**MONITORING AND DETECTING SOCIAL DISTANCING AMONG THE CROWD USING YOLO**

**A PROJECT REPORT**

***Submitted by***

**LAKSHMANAN K 513418104019**

**NITHISH L 513418104025**

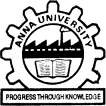
***in partial fulfillment for the award of the degree***

***of***

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE & ENGINEERING**



**UNIVERSITY COLLEGE OF ENGINEERING KANCHEEPURAM**

(A Constituent College of Anna University, Chennai) **ANNA UNIVERSITY CHENNAI – 600 025**

**APRIL 2021**

# ANNA UNIVERSITY: CHENNAI 600 025

# BONAFIDE CERTIFICATE

Certified that this Project report titled “**MONITORING AND DETECTING SOCIAL DISTANCING AMONG THE CROWD USING YOLO**” is the bonafide work of “**LAKSHMANAN.K (513418104019), NITHISH.L (513418104025)**” who carried out the project work under my supervision.

**SIGNATURE**

**Mr. M. Suresh, M.E,**

**SUPERVISOR,**

Teaching Fellow,

Department of CSE,

University college of Engineering,

Kancheepuram – 631552.

**SIGNATURE**

**Dr. V. Kavitha B.E., M.E., Ph.D.,**

**HEAD OF THE DEPARTMENT,**

Professor,

Department of CSE,

University college of Engineering,

Kancheepuram – 631552.

Kancheepuram - 631552.

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |

Submitted for the Project Viva Voice held on ………………

**INTERNAL EXAMINER** **EXTERNAL EXAMINER**

**ACKNOWLEDGEMENT**

We are always thankful to our college Dean **Prof. Dr. P. SAKTHIVEL B.E., M.E., Ph.D.,** who endorsed us throughout this project.

Our heartfelt thanks to HOD **Dr.V.KAVITHA, M.E, Ph.D., PROFESSOR,** Department of Computer Science and Engineering, University College of Engineering, Kancheepuram, for the prompt and limitless help in providing the excellent infrastructure to do the project and to prepare the thesis.

We express our deep sense of special guidance to our Project Guide, **Mr. M. SURESH, M.E., Teaching Fellow, Department of Computer Science and Engineering,** for his valuable guidance, ideas, inspiration and encouragement that was instrumental in the successful completion of the project.

We express our sincere thanks to the project committee members **Dr.K.SELVABHUVANESWARI, M.E, Ph.D., ASSISTANT PROFESSOR, Mrs.T.KALA, M.E, ASSISTANT PROFESSOR, Mr.M.SURESH, M.E, TEACHING FELLOW**, Department of Computer Science and Engineering, Kancheepuram, for their invaluable guidance and technical Support.

We thank all the faculty members of Department of Computer Science and Engineering for their valuable help and guidance. We are grateful to our family and friends for their constant support and encouragement. We thank the almighty whose showers of blessings made this project a reality.

## LAKSHMANAN.K NITHISH.L

## (513418104019) (513418104025)

## ABSTRACT

Social distancing is a recommended solution by the World Health Organization (WHO) to minimize the spread of COVID-19 in public places. The majority of governments and national health authorities have set the 2-meter physical distancing as a mandatory safety measure in shopping centers, schools and other covered areas. The public health bodies such as the Centers for Disease Control and Prevention (CDC) had to make it clear that the most effective way to slow down the spread of Covid-19 is by avoiding close contact with other people. To reduce the possibility of infection, it is advised that people should avoid any person-to-person contact such as shaking hands and they should maintain a distance of at least 1 meter from each other.

The current state-of-the-art object detectors with deep learning had their pros and cons in terms of accuracy and speed. The object might have different spatial locations and aspect ratios within the image. Hence, the real-time algorithms of object detection using the CNN model such as R-CNN and YOLO had further developed to detect multi-classes in a different region in images had been developed. YOLO (You Only Look Once) is the prominent technique for deep CNN-based object detection in terms of both speed and accuracy. The number of people in an image and video with bounding boxes can be detected via these existing deep CNN methods where the YOLO method was employed to detect the video stream taken by the camera. By measuring the Euclidean distance between people, the application will highlight whether there is sufficient social distance between people in the video.

The proposed system is to develop a generic Deep Neural Network-Based model for automated humans detection in video sequences with individual id’s, track person violating the social distancing threshold (2 meters) and physical distance measurement between the pixel and the given threshold value.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| CHAPTER NO | TITLE | PAGE  NO |
|  | ABSTRACT | iv |
|  | LIST OF TABLE CONTENTS | v |
|  | LIST OF FIGURES | viii |
|  | LIST OF ABBREVATIONS | x |
| 1 | INTRODUCTION |  |
|  | 1.1 BACKGROUND | 1 |
|  | *1.1.1 IMPORTANCE OF SOCIAL DISTANCING* | 2 |
|  | 1.2 OBJECTIVE | 3 |
|  | 1.3 SCOPE OF THE STUDY | 4 |
|  | 1.4 MOTIVATION FOR THE STUDY | 5 |
|  | 1.5 LIMITATIONS OF THE WORK | 6 |
|  | 1.6 OUTLINE OF THE CHAPTERS | 7 |
|  | 1.7 ORGANIZATION OF THE REPORT | 9 |
| 2 | PRELIMINARIES |  |
|  | 2.1 DEEP LEARNING | 10 |
|  | 2.2 COMPUTER VISION | 11 |
|  | 2.3 APPLICATIONS OF DEEP LEARNING FOR  COMPUTER VISION | 13 |
|  | 2.4 OBJECT DETECTION | 14 |
|  | *2.4.1 ROLE OF OBJECT DETECTION IN CV* | 15 |
|  | *2.4.2 WORKING OF OBJECT DETECTION* | 16 |
|  | *2.4.3 OBJECT DETECTION ALGORITHMS* | 17 |
|  | 2.5 ADVANCEMENT IN OBJECT DETECTION | 18 |
| 3 | LITERATURE SURVEY |  |
|  | 3.1 REALTIME SOCIAL DISTANCING  DETECTOR USING DEEP LEARNING | 20 |
|  | 3.2 SOCIAL DISTANCE DETECTION WITH DL | 20 |

|  |  |  |
| --- | --- | --- |
|  | 3.3 MONITORING SOCIAL DISTANCING  UNDER VARIOUS LOW LIGHT  CONDITIONS WITH DEEP LEARNING | 21 |
|  | 3.4 DEEP LEARNING-BASED SOCIAL  DISTANCE MONITORING  FRAMEWORK | 22 |
|  | 3.5 A SOCIAL DISTANCING DETECTOR  USING TENSORFLOW | 23 |
| **4** | **IMPLEMENTATION** |  |
|  | 4.1EXISTING SYSTEM | 24 |
|  | *4.1.1 DISADVANTAGES FACED* | 25 |
|  | 4.2 PROPOSED SYSTEM | 26 |
|  | *4.2.1 REASONS FOR CHOOSING YOLO*  *V3* | 27 |
|  | *4.2.2 DETECTING THROUGH YOLO V3* | 28 |
|  | *4.2.3 ADVANTAGES* | 30 |
|  | 4.3 MODULES AND LIBRARIES | 30 |
|  | 4.4 SYSTEM ARCHITECTURE | 31 |
|  | *4.4.1 PROCESS DIAGRAM* | 32 |
|  | *4.4.2 FLOW DIAGRAM* | 33 |
|  | *4.4.3 OUTLINE OF EACH STEPS* | 34 |
|  | 4.5 ALGORITHMS | 36 |
|  | *4.5.1 FASTER R-CNN WITH*  *ARCHITECTURE* | 36 |
|  | *4.5.2 SSD WITH ARCHITECTURE* | 37 |
|  | *4.5.3 YOLO V3 WITH ARCHITECTURE* | 38 |
| **5** | **DEVELOPMENT** |  |
|  | 5.1 INPUT VIDEO FRAME | 40 |
|  | 5.2 PEDESTRIAN DETECTION | 40 |
|  | 5.3 CAMERA VIEW CALIBRATION | 41 |
|  | 5.4 DISTANCE MEASUREMENT | 42 |
|  | 5.5 OUTPUT DEPICTION | 43 |

