visiting themes or amusement parks and museums - are all examples of crowded places of public gathering and interests. Crowd management has been becoming increasingly necessary in areas of public gathering [1]. This management is done by closely collaborating between event managers and planners, departments attending to emergencies, local tending and transportation authorities, supervisors, and participating public delegates and groups [2].

Crowd management is closely related to understanding precisely the participants' spatial behaviors and psychological conditions and interactions. It ensures security by managing comfortably the conditions that might be leading to the building up of disasters in consequence. It has been observed that opening exits at places of public convenience has

[*] **CONTACT**: kushsairy@unikl.edu.my

DOI: http://dx.doi.org/10.28991/esj-2021-SPER-14

proven to be more helpful than exercising police cordons preventing access to (or exit from) the events' sites [3]. It is understood that human individuals imitate each other. The incorrectly coordinated behaviors may lead to herding attitudes in the absence of centralized direction, as the social learning strategists dealing with human crowd dynamics [4]. The wearing of masks has been declared mandatory to reduce the spread of droplets and infectious aerosols. When infected individual coughs (or sneezes), the COVID-19 virus is spread through droplets of saliva or nasal secretions, which could cause adverse effects to others in crowded locations [5]. Though masks wearing in crowds, individuals protect themselves and others from contracting the virus, though, in the virus-infested environment, even the mask-wearing individuals could be susceptible to infection [6]. Since March 2020, when COVID-19 was declared a pandemic, measures from social distancing, wearing masks, and quarantining people for fourteen days (or longer) have been advised as part of SOP to break the chain of viruses' transmission. COVID-19 has infected over 199 million individuals worldwide and over 4.24 million deaths of individuals as of August 2, 2021. It is thus acknowledged to suggest strategies capable of restricting the transmission and spread of contagious sickness due to a significant number of verified cases arising daily under the existing ongoing pharmaceutical practices [7].

A study has suggested that reductions in physical activities because of COVID-19 restrictions lead us to have an effect on mental health in the form of severe anxiety and depression [8, 9]. The preventive measures said to be non-pharmaceutical interventions (NPIs), are necessary as they seem to be the only means for limiting the spread of disease by reducing an increase in the number of affected individuals, standing at 523,397 cases (with 8212 or 1.57% deaths) recorded in the Kingdom of Saudi Arabia. Such measures include 1) social distancing, 2) frequent washing of visible body surfaces, and 3) personal protective means of masks and head-covering, as the virus will last for a duration longer than the original forecasts [10]. The implementation of these measures is always essential because they produce demographically dependent results. An awareness campaign is suggested for creating the perception of COVID-19 among the public about avoiding crowded places, wearing masks, and washing visible surfaces of the body regularly and adequately after and during the time in public [11]. Artificial intelligence techniques are practical and can be employed in different ways to implement these measures to stop the COVID-19 virus from spreading [12]. Emerging technologies such as the Internet of Things (IoT), artificial intelligence (AI), big data, Deep-Learning (DL), and Machine Learning (ML) are utilized to diagnose the COVID-19 cases more quickly [13-16] for promoting awareness.

This paper proposes a model that uses the CNN technique, which is trained on the available dataset. Results are obtained by uploading photos and videos of people who need to be detected in numbers for those wearing masks and those not wearing the masks. Also, to detect those violating social distancing in crowded places. Such applications require that our Image enlargement techniques increase training data leading to improved performance to justify the speed of the proposed model.

2- Related Work

In the light of developments in mobile communication and advanced electronics, there lies the need that such social campaigns are supported by an architecture making use of pictures and videos for ensuring that in large gatherings, masks are worn all the time as a precautionary measure one and that social distancing of two meters gap is maintained in between individuals in a gathering as precautionary measure two.

Convolutional Neural networks (CNNs) have become popular among researchers and scientists as they are very efficient in feature extractions and complex object detection, such as human detection and face detection [17]. The Convolutional Neural Network is designed to process data made of grid patterns such as images and then automatically learn the spatial hierarchies of features and learn the low-to-high level patterns in an adaptable manner [18, 19]. A typical CNN structure is composed of convolution, pooling, pooling layers, and fully connected layers. The first two sets of layers extract the features, and the final layer maps the extracted features into becoming the final output. The hidden layers in CNN make it a more complex network to enhance feature learning and feature expression abilities [20]. The credits go to the more technologically developed CPUs, GPUs, and memory storage expansion for CNN's accurate and fast detection. The AlexNet [21], the 2012 ImageNet Competition winning model that used GPU to speed up computational operations, has given deep learning a new lease in life. However, there are still some challenges to be solved. It takes a long time during training and requires a significant computation time, slowing thus the speed of object detection and recognition. Object recognition has received a lot of attention as a crucial task in computer vision [22].

The two-stage detector method uses a heuristic algorithm like CNN to produce many candidate region recommendations for each image, then reviews to classify these regions. To build a sparse collection of candidate regions, R-CNN [23] initially uses selective search. The region characteristics are then extracted using CNN, and the class of each object is determined using SVM. It uses linear regression to fine-tune the bounding boxes. Although two-stage detectors have excellent detection performance, their training processes are complicated, and the testing rates are often slow, making them unsuitable for real-time applications. On the contrary, the one-stage detectors like YOLO series, SSD [24], RetinaNet [25], EfficientDet [26], and RefineDet [27] have directly classified and predicted on the entire image in use. They have a breakneck detecting speed, but all this comes at the expense of positioning correctness.

2-1- Face Detection

The Contextual Multi-Scale Region-based Convolutional Neural Network (CMS-RCNN) proposed by Zhu et al. [28] has substantially impacted face detection models. Ejaz et al (2019) [29] have used and implemented the Principal Component Analysis (PCA) method to recognize masked and unmasked facial surfaces. The model accuracy has been 96.25% in recognizing the faces without masks, but later on, the model accuracy has reduced to 68% when the model predicts faces with a mask. Loey et al. (2021) [30] have used a hybrid transfer learning model, and machine learning approaches for better feature extraction and classification. On the Real-Time Face Dataset (RMFD), 97% on the Masked Face Dataset (SMFD), the ultimate accuracy has been reached to 97% by Wang et al. (2020) [31]. Cabani et al. (2021) [32] have presented masked face images based on facial feature landmarks and developed a massive dataset of 137016 masked face photos, thus allowing for more training data. However, the data obtained may not be completely applicable to a real-world scenario, as they did not consider detection speed. Li et al. (2020) [33] have utilized YOLOv3 for face features detection, based on the DARKNET-19 deep learning network architecture. The model has been trained by using the FDDB dataset. This model has achieved an accuracy of 93%. Research on mask detection reported today has not been very extensive alongside the fact that improvement is required in every method. Zhao et al. (2020) [34] have proposed mixed YOLOv3-LITE, and it is aimed at embedded and mobile smart devices. To achieve the merging of shallow and deep features, Mixed YOLOv3-LITE employs the shallow backbone of YOLO-LITE to substitute the DarkNet-53 by adding a residual structure and parallel in high-to-low-resolution sub-networks.

The Global Context block introduces fusion between the feature extraction network and the feature pyramid network in GC-YOLOv3 [35]. The model can focus on different regions by using the global context block to create a relationship of long-range dependence between all feature pixels in the feature map. The output of the extraction network is used by the learnable semantic fusion component, which improves the head networks' ability to recognize objects in feature maps. With the addition of these components, the GC-YOLOv3 considerably improves accuracy while incurring only a minor increase in computation cost. Punn et al. [36] have presented a system for identifying and tracking persons that have used the YOLOv3 with the DeepSort algorithm.

