OpenCV Social Distancing Detector

ANIL DNYANDEV SAPKAL (49)
CHAITANYA RATHOD (78)
AJMAL SHAIKH (59)

Guide:

Prof. Mohd. Ashfaq

Abstract

Social distancing is a recommended solution by the **World Health Organisation (WHO)** to minimise the spread of COVID-19 in public places. The majority of governments and national health authorities have set the 2-m physical distancing as a mandatory safety measure in shopping centres, schools and other covered areas. In this research, we develop a hybrid Computer Vision and YOLOv4-based Deep Neural Network (DNN) model for automated people detection in the crowd in indoor and outdoor environments using common CCTV security cameras. The proposed DNN model in combination with an adapted inverse perspective mapping (IPM) technique and SORT tracking algorithm leads to a robust people detection and social distancing monitoring. The model has been trained against two most comprehensive datasets by the time of the research—the Microsoft Common Objects in Context (MS COCO) and Google Open Image datasets. The system has been evaluated against the Oxford Town Centre dataset (including 150,000 instances of people detection) with superior performance compared to three state-of-the-art methods. The evaluation has been conducted in challenging conditions, including occlusion, partial visibility, and under lighting variations with the mean average precision of 99.8% and the real-time speed of 24.1 fps. We also provide an online infection risk assessment scheme by statistical analysis of the spatio-temporal data from people's moving trajectories and the rate of social distancing violations. We identify high-risk zones with the highest possibility of virus spread and infection. This may help authorities to redesign the layout of a public place or to take precaution actions to mitigate high-risk zones. The developed model is a generic and accurate people detection and tracking solution that can be applied in many other fields such as autonomous vehicles, human action recognition, anomaly detection, sports, crowd analysis, or any other research areas where the human detection is in the centre of attention.

Keywords: social distancing; COVID-19; human detection and tracking; distance estimation; deep convolutional neural networks; crowd monitoring; pedestrian detection; inverse perspective mapping**.**

Introduction

The novel generation of the coronavirus disease (COVID-19) was reported in late December 2019 in Wuhan, China. After only a few months, the virus became a global outbreak in 2020. On May 2020 The World Health Organisation (WHO) announced the situation as pandemic [1,2]. The statistics by WHO on 8th October 2020 confirm 36 million infected people and a scary number of 1,056,000 deaths in 200 countries. With the growing trend of patients, there is still no effective cure or available treatment for the virus. While scientists, healthcare organisations, and researchers are continuously working to produce appropriate medications or vaccines for the deadly virus, no definite success has been reported at the time of this research, and there is no certain treatment or recommendation to prevent or cure this new disease. Therefore, precautions are taken by the whole world to limit the spread of infection. These harsh conditions have forced the global communities to look for alternative ways to reduce the spread of the virus. Social distancing, as shown in Figure 1a, refers to precaution actions to prevent the proliferation of the disease, by minimising the proximity of human physical contacts in covered or crowded public places (e.g., schools, workplaces, gyms, lecture theatres, etc.) to stop the widespread accumulation of the infection risk (Figure 1b).

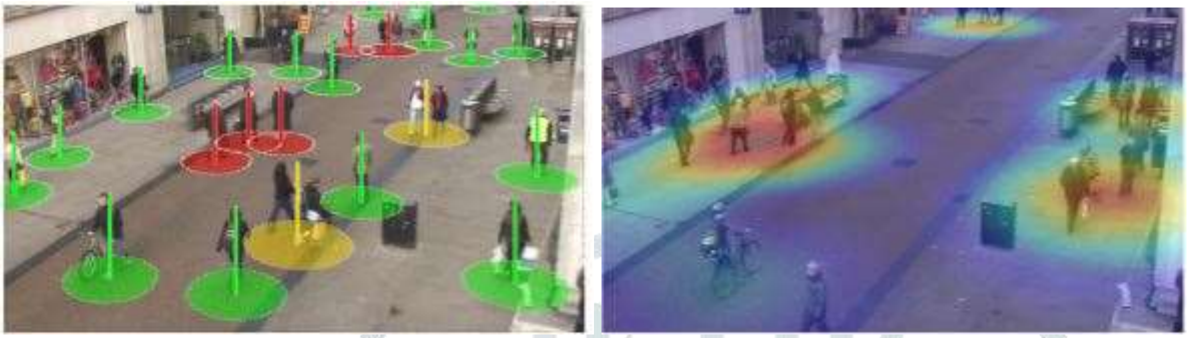


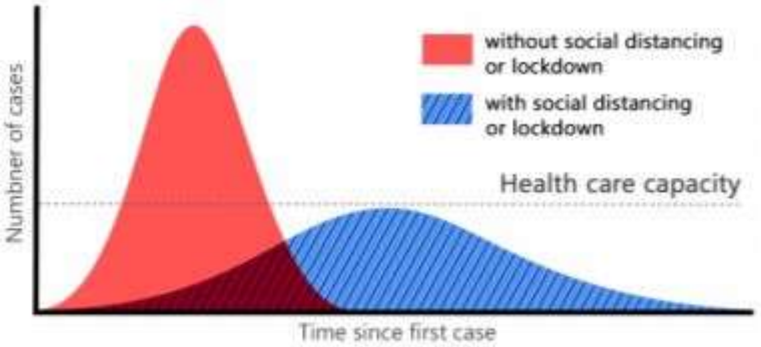
Figure 1. People detection, tracking, and risk assessment in Oxford Town Centre, using a public CCTV camera. (a) Social distancing monitoring; (b) Accumulated infection risk (red zones)

For several months, the World Health Organisation believed that COVID-19 was only transmittable via droplets emitted when people sneeze or cough and the virus does not linger in the air. However, on 8 July 2020, the WHO announced:

“There is emerging evidence that COVID-19 is an airborne disease that can be spread by tiny particles suspended in the air after people talk or breathe, especially in crowded, closed environments or poorly ventilated settings” [2].

Therefore, social distancing now claims to be even more important than thought before, and one of the best ways to stop the spread of the disease in addition to wearing face masks. Almost all countries are now considering it as a mandatory practice. According to the defined requirements by the WHO, the minimum distance between individuals must be at least 6 feet (1.8 m) in order to observe an adequate social distancing among the people [3]. Recent research has confirmed that people with mild or no symptoms may also be carriers of the novel coronavirus infection [4]. Therefore, it is important all individuals maintain controlled behaviours and observe social distancing. Many research works such as [5–7] have proved social-distancing as an effective non-pharmacological approach and an important inhibitor for limiting the transmission of contagious diseases such as H1N1, SARS, and COVID-19. Figure 2 demonstrates the effect of following appropriate social distancing guidelines to reduce the rate of infection transmission among individuals [8,9]. A wider Gaussian curve with a shorter spike within the range of the health system service capacity makes it easier for patients to fight the virus by receiving continuous and timely support from the health care organisations. Any unexpected sharp spike and rapid infection rate (such as the red curve in Figure 2), will lead to service failure, and consequently, exponential growth in the number of fatalities. During the COVID-19 pandemic, governments have tried to implement a variety of social distancing practices, such as restricting travels, controlling borders, closing pubs and bars, and alerting the society to maintain a distance of 1.6 to 2 m from each other [10].

However, monitoring the amount of infection spread and efficiency of the constraints is not an easy task. People require to go out for essential needs such as food, health care and other necessary tasks and jobs. Therefore, many other technology-based solutions such as [11,12] and AI related research such as [13–15] have tried to step in to help the health and medical community in copping with COVID-19 challenges and successful social distancing practices. These works vary from GPS-based patient localisation and tracking to segmentation, and crowd monitoring.



In such situations, Artificial Intelligence can play an important role in facilitating social distancing monitoring. Computer Vision, as a sub-field of Artificial Intelligence, has been very successful in solving various complex health care problems and has shown its potential in chest CT-Scan or X-ray based COVID-19 recognition [16,17] and can contribute to Social-distancing monitoring as well. Besides, deep neural networks enable us to extract complex features from the data so that we can provide a more accurate understanding of the images by analysing and classifying these features. Examples include diagnosis, clinical management and treatment, as well as the prevention and control of COVID-19 [18,19].

Possible challenges in this area are the importance of gaining a high level of accuracy, dealing with a variety of lighting conditions, occlusion, and real-time performance. In this work, we aim at providing solutions to cope with the mentioned challenges, as well. The main contribution of this research can be highlighted as follows:

- This study aims to support the reduction of the coronavirus spread and its economic costs by providing an AI-based solution to automatically monitor and detect violations of social distancing among individuals.
- We develop a robust deep neural network (DNN) model for people detection, tracking, and distance estimation called DeepSOCIAL (Sections 3.1–3.3). In comparison with some recent works in this area, such as [15], we offer faster and more accurate results
- We perform a live and dynamic risk assessment, by statistical analysis of spatio-temporal data from the people movements at the scene (Section 4.4). This will enable us to track the moving trajectory of people and their behaviours, to analyse the ratio of the social distancing violations to the total number of people in the scene, and to detect high-risk zones for short- and long-term periods.
- We back up the validity of our experimental results by performing extensive tests and assessments in a diversity of indoor and outdoor datasets which outperform the state-of-the-arts
- The developed model can perform as a generic human detection and tracker system, not limited to social-distancing monitoring, and it can be applied for various real-world applications such as pedestrian detection in autonomous vehicles, human action recognition, anomaly detection, and security systems.

Literature Survey

Survey Existing system :

Reuters spoke with 16 video analytics companies, many of them startups with a few million dollars in annual revenue, that have added offerings because of the coronavirus. Their systems can be set to produce daily reports, which site managers can use to correct recurring problems and document compliance.

Most work on a branch of AI technology known as computer or machine vision in which algorithms are trained on image libraries to identify objects with confidence of 80% or higher.