|  |  |  |
| --- | --- | --- |
|  | 5.6 PERFORMANCE EVALUATION | 43 |
| **6** | **SNAPSHOTS** |  |
|  | 6.1 VIDEO AS AN INPUT FRAME | 44 |
|  | 6.2 WEBCAM AS AN INPUT FRAME | 44 |
| **7** | **CONCLUSION** |  |
|  | 7.1 CONCLUSION TO OUR SOLUTION | 45 |
|  | 7.2 FUTURE IMPROVEMENTS | 45 |
| **8** | **REFERENCES** | 46 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO** | **NAME OF THE FIGURE** | **PAGE NO** |
| 1.1.1 | IMPORTANCE OF SOCIAL DISTANCING | **3** |
| 2.1 | DEEP LEARNIG CLASS OF MACHINE LEARNING | **11** |
| 2.2 | DIGITAL IMAGE PROCESSED | **12** |
| 2.4 | OBJECT DETECTION IN DL | **15** |
| 2.4.2 | DEEP LEARNING BASED OBJECT DETECTION IN VEHICLES | **18** |
| 2.5 | AN ADVANCEMENT IN OBJECT DETECTION BY GOOGLE | **19** |
| 4.1.1 | PERFORMANCE MEASURE OF FASTER R-CNN AND YOLO | **26** |
| 4.2.1.1 | OBJECT DETECTION USING YOLO V3 | **27** |
| 4.2.1.2 | ACCURACY MEASURE | **28** |
| 4.2.2 | YOLO V3 OBJECT DETECTION | **29** |
| 4.4.1 | PROCESS DIAGRAM FOR THE PROPOSED MODEL | **32** |
| 4.4.2 | FLOW DIAGRAM | **33** |
| 4.5.1 | SCHEMATIC REPRESENTATION OF FASTER R-CNN ARCHITECTURE | **36** |
| 4.5.2 | SCHEMATIC REPRESENTATION OF SSD ARCHITECTURE | **37** |
| 4.5.3 | SCHEMATIC REPRESENTATION OF YOLO V3 ARCHITECTURE | **38** |
| 5.2.1 | BOUNDING BOX REPRESENTATION | **41** |
| 5.2.2 | PEDESTRIAN DETECTED | **41** |
| 5.4.1 | EUCLIDEAN DISTANCE FORMULA FOR CENTROID TRACKING | **42** |
| 5.4.2 | PEDESTRIAN DETECTION | **43** |
| 5.6 | PERFORMANCE EVALUATION | **43** |
| 7.1 | MONITORING IN AN IMAGE | **44** |
| 7.2 | MONITORING IN A VIDEO FRAME | **44** |

|  |  |
| --- | --- |
| **ACRONYM** | **ABBREVATIONS** |
| CNN | Convolutional Neural Network |
| R-CNN | Regional based Convolutional Neural Network |
| SSD | Single Shot Detector |
| YOLO | You Only Look Once |
| WHO | World Health Organization |
| CDC | Centers for Disease Control and Prevention |
| COVID | Corona Virus Disease |
| GPS | Global Positioning Time |
| CPU | Central Processing Unit |
| GPU | Graphics Processing Unit |
| CV | Computer Vision |
| COCO | Common Object in Context |
| HOG | Histogram of Oriented Gradients |
| TOF | Time of Flight |
| YOLOR | You Only Learn One Representation |
| MAP | Mean Average Precision Value |
| FPS | Frames Per Second |
| ROI | Region Of Interest |
| TT | Time Taken |
| API | Application Interface |
| RPN | Region Proposal Network |
| OPENCV | Open Source Computer Vision |
| IOU | Intersection Of Union |
|  |  |

**LIST OF ABBREVATIONS**

**CHAPTER 1**

**INTRODUCTION**

**1.1 BACKGROUND OF THE PROJECT**

In late December 2019, Wuhan, China, identified a new generation of **coronavirus disease** (COVID-19). The virus spread globally in 2020 after just a few months. The condition was declared pandemic by the World Health Organization (WHO) in May 2020. According to **WHO** figures released on April 18, 2021, 141 million people are sick in 200 nations, with 3,000,000 fatalities. There is also **no successful vaccine** or therapy for the infection, despite the increasing number of patients. As a result, in addition to wearing face masks**, social distancing** now claims to be much more necessary than previously believed, and one of the most effective strategies to prevent the transmission of the disease. It is also considered a mandatory norm in almost all nations. According to WHO guidelines, persons must be separated by at least **6 feet (1.8 m)** in order to maintain sufficient social distance.

Governments attempted to introduce a series of social distancing policies during the COVID-19 pandemic, such as banning transport, policing boundaries, shutting pubs and bars, and alerting society to keep a gap of 1.6 to 2 meters between them. Monitoring the propagation of contamination and the effectiveness of the restraints, on the other hand, is a difficult challenge.

At the time when coronavirus is begun spreading across the individuals and society, research and scientists are starting to find out the best solution to eliminate the spread of this pandemic. Many researchers and scientists suggested tracking a person infected with COVID-19 **using GPS** and built-in applications in smartphones. However, this technology has limitations on tracking individuals who have no Wi-Fi or cell signals. On the other hand, some authorities utilize **drones** with mounted video cameras to track the gathering of individuals in the outdoor area. Such technology is suitable for monitoring COVID-19 which could amid the coronavirus outbreak.

The study aimed to derive an approximation that shows how early social distancing measures can reduce economic loss and the number of new infections significantly. Recent developments showed that the identification of individuals through **video surveillance cameras** can be achieved by face, and **a person’s manner of walking.** However, the detection of a person under crowds’ technique is difficult and hard to optimize. The solution approached over here is to detect out the pedestrians with a low-resolution camera and tracking them mathematically and analyzing out in the real time.

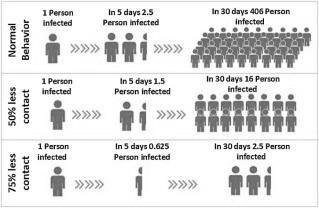
* + 1. **IMPORTANCE OF SOCIAL DISTANCING**

Social Distancing simply refers of avoiding close contact with the other

individuals in order to avoid catching the virus ourselves and to avoid passing it on. Social distancing now is one of the greatest calls for humanity in these critical times. In this way countries are not only using this is as a preventive measure for getting the preventive measures but also trying to **reduce the infection** and **flattening the curve of infection to community**. The lock down as it is being referred to will essentially reduce the viral load, reduce the burden of infected cases for treatment at various healthcare facilities and give time for the country to become prepared for any major outbreak.

The disease is spread via respiratory droplets that might end up on your hands if they are coughed on or you touch surfaces that have been coughed on. If you are too close, then you are likely to breath in the droplets, including COVID-19 virus if the person who coughs near you has the disease. So if movement of people in large numbers is reduced then it brings down the risk of community spread not only in the young but mainly in the elderly as well. That is why social distancing is such an important containment measure in this rapid spreading contagious disease**.**

Social distancing can also effectively extend to environmental precautions such as disinfecting often-touched surfaces that may pass on the virus. The best way to slow the spread is through public health measures that encourages and creating awareness about social distancing. Prevent infection by staying in your house, washing your hands, avoiding people who are sick. This measure can be life-saving – not necessarily for you if you are 30 years old and healthy, but maybe for your parents.

Fig 1.1.1 Importance of social distancing

**1.2 OBJECTIVE**

In this pandemic driven world, we people don’t even rely upon how come the disease is actually spreading out and knowing out the exact source of the transfer of a disease is what we people miss to interpret. In general we do not have the tendency or human vision to exactly know what type of distance we are practicing and how much feet or meters do we really follow upon. Casually we come to our own approximation from our vision and further unknowingly we would be the source of transfer. But there’s something beyond human vision and all the discrepancies faced by the human vision is actually solved through computer vision. With this motive what we have tried is to propose a model that could monitor out the crowd and measure out the exact social distancing between them. As a whole the above specified model will try to accomplish various case scenarios such as

* The higher authority officer who are employed as the security officers, don’t even need to bother about monitoring out the crowd and further taking necessary actions.
* It would also create an impact among the crowd that the transfer of non-crowded places to crowded places is also an essential and hidden source of transmission of covid-19.
* It would also help the less crowded places to maintain the social distancing so that it could be an early measure of not making out the crowd to violate the protocols.
* It reduces out the burden of the government which is actually using out the mechanism of drawing out the circles with 2mts difference and allowing people to stand in their respective circles.