2-2- Social Distance Detection

Ma et al. (2016) [37] have used a Deep Neural Network (DNN)-based approach for detection to monitor those in violation of social distancing. However, the outcome of their results has no statistical analysis. Moreover, no argument is provided for distance measurement. Ainslie et al. (2020) [38] have established the relationship between social distancing and the region's economic situation and suggested that all measures avoid extensive breakouts. Nguyen et al. (2020) [39] have given an overview of how a mix of emerging technologies of Bluetooth, Wi-Fi, GPS, and mobile phones equipped with related software of computer vision, image processing, and deep learning - all playing their essential role in ensuring social distancing in situations of varying scenarios.

Prem et al. (2020) [40] have looked into social isolation's mental and economic consequences during the prevailing pandemic. The analysis has revealed that establishing and maintaining social distance early on could help stop the pandemic from peaking. On the other hand, social distance is necessary for flattening the infection curve, though it might show up monetarily an inconvenient move. Pouw et al. (2020) [41] have suggested a graph-based method for mass management and monitoring of social distancing. Ahmad et al. [42] have described a top-view model for detecting and tracking individuals in crowds. The intensive care unit has been explored for the presence of SARS-CoV-2 viruses by Jin et al. [43]. Bhattacharya et al. (2021) [44] have investigated using a deep learning-based approach to identify COVID-19 infection using medical image processing techniques.

3- Research Methodology

To make our model more robust and accurate, the Spatial Pyramid Pooling (SPP) is introduced as a layer between convolutional layers and fully connected layers in Mobilenetv2. For fully connected layers, we have required fixed-size input. The SPP layer replaces the last pooling layer to produce a fixed-size input for fully connected layers. To put it in another way, information aggregation at a deeper level of the network structure in between convolutional layers and fully connected layers is presented to prevent the necessity for cropping or wrapping at the start. Also, the Mobilenetv2 is combined with a Single Shot Detector (SSD) for enhanced accuracy and speed of the proposed model in crowds. The SSD layers' task is to turn the pixels in the input image into features that characterize the image's contents and then pass those features on to the other layers in the row.

The SPP-SDD-Mobilenetv2 model consists of two phases. The first phase is about training the model for which we had loaded the required dataset from disk and then trained the model using TensorFlow and Keras on it before we saved the model. The second phase involves loading the detector, performing face detection, and detecting the social distancing features in images or video streams. A face detector model is required to recognize faces and detect masks [45] accurately. The Single Shot Detector (SSD) has been utilized, an algorithm primarily meant for detecting objects that is used here to recognize faces in images or real-time videos. This method recognizes faces in real-time, and even faces on embedded devices of Raspberry Pi. Following the successful detection of faces, we load our classifier of MobileNetV2 to categorize faces with (or without) masks.

For detecting the social distancing between two individuals in a gathering, a pre-trained MobileNetV2 model has been used that is trained on the ImageNet dataset of several classes. However, in this work, the only pedestrian-labeled class has been taken for detection purposes while keeping the rest of the labels for future applications. Each detected person is indicated by a box drawn around it, and the data is subsequently used for distance measurement between people. It is assumed that the target person is walking on the same flat plan in the video frame. Next is included the feature for counting the number of people who remain rowers to social distance requirement. The proposed model for social distance and mask detection is as shown in Figure 1.

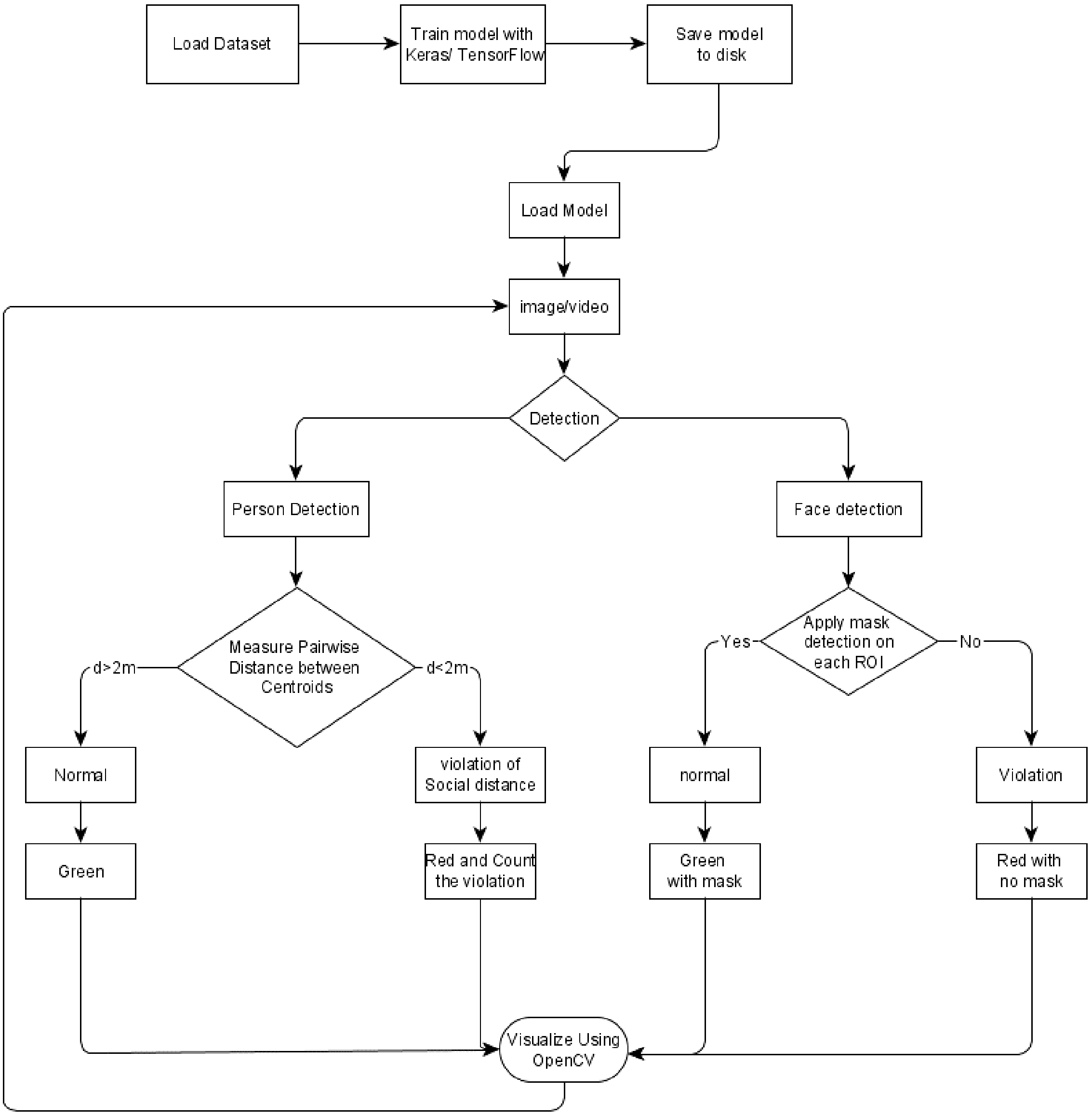

Figure 1. The proposed approach's model state diagram.

4- SPP-SSD-MobileNetV2

MobileNetV2 is a 53-layers convolution neural network architecture that is trained on the millions of images of a varying class dataset of ImageNet. MobileNetV2 significantly reduces the networks' complexity cost, memory utilization, and model size, making it appropriate for mobile devices or similar low-power processing devices while maintaining the same accuracy. MobileNetV2 is a feature extraction and classification network that shows state-of-the-art performance in detecting objects and semantic segmentation [46]. The linear bottleneck layer and inverted residual layer show improved accuracy and performance. To represent the input with reduced dimensions, we use the bottleneck layer, which has fewer nodes than the previous layers. To get the fixed information for the fully connected layers, we used the SPP layer instead of the last pooling layer. It has been assumed that manifolds of layers of interest could be embedded in the low dimensional spaces in neural networks.