Several customers said the technology, which can cost $1,000 or more annually to analyze data from a handful of off-the-shelf video cameras, is cheaper than dedicating staff to standing guard. It also can be safer, as some guards enforcing distancing have clashed with people protesting safety measures, they said.

Pepper Construction’s Suerth said its SmartVid system has not flagged crowding issues yet because staffing has been limited. But Suerth said that as more crews arrive, the company will look at trends to issue reminders at “tool box talks.”

“It’s another set of eyes on the site,” Suerth said, adding that software is less prone to mistakes than people and the “accuracy we’re seeing is really high.”

Samarth Diamond manager Parth Patel said he could adjust procedures when the software identifies spots where his 4,000 workers are clumping together in busy areas. People tagged as not having masks quickly would be offered one by a team reviewing camera feeds, Patel said.

“It will surely be helpful for the safety of employees and their comfort level, and it will be helpful to show it to authorities that we are adhering” to regulations, Patel said.

Patel said he has confidence in the algorithms after his family successfully used computer vision last year at supermarkets it owns to count female shoppers and decide where to stock a new line of dresses.

RPT Realty, which Chief Executive Brian Harper said had used camera software to count visitors over the past few months at two of the 49 open-air shopping centers it owns in the United States, is moving to assess tenants’ compliance with reduced occupancy regulations across five malls.

It also plans to help consumers decide when to shop by using technology from startup WaitTimes to analyze lines of people waiting to enter stores, a phenomenon that has become common during the pandemic as part of social distancing efforts. Signage will inform shoppers of the anonymous counting, according to Harper.

But calculating whether people are six feet (1.8 meters) apart and detecting objects such as face masks are all novel uses now being tested and launched on accelerated schedules. Some startups even promise to spot sneezing and coughing, claims that drew skepticism from some experts.

"Most solutions will be in uncharted territory, without a proven track record, and likely susceptible to false-positives and bugs," said Vinay Goel, a former Google Maps product leader who is now chief digital products officer at the tech unit of real estate services giant Jones Lang LaSalle Inc.

Beside costs, businesses are concerned AI will trigger too many reports of non-problems, like a family walking close together in an aisle, retail consultants said.

Indyme, a technology vendor that works with BevMo!, Office Depot and other U.S. retailers, said that its clients have preferred rudimentary boxes that can count people at entrances and automatically announce, “For your safety, please maintain a social distance of six feet, thank you.”

2. Limitation Existing system or research gap :

In this section, we provide a brief literature review on three types of research in this area: medical and clinical-related research, tracking technologies, and AI-based research. Although our research falls in the AI research category, due to the nature of the research questions, first, we will have a brief review on medical and technology-based research to have an in-depth understanding about the existing challenges. In Section 2.3 (AI-based Research) we gradually transit from object detection techniques to people detection, the existing methodologies, and research gaps for people detection using AI and computer vision.

2.1. Medical Research :

Many researchers in the medical and pharmaceutical fields are aiming at treatment of COVID-19 infectious disease; however, no definite solution has yet been found. One the other hand, controlling the spread of the virus in public places is another issue, where the AI, computer vision, and technology can step-in to help.A variety of studies with different implementation strategies [5,6,20] have proven that controlling the prevalence is a contributing factor, and social distancing is an effective way to reduce the transmission and prevent the spread of the virus in society. Several researchers such as [20,21] use Susceptible, Infectious, or Recovered (SIR) model. SIR is an epidemiological modelling system to compute the theoretical number of people infected with a contagious disease in a given population, over time. One of the oldest yet common SIR models is Kermack and McKendrick models introduced in 1927 [22]. Eksin et al. [21], have recently introduced a modified model of SIR by including a social distancing parameter, which can be used to determine the number of infected and recovered individuals. Effectiveness of social distancing practices can be evaluated based on several standard approaches. One of the main criteria is based on the reproduction ratio, Ro, which indicates the average number of people who may be infected from an infectious person during the entire period of the infection [23]. Any Ro

> 1 indicates an increasing rate of infection within the society and Ro < 1 indicates that every case will infect less than 1 person, hence, the disease rate is considered to be declining in the target population.Since the Ro value indicates the disease outspread, it is one of the most important indicators for selecting social distancing criteria. In the current COVID-19 pandemic, the World Health Organisation estimated the Ro rate would be in the range of 2–2.5 [14], which is significantly higher than other similar diseases such as seasonal flu with Ro = 1.4. In [11], a clear conclusion is drawn about the importance of applying social distancing for cases with a high amount of Ro. In another research based on the game theory on the classic SIR model, an assessment of the benefits and economic costs of social distancing has been examined [24]. The results also show that in the case of Ro < 1, social distancing would cause unnecessary costs, while Ro ≈ 2 implies that social distancing measures have the highest economic benefits. In another similar research, Kylie et al. [25] investigated the relationship between the stringency of social distancing and the region’s economic status. This study suggests although preventing the widespread outbreak of the virus is necessary, a moderate level of social activities could be allowed. Prem et al. [26] use location-specific contact patterns to investigate the effect of social distancing measures on the prevalence of COVID-19 pandemic in order to remove the persistent path of disease outbreak using susceptible-exposed-infected-removed models (SEIR).

2.2. Tracking Technologies :

Since the onset of coronavirus pandemic, many countries have used technology-based solutions, to inhibit the spread of the disease [12,27,28]. For example, some of the developed countries, such as South Korea and India, use GPS data to monitor the movements of infected or suspected individuals to find any possible exposure among the healthy people.

The India government uses the Aarogya Setu program to find the presence of COVID-19 patients in the adjacent region, with the help of GPS and Bluetooth. This may also help other people to maintain a safe distance from the infected person [29]. Some law enforcement agencies use drones and surveillance cameras to detect large-scale rallies and have carried out regulatory measures to disperse the population [30,31].

Other researchers such as Xin et al. [32] perform human detection using wireless signals by identifying phase differences and change detection in amplitude wave-forms. However, this requires multiple receiving antennas and can not be easily integrated in all public places.

2.3. AI-Based Research :

The utilisation of Artificial Intelligence, Computer Vision, and Machine Learning, can help to discover the correlation of high-level features. For example, it may enable us to understand and predict pedestrian behaviours in traffic scenes, sports activities, medical imaging, or anomaly detection, by analysing spatio-temporal visual information and statistical data analysis of the images sequences [13,19].

Among AI-Health related works, some researchers have tried to predict the sickness trend of specific areas [33], to develop crowd counting and density estimation methodologies in public places [34], or to determine the distance of individuals from the popular swarms [35] using a combination of visual and geo-location cellular information. However, such research works suffer from challenges such as skilled labour or the cost of designing and implementing the infrastructures.

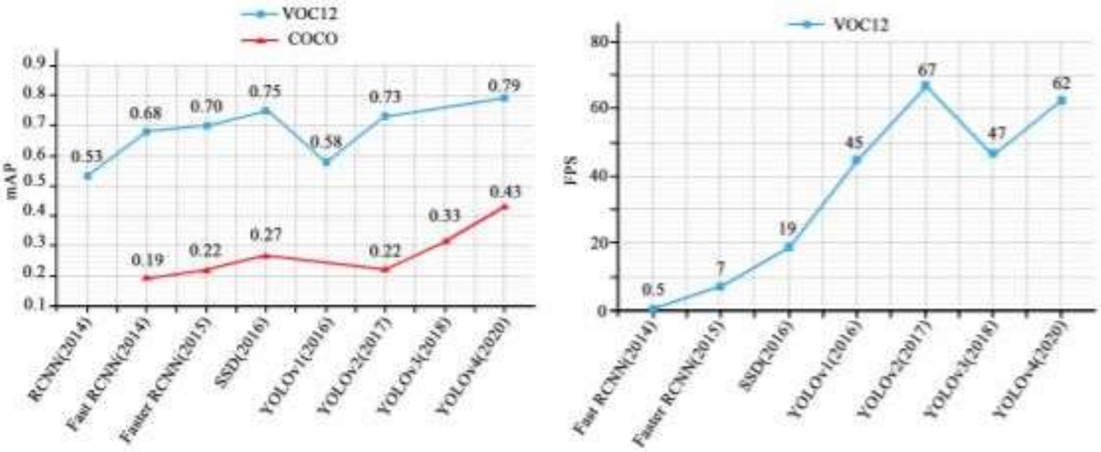
On the other hand, recent advances in Computer Vision, Deep Learning, and pattern recognition, as the sub-categories of the AI, enable the computers to understand and interpret the visual data from digital images or videos. It also allows computers to identify and classify different types of objects [36–38]. Such capabilities can play an important role in empowering, encouraging, and performing social distancing surveillance and measurements as well. For example, Computer Vision could turn CCTV cameras in the current infrastructure capacity into “smart” cameras that not only monitor people but can also determine whether people follow the social distancing guidelines or not. Such systems require very precise human detection algorithms.

People detection in image sequences is one of the most important sub-branches in the field of object detection and computer vision. Although many research works have been done in human detection [39] and human action recognition [40], the majority of them are either limited to indoor applications or suffer from accuracy issues under outdoor challenging lighting conditions. A range of other research works rely on manual tuning methodologies to identify people activities, however, limited functionality has always been an issue [41].

Convolutional Neural Networks (CNNs) have played a very important role in feature extraction and complex object classification, including human detection. With the development of faster CPUs, GPUs, and extended memory capacities, CNNs allow the researchers to make accurate and fast detectors compared to conventional models. However, the long time training, detection speed and achieving better accuracy, are still remaining challenges to be solved. Narinder et al. [15] used a deep neural network (DNN) based detector, along with Deepsort [42] algorithm as an object tracker for people detection to assess the distance violation index—the ratio of number of people who violated the social distancing measure to the total number of the assessed group. However, no statistical analysis of the outcome of their results is provided. Furthermore, no discussion about the validity of the distance measurements is provided.