**1.3 SCOPE OF STUDY**

The main scope of study is to estimate and monitor the social distancing among moving persons in a crowded locality through accurate person detection in video sequences using one of the highly performable object detecting algorithm called as YOLO. The key components of this work are as follows

* To present a deep learning-based social distance monitoring framework using an overhead view perspective.
* To deploy pre-trained YOLOv3 for human detection and computing their bounding box centroid information. In addition, a transfer learning method is applied to enhance the performance of the model. The additional training is performed with overhead data set, and the newly trained layer is appended to the pre-trained model.
* In order to track the social distance between individuals, the Euclidean distance is used to approximate the distance between each pair of the centroid of the bounding box detected. In addition, a social distance violation threshold is specified using a pixel to distance estimation.
* Utilizing a centroid tracking algorithm to keep track of the person who violates the social distance threshold.
* To assess the performance of pre-trained YOLOv3 by evaluating it on an overhead data set. The output of the detection framework is assessed with and without the transfer learning. Furthermore, the model performance is also compared with other deep learning models.

**1.4 MOTIVATION FOR THE STUDY**

It is believed and observed that the social distancing could reduce the number of infected patients. If someone is sick and there are no people around, a virus cannot spread.

Deep learning based neural system through computer vision would additionally increase the efforts of monitoring the social distance in the worst case scenario of the people who don’t stand in the circles and unknowingly or knowingly contact with the other people. Detecting and computing out the social distancing through deep learning would reduce the other workflows by humans.

The Centers for Disease Control and Prevention defines social distancing as remaining out of congregate settings, avoiding mass gatherings and maintaining distance ­approximately 6 feet or 2 meters among peoples.

Human detection using visual surveillance system is an established area of research which is relying upon manual methods of identifying unusual activities, however, it has limited capabilities. Monitoring out the distances between the moving persons in a crowded space is actually a cumbersome task for the humans and hence it would lead to the spread of covid-19. “**Social distancing is staying away from people not from your purpose.”**

**1.5 LIMITATIONS OF THE WORK**

* + - Since we have used the YOLO V3 algorithm for object detection, it actually has some difficulties such as
      * + Comparatively low recall and more localization error as compared to Faster R-CNN.
        + Struggles to detect close objects because each grid can propose only 2 bounding boxes.
        + Struggles to detect small objects.
* Also detecting out the objects as it is being trained on the Image Net dataset produces a less optimization as compared by training with our own custom datasets.
* The model works well with less training on a CPU but it works bad with huge number of datasets.
* On the other hand, YOLOv3 may not be ideal for using niche models where large datasets can be hard to obtain.
* There are other newer developments in YOLO such as YOLO V4 and YOLO V5 algorithms which actually performs well over YOLO V3.
* Another drawback is that for training the model we need a high end performing machines as this process crashes in the low end machines.
* The major drawback of processing out the video frame by frame over here is that the quality of pixels in each frame as we have reduced the frame height and frame width for efficient performance which actually degrades the quality of the pixels in the frame.

**1.6 OUTLINE OF THE CHAPTERS**

**Chapter 1: Introduction**

Each component of the study such as background, scope and objectives, methodology and data collection procedure is presented.

**Chapter 2: Preliminaries**

This chapter includes all the preliminaries required to know about the domain in a broader perspective. This is actually consists of a comprehensive discussion about the prerequisites in a theoretical manner.

**Chapter 3: Literature Survey**

This chapter presents a comprehensive coverage of various studies carried out in the analysis of social distancing measures by the world and each individual government. A detailed description of research papers and journals are listed regarding Deep learning measures for social distancing detection.

**Chapter 4: Implementation**

This chapter covers about all the implementation details and the ideas we employed into the project. It covers about the existing system we analyzed and the proposed system we made and a comparative analysis on both the systems is done in this chapter. Also a detailed descriptive system architecture is proposed with a flow diagram in this chapter. Various object detecting algorithms are discussed with their architecture.

**Chapter 5: Development**

This chapter includes the methodologies we used to train and develop our deep learning model.

**Chapter 6: Snapshots**

This Chapter includes all the snapshots snapped while running the model in the real time scenario. Also it includes some of the essential code snippets of our model.

**Chapter 7: Conclusion**

This chapter presents summary of the whole study with inferences. The scope for future research will also be discussed.

**Chapter 8: References**

This chapter provides a curated list of articles, research papers, journals and other surfed materials we referred during the development of our model.

**1.6 ORGANIZATION OF THE REPORT**

This report helps you to understand the problem of violating out the social distancing protocol among the people and also the solution which is provided through the project and the technologies that are used for the development of the project and the work which are yet to perform in future for the further enhancement of the project along with the impact which is going to be created by the proposed system called **MONITORING AND DETECTING THE SOCIAL DISTANCING AMONG THE CROWD USING YOLO V3.**

**CHAPTER 2**

**PRELIMINARIES**

**2.1 DEEP LEARNING**

Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.

Deep-learning architectures such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.

**APPLICATIONS OF DEEP LEARNING**

* Automatic speech recognition
* Image recognition
* Visual art processing
* Natural language processing
* Drug discovery and toxicology
* Customer relationship management
* Recommendation systems
* Bioinformatics
* Medical image analysis
* Mobile advertising
* Financial fraud detection

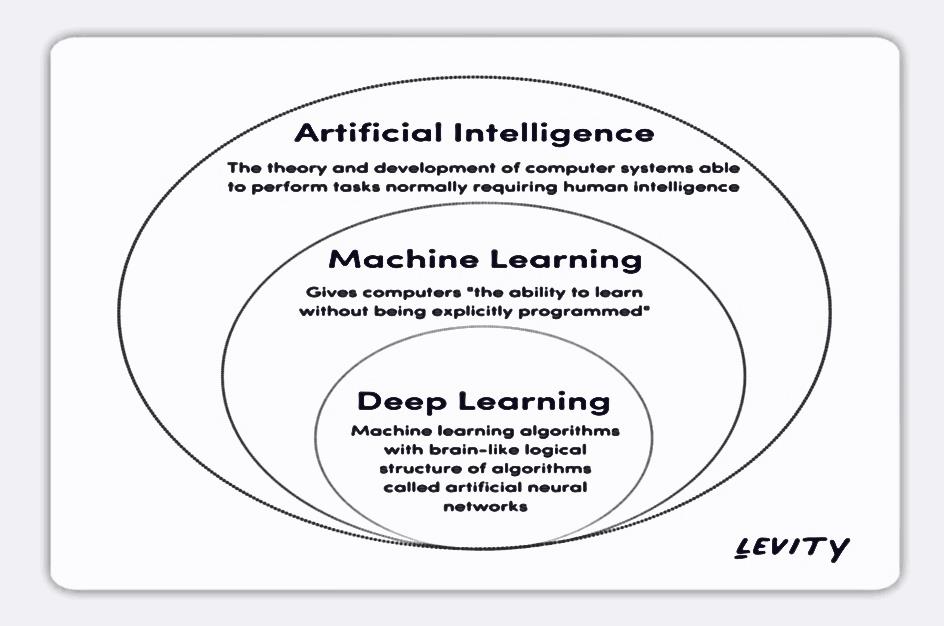


Fig.2.1 Deep learning class of machine learning

**2.2 COMPUTER VISION**

Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do. Computer vision tasks through methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information.

