sciendo

Transport and Telecommunication, 2022, volume 23, no. 2, 160–167
Transport and Telecommunication Institute, Lomonosova 1, Riga, LV-1019, Latvia
DOI 10.2478/ttj-2022-0014

SOCIAL DISTANCE EVALUATION IN TRANSPORTATION SYSTEMS AND OTHER PUBLIC SPACES USING DEEP LEARNING

Marco Guerrieri[1], Giuseppe Parla[2]

DICAM, University of Trento
Via Mesiano, 77, Trento, Italy
[1] marco.guerrieri@unitn.it, [2] peppeparla@yahoo.it

This research put forward an efficacious real-time deep learning-based technique to automate the process of monitoring the social distancing in transportation systems (e.g., bus stops, railway stations, airport terminals, etc.) and other public spaces with the purpose to mitigate the impact of coronavirus pandemic. The proposed technique makes use of the YOLOv3 model to segregate humans from the background of each image of a surveillance video and the linear Kalman filter for tracking the humans' motion even in case in which another object or person overlaps the trajectory of the person under analysis. The performance of the model in human detection is extremely high as demonstrated by the accuracy of the model that reaches values higher than 95%. The detection algorithm can be applied for alerting people to keep a safe distance from each other when they are in crowded places or in groups.

Keywords: COVID-19 pandemic, Social distance evaluation, Deep Learning, YOLOv3 model

1. Introduction

Since 2019 the Coronavirus disease 19 (COVID-19) has become the new global public health crisis. The World Health Organization (WHO) affirmed this novel virus disease as a pandemic on March 11, 2020. The virus is highly contagious and the incubation period ranges between 2-14 days depending on the variants of the virus. In accordance with the WHO, physical distancing helps limit the spread of COVID-19 and it is suggested to keep an adequate distance between people (at least 1m) and avert spending time in crowded places or in groups. Majority epidemiological models descend from the classical SIR model of Kermack and McKendrick (Kermack & McKendrick, 1927). In recent times Eksin *et al.* (2019) have presented a modified SIR model with the inclusion of a social distancing parameter. Social distancing is crucial, mainly for those people who are at higher risk of serious illness from COVID-19 and for unvaccinated people. Figure 1 shows how protective measures such as avoiding crowds and maintaining the minimum social distancing can slow down the development of new COVID-19 cases and reduce the risk of overwhelming the health care system (Harvard Medical School, 2020). It can be noted that social distancing is the primary strategy for reducing the number of infected people and mortality rates and brings potential benefits on the healthcare structures. Transportation systems play a significant role in promoting transmission given the crowded conditions into the stations and vehicles (Gossling, 2020; Musselwhite, 2020). Therefore, the control of the social distance between people, both in public and private places and also in transportation systems, is of fundamental interest for limiting the pandemic's damage to the health system and economy. However, social distancing monitoring for a large population cannot be done manually. Recent improvement in deep learning (DL) algorithms allows object detection in a very precise way (Brunetti *et al.*, 2018); therefore, DL can be used as a performing method for determining the instantaneous distancing between people and for checking the social distance violations (Ahmed *et al.*, 2021). In this field, the most common method for measuring the social distance is the top-down approach which requires an overhead view perspective and as a result is rarely applicable in real world cases. Motivated by this notion, this research proposes an analysis of frontal or side view video by Deep Learning and Yolov3 algorithms for human detection and tracking and for social distance calculations in each instant of time using surveillance video. In the detection phase each pedestrian is detected and a bounding box is associated. The social distance is evaluated by the Euclidean distance between each detected centroid pair. In addition, the application in the proposed model of the linear Kalman filter allows estimating the

pedestrian trajectories and the distance even in case of overlapping between with other pedestrians in the scene. The key objectives of this article are as follows: Section 2 explains the main characteristics of human detection and recognition methods based on deep learning approach and YOLOv3 algorithm, Section 3 briefly explains the whole proposed approach for social distancing monitoring. The experiments are presented in Section 4 along with results and discussions. Finally, conclusions are given in Section 5.

