AN AI BASED SOCIAL DISTANCE MONITORING SYSTEM

A Project Report

Submitted to the APJ Abdul Kalam Technological University

in partial fulfillment of requirements for the award of degree

in

Master of Computer Application

by

ABHISHEK M

TKM20MCA-2002

DEPARTMENT OF MCA

TKM COLLEGE OF ENGINEERING ,KOLLAM

KERALA

July 2022

DEPT. OF MCA

TKM COLLEGE OF ENGINEERING KOLLAM 2020 - 22

CERTIFICATE

This is to certify that the project report entitled **AN AI BASED SOCIAL DISTANCE MONITORING SYSTEM** submitted by **ABHISHEK M** (TKM20MCA-2002), to the APJ Abdul Kalam Technological University in partial fulfillment of the M.C.A degree in Master of Computer Application is a bonafide record of the project work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

(Internal Supervisor) **(Head of the Department)** **(External Examiner)**

DECLARATION

I ABHISHEK M hereby declare that the Project report **AN AI BASED SOCIAL DISTANCE MONITORING SYSTEM** , submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Application of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Dr. NADEERA BEEVI S

This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources.

I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Kollam

22-07-2022

ABHISHEK M

ACKNOWLEDGEMENT

I take this opportunity to express my deepest sense of gratitude and sincere thanks to everyone who helped me to complete this work successfully. I express my sincere thanks to **Dr. FOUSIA M SHAMSUDEEN**, Head of the Department, MCA, TKM COLLEGE OF ENGINEERING Kollam for providing me with all the necessary facilities and support.

I would like to place on record my sincere gratitude to my project guide **Dr. NADEERA BEEVI S**, Professor, MCA, TKM COLLEGE OF ENGINEERING for the guidance and mentorship throughout the course.

I profusely thank all other faculty members in the department and all other members of TKM College of Engineering, for their guidance and inspirations throughout my course of study. I owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this project.

ABHISHEK M

ABSTRACT

The virus that causes Covid19 is identified as SARS-CoV-2. The corona virus's devastating spread has brought forth a global catastrophe. Social isolation is thought to be a defense mechanism against the pandemic virus's rapid spread. By avoiding direct social contact with other people, the danger of the virus spreading can be reduced. In order to develop a deep learning platform for social distancing utilizing an aerial perspective, this work's goal is to achieve this. To detect people in video footage, the framework employs the YOLO object detection approach. The recognition and tracking of individuals in both indoor and outdoor environments is done using a deep learning detection approach. The users of the discovered bounding box information are identified using the detection model. The distance between two individuals from the centre of the observed bounding box is calculated using the Euclidean distance. Here utilised a pixel-physical distance calculation and a threshold to calculate the prevalence of social distance violations between individuals. To determine whether the distance value exceeds the minimal social distance criterion, violation thresholds are developed. Additionally, a tracking algorithm is employed to identify people in the video clip in order to follow anyone who violates or crosses the social threshold. The suggested method may be used for a low-cost embedded device with a fixed camera. The suggested method may be used to watch individuals from various cameras in a centralized surveillance system using a distributed CCTV system. This method is appropriate for establishing a surveillance system in smart cities to find individuals, categorize them, and assess social distance.

Contents

List of Figures

Chapter 1

INTRODUCTION

The SARS-CoV-2 virus spreads the infectious illness coronavirus disease. Most COVID-19 patients only have moderate to minor symptoms, and they often truly recover on their own. However, some individuals will become really sick and need medical care. According to the most current research, Most often, direct contacts are where the infection is disseminated, such as while they are conversing.

The virus may enter the mouth or nose and go through to the air when an infected person talks, sneezes, sings, or coughs. Following that, if airborne infected particles come into contact with another person's eyes, nose, or mouth, or if they are breathed in within a small distance (known as short-range aerosol or short-range airborne transmission), they may get the virus (droplet transmission). The virus may also spread in congested, unventilated indoor areas where people spend more time inside than outside. This is a result of aerosols' propensity to adhere to surfaces or travel further than necessary for conventional communication.

Close contact between people is restricted by the implementation of physical and social barriers in infection control to inhibit the spread of disease. Bans on travel, travel restrictions and the closure of businesses, schools, stadiums, theatres, or shopping malls are a few of the tactics used. By avoiding social settings, restricting their travel, avoiding popular places, utilizing no-contact greetings, and physically separating themselves from others, people can practice social distancing. Several governments in nations where the disease has spread either demand or support social seclusion.

The variation in the reported cases can be identified by looking in to Figure 1.1

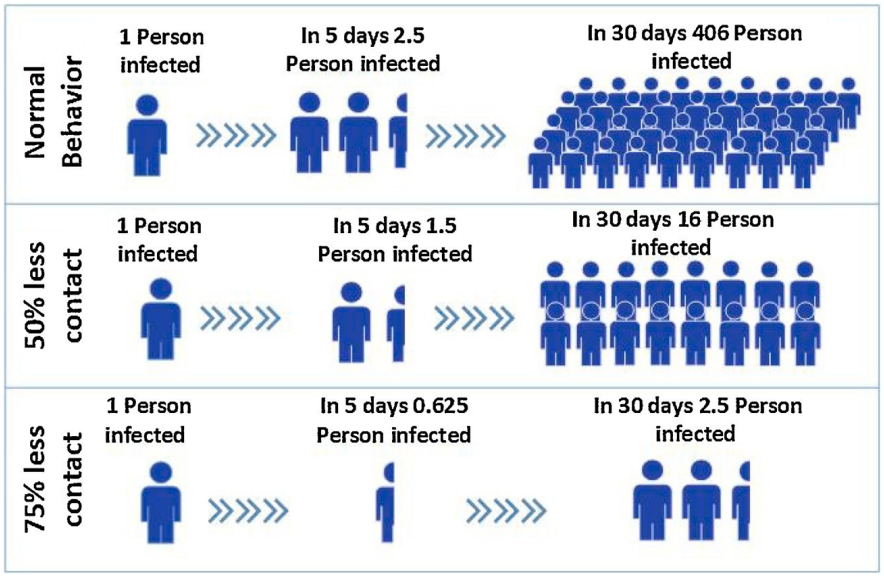

Figure 1.1: The significance of social distance

Deep learning, machine learning, and computer vision have all demonstrated promising outcomes in recent years for a variety of challenges encountered in daily life. Deep learning has recently improved, making object detection tasks more efficient. These techniques are frequently used to gauge social distance between individuals across moving frames. Clustering and distance-based algorithms are used to calculate how far apart people are from one another. A good camera calibration is necessary to map pixels to distance for truly easily measurable units because the majority of approaches were created using frontal or side view video sequences (i.e., feet, meters, etc.). Second, if chooses a top-down strategy, or an overhead view strategy, the distance calculations from the overhead view will result in an improved distance estimation and extensive coverage of the expansive scene.

