



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Social distance monitoring framework using deep learning architecture to control infection transmission of COVID-19 pandemic

Imran Ahmed^a, Misbah Ahmad^a, Gwanggil Jeon^{b,*}

^a Institute of Management Sciences, Center of Excellence in Information Technology, 1-A, Sector E-5, Phase VII, Hayatabad, Peshawar, Pakistan

^b Department of Embedded Systems Engineering, Incheon National University, Incheon, Republic of Korea

ARTICLE INFO

Keywords:

Internet of things
Deep learning
Social distancing
COVID-19
Faster-RCNN
Person detection
Transfer learning
Top view

ABSTRACT

The recent outbreak of the COVID-19 affected millions of people worldwide, yet the rate of infected people is increasing. In order to cope with the global pandemic situation and prevent the spread of the virus, various unprecedented precaution measures are adopted by different countries. One of the crucial practices to prevent the spread of viral infection is social distancing. This paper intends to present a social distance framework based on deep learning architecture as a precautionary step that helps to maintain, monitor, manage, and reduce the physical interaction between individuals in a real-time top view environment. We used Faster-RCNN for human detection in the images. As the human's appearance significantly varies in a top perspective; therefore, the architecture is trained on the top view human data set. Moreover, taking advantage of transfer learning, a new trained layer is fused with a pre-trained architecture. After detection, the pair-wise distance between peoples is estimated in an image using Euclidean distance. The detected bounding box's information is utilized to measure the central point of an individual detected bounding box. A violation threshold is defined that uses distance to pixel information and determines whether two people violate social distance or not. Experiments are conducted using various test images; results demonstrate that the framework effectively monitors the social distance between peoples. The transfer learning technique enhances the overall performance of the framework by achieving an accuracy of 96% with a False Positive Rate of 0.6%.

1. Introduction

The novel COVID-19 virus initially reported in Wuhan, China, during late December 2019, and after only in few months, the virus affected millions of people worldwide (Chan et al., 2020). The contagious virus is the kind of Severe Acute Respiratory Syndrome (SARS) spread through the respiratory system. It usually spreads through the air due to the direct exposure of infected (humans, animals) to healthy (humans and animals). The virus transmits through a cough or sneeze droplets of the infected (humans, animals) and travels up to 2 meters (6 feet). World Health Organization (WHO) announced it as a pandemic in March 2020 (WHO, 2020).

Till now, this deadly virus infected almost 8 million people around the globe. The recent number of confirmed cases (R. O. Q. W. R. O. O. G. U. K, 2021) around the globe is presented in Fig. 1. Medical experts, scientists, and researchers have been intensively working and seeking the vaccine and medicine for this lethal virus. To limit the spread of the virus, the global community is looking for alternative measures and

precautions. In the current situation, social distance management is affirmed as one of the best practices to limit the spread of this infection worldwide. It is related to decreasing the physical contact of individuals in crowded environments crowds (such as offices, shopping marts, hospitals, schools, colleges, universities, parks, airports, etc.) and sustaining enough distance between people (Ferguson et al., 2005), (E.C. for Disease Prevention, 2020). By reducing the close physical interaction between individuals, we can flatten the reported cases' curve and slow down the possibilities of virus transmission. Social distance management is crucial for those individuals that are at higher risk of severe sickness from COVID-19.

Social distance management can significantly reduce the risk of the virus' spread and disease severity, as explained in Fig. 2 (adopted from (whiteboxanalytics, 2020)). If a proper social distance mechanism is executed at the primary stages, it might play a central part in preventing the infection transmission and limiting the pandemic disease's peak (number of sick peoples), as evidenced in Fig. 2. Due to a large number of confirmed cases arising on daily basis, which leads to a shortage of

* Corresponding author.

E-mail addresses: imran.ahmed@imsiences.edu.pk (I. Ahmed), misbahahmad4872@gmail.com (M. Ahmad), gjeon@inu.ac.kr (G. Jeon).

<https://doi.org/10.1016/j.scs.2021.102777>

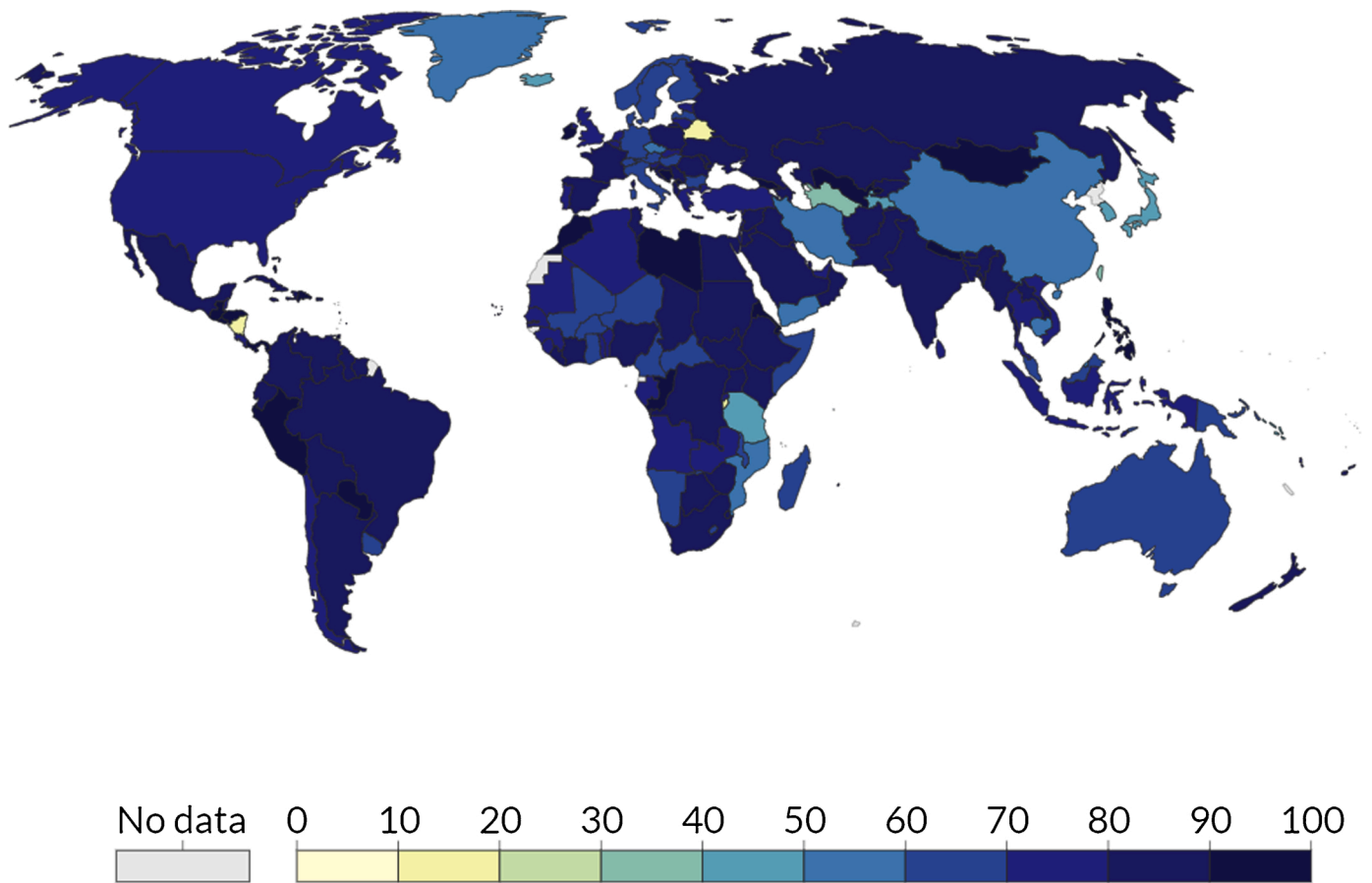


Fig. 1. Latest number of worldwide COVID-19 confirmed cases (R. O. Q. W. R. O. O. G. U. K, 2021).