4-1- Architecture of MobileNetV2

The Convolutional Neural Network's essential building element is the convolutional layer. Convolution is a mathematical term that refers to the mathematical combining of two functions to produce a third function. It operates on the basis of a sliding window mechanism that aids in the extraction of features from an image. This assists in the creation of feature maps. The output C is obtained by convolution of two functional matrices, one of which is the input image matrix A and the other is the convolutional kernel B as shown in Equation 1. The architecture of MobileNetV2 has two blocks, the inverted residual and linear bottleneck layer. Residual blocks with a skip connection connecting the end and start of the neural network that benefits the network to access the earlier activation that have not been modified in the convolution block [47]. For building a network of great depth, this approach is also essential. The inverted residual layers' hypothesis is feature mapping that can be encoded into low-dimensional sub-spaces and non-linear activation.

$$C(T) = (A * B)(x) = \int_{-\infty}^{\infty} A(T) \times B(T-x)dT \tag{1}$$

The layer takes on a low dimensional tensor which performs three separate convolutions with k-channels [48, 49]. The low dimensional input feature map is first expanded to a higher dimensional space suitable for non-linear activation using (1×1) convolution, and then ReLU6 is applied. RELU6 uses non-linearity because of its robustness when used with low precision computational. This encourages the model to learn sparse features earlier in the learning process and can be calculated using Equation 2. Then, 3×3 kernels, depth-wise convolution are conducted, followed by ReLU6 activation. Finally, the 1×1 convolution is used to spatially filter features that are projected back into a low dimensional subspace, as shown in Figure 2. The activation function should finally be linear activation to avoid losses of information.

$$y = min(max(0, x), 6) \tag{2}$$

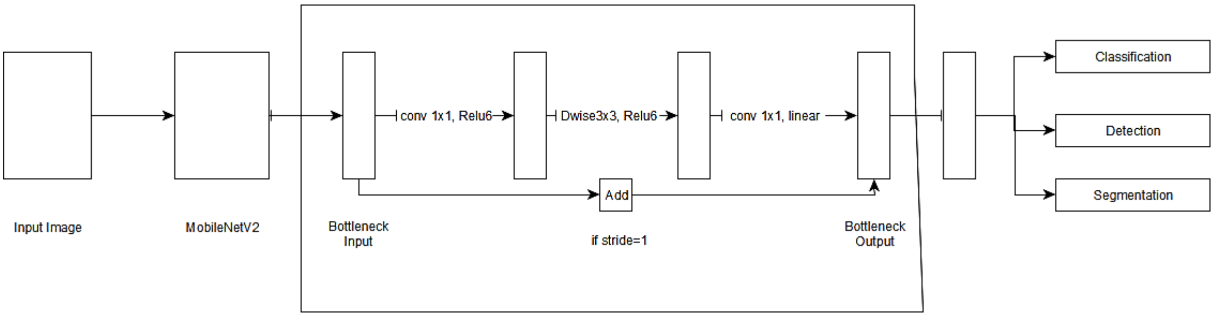

Figure 2. MobileNetV2 Architecture.

5- Dataset

There has been observed a lot of noise in recognition of masked faces and the application dataset, because of the duplicates in the photos. As an authentic dataset determines the model's accuracy, we analyzed the images and eliminated all repetitions manually through a manual cleaning process. At the same time, a wide range of essential data has been collected from different resources to make our dataset more robust. We have collected a wide range of human face images based on ethnicity, age, and gender obtained from many diverse sources. Images from the dataset and our proposed dataset are shown in Table 1.

Table 1. Different Datasets.

Datasets	Total Images
Face mask detection (FMD) by Kaggle	853
Real-Time Face Dataset (RMFD)	525
Proposed Dataset (RDO)	17000

We have acquired images from indoor and outdoor areas under different light and resolution conditions to encompass unconstrained situations and the subject diversity. In the end, we have left the data set consisting of 17000 images and we named it RDO dataset, in which 9000 images are masked while 8000 images are unmasked. So, 16000 images of both categories are used to train the model, while the remaining 2000 images are used for validation purpose. To avoid oversampling and to increase accuracy, the image augmentation technique is used. Image augmentation is done for increasing the size of the training dataset by digital systems artificially in the collection [50]. In this research, the training images are augmented by some operations such as horizontal flipping, contrasting, rotating, cropping, and zooming. Image augmentation also helps in reducing the difference between validation and training losses and accuracy. With image augmentation, we are able to train the model for more epochs without the need to address over-fitting [51].

The tool LabelMe has been used for labelling. Two mask detection classes have been considered, one with a mask and one without mask. Those who wear inappropriate masks are categorized as mask-less because it is clear that they do not intend to wear the mask. Our proposed data sets contain a lot of diversity, particularly in terms of gender, race, and head position.

6- Model Training

Deep learning techniques are commonly employed for image classification, object detection because of their better accuracy than other algorithms. However, training the deep neural networks is costly and time-consuming and require high computation and other expensive resources. To train the network faster and cost-effectively, deep learning-based Transfer Learning is evolved. Transfer learning is the technique in which a pre-trained model is reused on a new model. This technique is very efficient in deep learning as the new model gives efficient performance even on little data. It saves a lot of time and computation as instead of producing the learning process from scratches, we begin with patterns obtained from solving related tasks. There are several models trained from before such as MobileNetV2 [52], VGG16 [53], InceptionV3 [54], ResNet50 [55]. They are trained on the ImageNet dataset that is containing almost 14 million images.

In work the proposed here, the MobileNetV2 model is utilized for classifying people on the basis of the face mask. Five layers replace the penultimate layer of the MobileNetV2 model, keeping the base model weights frozen so that they cannot be changed during mask detection training, and later we use it for social distance detection. These layers consist of the average pooling layer having a pool size of 5x5, flattened layer, followed by a thickly populated layer of 128 neurons with Relu6 activation. After that, the drop out of 0.5 is added to avoid over-fitting. Next to that, Spatial Pyramid Pooling (SPP) layer is added to map any size entry down to a fixed size output. It is then followed by the decisive dense layer of 2 neurons and softmax activation function. Adam optimization has been used in our case. The categorical Cross-Entropy (CE) loss function is used which measures how good the prediction model is in terms of prediction [56]. The CE Loss can be described using Equation 3.

$$CE = -\log\left(\frac{e^{s_p}}{\sum_j^C e^{y_j}}\right) \tag{3}$$

Where Sp is the CNN score for the positive class. The categorical accuracy function is also utilized to judge the model's accuracy. A total of 25 epochs, each epoch consisting of 35 steps in each epoch as shown in Table 2, is used to train the model.

Table 2. Hyper parameters Used during Training.

Hyperparameters	Used
Loss Function	Categorical Cross-entropy
Activation Function	Softmax
Optimizer	Adam
Pooling Method	SPP
Learning Rate	0.018
Epochs	25
Bach size	35
Dropout	0.5
Weight Decay	0.0045

7- Mask Detection

The SSD layers' job is to convert pixels in the input image into features that describe the contents of the image, and then send those features to the other layers. Based on this information then classifier will decide whether the person wear mask or not. For our implementation of the SPP-SSD-MobileNetV2 model in detecting faces for the recognition of masks, first is employed the pre-trained SSD model using Tensorflow. Accordingly, a number of faces detected along with the location of their bounding boxes are obtained. Also, we got the confidence score in those predictions after implementing the face detection model. After that, the results are fed into the face mask classifier. The Deep Neural Network MobileNetV2 has then been used in the classifying task. Mobilenetv2 model then has been used to categorize each face as mask-wearing or without mask.