In order to develop a practical method for social distance monitoring, Here adopted an overhead perspective in this approach. To monitor social distance and calculate the distance between people, the stated viewpoint improves the field of vision and addresses the occlusion problem. Costs related to computers, communication, energy use, human resource utilization, and installation could be reduced. A deep learning-based system for social distance monitoring on an open campus is what this project aims to achieve. For human identification, the YOLOv5 (You Only Look Once) deep learning model is used. The detection model generates bounding box information and identifies people. Following pedestrian detection, the observed bounding box and any

related centroid data are used to determine the Euclidean distance between each pair of identified centroid pairs. A predetermined minimum social distance violation threshold is generated using pixel-to-distance presumptions. To determine whether the computed distance is included in the list of violations or not, the estimated data is compared to the violation threshold. The bounding box's original color of green is changed to red if it is discovered to be a part of the violation set. Using the centroid tracking method, a person who has passed the social separation barrier is also visible.

Chapter 2

LITERATURE STUDY

In November 2019, the Covid19 virus first appeared on a low scale. In December 2019, Wuhan, China, have seen the first significant cluster.The first suspected location of SARS-CoV-2 transmission in humans was one of Wuhan, China's outdoor wet markets. Later rumours cast doubt on the possibility that it began life as a biological weapon at a Chinese research facility. When the pandemic started to spread in late December 2019, Over the course of a month, the number of cases dramatically increases, During the first week of February 2020, daily estimates of two to four thousand additional confirmed cases had been reached.

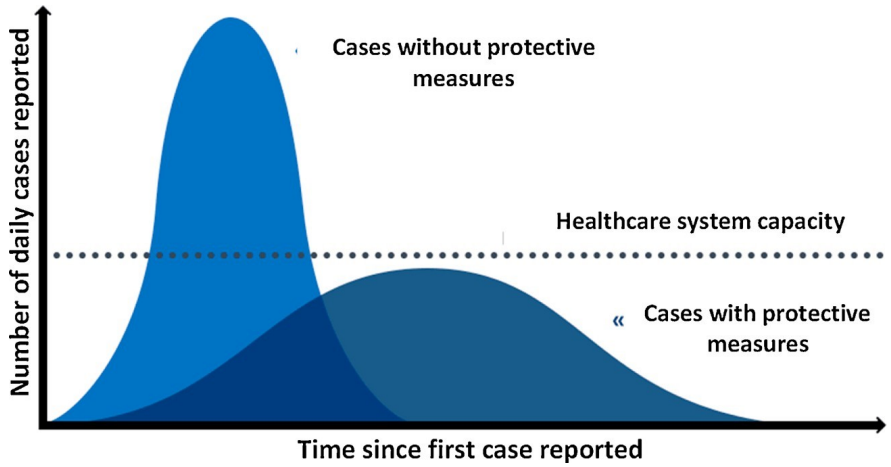

Figure 2.1: Impact of social isolation: pandemic instances are declining and are currently treated by the healthcare system's capacity ().

From the Figure 2.1 it is clear that Social distancing was found to be the most efficient way for halting the transmission of the contagious virus and was accepted as standard practice on January 23, 2020.When that happened, there was a first-ever sign

of respite that lasted for five straight days and continued until March 23, 2020, with no more confirmed instances. This is a result of the COVID-19 social distancing control strategy, which was first used in China and then spread internationally. Many wealthy countries are using GPS technology to locate sick people and potential victims.

A novel, relatively inexpensive vision-based technique that can detect human localizations in three dimensions and body orientation from a single image was presented in 2019 by Lorenzo Bertoni et al. [1]. By identifying the spatial configurations that characterize F-formations, they may recognize social interactions. By examining all potential pairings of individuals in a picture, the method takes into account groups of two persons in an ”all-vs-all” manner. The difficulty in localizing persons in 3D from a single image is the fundamental barrier to understanding social interactions.

November 2020. Sergio Saponara et al. suggested a paper[2]. They recommended classifying persons based on social distance using thermal images and artificial intelligence. The YOLOv2 (you look at once) methodology is used to construct a new deep learning detection method for spotting and following individuals in both indoor and outdoor settings. Additionally, an algorithm is employed to estimate and assess interpersonal distances as well as to automatically assess compliance with social distance norms. In order to lessen the spread of the COVID-19 virus, it is important to understand if and how people follow the social distancing rule. The recommended approach uses images from thermal cameras to build a comprehensive AI system for people tracking, social distance categorization, and body temperature monitoring. The suggested method is used to watch individuals from several cameras in a single centralised monitoring system on a distributed surveillance video system. The findings demonstrate the viability of the suggested approach for implementing a surveillance system in smart cities that is suited for people identification, social distance classification, and body temperature analysis.

In December 2021, ZHIMING CHEN et al . developed a model[3] that uses surveillance robots to promote social estrangement. In their study, they created a completely quadruped-based autonomous surveillance robot that can encourage physical distance in complex metropolitan situations. For the purpose of achieving autonomy, On the legged robot, they installed numerous cameras and a 3D LiDAR sensor specifically for the purpose. The robot then tracks surrounding pedestrian

groups using an integrated real-time social distance detection algorithm. Next, the robot navigates freely across highly dynamic environments by using a crowd-aware navigation algorithm. Finally, the robot uses a crowd-aware routing algorithm to offer suggestions to backed-up pedestrians using human-friendly vocal cues, thereby promoting social distance. The primary downsides of LiDAR are its expensive cost and difficulty to measure distance in the presence of significant amounts of snow, rain, or fog.

Drones and other surveillance cameras are used by Zhenfeng Shao et al. to identify crowd gatherings. In order to accurately identify pedestrians via human head identification in real-time and to determine the social distance between pedestrians on UAV photos, they presented a network for light-weight pedestrian detection. To improve the features of small objects, like human heads, the network uses multi-scale features and spatial attention in addition to using PeleeNet as its structural backbone.

