



Real time deep learning framework to monitor social distancing using improved single shot detector based on overhead position

Bharathi Gopal¹ · Anandharaj Ganesan²

Received: 23 July 2021 / Accepted: 24 December 2021 / Published online: 11 January 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

The current COVID 19 halo infection has caused a severe catastrophe with its deadly spread. Despite the implementation of the vaccine, the severity of the infection has not diminished, and it has become stronger and more destructive. So, the only solution to protect ourselves from infection is social-distancing. Although social-distancing has been in practice for a long time, in most places it is not effectively followed, and it is very difficult to find out manually at all times whether people are following it or not. Therefore, we introduced a newly developed framework of deep-learning technique to automatically identify whether people maintain social-distancing or not using remote sensing top view images. Initially, we are detecting the context of image which includes information about the environment. Our detection model recognizes individuals using the boundary box. Then centroid is determined over every detected boundary box. By means of applying Euclidean distance, the pair range distances of the detected boundary box centroid are determined. To evaluate whether the distance measurement exceeds the minimum social distance limit, the violation threshold is established. We used Improved Single Shot Detector model for detecting a person over an image. Experiments are carried out on widely collected remote sensing images from various environments. Based on the object detection algorithm of deep learning, a variety of performance metrics are compared to evaluate the efficiency of the proposed model. Research outcome shows that, our proposed model outperforms well while recognize and detect a person in a well excellent way.

Keywords Deep learning · Euclidean distance · Boundary box · Centroid · Threshold · SSD

Introduction

The epidemic situation of the covid-19 affects millions of people around the world every day. The spread of this epidemic is causing a great deal of distress in the lives of everyday people. Although high-level research and advancement treatment is undergoing, we are still far behind in diagnosing

the disease and providing the necessary treatment. Every day, the mortality rate among the affected population is increasing incredibly. Despite vaccinations, the spread of the disease has not diminished even slightly. So, we focused on finding some other solutions to escape from this epidemic. Meanwhile, medical experts from various fields have tried to date, no betterment treatment they found out so far. Hence situation among people is much more complicated; also the fear and plight of the people about the disease are increasing. On the one hand, some are infected with this deadly infection, but we must be careful that not to spread it further. The virus is highly contagious, especially among people who have been in close contact with one person to another person. In general, research indicates that the virus is transmitted to infected people when they speak, cough or sneeze. By means of an infected individual, the droplets transfer to the lungs via the respiratory system, where it begins to kill lung cells. On the other hand, in the present situation, those who have no symptoms of the disease are often affected (W. C. D. C. Dashboard 2020). Hence, this situation is forcing the global community to find some other

Communicated by: H. Babaie

✉ Bharathi Gopal
jvibara@gmail.com

Anandharaj Ganesan
younganand@gmail.com

¹ Department of Master of Computer Applications, Shanmuga Industries Arts and Science College, Tiruvannamalai, Tamilnadu, India

² Department of Computer Science and Applications, Adhiparasakthi College of Arts and Science, Tamilnadu Kalavai, India

effective steps to prevent further more spread of this life-threatening infectious virus. Therefore, in the present situation, social distancing is said to be the best anti-depressant. It is the necessary action, and it emphasizes that people who have no symptoms also, to keep at least 6 ft from others. This research work aims to encourage effective measures to overcome the further spread of covid-19 and diminish the further spread of corona virus infection with minimal loss of economic resources, also propose a technique to detect and avoid people crowding in public places.

The term “social distance” was coined to reduce the likelihood of the spread of COVID-19 and to prevent further outspread of infections. The main aim of introducing social distancing is to minimize the infection spreading rate, possibly reduce the vulnerability of the infection as well. According to WHO (Hensley 2020) regulations, the individual who is infected or un-infected must keep a minimum distance of 6 ft when they contact each other, also insists that must staying only at home and not coming out unnecessarily. It is a very vital action in the present situation, to concern seriously infected people. Figure 1 clearly emphasizes the necessity of

social distancing. When it is implemented at the preliminary stage, the situation might not be this dangerous as much, possibly the number of death cases reduced as well. And it is clinically proved and gets a show with help of literature work (Nguyen et al. 2020).

Nguyen et al., 2020 discussed that, by means of social distancing, the number of infected individuals decreases gradually and its effect on health organization burden has been reduced. It paves the way to reduce the number of suspected cases as well. And mortality rate eventually decreases, and it is clearly illustrated in Fig. 2. [Collected from literature work]. It shows the effect of following social distancing properly, in which initially cases reported as very high after some time range number of cases (infected cases) decreases tremendously which denotes the positive impact of following the social distancing.

Various research works are still in progress and researchers are working to improve the necessity and importance of social distancing. Also, their literary work replicates that how to follow effectively, what step needs to be taken, which advancement action needs to be done. That sort of fundamental

Fig. 1 Significance of maintaining social distancing

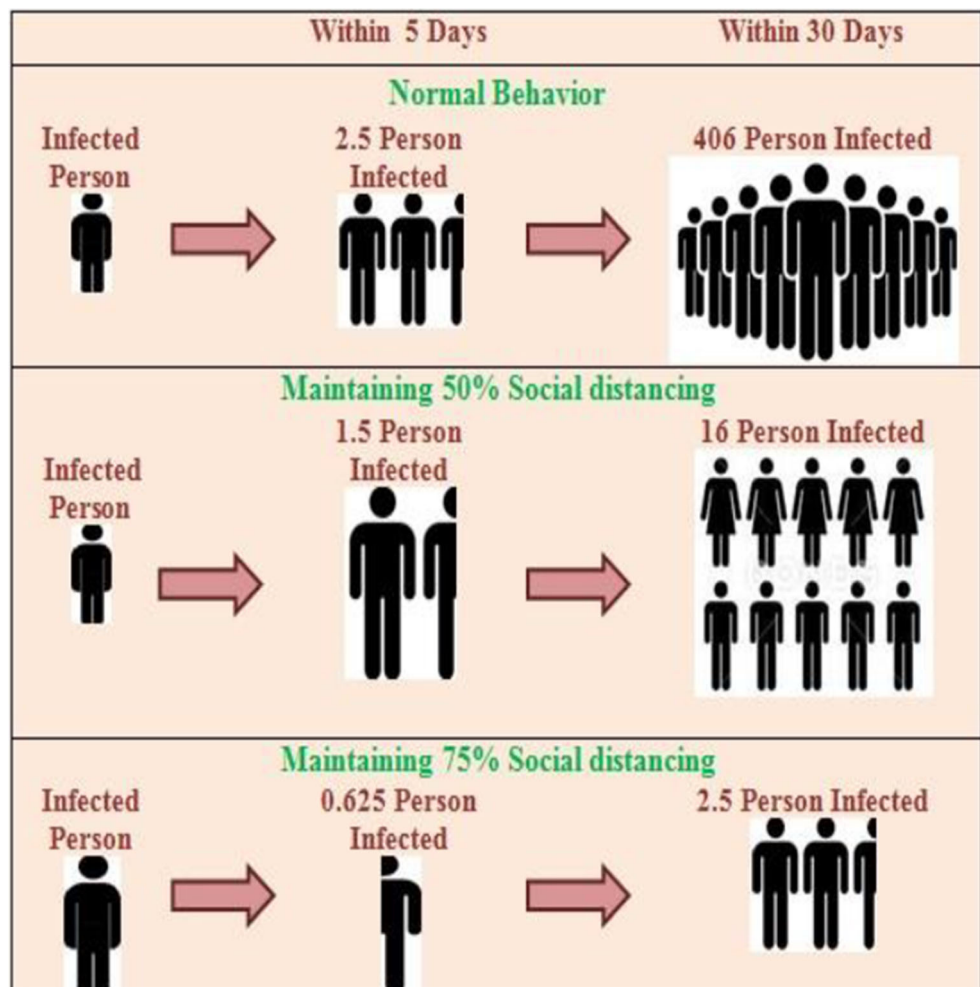
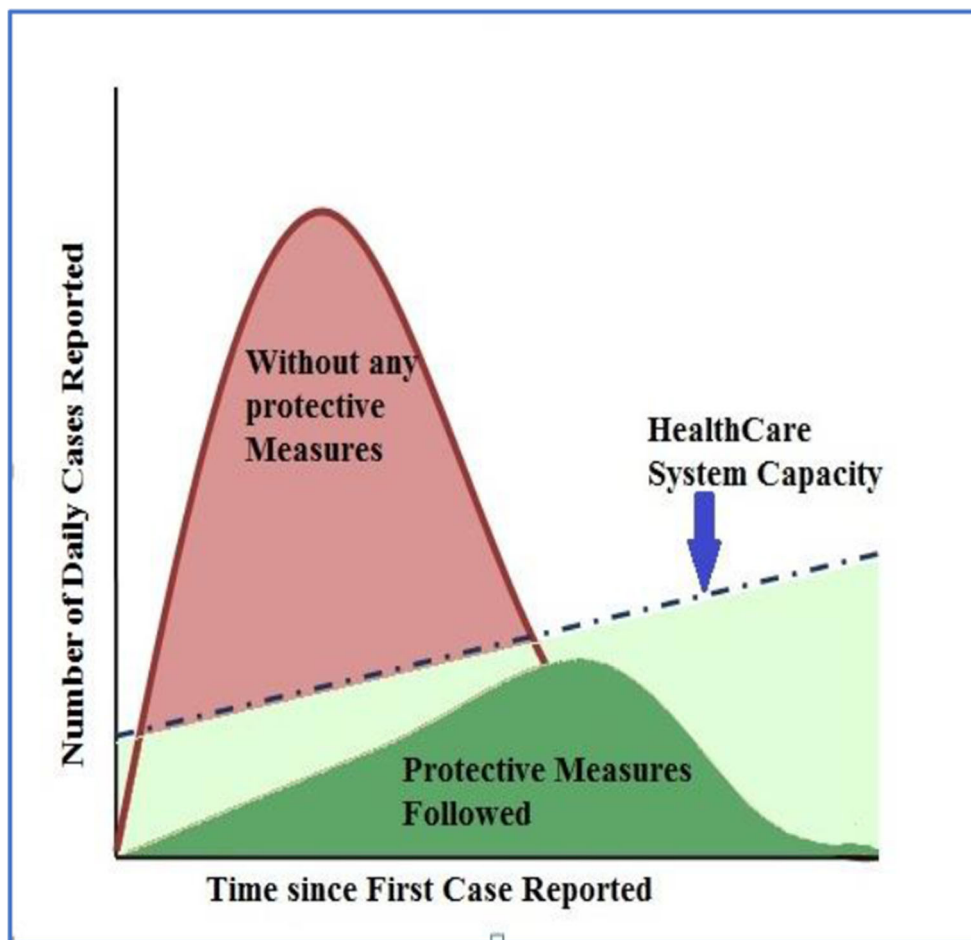


