Identifying the Social distancing from Video Surveillance Cameras using Deep Learning Architectures

MSc Research Project
Data Analytics

Karan veer Singh
Student ID: x20146248

School of Computing
National College of Ireland

Supervisor: Aaloka Anant

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Karan veer Singh
Student ID:	x20146248
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Aaloka Anant
Submission Due Date:	31/01/2022
Project Title:	Identifying the Social distancing from Video Surveillance Cameras using Deep Learning Architectures
Word Count:	7236
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Karan veer Singh
Date:	30th January 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Identifying the Social distancing from Video Surveillance Cameras using Deep Learning Architectures

Karan veer Singh
x20146248

Abstract

In order to break the chain of spreading the COVID-19, social distancing is considered as an one of the effective measure. In order to encourage the people for practising the social distancing, we have proposed a social distancing monitoring framework which can identify the number of people violating the social distancing norms from surveillance camera videos in the real time. Our proposed framework trains the 3 different deep learning architectures over Open Image dataset. The architectures used for this research are Faster-RCNN, SSD-ResNet50 and YOLOv3. After performing the certain set of experiments, we have identified the YOLOv3 as the optimal model for object detection paradigm in order to identify the human in video sequences. The architecture identifies the people by adding bounding box information around the people. Using euclidean distance, the pairwise distance between the bounding boxes of their centroid is determined. The states are determined as safe, unsafe and moderated based on the closeness ratio between the individual.

1 Introduction

With the emerging threat of infection due to the fatal virus around the world, people have gone remote. This deadly virus known as Coronavirus has instilled fear in the mind of people. It is also known as COVID-19 which came into existence in the year 2019 at Wuhan which is a state in the country of China. World Health Organization (WHO), an international public health promoting body classified the COVID-19 as a pandemic. This rapidly spreading contagious virus created a sense of epidemic worldwide. At the time, around 4.77 million fatalities and 233.5 million confirmed cases were recorded by (WHO). The outbreak of this contagious virus hampered the daily routine of people around the world. It contained the people from socializing and changed the pre-covid era perspective of the people. With the sudden outburst of such a contagious virus, the researchers and medical systems around the world were inefficacious. At the moment, there was no suitable medication or therapy to fight the deadly virus. The researchers and health experts continued to create the remedy but couldn't resolve it. No therapy or recommendation has been made for the prevention or remedial treatment of the lethal virus. Meanwhile, to cease the rapidly growing infection rate around the world, a better alternative had to be followed. A preventive method had to be followed until an efficient curing method is created.

Over time, to sway the growing upwards Covid infection rate countries around the world shut their borders. They hindered the movements of their citizens within the country. The physical mode of interactions was restricted over all the public places whether it be offices or schools. These restrictions in various countries were generally known as ‘Lockdown’. Most of the countries issued the protocol of lockdown which was the best possible and known measure to decline the growing infection curve. The governments and communities around the world are compelled to make the best alleviation intend to restrict Covid spread. Multiple precautionary measures were also introduced to hinder the infection such as Social Distancing, Wearing Face masks, Sanitization and various voluntary precautions. Covid’s unusualness, fundamentals and heterogeneity made foreseeing the life span and spread of this pandemic difficult. In the current situation, social distancing has shown to be one of the most effective alternative methods of halting the spread. Social distance can sometimes be referred to as ‘physical distance’, implying that the space separating you and those around you is reinforced. Social distancing aids in reducing physical touch or engagement with persons who may be infected with COVID-19. Other precautions, like wearing masks and washing hands, should also be taken. To avoid the pandemic, infected people should stay at home if they are ill.

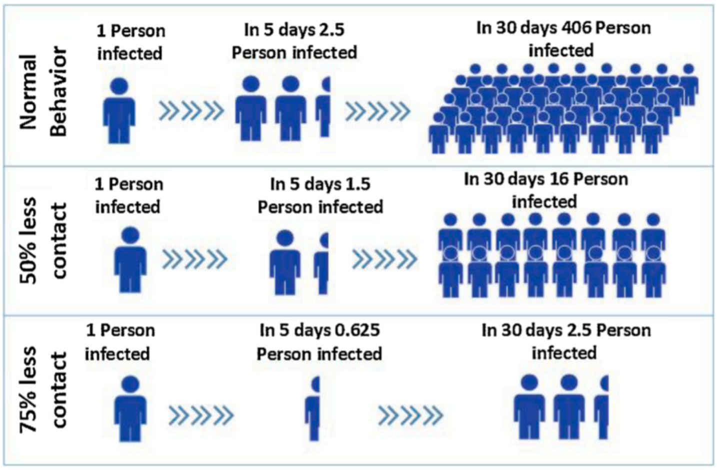

Figure 1: Importance of Social Distancing

But an extended set of restrictions or lockdown can have a concerning economical burden on the country due to which adhering to precautionary measures such as Social Distancing must be mandatory. These are some important approaches for breaking the cycle of infection and preserving the country’s dwindling economy. The issue now is to ensure that social etiquette is followed in real-time. During COVID-19 outbreaks, the government attempted to implement numerous social distancing measures, such as travel restrictions and border control. However, determining the amount of infection and the effectiveness of the limitations is not an easy process. People must cope with fundamental necessities such as food, health, and other tasks as well as work. At the initial times, the researchers and medical experts from the World Health Organization confirmed that the contagious virus could only be transmitted through the droplets sneezed from the infected person to the healthy person. But the studies performed on the later stage forfeited the previous study and showed the presence of virus spread through the air. Therefore, adhering to the precautionary measures was necessary for one own health benefits. Those seeking to assist the healthcare institution in dealing with COVID 19 issues and effective social distancing strategies have so investigated a variety of potential technological solutions, including Artificial Intelligence based technology.

As the guidelines laid by the WHO, every person following social distancing shall be 6 feet (1.83 metre) apart from every second person. To keep a check on the citizens to adhere to the protocols, a robust monitoring system must be available. In-person monitoring isn't suitable for the current situation and due to which a suitable technological system must be introduced. Real-time monitoring of protocol compliance is a challenge to the authorities. Utilization of the technologies such as Artificial Intelligence can play a key role in prevailing the challenges. With the help of Computer Vision, social distancing can be monitored and enforced remotely. It can work without human interventions. Computer Vision being a sub-field of Artificial Intelligence is very efficient with its output. With this system, a person can be monitored with certain objectives such as social distancing and locations. This system can be easily integrated with the existing monitoring appliances such as IP cameras and CCTV cameras. Through the real-time captured image sequence, the Computer Vision can monitor the social distancing adherence. This system can notify the authorities of the GPS location of the person who is not adhering to the protocols. The COVID-19 Social Distancing System complies with all the instruments of safety and security. It is a very convenient and easy-to-understand tool that can be operated without the guidance of an expert. This system may be used in public places such as hospitals, offices, government offices, schools and schools, construction sites, production facilities, airports, etc.

