



Social Distance and Face Mask Detection

Fahad Hassan, Soumya Roy, Mridula Prabhakar and
Himanshu Rawat

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 13, 2021

Social Distance and Face Mask Detection

ABSTRACT

With the recent outbreak and rapid transmission of the Covid-19 pandemic, the need for the public to comply with social distancing standards, and to wear masks in public places, is only growing. According to the World Health Organization (WHO), in order to comply with the necessary social distancing, people should maintain a distance of at least 3 feet from one another. This research paper focuses on a solution that will help you to secure the relevant social distance, by the use of the YOLO (You Only Look Once) object detection on in the video footages and images in real-time. The experimental results presented in the paper show that the detection of masked faces, faces and the human subjects, which is based on the YOLO, has a faster, safer and more secure, and better recognition rate when compared with its competitors. Our proposed object detection models have reached a high level of accuracy in excess of 90% with a very good output speed. The network is faster, and also makes sure that the output speed is capable of producing real-time results without compromising on accuracy. The model also provides promising results in variable scenarios.

INTRODUCTION

We are using You Only Look Once YOLO (You Only Look Once) algorithm for social distancing detection between people. YOLO algorithm is an object detection method and it is a neural network approach of object detection. Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class such as humans in digital images and videos. We have seen that there is a necessity for development of a model, particularly suitable for the detection of certain types of social distancing violations, in real-time. The first application is that our approach will be to actually discover a person's face in order to determine if they have a mask on their face or not. The second is to determine whether the social distance between two people, is being maintained or not. All of this is done in the most efficient, accurate, and simple ways to ensure minimal effort on the part of the concerned authorities. In order to implement the method(s) described above, we can use CCTV cameras to precisely differentiate between 3 sub-groups: the masked faces, unmasked faces (these two classes to be used for face mask detection) and the human subjects in general (this is to be used for social distancing detection between any two or more individuals. There have been other models that have tried to select the masked and the unmasked faces of people. Such models have favored object detection

networks like single-shot detector method to train their data from the dataset on. But such models were found to be ineffective to help the community to deal with the potential risks in real-time.

This research has been done keeping in mind the gravity of the current scenario and analysis has been done accordingly to come up with solutions to detect the violations. The solutions of the research include:

- Collecting the data from various data sets, and self -annotations on the photos, in order to test complex scenarios, whether the masks are worn, as well as the creation of the measured test sets, videos, pictures on social distancing.
- Perform face recognition autonomously, by training the model with the help of the user-created dataset with high built data-set comprising of a mixture of MAFA and WIDER-FACE data-sets to determine the accuracy with which masked faces can be detected.
- Development of an original method with minimum and user-friendly calibration so as to ensure that people are maintaining the recommended social distance or not.

Study of several object detection methods that give maximum accuracy and FPS, so that the model has applications in real-time usage.

Types of Methods

1. Regression based object Detectors(YOLO):

Detectors for pressure-based objects such as You Only Look Once (YOLO) and Single Shot Detector (SSD) multibox proved to be much faster than regional object detectors. Between the two, YOLO has long been the most popular choice among other items to find the same item. It takes to insert an entire image at once, unlike a region machines that reduce regional proposals provided by to separate. This makes it faster than other detectors in general images. Model pipeline awaits RGB image divided into $S \times S$ grid cells. Each cell grid is facing predicting boxes binding B. For each binding box 5 predicted values x, y, w, h and c . The coordinates of the centre point of the connecting box associated with the grid

cell x , y width and height of the bounding box w and h . Self-confidence score is available at the bounding box is c . In the case of class C , the effect of object detector is a tensor of shape:

$$(S \times S \times (B \times 5 + C)).$$

2. Detecting Masked Face in the Wild using LLE CNNs

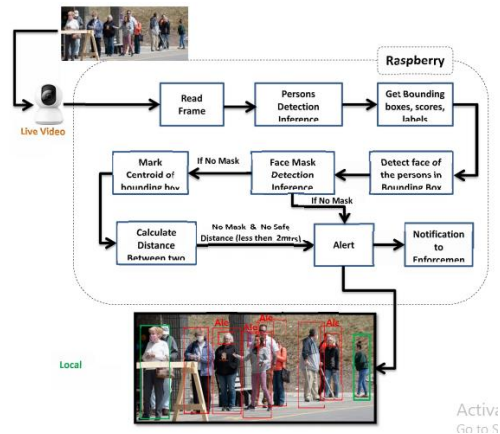
In, the authors not only proposed the MAFA database, however use of Locally Linear Embedding (LLE) CNNs. The model divides the face as it is covered by nature such as hands and other like various face masks types. This was an example of a blast as we hit everything else models with an Average Precision (AP) 76.4% in the MAFA exam is set to face detection.