Fig. 2 Effect of following the social distancing



action is explained, also explored to the best of their knowledge (ECDPC 2020; Fong et al. 2020; Ahmed et al. 2018; Kermack and McKendrick 1927). Most of the model utilizes the SIR mechanism also developed stochastic models with help of biological approaches (Eksin et al. 2019; Zhao and Zhao 2016). These respiratory diseases are highly contagious, among which the mode of transmission and the incidence of the virus are the most relevant factors. Precautionary measures and prevention steps must take to avoid the outspread of infection. Take considering this concern, an effect made by Eksin et al. (2019), which utilizes the SIR model also includes the boundaries of social distancing, boundaries fixes as $B(A, C)$. Here A denotes the individual which is affected by infection and C denotes convalesce individual from the infection. With help of A and C boundaries can be determined.

$$\begin{aligned} \frac{dS}{dt} &= -\beta S \frac{A}{T} B(A, T) \\ \frac{dA}{dt} &= -\delta A + \beta A \frac{A}{T} B(A, T) \\ \frac{dC}{dt} &= \delta A \end{aligned} \tag{1}$$

The above equation, where term ‘ β ’ denotes the rate of infection. ‘ δ ’ denotes the rate of convalesce. ‘ T ’ denotes the total population size by using $T = S + A + C$ formula, total

population is computed. The rate of transition is mapped from a vulnerable state (S) to an affected state (A), by using the expression $\frac{\beta SA}{T}$, transition is calculated.

Generally, the social distancing prototype consists of two types, i.e., long-term consciousness and short-term consciousness. The first of which prototype is said to be long-term consciousness, in which person’s contact with others is minimized proportionally to the overall percentage of affected (affected and convalesce) and equation indicates that,

$$B = \left(I - \frac{A + C}{T} \right)^k \tag{2}$$

On the way the second prototype which is known as short-term consciousness, which explains the minimization of contacts is directly proportional to the given instance of affected individual proportion and this is stated as follows.

$$B = \left(I - \frac{A}{T} \right)^k \tag{3}$$

In both cases, behavioral boundary is defined as ‘ k ’ and its range as, $-1 \leq k \leq 0$. If k value is higher which is the affected

individual becomes very sensitive and leads prevalence of disease.

On a similar background, Dr. Andrew Ng published an AI automation invention for identification and monitoring, whether social distancing is followed in workplaces or not and he collaborates this with the help of the most popularized company of AI and possible of highly recognizable people under their leadership (Alto 2020; Ng n.d.). The company releases an article, in that upcoming automation spontaneously detect and find out, if people gathered at the working places and maintaining a safe distance between the working employees, also it insists to maintain at least 6 ft between one another. It is possible with the help of a live stream surveillance camera. Also, that would be integrated into the entire existing surveillance camera available in the working places to maintain social distancing among laborers. Also, that states, it is possible which consist of detection, calibration, and quantification process carrier over to monitoring the safe distance between people. This great effort gets recognized and appreciated by various Vendors in the Core technology of Artificial intelligence (AI 2020).

In the domain of remote sensing environment, object detection is a very difficult task and it's a major research area. More research still in progress, to effectively detect objects in the environment. As considered our work, means the item of interest over the environment (individual person), in remote sensing image our item of interest (individual person), which is be less than 0.0001% over the environment, and finding this out in practice is something that is not possible. Inspired by artificial intelligence, many researchers are actively interested in putting this into practice. Also, over the past decades, Artificial intelligence has been proved to be a promising research area to overcome all existing issues in numerous domains, includes share market,

weather prediction, telecommunication, etc. Moreover, its role is huge, especially in the medical field and remote sensing material detection. In past decades, Machine Learning (ML) models are widely used, in the field of object detection. Also, the ML model has done a tremendous role in detecting an object in the environment by using remote sensing images. Furthermore, the evolution of data over the earth is increased massively; hence big data has been evolved to store and process data effectively. Consequently, ML model performance gets degraded for extracting features from the image, hence Deep Learning (DL) Model is deployed (Aires et al. 2001; Akhand et al. 2016). Deep learning has achieved its extent peak to extract the object from the image. And various researchers, still in progress to get better performance over recognize an object by using remote sensing images. By means of, Multilayer neural network or deep neural network all the things get possible in DL. Hence our research work focuses to propose a practical framework to effectively monitoring the social distance in public places with the help of a Deep

learning model using remote sensing static images. In a remote sensing image, the overhead position is used to detect our item of interest [individual person] from an image and it is determined using the DL model. The major objectives of this research work are as follows:

- Develop a deep learning framework in real-time to use the overhead position of remote sensing images to maintain social distance between people.
- By detecting an item of interest [individual person] from an image, we deploy pre-trained CNN, also a boundary box is applied over the images, and then centroid value is determined with help of the boundary box. Consequently, to intensify the model performance, a transfer learning approach is used. Also, we trained the layer for detecting an overhead position of the image and integrate them into pretrained model.
- Using the Euclidean distance, we estimate the close distance between each pair of centroids of the fixed boundary boxes, by this way we keep chase social distancing among people. Additionally, we fix a threshold value, in order to determine the violation among people by using pixel to space estimation.
- Centroid chasing algorithm is used, to find out those who do not follow the social distancing.
- Finally, evaluation metrics are used to determine the evaluation of model performance; also we evaluate our proposed practical framework with an additional state-of-art model in terms of evaluation metrics.

The paper is organized as follows: second section discussed the various related work, literature studies associated with sentimental analysis. Third section discussed the current problem statement along with the solution. The different methodology used for the proposed approach is discussed in fourth section. Experimentation and result discussion was on the fifth section. Finally, the conclusion of this paper and probable future work is discussed in sixth section.

Related study and literature survey

Various researches suggest that the deadly infectious disease covid-19 [Novel corona virus] began to spread in December 2019 from a place called Wuhan in China. At the end of the week in which the disease began to spread, the number of cases increased from 4000 to 5000 per day and reached its peak and spread to more than half of the population within a month of its inception, also taking millions of lives. Research also shows that people were most severely affected for up to 3 months [end of February 2020], with a slight decrease in the severity of the disease at the half of the third month and no new infections for the last 5 days following the end of that

month. Regarding this that people said they had no idea where the disease had started and that they had no medical equipment to control it, and that the method they used to control the disease by using social distancing as mentioned in the literature work (News 2020; N. H. C. of the People's Republic of China 2020).

Prem et al. began researching the spread of COVID-19 and the implications of using social distancing. By making use of the SEIR model [susceptible-exposed-irresistible-recuperate], the author's attempts were made to simulate the path of infectious also used artificial particular-location communication methods. Adherence to the social distance is said to be the same as the main reason for the spread of the disease if it is lifted before the infection subsides or at the same time as the infection subsides, and if the infection comes under control, the adherence to social distance should be gradually reduced (Prem et al. 2020). Though social distancing reduces the level of disease severity, on the concern of economic growth it's a very painful and backward economy as well. Considering the economic situation in the United States they were not allowed to observe the social distancing in the early days because of which the spread of the disease reached a very cruel peak, which is specifically observed and stated by Adolph et al. (2021). Later the research was towards how to adhere to the social distancing without affecting the economy, hence researchers working on concern that how social distancing is followed effectively without affecting the economic status of the country. Kylie et al. were into that of research, was made a comparison how economic growth is affected while following social distancing in (Ainslie et al. 2020).