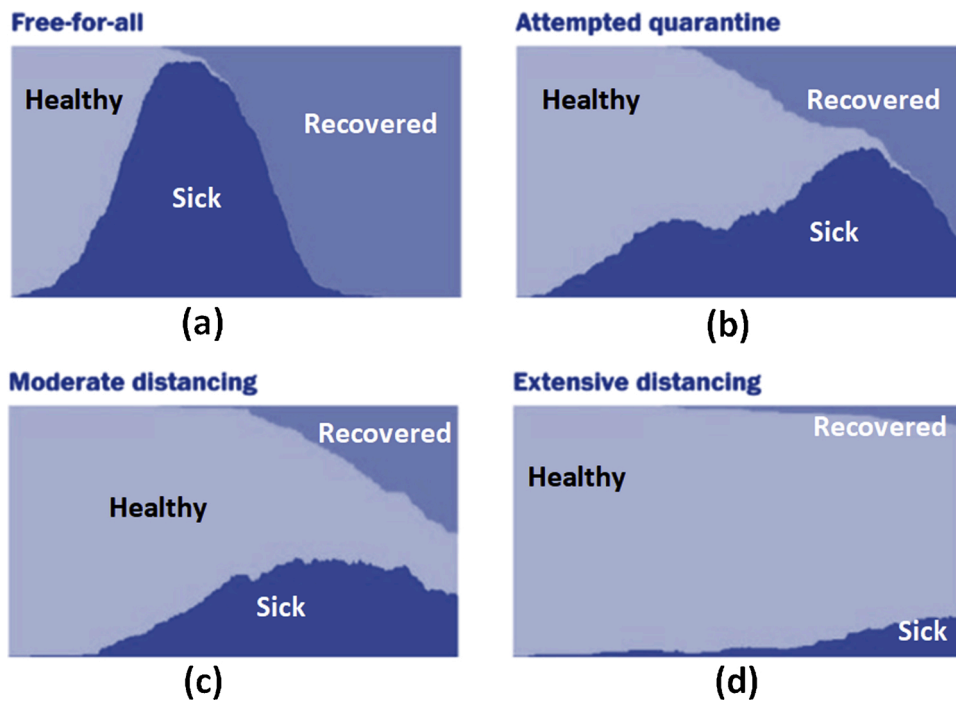


Fig. 2. Graphical illustration of social distance impact: blue curve describes the possible confirmed cases (sick people). It is noticed that with proper social distance management, the peak point of the pattern is at a lower point. The distribution of the peak is very high when no social distance measures are adapted (adapted from (whiteboxanalytics, 2020)).

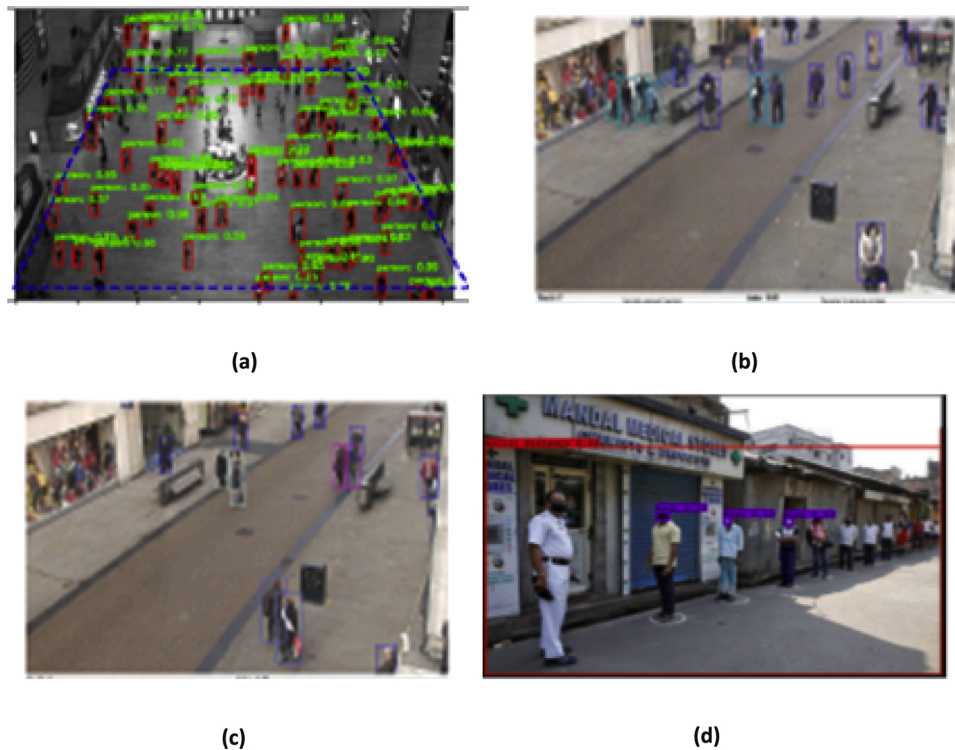


Fig. 3. Different social distance monitoring methods used in Literature. (a) (Yang D. & Renganathan V., 2020) implemented Faster-RCNN for monitoring social distance (b)-(d) (Punn N. S., 2020), (Ramadass et al., 2020), utilized YOLOv3 with Deepsort tracking algorithm for social distance monitoring.

pharmaceutical essentials, it is recognized as one of the most reliable measures to limit the transmission and spread of the contagious infection. From Fig. 2, the virus transmission across global becomes slow down. The graphs clearly indicate that virus transmission is controlled by social distance management as the number of sick people decreases with an increasing number of recovered peoples. With proper social distance mechanisms in public places, the number of infected people/cases and healthcare organizations' burden can be reduced and controlled. It decreases the fatality rates by ensuring that the number of infected cases does not exceed the capacity of healthcare organizations (Nguyen C. T., Van Huynh N., & Khoa T. V., 2020).

Researchers have made different efforts (Yang D. & Renganathan V., 2020), (Punn N. S., 2020) and (Ramadass, Arunachalam, & Sagayasree, 2020), to developed an efficient methods for social distance monitoring. They utilized different machine and deep learning-based approaches to monitor and measured the social distancing among people. Authors use various clustering and distance-based approaches to determine the distance between people; however, they mostly concentrated on the side or frontal view perspectives, as shown in Fig. 3. In such a perspective, peculiar camera calibration is required to map distance to pixels information for practical and measurable units (such as meters, feet). The developed approaches also suffer from occlusion problems as mostly concentrated on the side and frontal view camera perspectives. Conversely, if the same view is captured from the top view, then the distance calculations become better, and occlusion problems are also overcome with a wide coverage of the scene. Researchers, such as (Ahmed & Adnan, 2017), (Ahmad, Ahmed, Khan, Qayum, & Aljuaid, 2020), (Ahmed, Din, Jeon, & Piccialli, 2019), (Ahmed, Ahmad, Adnan, Ahmad, & Khan, 2019), (Ahmed, Ahmad, Piccialli, Sangaiah, & Jeon, 2018), (Ahmad M. & Ullah K., 2021), and (Migniot & Ababsa, 2016) adopted top perspective for detection, counting and tracking of people in different applications. A top view perspective gives a wide field of coverage and handles occlusion problems in a much better way than a side or frontal view.

Due to these discussed characteristics, it also plays a central role in

social distance monitoring. In this work, a deep learning framework is presented to maintain social distance between individuals in a public campus environment by utilizing a top view perspective. For human detection, deep learning architecture, i.e., Faster-RCNN, is applied. The model is initially tested without any training on the top perspective human data set. In some cases where the human body's visual feature is significantly changed compared to the frontal, it gives failure as the architecture was previously trained for the MS-COCO data set (frontal view). Therefore, the architecture is re-trained for top view human images, and with transfer learning, the new trained layer is added with the existing architecture; thereby, the efficiency of the model is improved. The detection architecture identifies people and provides their bounding box information. After the detection of human bounding box information, the central point, also called the centroid detected bounding box, is calculated. To measure the distance between people in the input image, the Euclidean distance is computed between each detected centroid. Using the distance to pixel approximation method, a distance threshold is defined for minimum social distance violation. The detected distance value is checked with a specified threshold value; either it appears under the violation threshold or not. The color of the bounding box is initialized as green; if its distance values appear in the violation list, its color is changed to red. The main contribution of the article is elaborated as follows:

- To develop a social distance monitoring framework utilizing deep learning architecture. The developed framework is used as a prevention measuring tool to reduce close interaction between peoples and limit COVID-19 virus spread.
- To extend pre-trained Faster-RCNN for human detection in top perspective and calculating information of bounding box centroid. Moreover, training of the model for top perspective human data set to enhance the architecture's performance.
- To utilize Euclidean distance approach to compute the distance among each pair of the detected bounding boxes.