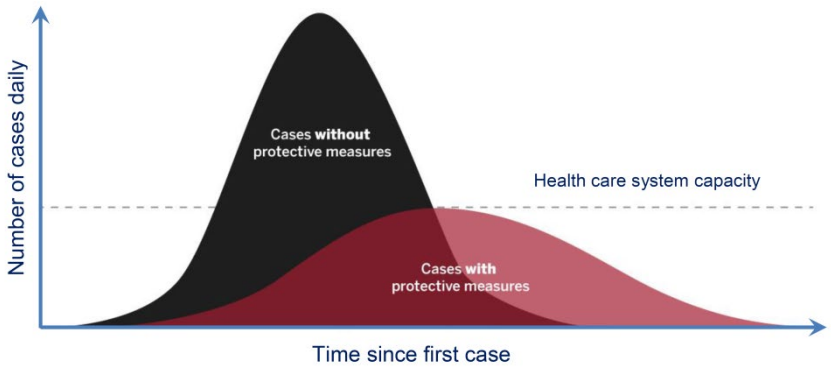

Figure 1. Qualitative correlation between the number of cases per day and the time with and without protective measures such as social distancing (adapted from Harvard Medical School, 2020)

2. Deep learning and Yolov3 model

Deep learning (DL) has shown numerous technical applications in several real-life areas including human detection. The models implemented in DL are designed using a large number of tagged data and neural network architectures that contain different layers (Lechgar *et al.*, 2019). Object detection systems like YOLO (You Only Look Once), SSD (single-shot detector) and Faster R-CNN (convolution neural network), not only classify images but also can locate and detect each object in images that contain multiple objects (Elgendy, 2020). Punn *et al.* (2021) have demonstrated that YOLO v3 achieve efficient performance (with balanced FPS and mAP score) in human detection with respect to Faster RCNN, SSD (Punn *et al.*, 2021).

YOLOv3 includes 53 convolutional layers and 23 residual layers and has proved significant advancement in real-time object detection, especially in the detection of small objects in the analysed scenes (Redmo & Farhadi, 2018). Therefore, in this research YOLOv3 is used as a real time human detection system. The model's overall architecture used in this research is presented in Figure 2.

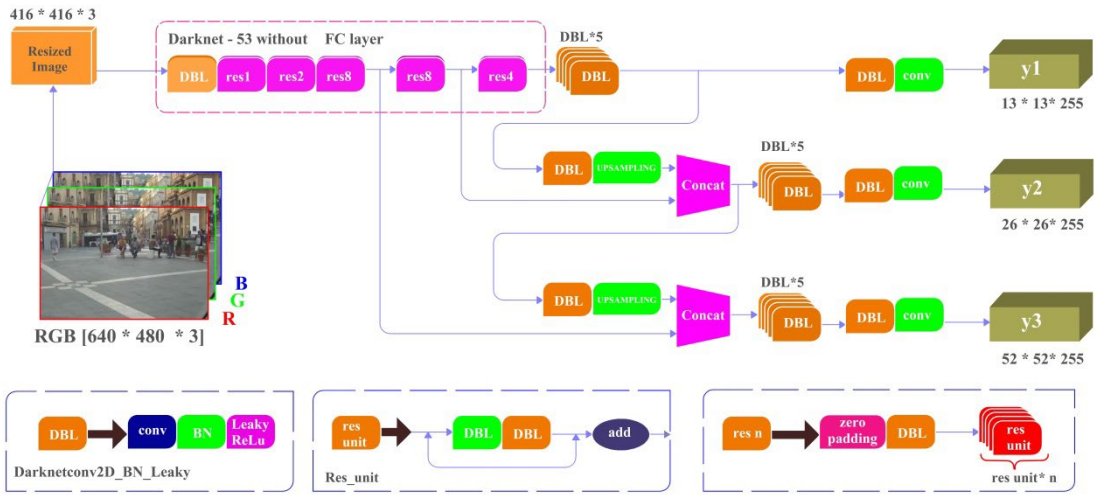

Figure 2. General architecture of YOLOv3 utilized for frontal view human detection