Various approaches like image classification, object detection are being utilized with Computer Vision, which can easily extract multiple characteristics of the image or video sequence. In our paper, we have developed an efficient Computer Vision-based Social distancing framework, which can identify the humans in the real time and also measure the distance between the individuals which can be used social distancing monitoring purposed. For object detection, our proposed work utilizes the 3 different transfer learning approaches, which includes the SSD-ResNet50, Faster-RCNN and YOLOv3. The outputs of these architectures will be compared in terms of mAP, FPS, Training time and loss score to provide a final overview about the most efficient model for object detection.

1.1 Research Question

Our research paper, will answer the following research questions.

- Which algorithm is most suitable for person detection in terms of performance and accuracy measures ?
- How does the Euclidean distancing method helps to detect the social distancing among the certain set of individuals ?

2 Literature Review

In this section, we will examine the latest developments by the researchers in the domain of our study. Here, each of the developments and their fallacies has been noted. This section is further divided into sub-sections namely, Object Detection, Image Classification, Machine Learning based Surveillance, and Social Distancing Interventions.

2.1 Study on Object Detection

Galvez et al. (2018) studied an efficient Convolutional Neural Network (CNN) algorithm for object detection. They compared two types of CNN model namely Single Shot Multibox detector (SSD) with MobileNetv1 and a faster region-based Convolution Neural Network (Faster-RCNN) with Inceptionv2. In their study, they utilized a dataset from personal videos and images which included 5 persons and 1 quadrotor. With their final interpretation, they concluded that SSD with Mobilenetv1 is a bit less efficient than Faster-RCNN with Inceptionv2. But the latter algorithm is faster than the counterpart which in this case could be considered for real-time. Although in this study, the drawback was that it couldn't show an algorithm that is both faster and efficient. In another study by Tripathi et al. (2018), they proposed a hybrid object detection algorithm. Generally, CNN is implemented for object detection but it has some fallacies which makes a need for a better algorithm. Object Detection is found to be more difficult to analyze, consumes massive energy which also makes it computation intensive. CNN requires a large dataset to extract features but its alternative, Scale Invariant Feature Transform (SIFT) algorithm can be implemented with a small dataset. Although, SIFT is less efficient than CNN. Therefore, the researcher of this study implemented a hybrid model of both CNN and SIFT algorithms to overcome the challenges.

Addapa et al. (2020) studied the Machine Learning (ML) approach in Amazon Web Service (AWS) for Object Detection. They used AWS Rekognition for Objection Detection as it is an easy-to-use tool that doesn't require any previous expertise in Machine Learning Technique. Here, the images or video dataset is inputted which in return gives output through recognition of objects, people, texts, etc. The AWS is competent enough for face comparisons, analysis and search. This system is built on robust Deep Learning (DL) techniques. Furthermore, custom labelling of the dataset can also be done using AWS Rekognition. In the study, it was found that the challenges of object detection are the dependency on various other computer vision techniques, which makes the model slow with lesser efficiency. V. N. Mandhala and N. (2021) studied Object Detection using ML for visually impaired people. The optimal objection detection model was tested upon various pre-trained approaches such as YOLOv3, Retina Net, YOLO Tiny. Object Detection plays a key role in the identification, classification of objects in images or video. This study would allow visually impaired people to identify objects and patterns in front of them. With an efficient model for this, challenges for visually impaired people can be overcome. This model detects accuracy based on the X-Y plane and indicates the person through a speech by converting input images. In the study, an efficient approach is suggested for object detection which provides an output on multiple factors for the productiveness.

2.2 Study on Image Classification

A paper proposed by Xin and Wang (2019) for an innovative training criterion of depth neural network for maximum interval minimum classification error based on the analysis of the error backpropagation algorithm. In the model, multiple deep learning models were implemented such as Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Naïve Bayes (NB), Random Forest (RF) and Decision Tree (DT). Furthermore, the dataset was fine-tuned by cross-entropy and M3CE for efficient output on the deep learning dataset, Modified National Institute of Standards and Technology database (MNIST) and Canadian Institute for Advanced Research (CIFAR-10). The comparison results in the study showed that the CNN algorithm has a higher efficient rate at 99.68% for the training set and 83.67% for the test set. Although, the RF algorithm consumes very little computation time. In another paper, Abhinav Patil (2021) studied image recognition using a Convolutional Neural Network algorithm. The dataset used in this were the images of cats and dogs. In the study, the results were obtained by using a custom neural network with the framework of CNN and Keras API. The image dataset of cats and dogs was obtained from the google repository. The model implemented here could easily classify images with multiple frames and dimensions. The accuracy of the model was an average of 92.5% with a layer filter of 256. This showed that a more robust machine and further enhancement of the model could increase the efficiency of the output.

In a paper, Singh (2019) discussed the implementation of Deep Learning (DL) for Image Classification. It was seen that the conventional approach has lower accuracy, non-optimal adaptiveness and effect. Therefore, the utilization of the DL approach can overcome these challenges. The author further stated that the image classification depends on four categories which are statistics, characteristics, shallow learning and deep learning. Image Classification through Machine Learning (ML) does have a limitation as it utilizes feature extraction for capturing features from images or videos. Here multiple algorithms were utilized such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Recursive Neural Network (RvNN). It was finally concluded that each algorithm has its own merits and demerits such as efficiency, performance and cost. Coming to another paper, Perez and Wang (2017) studied the data augmentation of Image Classification using deep learning techniques. Data augmentation helped increase the accuracy of classification output. Using the conventional way of augmentation is indeed efficient, but utilizing techniques such as Generative Adversarial Network (GAN), Neural Network Augmentation, VGG 16, Small Net will enhance the model output. The paper also suggested using multiple combinations of the approach to enhance the classification task. This technique can help overcome the challenges of overfitting, unbalanced data. The paper further showed the possibility of implementing transfer learning to extract and train a wide range of features in a dataset.

2.3 Study on Machine Learning-based Surveillance

Babanne et al. (2019) proposed a smart surveillance system that relied on the machine learning approach. This smart surveillance system is utilized to replace the conventional system which had higher costs and hardware. Not only that but this proposed system has a real-time alert generation system incorporated. As per the paper, this module can be utilized for various applications such as fire detection and alarm generation, Abnormal activity detection and prevention, Smart parking by using video sequences, and

Crowd accidental event identification. The greatest merit of this proposed approach is the earlier intimation of abnormal activates under the training of machine learning. This approach can be considered promising and can replace the current conventional approach. In another study by Saumya et al. (2020), a machine learning algorithm was utilized for the detection of bike riders without the helmet and triple riders in the Surveillance system. Human intervention for the monitoring of such a task will be a tiring scenario and wouldn't be efficient as well. In the paper, an approach is made to automate the monitoring system and making it furthermore efficient. Here a pre-trained model called YOLOv3 is implemented to identify the riders in the video or image sequence. This paper has explained the detailed methodology of the detection. The number of riders is tallied through the vertical binary projection to keep on the rider including the pillion exceeding two.