8- Distance Measurement

The detection architecture recognizes persons and delivers information about them in bounding boxes. The central point as the centroid is calculated after the detection of human bounding box information before the bounding box is located to be placed. After the detection of people, a box at located (x, y, w, h) is placed bounding each person accordingly. This is then transformed into a top-down view using OpenCV as shown in Figure 3. Position in top-down view is projected for each person based on the down-centered point of the bounding box. Two people in image/frame having two positional locations of (x1, y1) and (x2, y2), are considered to be at the distance (d) calculated by using Equation 4.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{4}$$

The minimum distance for the pair of people is 2 m. Now two people with distance below the minimum acceptable distance (z) are enclosed in red colored boxes while those in the safe zone are enclosed within the green colored boxes. For clearly visible people the distance in between them can be measured. In case of overlap, person detection may not good due to which distance cannot be calculated accurately. The bounding box's color threshold operation, g, can be defined as Equation 5:

$$g = \begin{cases} \text{red} & d < z \\ \text{green} & d \geq z \end{cases} \tag{5}$$

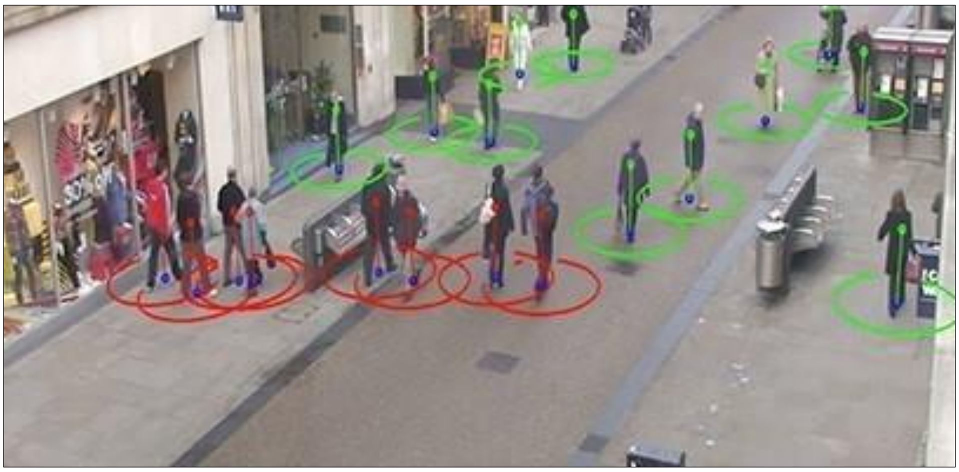

Figure 3. Top-Down Transformation.

9- Model Performance and Results

The experiment trails are conducted using visual studio, Keras/TensorFlow, OpenCV, and our trained model. OpenCV is used for display and image manipulations. Several performance parameters are used to assess the transfer learning model performance, such as Accuracy as given in Equation 6, Loss as given in Equation 7 sensitivity as given in Equation 8, intersection over union (IOA) as given in Equation 9, the mathematical equations reproduced follows.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{7}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{8}$$

$$IOU = \frac{TP}{TP+FP+FN} \tag{9}$$

Whereas TP, TN, FP, and FN stand for number of true positive, true negative, false positive, and false negative, in order mentioned. These are used as standard performance measures.

The TP, FP, TN, and FN are shown in the confusion matrix, a grid-like structure. Confusion matrices are made for evaluating the performance of the model during training and testing processes [57]. The model performance during training and testing are shown in Table 3.

Table 3. Performance of the proposed model.

Performance Parameters	Training	Testing
Accuracy	99.1%	99.3%
Precision	99.2%	99.2%
Loss	0.04%	0.02%
Intersection over union (IOU)	99.0%	99.1%

9-1- Non-max Suppression

It measures the performance of the algorithm being used for detection in measuring how much the bounding box predicted by the algorithm is different from the one actual box that needs to be around the target object [58]. In fact the bounding box is derived from the (x-y) coordinates of the image's source object. These coordinates are used to identify the image's source object [59]. The primary bounding box for an object in an image is thus hand-labelled and referred to as the Primary Boundary Box. The Predicted Boundary Box is the bounding box that the Deep Learning model places around the object predicted. In fact, the model's projected bounding box is extremely unlikely to be an exact primary bounding box [60]. As a result, we can use the metric Intersection Over Union (IOU) to determine how accurately the object has been identified in the Image/Frame.

9-2- Intersection over Union

Intersection over Union (IoU) is a well-known statistic for determining how much overlap exists between the ground truth bounding box and predicted bounding box. In actuality, the (x, y)-coordinates of our projected bounding box are exceedingly unlikely to match the (x, y)-coordinates of the ground-truth bounding box exactly. As a result, we'll need to develop a threshold value that rewards predicted bounding boxes that overlap strongly with the ground truth. In my case, the threshold value is 0.5 so, an Intersection over Union score of greater than 0.5 is regarded as a "good" prediction. It discards the rest of the bounding boxes with a threshold value less than 0.5 and chooses the one with the highest value at the end. In conclusion, the larger the IOU value, the more accurate the prediction is made. Mathematically it can be calculated using Equation 10.

$$\text{IOU} = \frac{\textit{Size of intersection}}{\textit{size of the union}} \tag{10}$$

When you look at Equation 8, you will note that Intersection Over Union is just a ratio. We compute the overlap region between the expected box and the ground-truth bounding box in the numerator. The field of union, or, to put it another way, the area covered by both the expected and ground-truth bounding boxes, is the numerator. Our final score is calculated by dividing the region of overlap by the area of union. We have developed a model that helps the relevant authorities to monitor the people regarding wearing masks. The accuracy of our model has reached 99% during training, and during testing, it provides 99% accurate results. Furthermore, the loss is thus significantly reduced to 0.04 and 0.02 during training and testing, respectively, as shown in Figure 4. We used matplotlib's pyplot command style function to plot these figures, making it work like MATLAB.

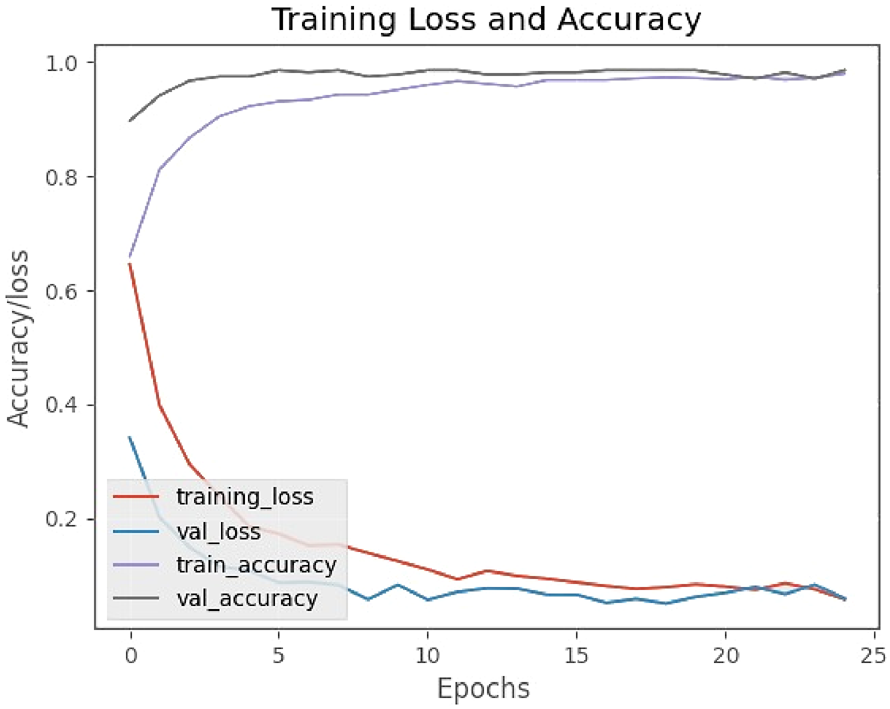

Figure 4. Model Performance chart.