Chapter 3

OBJECTIVE AND METHODOLOGY

3.1 OBJECTIVE

Social isolation is essential, especially for those who are more likely to become really sick with COVID-19. Keeping enough space between people and avoiding crowds in public places are two ways to increase social distance through reduced physical interaction. By lowering the chance of the virus spreading from infected people to healthy people, the virus and the severity of the illness can be considerably reduced.

3.2 METHODOLOGY

A method for tracking social distance based on deep learning is presented in this paper. In this work, persons in sequences are identified using the YOLOv5 deep learning-based detection paradigm. To calculate the bounding boxes and class probabilities, the model employed a single-stage network design. The COCO (Common Objects in Context) data set served as the model's first training ground. The centroid data for each bounding box is utilized to calculate the separation between them after detection. Each bounding box of the observed populations is measured here using the Euclidean distance. When the distance between any two bounding box centroids has been computed, a predefined threshold is used to decide whether or not it is less than the

predetermined pixel count. If two persons are close to one another but their distance value is greater than the required minimum social distance. The bounding box's color is updated to red and the bounding box's information is recorded in a violation set. The use of a centroid tracking algorithm makes it possible to monitor people who cross or exceed the social distance threshold. The model's output displays the total number of observed social distance violations as well as any discovered person bounding boxes and centroids.The flow diagram of the proposed method is shown in fig 3.1

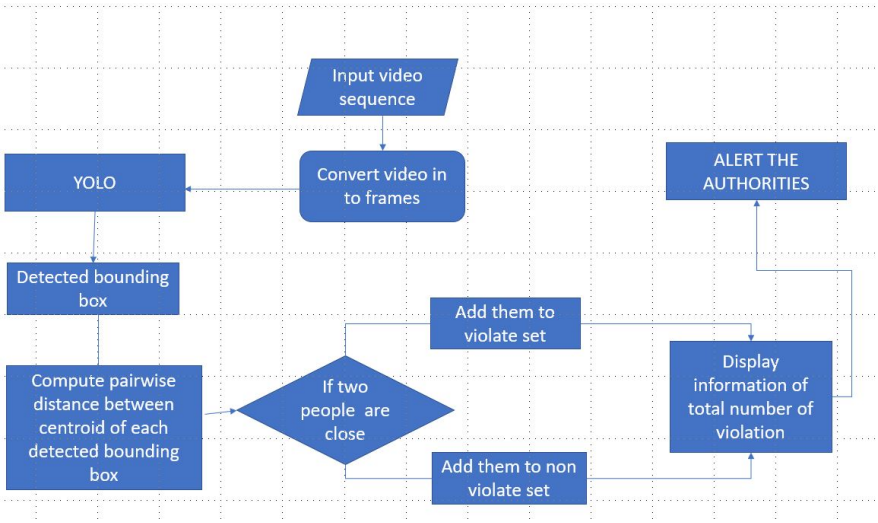

Figure 3.1: Flow diagram of social distance monitoring system

The primary aim of this effort are as follows:

• For providing a AI-based solution for measuring social distance from above.

• Pre-trained YOLOv5s will be utilised to identify persons and determine their bounding box centroid data.

The centroid of the observed bounding box of each pair has been separated exactly using the Euclidean distance, which has also been used to track social distance between people. A physical distance violation threshold is provided via a pixel-to-distance calculation.

• using centroid monitoring technique to track those who go above the physical distance restriction.

3.3 Computer vision

Computers and systems may gather information from digital photos, videos, and other visual inputs using computer vision, which is a kind of artificial intelligence that enables them to take actions or offer suggestions based on that data. The ability for robots to see, hear, and comprehend is provided by both artificial intelligence (AI) and computer vision (CV).

The long history of human eyesight gives it an edge over machine vision. Human sight has the benefit of being able to learn how to differentiate between objects, gauge their distance from the observer, assess if they are moving, and assess whether a picture is accurate throughout the course of a lifetime.

With the advent of computer vision, which relies on cameras, data, and algorithms rather of retinas, optic nerves, and a visual brain, computers can now learn to do identical tasks far more rapidly. Since a system that has been trained to inspect things or monitor a production asset is capable of analysing hundreds of products or processes per minute while identifying hidden defects or issues, it may quickly outperform people. Computer vision is used by many industries, including manufacturing, the car industry, and the energy and utility sectors.

3.3.1 Working of computer vision

For computer vision, a lot of data is needed. It repeatedly executes analyses of the data until it can distinguish between things and recognizes images. For instance, a computer needs to be fed a huge amount of tires photos and tire-related things in order to be trained to detect automotive tires. This is especially true of tires without any flaws.

Convolutional neural networks and deep learning, a kind of machine learning, are two fundamental technologies used in this (CNN).

a . Machine Learning

Machine learning, a area of computer science and AI, focuses on leveraging data and methods to mimic human learning processes and progressively improve the system's accuracy. Machine learning may be used to teach a computer how to comprehend the

context of visual input using algorithmic models. If sufficient information is fed into the model, the computer will ”look” at the data and teach itself to differentiate between various images. Algorithms enable the computer to learn on its own and eliminate the necessity for programming in order to recognise a picture.

Data science is a fast expanding area, and machine learning is a key component. Algorithms are taught to produce classifications or predictions using statistical methods in data mining operations. Ideally, decisions are taken that have an impact on significant growth indicators in applications and enterprises.

b . CNN

Deep learning uses a class of ANN called convolutional neural networks (CNN, or ConvNet). Deep learning is made use of by sophisticated artificial intelligence (AI) image processing systems like CNN to perform both generative and descriptive tasks. They frequently make use of computer vision technology, a combination of natural language processing, recommendation engines, and image and video identification (NLP).

A hardware or software system known as a neural network is modelled after how neurons in the human brain function. Traditional neural networks must be fed images in pixel-by-pixel, low-resolution chunks, which is not suitable for image processing. CNN’s ”neurons” are set up more like the frontal lobe, which in humans and other animals is where visual data are processed. The full visual field is covered by the layers of neurons, overcoming the issue with standard neural networks’ piecemeal picture processing.

A CNN uses a technology that resembles a multilayer perceptron and was built with minimal processing requirements. A CNN’s layers consist of an input layer, an output layer, and a hidden layer along with several convolutional, pooling, fully connected, and normalising layers. By eliminating constraints and enhancing image processing, a system that is far is produced.