Applications of CV are Automatic inspection, Assisting humans in identification tasks, Controlling processes, Detecting events, Navigation,

Organizing information



Fig 2.2 Digital Image Processed

**HUMAN VISION VS COMPUTER VISION**

Human vision revolves around light and does not involve repetition or patterns. In other words, we do not need to learn to see, it is biologically embedded in us. Human vision consists of several steps. First, light bounces off the image and enters the eyes through the cornea. Then, the cornea directs light to the pupils and iris, which work together to control the amount of light entering the eye. Once the light passes through the cornea, it enters the retina; the retina has special sensors called cones and rods, which are involved in color vision.

Computer vision is a field of artificial intelligence that aims to mathematically model the processes of visual perception in living beings, and to generate models, algorithms and programs that allow the simulation of these visual abilities using the capacity of computers. This means that computers can make inferences about images without human assistance. This seems simple because humans can effortlessly see the world around them; however, teaching a computer to see like a human is difficult because we still do not really understand how human vision works.

**2.3 APPLICATIONS OF DEEP LEARNING FOR CV**

**Image Classification**

Image classification involves assigning a label to an entire image or photograph. This problem is also referred to as “object classification” and perhaps more generally as “image recognition,” although this latter task may apply to a much broader set of tasks related to classifying the content of images.

**Image Classification with Localization**

Image classification with localization involves assigning a class label to an image and showing the location of the object in the image by a bounding box (drawing a box around the object). This is a more challenging version of image classification.

**Object Detection**

Object detection is the task of image classification with localization, although an image may contain multiple objects that require localization and classification. This is a more challenging task than simple image classification or image classification with localization, as often there are multiple objects in the image of different types. Often, techniques developed for image classification with localization are used and demonstrated for object detection**.**

**Object Segmentation**

Object segmentation, or semantic segmentation, is the task of object detection where a line is drawn around each object detected in the image. Image segmentation is a more general problem of spitting an image into segments. Object detection is also sometimes referred to as object segmentation. Unlike object detection that involves using a bounding box to identify objects, object segmentation identifies the specific pixels in the image that belong to the object. It is like a fine-grained localization.

**Image Colorization**

Image colorization or neural colorization involves converting a gray scale image to a full color image. This task can be thought of as a type of photo filter or transform that may not have an objective evaluation.

**Image Reconstruction**

Image reconstruction and image in painting is the task of filling in missing or corrupt parts of an image. This task can be thought of as a type of photo filter or transform that may not have an objective evaluation.

**Image Synthesis**

Image synthesis is the task of generating targeted modifications of existing images or entirely new images. This is a very broad area that is rapidly advancing. It may include small modifications of image and video (e.g. image-to-image translations).

**2.4 OBJECT DETECTION**

Object detection is an important computer vision task used to detect instances of visual objects of certain classes (for example, humans, animals, cars, or buildings) in digital images such as photos or video frames. The goal of object detection is to develop computational models that provide the most fundamental information needed by computer vision applications: “What objects are where?”

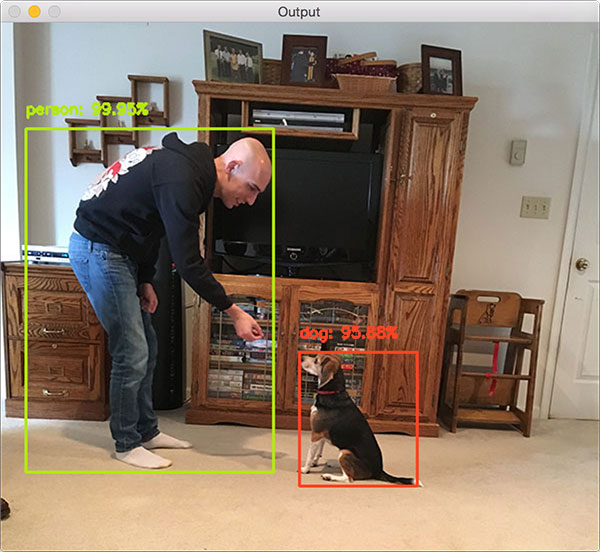
With time, the performance of this process has also improved significantly, helping us with real-time use cases. All in all, it answers the question: “What object is where and how much of it is there?

Fig 2.4 Object detection in deep learning

**2.4.1 ROLE OF OBJECT DETECTION IN CV**

Object detection is one of the fundamental problems of computer vision. It forms the basis of many other downstream computer vision tasks, for example, instance segmentation, image captioning, object tracking, and more. Specific object detection applications include pedestrian detection, people counting, face detection, text detection, pose detection, or number-plate recognition.

**2.4.2 WORKING OF OBJECT DETECTION**

Object detection can be performed using either traditional

* Image processing techniques or modern
* Deep learning networks

**Image Processing**

Image processing techniques generally don’t require historical data for training and are unsupervised in nature.

Hence, those tasks do not require annotated images, where humans labeled data manually (for supervised training). These techniques are restricted to multiple factors, such as complex scenarios (without unicolor background), occlusion (partially hidden objects), illumination and shadows, and clutter effect**.**

**Deep learning Networks**

Deep Learning methods generally depend on supervised training. The performance is limited by the computation power of GPUs that is rapidly increasing year by year. Deep learning object detection is significantly more robust to occlusion, complex scenes, and challenging illumination. A huge amount of training data is required; the process of image annotation is labor-intensive and expensive. For example, labeling 500’000 images to train a custom DL object detection algorithm is considered a small dataset. However, many benchmark datasets (MS COCO) provide the availability of labeled data.

**2.4.3 OBJECT DETECTION ALGORITHMS**

The field of object detection is not as new as it may seem. In fact, object detection has evolved over the past 20 years. The progress of object detection is usually separated into two separate historical periods.

**Traditional Object Detection period**

1. Viola-Jones Detector (2001), the pioneering work that started the development of traditional object detection methods.
2. HOG Detector (2006), a popular feature descriptor for object detection in computer vision and image processing.
3. DPM (2008) with the first introduction of bounding box regression.

**Deep Learning Detection period**

Most important two-stage object detection algorithms are

1. RCNN (2014).
2. Fast RCNN and Faster RCNN (2015).
3. Mask R-CNN (2017).
4. Pyramid Networks (2017).

Most important one-stage object detection algorithms are

1. YOLO (2016).
2. SSD (2016).
3. RetinaNet (2017).
4. YOLOv3 (2018).
5. YOLOv4 (2020).
6. YOLOR (2021).

****

Fig 2.4.2 Deep Learning based object detection for vehicles

**2.5 ADVANCEMENT IN OBJECT DETECTION**

Object detection is one of the most fundamental and challenging problems in computer vision. It has received great attention in recent years, especially with the success of deep learning methods that currently dominate the recent state-of-the-art detection methods. Object detection is increasingly important for computer vision applications in any industry. Currently the YOLO V5 and YOLOR are the algorithms invented in the domain of object detection and localization. A team of Google researchers have developed an advanced image classification and detection algorithm called GoogLeNet, that can recognize a wide range of objects. Google says in a blog post that ‘GoogLeNet’ algorithm can quickly recognize objects in images and can also locate and label multiple object sizes in one image. The software can even determine an object within or on top of an object.



Fig 2.5 An Advancement in object detection by Google

**CHAPTER 3**

**LITERATURE SURVEY**

**TITLE:** 3.1 Real-Time Social Distancing Detector Using Social Distancing Net-19 Deep Learning Network

**YEAR:** 2020

**AUTHOR:**  Rinkal Keniya

**CONCEPT:** With no doubt, the COVID-19 pandemic has put the world to halt. The world we lived in a few months prior is completely different than what it is now. The virus is spreading quickly and is a danger to the human race. Seeing the necessity of the hour one must always take certain precautions of which one being social distancing. Maintaining social distancing during COVID-19 is a must to ensure a slowdown in the growth rate of new cases. Our manuscript focuses on detecting if the people around are maintaining social distancing or not. Using our own self developed model named Social distancingNet-19 for detecting the frame of a person and displaying labels, they are marked as safe or unsafe if the distance is less than a certain value. This system can be used for monitoring people via video surveillance in CCTV.