Many countries are trying to come up with several effective programs by seeking the help of technologies in the face of this deadly coronavirus infection. Many developed countries, such as India and Korea, are taking precautionary measures with the help of GPS to prevent the spread of the disease to others and to protect infectious people. Considering the welfare of the people, the Government of India has launched the Arogya Setu app. With help of Bluetooth and GPS that app works, which helps to detect the COVID-19 affected people, in that way it safeguards some other people who have a chance of spreading infection. On the other hand, make use of drones and surveillance cameras, to detect people gathering unnecessarily at the public place also monitor whether people followed effectively social distancing or not. But this action increases the manual effort that makes the situation even more challenging. Hence highly advanced social distance monitoring system is needed with the help of trending technologies (Sombhadra et al. 2020; Punn and Agarwal 2020; Punn et al. 2020a; O 2020; Robakowska et al. 2017; Harvey and LaPlace 2019).

Identifying humans using surveillance cameras is a significant area of research that comes from the current situation. Because it's totally depends on the manual way to detect

numerous abnormal actions. Due to the advancement necessitate in various real-time applications, an intelligent system is needed over the area of object detection (Sulman et al. 2008). Therefore, detecting human unusual activity from a captured image is a challenging task and it leads to a situation worse due to ambiguity arising like camera angle, video image resolution, pose, background, etc. Hence high-level prior knowledge is demanded by the performance of a system to detect the real-time object (Wang 2013).

To detect an object in a moving position, two basic strategies are used. Detection and classification of techniques, the detection of an object at the primary stage includes filtering, background noise removal, and optical flow. Considering, background noise removal, variation of the present frame and the surrounding frame is determined at the block or pixel level. Some of the background noise removal approaches are temporal variation, background hierarchical model, adaptive Gaussian mixture, and non-parametric model. As for the optical flow detection method, to identify the motion region of a given image, and flow vector is correlated with the motion of an object. Also, techniques like outliers detection [to remove noise, color, etc...] and computational overheads are composed of optical flow techniques (Joshi and Thakore 2012; Javed and Shah 2002; Brutzer et al. 2011; Aslani and Mahdavi-Nasab 2013; Dollar et al. 2005; Piccardi 2004; Xu et al. 2016; Tsutsui et al. 2001). Aslani et al. stated that to identify parameters of motion using spatio-temporal three-dimensional image [3D] which is based on Spatio-temporal filtering technique. Due to less complexity in computation as well as simplicity in nature, this method is far better. But its performance is limited to uncertainty and noise factor (Agarwal et al. 2016; Niyogi and Adelson 1994).

Deep learning models are effectively addressing the problems involved in object detection. In about last decade, deep learning has produced promising outcomes over the domain of image detection and classification. CNN (Convolution neural network) model has been specifically invented for the purposes of image recognition. Because of layers like convolution, filtering, and max pooling it easily extracts features from an image, also detecting a wide range of features instantly with various dimensions. Also, the visualization purposed is well organized with help of statistical analysis (Zhao et al. 2019; Krizhevsky et al. 2017). CNN initially converts pixels into vectors; based on the vectors it takes features from an image. In terms of considering model training, take a larger time. To address this problem, the YOLO model is used. It has 25 layers and each layer run convolution as default, which means layers like filtering, max polling which run in every convolution layer. So, performance would as much as quick also take very less timing for training the model (Ren et al. 2015; Chen and Gupta

2017). A regression approach is used in YOLO, which effectively learns features from the boundary boxes; also probability score is associated with every class label. Hence this model provides excellent performance in terms of training and speed. So, the complexity of extract features from an image has been much quicker also its well-organized method exhibits its generalized power capabilities showed over an image (Redmon et al. 2016; Putra et al. 2018). The following Table 1 clearly demonstrates various techniques used so far to maintain social distancing.

We concluded from the literature that the researcher has done appreciable work to monitor, whether social distancing is followed in public places or not. Nearly maximum work focused only on the front or side camera view position of an image. Because of that, we build a framework using a Deep learning paradigm and consider the only overhead position of the remote sensing aerial image [Note: Overhead position offers a better field view of an image], Thus it plays a significant role in monitoring social distancing effectively through determining distance among people.

Methodology

Recent works, utilizes only front and side position of an image to detect social distancing among people. But our work is to detect and monitor social distance among people using overhead position of an image using deep learning framework. Our framework utilizes overhead recorded dataset, and furthermore subdivided that for training and testing purposes. Recent researches used numerous pre-trained models, and different frameworks for training dataset to detect an image. Nowadays different models are existing in order to extract and determine object from an image (Krizhevsky et al. 2012; Simonyan and Zisserman 2014; Girshick 2015; Girshick et al. 2014). On basis of considering effective strategy, improved SSD (Single Shot Multibox Detector) is used in our research work. The pretrained model was trained by using Common object in context dataset (COCO). By using transfer learning technique, efficiency of detection model is enhanced. Also new layer is trained and integrated with the pre-trained model in order to extract features from overhead position image.

Table 1 Analysis of various techniques involved in social distancing monitoring system based on literature review work

Existing solutions	Methods/techniques	Accuracy
Dalal and Triggs (2005)	Texture-based schemes using SVM	Acc=94.37 Sen=95.97 Spe=96.63 Pre=97.83 F1-score=97.37
Leibe et al. (2005)	Estimation based trajectory using CNN	Acc=97.03 Sen=96.05 Spe=100
Andriluka et al. (2008)	Tracklet-based detectors using RCNN	Acc=96 Sen=100 Spe=94
Eshel and Moses (2008)	Crowd detection faster RCNN	Acc=74 Sen=79.1 Spe=62.4
Su et al. (2009)	Counting model based on people vision using CNN with SSD	Acc=89.89 Sen=92.04 Spe=88.98 AUC=94.37
Punn et al. (2020b)	Detect humans and the Deepsort using the YOLOv3 model.	Acc=93 AUC=97 Pre=95 F1-score=93
Ramadass et al. (2020)	Autonomous drone-based model using YOLOv3 model	Acc=96.35 Spe=95.30 Sen=100 Pre=95.34
Pouw et al. (2020)	A framework for physical distancing and managing crowd using faster-RCNN and SSD.	Acc=70.82 Sen=69.81 Spe=92.92 Pre=68.78 F1-score=67.12

Step involved finding out social distance among people

Once an objects are detected by means of improved SSD, that information is treated as boundary box information. Also, a centroid is determined over every detected boundary box. Then find out meanwhile distance among each determined centroid over an image, a Euclidean distance is used. A threshold predefined value is equal to 6 ft is employed to verify that the distance between any two-boundary centroid is less than the number of pixels configured. If two persons are very close in contact means violation exists, that is, the distance between each of them is much shorter than the distance we specified [i.e., threshold]. Then that particular information is kept and stored as violation set. Because there is a violation, that particular boundary box color will be changed to red. To determine, those people who violate social distancing, a Centroid chasing algorithm is used. And the final output will be our model produce the total number of people who violates social distance along with its detected boundary boxes and centroid.

Boundary boxes

The conclusion derived from literature survey point out that, if we want to detect every object individually from the set the object over the image, the concept of boundary boxes is used. These boxes are covered in different spatial locations (for each filter) with different sizes or dimensions and proportions in the input image. In our work, we used width [W] × height [H] to generate boundary boxes over the dimension of image.

Consider any location of dimension of image can be $Wp\sqrt{r}$ X $Hp\sqrt{r}$. The ratio of aspect is $r > 0$ also the parameter is ‘p’ which ranges from (0,1]. For each model from literature, also our model enhanced SSD, are used the configured value of p and r which is given in Table 2. Then trained process is done over detection model in order to predict and determine which class, the generated boundary box is belonging to. Also, any adjustment is required on the boundary box dimension [adjust height and width], we have established a balancer to take care of it, also that enables to fit the ground truth object in a better way and paves the way to reduce the loss regression and classification. But in a real work environment, usually objects are closely in nature and kept in close contact with each other, so possibility of getting overlapping in the boundary boxes. So, we used non-max suppression (NMS), to determine the IOU [Intersection over union] boundary. That helps to retrieve our object of interest [individual persons] effectively. Then determine overlapping region ratio between ground truth value and the predefined boundary box. Finally, the determined

score value is compared with hyper-parameter of fixed threshold and the image of best boundary box is returned (Table 3).

Loss functions of boundary boxes

The step involved for training our model includes boundary box which is assigned as ‘b’ and the corresponding label such as ‘1’ means positive and ‘0’ means negative. Also, our object of interest [individual person] is associated with entire ground truth value box ‘g’. Furthermore, the assigned class label as positive means it works with $P_0 \in \{a_1, a_2, \dots, a_n\}$. Here a_n represent object belong to category n. On the other hand, the encoded vector also generated with respect to ‘b’ which is $f(g_b|b)$. Consequently, the negative boundary box is represented as $P_0 = 0$. An image R, consisting of some boundary box ‘b’ corresponding to trained model parameter ω and the class object belong to predicted class as $Z_{class}(R|b; \omega)$ along with equalizer box as $Z_{reg}(R|b; \omega)$. The following equation is used to determining loss function in order to detect boundary boxes,

$$L(b|R; \omega) = \alpha \cdot I_b^{obj} L_{reg} [f(g_b|b) - Z_{reg}(R|b; \omega)] + \beta \cdot Z_{class} [P_b, Z_{class}(R|b; \omega)] \tag{4}$$

From the above equation Z_{class} denotes prediction of single boundary box, L_{reg} denotes the regression loss with respect to boundary box. Also, $I_b^{obj} = 1$, and ‘b’ is a positive boundary. The weight combined with regression and classification as α, β respectively. Finally the overall boundary box losses is computed using $L(b|R; \omega)$.