Sreenu and Durai (2019) did a review on an intelligent surveillance system relying on the deep learning approach for crowd analysis. The concept of Big Data was discussed in the paper. With the crowd analysis, it had a wider range of applications such as violation detection, behaviour observation, social distancing, etc. Furthermore, a Deep Convolutional framework was utilized for these types of applications. The most used DL algorithms were You Only Look Once (YOLO), VGG-16 Net, Long Short-Term Memory (LSTM). In addition to these, Kalman filter-based object discrimination algorithm is used and RNN with backpropagation through time (BPTT) is used for posture detection. But the fallacies struck is the real-world application with the tremendous growth of data. Similarly, F. Turchini and Bimbo (2018) studied a surveillance system based on deep learning for critical open areas. With this system, several tasks such as abnormal car behaviour detection, parking monitoring, generic anomaly detection and localization, determining and detecting of objects. The proposed model followed the mode of object detection, tracking, and counting. For detection of the object, the pre-trained model of YOLOv2 was utilized with a dataset either obtained from COCO or manual mode. For tracking an object, a greedy association multi-target tracker algorithm is utilized. This paper suggested a robust surveillance system based on a deep learning algorithm with real-time capture of video or image sequences from cameras. This system could allow multiple installations without any complex restructuring and could overcome the challenges of occlusion.

2.4 Study on Social Distancing Interventions

Moosa (2020) assessed the efficiency of social distance in the containment of Covid-19. For that aim, a quantitative assessment of effectiveness is devised and utilized in ten countries with varying proportions of effectiveness or viral confinement based on the time-series features of verified infections. Seven of those countries experienced social distancing, whereas three did not. Fahim Aslam (2020) investigated the effectiveness and significance of Social Distancing. He went over numerous additional measures made during the epidemic, as well as the methods utilized by other governments. In another study, George Milne and Xie (2020) used simulation to investigate the efficiency of social distance in limiting the transmission of Coronavirus. The four widely used methods for smoothing out the infection and reducing optimum daily infection proportions are highly efficacious: school closure, workplace non-attendance, increased case insulation, and community reduction. They were also proven to be effective ten weeks just after the benchmark case arrived. Case isolation was found to become the most effective sole technique when it was raised to 100% of children and 90% of adults. Mishra and Majumdar (2020) investigated the role of social distance in Indian society. She discovered that social distance effectively reduced the rate of infection and that social distancing will be maintained as a regular practice from now on.

During the covid epidemic, Wismans et al. (2021) investigated the relevance of cleanliness and social distance for community healthcare. Gumbu et al. (2020) investigated the influence of social distance caused by coronavirus amongst international pupils in Wuhan, China, in similar research. Hussain (2020) examined several nations' covid and social distancing policies. He contrasted the danger in each civilization to the rules in place. Matrajt and Leung (2020) conducted a study that used an epidemiological statistical equation to examine the efficacy of a social distance strategy in a medium-sized city. This research recommended more proactive strategies to reduce SARS-CoV-2 propagation. Social distancing measures, in addition to screening and interaction monitoring, are necessary to reduce the COVID-19 impact. There is still fresh material on the biological characteristics of SARS-CoV-2. Incorporating such evidence into statistical equations like theirs is critical for providing public health officials with the best tools for making critical decisions. Di Domenico et al. (2021) also studied the impact of social distancing on curbing the powerful variant of coronavirus that evolved during the months after its outbreak. The authors compared the statistics of social distancing implemented by the authorities in France. Utilizing the two-strain mathematical model, the genomic development of the lethal virus was observed and the curbing rate of the novel coronavirus strain was kept in check. On the other hand, Günay and Kurtulmus (2020) studied the impact of physical distancing in the United States service economy. The impact of multiple sub-sectors such as hotels, entertainment, restaurants and airlines were assessed in contrast with the social distancing protocols adherence rate. Statistical evaluations such as Iterated Cumulative Sum of Square (ICSS) and Markov regime-switching regression analysis were implemented to the model to consider the inter-dependence of the impact due to social distancing protocols in each sub-sector. Furthermore, this paper suggested an efficacious approach to align the economics of disrupted sectors.

The comparative analysis for the various studies by different researchers is shown in Table 2.4

Paper Title	Publish Year	Method	Advantages	Future Scope / Disadvantages
Object detection using convolutional neural networks	2018	Single Shot Multi-box detector (SSD) with MobileNetv1 and a faster region-based Convolution Neural Net-work (Faster-RCNN) with Inceptionv2	SSD faster but less efficient than Faster-RCNN algo	Couldn't get an algorithm which is both efficient and faster
Real time object detection using CNN	2018	Scale Invariant Feature Transform (SIFT)	CNN requires a largedataset to extract features but its alternative, Scale Invariant Feature Transform (SIFT)algorithm can be implemented with a small dataset.	SIFT works with small dataset but is very less efficient than CNN
Object detection/recognition using machine learning techniques in aws.	2020	Machine Learning (ML) approach in Amazon Web-Service (AWS) for Object Detection	AWS Rekognition for Objection Detection is an easy-to-use tool that doesn't require any previous expertise in Machine Learning Technique	no disadvantage found
Object detection using machine learning for visually impaired people	2021	YOLOv3, Retina Net, YOLO Tiny	This study would allow visually impaired people to identify objects and patterns in front of them. With an efficient model for this, challenges for visually impaired people canbe overcome.	Although there was amazing outputs but room for improvement is necessary.

Table 1 : Comparison of Different Works for this Study

3 Methodology

In order to control the COVID-19 infection, social distancing is an important measure which needs to be followed strictly by the people. With the aim of ensuring the social distancing at workplace and public places, in this research we are aiming to develop a social distancing monitoring system which will provide an insight about the extent of social distancing measures has been followed in a specific area. The following objective can be achieved by analyzing the real-time video streams obtained from the surveillance cameras. The overall process of detecting the social distancing can be performed into the two phases. The first phase will be object detection phase, in this phase we will identify or detect the people from the video streams. In the second phase, the distance between the two people will be measured in order to identify the closeness. If people are very close to each other then the system will mark them as unsafe and if people are far from each other then system will mark them as safe. In this work, we have successfully implemented both the phases by implementing the various techniques and technologies. We will discuss each step in detail in the upcoming subsections.

3.1 Person Detection

Person detection is the first step of the proposed model, which can be achieved in the several ways. In order to accurately achieve this, various subset of steps needs to be followed which includes the data collection, data pre-processing, Model training and model evaluation etc. Each step plays an important role in order to achieve the accurate results.