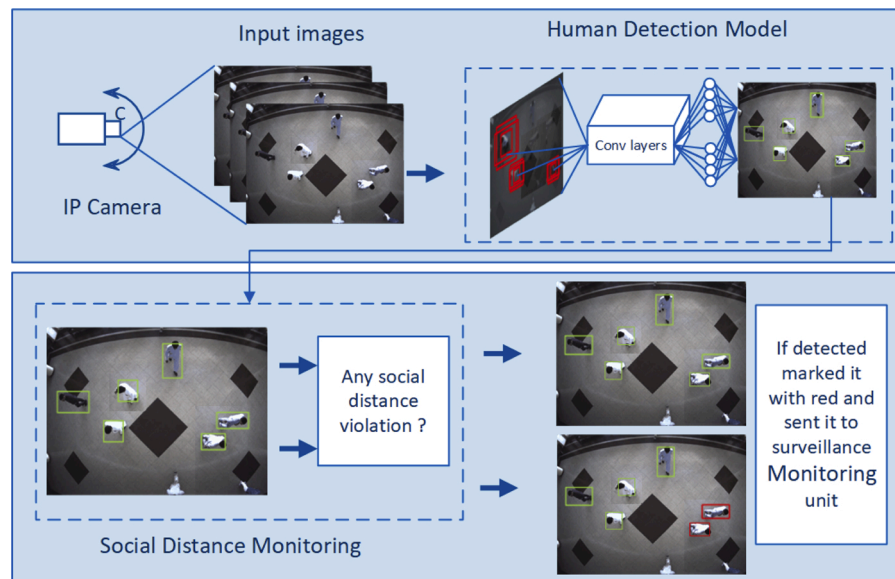


Fig. 4. Overview of framework used for social distance monitoring using top view human data set. The overall system is divided into two modules i.e., Human detection and social distance monitoring.

- To define a distance-based violation threshold for social distance utilizing a pixel to distance estimation approach to monitor people's interaction.
- To evaluate the architecture's performance by assessing it on a top view data set. The results of the presented framework are evaluated with transfer learning.

The remaining work presented in the article is organized as follows. Various methods developed for monitoring of social distance is elaborated in Section 2. In Section 3, the deep learning framework developed for monitoring of social distance is discussed. The data set utilized throughout the experimentation is concisely presented in Section 4. The comprehensive summary of the outcomes and evaluation performance of the framework is likewise explained in Section 4. Section 5 addresses the conclusion of the presented work with future trends.

2. Related work

Social distancing has been identified as the most important practice since late December 2019, after the growth of the COVID-19 pandemic. It is opted as standard practice to stop the infectious virus transmission on January 23, 2020 (B. News, 2020). During the first week of February 2020, the number of cases was increasing on an exceptional basis, with reported cases ranging from 2000 to 4000 per day. Later, there was a relief symbol for the first time, with no new positive cases for five successive days until March 23, 2020, (N. H. C. of the Peoples Republic of China, 2020). This happens because the social distance exercise, which was started in China and, lately, utilized globally to control the spread of COVID-19.

Kylie et al. (Ainslie et al., 2020) examined the correlation linking the social distancing strictness and the region's economic condition and reported that modest steps of this exercise could be adopted for avoiding a massive outbreak. Thus, until now, several nations have adopted solutions based on advanced technology (Punn, Sonbhadra, & Agarwal, 2020) to defeat the pandemic loss. Many advanced nations are exercising Global Positioning System technology to control suspected and infected people's movements. (Nguyen C. T. et al., 2020) presented a review of many developing technologies, consisting of Bluetooth, Wi-fi, GPS, mobile phones, computer vision, image processing, and deep learning that performed a vital function in various possible social distancing situations. Few employed drones and different monitoring

devices to identify people gatherings (Robakowska et al., 2017) and (Harvey & LaPlace, 2019).

Researchers take advantage of the Internet of Medical Things and presented a smart healthcare system for pandemic (Chinmay, Amit, Lalit, Joel, & J.P.C., 2021), and (Chakraborty C. & Garg L., 2021). Prem et al. (Prem et al., 2020) reviewed the effects and impacts of social distancing upon the pandemic outbreak. The investigations presumed that the initial and prompt exercise of social distance could progressively decrease the peak of the infection outbreak. However, social distance is essential for smoothing the infection curve; also, it is a financially bothersome action. (Sun & Zhai, 2020) introduced a study of two new indices: ventilation effectiveness and social distance probability, to quantify and predict the infection possibility of COVID-19. (Ge et al., 2020) and (Jin et al., 2020) investigated the presence of SARS-CoV-2 virus in the atmosphere of an intensive care unit. (Leng, Wang, & Liu, 2020) provided a sustainable design to control the airborne diseases in the courtyard environment. Bhattacharya et al., (Bhattacharya et al., 2020) provided a study of deep learning-based methods used for diagnosis of COVID-19 infection in medical image processing. (Rumpler, Venkataraman, & G”oransson, 2020) examined the result of COVID-19 recommendation measures on noise levels of central Stockholm, Sweden. (Rahman et al., 2020) studied the economic impacts of COVID-19 lockdown exercises using a data-driven dynamic clustering framework. Loey et al. (Loey, Manogaran, Taha, & Khalifa, 2020) used YOLO-v2 and ResNet-50 for facial mask detection.

Rehmani et al. (Rahmani & Mirmahaleh, 2020) provided a systematic review of different prevention and treatment techniques of COVID-19 disease. (Ahmed, Ahmad, Rodrigues, Jeon, & Din, 2020) introduced a social distance monitoring system for COVID-19. (Adolph, Amano, Bang-Jensen, Fullman, & Wilkerson, 2020), studied the situation of the United States during the pandemic. The study concluded that if the decision-makers supported the social distance practice at the initial stages and implemented it at earlier stages, human health might not be affected at such a higher rate.

Taking advantage of deep learning approaches, researchers presented efficient solutions for social distance monitoring. Punn et al. (Punn N. S., 2020) introduced a system utilizing the YOLOv3 with Deepsort algorithm to identify and track people; they practiced an Open Image Data set repository. The scholars also examined results with other deep learning paradigms. (Ramadass et al., 2020) presented a system based on a drone camera for monitoring of social distance. They also

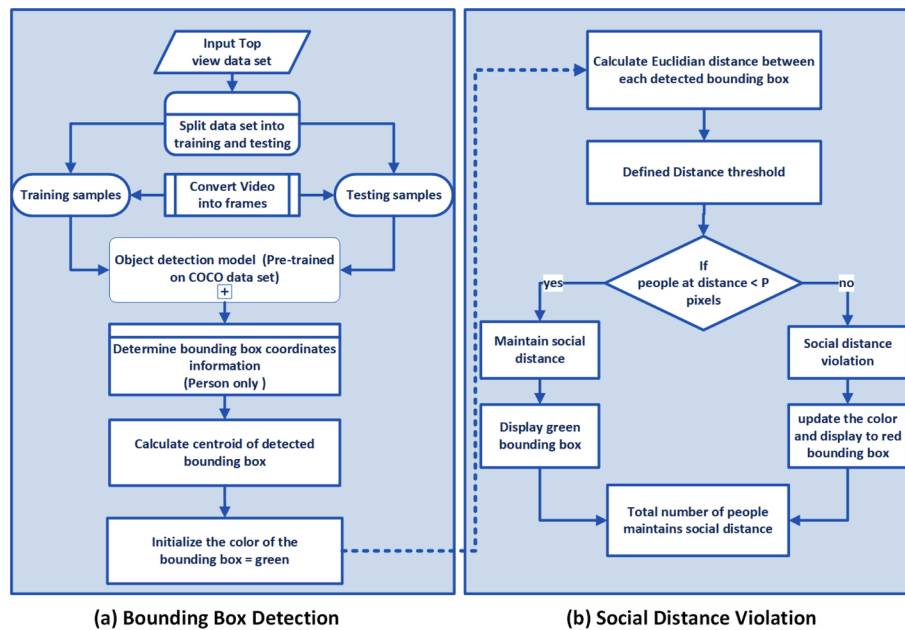


Fig. 5. The overall flow chart of the developed framework utilized for monitoring of social distance with top view human data set.

utilized the YOLOv3 and trained it with their own data set containing side and frontal view images. They also performed the monitoring of facial masks using facial images. Pouw et al., (Pouw C. A. & van Schadedewijk F., 2020) proposed an effective graph-based system for distancing monitoring and masses management. (Sathyamoorthy A. J. & Savle Y. A., 2020) produced a human detection model for a crowded environment. The system is developed for monitoring of social distance. The scholars considered a robot having an RGB depth camera with a 2-D lidar in crowd gatherings to perform collision free navigation.