In another study by Khandelwal et al. [43], the authors have addressed the people distancing in a given manufactory. They have used MobileNet V2 network [44] as a lightweight detector to reduce computational costs, which in turn provides less accuracy comparing to some other common models. Furthermore, the method only focuses on an indoor manufactory-setup distance measurement and does not provide any statistical assessment on the virus spread. Similar to the other research, no statistical analysis is performed on the results of the distance measurement in [45]. The authors have made a comparison between two common types of DNN models (You Only Look Once—YOLO and Faster RCNN). However, the system accuracy has been only estimated based on a shallow comparison on different datasets with non-comparable ground truths.

Figure, shows the outcome of our investigations and reviews in terms of mean Average Precisions (mAP), and the speed (Frame Per Second—FPS) on some of the most successful object detection models such as RCNN [46], fast RCNN [47], faster RCNN [48], SSD: Single Shot MultiBox Detector [49], YOLOv1-v4 [50–53] tested on the Microsoft Common Objects in Context (MS COCO) [54] and PASCAL Visual Object Classes (VOC) [55] data sets under similar conditions. Otherwise, the performance of the systems may vary depending on various factors such as backbone architecture, input image size, resolution, model depth, software, and hardware platform.



As can be seen from Figure 3, some of the models such as SSD and YOLOv2 perform in a contradictory manner in dealing with COCO and VOC12 datasets. They may seem good in one, and weak in another one. One of the possible reasons could be attributed to the different number of object categories in COCO and VOC12 (80 categories vs. 20). This makes the VOC12 dataset an easier goal to learn and less challenging. However, when it comes to the higher number of classes, the performance of the system may seem irregular, depending on the feature complexity of each object (good in some detections and weak in some other). Since the social distancing topic is very recent, there has not been much dedicated research regarding the accuracy of people detection and inter-people distance estimation in the crowd, no experiment on challenging datasets has been performed, no standard comparison has been conducted on common datasets, and no analytical studies or post-processing have been considered after the people detection-phase to analyse the risk of infection distribution.

3. Problem Statement and Objective :

Our objectives are Deep Neural Network-Based human detector model called DeepSOCIAL to detect and track static and dynamic people in public places in order to monitor social distancing metrics in COVID-19 era and beyond. Various types of state-of-the-art backbones, necks, and heads were evaluated and investigated. We utilised a CSPDarkNet53 backbone along with an SPP/PAN and SAM neck, YOLO head, Mish activation function. We applied the Complete IoU loss function and a Mosaic data augmentation on multi-viewpoint MS COCO and Google Open Image datasets to enrich the training phase, which ultimately led to an efficient and accurate human detector, applicable in various environments using any type of CCTV surveillance cameras.

The proposed method was evaluated for Oxford Town Centre dataset, including 7530 frames, and approximately 150,000 people detection and distance estimation. The system was able to perform in a variety of challenges including, occlusion, lighting variations, shades, and partial visibility, and proved a major development in terms of accuracy (99.8%) and speed (24.1 fps) compared to three state-of-the-art techniques. The system performed real-time using a basic GPU platform or a 10th generation multi-core/multi-thread CPU platform, or higher. We adapted an inverse perspective geometric mapping and SORT tracking algorithm for our application to estimate the inter-people distances, and to track the moving trajectories of the people, infection risk assessment and analysis to the benefit of the health authorities and governments.

DeepSOCIAL offered a viewpoint-independent human classification algorithm. Therefore, regardless of the camera angle and position, the outcome of this research is directly applicable for a wider community of researchers, not only in computer vision, AI, and health sectors but also in other industrial applications including pedestrian detection for driver assistance systems, autonomous vehicles, anomaly behaviour detections in public and crowd, surveillance security systems, action recognition in sports, shopping centres, public places; and generally, any applications that human detection falls in the centre of attention.

4. Scope :

Since this application is intended to be used in any working environment; accuracy and precision are highly desired to serve the purpose. Higher number of false positive may raise discomfort and panic situation among people being observed. There may also be genuinely raised concerns about privacy and individual rights which can be addressed with some additional measures such as prior consents for such working environments, hiding a persons identity in general, and maintaining transparency about its fair uses within limited stakeholders

The above mentioned use cases are only some of the many features that were incorporated as part of this solution. We assume there are several other cases of usage that can be included in this solution to offer a more detailed sense of safety. Several of the currently under development features are listed below in brief:

1) Coughing and Sneezing Detection: Chronic coughing and sneezing is one of the key symptoms of COVID-19 infection as per WHO guidelines and also one of the major route of disease spread to non-infected public. Deep learning based approach can be proved handy here to detect & limit the disease spread by enhancing our proposed solution with body gesture analysis to understand if an individual is coughing and sneezing in public places while breaching facial mask and social distancing guidelines and based on outcome enforcement agencies can be alerted.

2) Temperature Screening: Elevated body temperature is another key symptom of COVID-19 infection, at present scenario thermal screening is done using handheld contactless IR thermometers where health worker need to come in close proximity with the person need to be screened which makes the health workers vulnerable to get infected and also its practically impossible to capture temperature for each and every person in public places, the proposed use-case can be equipped with thermal cameras based screening to analyze body temperature of the peoples in public places that can add another helping hand to enforcement agencies to tackle the pandemic effectively.

Proposed System

3.1 Analysis/Framework/ Algorithm :

1) *Artificial Intelligence* :

Artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, Leading AI define the field as the study of"intelligent agents" any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. Colloquially, the term "artificial intelligence"is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving". As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI affect "AI is whatever hasn't been done yet. For instance, optical character recognition is frequently excluded from things considered to be AI, having become a routine technology. Modern machine capabilities generally classified as AI include understanding human speech, competing at the highest, autonomously operating cars, intelligent routing in content delivery networks, and military simulations.

2) *Python* :

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage-collected. Python is often described a language due to its comprehensive standard library. Python uses dynamic typing and a combination of reference counting and a cycledetecting garbage collector for memory management.It also features dynamic name resolution, which binds method and variable names during program execution.

3) *OpenCV* :

OpenCV Python is a library of Python bindings designed to solve computer vision problems. Python is a general purpose programming language started by Guido van Rossum that became very popular very quickly, mainly because of its simplicity and code readability. It enables the programmer to express ideas in fewer lines of code without reducing readability. Compared to languages like C/C++, Python is slower. That said, Python can be easily extendedwith C/C++, which allows us to write computationally intensive code in C/C++ and create Python wrappers that can be used as Python modules. This gives us two advantages:first,the code is as fast as the original C/C++ code (since itis the actual C++ code working in background) and second, it easier to code in Python than C/C++. OpenCV-Python makes use of Numpy, which is a highly optimized library for numerical operations with a MATLAB-style syntax.AlltheOpenCVarray structures are converted to and from Numpy arrays. This also makes it easier to integrate with other libraries that use Numpy such as SciPy and Matplotlib.

OpenCV-Python is the Python API for OpenCV, combining the best qualities oftheOpenCV C++ API and the Python language.

4) *TENSORFLOW*:

Tensor Flow is an open source library for fast numerical computing. It was created and is maintained by Google and released under the Apache 2.0 open source license. The API is nominally for the Python programming language, although there is access to the underlying C++ API. Unlike other numerical libraries intended for use in Deep Learning like Theano, Tensor Flow was designed for use both in research and development and in production systems, It can run on single CPU systems, GPUs as well as mobile devices and large scale distributed systems of hundreds of machines. Computation is described in terms of data flow and operations in the structure of a directed graph.

Nodes: Nodes perform computation and have zero or more inputs and outputs. Data that moves between nodes are known as tensors, which are multi-dimensional arrays of real values.

Edges: The graph defines the flow of data, branching, looping and updates to state. Special edges can be used to synchronize behavior within the graph, for example waiting for computation on a number of inputs to complete.

Operation: An operation is a named abstract computation which can take input attributes and produce output attributes. For example, you could define an add or multiply operation.

5) *YOLOV3* :

YOLOv3 is the latest variant of a popular object detection algorithm YOLO – You Only Look Once. The published model recognizes 80 different objects in images and videos, but most importantly it is super fast and nearly as accurate as Single Shot MultiBox (SSD). First, it divides the image into a 13×13 grid of cells. The size of these 169 cells varies depending on the size of the input. For a 416×416 input size that we used in our experiments, the cell size was 32×32. Each cell is then responsible for predicting a number of boxes in the image. For each bounding box,the network also predicts the confidence that the bounding box actually encloses an object, and the probability of the enclosed object being a particular class. Most of these bounding boxes are eliminated because their confidence is low or because they are enclosing the same object as another bounding box with very high confidence score. This technique is called non maximum suppression.

Easy integration with an OpenCV application: If your application already uses OpenCV and you simply wantto use YOLOv3, you don't have to worry about compiling and building the extra Dark net code. OpenCV CPU version is 9x faster: OpenCV's CPU implementation of the DNN module is astonishingly fast. For example, Dark net when used with OpenMP takes about 2 seconds on a CPU for inference on a single image. In contrast, OpenCV's implementation runs in a mere 0.22 seconds! Check out table below. Python support: Dark net is written in C, and it does not officially support Python. In contrast, OpenCV does.

PROPOSED SYSTEM :

The proposed system focuses on how to identify the person on image/video stream whether the social distancing is maintained or not with the help of computer vision and deep learning algorithm by using theOpenCV, Tensor flowlibrary.