By splitting images into pixels with labels or tags, a CNN aids a machine learning or deep learning model’s ability to ”see.” It creates predictions about what it is ”seeing” by performing convolutions on the labels, which is a mathematical operation on two

functions to create a third function. Until the predictions start to come true, the neural network conducts convolutions and evaluates the accuracy of its predictions repeatedly. Then, it is recognizing or viewing images similarly to how people do.

Similar to how a human would perceive a picture from a distance, a CNN first recognises sharp contours and basic forms before adding details as it iteratively tests its predictions. To comprehend individual images, a CNN is utilised. In a manner similar to this, recurrent neural networks (RNNs) are employed in video applications to assist computers in comprehending the relationships between the images in a sequence of frames.

3.3.2 Computer vision Application

Computer vision is an area of research that is being actively explored, but it extends farther than that. Applications in actual environments indicate how essential computer vision is to activities in commerce, entertainment, transportation, healthcare, and day-to-day life. One of the main driving forces behind the growth of these applications is the flow of visual data from smartphones, security systems, traffic cameras, and other visually instrumented devices. Even if it isn't being used right now, this information might be crucial to the functioning of many different firms. The information creates a learning environment for computer vision programs and a foundation for their incorporation into a range of human pursuits.

a . Image classification

Images that can be categorized using image classification include a dog, an apple, and a person's face. In more detail, it can correctly infer the class to which a given image belongs. For example, a social media company might want to employ it.

b . Object detection

In order to identify a specific class of image and then recognize and tabulate its existence in an image or video, object detection can employ image classification. Detecting damage on an assembly line or monitoring technology that needs maintenance are a few examples.

c . Object tracking

An item is tracked or followed after being discovered. This process is typically performed with real-time video streams or a collection of photos that were taken one after another. For instance, in order to prevent collisions and adhere to traffic laws, autonomous cars must track moving objects like people, other vehicles, and road infrastructure in addition to categorising and identifying them.

d . Content-based image retrieval

To explore, search, and retrieve images from enormous data repositories, content-based image retrieval uses computer vision rather than focusing on the metadata tags that are associated with the images. Manual image labeling can be replaced due to the faster annotation for this task. To increase the accuracy of search and retrieval, these activities may be applied to digital asset management systems.

3.4 Object Detection

The objective of object detection, a field of computer science related to computer vision and image processing, is to find instances of semantic objects belonging to a certain class (such as people, buildings, or automobiles) in digital images and videos. Face and pedestrian detection are two categories of object detection that have undergone substantial investigation. Many computer vision applications, such as image search and video monitoring, require object detection.

3.4.1 YOLOV5

The COCO dataset was used to train the YOLOv5 family of compound-scaled object identification models, which contains the fundamental skills for Test Time Augmentation, model assembly, hyperparameter development, and export to ONNX, CoreML, and TFLite. Size; model It is a brand-new convolutional neural network (CNN) that detects things quickly and correctly. A single neural network is used to process the entire image in this approach, after which it is divided into its component

Figure 3.2: Detection of people

parts and bounding boxes and probabilities are predicted for each one.Figure 3.2 shows the detection of people using yolov5

3.4.2 YOLOv5 Architecture

Like other single-stage object detectors, YOLO v5 contains three crucial components as it is a single-stage object detector.As seen in the Figure 3.3 it has

- Model Backbone
- Model Neck
- Model Head

Model Backbone's primary objective is to extract important details from an input image. In YOLO v5, the method for extracting highly informative features from an input picture is CSP—Cross Stage Partial Networks. CSPNet has shown a significant decrease in processing time with deeper networks.

The primary purpose of Model Neck is to produce feature pyramids. Pyramids enable models to scale objects successfully in general. The ability to recognize the same thing in various sizes and scales is helpful.

Models that use feature pyramids perform well on unobserved data. Other models, such as FPN, BiFPN, PANet, etc., employ other feature pyramid methodologies. PANet is utilized in YOLO v5 as a neck to obtain feature pyramids. It can be difficult to

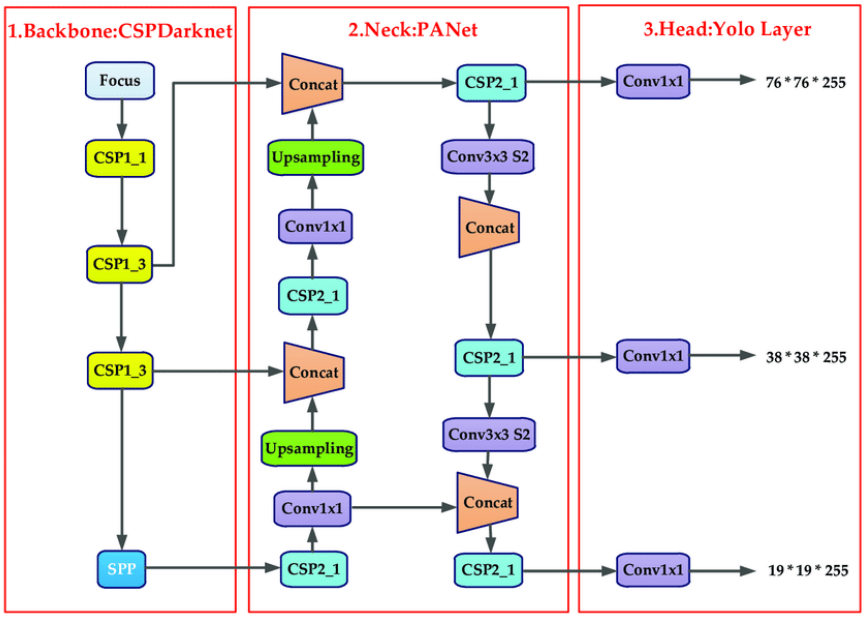

Figure 3.3: Yolov5 Architecture

find objects of various sizes, especially small ones. To find items, Here uses a pyramid of the same image at several scales. However, processing numerous scale images take time, and training end-to-end simultaneously would require too much memory. As a result,can only use it to conclude that accuracy should be pushed as high as possible, especially for contests, when speed is unimportant. Instead, here build a pyramid of features and use them to detect objects . Feature maps, on the other hand, that are closer to the image layer are made up of low-level structures that are ineffective for precise object detection.

A feature extractor called the Feature Pyramid Network (FPN) was developed with the precision and speed of a pyramid in mind. It substitutes detectors like Faster R-feature CNN's extractor for object recognition and creates several feature map layers with more precise information than the traditional feature pyramid (multi-scale feature maps).