The above study concluded that a considerable amount of work had been done by researchers in order to provide effective solutions to control the infection transmission of COVID-19. Some utilized medical images and presented a deep learning model for the detection of the COVID-19 virus. Few provided a systematic study of various prevention and treatment techniques for COVID-19. Some researchers presented a social distance monitoring system for public situations. However, mostly techniques concentrated on the side, and frontal camera perspective camera perspective e.g., (Ramadass et al., 2020), and (Punn N. S., 2020).

In this work, we introduced a deep learning framework to monitor social distance in a top view environment. Many researchers utilized top view for different surveillance applications and achieved good results. Ahmad et al., (Ahmad et al., 2020) presented a deep learning model for top view person detection and tracking. Ahmed et al., (Ahmed, Din, et al., 2019) different deep learning models for top view multiple object detection. (Ahmed, Ahmad, Adnan, et al., 2019) introduced a feature-based detector for person detection in different overhead views. Further in (Ahmed et al., 2018) authors proposed a robust features based method for overhead view person tracking. (Ahmad M. & Ullah K., 2021), presented a background subtraction-based person's counting system for an overhead view. Migniot et al. (Migniot & Ababsa, 2016), proposed a hybrid 3D and 2D human tracking system for a top view surveillance environment. Therefore, inspired by these works, we also developed a system that allows a better view of the scene and overcomes occlusion concerns by performing a pivotal role in social distance monitoring.

3. Social distance monitoring framework

In this work, a deep learning framework is introduced for top view social distance monitoring. The general overview of the work presented

in the paper is shown in Fig. 4. The overall work is divided into two modules, person detection and social distance monitoring. The top view installed IP camera captured images that are processed to the human detection module; the human detection module utilizes deep learning architecture and detects human bounding boxes. The bounding box information is further utilized by a social distance monitoring module that identifies the violations and further processed it to the surveillance unit. The detailed and technical explanation of each module is presented in Fig. 5. The overall framework is divided into two main modules, i.e., bounding box detection and social distance violation module. The first module detects the human bounding box, while the second estimate whether social distance violation is made or not. The recorded top view human data set is divided into training and testing samples.

For human detection, a deep learning paradigm is applied, as different kind of object detection algorithms are available, such as (Krizhevsky, Sutskever, & Hinton; Simonyan K., 2014; Girshick R. & Darrell T., 2021; Szegedy C., Jia Y., & Reed S., 2021) and (Girshick R., 2021). Due to the best performance, in this work, we used Faster-RCNN (Ren S. & Girshick R., 2021). As compared to other RCNN-based architectures in Faster-RCNN, the objection detection and region proposal generation tasks are performed by the same convolutional network. Due to this reason the object detection is much faster. The architecture applied a two-stage network to estimate object class probabilities and bounding box. The architecture is previously trained on the MS-COCO data set (Lin T.-Y., Belongie S., & Perona P., 2021).

The architecture is additionally trained for top view human detection. The detection algorithm detects human in the image and initializes its color to green. In the next step, after human detection, the center point, also called the centroid, is calculated utilizing information of detected bounding box coordinates. The distance between each center point of the bounding box is computed using Euclidean distance. A distance-based threshold is defined to identify whether peoples in the top view scene make the social distance violation or not, utilizing the estimated centroid distance information. The people do not maintain or violates the social distance if the estimated distance between people is less than the defined threshold (number of pixels); in such a scenario, the information of the bounding box is collected in violation list, as depicted from Fig. 5, and its color is updated to red. The developed framework displays the total number of violations caused by people and detected people bounding boxes along with centroid at the output. The detail of

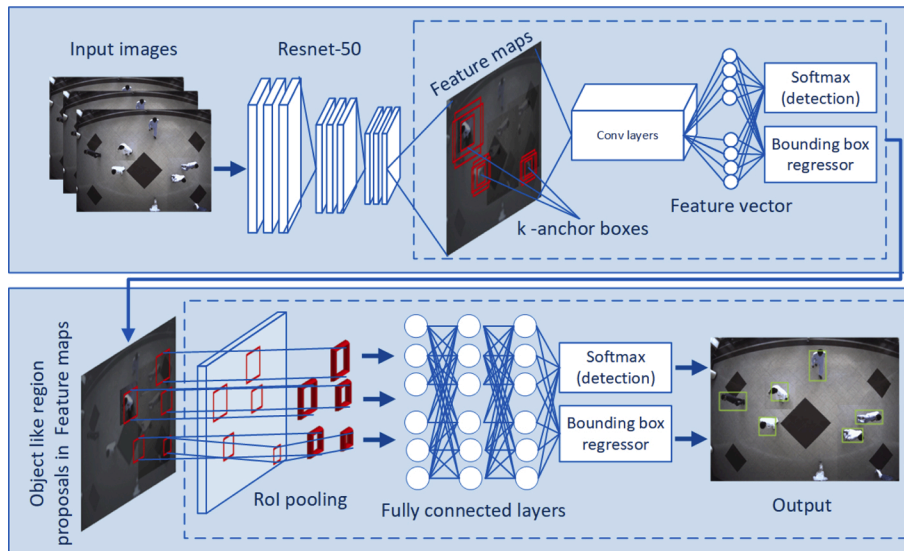


Fig. 6. Human detection module comprises of Faster-RCNN (Ren S. & Girshick R., 2021) with ResNet-(He K. & Ren S., 2021).

the human detection and social distance monitoring module is given in the following subsections:

3.1. Human detection module

A deep learning architecture, i.e., Faster RCNN (Ren S. & Girshick R., 2021), is adopted to perform human detection from the top view perspective. The detection paradigm has a two-stage architecture. The general architecture of Faster RCNN applied for top view human detection is given in Fig. 6. At the first stage, a Region Proposal Network (RPN) (Ren S. & Girshick R., 2021) is employed to produce region proposals or feature maps for the sample image. Convolution layers are used at the first stage to produce feature maps. The next stage utilizes a Fast-RCNN architecture (Wu, Huang, & Sun, 2016), a deep convolutional network to choose specific object features from the various region proposals, as described in Fig. 6. Finally, in the end, the softmax and bounding box classifier is used to detect the object (human) in the top view scene. The RPN utilizes Convolutional Neural Network (CNN) layers to section the entire input image. It outputs many orthogonal region proposals for an arbitrary size image.

In our case, as discussed earlier, the human body visual appearance is varying in the scene, and due to the highly flexible characteristic of CNNs, the RPN is tuned for various size region proposals. We used ResNet-50 as a backbone architecture for RPN generation that utilizes a sliding window approach over the last layer's feature maps. As compared to conventional deep learning architectures like VGG, which usually have convolution layers for classification, rather than fully connected layers, without any shortcut/skip connection defined as plain networks (for more explanation, readers are referred to (He K. & Ren S., 2021)). It utilizes a skip connection or termed as shortcut connection, which provides a more extensive network to map the preceding layer's input to the consequent layer input, without any modification/alteration in the input.