For some random samples, the model shows a green bounding box around the face that indicates the person putting a mask on and a red bounding box around the face that indicates the individual wearing no mask. These results are displayed with a confidence score, as shown in Figure 5.

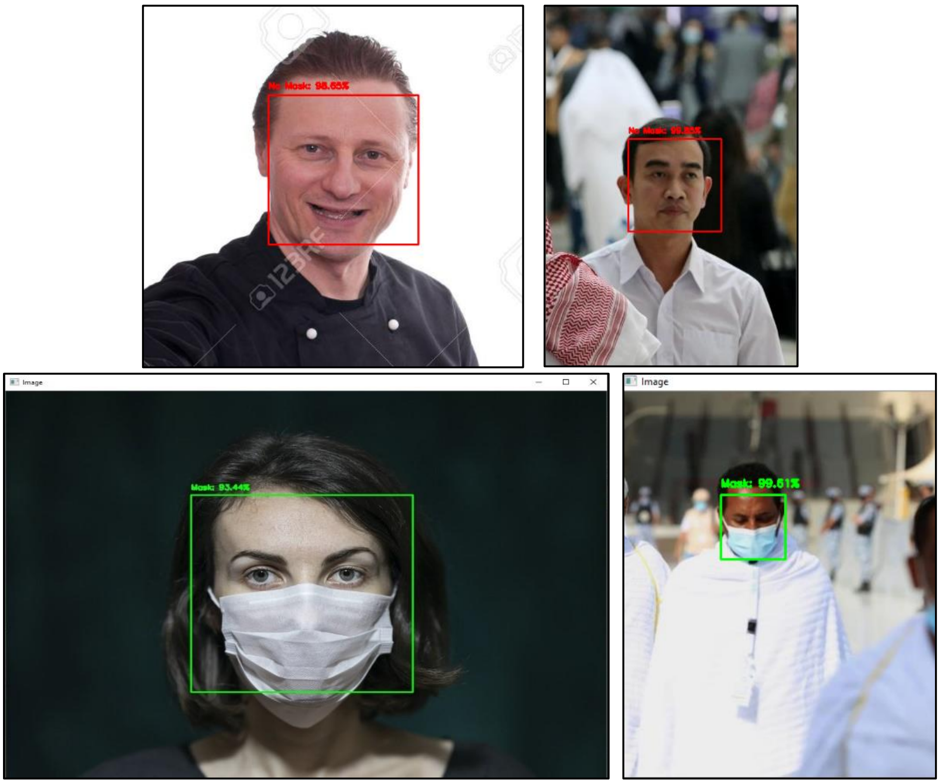

Figure 5. **Result of the proposed model for mask detection.**

The model recognizes people in a variety of scenic situations, as evidenced by the sample images. People are efficiently identified using a variety of features. However, the social distance between them is calculated. A red bounding box appeared around those closest to each other than the allowed safe distance, while a green box appeared around those following the social distancing SOP ideally and was considered to be in the safe zone. Also, a counter is activated that counts the number of persons who violate social distancing SOP, as shown in Figure 6. The framework is successful in detecting crowds accurately for the implementation of social distancing SOP.

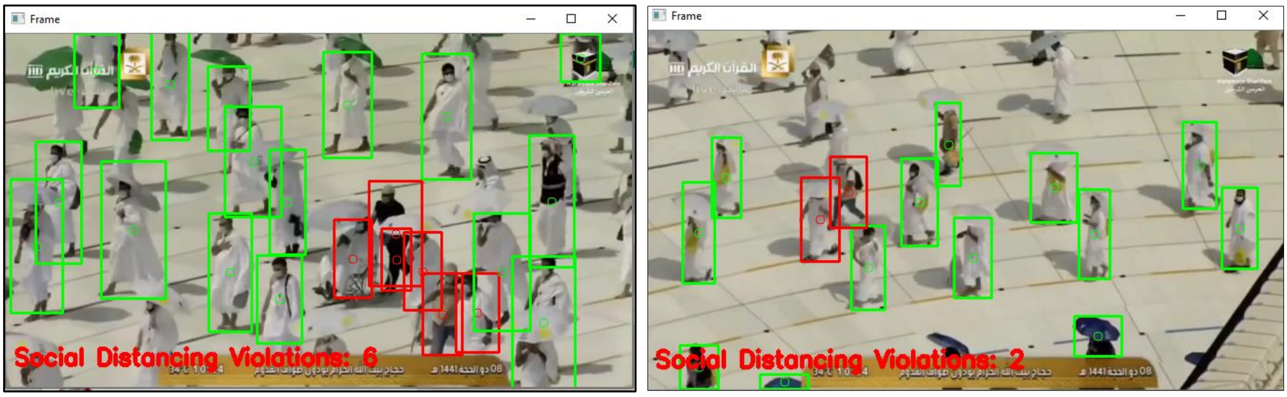
Frame
Social Distancing Violations: 6
Frame
Social Distancing Violations: 2

Figure 6. **Result of the proposed model for Social Distancing Detection.**

In this work are used four different videos to test each of the two sub-tasks to assess the overall performance of the proposed model. Table 4 shows the corresponding results of the experiments. The second column represents the number of frames, while the third column represents the frames per second. The last two columns indicate the classification accuracy for mask detection and social distance detection. Very high accuracy is achieved in detecting whether people maintain a social distance and whether or not a person is wearing a mask.

Table 4. **Evaluation of the proposed model on the test videos.**

Video	No of frames	FPS	Mask acc.	Distance acc
Video 1	237	14.43	99.3%	99.56%
Video 2	315	14.46	99.01%	99.48%
Video 3	297	14.44	99.2%	99.44%
Video 4	319	14.45	99.38%	99.58%

10- Models Comparison

Based on accuracy and detection of SOPs, the proposed model is compared with Faster-RCNN [61] ResNet50 [62], VGG16 [53], and InceptionV3 [14]. Table 5 shows that the proposed model has achieved the highest accuracy and can perform mask detection and social distance detection. Several state-of-the-art detecting models for mask detection and social distance detection, but none of them can detect both masks and social distances. Faster-RCNN was used only for social distance detection, with a 96% accuracy, while VGG16, ResNet50, and inceptionV3 were only used for mask detection, with 96.2, 97, and 98.3% accuracy, respectively. They are computationally costly and provide an average FPS of 2.89 in testing, making them difficult to deploy in real-time. Keeping this in mind, we created a model that can detect both masks and social distances with better accuracy. We make it lightweight so that it runs smoothly on low computational devices.

Table 5. Accuracy and Performance Comparison.

Models	Mask Detection	Social Distance Detection	Accuracy (%)	FPS (Average)
Faster-RCNN [61]	×	✓	96	8.2
ResNet50 [62]	✓	×	96.2	6.4
VGG-16 [53]	✓	×	97	3.08
InceptionV3[14]	✓	×	98.3	9
Proposed Model	✓	✓	99.1	14.46

11- Factors Affecting Performance

The process of face recognition from pictures and videos is a difficult task. Even though much study has been done to achieve 99% accuracy, we are still not getting satisfactory results due to this system's different factors. The factors that impair the accuracy of face recognition systems have been discovered.

11-1- Face Direction

Face detection performance dramatically relies on the direction of faces. Face detection performance accuracy is enhanced when the target people are looking towards the camera so that their eyes become observable, but performance decrease when the detected objects are from sides, as shown in Figure 7. However, our model also performs well even on pictures or videos detected from sides.