**TITLE:** 3.2 Social Distancing Detection with Deep Learning Model

**YEAR:** 2021

**AUTHOR:** Saharsh Arya, Leena Patil, Ayushi Wadegaonkar, Nishad Shinde, Palash Gorsia

**CONCEPT :** The paper presents a methodology for social distancing detection using deep learning to evaluate the distance between people to mitigate the impact of this coronavirus pandemic. The detection tool was developed to alert people to maintain a safe distance with each other by evaluating a video feed. The video frame from the camera was used as input, and the open-source object detection pre-trained model based on the YOLOv3 algorithm was employed for pedestrian detection. Later, the video frame was transformed into top-down view for distance measurement from the 2D plane. The distance between people can be estimated and any noncompliant pair of people in the display will be indicated with a red frame and red line. The proposed method was validated on a pre-recorded video of pedestrians walking on the street. The result shows that the proposed method is able to determine the social distancing measures between multiple people in the video. The developed technique can be further developed as a detection tool in real-time application.

**TITLE:** 3.3 Monitoring social distancing under various low light conditions with deep learning

**YEAR:** 2021

**AUTHOR:** Adina Rahim, Ayesha Maqbool, Tauseef Rana

**CONCEPT:** The purpose of this work is to provide an effective social distance monitoring solution in low light environments in a pandemic situation. The raging coronavirus disease 2019 (COVID-19) caused by the SARS-CoV-2 virus has brought a global crisis with its deadly spread all over the world. h toll. In this paper, a deep learning-based solution is proposed for the above-stated problem. The proposed framework utilizes the you only look once v4 (YOLO v4) model for real-time object detection and the social distance measuring approach is introduced with a single motionless time of flight (ToF) camera. The risk factor is indicated based on the calculated distance and safety distance violations are highlighted.

**TITLE:** 3.4 A deep learning-based social distance monitoring framework for COVID-19

**YEAR:** 2021

**AUTHOR:** ImranAhmed, Misbah Ahmada, Joel J.P.C. Rodrigues, Gwanggil Jeon, Sadia Din

**CONCEPT:** The ongoing COVID-19 corona virus outbreak has caused a global disaster with its deadly spreading. The risks of virus spread can be minimized by avoiding physical contact among people. The purpose of this work is, therefore, to provide a deep learning platform for social distance tracking using an overhead perspective. The framework uses the YOLOv3 object recognition paradigm to identify humans in video sequences. The transfer learning methodology is also implemented to increase the accuracy of the model. In this way, the detection algorithm uses a pre-trained algorithm that is connected to an extra trained layer using an overhead human data set. The detection model identifies peoples using detected bounding box information. Using the Euclidean distance, the detected bounding box centroid's pairwise distances of people are determined. To estimate social distance violations between people, we used an approximation of physical distance to pixel and set a threshold. A violation threshold is established to evaluate whether or not the distance value breaches the minimum social distance threshold. In addition, a tracking algorithm is used to detect individuals in video sequences such that the person who violates/crosses the social distance threshold is also being tracked. Experiments are carried out on different video sequences to test the efficiency of the model. Findings indicate that the developed framework successfully distinguishes individuals who walk too near and breaches/violates social distances; also, the transfer learning approach boosts the overall efficiency of the model.

**TITLE:** 3.5 A social distancing detector using a Tensorflow object detection model, Python and OpenCV

**YEAR:** 2020

**AUTHOR:** Basile Roth

**CONCEPT:** During the quarantine I was spending time on github exploring Tensorflow’s huge number of pre-trained models. While doing this, I stumbled on a repository containing 25 pre-trained object detection models with performance and speed indicators. Having some knowledge in computer vision and given the actual context, I thought it could be interesting to use one of these to build a social distancing application.

**CHAPTER 4**

**IMPLEMENTATION**

**4.1 EXISTING SYSTEM**

**Monitoring COVID-19 social distancing with person detection via Faster R-CNN and SSD**

The absence of any active therapeutic agents and the lack of immunity against COVID19 increase the vulnerability of the population. Since there are no vaccines available, social distancing is the only feasible approach to fight against this pandemic. Motivated by this notion, this article proposes a deep learning based framework for automating the task of monitoring social distancing using surveillance video.

The proposed framework utilizes the faster R-CNN and SSD model to segregate humans from the background. Here the results are compared by calculating the mean average precision value (mAp), frames per second (FPS) and loss values defined by object classification and localization. Later, the pairwise vectorized norm is computed based on the three-dimensional feature space obtained by using the centroid coordinates and dimensions of the bounding box. From the experimental analysis, it is observed that the Faster R-CNN displayed best results with balanced mAP and FPS score to monitor the social distancing in real-time. The great trade-off between speed and accuracy of the detection which is dependent on various factors like backbone architecture feature extraction network input sizes, model depth, varying software and hardware environment. The ratio of accuracy to the number of parameters is highest for Inception v2 model indicating that Inception v2 achieved adequate classification accuracy with minimal trainable parameters in contrast to other models.

**4.1.1 DISADVANTAGES FACED**

* Since R-CNN and SSD uses selective algorithm, it would take a lot of time to process out the frame.
* Faster R-CNN requires at least 100 milliseconds per image.
* Reduce image size by half in width and height lowers accuracy.
* SSD is fast but performs worse for small objects comparing with others.
* SSD requires a large amount of data for training purposes. This can be quite expensive and time-consuming depending on the application.
* Despite improvements over RCNN and Fast RCNN, it still requires multiple passes over a single image unlike YOLO.
* FRCNN has many components—the convolutional network, Regions of Interest (ROI) pooling layer and Region Proposal Network (RPN). Any of these can serve as a bottleneck for the others.
* RPN is trained where all anchors in the mini-batch, of size 256, are extracted from a single image.

**PERFORMANCE COMPARISON OF THREE ALGORITHMS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **TT (in sec)** | **ROI** | **mAP** | **TL** | **FPS** |
| Faster R-CNN | 9651 | 12135 | 0.969 | 0.02 | 3 |
| SSD | 2124 | 1200 | 0.691 | 0.22 | 10 |
| **YOLO** | **5659** | **7560** | **0.846** | **0.87** | **23** |

Fig 4.1.1 Performance measure of faster R-CNN, SSD and YOLO

**4.2 PROPOSED SYSTEM**

The proposed system uses out the YOLO v3 algorithm to detect the pedestrians in the video frame and further categorizing out only the person as the classifying object excluding out the other objects. Bounding boxes are drawn over the detected pedestrian and the annotated parameters of these boxes are taken and further used for distance measurement. The location for each pedestrian can be estimated based on the top-down view. The distance between pedestrians can be measured and scaled. Depending on the preset minimum distance, any distance less than the acceptable distance between any two individuals will be indicated with red lines that serve as precautionary warnings.

The major motive and the idea behind using YOLO algorithm is that it performs object detection in a single shot as compared with the other algorithms such as Faster R-CNN and SSD. The mAp and FPS values are well optimized and balanced for the YOLO implementation.

**4.2.1 REASONS FOR CHOOSING YOLO v3**

As a real-time object detection system, YOLO object detection utilizes a single neural network. The latest release of Image AI v2.1.0 now supports training a custom YOLO model to detect any kind and number of objects. Convolutional neural networks are instances of classifier-based systems where the system repurposes classifiers or localizers to perform detection and applies the detection model to an image at multiple locations and scales.

YOLO divides each image into a grid of S x S and each grid predicts N bounding boxes and confidence. The confidence reflects the accuracy of the bounding box and whether the bounding box actually contains an object (regardless of class).

YOLO also predicts the classification score for each box for every class in training so that combining it is easy to classify both the classes and to calculate the probability of each class being present in a predicted box. Darknet-53 has 53 convolutional layers, its deeper than YOLOv2 and it also has residuals or shortcut connections. As you can see, Darknet-53 is better than ResNet-101 but 1.5 times faster and is as accurate as ResNet-152 but 2x times faster.