Centroid over an image

Once boundary box is detected over an image, Next centroid is automatically calculated over an image. That paves the way to get the center position of all the boundary boxes. After getting center position of all the boundary boxes, then we apply Euclidean distance, over every box along with centroid point. Finally, if the values are less than our fixed threshold value, there exists a violation.

In general, for instance it works between each and every detected boundary boxes. Initially select any one boundary box out of an image, then applying the above equation is to find out centre position of an image.

To summarize a shape of given image R (a, b)

$$M_{ij} = \sum_a \sum_b a^i b^j R(a, b) \tag{5}$$

$$\mu_{xy} = \sum_a \sum_b (a - \bar{a})^x (b - \bar{b})^y R(a, b)$$

Where, $\bar{a} = \frac{M_{01}}{M_{00}}$ $\bar{b} = \frac{M_{10}}{M_{00}}$

Table 2 Feature extraction performance on ImageNet challenge collected from literature work

Used models	Accuracy	Parameters (p)	Ratio (a*100/p)
VGG-16 (Simonyan and Zisserman 2015)	0.70	16 M	4.72
ResNet-101(He et al. 2016)	0.75	43.6 M	1.77
Inception v2 (Szegedy et al. 2016)	0.73	11 M	7.39
Inception v3 (Szegedy et al. 2017)	0.77	23 M	3.57
Resnet v2 (Szegedy et al. 2017)	.079	55 M	1.47

Euclidean distance over a detected boundary box

Once a centroid is determined from a set of fixed boundary boxes. Then our search space reduces to single boundary box [individual person or single point]. Take that point as ‘G’. Then, from point ‘G’ our approach calculates Euclidean distance over determined boundary box along with point ‘G’. It is calculated by using the Eq. 6.

$$D(x, y) = \sqrt{2} \sum_{m=1}^n (y_m - x_m)^2 \tag{6}$$

where x, y point out two determined boundary boxes. y_m and x_m represent Euclidean vector, from starting from the origin of space; m represents total space (Fig. 3).

Convolution Neural Network (CNN)

CNN stands for convolution neural network, which is a notable type of deep learning neural network. It is used for processing the known data, a grid-like topology. CNN plays a significant role in processing the detection and classification of the image. It is one of the well-known practical applications and made a phenomenal success. Convolution, pooling, and fully-connected are the layers which are consist of the building block of CNN. The convolution layer performs convolution operation, which means instead of using normal matrix multiplication, its performance convolution operation within its layer and it’s just like a neural network. It comprises an input layer [where we import our features], an output layer, and numerous hidden layers [depends on model performance]. Each hidden layer holds a convolution layer, and then a pooling layer, and finally a fully connected layer.

1. Convolution layer: This is the primary layer of the convolution block, which consists of kernels and filters. It is independent of the feature of an image. All the layers are initialized randomly. Its operation of convolution is defined as

$$O(t) = (x*w)(t) \tag{7}$$

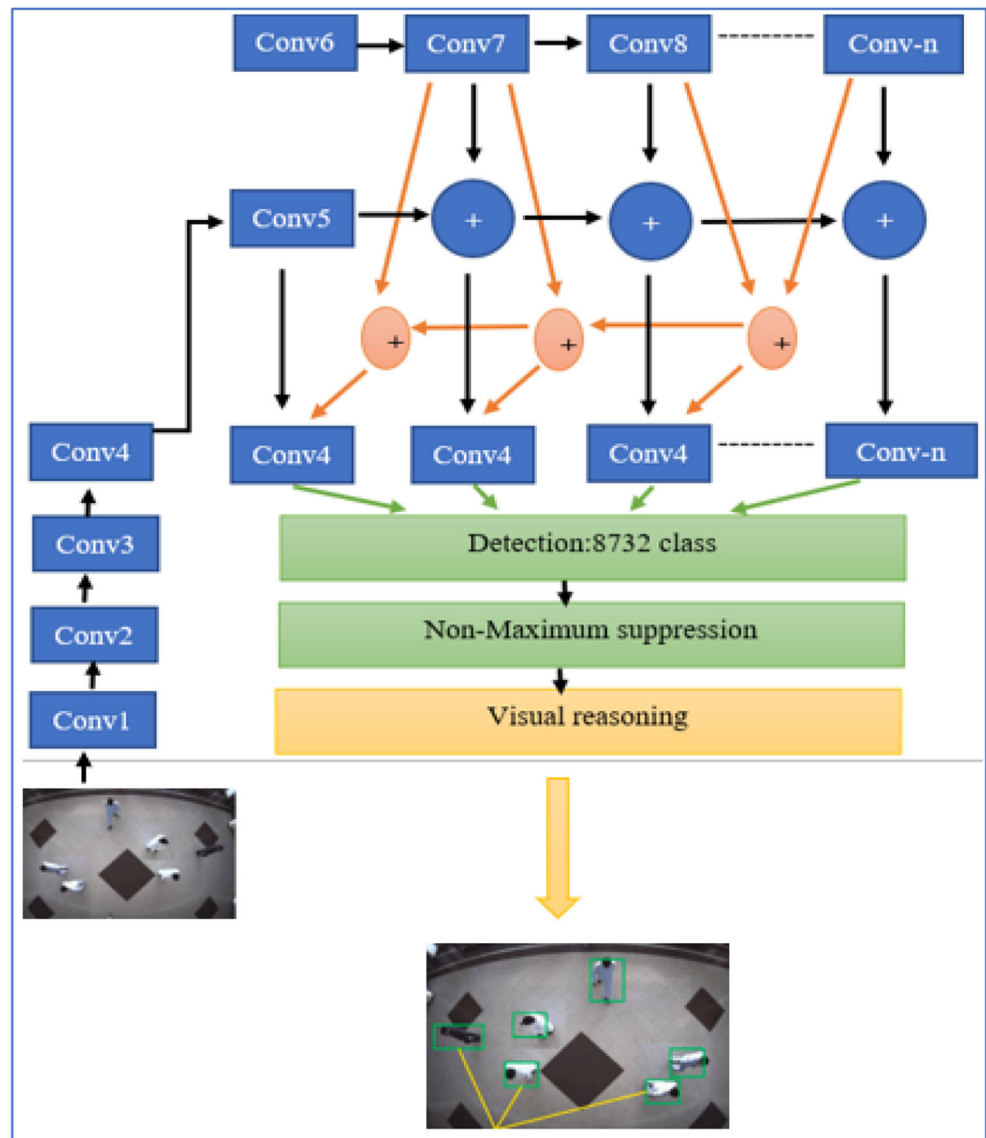
Where ‘x’ indicates the input vector and W indicates the kernel. This equation works with help of two inputs, which is matrix and filter or kernel of an image. By use of those inputs, it maps the output. If every image consists of matrix dimension such as [h*w*d], also filter size as [fh* fw * d]. And the corresponding output dimension volume is as {[h - fh + 1] * [w - fw + 1] * 1}. By detecting edge, sharpening and blurring kind of functions are possible over an image with help of different filters. And high-level feature extraction such as extract edge from an image with help of convolution operation. This layer depends on user requirements. Also, special operations like strides and padding are associated with every convolution operation. Once the performance of the convolution layer is done, the output is generated in the form of a feature map and that can be fed input to the pooling layer.

2. Pooling layer: CNN exactly consists of three stages. The first stage, corresponding to the linear activation and the second stage consists of a non-linear activation function. Accordingly, the third stage consists of pooling features. Pooling consists of two main kinds i.e., max pooling and average pooling. Input feature vector consisting of a window over within it along with stride. Consequently, the size is mentioned, which is known as the hyper-parameter

Table 3 Hyper-parameters to generate the boundary boxes

Models used for detection	Vector size (p)	Aspect ratio	Boundary boxes	Intersection of union
Faster RCNN (Punn et al. 2020c)	[0.26, 0.6, 1.0]	[0.6, 1.1, 2.0]	9	0.7
YOLO v3 (Punn et al. 2020c)	[0.26, 0.6, 1.0]	[0.4, 0.6, 1.0]	9	0.6
SSD (Punn et al. 2020c)	[0.2, 0.56, 0.94]	[0.4, 0.6, 1.0]	9	0.5
Improved SSD	[0.2, 0.57, 0.94]	[0.4, 0.6, 1.0]	9	0.5

Fig. 3 Improved single shot detector



of a given pooling layer. The maximum value of features is considered in the case of the max-pooling layer accordingly average value is considered in the case of the average pooling layer. It paves the way to increase the computational efficiency as well as decrease the feature representation.

3. Fully connected layer: The pooling layer output is fed input to the fully connected layer that layer helps to connect every neuron in that layer. The result of the fully connected layer indicates that the total classes in the classification. At last, by utilizing the soft-max activation function, the input vector gets normalized and that corresponding to each probability value of the input vector. This activation function performs exponential action that

smooths the data and ranges output between [0–1]. Similar to the traditional neural network, training error is determined with respect to weight W , also training error is minimized and back-propagation is done with help of a gradient descent optimizer.

$$L = -\frac{1}{|Z|} \sum_i \ln(p(y^i|Z^i)) \tag{8}$$

Where

- $|Z|$ Number of training example.
- y^i Training example corresponding to i^{th} label.
- $p(y^i|Z^i)$ Correctly classified class probability.