The RPN produces region proposals, also termed as k-anchors at every window location, with various aspect ratios, scale sizes, and provide a reliable translation-invariant feature. The resulting feature vector is later fed into convolution layers, referred to as softmax detection (used as class score layer) and a regression layer (applied for bounding box regression). The offset of each bounding box coordinate is predicted using a regression layer. For each detected bounding box, it outputs the bottom left and top corner coordinates information. For detection, Fast-RCNN is applied to detect human in each input image. As explained in Fig. 6, the corresponding region proposals are given as an

input to the softmax classifier to generate the class scores and bounding box. The base network, i.e., Resnet-50, generates these feature maps, which are provided to the Region of Interest (RoI) pool layer that manipulates the detected region proposals. It firstly matches the feature map to every region proposal; therefore, the features are at the corresponding location as the feature map detected. Later, it applies max pooling to modify the feature maps into a region of interest. It is a rectangular or quadrilateral window having (w, h, x, y) coordinates that indicate the width height, top left, and bottom corner. The bounding box coordinates are mathematically determined as follows:

$$\begin{aligned}
 t_x &= \frac{(x - x_a)}{w_a}, & t_y &= \frac{(y - y_a)}{h_a} \\
 t_x^* &= \frac{(x^* - x_a)}{w_a}, & t_y^* &= \frac{(y^* - y_a)}{h_a} \\
 t_w &= \log \frac{w}{w_a}, & t_h &= \log \frac{h}{h_a} \\
 t_w^* &= \log \frac{w^*}{w_a}, & t_h^* &= \log \frac{h^*}{h_a}
 \end{aligned} \tag{1}$$

In Eq. (1), x, y , expressed bounding box coordinates, and h, w determines the height and width of the detected bounding box. The ground truth, anchor box, and predicted bounding box is indicated with x^*, x_a , and x , sequentially. The object class score is obtained from the class score layer, i.e., person or human. It can be seen, in Fig. 6, that top view images are passed through convolutional layers that produce feature maps also shown with k-anchors. The anchors are divided into two classes (person and no-person (background)) for anchor selection. The Intersection over Union (IoU) is applied to describe where the anchor or proposed regions are projected with ground truth bounding boxes G_T . The IoU ratio is defined and provided as (Ren S. & Girshick R., 2021) and (Ahmad et al., 2020):

$$\text{IoU} = \begin{cases} \frac{B_{\text{Box}}(\text{anchor}) \cap G_T}{B_{\text{Box}}(\text{anchor}) \cap G_T} > 0.7 = \text{person} \\ \frac{B_{\text{Box}}(\text{anchor}) \cap G_T}{B_{\text{Box}}(\text{anchor}) \cap G_T} < 0.3 = \text{no - person.} \end{cases} \tag{2}$$

Finally, a function is determined at the edge of the RPN's. To represent the predicted bounding box positions, the regression function is used. The function is specified to calculate the loss of bounding box regression,

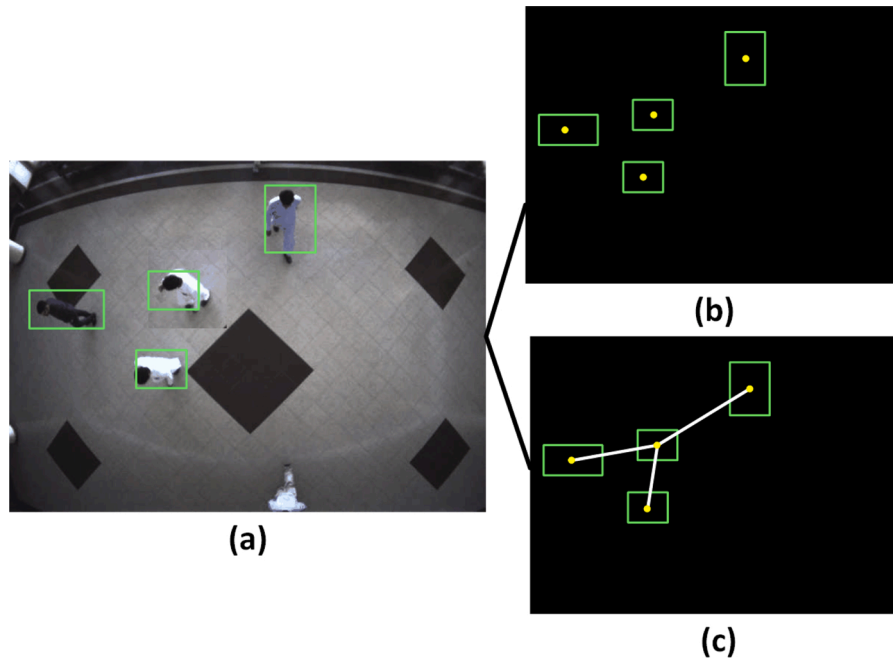


Fig. 7. (a) Detected human bounding boxes via detection module, (b) shows the calculated centroid/central point of the bounding box, and (c), illustrates the estimated distance of every detected bounding box. In the sample image, the distance between every pairwise detected bounding box is indicated with white lines; the green circles show the central point.

and classification (Ren S. & Girshick R., 2021) as:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (3)$$

In Eq. (3), L_{cls} indicates loss of classification, L_{reg} exhibits loss of bounding box regression. The coefficients of normalization are described with N_{cls} and N_{reg} , and the parameter used as the weight between two losses, and it is determined with λ . As an index for an anchor, i is used; the predicted probability of a human or person is described as p_i , while p_i^* is used as ground truth for classification. The anchor region associates with a positive class if $p_i^* = 1$, and the anchor region corresponds to a negative class if $p_i^* = 0$. t_i represents a predicted bounding box vector, where t_i^* represents ground truth. For L_{cls} classification loss of person and no-person, a logarithmic loss is determined and provided as:

$$L_{cls}(p_i, p_i^*) = -\log(p_i, p_i^*) + [(1 - p_i^*)(1 - p_i)]. \quad (4)$$

Applying Eq. (4), the regression loss is provided as (Ren S. & Girshick R., 2021):

$$L_{reg}(t_i, t_i^*) = \text{smooth}_{L_1}(t_i - t_i^*). \quad (5)$$

The model outputs bounding box information containing (person), as shown in green boxes in Fig. 6. These detected bounding box information is further processed to the social distance monitoring framework.

3.2. Social distance monitoring module

After human detection in top view images, the center point, also called the centroid of each detected bounding boxes is computed as shown with green boxes in Fig. 7(a). The estimated centroid of the bounding box is displayed in Fig. 7, the Fig. 7b illustrates a set of bounding box coordinates with its center point information. The centroid of the bounding box is computed as:

$$c_{(x,y)} = \frac{\hat{x}_{min} + \hat{x}_{max}}{2}, \frac{\hat{y}_{min} + \hat{y}_{max}}{2}. \quad (6)$$

In Eq. (6), \hat{x} represents the minimum and maximum value for width, and \hat{y} is the minimum and maximum height of the bounding box. After

computing, centroid, the distance is measured between each detected centroid, by applying the Euclidean distance approach. For every detected bounding box in the image, we firstly calculate the centroid of bounding boxes presented in Fig. 7(b); formerly the distance is measured (represented with red color lines) between every detected bounding boxes, Fig. 7(c). The distance between the two centroids is mathematically computed as;

$$D(c_{2(x,y)}, c_{1(x,y)}) = \sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2}. \quad (7)$$

A threshold value of T is defined using calculated distance values. This value is used to check whether any two or more people are at smaller distances than the defined T pixels threshold. The threshold T is applied using below equation:

$$V = \begin{cases} 1, & T \geq D \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

The threshold value T estimates the social distance violation V , in terms of the image pixels. If the computed distance value of two centroids is small, then the minimum social distance violation threshold is not maintained and detected people are close to each other. This information is collected and stored in the violation list. The detected bounding box is initialized as green color by detection architecture. The stored information is verified with a violation list; if the information exists and the detected bounding box are too close, the color is updated and changed to red. The developed framework presents information of total social distancing violations at the output, which is further processed to the surveillance unit.