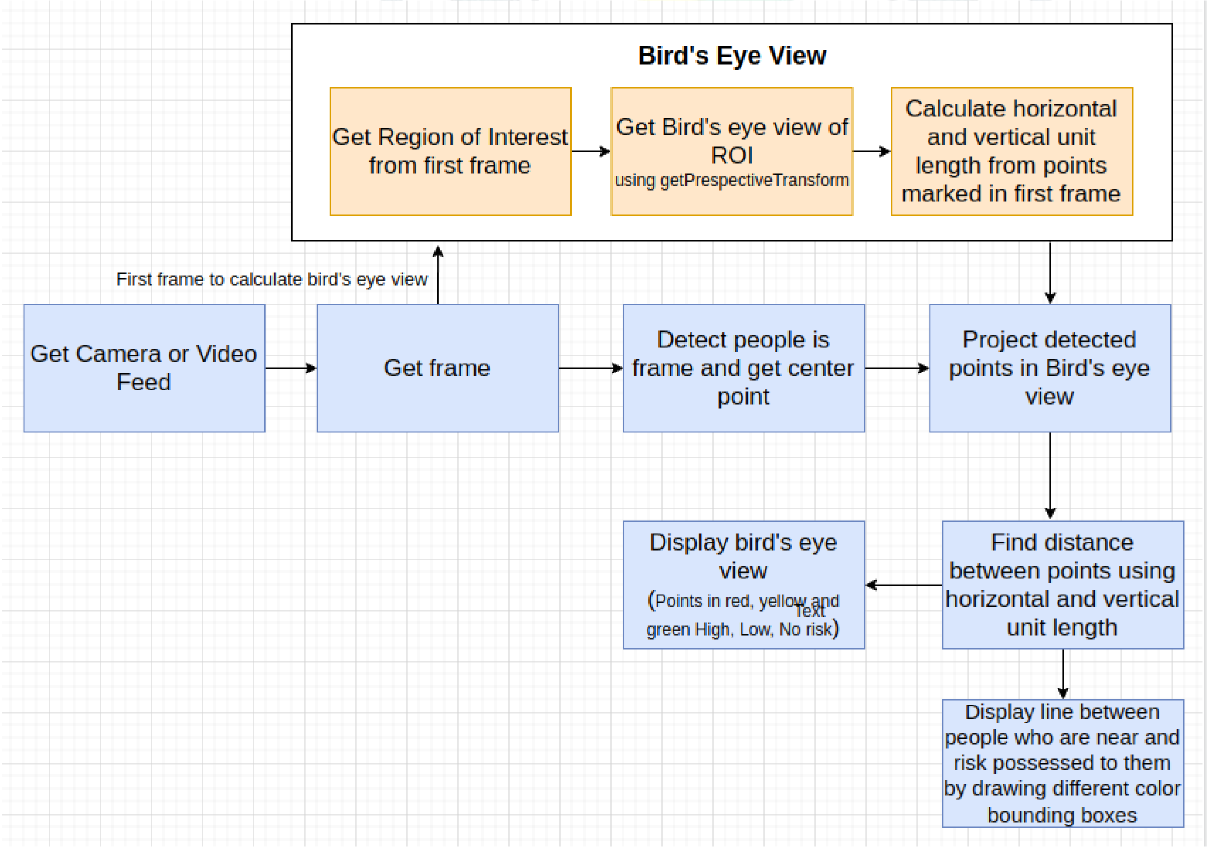
Approach

1. Detect humans in the frame with yolov3.

2. Calculates the distance between every human who is detected in the frame.

3. Shows how many people are at High, Low and Not at risk.

Camera Perspective Transformation or Camera Calibration: As the input video may be taken from an arbitrary perspective view, the first step is to transform perspective of view to a bird's-eye (top-down) view. As the input frames are monocular (taken from a single camera), the simplest transformation method involves selecting four points in the perspective view which define ROI where we want to monitor social distancing and mapping them to the corners of a rectangle in the bird's-eye view. Also these points should form parallel lines in real world if seen from above (bird's eye view). This assumes that every person is standing on the same flat ground plane. This top view or bird eye view has the property that points are distributed uniformly horizontally and vertically (scale for horizontal and vertical direction will be different).From this mapping, we can derive a transformation that can be applied to the entire perspective image.


Bird's Eye View
Get Region of Interest from first frame
Get Bird's eye view of ROI using getPrespectiveTransform
Calculate horizontal and vertical unit length from points marked in first frame
First frame to calculate bird's eye view
Get Camera or Video Feed
Get frame
Detect people is frame and get center point
Project detected points in Bird's eye view
Display bird's eye view (Points in red, yellow and green High, Low, No risk)
Text
Find distance between points using horizontal and vertical unit length
Display line between people who are near and risk possessed to them by drawing different color bounding boxes

Flow diagram of social distancing detector model

Software and libraries required :

1. **Python**
2. **Opencv**
3. **Numpy**
4. **Argparse**

3.2 Details of software design:

1. ARCHITECTURE OF PROPOSED SYSTEM:

The proposed system helps to ensure the safety of the people at public places by automatically monitoring them whether they maintain a safe social distance, and also by detecting whether or not and individual wears face mask. This section briefly describes the solution architecture and how the proposed system will automatically functions in an automatic manner to prevent the coronavirus spread.

The proposed system uses a transfer learning approach to performance optimization with a deep learning algorithm and a computer vision to automatically monitor people in public places with a camera integrated with a raspberry pi4 and to detect people with mask or no mask. We also do fine tuning, which is another form of transfer learning, more powerful than just the feature extraction.

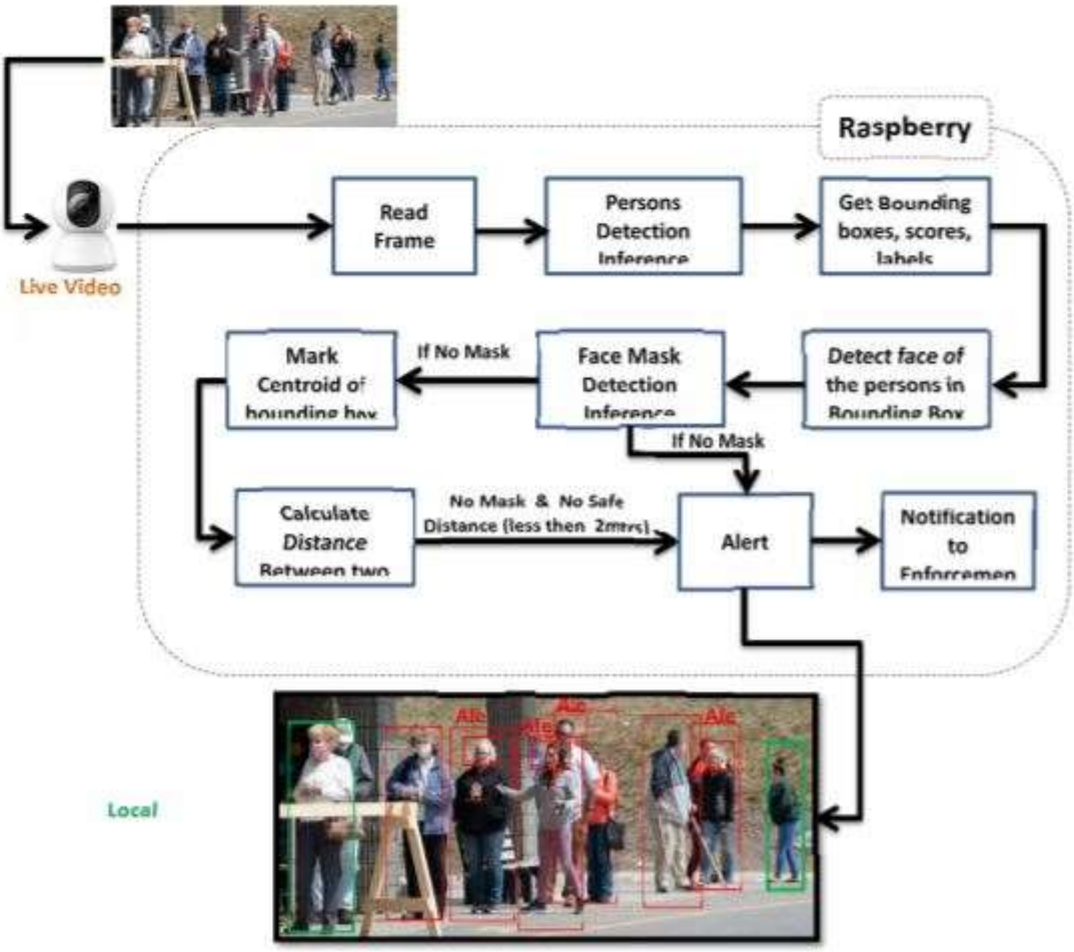

Raspberry
Live Video
Read Frame
Persons Detection Inference
Get Bounding boxes, scores, labels
Detect face of the persons in Bounding Box
Face Mask Detection Inference
If No Mask
Mark Centroid of bounding box
If No Mask
Calculate Distance Between two
No Mask & No Safe Distance (less then 2mtrs)
Alert
Notification to Enforcemen
Local