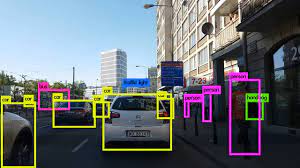


Fig 4.2.1.1 Object detection using YOLO v3

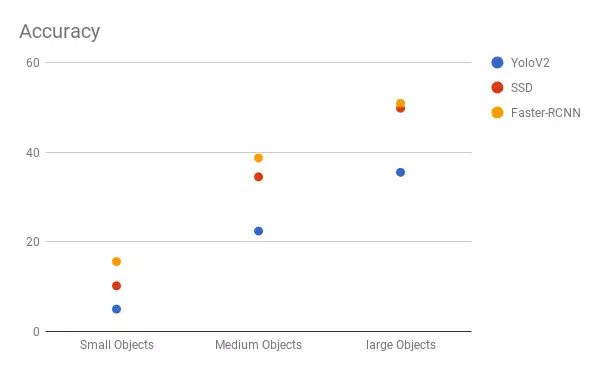
****

Fig 4.2.1.2 Accuracy measures

**4.2.2 DETECTING THROUGH YOLO V3**

We presented a computer vision technique for detecting people via a camera installed at the roadside or workspace. The camera field-of-view covers the people walking in a specified space. The number of people in an image and video with bounding boxes can be detected via these existing deep CNN methods where the YOLO method was employed to detect the video stream taken by the camera frame by frame. Pedestrian class was used and other object classes are ignored in this application. Hence, the bounding box best fits for each detected pedestrian can be drawn in the image, and these data of detected pedestrians will be used for the distance measurement

The most salient feature of v3 is that it makes detections at three different scales. YOLO is a fully convolutional network and its eventual output is generated by applying a 1 x 1 kernel on a feature map. In YOLO v3, the detection is done by applying 1 x 1 detection kernels on feature maps of three different sizes at three different places in the network. YOLO v3 makes prediction at three scales, which are precisely given by down sampling the dimensions of the input image by 32, 16 and 8 respectively. The first detection is made by the 82nd layer. For the first 81 layers, the image is down sampled by the network, such that the 81st layer has a stride of 32. If we have an image of 416 x 416, the resultant feature map would be of size 13 x 13.One detection is made here using the 1 x 1 detection kernel, giving us a detection feature map of 13 x 13 x 255.

Then, the feature map from layer 79 is subjected to a few convolutional layers before being up sampled by 2x to dimensions of 26 x 26. This feature map is then depth concatenated with the feature map from layer 61. Then the combined feature maps is again subjected a few 1 x 1 convolutional layers to fuse the features from the earlier layer (61). Then, the second detection is made by the 94th layer, yielding a detection feature map of 26 x 26 x 255. YOLO v3, in total uses 9 anchor boxes. Three for each scale. There is no more a need of soft maxing the classes for individual object’s probability of class score.

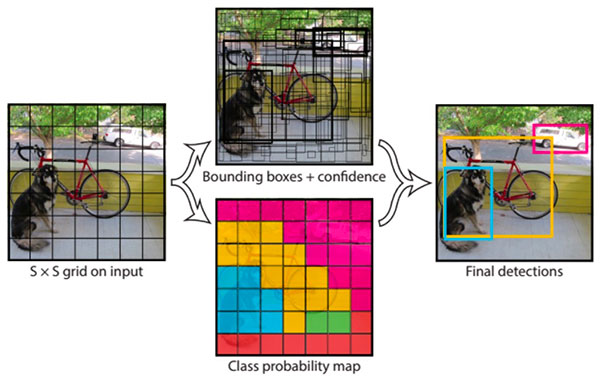


Fig 4.2.2 YOLO v3 object detection

**4.2.3 ADVANTAGES OF YOLO V3**

* YOLOv3 achieves 57.9 % mAP on the MS COCO dataset compared to DSSD513 of 53.3% and Retina Net of 61.1%.
* YOLOv3 uses multi-label classification with overlapping patterns for training. Hence it can be used in complex scenarios for object detection.
* Because of its multi-class prediction capabilities, YOLOv3 can be used for small object classification while it shows worse performance for detecting large or medium-sized objects.
* Darknet-53 has 53 convolutional layers, its deeper than YOLOv2 and it also has residuals or shortcut connections. ... As you can see, Darknet-53 is better than ResNet-101 but 1.5 times faster and is as accurate as ResNet-152 but 2x times faster.

**4.3 MODULES AND LIBRARIES**

**Numpy**

Numpy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of Numpy. Numpy is open-source software and has many contributors.

**Pandas**

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

**Matplotlib**

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API .

**OpenCV**

OpenCV is a highly optimized library with focus on real-time applications. OpenCV (Open Source Computer Vision Library) is a library of programming functions mainly aimed at real-time computer vision. OpenCV is written in C++ and its primary interface is in C++, but it still retains a less comprehensive though extensive older C interface.

**4.4 SYSTEM ARCHITECTURE**

This system which is proposed presents a method for detecting people using computer vision. Instead of using drone technology, the input is a stream of a video sequence from a CCTV camera installed. The camera’s range of view covers the pedestrians passing by in the range of the installed camera. The people in the frame are represented using a bounding box using the deep CNN models. The deep CNN based YOLO algorithm is used to detect the people in the sequence of video streams taken by the CCTV camera. The calculations are done by measuring the centroid distance between the pedestrians, this will represent whether the pedestrians in the video follow sufficient social distance.

**4.4.1 PROCESS DIAGRAM**

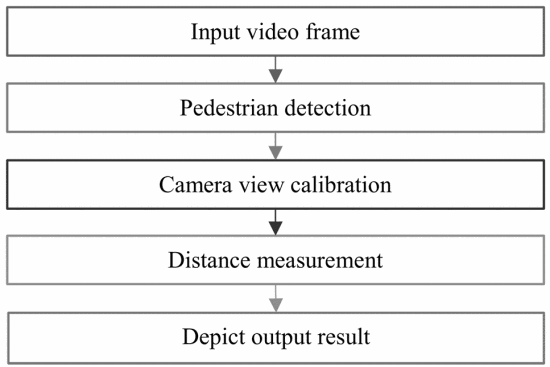
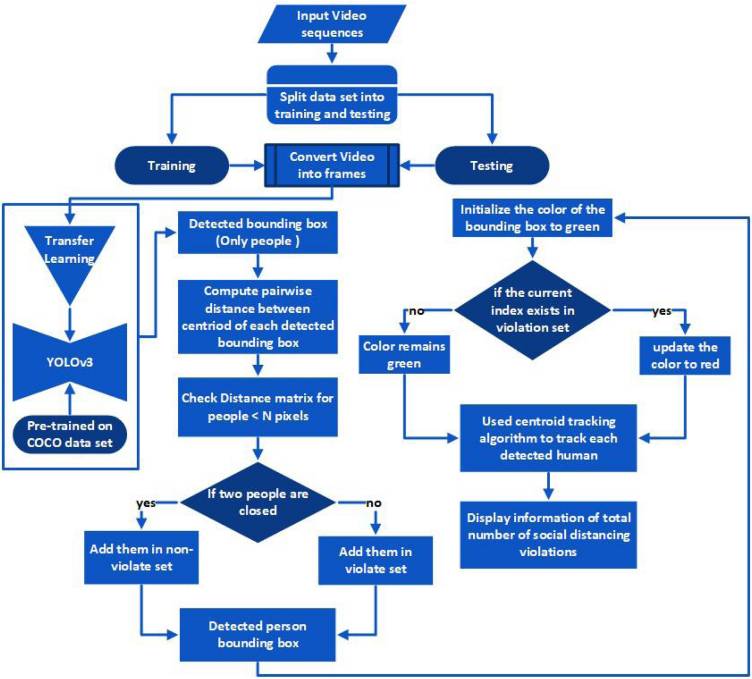
****

Fig 4.4.1 Process Diagram of the Proposed model

**4.4.2 FLOW DIAGRAM**

**** Fig 4.4.2Flow Diagram

**4.4.2 OUTLINE OF EACH STEPS**

**INPUT VIDEO FRAME**

The image captured and video recorded by the CCTV camera is given as the input. The given input will be of three cases such as good quality image, low resolution video, and live web cam recorded video in real time situation.