4. Experiments, results, and discussion

This section explains different experiments conducted in this work. For monitoring of social distance, a human data set (Ahmed, Ahmad, Adnan, et al., 2019), (Ullah K. & Ahmad M., 2019), and (Ahmed, Ahmad, Nawaz, et al., 2019) recorded at the Institute of Management Sciences, Hayatabad, Peshawar, Pakistan, is practiced. The data set includes multiple people images, captured utilizing a single top-view camera, mounted at 6m of height. The video sequences were recorded

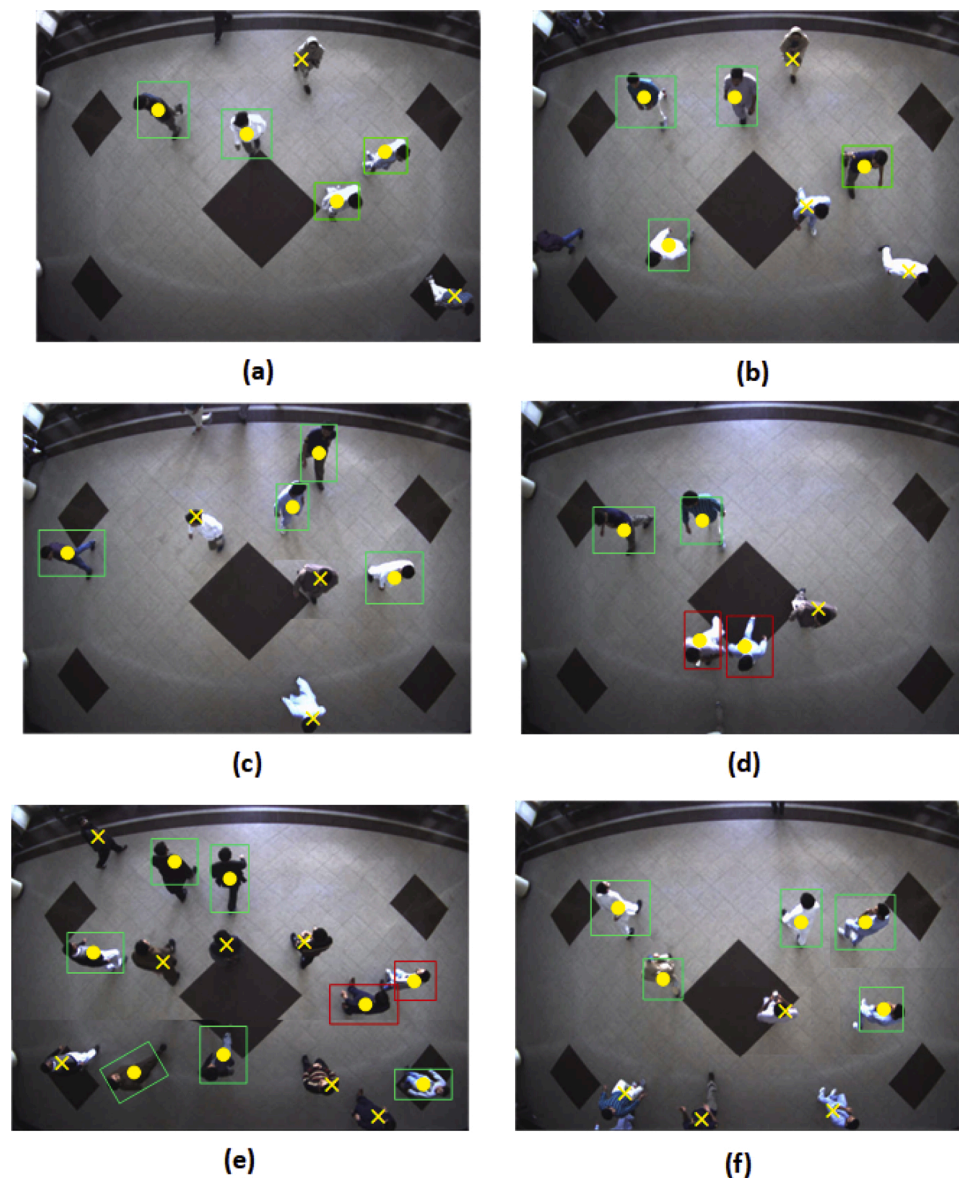


Fig. 8. Results of pre-trained detection utilized for monitoring of social distance from the top view. In sample images, the peoples who maintain social distancing are detected in green rectangles. The people who violate and do not maintain social distance are presented in red rectangles. The manually yellow crosses points not detected results.

at 20 frames per sec, with 640×480 pixels in PNG format. As because of the wide-angle lens practiced, the person's size or scale varies when moving away from the scene's center. For implementation purposes, OpenCV is utilized with Keras library. The overall experimental outcomes are classified into two subsections; the first section describes the visualization detection results, whereas the second subsection elaborates the performance evaluation. For a fair comparison, both pre-trained and trained models are tested, applying the same images.

4.1. Results of pre-trained architecture for social distance monitoring

The detection results of the pre-trained architecture have been shown in Fig. 8. The results are assessed utilizing different top view sample images. The people are smoothly walking in the scenes without any pre-defined constraints; it can be examined from samples that the person's visual features are not similar to the side or frontal perspective (Fig. 8). The size and scale are changing at various locations. The architecture only identifies the human class; thus, an object with features similar to a human is identified. The pre-trained architecture gives

better detection results with multiple sized bounding boxes, as presented by green color in Fig. 8.

For example, in images of Fig. 8(a), (b), and (d) people in green boxes are those who keep the social distance threshold. The detection architecture has also experimented for a different number of people's images, as shown in Fig. 8(e), and (f), different peoples in the scene are effectively identified and observed by the developed framework. In all images, it is observed that after human detection, the distance is estimated between the detected bounding boxes, in order to verify whether the people in the image or view maintain the defined social distance threshold or not. In Fig. 8(d), two individuals entering the scene indicated in red color bounding boxes are those who do not maintain the social distance threshold. Some not detected results are also shown with yellow crosses in sample images. From the samples, it is viewed that a person is adequately identified at various scenic locations. Despite, in a few cases, where the person's visual features are changing; hence, the architecture shows not detected results. The reason might be that pre-trained architecture is utilized, and human visual features from a top perspective are different, leading to not detecting results.

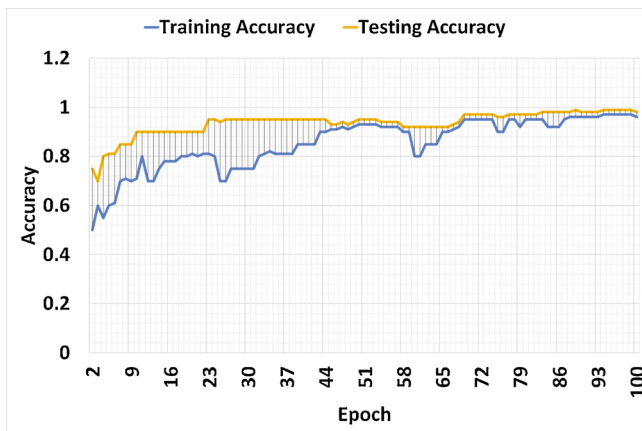


Fig. 9. Training accuracy of detection architecture after training on top view data set.

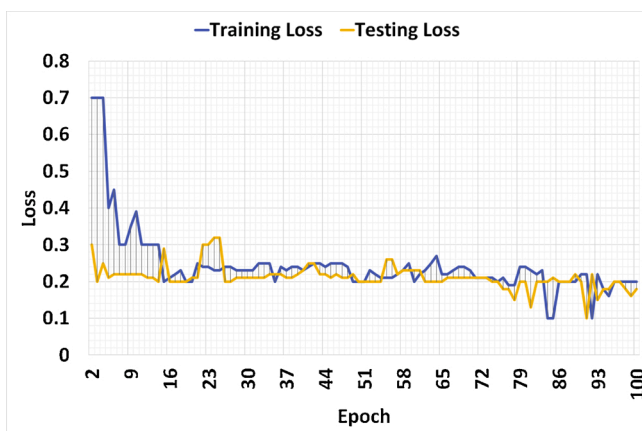


Fig. 10. Training loss of detection architecture after training on top view data set.

4.2. Results of the architecture for social distance monitoring with transfer learning

The transfer learning approach is adopted to increase the performance and efficiency of the human detection module. For training and testing purposes, a total of 1000 sample images are utilized. The training and testing samples are randomly splitted at the ratio of 70% and 30%, respectively. The detection architecture is further trained using 700 images of the top view data set. The batch size 64 and epoch size 100 is set for training. The accuracy and loss of the architecture are presented in Figs. 9 and 10. An additional layer is produced after training and is combined with a pre-trained architecture. It can be seen that both training and testing accuracy are raised considerably at the beginning of the 11th epoch because we studied only a particular class object, i.e., a person. From Fig. 10, it is also important to note that both training and testing loss reducing after the 10th epoch.