Fig 2: Solution architecture of proposed system

In this process camera video feeds from the Network Video Recorder (NVR) are streamed using RTSP and then these frames are converted to grayscale to improve speed and accuracy and are send to the model for further processing inside raspberry pi4. We have used the MobileNetV2 architecture as the core model for detection as MobileNetV2 provides a huge cost advantage compared to the normal 2D CNN model. The process also involves the SSD MultiBox Detector, a neural network architecture that has already been trained on a large collection of images such as ImageNet and PascalVOC for high quality image classification. We are loading the MobileNet V2 with pre-trained ImageNet weights, leaving the network head off and constructing a new FC head, attaching it to the base instead of the old head, and freezing the base layers of the network. The weights of these base layers will not be changed during the fine tuning phase of the backpropagation, while the head layer weights will be adjusted. After data is prepared and the model architecture is set up for fine tuning, then the model is compiled and trained. A very small learning rate is used during the retraining of the architecture to ensure that the convolutional filters already learned do not deviate dramatically and experiments have been carried out with OpenCV, TensorFlow using Deep Learning and Computer Vision in order to inspect the safe social distance between detected persons and face masks detection in real-time video streams. The main contribution of the proposed system is three components: person detection, safe distance measurement between detected persons, face mask detection. Real-time person detection is done with the help of Single Shot object Detection (SSD) using MobileNet V2 and OpenCV, achieves 91.2% mAP, outperforming the comparable state-of-the-art Faster R-CNN model. A bounding box will be displayed around every person detected. Although SSD is capable of detecting multiple objects in a frame, it is limited to the detection of a single person in this system. To calculate the distance between two persons first the distance of person from camera is calculated using triangle similarity technique, we calculate perceived focal length of camera, we assumed person distance D from camera and person's actual height H=165cms and with SSD person detection pixel height P of the person is identified using the bounding box coordinates. 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-point of the bounding boxes of all detected individuals. By doing this, we got x and y values, and these pixel values are converted into centimeters. 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 centimeters, a red bounding box is shown around them, indicating that they do not maintain a social distance. In the proposed system transfer learning is used on top of the high performing pre-trained SSD model for face detection with mobileNet V2 architecture as backbone to create a lightweight model that is accurate and computationally efficient, making it easier to deploy the model to raspberry pi. We used custom face crop datasets of about 3165 images annotated in mask and no mask. Annotated images are used to train a deep learning binary classification model that classifies the input image into the mask and no mask categories using the output class confidence. The result of the SSD model extracts a person mask and displays a bounding box. The proposed system monitors public places continuously and when a person without a mask is detected his or her face is captured and an alert is sent to the authorities with face image and at the same time the distance between individuals is measured in real time, if more than 20 persons have been identified continuously breaching safe social distance standards at the threshold time, then alert is sent to the control center at the State Police Headquarters to take further action. This system can be used in real-time applications requiring a secure monitoring of social distance between people and the detection of face masks for safety purposes due to the outbreak of Covid-19. Deploying our model to edge devices for automatic monitoring of public places could reduce the burden of physical monitoring, which is why we choose to use this architecture. This system can be integrated with edge device for use in airports, railway stations, offices, schools and public places to ensure that public safety guidelines are followed.

3.3 Methodology :

We propose a three-stage model including people detection, tracking, inter-distance estimation as a total solution for social distancing monitoring and zone-based infection risk analysis. The system can be integrated and applied on all types of CCTV surveillance cameras with any resolution from VGA to Full-HD, with real-time performance.

People Detection:

Figure 4 shows the overall structure of the Stage 1. A CCTV Camera collects the input video sequences, and passes them to our Deep Neural Network model. The output of the model would be the detected people in the scene with their unique localisation bounding boxes. The objective is to develop a robust human (people) detection model, capable of dealing with various types of challenges such as variations in clothes, postures, at far and close distances, with/without occlusion, and under different lighting conditions.

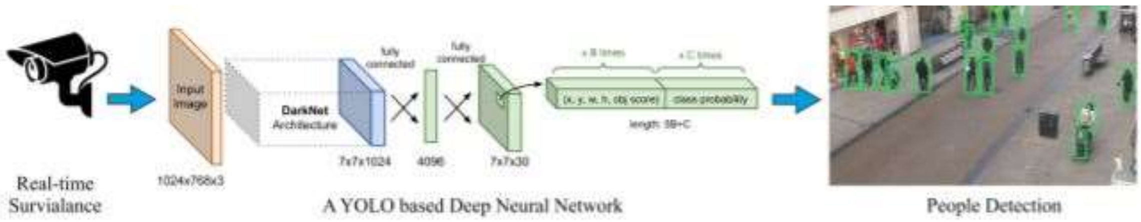
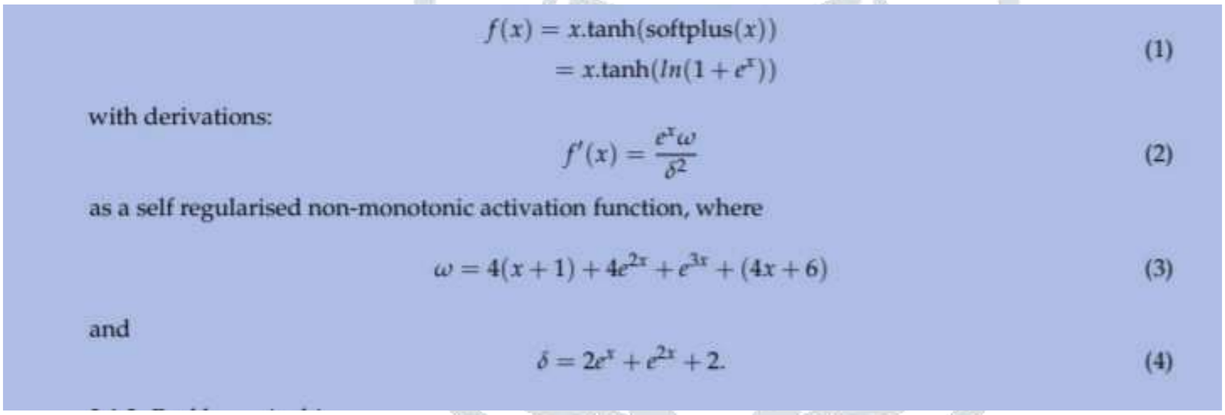


Figure 4. Stage 1—The overall structure of the people detection module.

1) *Inputs and Training Datasets* :

In order to have a robust detector, we would require a set of rich training datasets. This should include people with a variety in gender and age (man, women, boy, girl) with millions of accurate annotation and labelling. We selected two large datasets of MS COCO and Google Open Image dataset that satisfy the above-mentioed expectations, by providing more than 3.7 million annotated people. Further, details will be provided in Section 4 (Model Training and Experimental Results). In YOLOv4, the authors have dealt with two categories of training options for different parts of the network: “Bag of Freebies", which includes a set of methods to alter the model’s training strategy with the aim of increasing the generalisation; and “Bag of Specials” which includes a set of modules that can significantly improve the object detection accuracy in exchange for a small increase in training costs. Among various techniques of Bag of Freebies, we used the Mosaic data augmentations [53] which integrates four images into one, to increase the size of the input data without requiring to increase the batch size. On the other hand, in batch normalisation, the batch size reduction causes noisy estimation of mean and variance. To address this issue, we considered the normalised values of the previous k iterations instead of a single mini-batch. This is similar to Cross-Iteration Batch Normalisation (CBM) [95]. A set of possible activation functions for BoF are listed in Table 1. We also investigated the performance of our model against ReLU, Leaky ReLU, SELU, Swish, Parametric RELU, and Mish. Our preliminary evaluations confirmed the same results provided by Misra [86] for our human detection application. The Mish (Equation (1)) activation function converged towards the minimum loss, faster than Swish and ReLU, with higher accuracy. The result was consistent especially for diversity of parameter initialisers, regularisation methods, and lower learning rate values. Mish is in diagram.



$$\begin{aligned} f(x) &= x.\tanh(\mathrm{softplus}(x)) \\ &= x.\tanh(In(1+e^x)) \end{aligned} \qquad (1)$$

with derivations:

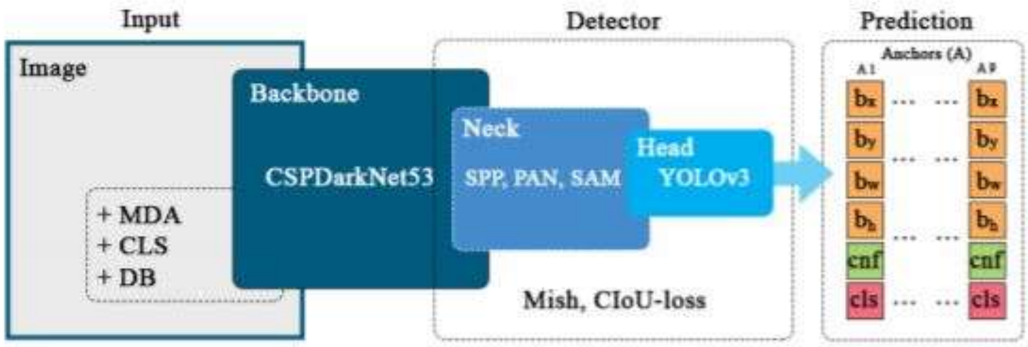
$$f'(x) = \frac{e^x \omega}{\delta^2} \qquad (2)$$

as a self regularised non-monotonic activation function, where

$$\omega = 4(x+1) + 4e^{2x} + e^{3x} + (4x+6) \qquad (3)$$

and

$$\delta = 2e^x + e^{2x} + 2. \qquad (4)$$



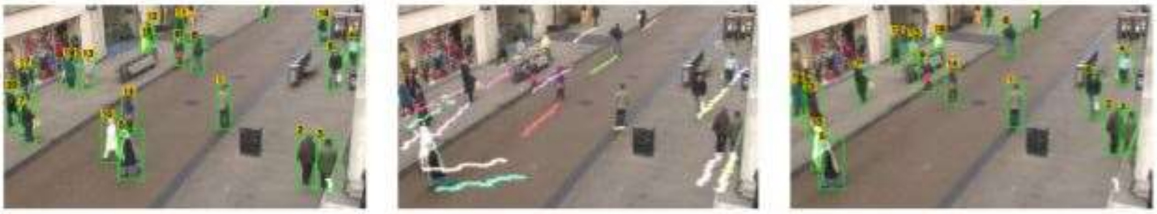
The network structure of the proposed three-level human detection module.

People Tracking :

The next step after the detection phase is people tracking and ID assignment for each individual. We use the Simple Online and Real-time (SORT) tracking technique [103] as a framework for the Kalman filter [104] along with the Hungarian optimisation technique to track the people. Kalman filter predicts the position of the human at time t + 1 based on the current measurement at time t and the mathematical modelling of the human movement. This is an effective way to keep localising the human in case of occlusion. The Hungarian algorithm is a combinatorial optimisation algorithm that helps to assign a unique ID number to identify a given object in a set of image frames, by examining whether a person in the current frame is the same detected person in the previous frames or not. Figure 7a shows a sample of the people detection and ID assignment, Figure 7b illustrates the tracking path of each individual, and Figure 7c shows the final position and status of each individual after 100 frames of detection, tracking, and ID assignment. We later use such temporal information for analysing the level of social distancing violations and high-risk zones of the scene. The state of each human in a frame is modelled as:

$$x = [u, v,s,r, u', v' ,s'] T'$$

where (u, v) represent the horizontal and vertical position of the target bounding box (i.e., the centroid); s denotes the scale (area), and r is the aspect ratio of the bounding box sides. u' , v' , and s' are the predicted values by Kalman filter for horizontal position, vertical position, and bounding box centroid, respectively.