3. YOLOv3 and Deepsort to track individuals in Surveillance footage

Here, the authors used the YOLOv3 object detection and the Deepsort object tracking algorithm people in watch videos. Individuals in the area(x , y) is why a 3-Dimensional feature space map(x , y , d), where d is the apparent depth of the person with its camera indicator. The standard for L2 is calculated in pairs_of people. Limit of intimacy between two people and then strongly updated according to the location of the person with a given width of pixels. Limit seen here that the range is set in pixels (90, 170), which means that there is no measurement based on camera stand.

Methodology

The proposed system helps to ensure the safety of people in public places by automatically monitoring their presence maintain a safe public space, and also by seeing if anyone or anyone is wearing a mask. This section explains it briefly the construction of a solution and how the proposed system will work automatically in a default way to prevent it the spread of coronavirus.



The proposed system uses a transfer learning method using an in-depth learning algorithm with a computer the idea of automatically monitoring people in public places with a camera and finding people with a mask or no mask. We also make good order, which is another way to learn to transfer, much more powerful than feature removal.

In this process video camera feeds from Network Video Recorder (NVR) are broadcast using RTSP and then these frames converted to grayscale to improve speed and accuracy. We use MobileNetV2 architecture as a key acquisition model as MobileNetV2 offers significant cost savings compared to in the standard 2D CNN model. This process includes the SSD MultiBox Detector, an existing neural network structure trained in a large collection of images such as ImageNet and PascalVOC with high quality image classification.

We download MobileNet V2 with a pre-weighted image of ImageNet, we leave the network upside down and build a new FC head, attaching it to the base instead of the old head, and freezing the basic layers of the network. The weights of these basic layers it will not be changed during the positive phase of the retrospective distribution, while the head layer components will be adjusted. After data is modified and model construction is set to good adjustment, then the model is assembled and trained. Very low level of reading used during structural retraining to ensure that the learned convolutional filters do not deviate too much and the test was done with OpenCV, TensorFlow using Deep Learning and Computer Vision to test a safe social distance between found people and face recognition in real-time video streaming. Great contribution of the proposed system is made up of three elements: human detection, measuring the safe distance between found people, facial detection.

Real-life detection is done with the help of Single Shot object Detection (SSD) using MobileNet V2 and OpenCV, access 91.2% mAP, surpasses the fastest art model compared to Faster R-CNN. A mandatory box will be displayed around everyone

found. While an SSD can get a lot of frames, it is limited to getting one person in this system. To calculate the distance between two people first a person's distance from the camera is calculated using a triangular similarity method, we calculate the camera's length, take a person's D distance from the camera and personal height $H = 165\text{cms}$ and pixel detection.

Using these values, the focal length of the camera can be calculated using the formula below:

$$F = (P \times D) / H$$

Then we use the real person's height H , the person's pixel height P , and the camera's focal length F to measure the person's distance from the camera. The distance from the camera can be determined using the following:

$$D1 = (H \times F) / P$$

After calculating the depth of the person in the camera, we calculate the distance between two people in the video. A number of people can be detected in a video. Thus, the Euclidean distance is measured between the mid-points of the bounding boxes of all detected individuals. By doing this, we got x and y values, and these pixel values are converted into centimetres. We have the x , y and z (the person's distance from the camera) coordinates for each person in cms. The Euclidean distance between each person detected is calculated using (x, y, z) coordinates. If the distance between two people is less than 2 meters or 200 centimetres, a red bounding box is shown around them, indicating that they do not maintain a social distance.

In the proposed system transfer system is used in addition to the highly trained SSD face detection model with mobileNet V2 expertise as the backbone to create a lightweight and efficient computer model. We have used custom face crop data sets for approximately 3165 images defined in the mask and no masks.

