

# FACE MASK AND SOCIAL DISTANCING VIOLATION DETECTION- A COMPUTER VISION APPROACH

Mahshiya VM, Manju VM, Dr G.Kiruthiga

<sup>1</sup> Student, Department of Computer Science, IES College of Engineering, Kerala, India

<sup>2</sup> Assistant Professor, Department of Computer Science, IES College of Engineering, Kerala, India

<sup>3</sup> Associate Professor, Department of Computer Science, IES College of Engineering, Kerala, India

## ABSTRACT

*This project aims to detect face masks and social distancing on a video feed using Machine Learning and Object Detection. TensorFlow and Keras were used to build a CNN model to detect face masks and it was trained on a dataset of 3800 images. YOLO Object detection was used to detect people in a frame and check for social distancing by calculating the Euclidean distance between the centroids of the detected boxes. Implemented a methodology for door opening for allowing entry only for the people who wear mask in work places. The hardware like sg90 servo motor and arduino nano are used for this purpose. And a voice message notification is also implemented when social distancing violation is detected. This two methodologies ensure more safety on the work places to prevent COVID-19 spread.*

**Keyword:** - Machine Learning, YOLO Object Detection, TensorFlow, Keras, SG90 servo motor, arduino nano, Euclidean distance.

## 1. INTRODUCTION

COVID-19 has brought us a lot of challenges. It affects different people in different ways. The working sectors became stacked. Many establishments are switching to a work from home-based environment. There are still some businesses that opened after lockdown and continuing its normal operations like restaurants, few schools and universities, construction sector, and a few offices all ensuring workplace safety such as wearing masks and social distancing.

The health authorities are working hard on ensuring that these businesses follow all the protocols for keeping the employees safe. They are conducting regular inspections and even shutting down these establishments if they keep violating the safety protocols. Many of these establishments are working towards the automation of detecting such violations, thereby reducing the time and labor spent for the same. The main aim of this project is to detect violations such as not wearing a mask or not following social distancing in a workplace and notify the officials through an Android app. Technology has advanced tremendously over the past century, everything starting from the Internet of Things (IoT) (Raj, 2021) to machine learning and deep learning. CNN is used in various fields like medical (Duran-Lopez, 2019), marine science (Sung, 2017) and many other applications (Hansen, 2017) and has become a prominent domain of machine learning.

This project was implemented using Keras and TensorFlow where the CNN (Convolutional Neural Network) model was trained for face mask detection and the YOLO Object Detection was used for social distancing detection. And this project can also be used in many establishments to allow entry only for people who wear mask. Door opening is established after detecting face mask using a sg90 servo motor connected to an arduino nano. An another methodology we used in this project is the voice message notification after detecting social distancing violation in an establishment. These two methodologies are used to ensure all safety protocols in enterprises to prevent Covid-19 spread.

## 2. RELATED WORKS

This section reviews some of the related works that implements CNN with the help of TensorFlow, Keras and YOLO Object detection and improvements in object detection.

Akanksha Soni et al. (Soni, 2020) developed a model that detects whether a person is wearing a helmet in real time thereby, detecting any violations. This project was also implemented with the help of TensorFlow, Keras and OpenCV. Their proposed model showed major improvements when compared to some previous models that gave wrong predictions whenever a rider wears clothes over their face. They achieved an overall accuracy of 98% when tested.

S Chen et al. (Chen, 2020) implemented a model with the help of TensorFlow to identify ID card numbers. With the help of OpenCV the image of an ID card is preprocessed and the number on the ID card is recognized and given as output with the help of a trained CNN model. When tested it was observed that training speed is fast and the accuracy is high.

Emily Caveness et al. (Caveness, 2020) developed TensorFlow Data Validation (TFDV) which offers a scalable solution for data analysis and validation for machine learning. It is deployed in production which is integrated with TensorFlow Extended (TFX), which is an end-to-end ML platform. Their system has gained a lot of traction ever since they open sourced their project. Other open-source data validation systems such as Apache Spark were also heavily inspired from their project. Apache Spark packs with built-in modules for streaming and has a fast, easy to use system for big data processing. (Nair, 2018)

Yonghui Lu et al. (Lu, 2020) proposed an efficient YOLO Architecture, YOLO-compact for a real time single category detection. As we know in most practical applications, the number of categories in object detection is always single and the authors aimed to make detections faster and more efficient for these scenarios. By performing a series of experiments, the authors were able to come up with an efficient and compact network with the help of YOLOv3. It was observed that YOLO-compact is only of 9MB size, about 26 times smaller than YOLOv3, 6.7 times smaller than tiny-yolov2 and 3.7 times smaller than tiny-yolov3. The average precision of YOLO-compact is 86.85% which is significantly higher than other YOLO model

## 3. METHODOLOGY AND IMPLEMENTATION

This section includes methodologies we used in this project and implementation part of the project.