(a) People detection and ID assignment at frame 0

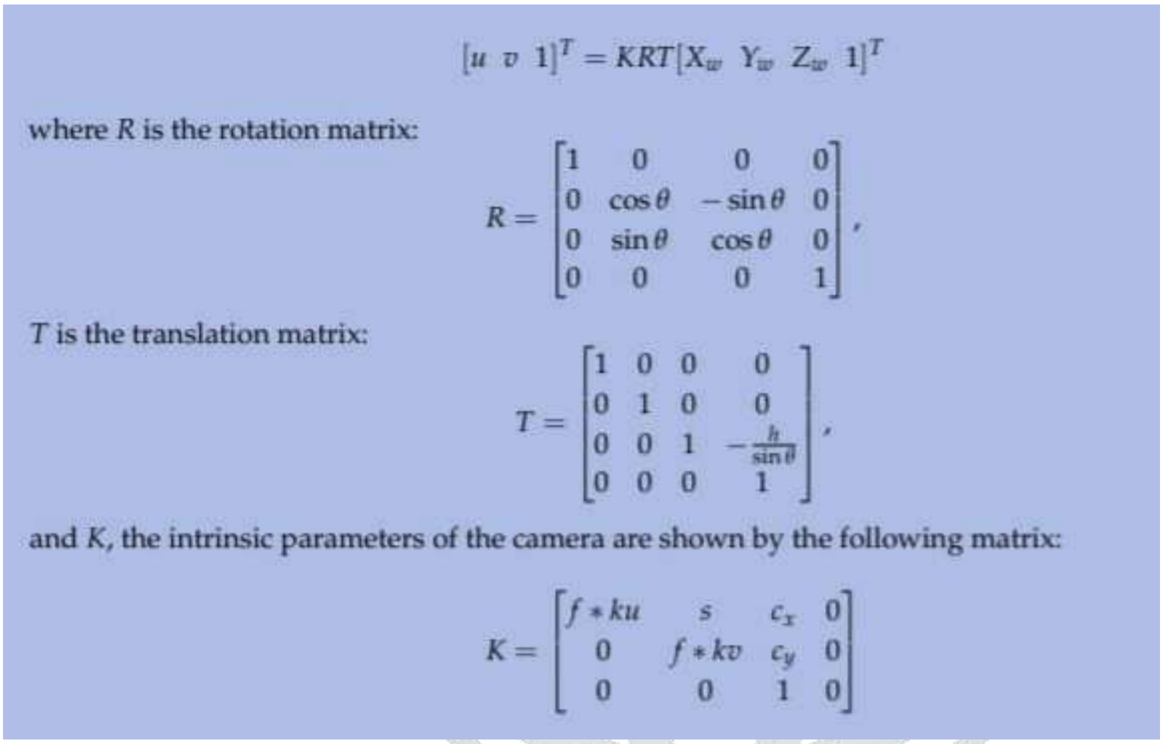
(b) Sample tracking for frames 0 to 100

(c) Final positions and IDs at frame 100

When an identified human associates with a new observation, the current bounding box will be updated with the newly observed state. This will be calculated based on the velocity and acceleration components, estimated by the Kalman filter framework. If the predicted identities of the query individual significantly differ from the new observation, almost the same state that is predicted by the Kalman filter will be used with almost no correction. Otherwise, the corrections weights will be split proportionally between the Kalman filter prediction and the new observation (measurement).

Inter-Distance Estimation :

Stereo-vision is a popular technique for distance estimation such as in [105]; however, this is not a feasible approach in our research when we aim at the integration of an efficient solution, applicable in all public places using only a basic CCTV camera. Therefore we adhere to a monocular solution. On the other hand, by using a single camera, the projection of a 3-D world scene into a 2-D perspective image plane leads to unrealistic pixel-distances between the objects. This is called perspective effect, in which we can not perceive uniform distribution of distances in the entire image. For example, parallel lines intersect at the horizon and farther people to the camera seem much shorter than the people who are closer to the camera coordinate centre. In 3-dimensional space, the centre or the reference point of each bounding box is associated with three parameters (x, y, z), while in the image received from the camera, the original 3D space is reduced to two-dimensions of (x, y), and the depth parameter (z) is not available. In such a lowered-dimensional space, the direct use of the Euclidean distance criterion to measure inter-people distance estimation would be erroneous. In order to apply a calibrated IPM transition, we first need to have a camera calibration by setting z = 0 to eliminate the perspective effect. We also need to know the camera location, its height, angle of view, as well as the optics specifications (i.e., the camera intrinsic parameters) [104].



$$[u \;\; v \;\; 1]^T = KRT[X_w \;\; Y_w \;\; Z_w \;\; 1]^T$$

where R is the rotation matrix:

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

T is the translation matrix:

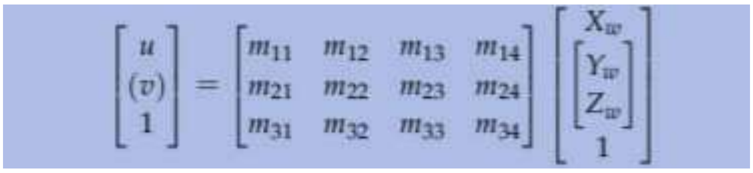
$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\frac{h}{\sin\theta} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and K, the intrinsic parameters of the camera are shown by the following matrix:

$$K = \begin{bmatrix} f * ku & s & c_x & 0 \\ 0 & f * kv & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

where h is the camera height, f is focal length, and ku and kv are the measured calibration coefficient values in horizontal and vertical pixel units, respectively. (c_x, c_y) is the principal point shifts that corrects the optical axis of the image plane.

The camera creates an image with a projection of three-dimensional points in the world coordinate that falls on a retina plane. Using homogeneous coordinates, the relationship between three-dimensional points and the resulting image points of projection can be shown as follows:



$$\begin{bmatrix} u \\ (v) \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_w \\ \begin{bmatrix} Y_w \\ Z_w \end{bmatrix} \\ 1 \end{bmatrix}$$

where M ∈ $R^{3\times4}$ is the transformation matrix with mij elements in Equation (16), that maps the world coordinate points into the image points based on the camera location and the reference frame, provided by the Camera

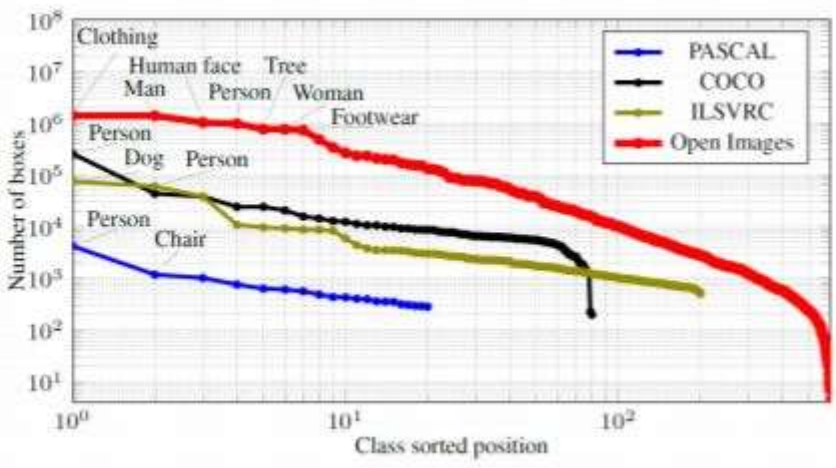
Intrinsic Matrix K (Equation (15)), Rotation Matrix R and the Translation Matrix T .And finally transferring from the perspective space to inverse perspective space (BEV) can also be expressed in the following scalar form:

$$(u,v) = \Big(\frac{m_{11} \times x_w + m_{12} \times y_w + m_{13}}{m_{31} \times x_w + m_{32} \times y_w + m_{33}}, \frac{m_{21} \times x_w + m_{22} \times y_w + m_{23}}{m_{31} \times x_w + m_{32} \times y_w + m_{33}}\Big)$$

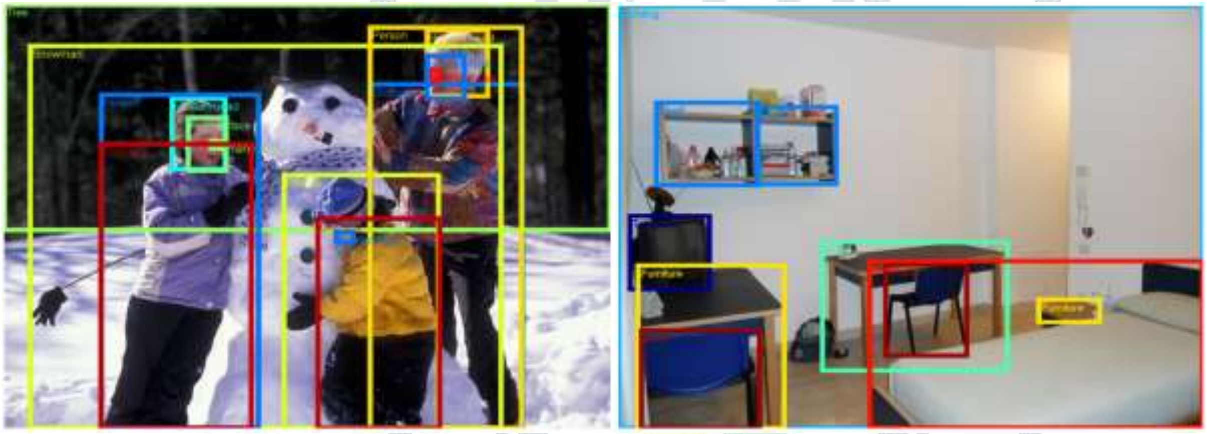
Model Training and Experimental Results :

In this section we discuss the steps taken to train our human detection model and the investigated datasets to train the model, followed by experimental results on people detection, social distancing measures, and risk infection assessment.