**PEDESTRIAN DETECTION**

Deep Convolutional Neural Networks model is a simple and efficient model for object detection. This model considers the region which contains only “Person” class and discards the regions that are notlikely to contain any object**.**

**CAMERA VIEW CALIBRATION**

The region of interest (ROI) of an image or a video frame focused on the person who is walking was captured using a CCTV camera was then changed into a two-dimensional bird’s view. The changed view’s dimension is 480 pixels on all sides. The calibration is done by transforming the view frame captured into a two-dimensional bird’s view. The camera calibration is done straightforwardly using OpenCV.

**DISTANCE MEASUREMENT**

The interval between the set of individuals in an input frame can be easily calculated once the bounding box for each person is mapped. By euclidean’s distance algorithm the center of each object is easily calculated and further the distance between a pair of people is computed and finally necessary alert is produced.

**DEPICT OUTPUT RESULT**

The view of the video is changed and every person within the given range of the camera’s view is detected. Every person who is detected in the frame is represented using points and circles. The individual whose distance is lower than the acceptable minimum threshold value is represented by red points and in other case it remains as the green bounded box.

**4.5 ALGORITHMS**

The object detection algorithms in the state of the art algorithm includes two-stage and single-stage detection algorithms. Out of which R-CNN , Fast R-CNN and Faster R-CNN comes under the category of two-stage whereas YOLO V3 through which we have proposed the model comes under the category of single-stage.

**4.5.1 FASTER R-CNN WITH ARCHITECTURE**

* The faster RCNN is derived from its predecessors RCNN and fast RCNN, which rely on external region proposal approach based on selective search (SS).
* Ren et al. proposed the Region Proposal Network (RPN) which uses CNN models to generate the region proposals that made faster RCNN 10 times faster than fast RCNN.
* RPN module performs binary classification of an object or not an object (background) while classification module assigns categories for each detected object (multiclass classification) by using the region of interest (RoI) pooling on the extracted feature maps with projected regions.
* From the experimental analysis, it is observed that the Faster R-CNN displayed best results with balanced mAP and FPS score to monitor the social distancing in real-time.

**Architecture**

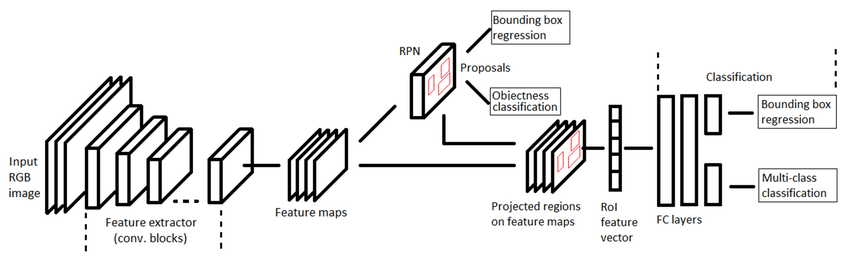
****

Fig 4.5.1 Schematic representation of Faster R-CNN architecture

**4.5.2 SSD WITH ARCHITECTURE**

* Single shot detector (SSD) [63] is also used as another object identification method to detect people in real time video surveillance system.
* For real-time processing, SSD further improves the accuracy and FPS by using multi scale features and default boxes in a single process.
* The principle of the feed-forward convolution network which generates bounding boxes of fixed sizes along with a score based on the presence of object class instances in those boxes.
* It consists of two steps: extracting feature maps and applying convolution filters to detect objects by using an architecture having three main parts. First part is a base pre-trained network to extract feature maps, whereas, in the second part, multi scale feature layers are used in which series of convolution filters are cascaded after the base network. The last part is a non-maximum suppression unit for eliminating overlapping boxes and one object only per box.

**Architecture**

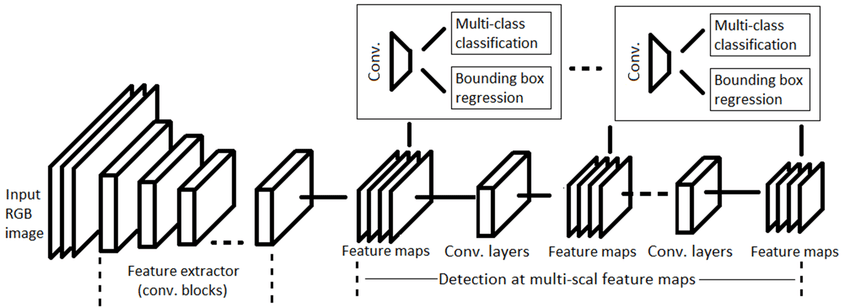
****

Fig 4.5.2 Schematic representation of SSD architecture

**4.5.3 YOLO V3 WITH ARCHITECTURE**

* YOLO considers the object detection problem as a regression task instead of classification to assign class probabilities to the anchor boxes. A single convolutional network simultaneously predicts multiple bounding boxes and class probabilities.
* YOLO v3 performs multi-label classification with the help of logistic classifiers instead of using soft max as in case of YOLO v1 and v2.
* Darknet-53 as a backbone architecture that extracts features maps for classification. In contrast to Darknet-19, Darknet-53 consists of residual blocks (short connections) along with the up sampling layers for concatenation and added depth to the network.
* YOLO v3 generates three predictions for each spatial location at different scales in an image, which eliminates the problem of not being able to detect small objects efficiently.

**Architecture**

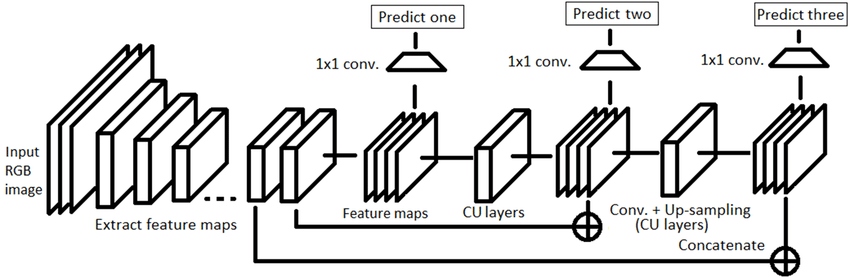
****

Fig 4.5.3 Schematic representation of YOLO v3 architecture

## 

## CHAPTER 5

## 

## DEVELOPMENT

**5.1 INPUT VIDEO FRAME**

The image captured and video recorded by the CCTV camera is given as the input. The camera is set up in a way it captures at a fixed angle and the video frame’s view was changed into a 2D bird’s view to accurately estimate the distance between each person. It is taken that the people within the frame are leveled on the horizontal plane. The interval between people is easily estimated, scaled, and measured by calculating the Euclidean distance between the centroids. A threshold value or a preset minimum value for the distance is set.

**5.2 PEDESTRIAN DETECTION**

This model considers the region which contains only “Person” class and discards the regions that are not likely to contain any object. This process of extracting the regions that contain the objects only is called as Region Proposals. The regions predicted by region proposal can vary in size and can be overlapping with other regions. So to ignore the bounding boxes surrounding the overlapping region, depending upon the Intersection Over Union (IOU) score maximum non suppression is used. The YOLOv3 is an object detection model that takes an image or a video as an input and can simultaneously learn and draw bounding box coordinates (tx, ty, tw, th), corresponding class label probabilities (P1 to Pc), and object confidence. The YOLOv3 is an already trained model on the Common Objects in Context dataset (COCO dataset). There are different objects present in a single frame, the goal is to identify “Only Person” class map bounding boxes related to only the people.