For the detected bounding box correlated with the object of interest the loss function is calculated. In Yolov3 the overall loss function comprises of localization loss (bounding box regressor), cross entropy and confidence loss for classification score, defined as follows (Punn *et al.*, 2021; Redmo & Farhadi, 2018):

$$\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=o}^{B}1_{i,j}^{obj}[(t_x-\hat{t}_x)^2+(t_y-\hat{t}_y)^2+(t_w-\hat{t}_w)^2+(t_h-\hat{t}_h)^2]+\sum_{i=0}^{S^2}\sum_{j=o}^{B}1_{i,j}^{obj}[(-\log(\sigma(t_0)] + \sum_{k=1}^{C}BCE[(\hat{y}_k,\sigma(s_k)]+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=o}^{B}1_{i,j}^{nobj}[(-\log(1-\sigma(t_0)] \quad (1)$$

In which λ_{coord} designates the weight of the coordinate error, S^2 indicates the number of grids in the image, and B is the number of generated bounding boxes per grid. $1_{i,j}^{nobj}=1$ describes that object confines in the j^{th} bounding box in grid *i*, otherwise it is 0.

For more details about the calculation of Loss function in various object detection models, including Yolov3, the interested reader may consult Punn *et al.* (2021).

Generally, the performances of a certain object detector are evaluated by the following metrics:

- Frame per second (FPS) to measure detection speed (number of images processed every second);
- Precision-recall curve (PR curve); the Precision and Recall are calculated by Eq. (2) and Eq. (3):

$$\text{Recall}=\frac{TP}{TP+FN}. \quad (2)$$

$$\text{Precision}=\frac{TP}{TP+FP}. \quad (3)$$

- Accuracy:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}. \quad (4)$$

The symbols TP, TN, FN and FP stand for True Positive, True Negative, False Negative and False Positive respectively.

3. Proposed approach

In this research, a deep learning-based framework is applied for monitoring COVID-19 social distancing with person detection and tracking via Computer Vision techniques (Guerrieri, *et al.*, 2013; Guerrieri *et al.*, 2018) and YOLO v3 algorithm. The social distance is calculated in each instant of time using surveillance video. The proposed method can be applied in real-world conditions as one of the most important remedies for dealing with the escalation of COVID-19 cases.

The essential phases undertaken to constitute a framework for monitoring social distancing are given in this section.

3.1. Camera calibration

Camera calibration aims at establishing two sets of parameters: intrinsic and extrinsic. The Zhang algorithm (Zhang, 2000) was implemented to determine the extrinsic parameters of the system. The calibration process was based on a set of 64 photos of a chessboard (Guerrieri & Parla, 2021). The validation of the calibrated model was obtained through tests concerning the comparison between estimated and measured values of the height of some people.

3.2. Training of the Neural Networks for people detection

The training of the neural networks is obtained by the implementation of the pre-existing dataset of MS COCO (Microsoft Common Objects in COntext) (Lin *et al.*, 2018) with a class label as “Person” along with the annotations. The dataset comprises nearly 270k segmented people and a total of 886k segmented objects (Lin *et al.*, 2018). In this research the dataset was divided into training and testing sets, in 7.5:2.5 ratio.

Figures 3 and 4 show the training process consisting of 2000 iterations. The performance of the proposed training model (cfr. Eq. 1 and 4 and Figure 3 and Figure 4) in human detection is extremely high: as shown in Figure 3 the accuracy reaches values higher than 95%.