Four common multi-object annotated datasets were investigated including PASCAL VOC [55], Microsoft COCO [54], Image Net ILSVRC [106], and Google Open Images Datasets V6+ [107] which included 16 Million ground-truth bounding boxes in 600 categories. The dataset was a collection of 19,957 classes and the major part of the dataset was suitable for human detection and identification. The dataset was annotated using the bounding-box labels on each image along with the corresponding coordinates of each label. Figure 8 represents the sorted rank of object classes with the number of bounding boxes for each class in each dataset. In Google Open Images dataset (GOI) the class "Person" shows the 4th rank, with nearly 10^6 annotated bounding boxes; richer than other three investigated datasets. In addition to the class person, we also adopted four more classes of "Man", "Woman", "Boy", and "Girl" from the GOI dataset for the "human detection" training purpose. This made a total number of 3,762,615 samples that we used from training, including 257,253 samples from the COCO dataset and 3,505,362 samples from the GOI dataset.



We also considered the category of human body parts such as the legs as we believe this allows the detector to learn a more general concept of a human being, particularly in occluded situations or in case of partial visibility e.g., at the borders of the input image where the full-body of the individuals can not be perceived. Figure 9 shows examples of annotated images from the Open Images Dataset. The figure illustrates the diversity of the annotated people including large and small bounding boxes, in far and near distances to the camera image plane, people occlusion, as well as variations in shades and lighting conditions.

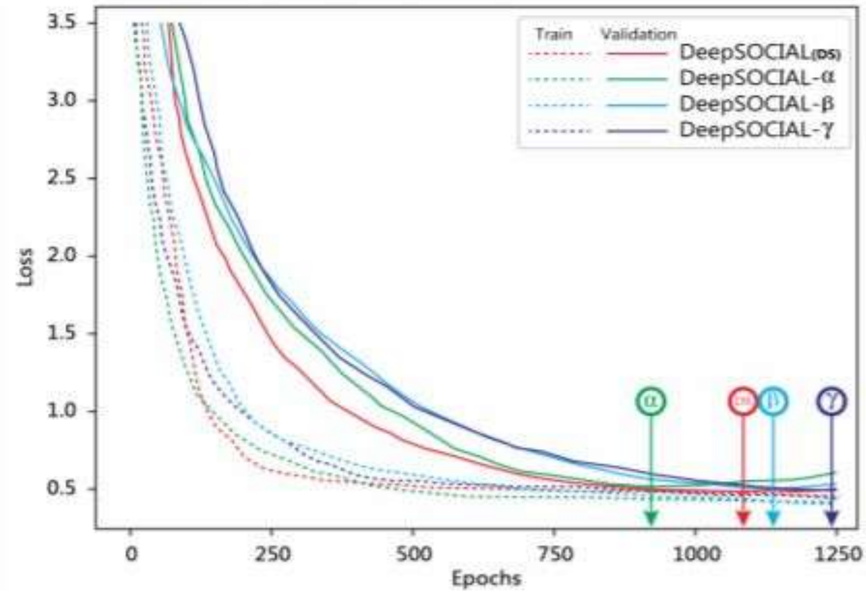


In order to train the developed model, we considered a transfer learning approach by using pre-trained models on Microsoft COCO dataset [54] followed by fine-tuning and optimisation of our YOLO-based model.

Performance Evaluation :

In order to test the performance of the proposed model, we used the Oxford Town Centre (OTC) dataset [31] as a previously unseen and challenging dataset with very frequent cases of occlusions, overlaps, and crowded zones. The dataset also contained a good diversity of human specimens in terms of clothes and appearance in a real-world public place. In order to provide a similar condition for performance analysis of YOLO based models, we fine-tuned each model on human categories of the Google Open Images (GOI) [107] data set. This was done by removing the last layer of each model and placing a new layer (with random values of the uniform probability distribution) corresponding to a binary classification (presence or absence of a human). Furthermore, in order to provide an equal condition for the speed and generalisability, we also tested each of the trained models against the OTC dataset [31]. We evaluated and compared our developed models against three common metrics of object detection in computer vision, including Precision Rate, Recall Rate, and FPS against three state-of-the-art human/object detection methods.

All of the benchmarking tests and comparisons were conducted on the same hardware and software: a Windows 10-based platform with an Intel c CoreTM i5-3570K processor and an NVIDIA RTX 2080 GPU with CUDA version 10.1. In terms of mass deployment of the system, the above hardware setup can handle up to 10 input cameras for real-time monitoring of e.g., different floors and angles of large shopping malls. However, for smaller scales, a cheaper RTX 1080 GPU or an 8-core/16-thread 10th generation Core^{TM} i7 CPU would suffice for real-time performance. Figure 10 illustrates the development of loss function in training and validations phases for four versions of our DeepSOCIAL model with different backbone structures. The graphs confirm a fast yet smooth and stable transition for minimising the loss function in DS version after 1090 epochs where we reached to an optimal trade-off point for both the training and validation loss. Table 3 provides the details of each backbone and the outcome of the experimental results against three more state-of-the-art model on the OTC Dataset.

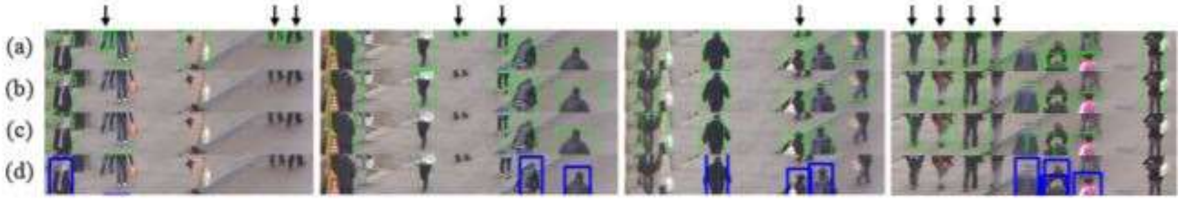


Training and validation loss of the DeepSOCIAL models over the Open Images dataset.

Interestingly, the Faster-RCNN model showed good generalisability; however, its low speed was an issue which seems to be due to the computational cost of the “region proposal” technique. Since the system required a real-time performance, any model with the speeds slower than 10 fps and/or a low level of accuracy may not be a suitable option for Social Distancing monitoring. Therefore, SSD and Faster-RCNN failed in this benchmarking assessment, despite their popularity in other applications. YOLOv3 and YOLOv4 based DeepSOCIAL–X methods provided relatively better results comparing the other models, and finally, the proposed DeepSOCIAL-DS model outperformed all of the assessed models in terms of both speed and accuracy

Detection performance of the DeepSOCIAL model in three different datasets.

sample footage of the challenging scenarios when the people either entered or exited the scene, and only part of their body (e.g., their feet) was visible. The figure clearly indicates the strength of DeepSOCIAL in Row (a), comparing to the state-of-the-art. The bottom row (d) with blue bounding boxes shows the ground-truth while some of the existing people with partial visibility are not annotated even in original ground-truth dataset. Row (c), YOLOv3 shows a couple of more detections; however, the IoU of the suggested bounding boxes are low and some of them can be counted as false positives. Row (b), the standard YOLOv4-based detector, shows a significant improvement comparing to row (c) and is considered as the second-best. Row (a), the DeepSOCIAL, shows 10 more true positive detections (highlighted by vertical arrows) comparing to the second best approach. Although the DeepSOCIAL model showed superior results even in challenging scenarios such as partial visibility and truncated objects, there could be some further challenges such as detections in extreme lighting conditions and lens distortion effects that may affect the performance of the model. This requires further investigations and experiments. Unfortunately, at the time of this research, there were no such datasets publicly available for our assessment.



Human detection with partial visibility (missing upper body parts). (a) DeepSOCIAL (b) YOLOv4 trained on MS-COCO, (c) YOLOv3 (d) Ground-truth annotations from the Oxford Town Centre (OTC) dataset.

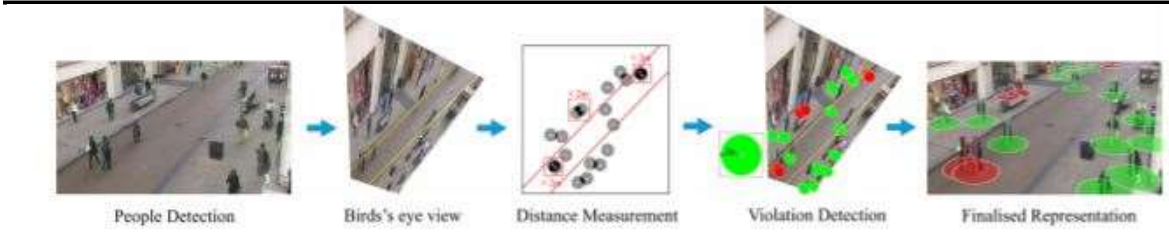
Social Distancing Evaluations :

We considered the midpoint of the bottom edge of the detected bounding boxes as our reference points (i.e., shoes' location). After the IPM, we would expect to have the location of each person, in the homogeneous space of BEV with a linear distance representation. Any two people Pi , Pj with the Euclidean distance of smaller than r (i.e., the set restriction) in the BEV space were considered as contributors in social distancing violation:

Left (from Oxford Town Centre Dataset [31]) shows the detected people followed by the steps we have taken for inter-people distance estimation including tracking, IPM, homogeneous 360◦ distance estimation, safe movements (people in green circles) and the violating people (with overlapping red circles):

$$\Lambda_{\xi}(P_i, P_j, r) = \begin{cases} 1 & \text{if } \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq r \\ 0 & \text{if } \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} > r \end{cases}$$

Regarding the Oxford Town Centre (OTC) dataset, every 10 pixels in the BEV space was equivalent to 98 cm in the real world. Therefore, r ≈ 2 × ξ and equal to 20 pixels. The inter-people distance measurement was measured based on the Euclidean L2 norm distance.