The model Head is primarily used for the final detecting step. It produced final output coordinates with bounding boxes, scores, and class probabilities using bounding boxes on the features. The heads of the YOLO V3 and V4 models are similar to the head of the YOLO v5.

In YOLO v5, the last detection layer uses the sigmoid as activation function whereas the middle and hidden layers use the Leaky ReLU as activation function.

SGD is the standard optimization function for training.

3.5 Object Tracking

Tracking an object involves:

- Identifying the people in the image and placing a bounding box around them
- Giving each of the initial detections a distinct ID
- Thereafter keeping track of each object's assignment of a distinct ID as it moves between frames in a video.

This can count distinct items in a video with object tracking, which enables to assign an individual ID to each tracked object. When creating a person counter, object tracking is essential.

3.5.1 Different Object Tracking Types

Image tracking and video tracking are the two methods of object tracking.

- Image tracking:The method of automatically detecting and following visuals is known as image tracking.

  It mostly pertains to augmented reality (AR). The method, for instance, finds two-dimensional planar pictures that may be used to overlay a three-dimensional graphic object when a two-dimensional image is entered by a camera.

  After superimposing the 3D graphic, the user may change the camera angle without really losing sight of the 2D planar surface underneath it.

- Video tracking:Tracking a moving item in a video is known as "video tracking."

  The goal of video tracking is to connect or link target items as they occur in each frame of the video. In other words, video tracking involves progressively scanning the video frames and anticipating and building a bounding box around the item to connect its former location with its present location.

  Due to its ability to process real-time video, video tracking is frequently utilized in security, self-driving automobiles, and traffic monitoring.

3.5.2 4 stages of the Object Tracking process

- Target initialization: Identifying the target(s) of interest is the initial stage.

  In the first frame of the video, it includes the act of drawing a bounding box around it. The next step is for the tracker to forecast or estimate the object's position in the remaining frames while drawing the bounding box concurrently.

- Appearance modeling:The visual appearance of the object is analyzed in appearance modelling. The look of the targeted object may change as it moves through different scenes, such as illumination, angle, speed, etc. This could cause disinformation causing the algorithm to lose track of the object.

  It is necessary to do appearance modelling so that the numerous changes and distortions brought about as the target object moves can be captured by the modelling algorithms.

- Motion estimation:The ability of the model to accurately forecast an object's future position is typically implied by motion estimate.

- Target positioning:Motion estimation gives a rough idea of the area where the object is most likely to be found. Once the object's location has been roughly determined, we can utilise a visual model to pinpoint the target's precise location.

3.5.3 Levels of Object Tracking

Object tracking can be described in two levels:

a . Single Object Tracking

Instead of tracking several objects, Single Object Tracking (SOT) seeks to track one object from a single class. It is also known as Visual Object Tracking on occasion.

The target object's bounding box in SOT is established in the opening frame.The objective of the method is to locate the same item in the subsequent frames.

SOT falls under the category of detection-free tracking because the tracker requires the user to manually supply the first bounding box. Because of this, even if a trained

categorization model doesn't exist for an object, Single Object Trackers should be able to track it.

b . Multiple Object Tracking

The method known as "multiple object tracking" (MOT) uses a tracking algorithm to follow each and every object of interest in the video.

The tracking method first establishes how many items will be in each frame, and then it tracks each object's identification from one frame to the next until it leaves the frame.

3.6 The centroid tracking algorithm

The centroid tracking algorithm involves several stages.They are

3.6.1 Calculate the centroids by using bounding box's coordinates.

The centroid tracking method assumes that each recognised object's (x, y) coordinates are provided for each frame.

Any object detector we choose (colour thresholding + pattern extraction, Haar cascades, HOG + linear SVM, SSDs, Faster R-CNNs, etc.) may provide these bounding boxes as long as they compute them for every frame of the video.

We should determine the "centroid," or more accurately, the center(x, y)-coordinates of the bounding box, once we obtain the bounding box coordinates. Since these initial bounding boxes are the first ones that our algorithm has seen, we will give each one a unique ID, as shown in figure 3.4.

3.6.2 Determine the Euclidean distance between newly created bounding boxes and preexisting item.

If it unable to connect the newly computed object centroids (yellow) with the previously computed object centroids (red) after Step 1 of computing object centroids for each frame in our video stream, object tracking would be successful (purple). Then,

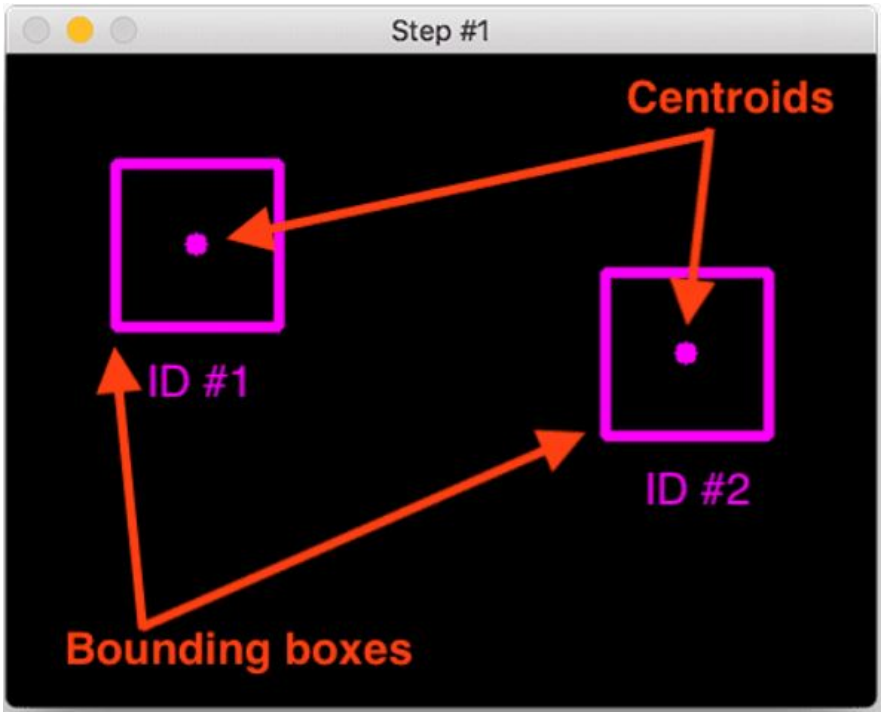

Figure 3.4: Accepting bounding box values from an object detector and calculating centroids.

for each pair of input and existing item centroids, we compute the Euclidean distance between them (depicted by green arrows). Then, each pair of the new centroids (green) and the old centroids (yellow) are separated by their Euclidean distances (shown in figure 3.5).