Proposed approach

In this section, we will point out about proposed detection model along with clear architecture. We will use an improved SSD for detecting every object over an image. Also proposed flow framework of social distancing monitoring is demonstrated in Fig. 4.

Improved single shot detector with feature fusion

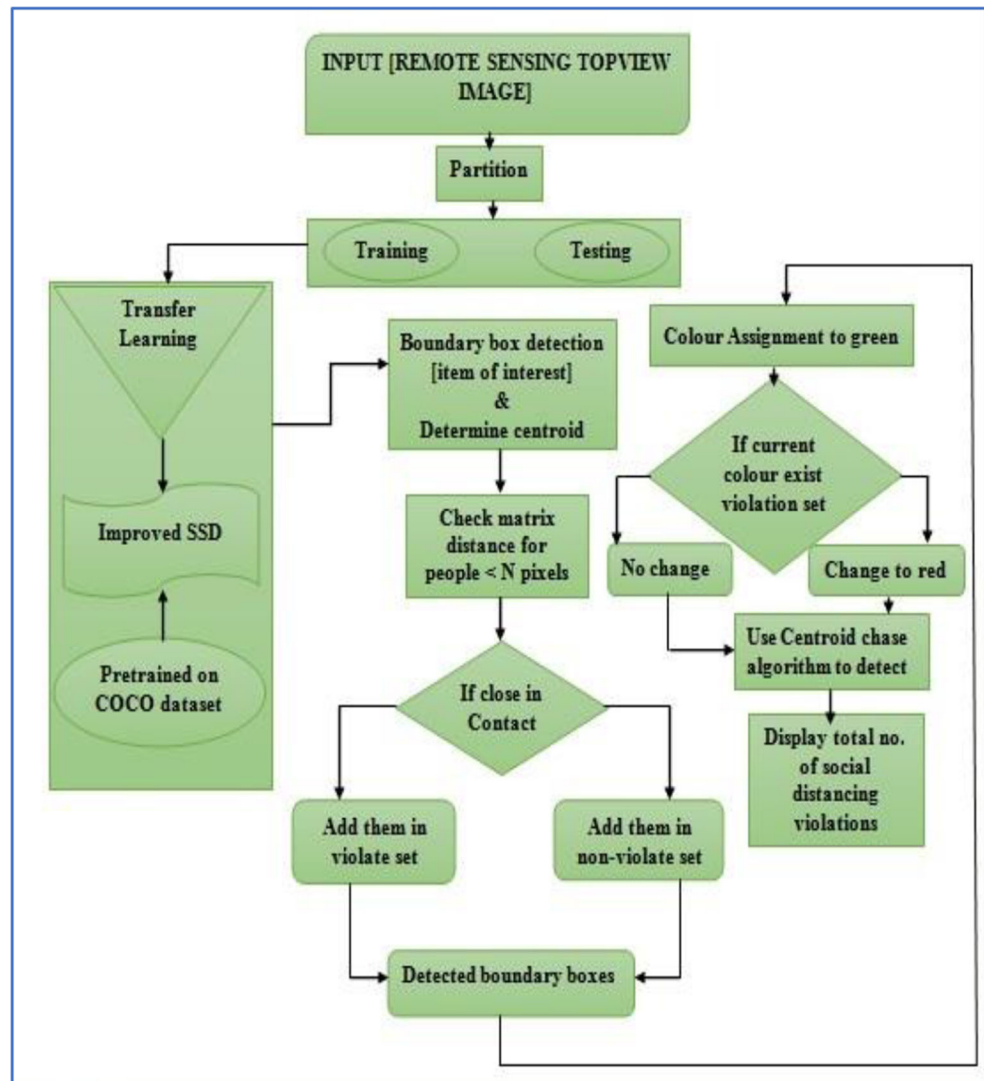
Considering traditional single shot detector [SSD], it achieves better performance by making use of various layers involved in features maps to detect an object in feature pyramid. By utilizing, each layer in the feature pyramid detects independently. Therefore, the traditional SSD model has not the ability to find out the relationship between different layers. Each layer in the pyramid consists of unique information regarding the feature. Consequently, Low-level and High-level layer consists of rich information and powerful semantic features.

Hence a high level of importance needs to combine the features of various layers so that the feature pyramid obtaining wealthy feature details regarding semantics.

We demonstrate the structure of an improved single-shot detector in Fig. 3. Our suggested improved SSD has differed from the traditional SSD. Because in improved SSD, used feature information in two-way transfer learning in order to accomplish various output layer feature fusion. In feature fusion pyramid consists of one layer in middle -level. And the reason for that, rich details and well semantic features are present in the middle level. The upper and lower level in the middle layer is known as higher and lower-level layer respectively. It is determined that it is not a traditional SSD also by making use of higher and lower-level layers to get feature fusion results.

To accomplish feature fusion; initially make to set the same size as all feature maps. And utilizing the up and down-sampling technique, the shape of each layer has been resized. The higher and lower-level layer has resized the shape in the

Fig. 4 Proposed flow to detect social distance monitoring framework using overhead position



form of $2H \times 2 W$, which is equal to the middle layer i.e., target layer, and showed in Fig. 3. Furthermore, to combine the dimension of the channel, a 1×1 convolution layer is used [which is equal to 512]. To combine feature value distribution before feature map fusion, a normalization technique is required. Because every layer features value distribution is different. Hence batch normalization is used to accomplish uniform distribution before fusion. At the last feature, a map is obtained with the help of various layers. The obtained feature map consists of well semantic features and finest details because of invoked feature fusion technique.

Object detection in improved single-shot detector

By using object features, the object’s category is done in the existing detector. If the object feature is too small and not apparent means detecting its own feature from that category is very difficult which leads to an increase the complexity. When detecting an object from a remote sensing aerial image is uncertain, it is necessary to examine high-level attention between objects.

When we determine the type of object that is not clearly visible, it is natural to look at the scene and the objects around it to help make a decision. For instance, if you find objects on the road, you can easily determine the road, vehicles, and trees over the image. However, as shown in Fig. 4, due to the lack of existing features, people in some other cars can be missing or misidentified. In this scenario, the detected objects can be used with more confidence, helping us to determine the obscure objects.

For instance, we have some object $A = \{A_1, A_2, \dots, A_n\}$ over a given image R , and n is the maximum object that is present. Our objective is to detect every object over an image. Hence the objective function for model training can be defined in eq. 7:

$$\arg_X \max L = \log Q(AA_{1:n}|X, R) \tag{9}$$

From the equation X is the model, in which maximize the log function L and $A_{1:n}$ stands for $\{A_1, A_2, \dots, A_n\}$. By means of using relation among objects, we create an equivalent transformation of eq.[7] which is given in eq. [8].

$$\begin{aligned} \arg_X \max L &= \log Q(A_{1:n}|X, R) \\ &= \sum_{m=1:n} \log Q(A_m|A_{1:n-1}, X, R) \end{aligned} \tag{10}$$

We integrate our visual reasoning to the objective function and that approximation will be,

$$\arg_X \max L \approx \sum_{m=1:n} \log Q(A_m|S_{n-1}, X, R) \tag{11}$$

S denotes visual model reasoning that able to note all the relation among objects. The detection model is represented as X that can be trained with help of the network.

Two sections comprise up the overall visual reasoning model. The first section enables the connection among objects. In case we have categories of objects as ‘ n ’ which means we employ relationship matrix $F = (f_1, f_2, \dots, f_n) \in T^{n \times n}$ which is symmetric. The relationship among the object i and j can be denoted by f_{ij} obtained by make of the following equation,

$$\begin{aligned} f_{ij} &= \sum_{n=1}^N R_n(i, j) \\ R_n(i, j) &= \begin{cases} 1 & \text{if } d_{ij} \end{cases} \end{aligned} \tag{12}$$

N indicates the total training items, the center distance among the objects is indicated as d_{ij} and our region of interest [individual person] is denoted as T . Once a relationship matrix is built, it is used to provide to model to find out the object over an image R .

The second stage manifests how we use the relationship to help the detection model. By making use of an enhanced SSD algorithm, we detect an entire object over a single image with help of confidence value. The confidence value of the detected item is greater than 0.6 is treated as trustable detection results. Furthermore, make use of those trustable detection results are less belongs to the confidence value ranges from 0.4 to 0.6. Hence less trustable detection outcome’s confidence value gets updated by using the below formula,

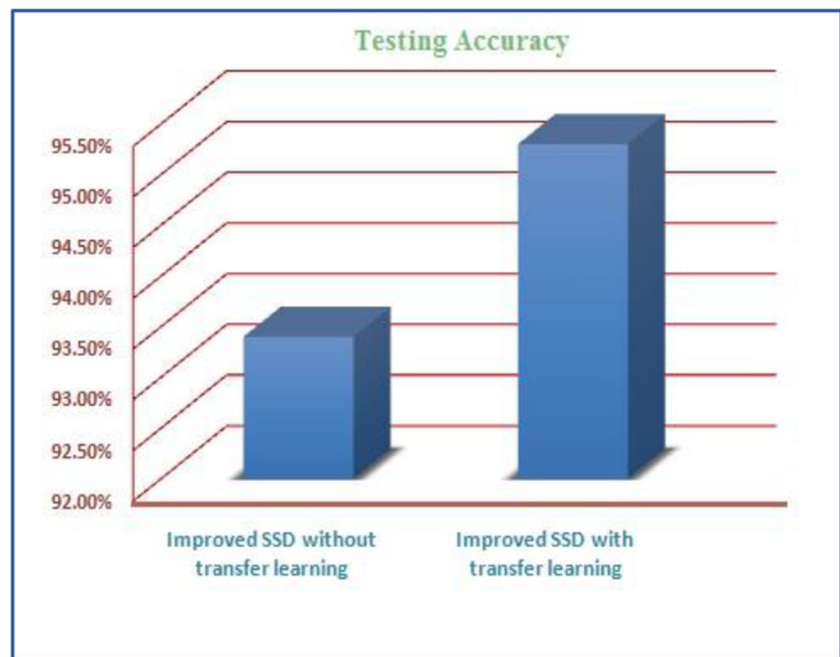
$$\begin{aligned} s_i(x) &= s_i(x) + \lambda \frac{e^{p_i(x)} - e^{-p_i(x)}}{e^{p_i(x)} + e^{-p_i(x)}} \\ p_i &= \sum_{j \in T} e_{ij} - Q(e_{ij} = 0) \end{aligned} \tag{13}$$