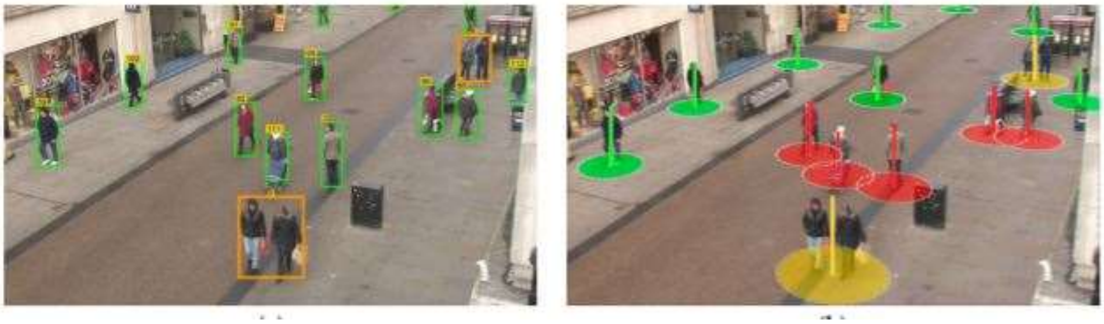


The summary of the steps taken for detection, tracking, and distance estimation.

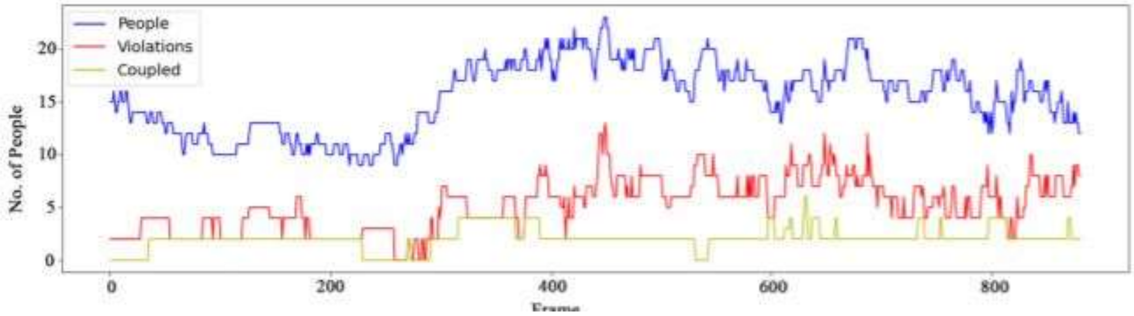
One of the controversial opinions that we received from health authorities was the way of dealing with family members and couples in social distancing monitoring. Some researchers believed social distancing should apply on every single individual without any exceptions and others were advising the couples and family members can walk in a close proximity without being counted as a breach of social distancing. In some countries such as in the UK and EU region the guideline allows two family members or a couple walk together without considering it as the breach of social distancing. We also considered a solution to activate the couple detection. This will be helpful when we aim at recognizing risky zones based on the statistical analysis of overall movements and social distancing violations over a mid or long period (e.g., from few hours to few days). Applying a temporal data analysis approach, we consider two individuals (pi , pj) as a couple, if they are less than d meters apart in an adjacency, for a tΛ of more than ε seconds. As an example, in Figure 14a, we have identified people who have been less a meter apart from each other for more than ε = 5 s, in the same moving trajectory:

$$\begin{aligned} C(p_i,p_j)=1 \quad & \text{IF} \quad p_i,p_j \in D \\ & \text{AND} \quad \Lambda_\zeta(p_i,p_j,d)=1 \\ & \text{AND} \quad t_\Lambda(p_i,p_j) > \varepsilon \end{aligned}$$

In cases where a breach occurs between two neighbouring couple, or between a couple and an individual, all of the involved people will turn to red status, regardless of being a couple or not. The flexibility of our algorithm in considering different types of scenarios enables the policymakers and health authorities to proceed with different types of investigations and evaluations for the spread of the infection. For example, Figure 15 from the Oxford Town Centre dataset, provides a basic statistics about the number of people in each frame, the number of people who do not observe the distancing, the number of social distancing violations without counting the coupled groups as violations.



Social distancing violation detection for coupled people and individuals. (a) Examples of coupled people detections-orange bounding boxes; (b) Three types of detections: safe, violations, coupled.



A 2D recording of the number of detected people in 900 frames from the OTC Dataset, as well as the number of violations and number of couples with no violations.

Regarding the coupled group we reached to an accuracy and recall rate of 98.7% and 23.9 fps, respectively which is slightly lower than our results of normal human detection as per the Table 3. This was expected due to added complexity in tracking two side by side people and possibly more complex occlusion scenarios.

Implementation Plan for next semester

In next semester we work on plan and add extra features in our project , also we add modules from OPENCV . We work on the ideas and concepts as well extra features which is mention in future scope. We ensures that project is error free and which produce correct result and gives correct output for every possible case.

Running the program will give you frame (firstframe) where you need to draw ROI and distance scale. To get ROI and distance scale points from first frame Code to transform perspective to Bird's eye view (Top view) and to calculate horizontal and vertical 180 cm distance in Bird's eye view ROI and Scale points' selection for first frame. The second step to detect pedestrians and draw a bounding box around each pedestrian. To detect humans in video and get bounding box details. Now we have bounding box for each person in the frame. We need to estimate person location in frame. i.e we can take bottom center point of bounding box as person location in frame. Then we estimate (x,y) location in bird's eye view by applying transformation to the bottom center point of each person's bounding box, resulting intheirpositioninthebird's eye view. To calculate bottom center point for all bounding boxes and projecting those points in Bird's eye view. Last step is to compute the bird's eye viewdistancebetweenevery pair of people (Point) and scale the distances by the scaling factor in horizontal and vertical direction estimated from calibration.

Lastly we can draw Bird's Eye View for region of interest (ROI) and draw bounding boxes according to risk factor for humans in a frame and draw lines between boxes according to risk factor between two humans.Red,Yellow,Greenpoints represents risk to human in Bird's eye view. Red: High Risk, Yellow: Low Risk and Green: No Risk. Red, Yellow lines between two humans in output tell they are violating social distancing rules.

CONCLUSION

The article proposes an efficient real-time deep learning based framework to automate the process of monitoring the social distancing via object detection and tracking approaches, where each individual is identified in the real-time with the help of bounding boxes. The generated bounding boxes aid in identifying the clusters or groups of people satisfying the closeness property computed with the help of pair wise vector approach. The number of violations are confirmed by computing the number of groups formed and violation index term computed as the ratio of the number of people to the number of groups. The extensive trials were conducted with popular state-of-the-art object detection models: Faster RCNN, SSD, and YOLO v3, where YOLO v3 illustrated the efficient performance with balanced FPS and MAP score. Since this approach is highly sensitive to the spatial location of the camera, the same approach can be fine tuned to better adjust with the corresponding field of view.

The emerging trends and the availability of intelligent technologies make us to develop new models that help to satisfy the needs of emerging world. So we have developed a novel social distancing detector which can possibly contribute to public healthcare. The model proposes an efficient real-time deep learning based framework to automate the process of monitoring the social distancing via object detection and tracking approaches, where each individual is identified in the real-time with the help of bounding boxes. Identifying the clusters or groups of people satisfying the closeness property computed with the help of Bird's eye view approach. The number of violations is confirmed by computing the number of groups formed and violation index term computed as the ratio of the number of people to the number of groups. The extensive trials were conducted with popular state-of-the-art object detection models Faster RCNN, SSD, and YOLO v3, since this approach is highly sensitive to the spatial location of the camera, the same approach can be fine tuned to better adjust with the corresponding field of view.

This system works very effectively and efficiently in identifying the social distancing between the people and generating the alert that can be handled and monitored.This solution can be used in places like temples, shopping complex, metro stations, airports, etc.

REFERENCES:

[1] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition ´ via sparse spatio-temporal features," in 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. IEEE, 2005, pp. 65–72.

[2] M. Piccardi,"Background subtraction techniques: a review," in 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583), vol. 4. IEEE, 2004, pp. 3099–3104.

[3] Y. Xu, J. Dong, B. Zhang, and D. Xu, "Background modeling methods in video analysis: A review and comparative evaluation," CAAI Transactions on Intelligence Technology, vol. 1, no. 1, pp. 43–60, 2016.

[4] Pias, “Object detection and distance measurement,” https://github.com/ paul-pias/Object-Detectionand-Distance-Measurement, 2020.

[5] https://blog.usejournal.com/social-distancing-ai-using-python-deep-learning-c26b20c9aa4c

[6] https://www.learnopencv.com/deep-learning-based-object-detection-using-yolov3-with-opencv-python-c/

[7] https://pjreddie.com/media/files/yolov3.weights

[8] Yolov3 Object detection: https://www.learnopencv.com/deep-learning-based-object-detection-using-yolov3-with-opencv-python-c/

[9] https://www.analyticsvidhya.com/blog/2020/05/social-distancing-detection-tool-deep-learning/

[10] https://www.sciencedirect.com/science/article/pii/S2210670720307897

[11] https://www.pyimagesearch.com/2020/06/01/opencv-social-distancing-detector/

[12] https://github.com/topics/social-distancing-detection.