### 3.1 Face Mask Detection

This project uses TensorFlow and Keras to train a CNN model for detecting face masks.

### 3.2 TensorFlow and Keras

TensorFlow is an open-source platform that is used for Machine Learning, created by the Google Brain team. It is explicitly used for complex numerical computation, that packs together a bunch of machine learning and deep learning models and algorithms. It can be used for a variety of applications such as classifying handwritten digits, object detection, image recognition, natural language processing (Natraj, 2019) by training and running deep neural networks. Application Development for Mask Detection and Social Distancing Violation Detection using Convolutional Neural Networks Keras which acts as an interface for TensorFlow is an open-source library that provides an efficient way of implementing neural networks. It consists of useful functions such as activation functions, and optimizers.

#### 3.2.1 How Does TensorFlow Work?

With the help of TensorFlow, developers can create dataflow graphs which are structures that show how data passes through the graph, or a series of nodes. Think of each node as a mathematical operation and each edge representing a multidimensional data array or a tensor. This can be easily implemented in python where these nodes and tensors act as objects. However, the mathematical operations are performed in C++ binaries which shows an optimal

performance. Python takes care of directing the traffic and combines them to work together as a unit. TensorFlow can be run on multiple platforms such as in a cloud, a local machine, CPUs or GPUs, iOS, and Android devices. It can also be run on Google's custom TensorFlow Processing Unit (TPUs). The trained models can be run on any system for predicting results. TensorFlow 2.0 which was released in October 2019 made many significant changes from user feedback. It works more efficiently and is more convenient with simple Keras API for training models and better performance. With the help of TensorFlow Lite, it is possible to train models on a wide variety of devices.

### 3.3 Training the Face Mask Detection Model

In this project a convolutional neural network is used to detect face masks. The neural network takes in the input image as a frame from the video, processes it and classifies it under two categories: mask and no mask. The model was trained using 3800 images, 1900 images each for "with mask" and "without mask" categories.

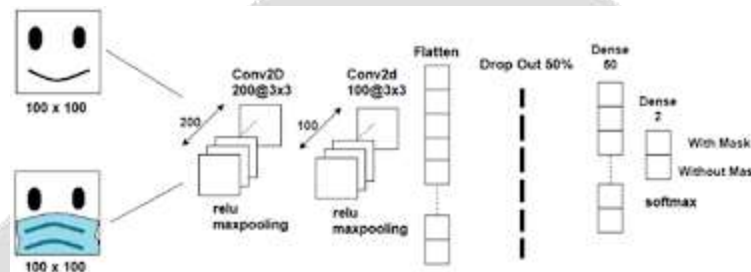


Fig 1: Network architecture

The network consists of two convolution layers each followed by a relu activation function and a max pooling layer. Figure 1 shows the network architecture. Relu function (Rectified Linear Unit) is introduced for non-linearity in the convolutional network. It is shown as:

$$F(x) = \max(0, x)$$

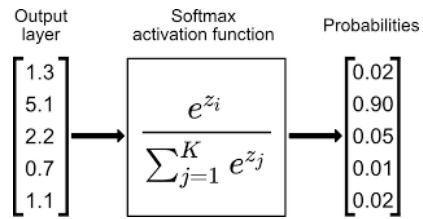
Pooling layers will reduce the number of parameters if the images are large. Spatial pooling also referred to as sub-sampling or down-sampling helps to reduce the dimensionality of each map making sure important information is retained. It can be of different types:

- Sum Pooling
- Max Pooling
- Average Pooling

Max pooling will take the largest element from the rectified feature map.

The data is then flattened by converting it to 1-dimensional array which is passed as the input to the final output layer. To help prevent overfitting, the network ignores a certain percentage of neurons during training. In this case the network drops out 50% of neurons. These units are not considered during some forward or backward pass. The final output layer takes the values and transforms them into a probability distribution, this is achieved with the help of softmax function. This function is helpful when it comes to classification problems or when dealing with multi-class classification problems. The final predicted class is the item in the list whose confidence score is the highest. It is represented as:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$



**Fig 2 :** Softmax activation function

The final prediction will be based on the class that has the highest probability. The model was then compiled using Adam optimizer and trained for 20 epochs with a learning rate of 0.001 and an accuracy of 96.35% was observed on the validation set.