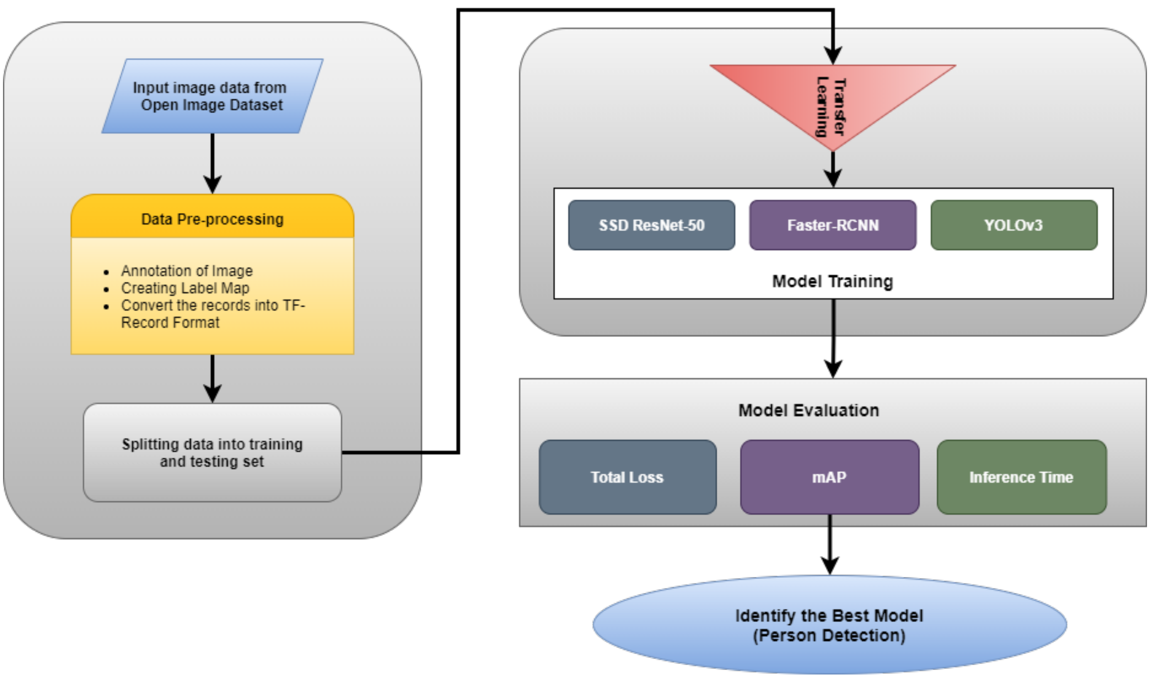

Figure 2: Proposed Methodology for Person Detection

3.1.1 Data Collection

Data collection is the first and primary stage, where we have collected the data from Open Image dataset *Projects* (n.d.). In order to extract the dataset, we have used the open image dataset toolkit. This toolkit provides the dataset of 9 million images with their rich annotations, defined in the multiple categories. Among those categories we have only collected the 5,000 images of person category. Each image will be associated with some label, which will contains the information about bounding boxes, localized narrative and object segmentation. At initial stage, the obtained image dataset is not tensorflow compatible. To utilize the dataset for training using tensorflow library, it needs to be processed.

3.1.2 Data Pre-processing

The obtained dataset of images needs to be converted into tensorflow record. In order to achieve that first annotation of image is performed and after that label map is create for each, which contains the information such as label id and label name. In our case, we are only considering the one label that is Person. After successful annotation the data needs to be converted into TFRecord format, for which we have used the certain scripts provided by the tensorflow zoo and later, the data has been splitted into training and test set.

3.1.3 Model Training

In order to train the obtained images from Open image dataset, the certain set of pre-processing the data needs to be splitted into training and test set. Training set will be utilized for model training, whereas, the test set will be utilized for evaluating the model performance. We have splitted the 70% of data for training and remaining 30% of data for testing. We are using the 3 different transfer learning models in order to detect bounding box on the person. The Transfer learning models used for this research are :

- SSD ResNet50 (Pre-Trained)
- Faster-RCNN (Pre-Trained)
- YOLOv3 (Pre-Trained)

The following transfer learning model with pre-defined weights will be utilized for training in order to detect the person with better accuracy. In order to train these model on the custom dataset, the training pipeline of these model needs to be configured. The configuration can be done by changing the various information in the configuration file such as number of classes, batch size, activation function, number of epochs etc. After successful configuration, the training process needs to be initiated by providing the following command as shown in Figure 3.

```
# Training the Model SSD ResNet50
python model_main_tf2.py --model_dir=models/my_ssd_resnet50_v1_fpn
--pipeline_config_path=models/my_ssd_resnet50_v1_fpn/pipeline.config

# Training the Model Faster-RCNN
python model_main_tf2.py --model_dir=models/my_faster_rcnn_resnet50_v1
--pipeline_config_path=models/my_faster_rcnn_resnet50_v1/pipeline.config

# Training the YOLOv3 Model
!./darknet detector train data/obj.data cfg/yolov4-obj.cfg
yolov4.conv.137 -dont_show -map
```

Figure 3: Commands to train the Pre-trained models on Open Image Dataset

3.1.4 Model Evaluation

After successful training of models over the training data, the performance of model needs to be analyzed using various performance metrics. The metrics that will be used to evaluate the model performance are total loss, mean average precision (mAP) and inference time. The model with minimum loss, maximum mAP and with low inference time will be considered as the most optimal model for person detection. The results obtained after evaluation will be discussed in chapter 6.

3.2 Social Distancing Monitoring

Model selection is the first and primary stage in order to detect the violation of social distancing. In the first step of social distance monitoring system is to select the best performing model for identifying the bounding boxes around the people. Once the bounding boxes are identified, the frame needs to be converted into Bird-eye view. Bird-eye view mainly represents top-down view of the scene. The transformation from perspective view to bird-eye view is shown in Figure 4. In order to achieve this, first 4 points on original image needs to be selected which needs to be transformed.

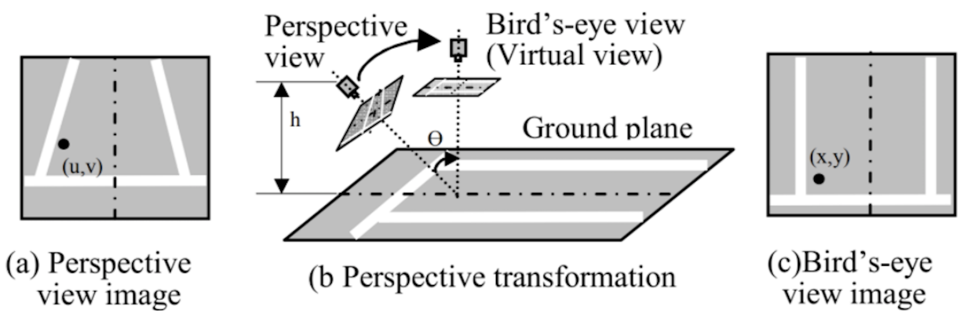

Figure 4: Transformation from Perspective view to bird-eye view

After successful detection of bounding box and frame transformation, all the centroid from the results has been extracted and the euclidean distance between all the pairs of centroid is calculated. If the centroid distance between any two centroid pair is less than threshold, then social distancing is violated. We are using the indexing mechanism, which justifies the closeness between the two people. If the closeness between between the set of individuals is very high, they will be represented with index0 and will be counted into unsafe zone. On the other hand, if the closeness is medium they are added into moderate zone (index=1) as those individuals may enter into unsafe zone, after certain period of time. Similarly, if the closeness between individual is low, then those individuals are considered into the safe zone (index=2). The count obtained from each individual will give a insight about the people violating the social distancing.