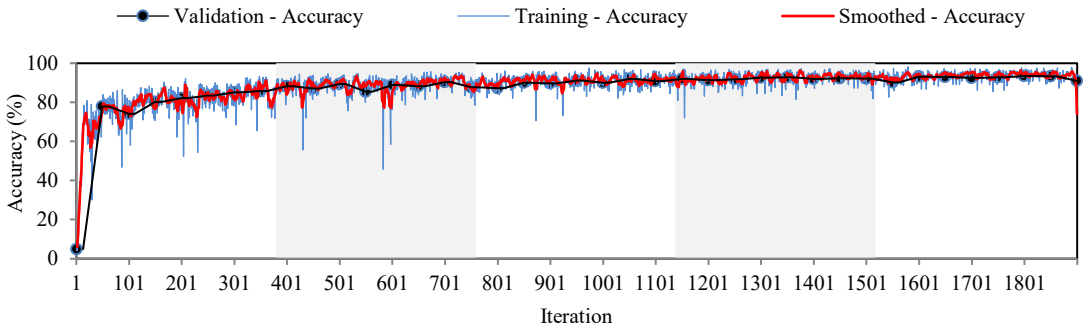

Figure 3. Accuracy evolution related to the number of iterations

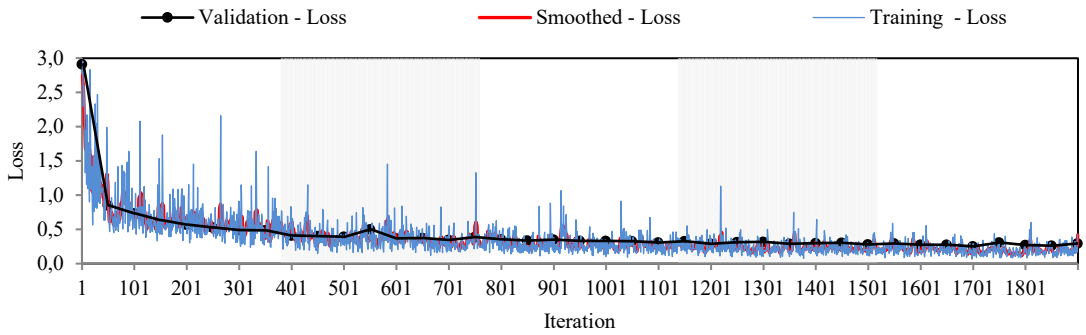

Figure 4. Loss evolution related to the number of iterations

3.3. Top view image transformation of the detected people

The proposed technique allows estimating the distance between each pair of people in a given image of a video and therefore to identify the instantaneous social distance in each instant of time.

The distance considered is that of the people's centroid projected on the ground plane (road pavement surface). The technique for the social distance estimation is based on the inverse perspective mapping (IPM) procedure (Dorj & Lee, 2016). The schematic illustration of the top view image transformation is given in Figure 5. Then, the location of the bounding box for each detected person in the perspective view is transformed into a top-down view.

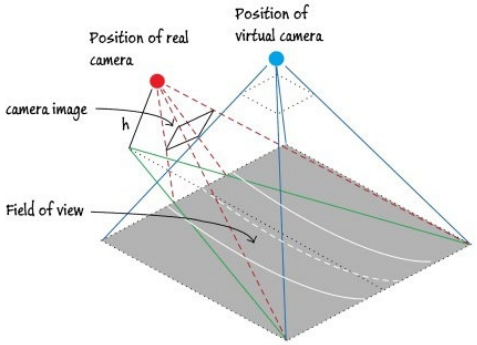

Figure 5. Schematic illustration of the top view image transformation (adapted from Dorj and Lee, 2016)

3.4. People tracking

The process of tracking the projection on the pavement surface of the centroid of the people is based on the use of the linear Kalman filter (Kalman, 1960) that is a recursive predictive filter able to evaluate the state of a dynamic system. The dynamic equation of the linear Kalman filter is (Welch & Bishop, 2006; Niu, 2018):

$$x_{n+1} = A_n x_n + B_n u_n. \quad (15)$$

Considering the error covariance (Welch & Bishop, 2006; Niu, 2018):

$$P_{n+1} = A_n P_n A_n{}^T + Q_n, \quad (16)$$

where x_n is the state value (people coordinates) at step n, A_n is the state transition matrix, u_n is the measurement and the input and at step n. Q_n is the white noise covariance (Niu, 2018). This step is called the "prediction step" since it estimates the n+1 state. Kalman gain value is given by the following relationship:

$$K_n = P_n C^T (C P_n C^T + R_n)^{-1}, \quad (17)$$

where C is the measurement matrix and R is measurement noise.