### 3.4 Detection

OpenCV was used to capture the video and the trained model was loaded using TensorFlow.

A pre-trained model was used to detect faces in a video. The weights were initialized from the configuration file with the help of OpenCV.

After getting the bounding boxes for the faces in the frame, that region of interest is cropped out from the main frame, reshaped, and is then passed to the model.



**Fig 3:** Detecting face masks

### 3.5 Door Opening

In this project we implemented a methodology for door opening for allowing entry only for the people who wear mask in work places. The hardware like SG90 servo motor and Arduino nano are used for this purpose



**Fig 4:** SG90 Micro Servo Motor

Micro Servo Motor SG90 is a tiny and light weight server motor with high output power. Servo can rotate approximately 180 degrees (90 in each directions), and works just like the standard kinds but smaller. We can use any Servo code, hardware or library to control these servos.



**Fig 5:** Arduino nano

Arduino Nano is used to produce a clock of precise frequency using constant voltage. There is one limitation of using Arduino Nano i.e. it doesn't come with a DC power jack, which means you can not supply an external power source through a battery.

#### 4. SOCIAL DISTANCING DETECTION

This project uses YOLO Object detection for detecting people in a frame and finds the Euclidean distance between the centroids of the detected boxes.

##### 4.1 YOLO Object Detection

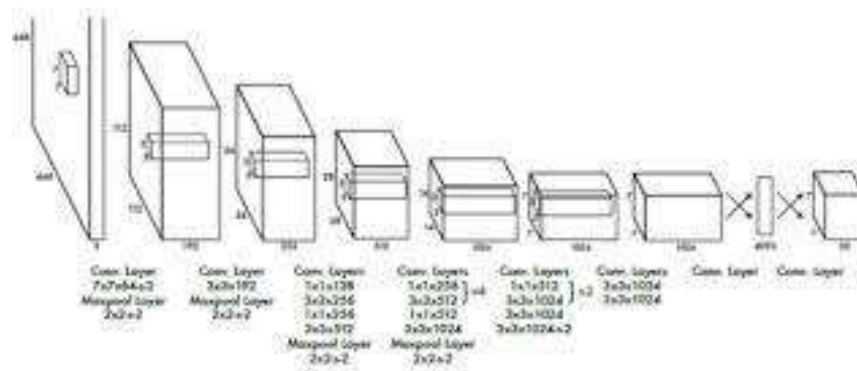
You Only Look Once (YOLO) is an effective real time object detection system. It considers object detection as a regression problem and finds the class probabilities for each of the bounding boxes. In one evaluation the neural network predicts bounding boxes and class probabilities from the image, hence the name YOLO. The base model detects images at an astonishing speed of 45 frames per second whereas a smaller version called Fast YOLO detects at 155 frames per second. It performs better than other detection methods such as Deformable Part Models (DPM) and Region-based convolutional neural networks (R-CNN). There are two types of algorithms that work towards object detection- Algorithms based on classification, and Algorithms based on regression. YOLO falls under the latter category.

##### 4.1.1 How Does YOLO Work?

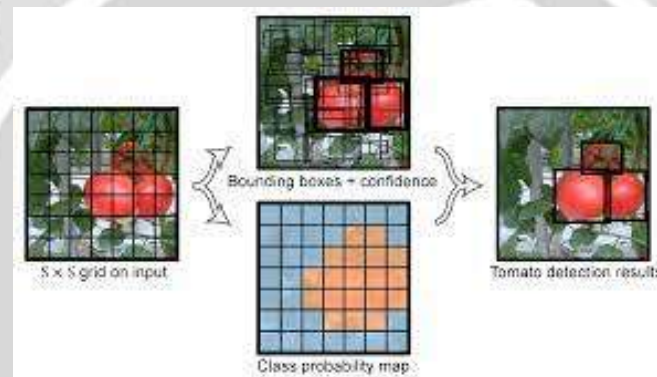
A single neural network is used to combine separate parts of object detection. It takes in features from an image to predict each bounding box. The detection is modeled as a regression problem where an image is divided into  $S \times S$  grids. In Fig. 7, the authors set the value of  $S$  as 7. This can be changed in the YOLO configuration file. Each grid predicts bounding boxes (B), their confidence scores, and their class probabilities (C). These predictions are encoded as an  $S \times S \times (B \cdot 5 + C)$  tensor.