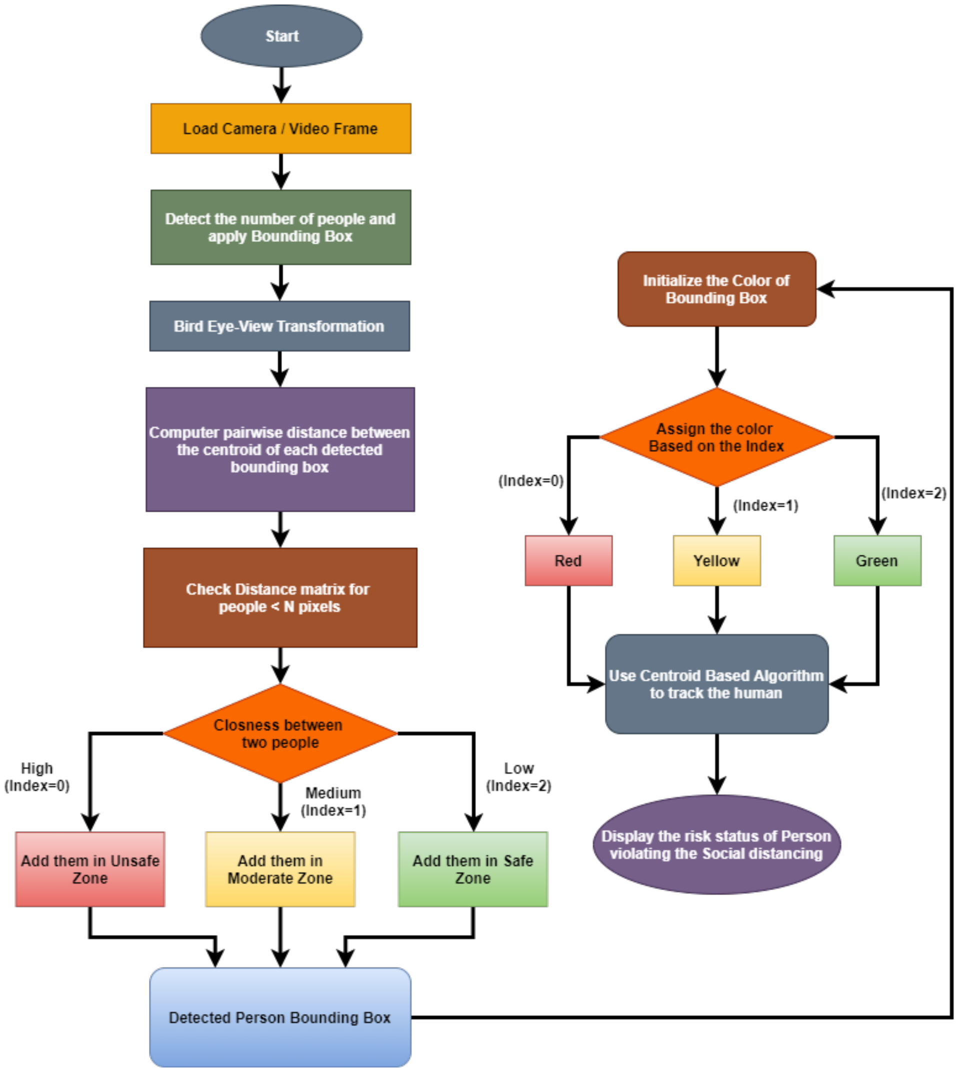

Figure 5: Social Distancing Monitoring Framework

Later, based on the index the color of boundary box has been updated as Red, yellow and green. The bounding box with red color are the individuals, who are violating the social distancing norms. Bounding box with yellow color indicates that individual is following the social distancing but may enter into unsafe zone and the bound box with green color represents the safe zone, where people are strictly following the social distancing.

4 Design Specification

In this work, the most tedious and challenging task is to identify the person from the video frames and create the bounding box around them, which will help to calculate the social distancing between the set of individuals. We have used 3 different methods in order to identify the best approach. In this section, we will discuss about the design and architecture part of each algorithm.

4.1 SSD-ResNet50

SSD-ResNet50 is a object detection architecture, where SSD stands for single short multibox detector which performs the localization and classification task in the single forward pass of the network. Whereas the ResNet-50 stands for residual network, which utilizes the 50 neural network layers. SSD-ResNet50 is also called as RetinaNet. In this work, we have utilized the pre-defined weights of this architecture in order to train the dataset. The convolutional layers works as a backbone in the SSD-ResNet50 architecture. The overall architecture of SSD-ResNet50 is shown in Figure 6. In the following architecture, the ResNet50 architecture is used for deep feature extraction and FPN (Feature pyramid network) is used for constructing a rich multi-scale feature pyramid from one single resolution of input image, it is available on the top of ResNet. Ca

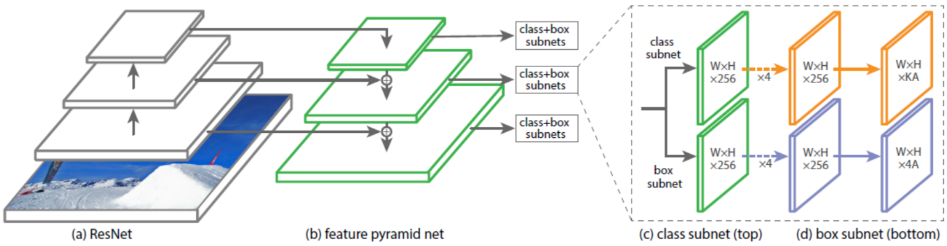

Figure 6: SSD-ResNet50 Architecture

4.2 Faster RCNN

Faster RCNN is one of the most popular architecture for object detection, which utilizes the concept of Convolutional neural network. Faster RCNN utilizes the region proposal network, which take the convolution feature map generated by CNN and generates the proposal. In the Faster RCNN a new layers called ROI pooling has been introduced which extracts the features vector from all the proposals in the same image. Faster RCNN more accurate as compared to the RCNN as it does not cache the extracted features also it build the network with single stage. On the other hand, RCNN has the different stages and requires large disk storage to store the extracted features. The another concept which makes the faster RCNN better as compared to RCNN as it does not store the pyramid of the image to different scales. Faster RCNN uses the concept of anchor box, each anchor box represents the different scales and aspect ratios. The architecture of Faster RCNN is shown in Figure 7.

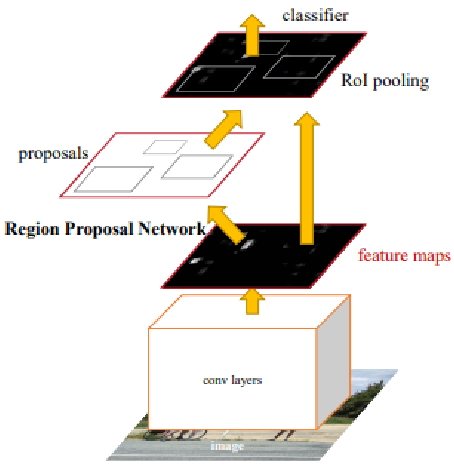

Figure 7: Faster RCNN Architecture

4.3 YOLOv3

You only look once (YOLO) is among the best models of deep learning architectures designed for object detection. It uses the variant of DarkNet, which has 53 layers trained on the ImageNet. This model utilizes the binary cross entropy for calculating loss for each label. The architecture of YOLOv3 is shown in Figure 8. YOLO is a fully convolutional neural network, which extract the features using DarkNet-53. Before feeding the images into YOLOv3, the images needs to be resized with the pixel ratio of 416 X 416. The output of bounding box can be represented by 6 numbers. YOLOv3 is pre-trained model over the COCO dataset, in our work we are training YOLOv3 also on Open Image data for person class.