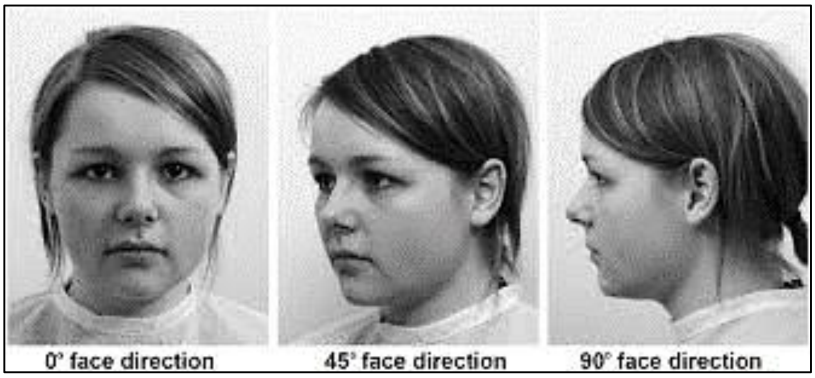

Figure 7. Different directions of faces.

11-2- Face Size

Face detection performance also changes when the size of the face changes. Smaller faces are likely to be detected poorly compared to more prominent ones, especially in low-resolution images or compressed images. Our model is trained for producing satisfactory results both with low-resolution images and videos and on smaller-size faces.

11-3- Media Quality

Face detection in low-resolution images is not always accurate. Low-quality images include the following categories.

- ***Blurry image:***

  The significant factors for blurry images include: 1) a camera with an out-of-focus lens, 2) interlacing and interleaving, 3) camera moving relative to object, and 4) turbulence in the vicinity, and other factors related to camera or users of the camera.

- ***Low resolution (LR):***

  Low resolution (LR) is caused by a wide camera standoff gap and a low spatial resolution camera sensor.

- ***Artifacts:***

  Low compression settings, motion between interlaced scans, and other factors can cause artifacts.

- ***Acquisition:***

  Acquisition conditions that cause image noise, for example, when the illumination level is low.

Face detection dramatically depends on the quality of images or videos. Its performance is affected when the photo or video is blurry, and the face should not focus, as shown in Figure 8. As a result, we use some low-quality images in our phase of training by images to ensure that our model performs well even on low-resolution images.

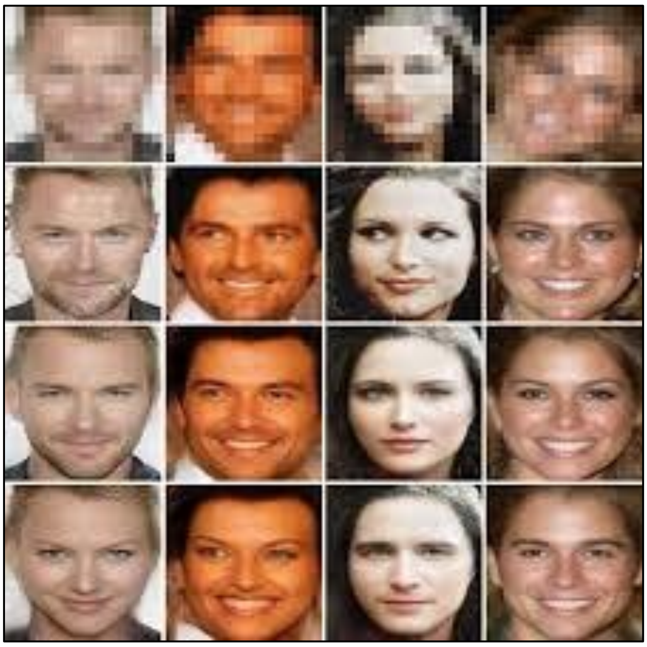

Figure 8. Low-resolution images.

11-4- Camera Calibration

Camera calibration aims to figure out the image forming process' geometric parameters. Many computer vision applications require this step, particularly when metric information about the scene is needed. For both mask detector and distance measurement, proper camera calibration is essential. For better accuracy, it is required that camera calibration must be appropriate

12- Conclusion

The world is in an alarming situation due to the COVID-19 pandemic. Every nation is trying its level best to stop the virus from spreading. Crowded places provide the main spots for rapid transmission of the virus. Wearing a mask and keeping a safe distance from each other in public and crowded areas is the only way to control virus transmission and ensure other people's safety from this deadly disease. As per WHO guidelines, infected people are advised to stay enclosed and away from others. For healthy people attending public places wearing masks in public places is mandatory and keeping social distancing is compulsory in many countries, and some countries have severe penalties for their violations. So, monitoring people through human resources is very tough. We have proposed a model that can detect the person without the mask and those who are not following the social distancing SOP by keeping this in mind. The model presents a count of those in a crowd violating these requirements. The model has used in this work is SPP-SSD-MobileNetV2, which gives the best results on low computation and saves much of our time and memory space. The image augmentation techniques have been used to train the model very well, thus enabling an increase in the diversity of the dataset being used in training. The model has achieved an accuracy of 99% during testing and an accuracy of 99% during training. The loss is also reduced to a very minimum during both testing and training. The model is doing well for real-time use applications as it has the same accuracy on video streams. There is always room for improvement, so the exact model can also be used for advanced detection of facial part detection and facial landmarks process.

13- Declarations

13-1- Author Contributions

Conceptualization, M.H.K.K. and S.K.; methodology, M.H.K.K., S.K., K.K. and W.A.; software, M.I. and S.H.; validation, K.K., M.I. and W.A.; formal analysis, S.K, and W.A.; investigation, M.H.K.K. and S.K.; resources, K.K. and W.A.; data curation, M.K.K.K., W.A., and K.K.; writing—original draft preparation, M.H.K.K., S.K., and K.K.; writing—review and editing, W.A., S.K.; visualization, M.H.K.K.; supervision, K.K., S. K.; project administration, W.A. and K.K.; funding acquisition, K.K. All authors have read and agreed to the published version of the manuscript.

13-2- Data Availability Statement

Publicly available datasets were analyzed in this study. This data can be found here: [https://www.kaggle.com/andrewmvd/face-mask-detection] and [https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset].

13-3- Funding

This research work has been funded for researcher by the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia. This publication is funded by the Universiti Kuala Lumpur Journal Publication Grant.

13-4- Acknowledgements

The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia for funding the researcher on this through Project No. QURDO001 titled "Intelligent Real-time Crowd Monitoring System Using Unmanned Aerial Vehicle (UAV) Video and Global Positioning Systems (GPS) Data"

13-5- Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

14- References

[1] Martella, C., J. Li, C. Conrado, and A. Vermeeren. “On Current Crowd Management Practices and the Need for Increased Situation Awareness, Prediction, and Intervention.” Safety Science 91 (2017): 381–93. doi:10.1016/j.ssci.2016.09.006.

[2] Filingeri, Victoria, Ken Eason, Patrick Waterson, and Roger Haslam. “Factors Influencing Experience in Crowds – The Participant Perspective.” Applied Ergonomics 59 (2017): 431–41. doi:10.1016/j.apergo.2016.09.009.

[3] Zhao, Hantao, Tyler Thrash, Mubbasir Kapadia, Katja Wolff, Christoph Holscher, Dirk Helbing, and Victor R. Schinazi. “Assessing Crowd Management Strategies for the 2010 Love Parade Disaster Using Computer Simulations and Virtual Reality.” Journal of the Royal Society Interface 17, no. 167 (2020): 20200116. doi:10.1098/rsif.2020.0116.

[4] Toyokawa, Wataru, Andrew Whalen, and Kevin N. Laland. “Social Learning Strategies Regulate the Wisdom and Madness of Interactive Crowds.” Nature Human Behaviour 3, no. 2 (2019): 183–93. doi:10.1038/s41562-018-0518-x.

[5] Cai, Qingxian, Minghui Yang, Dongjing Liu, Jun Chen, Dan Shu, Junxia Xia, Xuejiao Liao, et al. “Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study.” Engineering 6, no. 10 (2020): 1192–98. doi:10.1016/j.eng.2020.03.007.