Each bounding box has 5 predictions: (x, y, w, h) and confidence. Each grid cell predicts conditional class probabilities (C) as  $P(\text{Class} | \text{Object})$ .

The architecture for this Convolutional Neural Network was inspired from GoogLeNet model that is used for image classification. Just like GoogLeNet their model was implemented using 24 convolutional layers that helps to extract features from the image followed by 2 fully connected layers to predict the output probabilities and coordinates. They also came up with a faster version on YOLO named Fast YOLO that uses a CNN with less layers (9 instead of 24) and less filters for those layers. Other than that, the training and testing parameters were the same between YOLO and Fast YOLO.



**Fig 6:** The YOLO architecture



**Fig 7:** The YOLO Model (Redmon, 2016).

## 4.2 Detection

This project uses YOLO v3 to detect people. OpenCV python package was used to read the weights and set up the model using the configuration file.

```
net=cv2.dnn.readNet("./YOLO/yolov3.weights", "./YOLO/yolov3.cfg")
```

The classes were read from “coco.names” file which contains the names of 80 different classes such as person, bicycle, car etc. Open CV was used to capture the video as shown below. This project was tested on some sample videos, on the webcam and on an IP camera as shown below.

```
cap=cv2.VideoCapture('./Videos/Example 4.mp4')
```

```
cap = cv2.VideoCapture(0)
```

```
cap=cv2.VideoCapture("http://192.168.2. 2:8080/video")
```

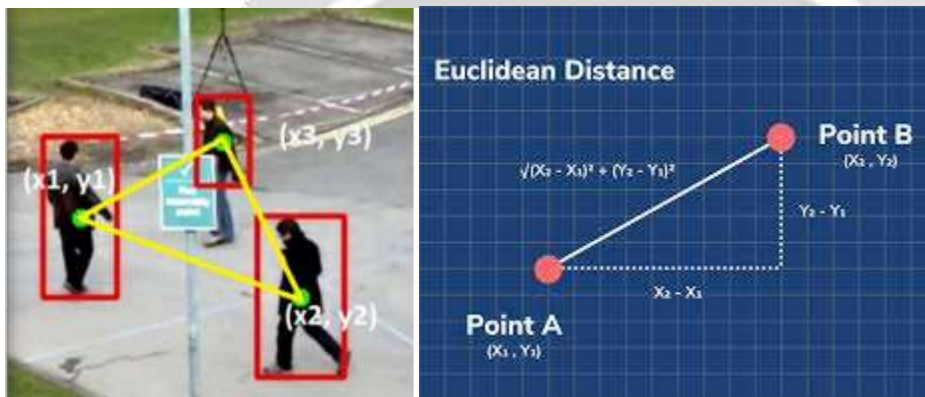
In each frame, the people are detected, and boxes are drawn along with their centroids. The centroids for each box were calculated and appended to a list. Red boxes indicate that the person is violating social distancing as shown in Fig.8 below.



**Fig 8:** Detecting Social Distancing Violation.

Using Euclidean distance, the distance between the centroids can be calculated with the help of the scipy package.

```
D = dist.cdist(centroids,centroids,metric="euclidean")
```



**Fig 9:** Calculating the Euclidean Distance between centroids.

Where D is a 2-D array as shown in the table below:

**Table 1:** 2-D array containing the Euclidean distance between each point

	(x1,y1)	(x2,y2)	(x3,y3)
(x1,y1)	0	5	3
(x2,y2)	5	0	5
(x3,y3)	3	5	0

Where each element in the array is calculated according to the formula mentioned below:

$$\text{Dist}((x_i, y_i), (x_j, y_j)) = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]}$$

$i, j = 1, \dots, N$  detections

From the above equation, i and j represents rows and columns in the 2-D array, respectively.

If this distance is less than a certain threshold value, the person is violating social distancing. We can estimate this threshold value by finding the ratio as follows:

Assuming the pixel density to be 96 dpi, we can deduce that 1 pixel = 0.26458333 mm or 1 mm = 3.7795 pixels. Taking the ratio,

$Ratio = \frac{\text{Real Height of the person (mm)}}{\text{Pixel Height of the person in the frame (mm)}}$

Now to estimate the pixel distance for 1 meter (1000 mm):

$1000 / Ratio * 3.7795$

This threshold value may or may not be accurate and it needs to be increased or decreased accordingly.