Actual measurement value at the updated time and error covariance is (Niu, 2018):

$$P_n = (I - K_n H) P_n. \quad (18)$$

In which K_n is the measurement value and H is the mapping matrix from true state to observation.

The filter is able to approximate the trajectory of a person even it overlaps with another object or person in the analysed frames of a video.

4. Experiments, results and discussion

The detection and tracking algorithms proposed in the present research allow estimating the mutual social distance of two or more people who are in the same pedestrian area (e.g., bus stop or railway station, Corriere *et al*., 2013). The proposed algorithm identifies clusters of people (minimum two people) who maintain a social distance (d) lower than the minimum social distance (d_{min}) established for prolonged periods of time. For each detected person the model calculates a circle on the surface of the pavement whose radius is equal to the half of the minimum social distance to be respected ($r = d_{min}/2$). Denoting with (x_1; y_1) the centroid coordinates of the first detected person (Id_1) and with (x_2; y_2) the centroid coordinates of the second detected person (Id_2) (Figure 6) the social distance violation occurs if it results $d < d_{min}$ where the Euclidean distance d can be calculated using Eqs. (19), (20):

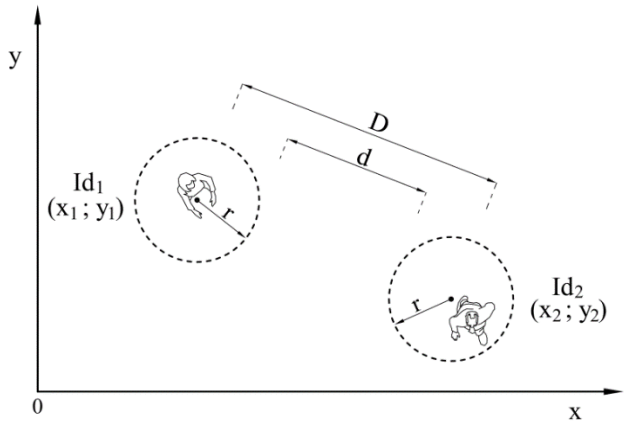

Figure 6. Social distance violation evaluation in function of the Euclidean distance

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - 2r = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - 2\frac{d_{min}}{2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d_{min}. \quad (19)$$

$$\begin{cases} d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d_{min} \geq 0 \rightarrow \text{no social distance violation} \\ d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d_{min} < 0 \rightarrow \text{social distance violation} \end{cases}. \quad (20)$$

Consequently, after setting a threshold distance (i.e. the minimum social distance d_{min}), the social distance can be estimated, and the bounding boxes of the detected persons will change color and can raise an alarm if two or more people come closer than the minimum social distance.

As shown in Figure 7, in the "Bird's Eye View" representation of each frame, a pedestrian is marked with a red circle if the spacing is less than d_{min} or with a blue circle if the spacing is greater than d_{min}.

The following Figures 8 and 9 show the final output of the model concerning a scenario in the historic center of Caltanissetta (Italy).

In particular, in Figures 8 and 9, the algorithm identifies among all the framed persons (moving or stationary) those who violate the minimum social distancing. These violations are identified with red Bounding Boxes, instead the pair of pedestrians whose distance is higher than the minimum social distancing is marked in green. In addition, the proposed method evaluates the residence time of the cluster of pedestrians in the crowded place. For each frame of the video sequence, the algorithm records a matrix that contains the parameters given in Table 1.

In each row of Table 1 (pedestrian under examination), in addition to the coordinates and dimensions of the Box in the image plane (columns x_{min}, y_{min}, L_{box} and H_{box}), are inserted the order numbers (Id) of the pedestrians next to the under examination that violate the minimum safety distance (IdT column). This condition is marked in the "d <d_{min}" column.