(X1, Y1)

**Person**

(X2, Y2)

Fig 5.2.1 Bounding Box representation

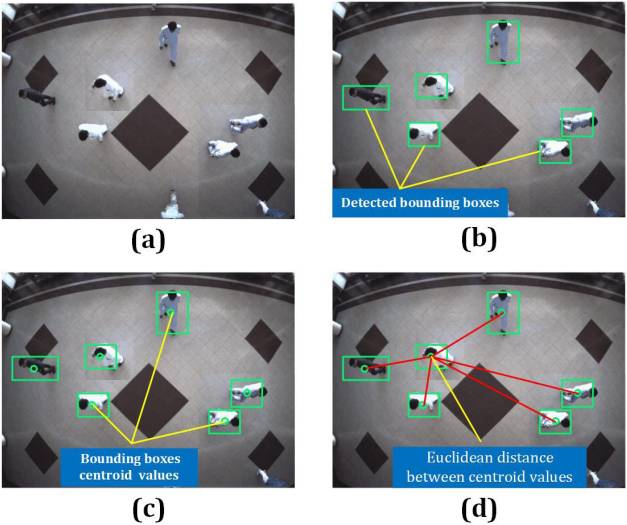


Fig 5.2.2 Pedestrian detected

**5.3 CAMERA VIEW CALIBRATION**

The region of interest (ROI) of an image or a video frame focused on the person who is walking was captured using a CCTV camera was then changed into a two-dimensional bird’s view. The changed view’s dimension is 480 pixels on all sides. The calibration is done by transforming the view frame captured into a two-dimensional bird’s view. The camera calibration is done straightforwardly using OpenCV. The transformation of view is done using a calibration function that selects 4 points in the input image/video frame and then mapping each point to the edges of the rectangular two-dimensional image frame. On performing this transformation, every person in the image/frame is considered to be standing on a leveled horizontal plane.

**5.4 DISTANCE MEASUREMENT**

The interval between the set of individuals in an input frame can be easily calculated once the bounding box for each person is mapped. To do so the bottom center of the box mapped to every person within the range is considered. For each person in the input frame, the orientation in the bird’s view transformation is calculated based on the central axis point of every person in the input frame. The distance interval of every set of people can be estimated from the bird’s view by calculating the Euclidean distance between centroids. As the camera is calibrated, more accurate results can be obtained.

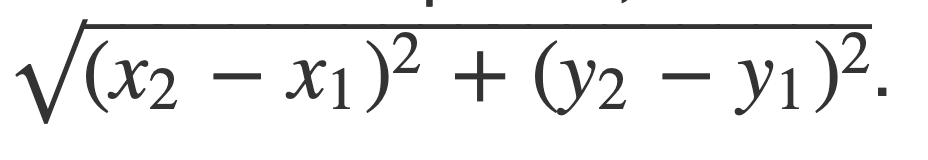


Fig 5.4.1 Euclidean distance formula for centroid tracking

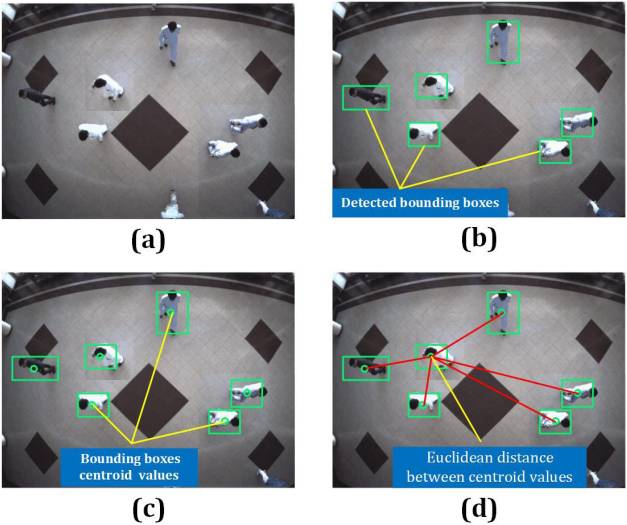


Fig 5.4.2 Euclidean distance metric calculation

**5.5 OUTPUT DEPICTION**

The pre-filmed video of people in a crowded area is taken as the input. As the input video is at an angle, the perspective of the recorded video is changed into a two-dimensional bird’s view frame by frame for the precise calculation of the pairwise distances between all detected people in a frame. The view of the video is changed and every person within the given range of the camera’s view is detected. Every person who is detected in the frame is represented using points and circles. The individual whose distance is lower than the acceptable minimum threshold value is represented by red points as shown in Figure 8 and the individuals who keep a safe distance from others are represented by green points.

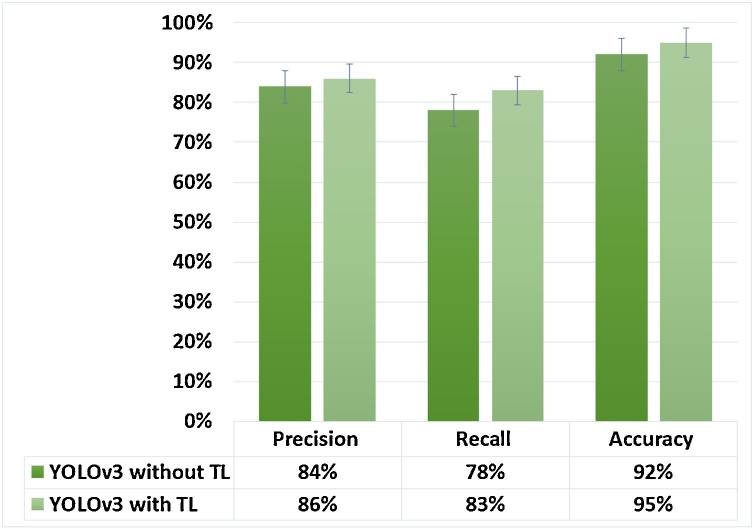
**5.6 PERFORMANCE EVALUATION**

Fig 5.6 Performance evaluation

CHAPTER 6

## SNAPSHOTS

## 6.1 IMAGE AS AN INPUT FRAME

## 

Fig 7.1 Monitoring in an image

**6.2 VIDEO AS AN INPUT FRAME**

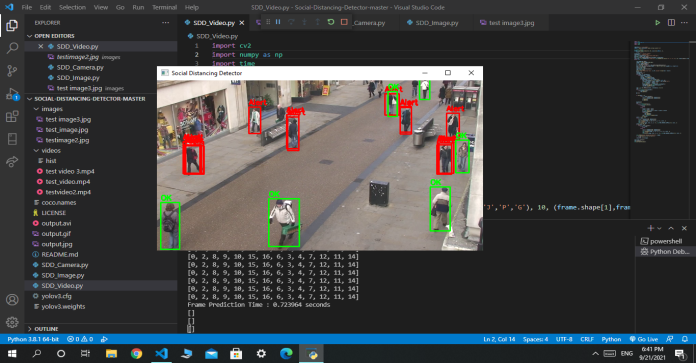
****

Fig 7.2 Monitoring in a video frame

**CHAPTER 7**

**CONCLUSIONS**

**7.1 CONCLUSION TO OUR SOLUTION**

* A methodology of social distancing detection tool using a deep learning model is proposed. By using computer vision, the distance between people can be estimated and any noncompliant pair of people will be indicated with a red frame and a red line**.**
* The developed system uses a pre-filmed video of people on a crowded street. The proposed system is capable of estimating the distance between people. The social distancing patterns are distinguished and classified as “Safe” and “Unsafe” distance
* The classifier can be implemented for live video streams and can be used for developing real-time applications. This system can be integrated with CCTV for surveillance of people during pandemics.

**7.2 FUTURE SCOPE**

Furthermore, the work can be further improved by optimizing the pedestrian detection algorithm, integrating other detection algorithms such as mask detection and human body temperature detection.

**CHAPTER 8**

# REFERENCES

1. Yew Cheong Hou - “Social distancing detection with deep learning model “ - 4th March 2020.
2. Sneha Madane - “Social Distancing Detection and Analysis through Computer Vision” - 10th May 2021
3. Adina Rahim - “Monitoring Social Distancing Using Artificial Intelligence for Fighting COVID-19 Virus Spread” - February 5 2021.
4. Jonathan Hui - “Object detection: speed and accuracy comparison (Faster R-CNN, SSD and YOLOv3” - March 28 2018
5. OpenCV - “https://opencv.org/”
6. Narinder Singh Punn - “Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3” - April 27 2021
7. Rohith Gandhi - “R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms” - July 10 2018
8. Adrian RoseBook - “ OpenCV Tutorial: A Guide to Learn OpenCV”