## 5. CONCLUSIONS

This project has been developed to come up with an efficient way for detecting and notifying officials ICEIS 2021 - 23rd International Conference on Enterprise Information Systems 766 when a person does not follow the COVID 19 safety protocols in a workplace, business establishments etc. In this work, we have trained a model for face mask detection using TensorFlow and Keras and used YOLO Object detection for detecting social distancing. The proposed CNN architecture comprises two convolutional layers followed by relu activation function and a max pooling layer. YOLOv3 was used to detect people in a frame and find the Euclidean distance between them. With the help of OpenCV we were able to capture the video feed from different sources like webcam, video file or an IP camera.

In this work we use an SG90 servo motor for door opening function to allow entry only for the people who wear mask. And a voice message notification is also implemented when detecting social distancing violation. These two methods ensure more safety on working places

## 5. REFERENCES

- [1]. Soni, A., & Singh, A. P., 2020. Automatic Motorcyclist Helmet Rule Violation Detection using Tensorflow & Keras in OpenCV. In 2020 IEEE International Students Conference on Electrical, Electronics and Computer Science (SCEECS).
- [2]. Chen, S., Wei, Y., Xu, Z., Sun, P., & Wen, C., 2020. Design and Implementation of Second-generation ID Card Number Identification Model based on TensorFlow. In IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA).
- [3]. Soni, A., & Singh, A. P., 2020. Automatic Motorcyclist Helmet Rule Violation Detection using Tensorflow & Keras in OpenCV. In 2020 IEEE International Students Conference on Electrical, Electronics and Computer Science (SCEECS).
- [4]. Chen, S., Wei, Y., Xu, Z., Sun, P., & Wen, C., 2020. Design and Implementation of Second-generation ID Card Number Identification Model based on TensorFlow. In IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA).
- [5]. Caveness, E. C., P. S., Peng, Z., Polyzotis, N., Roy, S., & Zinkevich, M., 2020. TensorFlow Data Validation: Data Analysis and Validation in Continuous ML Pipelines. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data.
- [6]. Lu, Y., Zhang, L., & Xie, W., 2020. YOLO-compact: An Efficient YOLO Network for Single Category Realtime Object Detection. In 2020 Chinese Control and Decision Conference (CCDC).
- [7]. Ullah, M. B., 2020. CPU Based YOLO: A Real Time Object Detection Algorithm. In 2020 IEEE Region 10 Symposium (TENSYP).
- [8]. Raj, A., Maji, K., & Shetty, S. D., 2021. Ethereum for Internet of Things security. In Multimedia Tools and Applications.
- [9]. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [10]. Natraj, L., & Shetty, S. D., 2019. A Translation System That Converts English Text to American Sign Language Enhanced with Deep Learning Modules. In International Journal of Innovative Technology and Exploring Engineering Regular Issue, 8(12), 5378-5383.
- [11]. Khawas, C., & Shah, P., 2018. Application of Firebase in Android App Development-A Study. In International



- Journal of Computer Applications, 179(46), 49-53.
- [12]. Fatima, N. S., Steffy, D., Stella, D., & Devi, S. N., 2020. Enhanced Performance of Android Application Using RecyclerView. In *Advances in Intelligent Systems and Computing* Advanced Computing and Intelligent Engineering, 189-199.
- [13]. Nair, L. R., Shetty, S. D., & Shetty, S. D., 2018. Applying spark based machine learning model on streaming big data for health status prediction. In *Computers & Electrical Engineering*, 65, 393-399.
- [14]. Duran-Lopez, L., Dominguez-Morales, J., AmayaRodriguez, I., Luna-Perejon, F., Civit-Masot, J., Vicente-Diaz, S., & Linares-Barranco, A., 2019.
- [15]. Breast Cancer Automatic Diagnosis System using Faster Regional Convolutional Neural Networks. In *Proceedings of the 11th International Joint Conference on Computational Intelligence*.
- [16]. Sung, M., Yu, S., & Girdhar, Y., 2017. Vision based realtime fish detection using convolutional neural network. In *OCEANS 2017 - Aberdeen*. Hansen, D. K., Nasrollahi, K., Rasmusen, C. B., & Moeslund, T. B. (2017).
- [17]. Real-Time Barcode Detection and Classification using Deep Learning. In *Proceedings of the 9th International Joint Conference on Computational Intelligence*.
- [18]. Goodfellow, I., Bengio, Y., & Courville, A., 2016. *Deep learning*, Cambridge (EE. UU.). MIT Press. Belciug, S., 2020. *Artificial intelligence in cancer: Diagnostic to tailored treatment*, London, United Kingdom.
- [19]. Academic Press. *Application Development for Mask Detection and Social Distancing Violation Detection using Convolutional Neural Networks*