Annotated images are used to train the in-depth binary separation classification model that separates the input image from the mask and no mask categories use the output phase confidence. The effect of the SSD model removes the human mask and displays the connecting box.

The proposed system continuously monitors public areas and when a person without a mask is found his or her face is captured and a warning is sent to the authorities with a face mask and at the same time the distance between people is measured in real time, if there are 20 people identified continuously. a warning has been sent to the control centre at the State Police Headquarters for further action. This system can be used for real-time applications that require safe public-distance monitoring and access

to face masks for the purpose of the Covid-19 outbreak. Placing our model on the edges of automatic public monitoring devices can reduce the burden of physical monitoring, which is why we choose to use this technique. This system can be integrated with the device at the edge of use at airports, train stations, offices, schools and public places to ensure compliance with public safety guidelines.

Experimental Results

The proposed program is an in-depth learning solution that uses OpenCV and TensorFlow, model training. We combine MobileNetV2 in-depth learning mode with SSD frame for fast and efficient real-time detection in video streams and use a triangular parallel process to measure the distance between real-time camera people in public places and contains customized data collection to resolve the discovery model. in real-time face-to-face-to-face transmission through learning transfer to a pre-trained SSD face detector. This model combine's social distance detection and face mask detection.

In the proposed system, four steps are followed, such as:

- 1) Data collection and pre-processing
- 2) Model development and training
- 3) Model testing
- 4) Model implementation

1. Data Collection and Pre-processing

The proposed system uses a custom data set with images of different face faces labelled for use training of our models. We are using an existing back-end removal algorithm [21, 22] in the pre-processing step. Real time automatic detection of public distance maintenance and verification of people wearing a mask or not by SSD algorithm. The database used to train our proposed face detector has 3165 images. Before the custom face mask image the database is labelled, the data set is divided into a set of training data and test data is set. Training data set must contain 80% images for effective algorithm training and prediction comprehension and the test data set should contain 20% images depending on check of the prediction accuracy of the algorithm. Images in the data training collection are divided into two categories: mask and no mask.

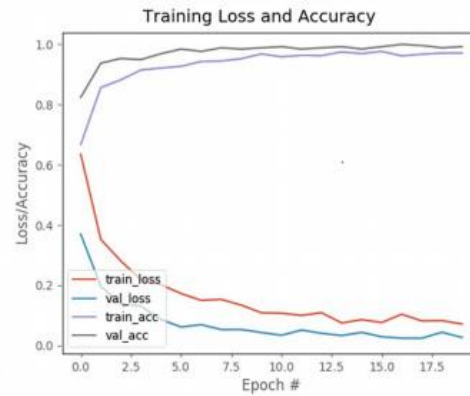


2. Model building and Training

Our proposed framework uses a transfer learning method and will optimize the MobileNetV2 model, which is the most efficient build method that can be used on moderate computing power devices. We used 80% of our complete custom data to train our model with a single shot detector, which takes only one shot to find multiple objects in the image using multibox. A custom data set is loaded into the project directory and the algorithm is trained on the basis of labelled images. In pre-processing steps, the image has been resized to 224×224 pixels, converted to the same member format and the corresponding labels are inserted into the databases before using our SSD model as input build our custom model with MobileNetV2 as the backbone and train our model using the TensorFlow Object Detection API.

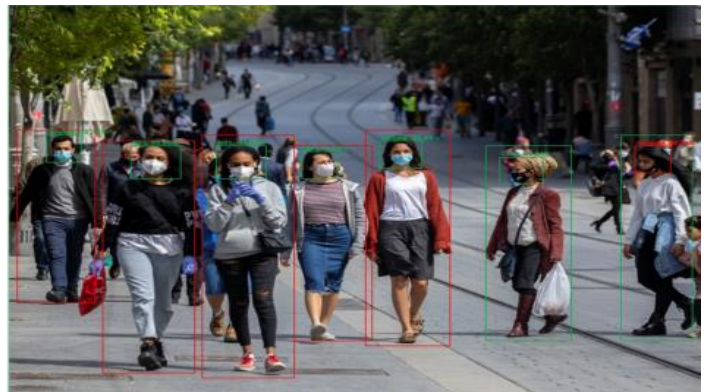
Prior to the start of the modelling training, tensorflow assists with data extensions and downloads of ImageNet's pre-trained tools to make the algorithm's predictive performance more accurate. After downloading pre-trained instruments and building a fully integrated header (FC), the SSD algorithm is trained with both ImageNet pre-trained tools and animated images in custom data set by tuning the header hardware without updating the basic layer hardware. We trained our 1000 step model using the Adam optimizing algorithm, the decay rate of learning network hardware updates, and the cross-entropy binary mask-type editing.