3.6.3 Modify the object's (x, y) coordinates that already exist

The essential tenet of the centroid tracking method is that, even if an object may move among two frames, the distance between its centroids in frames Ft and Ft + 1 will always be less than all other distances.

In light of this, suppose here choose to build an object tracker by connecting centroids to the shortest distances between subsequent frames as shown in Figure 3.6

3.6.4 Adding new objects

Input detections must be recorded if there are more than the number of monitored objects. Adding a new object to our database of monitored items by ”registering” just

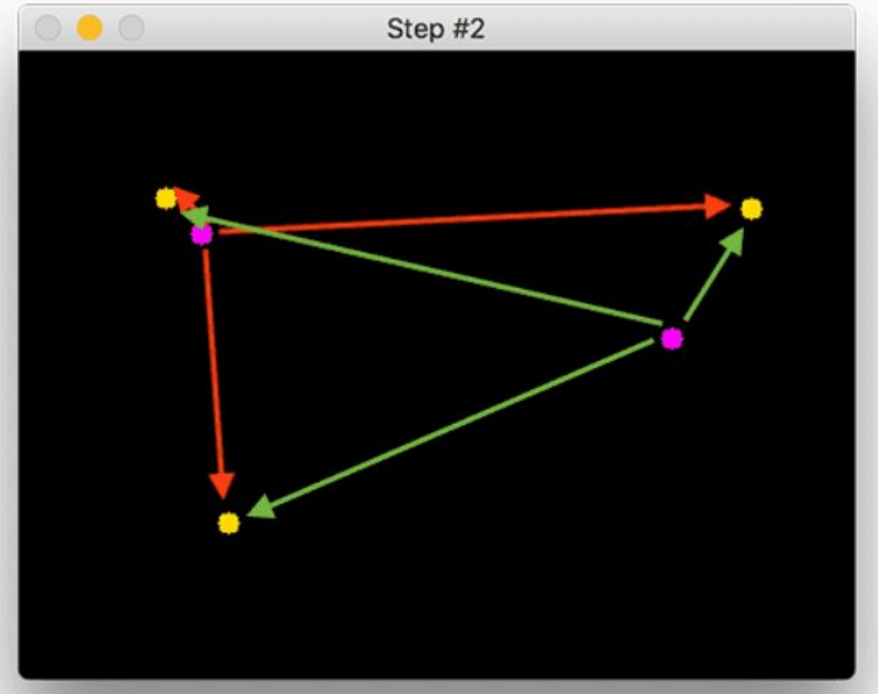

Figure 3.5: The distance measure between every pair of the original item and the updated item.

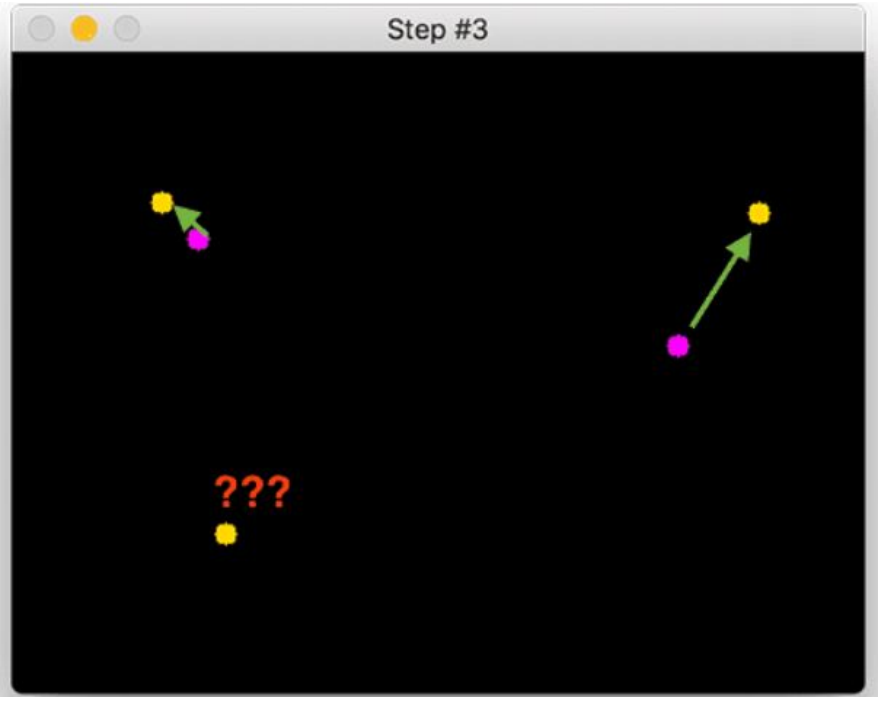

Figure 3.6: Using the minimum distances amongst frames to connect centroids

means doing that.