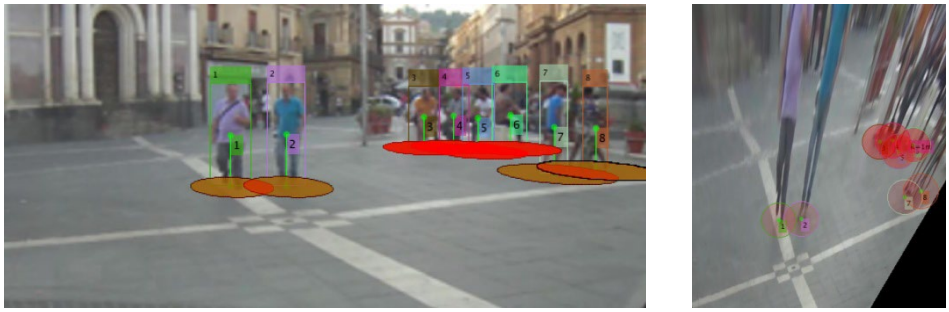

Figure 7. Algorithm output (the red circles denote the pedestrians whose d < d_{min})

In addition, for the group of pedestrians examined, the algorithm records whether the time spent during the condition in which the minimum social distance is not respected has exceeded a certain minimum time value T_{min} established by the analyst. The research assumed the threshold values d_{min} = 2 m and T_{min} = 3 sec (Marcon, 2021). This is because the risk of infection is related both to the distance between people and to the time of exposure. For instance, Figure 9 shows the words "High risk of infection time = 1.04 minutes".

Table 1. Algorithm output

Id	x_{min}	y_{min}	L_{box}	H_{box}	Id in the cluster	d<d_{min}	d($_{camera}$)	T_{min}	True false	Box	T[sec]	Frame
1	210	70	30	75	0	0	0	0	0	0	"---"	1923
2	392	76	33	64	[2, 3]	[1,1]	[20, 91, 20, 34]	1	1	[392 73 203 70]	1,04	1923
3	429	77	30	63	[3, 4]	[1,1]	[20, 34, 20, 38]	1	1	[392 73 203 70]	1,04	1923
4	482	70	35	69	[4, 5]	[1,1]	[20, 38, 19. 59]	1	1	[392 73 203 70]	1,04	1923
5	537	58	27	81	[5, 4]	[1, 1]	[19, 60, 20, 38]	1	1	[392 73 203 70]	1,04	1923
6	619	51	22	89	0	0	0	0	0	0	"---"	1923

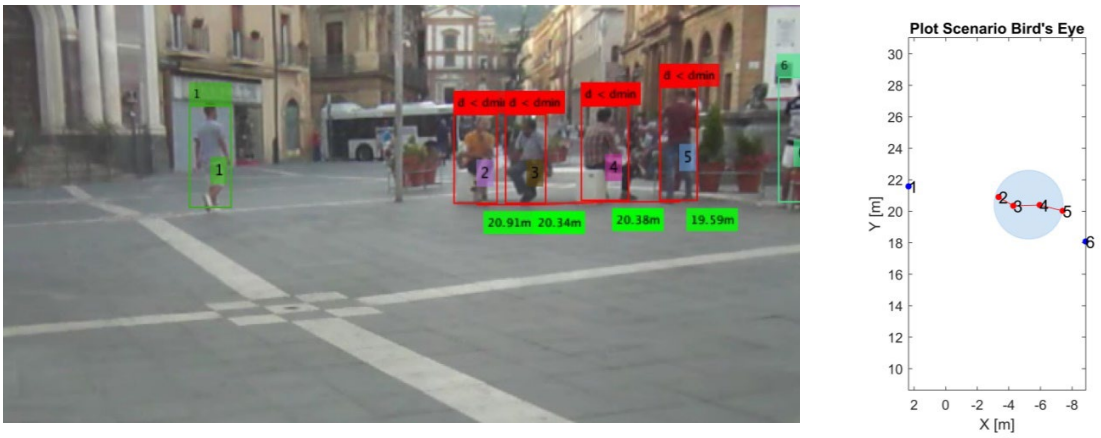

Figure 8. Example of social distance violations (in red the bounding Box of the pedestrians whose d < d_{min})