The detection architecture is again tested for similar test images, as presented in the previous section. The experimental outcomes show that transfer learning considerably enhances the results of human detection, as observed in Fig. 11. In sample images, it is observed that the detection architecture identifies the multiple peoples at different locations of the scene. Multiple people with different visual features are efficiently classified, and the social distance among individuals is also measured, as

displayed in the sample images. In Fig. 11(a), and (b), no violation is observed as all individual are labeled with green color bounding boxes detected by the presented framework. In example images Fig. 11(d) and (e), the violation is observed. As compared to Fig. 11(c) and (f), the number of individuals in the scene is small, and all people keep a social distance threshold, and there is no violation noted and observed. In Fig. 11(d), because of close interactions among people, the violation is reported by a deep learning-based automated framework. Similar behavior is found in Fig. 11(e), where there is a large number of people. The framework efficiently detected people and social distance violations with red color bounding boxes if the estimated distance between them is small and close to each other.

4.3. Performance evaluation

Different parameters have been utilized for evaluation purposes, i.e., true-negative, false-positive, and false-negative. Furthermore, these parameters are applied to estimate Accuracy, Recall, F1-Score, Precision, True Positive Rate (TPR), and False Positive Rate (FPR). The Accuracy, Recall, F1-score, and Precision results of both pre-trained and trained detection architecture are reflected in Fig. 12. We used the standard error method to show the average values. It can be observed that the accuracy of the detection architecture is 96%, Recall is 92%, the F1-score is 94%, and the Precision is 95%.

These findings are evident that transfer learning and additional training enhances detection results. The TPR and FPR of the detection architecture after training for the top view human data set is depicted in Fig. 13 and Fig. 14. The TPR of the framework increased from 91% to 96%. Furthermore, the FPR of the framework is improved and reduced from 0.9% to 0.6%, which reveals the efficiency of the deep learning architecture.

5. Conclusion and future work

A social distance monitoring framework is introduced based on deep learning architecture to control infection transmission of COVID-19 pandemic. The overall framework is categorized into two modules: the human detection and social distance monitoring module. The pre-trained Faster-RCNN paradigm is utilized for top view human detection purposes. The top perspective is substantially different from the frontal perspective; thus, the transfer learning strategy is performed to boost human detection results. The architecture has been trained on the top view human data set, and the new layer is fused with the existing architecture. As far as we know, it may be the first effort that employed transfer learning for Faster-RCNN, i.e., and practiced social distance monitoring in a top perspective. The information of the bounding box obtained from the human detection module is used to determine the centroid or central point coordinates information. Applying Euclidean distance, the distance between the detected bounding boxes is estimated. A distance-based threshold is determined, and the pixel to physical distance approach is applied to examine social distance violations. Experimental outcomes revealed that the presented framework effectively and efficiently recognizes close interactions between peoples and those who do not maintain a social distance violation threshold. In addition, transfer learning enhances the general efficiency and robustness of the detection architecture. For a pre-trained architecture, the accuracy of the framework is 91%, and after training and transfer learning, it archives an accuracy of 96%. In the future, the work may be enhanced for various outdoor and indoor situations. Various detection architectures might be practiced for social distancing monitoring.

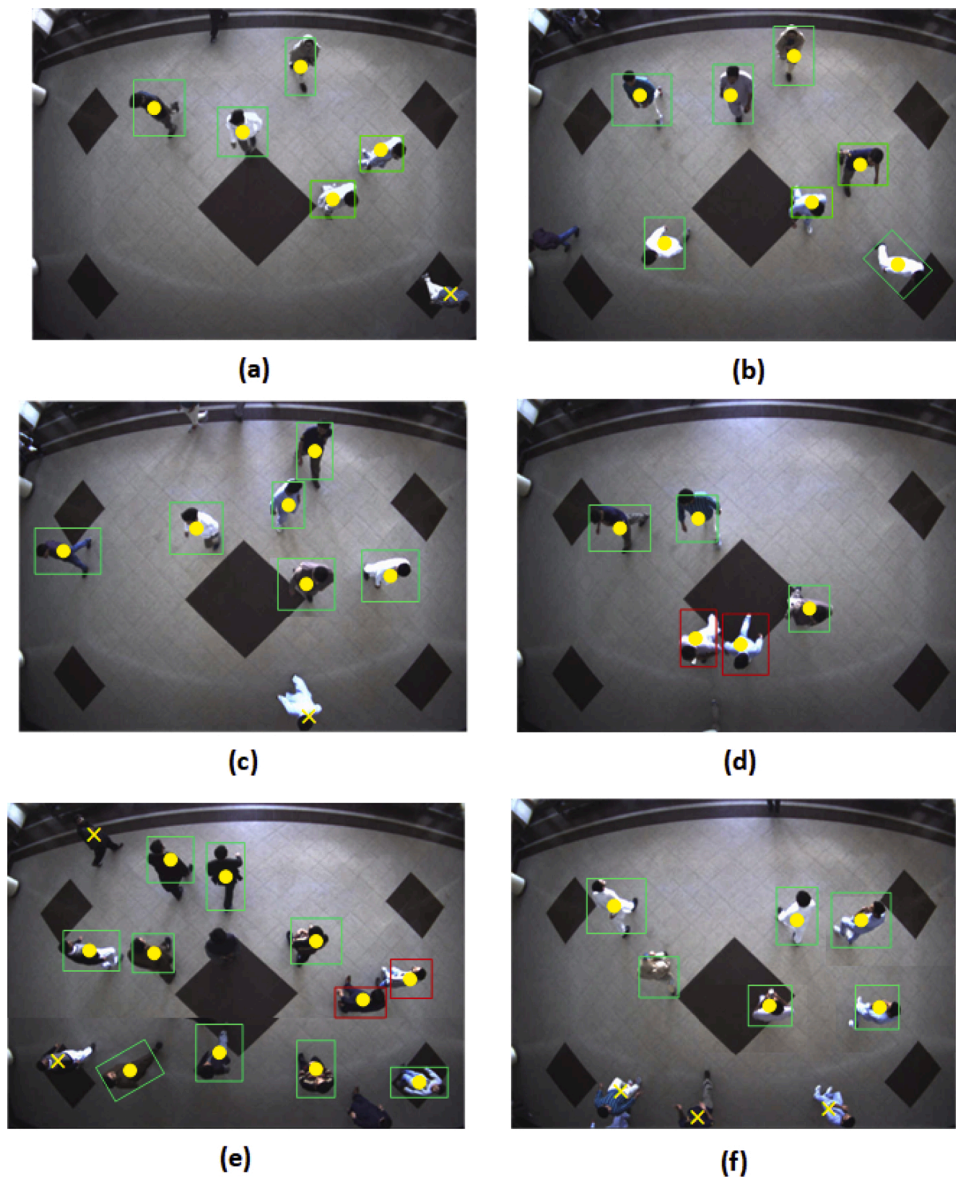


Fig. 11. Results of social distance monitoring after training and applying transfer learning. It can be observed that the performance of the detection architecture is enhanced. In sample images, the people who maintain social distance are detected with green bounding boxes, whereas those who do not maintain the social distance threshold are presented in red rectangles.

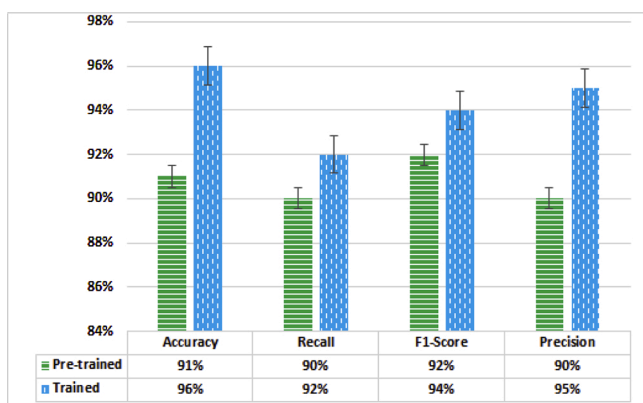


Fig. 12. Accuracy, recall, F1-score and precision results.