Where $s_i(x)$ is the total value of confidence of the detected objected x which is contained in class i , T is the trustable objects that are around the object ‘ x ’ in an image. The parameter of trade-off among the model detection is indicated as. The total number of times $=0$ that can be returned by the function Q . The object’s final value of confidence is no more dependent on a feature of its own, but also on the objects by its surrounding. Indicates that the object consists of class I , if the value is greater than 0 it means the value of confidence over objects gets increased otherwise will decrease. In this way, an object has to be detected with less relevant features.

Table 4 Testing accuracy between improved SSD with and without transfer learning

Model	Testing accuracy
Improved SSD without transfer learning	93.4%
Improved SSD with transfer learning	95.3%

Fig. 5 Comparison of testing accuracy between improved SSD with and without transfer learning



Experimentation

Description about dataset and performance metrics

This section explores a clear description regarding different approaches we are used for our experimentation purposes, and we compared our approach with different state-of-art approaches also resultant is clearly demonstrated.

Utilizing single-stage network architecture, the boundary box and the different class probabilities are estimated. Also, the model utilizes COCO [Common objects in context] dataset for training purposes and training is done with help of these pre-trained datasets.

Furthermore, a transfer learning technique is implemented in order to detect the overhead position of an image that enhances the model detection efficiency. The new layer is trained and integrated with existing architecture using improved SSD.

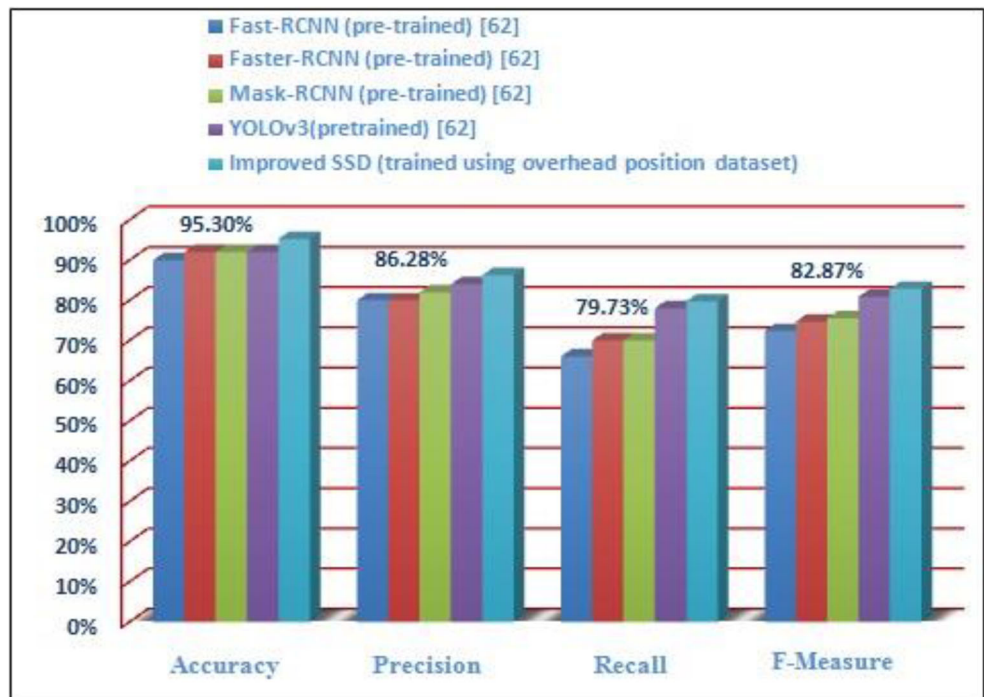
For training, the detecting model of overhead position, we used the dataset taken from the Institute of Management Sciences, Hyderabad, and Peshawar Pakistan, which is an indoor recorded dataset (Ahmad et al. 2019). Further split the dataset 80% and 20% accordingly for training and testing purposes.

Our experimentation result is estranged into two phases. In the First phase, we discussed the result based on testing the pre-trained model either with or without the transfer learning technique. Then in the Second phase discussion is based on applied transfer learning for the overhead position detection model is evaluated with other state-of-art approaches. The main purpose of comparing the results of both phases, the model is tested using the same remote sensing top view images. We used the metrics like accuracy, precision, and recall for performance evaluation.

Table 5 Testing accuracy, precision, recall between Improved SSD with transfer learning and other state-of-art approaches [collected from literature work]

Model	Accuracy	Precision	Recall	F-measure
Fast-RCNN(pre-trained) (Imran et al. 2021)	90%	80%	66%	72.33%
Faster-RCNN (pre-trained)(Imran et al. 2021)	92%	80%	70%	74.67%
Mask-RCNN (pre-trained)(Imran et al. 2021)	92%	82%	70%	75.53%
YOLOv3 (pre-trained)(Imran et al. 2021)	92%	84%	78%	80.89%
Improved SSD (trained using overhead position dataset)	95.3%	86.28%	79.73%	82.87%

Fig. 6 Comparison of performance evaluation between improved SSD with transfer learning and other state-of-art approaches



Performance metrics used for evaluation purposes

Accuracy: Accuracy can be evaluated as follows

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{15}$$

Where,

- TP An instance which is positive also predicted as positive.
- TN An instance which is negative also predicted as negative.
- FP An instance which is negative but predicted as positive.
- FN An instance which is positive but predicted as negative.

Precision: Precision can be evaluated as follows.

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

Recall: Recall can be evaluated as follows.

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

F-Measure: F-Measure can be evaluated as follows.

$$F - Measure = 2 \frac{Precision \cdot recall}{precision + recall} \tag{18}$$

Fig. 7 Front and side view tested sample input

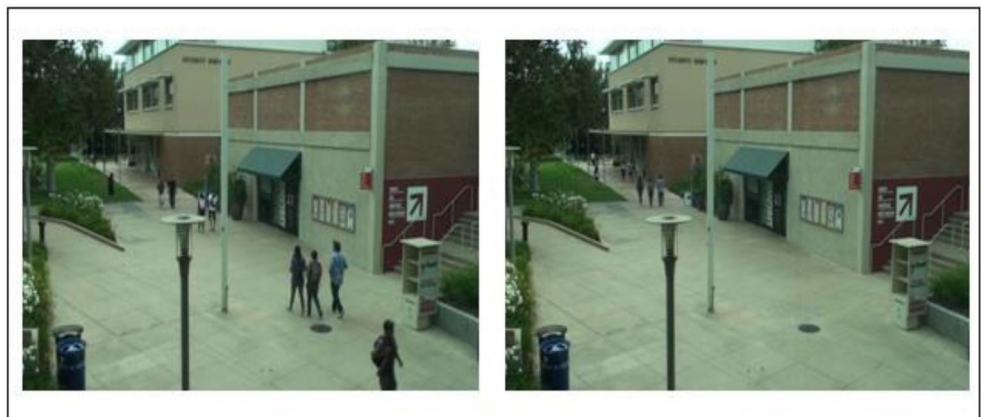
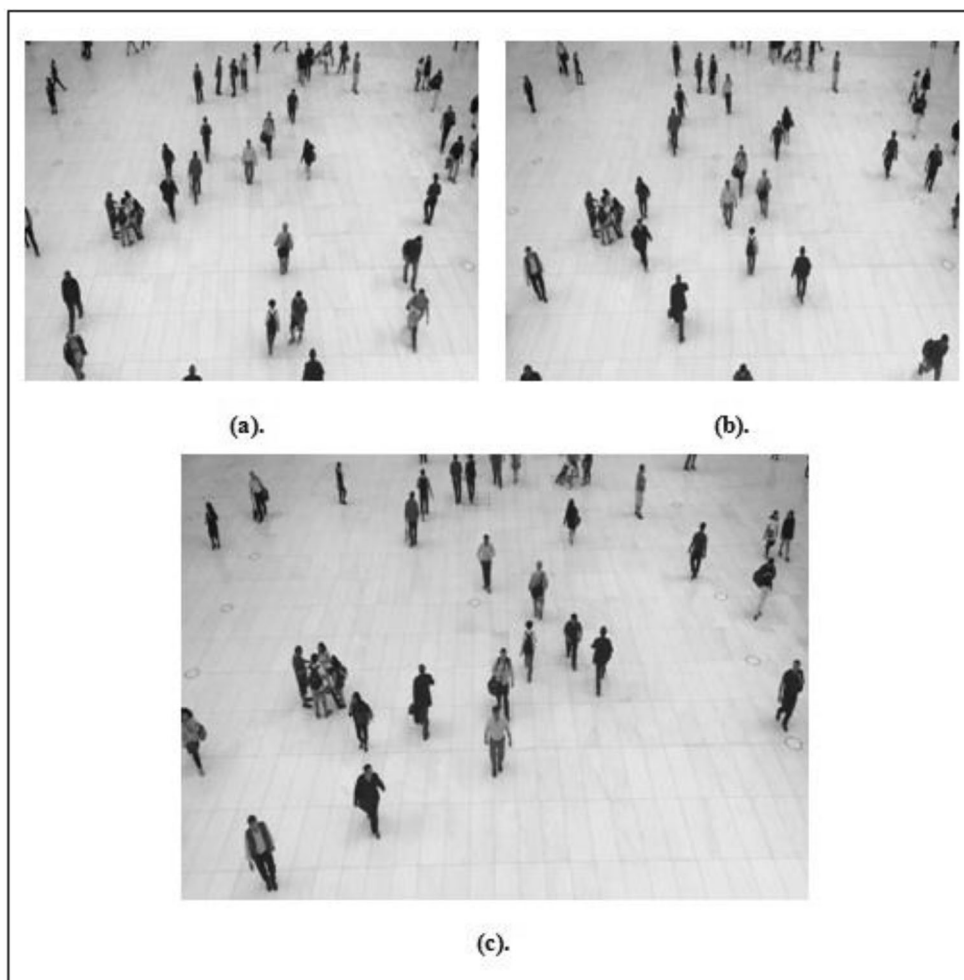