[6] Betsch, Cornelia, Lars Korn, Philipp Sprengholz, Lisa Felgendreff, Sarah Eitze, Philipp Schmid, and Robert Böhm. “Social and Behavioral Consequences of Mask Policies during the COVID-19 Pandemic.” Proceedings of the National Academy of Sciences of the United States of America 117, no. 36 (2020): 21851–53. doi:10.1073/pnas.2011674117.

[7] Eikenberry, Steffen E., Marina Mancuso, Enahoro Iboi, Tin Phan, Keenan Eikenberry, Yang Kuang, Eric Kostelich, and Abba B. Gumel. “To Mask or Not to Mask: Modeling the Potential for Face Mask Use by the General Public to Curtail the COVID-19 Pandemic.” Infectious Disease Modelling 5 (2020): 293–308. doi:10.1016/j.idm.2020.04.001.

[8] Meyer, Jacob, Cillian McDowell, Jeni Lansing, Cassandra Brower, Lee Smith, Mark Tully, and Matthew Herring. “Changes in Physical Activity and Sedentary Behavior in Response to Covid-19 and Their Associations with Mental Health in 3052 Us Adults.” International Journal of Environmental Research and Public Health 17, no. 18 (2020): 1–13. doi:10.3390/ijerph17186469.

[9] Hsieh, Chih Chia, Chih Hao Lin, William Yu Chung Wang, David J. Pauleen, and Jengchung Victor Chen. “The Outcome and Implications of Public Precautionary Measures in Taiwan–Declining Respiratory Disease Cases in the COVID-19 Pandemic.” International Journal of Environmental Research and Public Health 17, no. 13 (2020): 1–10. doi:10.3390/ijerph17134877.

[10] Ul Haq, Shamsheer, Pomi Shahbaz, and Ismet Boz. “Knowledge, Behavior and Precautionary Measures Related to COVID-19 Pandemic among the General Public of Punjab Province, Pakistan.” Journal of Infection in Developing Countries 14, no. 8 (2020): 823–35. doi:10.3855/jidc.12851.

[11] Khan, Akbar, Jawad Ali Shah, Kushsairy Kadir, Waleed Albattah, and Faizullah Khan. “Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review.” Applied Sciences (Switzerland) 10, no. 14 (2020): 4781. doi:10.3390/app10144781.

[12] Loey, Mohamed, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and Nour Eldeen M. Khalifa. “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 65 (2021): 102600. doi:10.1016/j.scs.2020.102600.

[13] Meenpal, Toshanlal, Ashutosh Balakrishnan, and Amit Verma. “Facial Mask Detection Using Semantic Segmentation.” 2019 4th International Conference on Computing, Communications and Security, ICCCS 2019, 2019. doi:10.1109/CCCS.2019.8888092.

[14] Jignesh Chowdary, G., Narinder S. Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal. “Face Mask Detection Using Transfer Learning of InceptionV3.” In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12581 LNCS (2020):81–90. Cham: Springer. doi:10.1007/978-3-030-66665-1_6.

[15] Singh, Sunil, Umang Ahuja, Munish Kumar, Krishan Kumar, and Monika Sachdeva. “Face Mask Detection Using YOLOv3 and Faster R-CNN Models: COVID-19 Environment.” Multimedia Tools and Applications 80, no. 13 (2021): 19753–68. doi:10.1007/s11042-021-10711-8.

[16] Chavda, Amit, Jason Dsouza, Sumeet Badgujar, and Ankit Damani. “Multi-Stage CNN Architecture for Face Mask Detection.” In 2021 6th International Conference for Convergence in Technology, I2CT 2021, 1–8. IEEE, 2021. doi:10.1109/I2CT51068.2021.9418207.

[17] Albattah, Waleed, Muhammad Haris Kaka Khel, Shabana Habib, Muhammad Islam, Sheroz Khan, and Kushsairy Abdul Kadir. “Hajj Crowd Management Using CNN-Based Approach.” Computers, Materials and Continua, 2020. doi:10.32604/cmc.2020.014227.

[18] Gu, Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, et al. “Recent Advances in Convolutional Neural Networks.” Pattern Recognition 77 (May 2018): 354–377. doi:10.1016/j.patcog.2017.10.013.

[19] Yamashita, Rikiya, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. “Convolutional Neural Networks: An Overview and Application in Radiology.” Insights into Imaging 9, no. 4 (2018): 611–29. doi:10.1007/s13244-018-0639-9.

[20] Hamouda, Maissa, and Med Salim Bouhlel. “Modified Convolutional Neural Networks Architecture for Hyperspectral Image Classification (Extra-Convolutional Neural Networks).” IET Image Processing, 2021. doi:10.1049/ipr2.12169.

[21] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks.” Communications of the ACM 60, no. 6 (2017): 84–90. doi:10.1145/3065386.

[22] Zhao, Zhong Qiu, Peng Zheng, Shou Tao Xu, and Xindong Wu. “Object Detection with Deep Learning: A Review.” IEEE Transactions on Neural Networks and Learning Systems 30, no. 11 (2019): 3212–32. doi:10.1109/TNNLS.2018.2876865.

[23] Uijlings, J. R.R., K. E.A. Van De Sande, T. Gevers, and A. W.M. Smeulders. “Selective Search for Object Recognition.” International Journal of Computer Vision 104, no. 2 (2013): 154–71. doi:10.1007/s11263-013-0620-5.

[24] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng Yang Fu, and Alexander C. Berg. “SSD: Single Shot Multibox Detector.” In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9905 LNCS (2016):21–37. Cham: Springer. doi:10.1007/978-3-319-46448-0_2.

[25] Lin, Tsung Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. “Focal Loss for Dense Object Detection.” In IEEE Transactions on Pattern Analysis and Machine Intelligence, 42:318–27, 2020. doi:10.1109/TPAMI.2018.2858826.

[26] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. “EfficientDet: Scalable and Efficient Object Detection.” In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 10778–87, 2020. doi:10.1109/CVPR42600.2020.01079.

[27] Zhang, Shifeng, Longyin Wen, Zhen Lei, and Stan Z. Li. “RefineDet++: Single-Shot Refinement Neural Network for Object Detection.” IEEE Transactions on Circuits and Systems for Video Technology 31, no. 2 (2021): 674–87. doi:10.1109/TCSVT.2020.2986402.

[28] Zhu, Chenchen, Yutong Zheng, Khoa Luu, and Marios Savvides. “CMS-RCNN: Contextual Multi-Scale Region-Based CNN for Unconstrained Face Detection.” In Advances in Computer Vision and Pattern Recognition, PartF1:57–79. Cham: Springer, 2017. doi:10.1007/978-3-319-61657-5_3.

[29] Ejaz, Md Sabbir, Md Rabiul Islam, Md Sifatullah, and Ananya Sarker. “Implementation of Principal Component Analysis on Masked and Non-Masked Face Recognition.” In 1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019, 1–5. IEEE, 2019. doi:10.1109/ICASERT.2019.8934543.

[30] Loey, Mohamed, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and 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.” Measurement: Journal of the International Measurement Confederation 167 (2021): 108288. doi:10.1016/j.measurement.2020.108288.

[31] Wang, Zhongyuan, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, et al. “Masked Face Recognition Dataset and Application,” 2020. http://arxiv.org/abs/2003.09093.

[32] Cabani, Adnane, Karim Hammoudi, Halim Benhabiles, and Mahmoud Melkemi. “MaskedFace-Net – A Dataset of Correctly/Incorrectly Masked Face Images in the Context of COVID-19.” Smart Health 19 (2021): 100144. doi:10.1016/j.smhl.2020.100144.

[33] Li, Chong, Rong Wang, Jinze Li, and Linyu Fei. “Face Detection Based on YOLOv3.” In Advances in Intelligent Systems and Computing, 1031 AISC:277–84. Singapore: Springer, 2020. doi:10.1007/978-981-13-9406-5_34.