- Giving it a new object ID
- maintaining the bounding box centroid for the item

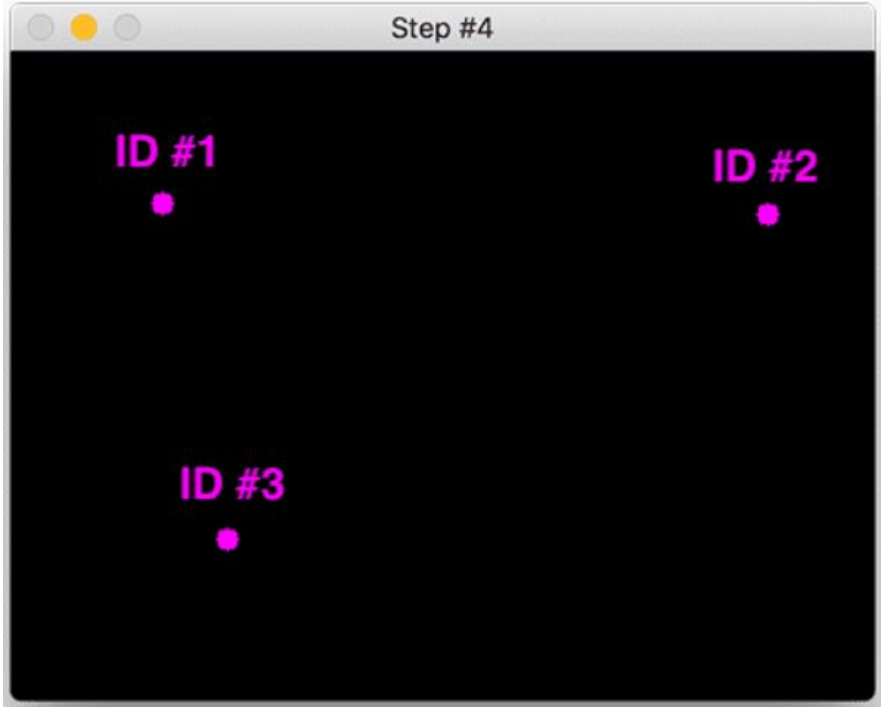

Figure 3.7: Registration of a new item

Figure 3.7 represents the creation of a new object with object ID 3

3.6.5 Deregister older items

Any system that tracks objects logically must be able to deal with circumstances when an object has disappeared, become misplaced, or gone out of sight.

In contrast, for the sake of the approach, here deregister old objects when they cannot be matched to any other objects for a total of N subsequent frames. The deployment location of the object tracker truly affects how the system responds to certain situations.

3.7 Distance Calculation Between Detected Objects

Python's Dist() method is used to calculate the Euclidean distance between two points, p and q, each of which is specified as a list of coordinates. The dimensions of the two points must match.

The length of a line segment connecting any two points in Euclidean space is known as the Euclidean distance between them. Since it can be calculated from the

points' Cartesian coordinates using the Pythagorean theorem, it is also referred to as the Pythagorean distance. The minimum distance between two pairs of points from the two items is typically used to determine the distance between two objects that are not point is common knowledge to use formulas to determine distances between various types of objects, such as the separation between a point and a line.

The formula for the Euclidean distance says

d = [(x22 – x11)2 + (y22 – y11)2]

Figure 3.8 represents Euclidean distance between two points where, The coordinates of one point are (x11, y11). The other point's coordinates are (x22, y22). The distance between (x11, y11) and (x22, y22) is represented by d.

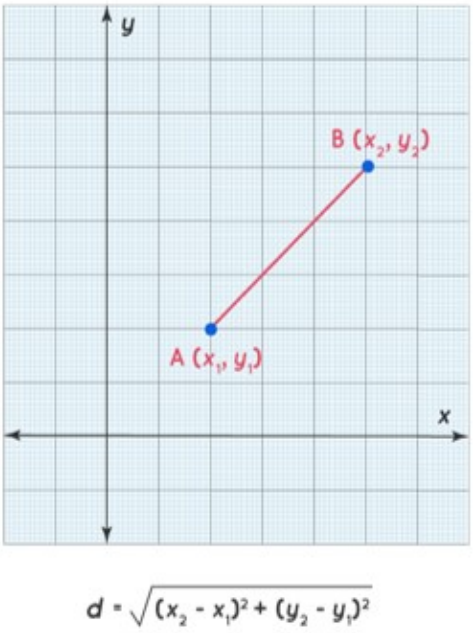

Figure 3.8: Distance calculation between two points

3.8 Dataset Used

3.8.1 COCO (Microsoft Common Objects in Context)

Microsoft Common Objects in Context, or MS COCO, is a large dataset for object recognition, segmentation, key-point detection, and captioning. In the dataset, there are 328K photos. In 2014, the MS COCO dataset saw its initial release. Total photographs

are 164K, with 83K being training images, 41K being validation images, and 41K being test shots. The 2015 release of the 81K image additional test set contained all of the earlier test photos in addition to 40K new ones.

In 2017, the validation/training split was modified from 83K/41K to 118K/5K based on community feedback. The new split makes use of the same images and annotations. The 2017 test set is made up of a subset of the 41K photographs from the 2015 test set. A brand-new, unannotated dataset of 123K images is also included in the 2017 version..

The COCO Object Detection Task is intended to advance object detection technology. Two object detection tasks are available in COCO: object segmentation output or bounding box output (the latter is also known as instance segmentation). the 80 object categories and more than 200,000 photos in the COCO train, validation, and test sets. A thorough segmentation mask is annotated on each object instance. Public annotations on the training and validation sets (with more than 500,000 segmented object instances) are provided.

3.8.2 Annotations:

Annotations to the dataset are for

- object detection:80 item categories are included in the bounding boxes and per-instance segmentation masks.
- captioning:images with explanations in natural language
- keypoints detection has almost 200,000 images and 25,000 instances of persons with keypoint labeling (17 potential key points, including left eye, nose, right hip, and right ankle).
- stuff image segmentation - 91 thing categories with per-pixel segmentation masks, including grass, wall, and sky.
- panoptic: entire scene segmentation, with a subset of 91 stuff categories and 80 thing categories (such as human, bicycle, and elephant) (grass, sky, road)

- dense pose:More than 39,000 images and 56,000 person instances had Dense-Pose annotations added to them. For each labelled person, a mapping between the image pixels that make up that person's body and a template 3D model has been produced. The annotations are only made public for the training and validation photos.

Chapter 4

RESULTS

AN AI BASED SOCIAL DISTANCE MONITORING SYSTEM detects the social distance violation in a video frame or an image.After the Detection the model places a bounding box around the people and assign red or green color to the box according to violation. To detect people in video footage, the framework employs the YOLO object detection approach. The recognition and tracking of individuals in both indoor and outdoor environments is done using a deep learning detection approach.After reaching a particular limit of violation on a frame, the model displays a alert message notifying about the violation and automatically sends a mail notification to the authorities warning them about the violation.The model got detection accuracy of 95 percentage.The method can be used for a low-cost embedded device with a fixed camera. The method can be used to watch individuals from various cameras in a centralized surveillance system using a distributed CCTV system. This method is appropriate for establishing a surveillance system in smart cities to find individuals, categorize them, and assess social distance.