Fig. 8 a, b & c Overhead position tested sample input images



Result analysis

PHASE I: result comparison between improved SSD with and without transfer learning technique

The testing result based on the framework of social distancing using a pre-trained model either with or without transfer learning has been reported and visualized as a graph for

understanding purposes, which is shown in Fig. 5 and corresponding tested accuracy as listed in Table 4. As the result showed that our suggested improved SSD approach with transfer learning outperforms well, to detect and monitor social distancing effectively over remote sensing top view images and our suggested approach produces better results compared to improved SSD without transfer learning. For testing purposes, we used same static images for both the model. Our

Fig. 9 Detecting boundary boxes



Fig. 10 Calculating centroid over an Image



research work is only based on the top-view position of remote sensing image, but in testing, we include few frontals and side view position images, in order to trace out how our approach effectively works in that context also and how to detect our item of interest [individual person], by the way, all the detected result as displayed as follow, which clearly demonstrated the output result of our model.

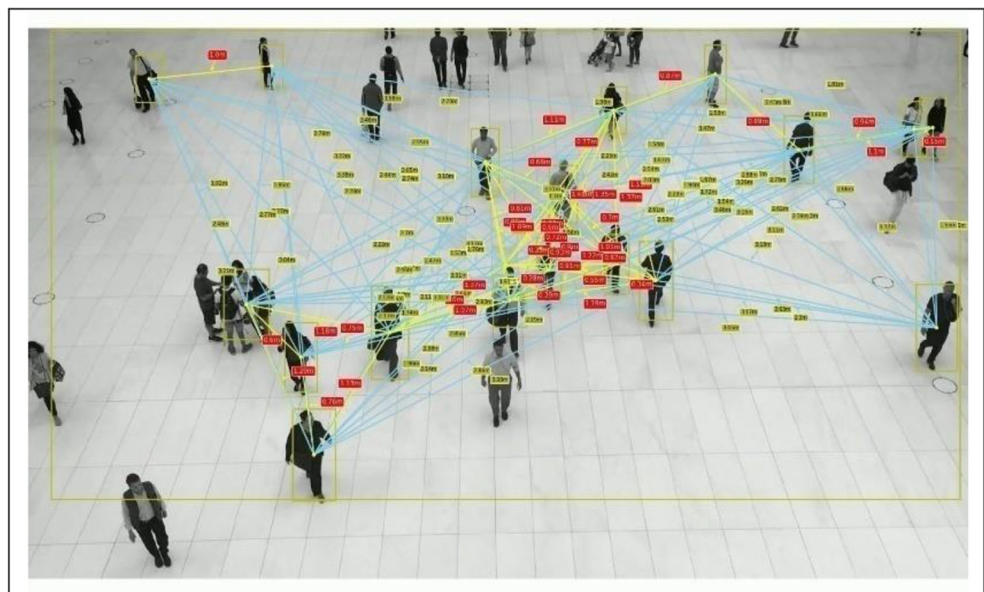
PHASE II: result comparison between improved SDD with transfer learning and other State-of-art approaches

Transfer learning technique improves the accuracy of our detection model. By means of detecting overhead position, in addition overhead dataset is trained to detect overhead position of an image. We fix epoch and batch size for training the dataset as 50 and 64 respectively. Furthermore, we appended a middle layer over an architecture that paves the way to minimize the training losses and testing error. In this phase, comparison is done with help of state-of-art approaches like Fast-

RCNN (pre-trained), Faster-RCNN (pre-trained), Mask-RCNN (pre-trained), YOLOv3 (pre-trained), Improved SSD (trained using overhead position dataset). As best of our knowledge, state-of-art approaches is the best approaches so far to detect objects over an image and from literature work we used those approaches (Imran et al. 2021). We listed various performance evaluation measures used in our experimentation in Table 5 and also showed the corresponding graph in Fig. 6.

From Fig. 6, illustrates clearly that our suggested improved SSD with transfer learning outperforms well and achieves an accuracy of 96.7% while detecting our item of interest [individual person] from a remote sensing top view image. Hence our proposed social distancing framework using improved SSD is the best among other state-of-art approaches collected from literature work. The front and side view, overhead position tested sample input images, then detecting bounding boxes and calculating centroid over an images finally applying Euclidean distance over centroid points are shown in figures (Figs. 7, 8, 9, 10 and 11).

Fig. 11 Applying Euclidean distance, over every boundary boxes along with centroid point



Conclusion

In this research work, we present a real-time deep learning framework to effectively monitor the social distancing among people using the Improved SSD model. For experimentation, purposes a pre-trained COCO dataset along with an overhead position dataset is used. To improve the performance of the pre-trained model transfer learning technique is adopted. By means of integrating a new layer over the existing architecture, an overhead dataset is trained. Our detection model uses a centroid chasing algorithm to identifies, whether people violate social distancing or not. Centroid chasing algorithm works on the basis of the fixed threshold value. The experimentation result clearly demonstrates that our deep learning framework works efficiently identifies people who have close contact and violate social distancing. Our detection model performs with better accuracy of 96.7%. Also, overall model detection efficiency and accuracy have been improved with help of the transfer learning technique. Our detection accuracy has been compared with various existing state-of-art approaches also comparison graph is generated. In the future, We aim to extend our research work to use real-time surveillance camera captured video to improve the detection of boundary box rate and real-time performance measures and also plan to integrate a new layer to train the model even more effectively.

Code availability Available on Request.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Adolph C, Amano K, Bang-Jensen B, Fullman N, Wilkerson J (2021) Pandemic politics: timing state-level social distancing responses to covid-19. *J Health Polit Policy Law* 46:211–233. <https://doi.org/10.1215/03616878-8802162>
- Agarwal A, Gupta S, Singh DK (2016) Review of optical flow technique for moving object detection. 2nd International Conference on Contemporary Computing and Informatics IEEE, pp 409–413. <https://doi.org/10.1109/IC3I.2016.7917999>
- Ahmad M, Ahmed I, Ullah K, Khan I, Khattak A, Adnan A (2019) Energy efficient camera solution for video surveillance. *Int J Adv Comput Sci Appl* 10(3). <https://doi.org/10.14569/IJACSA.2019.0100367>
- Ahmed F, Zviadrine N, Uzicanin A (2018) Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review. *BMC Public Health* 18(1):518. <https://doi.org/10.1186/s12889-018-5446-1>
- AI L (2020) Landing AI Named an April 2020 Cool Vendor in the Gartner Cool Vendors in AI Core Technologies. <https://www.pnwswire.com/news-releases/>. Online; Accessed 22 April 2020
- Ainslie KE, Walters CE, Fu H, Bhatia S, Wang H, Xi X, Baguelin M, Bhatt S, Boonyasiri A, Boyd O et al (2020) Evidence of initial success for china exiting covid-19 social distancing policy after achieving containment. *Wellcome Open Res* 5(81):81. <https://doi.org/10.12688/wellcomeopenres.15843.2>
- Aires F, Prigent C, Rothstein M (2001) A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations. *J Geophys Res* 106:14887–14907. <https://doi.org/10.1029/2001JD900085>
- Akhand K, Nizamuddin M, Roytman L, Kogan F (2016) Using remote sensing satellite data and artificial neural network for prediction of potato yield in Bangladesh. In *Remote Sensing and Modeling of Ecosystems for Sustainability XIII*. Vol. 9975, p. 997508. International Society for Optics and Photonics. <https://doi.org/10.1117/12.2237214>
- Alto P (2020) Landing AI Named an April Cool Vendor in the Gartner Cool Vendors in AI Core Technologies. <https://www.yahoo.com/lifestyle/landing-ai-named-april-2020-152100532.html>. Online; Accessed 21 April 2020
- Andriluka M, Roth S, Schiele B (2008) People-tracking-by-detection and people-detection-by-tracking. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 1–8
- Aslani S, Mahdavi-Nasab H (2013) Optical flow based moving object detection and tracking for traffic surveillance. *Int J Electr Comput Eng Electron Commun Eng* 7:1252–1256
- Brutzer S, Hoferlin B, Heidemann G (2011) Evaluation of background subtraction techniques for video surveillance. *CVPR* 2011. IEEE pp1937–1944. <https://doi.org/10.1109/CVPR.2011.5995508>
- Chen X, Gupta A (2017) An implementation of faster rcnn with study for region sampling. <https://arxiv.org/abs/1702.02138v2>
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society Conference on Computer vision and Pattern Recognition (CVPR'05), vol. 1. IEEE, pp 886–893
- Dollar P, Rabaud V, Cottrell G, Belongie S (2005) Behavior recognition via sparse spatio-temporal features. IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. IEEE, pp 65–72. <https://doi.org/10.1109/VSPETS.2005.1570899>
- ECDPC (2020) Considerations relating to social distancing measures in response to COVID-19—second update. <https://www.ecdc.europa.eu/en/publications-data/considerations>. Online; Accessed 23 March 2020
- Eksin C, Paarporn K, Weitz JS (2019) Systematic biases in disease forecasting—the role of behavior change. *Epidemics* 27:96–105. <https://doi.org/10.1016/j.epidem.2019.02.004>
- Eshel R, Moses Y (2008) Homography based multiple camera detection and tracking of people in a dense crowd. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 1–8
- Fong MW, Gao H, Wong JY, Xiao J, Shiu EY, Ryu S, Cowling BJ (2020) Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—social distancing measures. *Emerg Infect Dis* 26:976–984. <https://doi.org/10.3201/eid2605.190995>
- Girshick R (2015) Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, pp 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>
- Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp 580–587. <https://doi.org/10.1109/CVPR.2014.81>
- Harvey J, LaPlace A (2019) Megapixels.cc: Origins, ethics, and privacy implications of publicly available face recognition image datasets. Online Available: <https://megapixels.cc/>

- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 770–778
- Hensley L (2020) Social distancing is out, physical distancing is in—here’s how to do it. *Global News—Canada* (27 March 2020)
- Imran A, Ahmad M, JJPC R, Jeon G, Din S (2021) A deep learning-based social distance monitoring framework for COVID-19. *Sustain Cities Soc* 65:102571. <https://doi.org/10.1016/j.scs.2020.102571>
- Javed O, Shah M (2002) "Tracking and object classification for automated surveillance." In *European Conference on Computer Vision*, pp. 343–357. Springer, Berlin, Heidelberg
- Joshi KA, Thakore DG (2012) A survey on moving object detection and tracking in video surveillance system. *Int J Soft Comput Eng* 2(3): 44–48
- Kermack WO, McKendrick AG (1927) Contributions to the mathematical theory of epidemics. *Proc R Soc A* 115:700–721. <https://doi.org/10.1098/rspa.1927.0118>
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *NIPS’12: Proceedings of the 25th International Conference on Neural Information Processing Systems*. vol.1, pp 1097–1105
- Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. *Commun ACM* 60:84–90. <https://doi.org/10.1145/3065386>
- Leibe B, Seemann E, Schiele B (2005) Pedestrian detection in crowded scenes. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 1. IEEE, pp 878–885
- N. H. C. of the People’s Republic of China (2020) Daily briefing on novel coronavirus cases in China. <http://en.nhc.gov.cn/2020-03/20/c78006.htm>. Online; Accessed 20 March 2020
- News B (2020) China coronavirus: lockdown measures rise across Hubei province. <https://www.bbc.co.uk/news/world-asia-china51217455>. Online; Accessed 23 Jan 2020
- Ng AY (n.d.) Curriculum Vitae. <https://ai.stanford.edu/~ang/curriculum-vitae.pdf>
- Nguyen CT, Saputra YM, Van Huynh N, Nguyen N-T, Khoa TV, Tuan M et al (2020) Enabling and emerging technologies for social distancing: a comprehensive survey and open problems. *Physics and society*. arXiv:2005.02816
- Niyogi SA, Adelson EH (1994) Analyzing gait with spatiotemporal surfaces. In: *Proceedings of 1994 IEEE Workshop on Motion of Nonrigid and Articulated Objects*. IEEE, pp 64–69. <https://doi.org/10.1109/MNRAO.1994.346253>
- O (2020) Website of Indian Government. Distribution of the novel coronavirus-infected pneumoni Aarogya Setu Mobile App. <https://www.mygov.in/aarogya-setu-app/>
- Piccardi M (2004) Background subtraction techniques: a review. In: *2004 IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp 3099–3104. <https://doi.org/10.1109/ICSMC.2004.1400815>
- Pouw CA, Toschi F, van Schadewijk F, Corbetta A (2020) "Monitoring physical distancing for crowd management: Real-time trajectory and group analysis." *PloS one* 15, no. 10: e0240963. <https://doi.org/10.1371/journal.pone.0240963>.
- Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CA, Munday JD et al (2020) The effect of control strategies to reduce social mixing on outcomes of the covid19 epidemic in Wuhan, china: a modelling study. *Lancet Public Health* 5:261–270. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6)
- Punn NS, Agarwal S (2020) Automated diagnosis of covid-19 with limited posteroanterior chest x-ray images using fine-tuned deep neural networks. *Appl Intell* 51:2689–2702. <https://doi.org/10.1007/s10489-092-01900-3>
- Punn NS, Sonbhadra SK, Agarwal S (2020a) COVID-19 epidemic analysis using machine learning and deep learning algorithms. *Med Comput Sci*. <https://doi.org/10.1101/2020.04.08.20057679>
- Punn NS, Sonbhadra SK, Agarwal S (2020b) Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques. *arXiv preprint arXiv: 2005.01385*
- Punn NS, Sonbhadra SK, Agarwal S, Rai G (2020c) Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques. *arXiv preprint arXiv: 2005.01385*
- Putra M, Yusoff Z, Lim K, Salim S (2018) Convolutional neural network for person and car detection using yolo framework. *J Telecommun Electron Comput Eng (JTEC)* 10(1–7):67–71
- Ramadass L, Arunachalam S, Sagayasree Z (2020) Applying deep learning algorithm to maintain social distance in public place through drone technology. *Int J Pervasive Comput Commun* 16(3):223–234
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 779–788. <https://arxiv.org/abs/1506.02640v5>
- Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. *Comput Vis Pattern Recog* 91–99. <https://arxiv.org/abs/1506.01497v3>
- Robakowska M, Tyranska-Fobke A, Nowak J, Slezak D, Zuratynski P, Robakowski P, Nadolny K, Ładny JR (2017) The use of drones during mass events. *Disaster Emerg Med J* 2(3):129–134. <https://doi.org/10.5603/DEMJ.2017.0028>
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>
- Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations. ICLR 2015* <https://arxiv.org/abs/1409.1556>
- Sonbhadra SK, Agarwal S, Nagabhushan P (2020) Target specific mining of covid-19 scholarly articles using one class approach. *Chaos, Solitons Fractals* 140:110155. <https://doi.org/10.1016/j.chaos.2020.110155>
- Su C-W, Liao H-YM, Tyan H-R (2009) A vision-based people counting approach based on the symmetry measure. In: *IEEE International Symposium on Circuits and Systems*. IEEE, pp 2617–2620
- Sulman N, Sanocki T, Goldgof D, Kasturi R (2008) How effective is human video surveillance performance? *19th international conference on pattern recognition*. IEEE pp 1–3. <https://doi.org/10.1109/ICPR.2008.4761655>
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. *Proc IEEE Conf Comput Vis Pattern Recog* 1:2818–2826. <https://doi.org/10.1109/CVPR.2016.308>
- Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI’17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. pp 4278–4284. <https://doi.org/10.5555/3298023.3298188>
- Tsutsui H, Miura J, Shirai Y (2001) Optical flow-based person tracking by multiple cameras. In: *Conference Documentation International Conference on Multisensor Fusion and Integration for Intelligent Systems*. IEEE, pp. 91–96. <https://doi.org/10.1109/MFI.2001.1013514>
- W. C. D. C. Dashboard (2020). <https://covid19.who.int/>. (Online); Accessed 23 Aug 2020
- Wang X (2013) Intelligent multi-camera video surveillance: a review. *Pattern Recog Lett* 34(1):3–19. <https://doi.org/10.1016/j.patrec.2012.07.005>

- Xu Y, Dong J, Zhang B, Xu D (2016) Background modeling methods in video analysis: A review and comparative evaluation. *CAAI Trans Intell Technol* 1:43–60. <https://doi.org/10.1016/j.trit.2016.03.005>
- Zhao M, Zhao H (2016) Asymptotic behavior of global positive solution to a stochastic sir model incorporating media coverage. *Adv Differ Equ* 149(1):1–17. <https://doi.org/10.1186/s13662-016-0884-5>
- Zhao Z-Q, Zheng P, Xu S-t, Wu X (2019) Object detection with deep learning: a review. *IEEE Trans Neural Netw Learn Syst* 30(11): 3212–3232. <https://doi.org/10.1109/TNNLS.2018.2876865>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.