[34] Zhao, Haipeng, Yang Zhou, Long Zhang, Yangzhao Peng, Xiaofei Hu, Haojie Peng, and Xinyue Cai. “Mixed YOLOv3-LITE: A Lightweight Real-Time Object Detection Method.” Sensors (Switzerland) 20, no. 7 (2020). doi:10.3390/s20071861.

[35] Yang, Yang, and Hongmin Deng. “Gc-Yolov3: You Only Look Once with Global Context Block.” Electronics (Switzerland) 9, no. 8 (2020): 1–14. doi:10.3390/electronics9081235.

[36] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, Sonali Agarwal, and Gaurav Rai. “Monitoring COVID-19 Social Distancing with Person Detection and Tracking via Fine-Tuned YOLO v3 and Deepsort Techniques,” 2020. http://arxiv.org/abs/2005.01385.

[37] Ma, Tao, Fen Wang, Jianjun Cheng, Yang Yu, and Xiaoyun Chen. “A Hybrid Spectral Clustering and Deep Neural Network Ensemble Algorithm for Intrusion Detection in Sensor Networks.” Sensors (Switzerland) 16, no. 10 (2016): 1701. doi:10.3390/s16101701.

[38] Ainslie, Kylie E. C., Caroline E. Walters, Han Fu, Sangeeta Bhatia, Haowei Wang, Xiaoyue Xi, Marc Baguelin, et al. “Evidence of Initial Success for China Exiting COVID-19 Social Distancing Policy after Achieving Containment.” Wellcome Open Research 5 (October 1, 2020): 81. doi:10.12688/wellcomeopenres.15843.2.

[39] Nguyen, Cong T., Yuris Mulya Saputra, Nguyen Van Huynh, Ngoc-Tan Nguyen, Tran Viet Khoa, Bui Minh Tuan, Diep N. Nguyen, et al. “A Comprehensive Survey of Enabling and Emerging Technologies for Social Distancing—Part I: Fundamentals and Enabling Technologies.” IEEE Access 8 (2020): 153479–153507. doi:10.1109/access.2020.3018140..

[40] Prem, Kiesha, Yang Liu, Timothy W. Russell, Adam J. Kucharski, Rosalind M. Eggo, Nicholas Davies, Stefan Flasche, et al. “The Effect of Control Strategies to Reduce Social Mixing on Outcomes of the COVID-19 Epidemic in Wuhan, China: A Modelling Study.” The Lancet Public Health 5, no. 5 (2020): e261–70. doi:10.1016/S2468-2667(20)30073-6.

[41] Pouw, Caspar A.S., Federico Toschi, Frank van Schadewijk, and Alessandro Corbetta. “Monitoring Physical Distancing for Crowd Management: Real-Time Trajectory and Group Analysis.” PLoS ONE 15, no. 10 October (2020): 240963. doi:10.1371/journal.pone.0240963.

[42] Ahmad, Misbah, Imran Ahmed, Fakhri Alam Khan, Fawad Qayum, and Hanan Aljuaid. “Convolutional Neural Network–Based Person Tracking Using Overhead Views.” International Journal of Distributed Sensor Networks 16, no. 6 (2020): 1550147720934738. doi:10.1177/1550147720934738.

[43] Jin, Tingxu, Jun Li, Jun Yang, Jiawei Li, Feng Hong, Hai Long, Qihong Deng, et al. “SARS-CoV-2 Presented in the Air of an Intensive Care Unit (ICU).” Sustainable Cities and Society 65 (2021): 102446. doi:10.1016/j.scs.2020.102446.

[44] Bhattacharya, Sweta, Praveen Kumar Reddy Maddikunta, Quoc Viet Pham, Thippa Reddy Gadekallu, Siva Rama Krishnan S, Chiranji Lal Chowdhary, Mamoun Alazab, and Md Jalil Piran. “Deep Learning and Medical Image Processing for Coronavirus (COVID-19) Pandemic: A Survey.” Sustainable Cities and Society 65 (2021): 102589. doi:10.1016/j.scs.2020.102589.

[45] Anisimov, Dmitriy, and Tatiana Khanova. “Towards Lightweight Convolutional Neural Networks for Object Detection.” In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017, 1–8. IEEE, 2017. doi:10.1109/AVSS.2017.8078500.

[46] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2017. http://arxiv.org/abs/1704.04861.

[47] Sanjaya, Samuel Ady, and Suryo Adi Rakhmawan. “Face Mask Detection Using MobileNetV2 in the Era of COVID-19 Pandemic.” In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020, 1–5. IEEE, 2020. doi:10.1109/ICDABI51230.2020.9325631.

[48] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang Chieh Chen. “MobileNetV2: Inverted Residuals and Linear Bottlenecks.” In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 4510–20, 2018. doi:10.1109/CVPR.2018.00474.

[49] Howard, Andrew, Mark Sandler, Bo Chen, Weijun Wang, Liang Chieh Chen, Mingxing Tan, Grace Chu, et al. “Searching for MobileNetV3.” In Proceedings of the IEEE International Conference on Computer Vision, 2019-October:1314–24, 2019. doi:10.1109/ICCV.2019.00140.

[50] Buslaev, Alexander, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. “Albumentations: Fast and Flexible Image Augmentations.” Information (Switzerland) 11, no. 2 (2020): 125. doi:10.3390/info11020125.

[51] Saleh, Abeer M., and Talal H. “Analysis and Best Parameters Selection for Person Recognition Based on Gait Model Using CNN Algorithm and Image Augmentation.” Journal of Big Data 8, no. 1 (2021): 1–20. doi:10.1186/s40537-020-00387-6.

[52] Fu, Xiaomeng, and Huiming Qu. “Research on Semantic Segmentation of High-Resolution Remote Sensing Image Based on Full Convolutional Neural Network.” 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE) (December 2018). doi:10.1109/isape.2018.8634106.

[53] Simonyan, Karen, and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.

[54] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. “Rethinking the Inception Architecture for Computer Vision.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016). doi:10.1109/cvpr.2016.308.

[55] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition.” In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December: 770–78, 2016. doi:10.1109/CVPR.2016.90.

[56] Inamdar, Madhura, and Ninad Mehendale. “Real-Time Face Mask Identification Using Facemasknet Deep Learning Network.” SSRN Electronic Journal 3663305 (2020). doi:10.2139/ssrn.3663305.

[57] Xu, Jianfeng, Yuanjian Zhang, and Duoqian Miao. “Three-Way Confusion Matrix for Classification: A Measure Driven View.” Information Sciences 507 (2020): 772–94. doi:10.1016/j.ins.2019.06.064.

[58] Zhang, Xiaosong, Fang Wan, Chang Liu, Xiangyang Ji, and Qixiang Ye. “Learning to Match Anchors for Visual Object Detection.” IEEE Transactions on Pattern Analysis and Machine Intelligence (2021): 1–1. doi:10.1109/tpami.2021.3050494.

[59] Rezatofighi, Hamid, Nathan Tsoi, Junyoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. “Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression.” In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2019-June):658–66, 2019. doi:10.1109/CVPR.2019.00075.

[60] Hou, Feifei, Wentai Lei, Shuai Li, Jingchun Xi, Mengdi Xu, and Jiabin Luo. “Improved Mask R-CNN with Distance Guided Intersection over Union for GPR Signature Detection and Segmentation.” Automation in Construction 121 (2021): 103414. doi:10.1016/j.autcon.2020.103414.

[61] Ahmed, Imran, Misbah Ahmad, and Gwanggil Jeon. “Social Distance Monitoring Framework Using Deep Learning Architecture to Control Infection Transmission of COVID-19 Pandemic.” Sustainable Cities and Society 69 (2021): 102777. doi:10.1016/j.scs.2021.102777.

[62] Loey, Mohamed, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and 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.” Measurement 167 (January 2021): 108288. doi:10.1016/j.measurement.2020.108288.