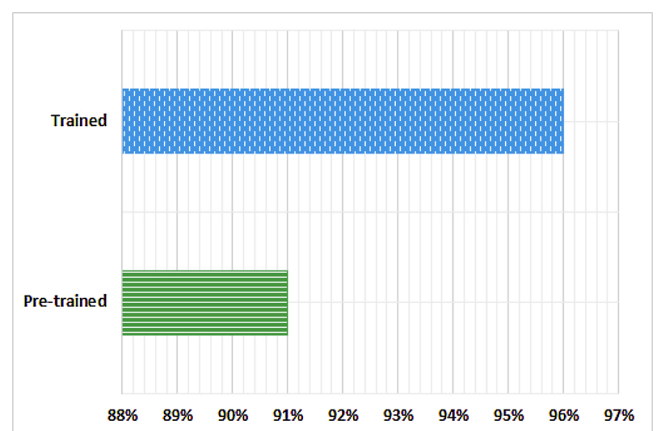


Fig. 13. True Positive rate of trained an pre-trained architecture.

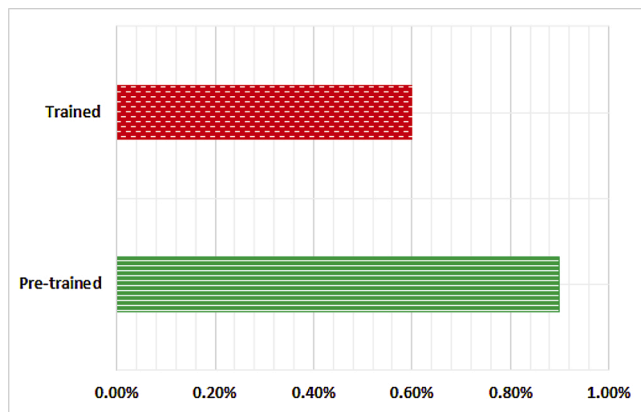


Fig. 14. False Positive rate of trained and pre-trained architecture.

Conflict of interest

None.

Acknowledgment

This work was supported under the framework of international cooperation program managed by the National Research Foundation of Korea (2019K1A3A1A8011295711).

References

- Adolph, C., Amano, K., Bang-Jensen, B., Fullman, N., & Wilkerson, J. (2020). *medRxiv*.
- Ahmad, M., Ahmed, I., Khan, F. A., Qayum, F., & Aljuaid, H. (2020). *International Journal of Distributed Sensor Networks*, 16, 1550147720934738.
- Ahmad M., Ahmed I., Ullah K., Khan I., Adnan A. 2018 9th IEEE annual ubiquitous computing, electronics mobile communication conference (UEMCON) (pp. 746–752). doi:10.1109/UEMCON.2018.8796595.
- Ahmed, I., & Adnan, A. (2017). *Cluster Computing*, 1–22.
- Ahmed, I., Ahmad, A., Piccialli, F., Sangaiah, A. K., & Jeon, G. (2018). *IEEE Internet of Things Journal*, 5, 1598–1605.
- Ahmed, I., Ahmad, M., Nawaz, M., Haseeb, K., Khan, S., & Jeon, G. (2019). *Computer Communications*, 147, 188–197.
- Ahmed, I., Din, S., Jeon, G., & Piccialli, F. (2019a). *IEEE Internet of Things Journal*.
- Ahmed, I., Ahmad, M., Adnan, A., Ahmad, A., & Khan, M. (2019b). *International Journal of Machine Learning and Cybernetics*, 1–12.
- Ahmed, I., Ahmad, M., Rodrigues, J. J., Jeon, G., & Din, S. (2020). *Sustainable Cities and Society*, 102571.
- Ainslie, K. E., Walters, C. E., Fu, H., Bhatia, S., Wang, H., Xi, X., et al. (2020). *Wellcome Open Research*, 5.
- Bhattacharya, S., Maddikunta, P. K. R., Pham, Q.-V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., et al. (2020). *Sustainable cities and society*, 102589.
- R. O. Q. W. R. O. O. G. U. K. Blavatnik School of Government, University of Oxford (2020). <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>. Accessed 20.01.21.
- Chakraborty C., Banerjee A., Garg L., Rodrigues J. (2021). Springer.
- Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., Chu, H., Yang, J., et al. (2020). *The Lancet*, 395, 514–523.
- Chinmay, C., Amit, B., Lalit, G., Joel, R., & J.P.C. (2021). *Series Studies in Big Data*, 80, 98–136. <https://doi.org/10.1007/978-981-15-8097-0>
- E.C. for Disease Prevention, S. Control (2020). <https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-social-distancing-measures-guide-second-update.pdf>.
- Ferguson, N. M., Cummings, D. A., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., et al. (2005). *Nature*, 437, 209–214.
- Ge, X.-Y., Pu, Y., Liao, C.-H., Huang, W.-F., Zeng, Q., Zhou, H., et al. (2020). *Sustainable cities and society*, 61, 102413.
- Girshick R., Donahue J., Darrell T., Malik J. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580–587).
- Girshick R. Proceedings of the IEEE international conference on computer vision (pp. 1440–1448).
- Harvey, A., & LaPlace, J. (2019). *Megapixels: Origins, ethics, and privacy implications of publicly available face recognition image datasets*.
- He K., Zhang X., Ren S., Sun J. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778).
- Jin, T., Li, J., Yang, J., Li, J., Hong, F., Long, H., et al. (2020). *Sustainable cities and society*, 102446.
- Krizhevsky A., Sutskever I., Hinton G.E. Advances in neural information processing systems (pp. 1097–1105).
- Leng, J., Wang, Q., & Liu, K. (2020). *Sustainable Cities and Society*, 62, 102405.
- Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., et al., in: European conference on computer vision, Springer (pp. 740–755).
- Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2020). *Sustainable Cities and Society*, 102600.
- Migniot, C., & Ababsa, F. (2016). *Journal of Real-Time Image Processing*, 11, 769–784.
- B. News (2020). <https://www.bbc.co.uk/news/world-asia-china-51217455> Accessed 23.01.20.
- Nguyen C. T., Saputra Y. M., Van Huynh N., Nguyen N.-T., Khoa T. V., Tuan B. M., et al. (2020). arXiv preprint arXiv:2005.02816.
- N. H. C. of the Peoples Republic of China (2020). <http://en.nhc.gov.cn/2020-03/20/c78006.htm> Accessed 20.03.20.
- Pouw C. A., Toschi F., van Schadewijk F., Corbetta A. (2020). arXiv preprint arXiv:2007.06962.
- Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., et al. (2020). *The Lancet Public Health*.
- Punn N. S., Sonbhadra S. K., Agarwal S. (2020). arXiv preprint arXiv:2005.01385.
- Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020b). *medRxiv*.
- Rahman, M. A., Zaman, N., Asyhari, A. T., Al-Turjman, F., Bhuiyan, M. Z. A., & Zolkipli, M. (2020). *Sustainable cities and society*, 62, 102372.
- Rahmani, A. M., & Mirmahaleh, S. Y. H. (2020). *Sustainable cities and society*, 102568.
- Ramadass, L., Arunachalam, S., & Sagayaraj, Z. (2020). *International Journal of Pervasive Computing and Communications*.
- Ren S., He K., Girshick R., Sun J. Advances in neural information processing systems (pp. 91–99).
- Robakowska, M., Tyranska-Fobke, A., Nowak, J., Slezak, D., Zuratynski, P., Robakowski, P., et al. (2017). *Disaster and Emergency Medicine Journal*, 2, 129–134.
- Rumpler, R., Venkataraman, S., & G”oransson, P. (2020). *Sustainable cities and society*, 63, 102469.
- Sathyamoorthy A. J., Patel U., Savle Y. A., Paul M., Manocha D. (2020). arXiv preprint arXiv:2008.06585.
- Simonyan K., Zisserman A. (2014). arXiv preprint arXiv:1409.1556.
- Sun, C., & Zhai, Z. (2020). *Sustainable cities and society*, 62, 102390.
- Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., et al. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1–9).
- Ullah K., Ahmed I., Ahmad M., Khan I. 2019 9th annual information technology, electromechanical engineering and microelectronics conference (IEMECON), IEEE (pp. 284–289).
- whiteboxanalytics. <https://www.whiteboxanalytics.com.au/white-box-home/flattening-the-coronavirus-curve> Accessed 12.03.20.
- WHO. <https://www.who.int/dg/speeches/detail/2020>. Accessed 12.03.20.
- Wu, X., Huang, G., Sun, L., et al. (2016). *Robot*, 38, 711–719.
- Yang D., Yurtsever E., Renganathan V., Redmill K.A., ”Ozg“uner “U. (2020). arXiv preprint arXiv:2007.03578.