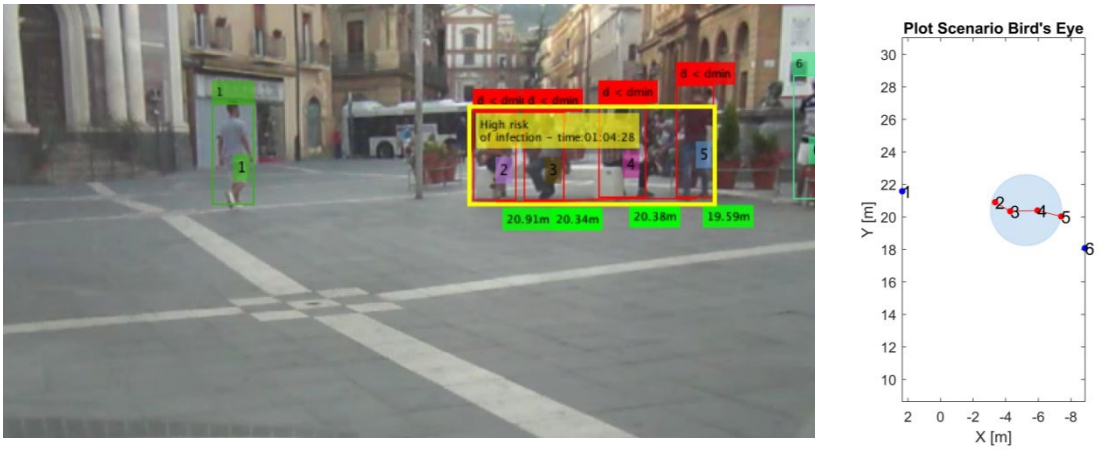

Figure 9. Example of social distance violations (in yellow the bounding Box of a cluster)

Ultimately, the procedure described generates statistics relating to:

- the number of clusters present in the scene at the same instant of time;
- the size of the smallest cluster;
- the size of the largest cluster;
- the average size of the clusters present in the scene;
- the median value of the size of the clusters present in the scene;
- the number of people who are not part of any cluster.

The experimental results of the research revealed that the proposed methodology reliably and in real time recognizes the interactions between people who maintain (or do not maintain) the minimum social distancing required in the event of pandemic or post-pandemic events. Therefore this procedure lends itself to being used for the control of social distancing in public spaces such as those of stops and stations of public transportation systems.

5. Conclusions

The control of the social distances between people plays a fundamental role for limiting the pandemic's damage to the health system and economy. Considering that the control of social distancing for a large population cannot be done manually, this research proposes a method based on deep learning (DL) and Yolov3 model to segregate humans from the background of each image of a surveillance video. The training of the neural networks is obtained by the use of the pre-existing dataset of MS COCO that comprises nearly 270k segmented people. The overall dataset was splitted in two parts: one for model training another for validation of results, in 7.5:2.5 ratio.

In addition, the linear Kalman filter for tracking the humans' motion was implemented. Experiments were carried out considering a case study in the historic center of Caltanissetta (Italy).

The results showed the success of the proposed techniques in obtaining accurate measurements of the distances and positions of persons in real-world coordinates. In fact, the accuracy in people detection reaches values higher than 95%. Starting from the analysis of a video surveillance, the algorithm identifies clusters of people who maintain for prolonged periods of time a social distance (i.e Euclidean distance) lower than the minimum social distance established. The social distancing violations are identified with red Bounding Boxes. In conclusion, the detection algorithm can be applied for alerting people to keep a safe distance from each other when they are in crowded places or in groups (e.g., bus and railway stations, airport terminals, etc.) and hence can provide better control over the transmission of infectious diseases.

Acknowledgments

This research was developed within the Project SAFEMOBILITY-C19 financed by the University of Trento, Internal Call 2020 "Covid 19".

References

1. Ahmed, I., Ahmad, M., Rodrigues, J.J.P.C., Jeon, G., Din, S. (2021) A deep learning-based social distance monitoring framework for COVID-19. *Sustainable Cities and Society*, 65, art. no. 102571.