The parameters are started with an initial reading rate of $INIT_LR = 1$ of 4, epoch number $EPOCHS = 20$ and batch size $BS = 32$. We used a webcam to monitor the public distance using cv2 and after the person is identified, we start by tying the box links and inserting the space between the top left and bottom left and the top right and bottom right points. We measure the Euclidean distance between points to determine the distance between the people in the frame.



3. Model Testing

The proposed system operates automatically and helps to automate the community-level assessment process. Once the model has been trained with the custom data set and pre-trained weights provided, we check the accuracy of the model in the test database by displaying the binding box with the brand name and the confidence level above the box. The proposed model first sees everyone in the camera range and shows a solid green box around each person far away from each other by showing binding boxes on the face of a person identified with a mask or non-labelled mask and self-confidence scores. If the mask is not visible on the face, and if the public distance is not maintained, the system creates a warning and sends a warning to the monitoring authorities with a facial image. The system detects social distances and masks with 91.7% accuracy points with a confidence level of 0.7, an accuracy rate of 0.91 and a memory value of 0.91 with $FPS = 28.07$.



4. Model Implementation

The proposed system uses a piercing pi4 with a camera to automatically track public spaces in real time to prevent the spread of Covid-19. The camera feeds real-time videos of public places, which constantly monitors and monitors public locations and detects whether people are keeping public safe

distances and also checks whether those people are wearing a mask or not. Our solution works in two stages: first, when a person is identified anonymously whose picture is taken and sent to the control centre at the Provincial Police Headquarters; and secondly, when the detection of a public nuisance by the public is progressively detrimental, there will be an alarm instructing the public to maintain public order and a critical warning will be sent to the State Police Headquarters for further action.

Applications, Conclusions & Future Scope

We have proposed a way to help maintain a safe environment and ensure the protection of people by automatically monitoring public areas to prevent the spread of the COVID-19 virus and to assist the concerned authorities in reducing their physical activity in areas where there is a need for reliable surveillance and real-time surveillance. Besides, the system is easy to install on any existing business system while maintaining the security and privacy of user information. This proposed system will help in tracking people in public places easily in an automated way. We have talked in detail about the pursuit of social exclusion and the identification of face masks that help ensure human health. Therefore, the Social Distancing Face program will be a leading digital solution for many industries, especially retail, health, temples, shopping malls, municipal stations, airports and corporate sectors.

Things are often seen with their distinctive features. There are too many features in a human face, which can be seen between the face and many other things. It gets a face by removing structural features such as eyes, nose, mouth etc. then use them to find faces. Generally, some form of mathematical division is appropriate and helps to distinguish between face and non-facial regions. Human faces have certain textures that can be used to distinguish between faces and others things. In addition, a series of features can help find things on the face. The above usage cases are one of the many features included as part of this solution. We think there are other use cases that can be included in this solution that provides a detailed sense of security.

References

- [1] <https://www.pyimagesearch.com/2020/06/01/opencv-social-distancing-detector/>
<https://www.goeduhub.com/10498/social-distancing-detection-python-machine-learning/>
- [2] WHO, "Coronavirus disease (COVID-19) advice for the public", 2020, Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public>
- [3] Prateek Khandelwal, Anuj Khandelwal, Snigdha Agarwal, Deep Thomas, Naveen Xavier and Arun Raghuraman, "Using Computer Vision to enhance Safety of Workforce in Manufacturing in a Post COVID World", 2020.
- [4] Ge, J. Li, Q. Ye and Z. Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 426-434, doi: 10.1109/CVPR.2017.53.
- [5] Narinder Singh Punn, Sanjay Kumar Sonbhadra and Sonali Agarwal, "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques", 2020, arXiv:2005.01385.