4.1 Screenshots

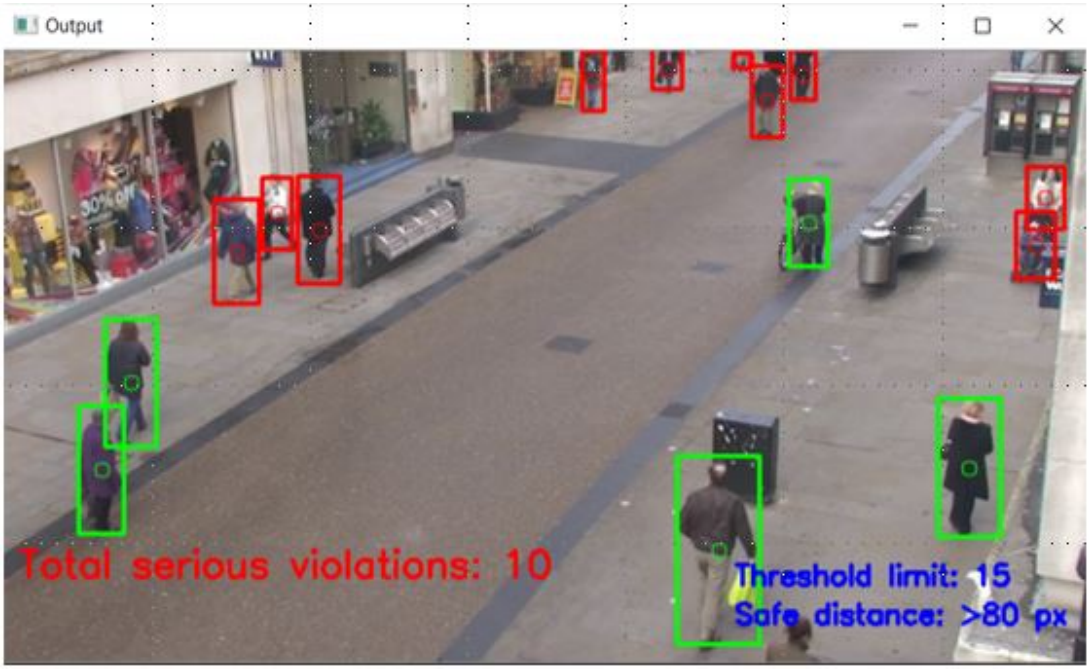

Figure 4.1: Detection of people and drawing bounding box over them

Figure 4.2: Alert message will be shown after the violation reaches a particular limit

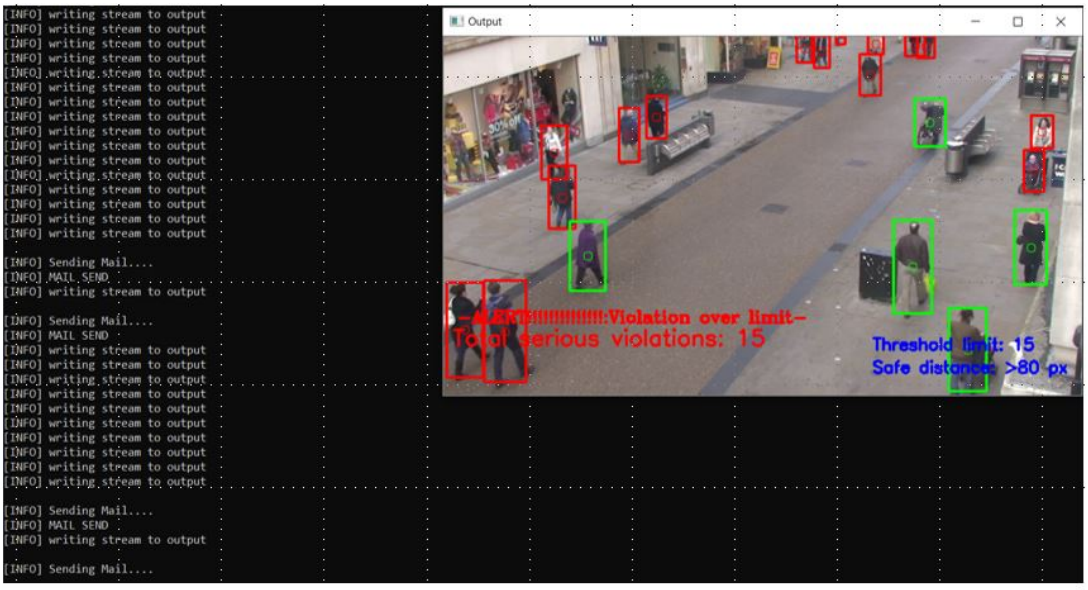

Figure 4.3: Sending mail notification

Chapter 5

CONCLUSION

This work utilizes a deep learning-based methodology for social distance monitoring. Pre-trained YOLOv5 algorithm is used to detect pedestrians. The centroid distances between both the established bounding boxes should then be computed using the Euclidean distance. A limit and an approximation of the social distance to the pixel are used to examine the social distance breaches among individuals. An alert notification will be sent to the authorities advising them of the infringement if the distance value exceeds the minimal social distance that was chosen, according to a violation threshold.

5.1 Future Enhancement

• Future iterations of the work might make it more suitable for various indoor and outdoor settings.

• To assist find the violator or violators, several detection and tracking techniques may be utilised.

REFERENCES

[1] Lorenzo Bertoni , Sven Kreiss , and Alexandre Alahi,"Perceiving Humans: From Monocular3D Localization to Social Distancing,IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 23, NO. 7, JULY 2022

[2] Sergio Saponara,Abdussalam Elhanashi,Alessio Gagliardi."Implementing a real-time, AI-based, people detection and social distancing measuring system for Covid-19",Journal of Real-Time Image Processing (2021) 18:1937–1947 https://doi.org/10.1007/s11554-021-01070-6

[3] ZHIMING CHEN1 , TINGXIANG FAN2 , XUAN ZHAO3 , JING LIANG 4 , CONG SHEN 5 , HUA CHEN 1 , DINESH MANOCHA 4 , JIA PAN2 , AND WEI ZHANG 1,"Autonomous Social Distancing in Urban Environments Using a Quadruped Robot", IEEE Access(Digital Object Identifier 10.1109/ACCESS.2021.3049426)

[4] Zhenfeng Shao , Gui Cheng , Jiayi Ma , Zhongyuan Wang , Jiaming Wang and Deren Li,"Real-Time and Accurate UAV Pedestrian Detection for Social Distancing Monitoring in COVID-19 Pandemic,IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 24, 2022.

[5] Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1–9). W. C. D. C. Dashboard (Online). https://covid19.who.int/ (Accessed 23 August 2020). W.H. Organization (2020).

[6] Yash Chaudhary, D. G., Mehta, M. (2020). In 22nd international conference on E-health networking, applications and services (IEEE Healthcom 2020).