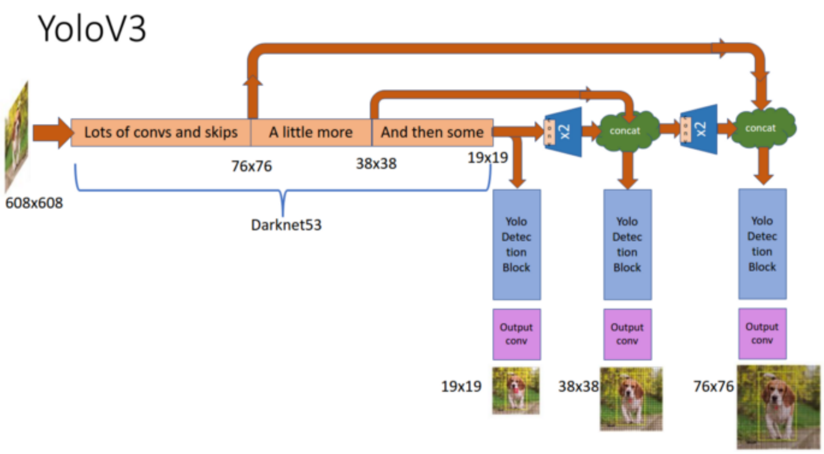

Figure 8: YOLOv3 Architecture

5 Implementations

The main objective of this research is to implement a social distancing monitoring tool, in order to control the spread of COVID-19. In this work, firstly, the main task of person detection with bounding boxes has been achieved using YOLOv3 model along with 2 other algorithm. After successful detection of person, the image frame has been converted into Bird-eye view and distance among the individual bounding boxes has been calculated using Euclidean Distance . The complete framework has been implemented using the python and there are several set of python libraries has been used in order to achieve the task. The libraries are OpenCV, numpy, pandas, tensorflow and tensorboard. Along with it as we are using the transfer learning algorithms, the pre-trained weight of all the 3 models has been downloaded and then code has been executed. We have changed the configuration files of these pre-trained models such as batch size, number of classes etc. After successful training of all these model over Open Image Dataset, these models has been saved and best model has been considered for detection of social distancing. As the model requires the large amount of image data for training, a system with good GPU capabilities is required to train the model. The output obtained after implementing the application is shown in Figure 9.

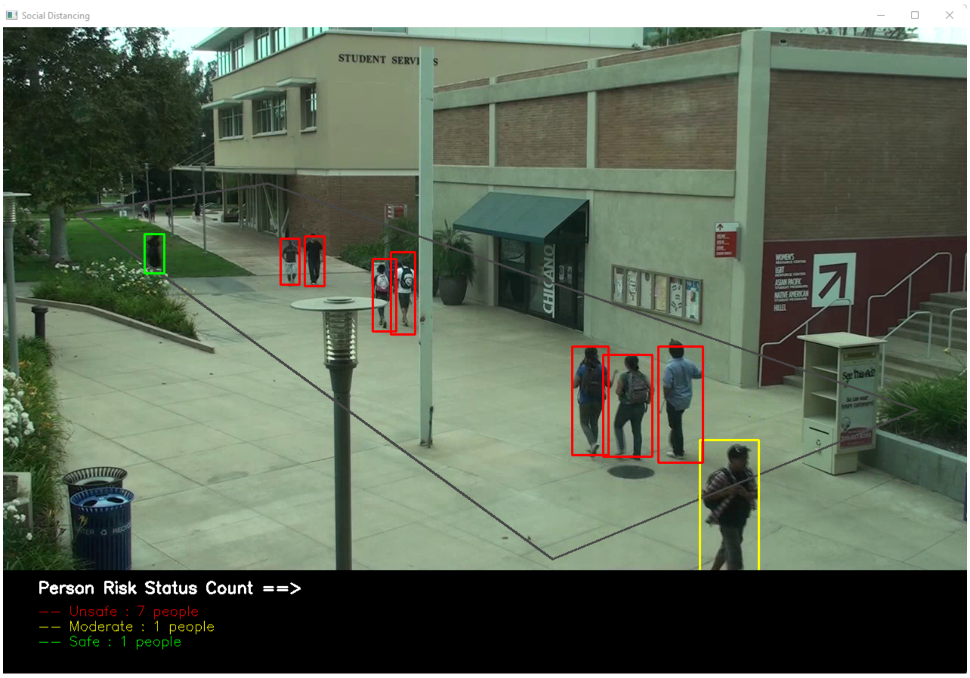

Figure 9: Implementation of Social distancing Monitoring application

After defining the path of video the application identifies the human in the real time and generate bounding box around them. In the application we are also calculating the Person risk status count, which informs about the number of people who are safe, unsafe and in moderate state. These states has also been represented with the color of bound boxes as Green, red and Yellow. The red color represents the unsafe state, green indicates the safe state and Yellow color indicates the moderate state.

6 Evaluation

In this work, we have used 3 different object detection model for identifying the persons class. In order to identify the best model, the performance comparison has been performed between the each algorithm. All the model have been trained over NVIDIA GTX 1660 GPU, where the data is acquired from open image dataset (OID) toolkit. Where almost 5000 images of the dataset has been acquired and divided into training and test set with ratio of 80:20. Where 80% of data is utilized for training and 20% of remaining data is used for testing. During the training phase, the performance each model is monitored along with total loss, mAP score, training time and FPS. The results obtain after all the certain set of experiments will be discussed in the upcoming subsection.

6.1 Experiment 1 / Evaluation Based on mAP Score

mAP mainly stands for the Mean Average precision. In order to evaluate the performance of object detection models mAP gives us a better insight. As it compares detected box with ground truth bounding box and provides a score accordingly. Model having high mAP score indicates the better accuracy score. In this work, we have calculated the mAP score over the test set for all the 3 models. Obtained mAP Score for all the models has been represented in the Figure 10.

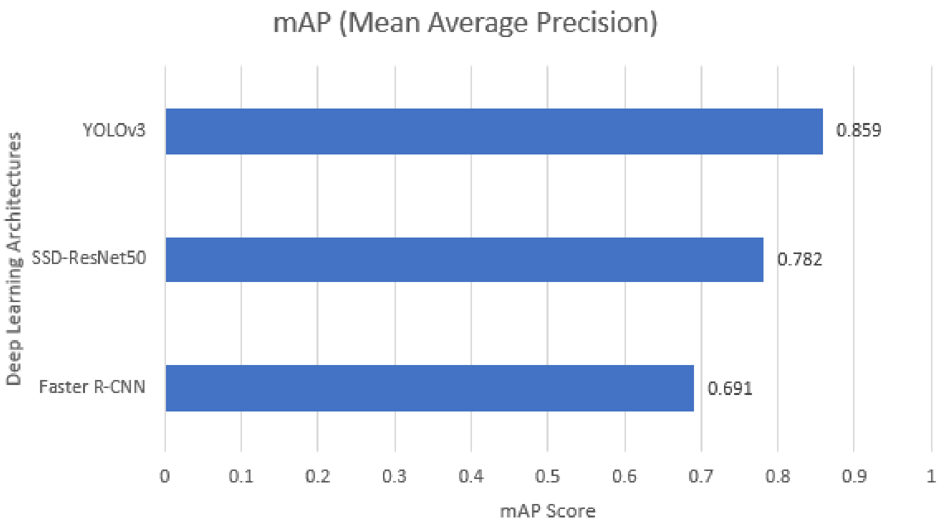

Figure 10: Mean Average Precion (mAP) Comparison

After analyzing the graph, it has been noted that the most accurate model is YOLOv3, as it has the highest mAP score of 0.859, followed by the SSD-ResNet50 with mAP score of 0.782 and faster-RCNN model provides the lowest mAP score of 0.691. The lowest mAP score of Faster R-CNN indicates that it does not classify the person class accurately.