2. Brunetti, A., Buongiorno, D., Trotta, G. F., & Bevilacqua, V. (2018) *Neurocomputing*, 300, 17–33.
3. Corriere, F., Di Vincenzo, D., Guerrieri, M. (2013) A logic fuzzy model for evaluation of the railway station's practice capacity in safety operating conditions. *Archives of Civil Engineering*, 59(1), 3–19.
4. Dorj, B., Lee, D.J. (2016) A Precise Lane Detection Algorithm Based on Top View Image Transformation and Least-Square Approaches. *Journal of Sensors*, art. no. 4058093.
5. Eksin, C., Paarporn, K., Weitz, J. S. (2019) Systematic biases in disease forecasting–the role of behavior change. *Epidemics*, 27, 96–105.
6. Elgendy, M. (2020) *Deep Learning for Vision Systems; Simon and Schuster*. New York, NY, USA.
7. Gossling, S., Scott, D., Hall, M. (2020) Pandemics, tourism and global change: A rapid assessment of COVID-19. *Journal of Sustainable Tourism,* 1–20.
8. Guerrieri, M., Parla, G. (2021) Deep learning and yolov3 systems for automatic traffic data measurement by moving car observer technique. *Infrastructures*, 6(9), 134.
9. Guerrieri, M., Mauro, R., Parla, G., Tollazzi, T. (2018) Analysis of kinematic parameters and driver behavior at turbo roundabouts. *Journal of Transportation Engineering Part A: Systems*, 144(6), 04018020.
10. Guerrieri, M., Parla, G., Corriere, F. (2013) A new methodology to estimate deformation of longitudinal safety barriers. *ARPN Journal of Engineering and Applied Sciences*, 8(9), 763–769.
11. Harvard Medical School. (2020) Preventing the spread of the coronavirus. Available: https://www.health.harvard.edu/diseases-and-conditions/preventing-the-spread-of-the-coronavirus (Accessed 15 December 2021).
12. Kalman, R. E. (1960) A new approach to linear filtering and predictions problems. *Journal of Basic Engineering*. 82(D), 35-45.
13. Kermack W. O., McKendrick, A. G. (1927) Contributions to the mathematical theory of epidemics–i.
14. Lechgar, H., Bekkar, H.,Rhinane, H. (2019) Detection of cities vehicle fleet using YOLO V2 and aerial images, International Archives of the Photogrammetry. *Remote Sens. Spat. Inf. Sci.-ISPRS Arch*., 42, 121–126.
15. Lin, T.Y. *et al.* (2018) Microsoft COCO: Common Objects in Context.
16. Marcon, Y. (2021) Distance-based clustering of a set of XY coordinates. Available: https://www.mathworks.com/matlabcentral/fileexchange/56150-distance-based-clustering-of-a-set-of-xy-coordinates, MATLAB Central File Exchange. Retrieved September 15.
17. Musselwhite, C., Avineri, E., Susilo, Y. (2020). Editorial JTH 16 –The Coronavirus disease COVID-19 and implications for transport and health. *Journal of Transport and Health,* 16, 100853.
18. Niu, M. (2018) Object Detection and Tracking for Autonomous Driving by MATLAB toolbox. Ohio State University, Available: https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1523978995395614&disposition=inline.
19. Punn, N. S., Sonbhadra, S. K., Agarwal S., Rai, G. (2021) *Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLOv3 and Deepsort techniques*. (arXiv:2005.01385).
20. Redmon, J., Farhadi, A. (2018) *YOLOv3: An Incremental Improvement*. Available online: https://arxiv.org/pdf/1804.02767v1.pdf (accessed on 1 August 2021).
21. Welch, G., Bishop, G. (2006) *An Introduction to the Kalman Filter. Department of Computer Science University of North Carolina at Chapel Hill Chapel Hill*, NC 27599-3175. Available: https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf.
22. Zhang, Z. (2000) A flexible new technique for camera calibration. In IEEE Transactions on Pattern Analysis and Machine Intelligence. *IEEE: Piscataway*, 22, 1330–1334.