6.2 Experiment 2 / Evaluation Based on Total Loss

Loss is another important metric which is required to be calculated for evaluating the model performance. The model with minimum loss is considered as the most optimal model in terms of performance. In this work the total loss has been calculated over every epoch by applying the early stopping mechanism. Each model has been trained over multiple set of iterations and then results has been collected. Total loss obtained after training the following algorithms are as follow as shown in Figure 11.

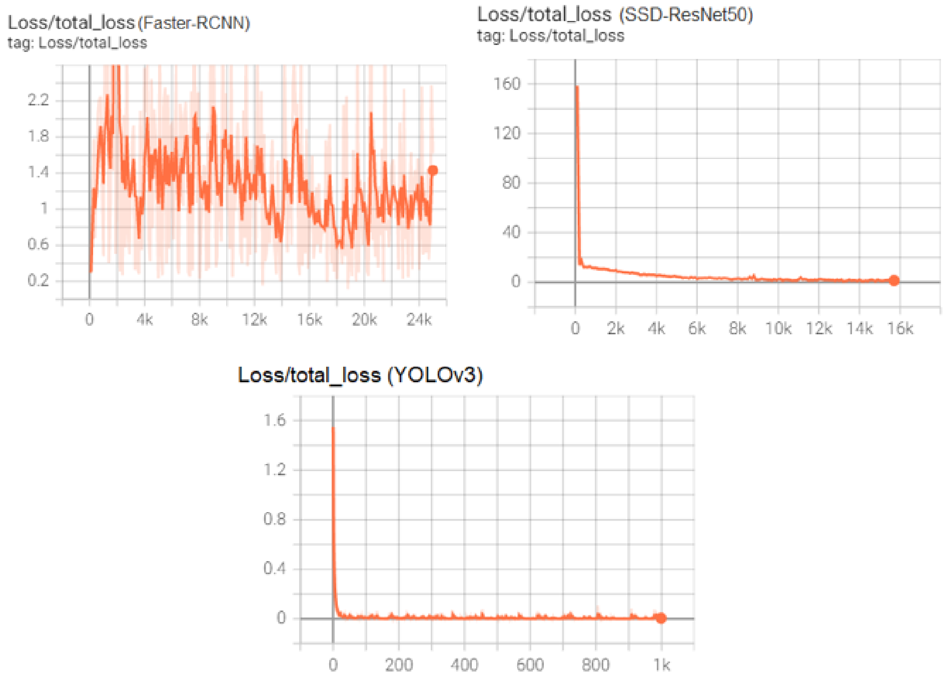

Figure 11: Total Loss Comparison

With the Faster RCNN model, we have achieved the minimum loss score of 1.618. Whereas, with the SSD-ResNet50 the minimum loss obtained is 1.561. On the other hand, the minimum loss obtained using YOLOv3 architecture is 0.0931. On performing the comparative analysis between all the models it has been found that YOLOv3 provide the minimum value of loss as compared to all the other models. The Faster-RCNN provides the highest loss score and hence it is not preferable model to detect the person class.

6.3 Experiment 3 / Evaluation Based on Training Time

Training time of deep learning architecture mainly refers to the time taken by each algorithm for training the input data. The model which can be trained in the minimum time and also can provide the good performance will be considered as the most efficient model. The comparison graph of all the architecture for training time is shown in Figure 12

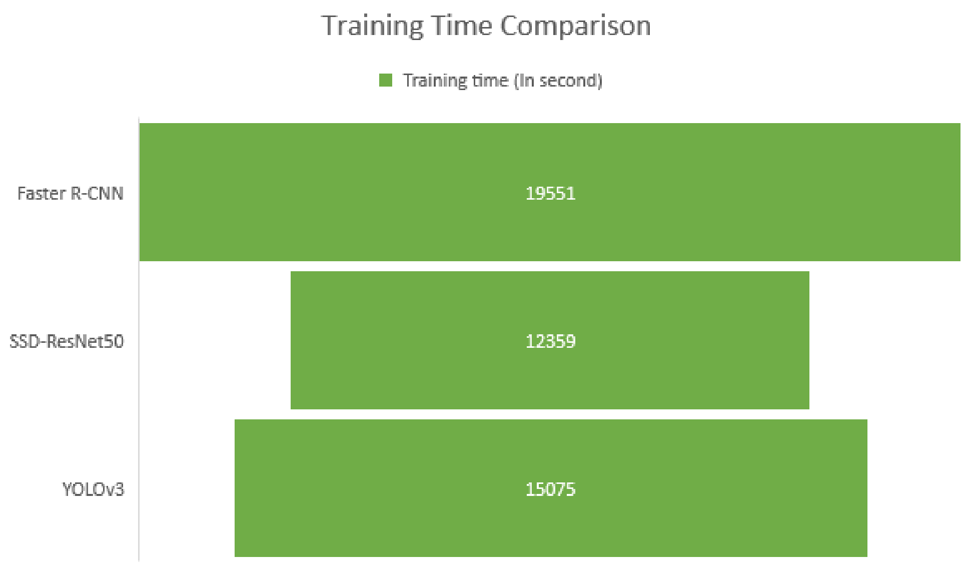

Figure 12: Training Time Comparison

After analyzing the graph it has been found that time consumed by Faster-RCNN for training is 19551 seconds, which is approximately equal to 6 hours. On the other hand, the SSD-ResNet50 consumed the training time of 12359 seconds, which is approximate equals to 4 hours and training time consumed by YOLOv3 is 15075 seconds, which is nearly equal to 4.2 hours. However, it can be said that SSD-ResNet50 take the minimum time for training, followed by YOLOv3 and Faster R-CNN.

6.4 Experiment 4 / Evaluation Based on FPS

For real time application, the FPS value should be high. An architecture with more FPS score will be able to process the number of frames from input video on per second basis. Model with lower FPS rate is not advised to be used for real-time application. The comparative analysis of proposed architecture is shown in Figure 13.

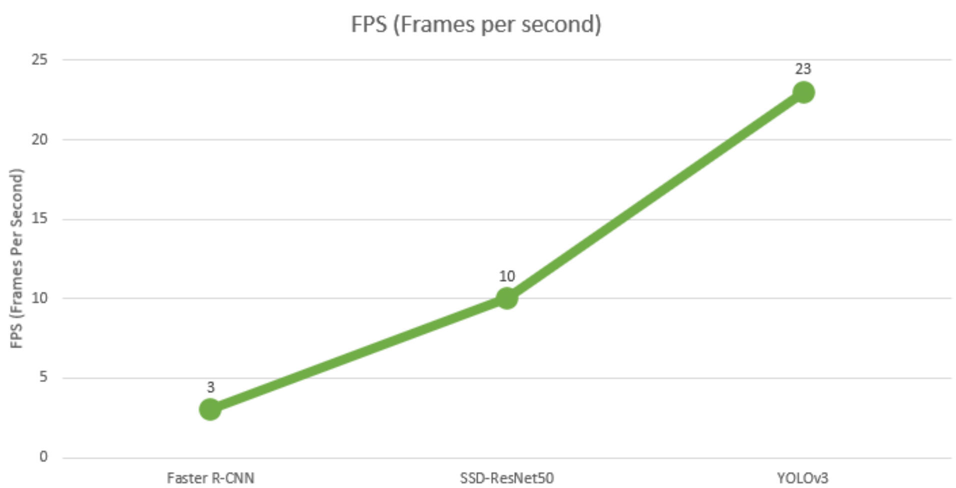

Figure 13: FPS Comparison

After analysing the frames per second record over the sample video, it has been found that Faster R-CNN can process the maximum 3 frames in a second. On the other hand, SSD-ResNet50 architecture achieved the FPS of 10 and the highest frame rate score has been obtained using YOLOv3, which can process the 23 frames in a second. Therefore, YOLOv3 will be the ideal choice for real-time application.

6.5 Discussion

The identification of person is the most crucial task, in order to identify the social distancing among the set of individuals. Therefore,after performing certain set of experiments for person detection it has been found that, YOLOv3 model provides us the most accurate results in terms of accuracy and performance as the calculated mAP score is highest, total loss is minimum and it can process the maximum number of frames in a second. The YOLOv3 can be the ideal choice for real-time object detection application. However, the training time of YOLOv3 is comparatively high as compared to the models like SSD-ResNet50. But it can be neglected, as training the deep learning architecture over input image is one time process. Discussing about the other 2 architectures, it can be said that in terms of performance and training time Faster R-CNN generates the poor results, it consumes the highest training time and generates the low mAP score and as the FPS rate of Faster R-CNN is very low, it can not be used to real-time object detection application. SSD-ResNet50 architecture somewhere lies in the middle of both the models in terms of accuracy as well as performance. However, the training time consumed using SSD-ResNet50 is minimum as compared to YOLOv3 and Faster-RCNN architecture.

7 Conclusion

The following research proposes an efficient, deep learning architecture based framework for monitoring the social distancing practices in the real-time. Using the proposed architecture, the individual has been identified in the real time with bounding boxes and closeness between the two objects has been calculated using euclidean distance. The current system utilizes the computer vision technology and the different deep learning architectures for person detection and also identifies the number of people violating the social distancing norms in the real-time. After certain set of experiments and analysis it has been found that among the 3 different deep learning architectures Faster-RCNN, SSD-ResNet50 and YOLOv3. YOLOv3 architecture outperforms in terms of both performance and accuracy with highest FPS and mAP score. The visualization results has shown that the our proposed framework correctly identifies the social distancing measure between the people, the following framework can be further implemented for various kind of places such as Office, Markets, Traffic signals, public places, schools and colleges which can help to reduce the spread of COVID-19. Further, this work can be extended by integrating the other features of application such as mask detection, human body temperature detection and so on. However, there are some concerns which can be raised about the privacy and individual rights which may require the personal consent from public. Maintaining the transparency and taking care about the individual's identity in general are the another area of debates, which needs to be discussed.

References

Addapa, A., Ramaswamy, P. and Mungara, K. (2020). Object detection/recognition using machine learning techniques in aws.

Aslam, F. (2020). Covid-19 and importance of social distancing.

Babanne, V., Mahajan, N. S., Sharma, R. L. and Gargate, P. P. (2019). Machine learning based smart surveillance system, *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 84–86.

Di Domenico, L., Pullano, G., Sabbatini, C., Lévy-Bruhl, D. and Colizza, V. (2021). Impact of january 2021 social distancing measures on sars-cov-2 b.1.1.7 circulation in france.

F. Turchini, L. Seidenari, T. U. and Bimbo, A. D. (2018). Deep learning based surveillance system for open critical areas, *Inventions* **3**.

Galvez, R. L., Bandala, A. A., Dadios, E. P., Vicerra, R. R. P. and Maningo, J. M. Z. (2018). Object detection using convolutional neural networks, *TENCON 2018 - 2018 IEEE Region 10 Conference* pp. 2023–2027.

Gumbu, Y., Marango, T. and Chitongo, L. (2020). The impact of social distancing as a response to covid19 among foreign students in wuhan, china, **61**: 190–206.

Günay, S. and Kurtulmus, E. (2020). Covid-19 social distancing and the us service sector: What do we learn?, *Research in International Business and Finance* **56**: 101361.

Hussain, A. H. M. B. (2020). Social distancing in risk society: a cross-national analysis of policy responses to the covid-19 pandemic.

Matrajt, L. and Leung, T. (2020). Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of coronavirus disease, *Emerging Infectious Diseases* **26**.

Milne, G. and Xie, S. (2020). The effectiveness of social distancing in mitigating covid-19 spread: a modelling analysis.

Mishra, M. and Majumdar, P. (2020). Social distancing during covid-19: Will it change the indian society?, *Journal of Health Management* **22**: 224–235.

Moosa, I. (2020). The effectiveness of social distancing in containing covid-19, *Applied Economics* **52**: 1–14.

Patil, A. (2021). Image recognition using machine learning, *Social Science Research Network* .

Perez, L. and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning, *ArXiv* **abs/1712.04621**.

Projects (n.d.).
URL: *https://opensource.google/projects/*

Saumya, A., Gayathri, V., Venkateswaran, K., Kale, S. and Sridhar, N. (2020). Machine learning based surveillance system for detection of bike riders without helmet and triple rides, *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 347–352.

Singh, S. (2019). Image classification using deep learning.

Sreenu, G. and Durai, M. A. (2019). Intelligent video surveillance: a review through deep learning techniques for crowd analysis, *Journal of Big Data* **6**: 48.

Tripathi, A., Kumar, T., Dhansetty, T. and Kumar, J. (2018). Real time object detection using cnn, *International Journal of Engineering and Technology(UAE)* **7**: 33–36.

V. N. Mandhala, D. Bhattacharyya, V. B. and N., T. R. (2021). Object detection using machine learning for visually impaired people.

(WHO), W. H. O. (2021). Who coronavirus (covid-19) dashboard.
URL: *https://covid19.who.int/*

Wismans, A., Letina, S., Wennberg, K., Thurik, R., Baptista, R., Burke, A., Dejardin, M., Janssen, F., Santarelli, E., Torrès, O. and Franken, I. (2021). The role of impulsivity and delay discounting in student compliance with covid-19 protective measures, *Personality and Individual Differences* **179**: 110925.

Xin, M. and Wang, Y. (2019). Research on image classification model based on deep convolution neural network, *EURASIP Journal on Image and Video Processing* **2019**.